STATIC DECISION MODELS FOR QUEUEING SYSTEMS
WITH NON-LINEAR WAITING COSTS

BY
SHALER STIDHAM, JR.

TECHNICAL REPORT NO. 112
September 3, 1968

SUPPORTED BY THE ARMY, NAVY, AIR FORCE AND NASA UNDER
CONTRACT Nonr-225(53)(NR-042-002)
WITH THE OFFICE OF NAVAL RESEARCH

DEPARTMENT OF OPERATIONS RESEARCH
AND
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
STATIC DECISION MODELS FOR QUEUEING SYSTEMS
WITH NON-LINEAR WAITING COSTS*

by
Shaler Stidham, Jr.

TECHNICAL REPORT NO. 112
September 3, 1968

Supported by the Army, Navy, Air Force and NASA under
Contract Nonr-225(53) (NR-042-002)
with the Office of Naval Research

*This research was partially supported by the Office of Naval
Research under Contract Nonr-225(89* (NR-047-061)

Gerald J. Lieberman, Project Director

Reproduction in Whole or in Part is Permitted for
any Purpose of the United States Government

DEPARTMENT OF OPERATIONS RESEARCH
and
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
ACKNOWLEDGEMENTS

My deepest gratitude goes to Professor Frederick S. Hillier, whose careful balance of patience and prodding were in no small way responsible for the completion of this dissertation. I am also indebted to Professor Arthur F. Veinott, Jr., for valuable suggestions which helped to clarify my thinking and improve the proofs of many of the theorems. Special thanks are due to Jerri Rudnick, who did most of the typing and who was willing to sacrifice several evenings and a weekend to it near the end, and especially to my wife, who bore my preoccupation with good grace and good cheer.

The research was supported in part by the Office of Naval Research under Contracts Nonr-225(89) and Nonr-225(53).
# TABLE OF CONTENTS

LIST OF ILLUSTRATIONS vi

CHAPTER

I Introduction 1
  1.1 Review of Optimization of Queueing Systems 1
  1.2 General Stochastic Decision Model for Queueing Systems 5
  1.3 Applications of General Stochastic Decision Model 10
  1.4 Models to be Studied in this Paper 12

II Single-Station Models with Linear Costs 27
  2.1 Introduction and Motivation 27
  2.2 Proof of Theorem 2.1 for Degenerate Service-Time Distribution 37
  2.3 Proof of Theorem 2.1 for Exponential Service-Time Distribution 40
  2.4 Proof of Theorem 2.1 for k-Erlang Service-Time Distribution 44
  2.5 Remarks about Theorem 2.1 50

III Single-Station Models with Non-Linear Service Costs 58
  3.1 Dominance by l-System 59
  3.2 Non-Dominance by l-System 72

IV Single-Station Models with Non-Linear Waiting Costs 78
  4.1 Introduction 78
  4.2 Generalization of $E[L] = \lambda E[W]$ 83
  4.3 Proof of Ordering of Customer Waiting Costs: Transient Case, k-Erlang Service-Time Distribution, Convex Costs 89

V Network Queueing Systems 107
  5.1 Introduction 107
5.2 Proof of Theorem 5.1 for Type II-A Systems 114
5.3 Proof of Theorem 5.1 for Type II-B Systems 122
5.4 Remarks 124

VI
Single-Station Models: Some Generalizations of Operating Policy, Decision Space, and Cost Structure; Summary and Conclusions 126

6.1 Introduction and Review of Previous Chapters 126
6.2 Effect of Queue Discipline 128
6.3 Models where the Service-Cost is Charged only for Server's Busy Time 132
6.4: Models where each Server Works at a Different Mean Rate 135
6.5 Models with State-Dependent Service Rate 137
6.6 Models with an Upper Bound on the Mean Service Rate 143
6.7 Team Model 154
6.8 Summary and Conclusions 163

Appendix Numerical Analysis of M|M|c System 166

References 203
# LIST OF ILLUSTRATIONS

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Test for Step Functions</td>
<td>73</td>
</tr>
<tr>
<td>2</td>
<td>Graph of $E[L(c,\rho)]$</td>
<td>76</td>
</tr>
<tr>
<td>3</td>
<td>Type II-A Network: Series System</td>
<td>109</td>
</tr>
<tr>
<td>4</td>
<td>Type II-B Network: Partially Ordered System</td>
<td>109</td>
</tr>
<tr>
<td>5</td>
<td>Optimal $\hat{\mu}_{c}$ as Function of $K$, $c = 1,2,3,4,5$ ($A = 1$)</td>
<td>151</td>
</tr>
<tr>
<td>6</td>
<td>Optimal $\hat{\mu}_{c}$ as Function of $A$, $c = 1,2,3,4,5$ ($K = 1$)</td>
<td>152</td>
</tr>
<tr>
<td>7</td>
<td>$E[C(c,\bar{\mu})]$ as Function of $c = 2,3,4,5,6,7$ ($A = 1$, $K = 0.25$, $\bar{\mu} = 0.75$)</td>
<td>155</td>
</tr>
</tbody>
</table>
NONTECHNICAL SUMMARY

Some models for the optimal design of queueing systems are presented. The decision variables are measures of the service capacity of the service mechanism; they must be determined and fixed before the system begins to operate and cannot be dynamically varied thereafter. In most of the models studied, the decision variables are the number of servers \( c \) and the mean rate \( \mu \) at which each serves. The objective function is the steady-state total expected cost rate of operating the system, which is assumed to be the sum of a cost of operating the service mechanism and a cost due to customers waiting in the system.

It is shown that a single-server system is optimal for a wide class of arrival processes and service-time distributions, a wide variety of service and waiting cost functions, and a wide variety of system structures and operating policies.

The optimality of the single-server system is first demonstrated for single-station models with general arrival process and degenerate, exponential, or Erlang service-time distribution, where the service-cost rate is proportional to both \( c \) and \( \mu \) and the waiting-cost rate is proportional to the number of customers in the system. This result is proved by showing that the single-server system stochastically minimizes the number of customers in the system among all systems with equal total service capacity \( c \mu \).
Several generalizations of this model are presented. In most cases the single-server system remains optimal for the generalized model. The models considered are: (1) single-station models with service-cost rate proportional to \(c\) and concave, quasi-homogeneous, or subadditive in \(\mu\); (2) single-station models with customer waiting costs which are monotone increasing in the time spent by a customer in the system; (3) multi-station models in which the service times at each station are exponentially distributed and the stations are arranged in a series or as the nodes of a partially ordered graph; (4) single-station models (a) with non-FCFS queue disciplines, (b) with service-cost rate which is charged only when the server is busy, (c) where each server may work at a different rate, (d) where the service mechanism acts as a variable-rate Poisson output process with parameter depending on the number of customers in the system, (e) with upper bounds on the mean service rate, and (f) where the number of servers is fixed and the decisions are the mean rate at which they should operate and whether they should work separately or as a team.
CHAPTER I

Introduction

1.1 Review of Optimization of Queueing Systems

Most of the literature of queueing theory has been concerned with mathematical description of queueing systems with predefined structures and operating policies. Until recently very little research had been done on the optimization of queueing systems.

This situation is perplexing when one compares queueing theory with inventory theory, where the emphasis has always been on optimization. Both theories describe what might be called stochastic input-output systems: systems where demands (customers) arrive in some probabilistic fashion, request and perhaps wait for service, are satisfied (or turned away) in some probabilistic fashion, and finally depart from the system (or return to demand more service). As Maxwell [7] has observed, the differences are not in the systems modeled but in the assumptions which are made about the structure and operating policy of the system (e.g., whether it is observed in continuous or discrete time, whether production (service) can occur before demands (customers) arrive, whether backlogging of demands (waiting of customers) is allowed, whether demand occurs in discrete units or continuously variable amounts).

The ordinary one-server, one-queue queueing model can be viewed as an (albeit unorthodox) inventory model in which demands occur at intervals whose length is a random variable, each demand is for one unit, production only starts when a demand arrives, one unit is produced at a time, the time required to produce a unit is a random variable,
and production stops as soon as all demand currently in the system is filled. Thus we have a production-inventory model in which inventory is always negative, i.e., we are always backlogging. Similarly, most inventory systems can be viewed as queueing systems, if similar latitude is allowed in the definition of a queueing system, such as permitting customers to be serviced before they arrive.

Although such identifications of the standard queueing system with an unstandard inventory system, and vice versa, can be made, they may not be too useful. But, between the two extreme cases of the "pure queueing" and the "pure inventory" system are many systems which can be usefully viewed as either queueing systems or inventory systems.*

The correspondence suggested by Maxwell helps us in our study of the optimization of queueing systems. For, when we view queueing and inventory systems as two examples of stochastic input-output systems, we see that there are quite often two natural types of cost to be balanced against each other in both cases: the cost of operating the service mechanism and the cost of delaying the satisfaction of demand. If $x$ represents some measure of the capacity or efficiency of the service mechanism, it is also natural in both cases for the (expected)

* Another, more commonly cited, equivalence between queueing and inventory systems is obtained by viewing items arriving from production and going into inventory as customers arriving and waiting for service; service occurs when a demand depletes the inventory. This approach is useful when it comes to exploiting similarities in the mathematical analysis of the stochastic processes underlying the two systems (cf. Smith [31] and Prabhu [25]), but Maxwell's approach is more instructive in pointing out the physical and purposeful similarities between them.
service cost to be a monotone increasing function of \( x \) and the (expected) waiting (delay) cost to be a monotone decreasing function of \( x \).

In this paper we study a particular problem of this type for queueing systems. This sort of problem — choosing values of variables, or a policy for setting values, to minimize the sum of an increasing and a decreasing cost — is classical in operations research. It is basic to most inventory models, but until recently has received only casual attention in queueing research. In recent efforts, however, more and more attention is being given to this problem. *

Typically, in these optimization efforts, the author establishes the decision variables of the service mechanism (i.e., measures of service efficiency: e.g., the number of servers and/or the mean service rate of each) and imposes a reasonable cost structure on the system, where by "reasonable" we mean at least that the service cost is monotone increasing in the service-efficiency measures and the waiting cost is monotone decreasing.

* One of the earliest optimization results is in Morse [23], and cited in Hillier [12], [13], and Hillier and Lieberman [14], all of which are good introductions to optimization of queueing systems. Other relevant references are Conway, Maxwell, Miller [7], Yadin and Naor [38], [39], Dogrusoz [8], Bowman and Fetter [5], Heyman [11], Moder and Phillips [22], and Miller [21].

It is interesting to note in passing the number of times that optimization models have yielded formulas for the optimal value of the decision variable which are similar to the classical EOQ formula in inventory. (Cf. Morse [23], Yadin and Naor [39], Heyman [11], and Conway, Maxwell, and Miller [7].)
As in inventory theory, we are really concerned with determining the optimal policy (i.e., rule which sets the values of the decision variables for each state that the system can be in). But, as in inventory theory, most of the early efforts in optimization of queueing systems have begged this basic question and instead solved a simpler problem: what are the parameters of an optimal policy within the class of policies of a certain simple form? (Cf. [8], [38], [39].) Thus, for example, Yadin and Naor [39] solved for the optimal value of the queue size at which to turn on a single server, assuming that the form of the policy is to turn on the server when the queue size reaches a certain figure and turn him off when the system is empty. Such efforts have paralleled the stage in the development of inventory theory when researchers were restricting their attention to policies of the base-stock or $(s,S)$ type and solving for the optimal values of the parameters.

More recent efforts, however, (cf. [11], [21]) have involved attempts to discover the form of the optimal policy (cf. proofs of the optimality of $(s,S)$ policies in inventory theory). For example, Heyman [11] showed that a policy of the form assumed by Yadin and Naor is optimal for their problem under fairly general conditions.

When we concern ourselves with discovering the form of the optimal policy, we get into the theory of **stochastic decision processes** (sequential control processes). As in inventory theory, Markovian properties are exploited wherever possible. But in general, the costs, decisions, and evolution of the system at any time may depend on the complete history of the system up to that time, rather than just on the
current state.

We give below a general model of a queueing system as a stochastic decision process; most of the optimization models which have been cited can be viewed as special cases of this formulation.

1.2 **General Stochastic Decision Model for Queueing Systems**

1.2.1 Definition of system

In general we consider a queueing system to be defined by two stochastic processes: an **input process**, and an **output process**.

The input process is defined by a sequence of random variables, \( \{ A_i, i = 1, 2, \ldots \} \), where \( 0 \leq A_1 \leq A_2 \leq \ldots \), and \( A_i \) the arrival time of the \( i \)th customer to enter the system. The output process is defined by the sequence of random variables, \( \{ D_i, i = 1, 2, \ldots \} \), where \( D_i \geq A_i \) for all \( i \), and \( D_i \) the departure time of the \( i \)th customer to enter the system. Note that the sequence \( \{ D_1, D_2, \ldots \} \) need not be in monotone increasing order, since the service mechanism (which we have not yet explicitly described) may not eject customers in the order in which they arrive.

This definition is essentially equivalent to those of Little [20] and Jewell [15]. They define the system in terms of an input process and a waiting-time-in-system process, \( (w_1, w_2, \ldots) \). In our notation, \( D_i = A_i + w_i, i = 1, 2, \ldots \).

For generality, we specify another sequence of random variables associated with the arrival process: \( \{ P_i, i = 1, 2, \ldots \} \), where \( P_i \) the information parameter for the \( i \)th customer, and may be used to
indicate the priority class, predicted service time, or some cost information about the customer.

1.2.2 Decision variables

The output process (and perhaps the input process) depend stochastically on the structure and operating policy of the service mechanism, which may be viewed as a black box which accepts items from the input process, \( \{A_i\} \), and ejects them according to the output process \( \{D_i\} \).

The service mechanism may consist of one or several servers, fed by one or several queues, arranged in series or in parallel or in a complex network. It may have complex rules for determining whether or not to accept a customer, which customer to serve next, which route for a customer to follow through the network.

In the general model, we do not explicitly account for these differences in structure or operating policy. Instead we simply postulate that there are two types of decision parameters: design parameters and action parameters. Design parameters (designated \( \delta \)) relate to the static structure of the system, e.g., how many servers to provide, what arrangement (series, parallel, network) to place them in. Action parameters (designated \( \alpha(t) \)) describe the dynamic operating policy of the system and may be varied through time. They may represent such things as the mean rate at which a server is operating at a certain time, how many servers are turned on, which classes of customers to accept into the system.

The design parameters \( \delta \) must belong to a feasible set \( \Delta \), and
the action parameters $\alpha(t)$ must belong to a feasible set $A(t)$, for all $t \geq 0$.

1.2.3 Evolution of the system

The effect of the service-mechanism black box on the input and output processes, $\{A_i\}$ and $\{D_i\}$, is determined, as a function of the design and action parameters, by certain evolutionary rules. These are best defined by birth and death transition intensity vectors. Because the system may not be Markovian (at least under a simple state description such as the number in the system), these intensities will generally depend on the history of the system.

For some systems a description of the history sufficiently detailed to specify unambiguously the transition intensities is quite complex; for example, in network systems it might have to include the time at which each customer entered and left each facility in the network. However, for our analysis, it will not be necessary to formulate the evolutionary rules precisely for each system. Therefore we will give a definition for the history of the system which is sufficient for specifying the evolution of most of our models and for specifying the cost in all our models (see Section 1.2.4).

We define the history of the system at time $t$ as

$$H(t) = \{A_i, P_i | A_i \leq t\}, \{D_i | D_i \leq t\}, \alpha(\tau), 0 \leq \tau < t\}.$$

Then there are two transition intensity vectors, one governing inputs and the other governing outputs, at time $t$: 

7
\( q_A(p|t, \delta, \alpha, h) = \text{Prob} \{ \text{an arrival of class } p \text{ occurs in} \) \\
(t, t+dt) | \text{design parameter } \delta, \alpha(t) = \alpha, H(t) = h \}, \) \\
for all \( p \in \text{set of possible values of information} \) \\
parameter; \( q_D(i|t, \delta, \alpha, h) = \text{Prob} \{ i^{\text{th}} \text{arrival departs in} \) \\
(t, t+dt) | \text{design parameter } \delta, \alpha(t) = \alpha, H(t) = h \} \) for \( i = 1, 2, \ldots \). \\

1.2.4 Cost of operating system \\
We assume that there is a cost rate, at any time \( t \), of operating 
the system, which depends on the history at time \( t \), the value of the 
design parameter (fixed throughout time), and the value chosen for the 
action parameter at \( t \). 

\( C(t|\delta, \alpha, h) = \text{operating cost rate at time } t, \text{ given design} \) 
parameter, \( \delta, \alpha(t) = \alpha, H(t) = h. \) 

Note that, with the history defined as in Section 1.2.3, the cost 
rate, and in particular its waiting-cost component, depends only on the 
times customers enter and leave the system as a whole. This rules out 
waiting costs, for example, which depend on the number of customers or 
the length of time a customer spends in the queue (exclusive of service). 
In fact, all the waiting costs studied in this paper will be functions 
of the number of customers or the waiting time of a customer in the 
system. 

We accommodate set-up costs (cf. Heyman [11]) by "spikes" added to
the cost rate at the point, $t$, where the set-up cost is incurred.

1.2.5 Optimization

For the purposes of optimally controlling the queueing system, we define a policy for designing and operating the system as follows:

$$\pi = (\delta_\pi; f_\pi(t,h), \ t \geq 0, \ \text{all possible histories, } h, \ \text{at time } t) ,$$

where $\delta_\pi$ specifies a value of the design parameter, and $f_\pi(t,h)$ specifies an action $\alpha(t)$ to be taken at time $t$, if $H(t) = h$, for every $t \geq 0$ and all possible histories, $h$, at time $t$.

Given any $\pi$ we have a stochastic process without decisions. Let $C_t(\pi)$ be the cost rate of this process at time $t$ ($C_t(\pi)$ is a random variable). Then our objective function, to be minimized by proper choice of $\pi$, is expected long-run average cost:

$$(1) \quad \lim_{T \to \infty} \frac{1}{T} \int_0^T E[C_t(\pi)] \, dt .$$

Suppose that we can prove that an optimal policy exists in the class of all $\pi$ such that $C_t(\pi)$ approaches a limit $C(\pi)$ in distribution as $t \to \infty$. In this case $C(\pi)$ is the steady-state cost rate of policy $\pi$ and minimizing (1) is equivalent to minimizing

$$(2) \quad E[C(\pi)] .$$
1.3 Applications of General Stochastic Decision Model

In the degree of generality presented, the stochastic decision model may not be of much use in solving problems. In particular, the dependence of (i) the arrival process, (ii) the evolution of the system, or (iii) the operating cost rate, on the history of the system makes explicit calculation and optimization of the expected long-run average cost difficult. However, there are many practical applications where this dependence is required in at least one of the three cases if the model is to be realistic.

As an example of (i), consider closed queueing systems, where departing customers return later, or systems where some arrivals balk or where arrivals may be either accepted or rejected by the service mechanism. As an example of (ii), consider systems where the arrival process is not Poisson or the service-time distributions are not exponential. As an example of (iii), consider systems where the system waiting-cost rate depends, not only on the number of customers in the system, but also on how long each has been in the system.

One of the difficulties of using the complete history of the system as a state description is that, although the arrivals, evolution, and cost rate are Markovian with such a state space, the state description is not homogeneous in time. Having a homogeneous state description is convenient because it allows for the optimality of stationary policies for infinite-horizon problems (cf. Blackwell [4], Miller [21], Heyman [11]).

If the state space is homogeneous and it can be shown that there are optimal policies among stationary policies of a certain form (i.e.,
defined by certain parameters) for which steady state exists, then the optimal values of the parameters of the optimal policy can be found by minimizing an expression for the steady-state expected cost rate as a function of the parameters. This is what is done in the analysis of \((s,S)\) policies in inventory theory and in Heyman's analysis of queueing systems where the server can be turned on and off.

There are several ways of obtaining a homogeneous state description while maintaining the Markovian character of the arrivals, evolution, and cost rate in the face of (i), (ii), (iii). One way is to assume the difficulties of (i), (ii), and (iii) do not exist. That is, we employ a simple state description, such as the number of customers in the system or the number of busy servers, and then we assume that the arrival process, evolution, and cost rate are Markovian with respect to this state description. This usually amounts to assuming that the arrivals are from one or more Poisson processes and that the service-time distributions are exponential. Thus we have a continuous-time Markovian decision process. This is the approach of Miller [21].

A second approach is to enlarge the state space slightly (beyond a simple description such as the number in the system), enough to regain the Markovian properties but not enough to lose the homogeneity of the state description. This is the augmentation method described by Kendall [17], also known as the method of supplementary variables. An example where this approach is applicable is the case where the evolution of the departure process and/or the waiting-cost rate depend not only on the number of customers in the system but also on how long each has been in the system. In this case an appropriate state description is the vector
(m; T_1, T_2, ..., T_m), where m = the number in the system and T_i = the elapsed time in the system of the i^{th} customer, i = 1, 2, ..., m. For further discussion of this approach see Chapter IV, Section 4.1.

A third approach is to restrict the times at which actions may be taken to a set of regeneration points, an imbedded Markov chain (cf. Kendall [17]). The imbedded-chain technique for analysis of queueing systems is a contraction method, in the terminology of Kendall. In this case the theory of Markov renewal decision processes becomes relevant (cf. Jewell [15a]) and the key renewal theorem (cf. Smith [32]) can be applied to the sequence of busy cycles of the queueing process, considered as a renewal process, to derive the steady-state expected cost rate. This is the approach of Heyman [11].

1.4 Models to be Studied in this Paper

In this paper we will be concerned with special cases of the general stochastic decision model in which (i) the decision variables δ and α(t) are measures of the capacity (or available efficiency) of the service mechanism, and (ii) the operating-cost rate of the system is the sum of a service-cost rate and a waiting-cost rate:

\[ C_t(\pi) = C_{s,t}(\pi) + C_{w,t}(\pi) \] .

The design parameter δ is interpreted as the amount of capacity (e.g., number of servers, mean service rate of each) built into the service mechanism. It is the maximum capacity that can be made available for use at any time \( t \geq 0 \). We interpret α(t) as the amount of capacity
actually made available at time $t$. Thus $\alpha(t) \leq \delta$ for all $t \geq 0$.*

The model of Heyman [11] fits this pattern, with the number of servers as the measure of capacity. In his model $\delta$ is fixed at 1 (i.e., the maximum number of available servers is 1) and $\alpha(t)$ may be either 0 or 1 (i.e., the server may be either on (available) or off (not available) at time $t$). Thus his problem is one in which the design of the system is fixed and the concern is with choosing an optimal policy for operating the system.

The models which will be studied in this paper, on the other hand, are all characterized by the property that the question of optimal design is the only relevant one. That is, the search for optimal policies can be restricted to those in which the available capacity is fixed at its maximum value at all times, i.e., $\alpha(t) = \delta$, for all $t \geq 0$. There are at least two cases in which the model structure justifies such a restriction: (i) the set of feasible actions $A(t)$ consists of the single point $\delta$, for all $t \geq 0$, or (ii) the service-cost component of the cost rate $C_t(\pi)$ of a policy $\pi$ is independent of $f_t(t,h)$, the action prescribed by $\pi$ at time $t$ given history $h$, for all $t \geq 0$ and all possible values of $h$. In the latter case, if the expected waiting-cost rate at time $t$ is monotone decreasing in $\alpha(t)$, then it is clear that one may as well have the maximum capacity.

* Note that the amount of capacity actually in use at time $t$ may be less than $\alpha(t)$. For example, if the measure of capacity is the number of servers, there may be $n$ servers "turned on" (available) at time $t$, but only $m < n$ customers in the system and hence only $m$ servers actually busy (in use) at time $t$.  

13
available at all times, even when it is not all being used (e.g., when there are idle servers, for the case where the measure of capacity is the number of servers). There is nothing to be gained by dynamically varying the capacity in such cases. With one exception (see Section 6.3), the models considered in this paper satisfy criterion (ii).

Whichever the justification, however, for all our models we assume that the operating action is fixed: \( \alpha(t) = \delta \). Thus the models we consider are static decision models, in contrast to the dynamic decision models discussed in the previous section.

In light of this discussion, the following assumptions will be made about all the models which we study:

**ASSUMPTION I:** An optimal policy for designing and operating the system exists in the class of policies with constant action, \( \alpha(t) = \delta, \ t \geq 0 \).

Henceforth we will consider only the decision variable \( \delta \), and the operating-cost rate will be expressed as a function of \( \delta \): \( C_t(\delta) \).

**ASSUMPTION II:** The operating-cost rate at time \( t \), as a function of the decision variable \( \delta \), has the form:

\[
C_t(\delta) = C_{s,t}(\delta) + C_{w,t}(\delta),
\]

where \( C_{s,t}(\delta) \) is the service-cost rate and \( C_{w,t}(\delta) \) is the waiting-cost rate. Moreover, \( E[C_{s,t}(\delta)] \) is a monotone increasing function of \( \delta \), and \( E[C_{w,t}(\delta)] \) is a monotone decreasing function of \( \delta \).

Because we can restrict our attention to policies with constant
action \( \alpha(t) = \delta \), we know that, for any homogeneous state description, an optimal policy exists in the class of policies which are stationary with respect to that state description. We will also make the following assumption:

**ASSUMPTION III:** An optimal value of \( \delta \) exists in the set of all \( \delta \in \Delta \) such that a steady-state distribution of \( C_t(\delta) \) exists.

Under these assumptions, we know from (2) that the objective function takes the form:

\[
E[C(\delta)] = E[C_s(\delta)] + E[C_w(\delta)],
\]

where \( C(\delta) \equiv \text{steady-state operating-cost rate of the system}, \)
\( C_s(\delta) \equiv \text{steady-state service-cost rate of the system}, \)
\( C_w(\delta) \equiv \text{steady-state waiting-cost rate of the system}. \)

The objective is to minimize \( E[C(\delta)] \) over \( \delta \in \Delta \).

If the design parameter \( \delta \) is the pair \((c, \mu)\), where \( c \equiv \text{the number of servers}, \) and \( \mu \equiv \text{the mean rate at which each serves}, \) then our cost model is a slight generalization of the one considered by Hillier [12, p. 127]. Most of the cost models discussed in this paper will be of this type. An example of a practical application where this type of cost model is appropriate is the selection of the type and number of machines to handle jobs in a job shop. One may have the option in such a case of selecting one fast machine or several slow ones. The possible values of \( c \) are thus \( c = 1, 2, \ldots \); at one extreme, \( \mu \) may be restricted to a small finite set of values or (at the other
extreme) it may be allowed to assume any positive real value, at least as a good first approximation. Another example is where the server is actually a team of men and one must select the optimal number of teams and the optimal team size. Here the mean service rate of a server (team) may be an integer multiple of the (fixed) rate of each man; hence $\mu$ may be restricted to a discrete set of values. For further discussion of applications, see Hillier [13].

The following four assumptions will also be made about all the systems considered:

ASSUMPTION IV: The arrival times $\{A_i, i = 1, 2, \ldots\}$ to the system are independent of the structure, history, or policy of the system.

In particular, this assumption rules out closed queueing systems (e.g., machine-interference systems), systems with a finite source population, and systems with a state-dependent arrival rate.

ASSUMPTION V: All arrivals which do not enter service immediately join a single queue in front of the service mechanism. There is no upper bound on the size of the queue.

This assumption rules out balking and limited-waiting-room systems.

ASSUMPTION VI: The queue discipline is strict. That is, (i) an arrival enters service immediately if there is capacity available in the service mechanism and otherwise enters the queue, (ii) all arrivals which enter the queue stay in the queue until served, (iii) once a customer enters service, he departs from the system when and only when his service is completed, (iii) when a server becomes free, a customer is removed
immediately from the queue and placed in service at that server.
(Cf. Morse [23], Takacs [34], Kleinrock [18].)

In particular, this assumption rules out reneging and preemption
of the server by customers.

**ASSUMPTION VII:** The service-time distribution does not depend on the
class of the customer.

The following notation and definitions will be used, in addition
to the quantities already defined:

\[ L_t(\delta) = \text{number of customers in the system (in queue and in service) at time } t, \text{ given } \delta; \]
\[ L(\delta) = \text{steady-state number in the system, given } \delta; \]
\[ W_n(\delta) = \text{waiting time in the system (in queue and in service) of the } n^{\text{th}} \text{ arrival, given } \delta; \]
\[ W(\delta) = \text{steady-state waiting time in the system of an arbitrary arrival, given } \delta; \]
\[ W_{q,n}(\delta) = \text{waiting time in the queue of the } n^{\text{th}} \text{ arrival, given } \delta; \]
\[ W_q(\delta) = \text{steady-state waiting time in the queue of an arbitrary customer}; \]
\[ \lambda = \text{mean arrival rate (when it exists)}. \]

**DEFINITION:** A single-station queueing system is one in which the
collapsed time between a customer's departure from the queue and his
departure from the system (i.e., the time he spends in the service
mechanism) is described by a single random variable. A network queueing
system is one in which a customer's time in the service mechanism is
divided into time spent at several facilities in succession, including (possibly) time spent at internal queues between the facilities. (The classification may be quite arbitrary and depend on how detailed a model is desired.)

Where applicable, the standard notation for single-station queueing systems will be employed: A/S/n, where the symbol A specifies the type arrival process, S specifies the type of service-time distribution, and n specifies the number of servers. The possible symbols in the A-position are: G, for a completely general stochastic process; GI, for a renewal process; E_k, for a renewal process in which the interarrival times have k-Erlang distribution (i.e., density function is given by

\[ f(t) = \frac{(k\lambda)^k t^{k-1}}{(k-1)!} e^{-k\lambda t} , \]

where \( \lambda = \text{mean arrival rate} \); M, or E_1, for a renewal process in which the interarrival times have exponential distribution; D, for the case where the interarrival times are constant. The possible symbols for the S-position are: G, for a general service-time distribution; E_k, for a k-Erlang service-time distribution; M, for an exponential service-time distribution; D, for a degenerate service-time distribution (constant service time). The system given central consideration in this paper is G/M/c: a queueing system with general arrival process and c servers with mutually independent, identical exponential service-time distributions.
DEFINITION: A set $S$ of non-negative real numbers is called $n$-complete, where $n$ is a positive integer, if for all $x \geq 0$,

$$x \in S \implies i x \in S, \ i = 1, 2, \ldots, n.$$ 

$S$ is called $\omega$-complete if $S$ is $n$-complete for $n = 1, 2, \ldots$.

DEFINITION: Let $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_n)$ be random vectors. We say $X$ is stochastically smaller than $Y$ (denoted $X \preceq Y$) if, for every $t = (t_1, \ldots, t_n)$,

$$\text{Prob}[X > t] \leq \text{Prob}[Y > t].$$

(Cf. Lehman [19], Veinott [35], [36], Bessler and Veinott [2].)

The following properties of stochastic ordering will be used frequently in this paper (for proofs, see Veinott [36]):

Property 1°: If $X \preceq Y$, then $E[X] \leq E[Y]$.

Property 2°: Let $n = 1$. Then $X \preceq Y$ iff $h(X) \preceq h(Y)$, for every monotone increasing Borel function $h$.

The following lemma will be used repeatedly, and so is proved here.

LEMMA 1.1: Let $Z$ be a non-negative real-valued random variable and let $P(t) = \text{Prob}[Z > t]$, all $t \geq 0$. Suppose there exists a function $\eta(t)$ satisfying

$$(4) \quad \text{Prob}[Z \leq t + \Delta t | Z > t] = \eta(t) \Delta t + o(\Delta t),$$

for all $t \geq 0$, $\Delta t \geq 0$. 

19
Then

\( P(t) = \exp(-\int_0^t \eta(\tau)d\tau), \ t \geq 0. \)

**Proof.**

First observe that

\[
P(t) - P(t+\Delta t) = \text{Prob}\{t < Z \leq t + \Delta t\}
= \text{Prob}\{Z > t\} \text{Prob}\{Z \leq t + \Delta t | Z > t\}
= P(t) \text{Prob}\{Z \leq t + \Delta t | Z > t\}.
\]

Therefore, by (4),

\[
P(t) - P(t+\Delta t) = P(t)\eta(t)\Delta t + o(\Delta t);
\]

therefore, dividing by \( \Delta t \) and letting \( \Delta t \to 0 \),

\[
P'(t) = -\eta(t)P(t).
\]

The unique solution to this differential equation, with initial condition \( P(0) = 1 \), is

\[
P(t) = \exp[-\int_0^t \eta(\tau)d\tau].
\]

Q.E.D.

We may interpret \( Z \) as the time until a certain event takes place, and \( \eta(t) \) as the instantaneous rate of occurrence of the event at time \( t \).
In the succeeding chapters of this paper we will be dealing with many different queueing systems of the type defined by Assumptions I-VII, where the objective function is a special case of (5). It will be convenient to have a concise terminology for the systems which we will be considering, since they often differ from each other in rather minute respects. Therefore we classify the systems into types. In all cases, the arrival process is defined by the sequence of arrival times \( A = (A_1, A_2, \ldots) \).

**TYPE I-A:** A single-station queueing system with \( c \) servers whose service times are mutually independent with mean rate \( \mu \) and cumulative distribution function (c.d.f.) \( F(\cdot) \). The queue discipline is first-come, first-served (FCFS). The decision variables are \( c \) and \( \mu \), and the feasible sets for each are:

\[
\begin{align*}
c & \leq \bar{c}, \quad \bar{c} \text{ a positive integer or } \infty; \\
\mu & \in \mathbb{R}, \quad \mathbb{R} \text{ a } \bar{c}\text{-complete set},
\end{align*}
\]

**TYPE I-B:** A single-station queueing system with \( c \) servers whose service times are mutually independent with mean rate \( \mu \) and c.d.f. \( F(\cdot) \). The queue discipline is strict, but not necessarily FCFS. The decision variables are \( c \) and \( \mu \), and the feasible sets for each are:

\[
\begin{align*}
c & \leq \bar{c}, \quad \bar{c} \text{ a positive integer or } \infty; \\
\mu & \in \mathbb{R}, \quad \mathbb{R} \text{ a } \bar{c}\text{-complete set},
\end{align*}
\]

**TYPE I-C:** A single-station queueing system with \( c \) servers whose service times are mutually independent. The \( i \)th server serves all
customers according to a service-time distribution with mean rate $\mu_i$ and c.d.f. $F_i(\cdot)$, where all the $F_i(\cdot)$ have the same form, $i = 1, 2, \ldots, c$. The queue discipline is FCFS. The decision variables are $c$ and $\mu_1, \mu_2, \ldots, \mu_c$, and the feasible sets for each are:

$$c = 1, 2, \ldots; \quad 0 < \mu_i < \infty, \quad i = 1, 2, \ldots, c.$$

**TYPE I-D**: A single-station queueing system whose service mechanism is a Poisson output process with variable rate $\mu_n$, depending on $n$, the number of customers in the system, $n = 1, 2, \ldots$. The decision variables are $\mu_1, \mu_2, \ldots$, and the feasible sets have the form:

$$0 < \mu_i < \infty, \quad i = 1, 2, \ldots.$$

**TYPE I-E**: A single-station queueing system with $c$ servers whose service times are mutually independent with mean rate $\mu$ and c.d.f. $F(\cdot)$. The queue discipline is FCFS. The decision variables are $c$ and $\mu$ and the feasible sets are of the form:

$$c = 1, 2, \ldots; \quad \mu \leq \bar{\mu}.$$

**TYPE II-A**: A queueing system consisting of $r$ stations, $i = 1, 2, \ldots, r$, arranged in series. The output of station $i$ is the input to station $i+1$, $i = 1, 2, \ldots, r-1$, and there is no limit on the size of the queues between stations. The input to station 1 is the input to the system. The output from station $r$ is the output from the system. At station $i$ there are $c_i$ parallel servers whose service times are mutually
independent and independent from the service times at other stations. Each service time at station $i$ is distributed with mean rate $\mu_i$ and c.d.f. $F_i(\cdot)$. The queue discipline at each station is FCFS. The decision variables are $c_1, \ldots, c_r$ and $\mu_1, \ldots, \mu_r$, and the feasible sets for each are:

$$c_i = 1, 2, \ldots; \quad 0 < \mu_i < \infty; \quad i = 1, 2, \ldots, r.$$ 

**TYPE II-B:** A network of $r$ stations which is the graph of a partially ordered finite set. That is, (i) the $r$ stations are the nodes of the graph, (ii) a directed arc from node $j$ to node $k$ indicates that a customer may go from station $j$ to station $k$, and (iii) the nodes are numbered, $i = 1, 2, \ldots, r$, in such a way that there is a directed arc from node $j$ to node $k$ only if $j < k$. The decision variables and feasible sets are the same as in Type II-A networks. The queue discipline is FCFS.

The main result of this paper is that

$$\hat{c} = 1,$$

for a wide variety of queueing systems with objective function given by

(3).

In Chapter II systems of Type I-A are studied for which the objective function takes the form

$$E[C(c, \mu)] = c_8 c_\mu + c_w E[L(c, \mu)].$$
We prove (6) for such systems by proving the stochastic ordering

\[(L_t(1,c\mu)|A = a) \subseteq (L_t(c,\mu)|A = a),\]

for all feasible pairs \((c,\mu)\), all realizations, \(a\), of the arrival process, \(A\), and all times \(t \geq 0\). The proof is given for the cases of degenerate, exponential, and \(k\)-Erlang service-time distribution. The optimal value of \(\mu\) is derived for \(M/M/c\) systems, where \(R = [\mu|\mu > 0]\).

In Chapter III, the main results of Chapter II are extended to Type I-A queueing systems with objective functions of the form

\[(9) \quad E[C(c,\mu)] = C_s(c,\mu) + C_wE[L(c,\mu)],\]

where

\[(10) \quad C_s(1,\mu) \leq C_s(c,\mu), \text{ for all } c \text{ and } \mu.\]

The case where (10) is violated is also discussed.

In Chapter IV, we consider Type I-A systems with objective function

\[(11) \quad E[C(c,\mu)] = C_s(c,\mu) + E[C_w(c,\mu)],\]

where \(C_s(c,\mu)\) satisfies (10) and \(C_w(c,\mu)\) depends not only on the number of customers but also on the length of time each has been in the system. It is shown that, under very general conditions, the expected system waiting-cost rate obeys
\[(12) \quad E[\mathcal{C}_W(c, \mu)] = \lambda E[\mathcal{C}_W(h(W(c, \mu)))], \]

where \( h \) is a monotone increasing, differentiable Borel function with \( h(0) = 0 \). The relation,

\[(13) \quad E[h(W_n(1, c\mu))|A = a] \leq E[h(W_n(c, \mu))|A = a] \]

is proved for the case where the service-time distribution is \( k \)-Erlang and \( h \) is convex. If (12) holds, \( \hat{c} = 1 \) for this case as a consequence of (13). It is also shown that \( \hat{c} = 1 \) for exponential service-time distribution if (12) holds, where in this case \( h \) need not be convex.

This result is proved by demonstrating algebraically that, for all \( t \geq 0 \), \( \text{Prob} \{ W(1, c\mu) > t \} \leq \text{Prob} \{ W(c, \mu) > t \} \).

In Chapter V, network queueing systems of Type II-A and Type II-B are considered. In the case of exponential service-time distributions, the optimal value of \( c_i \) is shown to be \( \hat{c}_i = 1, \ i = 1, 2, \ldots, r \), for objective functions of the form

\[(14) \quad E[C(c, \mu)] = \sum_{i=1}^{r} C_{Si}(c, \mu) + C_{W}E[L(c, \mu)] \]

where \( c = (c_1, c_2, \ldots, c_r), \, \mu = (\mu_1, \mu_2, \ldots, \mu_r) \) and \( C_{Si}(1, c\mu) \leq C_{Si}(c, \mu) \), all \( c, \mu, \ i = 1, 2, \ldots, r \).

In Chapter VI we consider extensions of the main results of Chapter II to single-station models of greater generality. In particular, we study models of Type I-B, of Type I-A where the service-cost rate is charged only when servers are busy, of Type I-C, I-D, and I-E, and a
special team model.

The chapters beyond Chapter II may be read independently of each other. Chapter II contains the basic results upon which the work in the other chapters is based, and should be read first.
CHAPTER II

Single-Station Models with Linear Costs

1.1 Introduction and Motivation

In this chapter we study Type I-A queueing systems with a simple linear type of objective function. Recall the definition of Type I-A systems given in Section 1.4:

TYPE I-A: A single-station queueing system with \( c \) servers whose service times are mutually independent, with mean rate \( \mu \) and c.d.f. \( F(\cdot) \). The queue discipline is FCFS. The decision variables are \( c \) and \( \mu \) and the feasible sets are:

\[
c \leq \overline{c}, \quad \overline{c} \text{ a positive integer or } \infty;
\]

\[
\mu \in \mathbb{R}, \quad R \text{ a } \overline{c}\text{-complete set}.
\]

The objective function is the following special case of (3) in Section 1.4:

\[
E[C(c,\mu)] = C_s c \mu + C_w E[L(c,\mu)].
\]

Implicit in this cost equation are the following assumptions:

(a) The service-cost rate is a deterministic function of \( c \) and \( \mu \), constant through time. We are modeling a system where one must pay for a server whether he is busy or idle, where the total service-cost rate is proportional to the number of servers, and where the cost rate for each server is proportional to his mean rate of service. (See
Chapter III for more general service-cost functions.)

(b) The waiting-cost rate, on the other hand, is a random function of $c$ and $\mu$, through its dependence on the random variable $L(c,\mu)$. The waiting-cost rate for any realization of the queueing process, however, is proportional to the number of customers in the system. Implicit in this model, therefore, is the assumption that the waiting-cost rate for each customer is the same and that this rate does not depend on the length of time he has been in the system; that is, the total waiting cost incurred by a customer is proportional to the time he spends in the system. (See Chapter IV for more general waiting-cost functions.)

We are looking for the pair $(\hat{c},\hat{\mu})$ which will minimize (1). The main result of this chapter will be a proof that, under quite general assumptions about the arrival process and service-time distribution, the optimal value of $c$ is $\hat{c} = 1$.

2.1.1 Informal argument for $\hat{c} = 1$

For the $M|\mu|c$ system, the result, $\hat{c} = 1$, has been demonstrated empirically by Morse [23] and cited in Hillier [12], and Hillier and Lieberman [14]. It has a certain intuitive plausibility, at least for systems with exponential service times, as the following informal argument should demonstrate.

Suppose we are given a pair $(c,\mu)$ which is supposed to minimize (1) for a system with exponential service times, and suppose $c > 1$. Then consider the pair $(1,c\mu)$. This pair yields the same service-cost rate. Thus it remains to demonstrate that it will yield at least as small a value for the expected number in the system. To see that
this is true, consider the queue and the servers together as a black box which accepts input from the arrival process and produces output as a function of the arrival stream, the number of servers, and the mean service rate. Denote the black box specified by the pair \((c, \mu)\) the c-system, and the system specified by the pair \((l, c\mu)\) the l-system.

Now in the l-system, as long as there are customers present, the output process is Poisson with mean rate \(c\mu\); the input-output situation is therefore as illustrated below, where \(n \geq 1\) is the number of customers present:

```
\[
\begin{array}{c}
\text{input process} \\
A
\end{array}
\xrightarrow{\text{l-system}}
\[
\begin{array}{c}
\frac{n(\geq 1)}{n(\geq 1)}
\end{array}
\xrightarrow{\text{output process}}
\[
\begin{array}{c}
\text{Poisson (} c\mu \text{)}
\end{array}
\]
```

Let \((x \land y)\) denote the minimum of two real numbers, \(x\) and \(y\). In the c-system, if there are \(n\) customers present, then there are \((n \land c)\) busy servers, each working at mean rate \(\mu\). It is a well-known fact that the combined output from \(l\) independent and identically distributed Poisson processes, each with mean rate \(\mu\), is a Poisson process with mean rate \(l\mu\). Therefore, the output process of the c-system, conditioned on the number, \(n\), of customers in the system, is a Poisson process with mean rate \((n \land c)\mu\).

```
\[
\begin{array}{c}
\text{input process} \\
A
\end{array}
\xrightarrow{\text{c-system}}
\[
\begin{array}{c}
\frac{n(\geq 1)}{n(\geq 1)}
\end{array}
\xrightarrow{\text{output process}}
\[
\begin{array}{c}
\text{Poisson ((} n \land c \text{)}\mu\)}
\end{array}
\]
```

Thus, no matter how many customers are present, the l-system and
the c-system are identical except that the mean output rate of the
1-system is at least as large as that of the c-system. This fact
suggests that, on the average, the 1-system will be ejecting customers
greater than the c-system and hence will have less congestion and fewer
customers in the system.

There are pitfalls in this argument. For one thing, it is some-
what circular, since it involves conditioning on the number in the
system. However, the formal reasoning in the proof of our main theorem
(Theorem 2.1) proceeds inductively and thereby avoids circularity.
Moreover, the circularity is not essential to the informal argument,
for observe that the 1-system not only dominates (in terms of mean
output rate) a c-system with the same number of customers in it, it also
dominates a c-system with any other number of customers present.

The significance of this dominance can be readily appreciated if
one imagines that the black-box service mechanism, instead of consisting
of c parallel servers, consists of one server whose service-time
distribution is exponential, but with instantaneously variable mean
rate, \((n \wedge c) \mu\). Now we are comparing two single-server systems which
are alike in all respects except that the mean service rate of the
server is always greater in the system with the server who serves at a
constant rate \(c \mu\). Clearly, we can expect less congestion in the latter
system. But the variable-rate single-server system we have just
described is certainly equivalent (in terms of the distribution of the
number in the system) to the original c-system, since the service-
mechanisms of both can be represented by the same black box. The reader
who objected to the previous argument may find this version more
persuasive.*

For non-exponential service-time distributions, we will have to take into account the higher moments of the service-time distribution in order to make the above argument rigorous. In particular, we need to be certain that the variance of the service time in the l-system is not too large relative to the variance in the c-system, since the expected number in the system is generally an increasing function of the variance of the service time.

It should be noted that so far, and in what follows, very little is said about the nature of the arrival process. Essentially this is because we are comparing two systems which differ only in the nature of the service mechanisms, so that the complicating effects of pathological arrival processes are washed out in the comparison, since these effects are felt equally by both systems. We need assume only that the arrival process is independent of the service mechanism in the sense that the times of arrivals in the future are not affected by how the service mechanism has operated in the past (Assumption IV).

2.1.2 Statement of main theorem and corollaries

The general form of the main theorem is given below. It will be proved in Sections 2.2, 2.3, and 2.4, for the specific types of service-time distribution: degenerate, exponential, and k-Erlang. The statement of the theorem is followed by a remark and some corollaries.

* This form of the argument was suggested by Professor A. F. Veinott, Jr., who also suggested the method of proof of Theorem 2.1 for the case of exponential service-time distribution.
Let \( a = (a_1, a_2, \ldots) \) be a realization of the arrival process, \( A = (A_1, A_2, \ldots) \) to a system of Type I-A. We will use the following notation for the number of customers in the system at time \( t \), given \( a \), as a function of \( c \) and \( \mu \):

\[
L^a_t(c, \mu) = (L_t(c, \mu)|A = a).
\]

**Theorem 2.1:** Suppose the service-time distribution in a queueing system of Type I-A is degenerate, exponential, or \( k \)-Erlang \( (k = 1, 2, \ldots) \).

Then, for any arbitrary realization \( a \) of the arrival process, and any feasible pair \( (c, \mu) \),

\[
(2) \quad L^a_t(1, c\mu) \subset L^a_t(c, \mu), \quad t \geq 0.
\]

It should be noted in the proofs of Theorem 2.1 which follow in the succeeding sections that we do not make use of Assumption III, which restricts the arrival process to the class of those for which there is an optimal policy for the system in question among those for which a steady state exists. In fact, for Theorem 2.1 we need make no assumptions about the arrival process except Assumption IV: that it is independent of the service mechanism. The sequence of interarrival times need not be a renewal process nor even a stationary process. This fact is useful in extending the result to networks of queues (see Chapter IV), where the output process of one station is the input process to another, since the output processes of queueing systems are not generally renewal processes. Also, the realization \( a \) could
include some batches (cases, e.g., where \( a_m < a_{m+1} = a_{m+2} = \cdots = a_{m+k} < a_{m+k+1} \), yielding a batch of size \( k \)). This fact is useful in the proof of the theorem for k-Erlang service-time distribution, in which the familiar device is employed of relating a system with k-Erlang service-time distribution and unit arrivals to a system with exponential service-time distribution and batch arrivals of fixed batch-size \( k \).

The following two corollaries give the result of Theorem 2.1 for the two most obvious special cases of \( \bar{c} \)-complete sets, for \( \bar{c} = \infty \).

**COROLLARY 2.2:** Suppose the feasible sets for \( c \) and \( \mu \) in a queueing system of Type I-A have the form:

\[
c = 1, 2, \ldots \; ; \quad 0 < \mu < \infty .
\]

Then, if the service-time distribution is degenerate, exponential, or k-Erlang \( (k = 1, 2, \ldots) \),

\[
L_t^a(1, c\mu) \subset L_t^a(c, \mu), \quad t \geq 0 ,
\]

for any \( a \) and any feasible pair \( (c, \mu) \).

**COROLLARY 2.3:** Suppose the feasible sets in a queueing system of Type I-A have the form:

\[
c = 1, 2, \ldots ;
\]

\[
\mu = i\mu_0, \quad \text{for some} \quad \mu_0 > 0 \; , \; i = 1, 2, \ldots .
\]

Then, if the service-time distribution is degenerate, exponential, or
k-Erlang \((k = 1, 2, \ldots)\),

\[ L_t^a(1, c, \mu) \subseteq L_t^a(c, \mu), \quad t \geq 0, \]

for any \(a\) and any feasible pair \((c, \mu)\).

The following six corollaries culminate in a proof that \(c = 1\) and give several intermediate results.

COROLLARY 2.4: Under the conditions of Theorem 2.1, for any \(a\) and any feasible pair \((c, \mu)\),

\[ \mathbb{E}[L_t^a(1, c, \mu)] \leq \mathbb{E}[L_t^a(c, \mu)], \quad t \geq 0. \]

Proof.

Apply property \(1^\circ\) of stochastic ordering (cf. Section 1.4) to (2), and obtain (3) immediately.

Q.E.D.

COROLLARY 2.5: Suppose the service-time distribution in a queueing system of Type I-A is degenerate, exponential, or k-Erlang \((k = 1, 2, \ldots)\). Then, for any feasible pair \((c, \mu)\),

\[ L_t(1, c, \mu) \subseteq L_t(c, \mu), \quad t \geq 0. \]

Proof.

We must show that, for any \(t \geq 0\), and for any \(x \geq 0\),

\[ \text{Prob}[L_t(1, c, \mu) > x] \leq \text{Prob}[L_t(c, \mu) > x]. \]
By the law of total probability and (2),

$$\text{Prob}[L_t(1, c\mu) > x] = \int \text{Prob}[L_t^a(1, c\mu) > x] \, d \text{Prob}[A \leq a]$$

$$\leq \int \text{Prob}[L_t^a(c, \mu) > x] \, d \text{Prob}[A \leq a]$$

$$= \text{Prob}[L_t(c, \mu) > x].$$

Q.E.D.

COROLLARY 2.6: Under the conditions of Corollary 2.5, for any feasible pair \((c, \mu)\),

$$E[L_t(1, c\mu)] \leq E[L_t(c, \mu)], \quad t \geq 0.$$  \hspace{1cm} (5)

Proof.

Apply property 1° to (4), and obtain (5) immediately.

Q.E.D.

COROLLARY 2.7: Under the conditions of Corollary 2.5, let \((c, \mu)\) be a feasible pair for which steady-state distributions of \(L(c, \mu)\) and \(L(1, c\mu)\) exist. Then,

$$L(1, c\mu) \subseteq L(c, \mu).$$ \hspace{1cm} (6)

Proof.

An immediate consequence of (4) and the fact that \(L_t(c, \mu) \to L(c, \mu)\) and \(L_t(1, c\mu) \to L(1, c\mu)\) in distribution as \(t \to \infty\).

Q.E.D.

COROLLARY 2.8: Under the conditions of Corollary 2.7,
\[(7) \quad E[L(1, c\mu)] \leq E[L(c, \mu)] \, .\]

**Proof.**

Apply property \(1^{\circ}\) to (6), and obtain (7) immediately.

Q.E.D.

*COROLLARY 2.9: Suppose the service-time distribution in a queueing system of Type I-A is degenerate, exponential, or \(k\)-Erlang \((k = 1, 2, \ldots)\). Then the optimal value of \(c\) for the objective function given by (1) is

\[(8) \quad \hat{c} = 1 \, .\]

**Proof.**

Suppose \((c, \mu)\) are an arbitrary feasible pair, claimed to be optimal for (1), with \(c > 1\). Since \(R\) is a \(\overline{c}\)-complete set, \(R\) contains \(c\mu\). Therefore, the pair \((1, c\mu)\) is feasible. But, by (7),

\[E[L(1, c\mu)] \leq E[L(c, \mu)] \, .\]

Therefore, by (1),

\[E[C(1, c\mu)] = c_s c\mu + E[L(1, c\mu)] \leq c_s c\mu + E[L(c, \mu)] = E[C(c, \mu)] \, .\]

Therefore, \(\hat{c} = 1\).

Q.E.D.
In the following corollary, the optimal value of $\mu$ is derived for a special case.

COROLLARY 2.10: Consider a $M|M|c$ system of Type I-A with objective function given by (1) and feasible set, $0 < \mu < \infty$. Then the optimal values for $c$ and $\mu$ are:

\[
\hat{c} = 1; \\
\hat{\mu} = \lambda + \sqrt{\frac{\lambda C_V}{C_S}}.
\]

Proof.

See Morse [23] for a proof of the formula for $\hat{\mu}$.

Q.E.D.

2.2 Proof of Theorem 2.1 for Degenerate Service-Time Distribution

In this section we provide a proof of Theorem 2.1 for the case of constant service time, i.e., where

\[
F(t) = \begin{cases} 
0, & t < \frac{1}{\mu} \\
1, & t \geq \frac{1}{\mu}
\end{cases}
\]

for a server with mean service rate $\mu$.

Let $T = \frac{1}{c\mu}$ be the constant service time of the server in the $1$-system. Then $cT = \frac{1}{\mu}$ is the constant service time of the servers in the $c$-system. Since we are dealing with a deterministic stream of arrivals and deterministic service times, the systems as a whole are
deterministic and the relation we want to prove is:

\[(10) \quad L_t^{a}(1,c\mu) \leq L_t^{a}(c,\mu), \quad t \in [0,\infty).\]

It suffices to consider only those periods of time in which the server in the \(l\)-system is actually serving a customer (\(l\)-system busy periods), since when he is idle, there is no one in the \(l\)-system \((L_t^{a}(1,c\mu) = 0)\) and \((10)\) is trivially satisfied, since \(L_t^{a}(c,\mu) \geq 0\) for all \(t \in [0,\infty)\).

Therefore let us consider an arbitrary \(l\)-system busy period of length \(\bar{t}\) and suppose, for convenience, that it begins at time \(t = 0\) and ends at time \(t = \bar{t}\). Then we need to show:

\[(11) \quad L_t^{a}(1,c\mu) \leq L_t^{a}(c,\mu), \quad t \in [0,\bar{t}].\]

Now the \(l\)-system busy period begins with an arrival at time \(t = 0\). Just before \(t = 0\), the number in the \(l\)-system is \(0\), by definition of a busy period. However, at \(t = 0-\), the \(c\)-system may be in the middle of a busy period and hence have some number in the system. Therefore, let us denote by \(L_0 (\geq 0)\) the number in the \(c\)-system just before the arrival at \(t = 0\). Thus for any \(t \in [0,\bar{t}]\).

\[(12) \quad L_t^{a}(1,c\mu) = n_a(t) - n_d(t|1,c\mu)\]

\[L_t^{a}(c,\mu) = L_0 + n_a(t) - n_d(t|c,\mu),\]

where \(n_a(t) \equiv \text{number of arrivals in } [0,t]\)

\(n_d(t|1,c\mu) \equiv \text{number of departures from } l\text{-system in } [0,t]\)
\( n_d(t|c, \mu) \equiv \text{number of departures from c-system in } [0, t]. \)

Now since the server in the l-system starts a service at \( t = 0 \) and is always busy in \([0, t]\), \( n_d(t|l, c \mu) \) will just be the maximum number of service times of length \( T \) that can be fit into an interval of length \( t \), or \( \lfloor \frac{t}{T} \rfloor \), where \( \lfloor x \rfloor \) denotes the largest integer \( \leq x \). Therefore,

\[
L_t^a(l, c \mu) = n_a(t) - \lfloor \frac{t}{T} \rfloor .
\]

In the c-system, the number of services which can be started and finished by each server is, by a similar argument, less than or equal to \( \lfloor \frac{t}{cT} \rfloor \). Therefore, the total number of services which can be started and finished in the c-system in \([0, t]\) is less than or equal to \( c \lfloor \frac{t}{cT} \rfloor \). In addition to these new services, one or more of the services in progress at time \( t = 0 \) may terminate in the interval \([0, t]\). Since the number of services in progress at \( t = 0 \) is \( L_0 \lor c \), the number of such terminations must be less than or equal to \( L_0 \lor c \). Therefore,

\[
n_a(t|c, \mu) \leq c \lfloor \frac{t}{cT} \rfloor + L_0 \lor c
\]

\[
\leq \lfloor \frac{t}{T} \rfloor + L_0 .
\]

Therefore, using (12) and (13),

\[
L_t^a(c, \mu) \geq L_0 + n_a(t) - \lfloor \frac{t}{T} \rfloor - L_0
\]

\[
= n_a(t) - \lfloor \frac{t}{T} \rfloor
\]

\[
= L_t^a(l, c \mu) .
\]

Q.E.D.
2.3 **Proof of Theorem 2.1 for Exponential Service-Time Distribution**

In this section we give a proof of Theorem 2.1 for the case of exponentially distributed service time, i.e., \( F(t) = 1 - e^{-\mu t}, \ t \geq 0, \) for a server with mean service rate \( \mu. \)

As in the proof for constant service time, it suffices to prove that (2) holds for \( t \) in a \( l \)-system busy period. The proof here, however, is considerably more involved and requires the following notation and a lemma, which will be proved later in the section:

\[
U_n = n^{th} \text{ departure time from the } l\text{-system, } n = 1,2,\ldots ,
\]

\[
V_n = n^{th} \text{ departure time from the } c\text{-system, } n = 1,2,\ldots .
\]

(Note that the \( n^{th} \) departure time from the \( c\)-system is not necessarily the departure time of the \( n^{th} \) arrival, since an item that arrives later than the \( n^{th} \) arrival may finish service earlier, even though it must enter service later.)

**Lemma 2.11:**

\[
(14) \quad U_1 \subseteq V_1 ;
\]

\[
(U_n | U_{n-1} = z) \subseteq (V_n | V_{n-1} = z),
\]

for \( n = 2,3,\ldots, \) and all feasible \( z \geq 0. \) Moreover,

\[
(15) \quad (U_n | U_{n-1} = z) \text{ is a stochastically monotone increasing function of } z, \ n = 2,3,\ldots
\]

We now give a proof of Theorem 2.1 for exponential service-time distribution, using this lemma.
Fix $t > 0$, and assume $t$ is in a $l$-system busy period which began at time $0$. As in Section 2.1, let $L_0 = \text{number in } c$-system just before time $0$. Then,

$$L_t^a(1,c) = n_a(t) - n_d(t|1,c)$$
$$L_t^a(c) = L_0 + n_a(t) - n_d(t|c,c)$$

We wish to show, for $m = 0,1,2,...$,

$$\text{Prob}[L_t^a(1,c) > m] \leq \text{Prob}[L_t^a(c) > m]$$

Now for $m \geq n_a(t)$, the left-hand side of this expression is 0 and the relation is trivially true. Therefore we assume that $0 \leq m < n_a(t)$.

In the $l$-system, using (16) and letting $n = n_a(t) - m$:

$$L_t^a(1,c) > m \iff n_a(t) - n_d(t|1,c) > m$$
$$\iff n_d(t|1,c) < n_a(t) - m$$
$$\iff n_d(t|1,c) < n$$
$$\iff U_n > t$$

Similarly, it can be shown that $L_t^a(c,c) > m$ if $V_n > t$.

Therefore it suffices to show, for all $t > 0$, that

$$\text{Prob}[U_n > t] \leq \text{Prob}[V_n > t], \quad n = 1,2,...,n_a(t)$$

or

$$U_n \subseteq V_n, \quad n = 1,2,...$$
We will prove (17) by induction on $n$, using the lemma and properties $1^\circ$ and $2^\circ$ of stochastic ordering.

**Case $n = 1$** $U_1 \subseteq V_1$ by Lemma 2.11.

Q.E.D.

**Case $n > 1$** Assume that (17) is true for $n - 1$, i.e., $U_{n-1} \subseteq V_{n-1}$. Fix $t > 0$. We wish to show that $\text{Prob}(V_n > t) \leq \text{Prob}(V_{n-1} > t)$. Observe that

$$
\text{Prob}(V_n > t) = \int \text{Prob}(V_n > t \mid U_{n-1} = z) \, d \, \text{Prob}(U_{n-1} \leq z).
$$

Therefore, using (15) of Lemma 2.11, the induction assumption, and properties $1^\circ$ and $2^\circ$ of stochastic ordering,

\begin{equation}
\text{Prob}(V_n > t) = \int \text{Prob}(V_n > t \mid U_{n-1} = z) \, d \, \text{Prob}(U_{n-1} \leq z) \\
\leq \int \text{Prob}(V_n > t \mid U_{n-1} = z) \, d \, \text{Prob}(V_{n-1} \leq z).
\end{equation}

But, from (14) of Lemma 2.11,

\begin{equation}
\int \text{Prob}(V_n > t \mid U_{n-1} = z) \, d \, \text{Prob}(V_{n-1} \leq z) \\
\leq \int \text{Prob}(V_n > t \mid V_{n-1} = z) \, d \, \text{Prob}(V_{n-1} \leq z)
\end{equation}

Therefore, combining (18) and (19),

$$
\text{Prob}(V_n > t) \leq \int \text{Prob}(V_n > t \mid V_{n-1} = z) \, d \, \text{Prob}(V_{n-1} \leq z) = \text{Prob}(V_n > t).
$$

Q.E.D.
We now prove Lemma 2.11.

Proof of Lemma 2.11.

We will prove (14) for n = 2, 3, ...; the same method of proof works for the special case, n = 1.

We must show, for all t > z, that

\[(20) \quad \text{Prob}\{U_n > t | U_{n-1} = z\} \leq \text{Prob}\{V_n > t | V_{n-1} = z\} \quad .\]

Let \(b_1(\tau)\) be the number of busy servers in the l-system at time \(\tau\), \(z \leq \tau \leq t\), given \(U_{n-1} = z\) and \(U_n > \tau\); let \(b_c(\tau)\) be number of busy servers in the c-system at time \(\tau\), \(z \leq \tau \leq t\), given \(V_{n-1} = z\) and \(V_n > \tau\). Then, since we are in a l-system busy period,

\[(21) \quad b_1(\tau) = 1 \quad , \quad \text{for all z \leq \tau \leq t} \quad . \]

Since the maximum number of busy servers in the c-system is \(c\),

\[(22) \quad b_c(\tau) \leq c \quad , \quad \text{for all z \leq \tau \leq t} \quad . \]

At time \(\tau\), \(z \leq \tau \leq t\), given \(U_{n-1} = z\) and the \(n^{th}\) departure has not yet occurred, the output from the l-system is a Poisson process with mean rate \(b_1(\tau)c_\mu\). Therefore, if we let \(\eta(\tau - z) = b_1(\tau)c_\mu\), and \(Z = U_n - z\), equation (4) of Lemma 1.1 is satisfied with \(t\) replaced there by \(\tau - z\). Hence, Lemma 1.1 applies, and using (21), we obtain

\[(23) \quad \text{Prob}\{U_n > t | U_{n-1} = z\} = \exp\left(-\int_0^{t-z} b_1(\tau)c_\mu d\tau\right) \quad . \]
Similarly, at time $\tau$, $z \leq \tau \leq t$, given $V_{n-1} = z$, and the $n^{th}$ departure from the c-system has not yet occurred, the output from the c-system is a Poisson process with mean rate $b_c(\tau)\mu$. Therefore, again applying Lemma 1.1 and using (22), we obtain

$$\text{Prob}\{V_n > t | V_{n-1} = z\} = \exp(-\int_0^{t-z} b_c(\tau)\mu d\tau)$$

$$\leq \exp(-\int_0^{t-z} c\mu d\tau)$$

$$= \exp(-c\mu(t-z)) .$$

Therefore, combining (23) and (24), we obtain (20), the desired result.

Finally, the proof of (15), that $(U_n | U_{n-1} = z)$ is a stochastically monotonic increasing function of $z$, follows directly from equation (23).

Q.E.D.

2.4 Proof of Theorem 2.1 for k-Erlang Service-Time Distribution

We now treat the case where the service-time distribution is k-Erlang; that is, if $\mu$ is the mean service rate of a server, then

$$F'(t) = \frac{(k\mu)^k}{(k-1)!} \exp(-k\mu t), \quad t \geq 0 .$$
In order to prove Theorem 2.1 for k-Erlang service-time distribution, it suffices as in the previous cases to prove that (2) holds for \( t \) in a \( l \)-system busy period beginning at time \( 0 \). This will be done by the following steps:

(a) We consider each customer as a batch of \( k \) phases which must be served sequentially by the same server, independently and according to the same exponential service-time distribution, with mean service rate \( kc\mu \) for the \( l \)-system and \( k\mu \) for the \( c \)-system. A customer is considered to begin service when the first of the phases associated with it begins service and to leave service and depart from the system when the last of its phases departs.* The service-time distribution for a customer, defined in this way, is k-Erlang with mean rate \( c\mu (\mu) \) in the \( l \)-system (\( c \)-system), since it is the convolution of \( k \) independent exponential distributions with mean rate \( kc\mu (k\mu) \).

(b) By an argument similar to that used in the proof of Theorem 2.1 for exponential service-time distributions, we prove that the number of phases in the k-Erlang \( l \)-system is stochastically smaller than the number of phases in the k-Erlang \( c \)-system, at any time \( t \geq 0 \).

(c) Finally, we show that having stochastically fewer phases implies having stochastically fewer customers.

We will use the following notation: let \( P_t^l(1, c\mu) \) = the number of phases in the k-Erlang \( l \)-system at time \( t \), and \( P_t^c(c, \mu) \) = the number

* A phase is treated as if it were a physical entity which enters the system with its associated customer and departs from the system when its service is completed.
of phases in the k-Erlang c-system at time $t$; let $X_n$ the departure time of the $n$th phase to depart from the l-system, and $Y_n$ the departure time of the $n$th phase to depart from the c-system.

The following lemmas will be required:

**LEMMA 2.12:**

\[(25)\]

\[X_1 \subseteq Y_1 ;\]

\[(X_n | X_{n-1} = z) \subseteq (Y_n | Y_{n-1} = z) ,\]

for $n = 2, 3, \ldots$, all feasible $z \geq 0$ ;

\[(26)\]

$(X_n | X_{n-1} = z)$ is a stochastically monotone increasing function of $z$, for $n = 2, 3, \ldots$ .

**LEMMA 2.13:** $P_t^{a}(1, c \mu) \subseteq P_t^{a}(c, \mu)$, $t \geq 0$.

**LEMMA 2.14:** For all $t \geq 0$, if $P_t^{a}(1, c \mu) \subseteq P_t^{a}(c, \mu)$, then $L_t^{a}(1, c \mu) \subseteq L_t^{a}(c, \mu)$.

Lemmas 2.13 and 2.14 together imply

\[(2)\]

$L_t^{a}(1, c \mu) \subseteq L_t^{a}(c, \mu)$, all $t \geq 0$ ,

the desired result, for k-Erlang distributions. Therefore, it remains to prove the three lemmas.

**Proof of LEMMA 2.12.**

The proof is exactly the same as the proof of Lemma 2.11, with
$U_n$ replaced by $X_n$, $V_n$ by $Y_n$, $\mu$ by $k\mu$, and $c\mu$ by $kc\mu$. Thus,

$$\operatorname{Prob}(X_n > t | X_{n-1} = z) = \exp(-kc\mu(t-z)),$$

and

$$\operatorname{Prob}(Y_n > t | Y_{n-1} = z) \geq \exp(-ck\mu(t-z)).$$

Combining (27) and (28) yields (25) for $n = 2, 3, \ldots$; the special case $n = 1$ is proved by the same method, mutatis mutandis. We obtain (26) as a direct consequence of (27).

Q.E.D.

Proof of Lemma 2.13.

Our proof follows the lines of the proof of Theorem 2.1 for exponential service-time distributions. By the same reasoning as used there, it suffices to prove

$$X_n \subseteq Y_n, \quad n = 1, 2, \ldots,$$

by induction on $n$.

We use Lemma 2.12 to prove the case $n = 1$ directly.

Consider the case $n > 1$, and assume (29) is true for $n - 1$, i.e., $X_{n-1} \subseteq Y_{n-1}$. Fix $t > 0$. We wish to show that

$$\operatorname{Prob}(X_n > t) \leq \operatorname{Prob}(Y_n > t).$$

Observe that
\[(30) \quad \text{Prob}\{X_n > t\} = \int \text{Prob}\{X_n > t | X_{n-1} = z\} \, d\, \text{Prob}\{X_{n-1} \leq z\}.\]

Using (30), (26), the induction assumption, and properties 1° and 2°, we obtain:

\[(31) \quad \text{Prob}\{X_n > t\} = \int \text{Prob}\{X_n > t | X_{n-1} = z\} \, d\, \text{Prob}\{X_{n-1} \leq z\} \leq \int \text{Prob}\{X_n > t | X_{n-1} = z\} \, d\, \text{Prob}\{Y_{n-1} \leq z\}.\]

But, by (25),

\[(32) \quad \int \text{Prob}\{X_n > t | X_{n-1} = z\} \, d\, \text{Prob}\{Y_{n-1} \leq z\} \leq \int \text{Prob}\{Y_n > t | Y_{n-1} = z\} \, d\, \text{Prob}\{Y_{n-1} \leq z\}.\]

Therefore, combining (31) and (32),

\[
\text{Prob}\{X_n > t\} \leq \int \text{Prob}\{Y_n > t | Y_{n-1} = z\} \, d\, \text{Prob}\{Y_{n-1} \leq z\} = \text{Prob}\{Y_n > t\},
\]

which completes the proof.

Q.E.D.

Proof of LEMMA 2.14.

Fix \( t \geq 0 \) and let

\[(33) \quad F^a_t(1, c, \mu) \subset F^a_t(c, \mu).\]
We wish to show that

\[ L_t^\alpha(l, c, \mu) \subseteq L_t^\alpha(c, \mu) \, . \]  

Now, in the \( l \)-system, the number of customers in the system is unambiguously determined by the number of phases in the system; it is found by dividing the number of phases by \( k \) and taking the smallest integer greater than or equal to the quotient. Therefore

\[ L_t^\alpha(l, c, \mu) = \langle P_t^\alpha(l, c, \mu)/k \rangle \, , \]

where \( \langle x \rangle \) denotes the smallest integer \( \geq x \). However, clearly we have, for any realization of the \( c \)-system

\[ L_t^\alpha(c, \mu) \geq \langle P_t^\alpha(c, \mu)/k \rangle \, , \]

since each customer in the \( c \)-system can have at most \( k \) phases associated with it and still in the system.*

Now \( \langle x \rangle \) is a non-decreasing function of \( x \). Therefore, using property 2° of stochastic ordering and combining (35), (33), and (36)

---

* The following examples show that (36) can occur with both strict inequality and equality. Let \( c = 2, k = 2, P_t^\alpha(c, \mu) = 4 \). Therefore \( \langle P_t^\alpha(c, \mu)/k \rangle = \langle 4/2 \rangle = 2 \).

**Case 1:** There is one phase at each server and two phases (one customer) in the queue. Hence \( L_t^\alpha(c, \mu) = 3 > 2 = \langle P_t^\alpha(c, \mu)/k \rangle \).

**Case 2:** There are two phases (one customer) at each server, one in service and one in the subqueue. Hence \( L_t^\alpha(c, \mu) = 2 = \langle P_t^\alpha(c, \mu)/k \rangle \).
we obtain (34), our desired result.

2.4 Remarks about Theorem 2.1

(i) The validity of Theorem 2.1 for constant and exponential service-time distributions can be deduced from its validity for $k$-Erlang service-time distributions by letting $k \to \infty$ and setting $k = 1$, respectively. Separate proofs were given for pedagogical reasons.

(ii) Following is a sketch of an alternative method of proof of Lemma 2.13 for $k$-Erlang service-time distributions.

We observe that, with the state of the system defined as the number of phases in the system, the $k$-Erlang $l$-system is equivalent to an exponential $l$-system with batch arrivals of fixed batch size $k$. Moreover, with the new state space, the $k$-Erlang $c$-system, while not equivalent to an exponential batch-arrival $c$-system (they have different queue disciplines), is dominated by such a system (i.e., has stochastically more items in it at all times).

Using Theorem 2.1 for exponential service-time distributions, together with these equivalence and dominance relations, we conclude that the $c$-system has stochastically more phases in it at all times than the $l$-system (this is the conclusion of Lemma 2.13).

Formal proofs of the equivalence and dominance relations can be constructed, following again as a model the proof of Theorem 2.1 for exponential service-time distributions, but the following informal argument should satisfy the reader.

First consider the two $l$-systems.
We wish to show that the number of phases at any time in a k-Erlang l-system with mean service rate \( c_\mu \) is identically distributed as the number of customers in an exponential l-system with mean service rate \( kc_\mu \), and with batch arrivals of fixed batch size \( k \), where the arrival times of the batches are the same as the arrival times of the customers to the k-Erlang system.

We note that batches of \( k \) customers arrive at the exponential system at exactly the same times as groups of \( k \) phases at the k-Erlang system. The service times of the customers in the exponential system are independent and identically distributed, exponentially with mean rate \( kc_\mu \); the service times of the phases in the k-Erlang system are also independent and identically distributed with the same exponential distribution. Finally, the queue discipline for customers in the exponential system is the same as that for phases in the k-Erlang system, namely first-come, first served. Hence, the distributions of the number in the system must be identical.

Now consider the two c-systems.

We wish to show that the number of phases at any time in a k-Erlang c-system with mean service rate \( \mu \) for each server is stochastically larger than the number of customers in an exponential c-system with mean service rate \( k\mu \) for each server and with arrivals occurring in batches of fixed size \( k \), where the arrival times of the batches are the same as the arrival times of the customers to the k-Erlang system.

As in the case of the l-systems, we note that batches of \( k \) customers arrive at the exponential system at exactly the same times as groups of \( k \) phases at the k-Erlang system, and that each of the c
servers in the exponential system serves customers according to the
same service-time distribution — exponential with mean rate $k\mu$ — as
each of the $c$ servers in the $k$-Erlang system serves phases.

The difference between the two systems lies in the manner in which
the next item to be served is selected from the queue. In the exponential
system, whenever a server becomes free, the customer at the head of the
queue (if there is one) goes to that server and immediately begins
service. Thus items belonging to the same batch are not necessarily
served by the same server. In the $k$-Erlang system, however, all phases
associated with the same customer are served sequentially by the same
server. Whenever a server finishes on the last phase of a customer,
all the phases associated with the customer at the head of the queue
move to that server and form a subqueue in front of him, the first phase
going immediately into service. The server serves all the phases in a sub-
queue before any other phases and phases in a subqueue are not allowed
to leave and be served by another server, even if he is idle.

Note that this situation — simultaneously having items waiting to
be served and idle servers — is not possible in the exponential system.
This fact strongly suggests that the exponential $c$-system with batch
arrivals is more efficient in processing customers than the $k$-Erlang
$c$-system is in processing phases and thus has stochastically fewer
customers in it at all times than the $k$-Erlang $c$-system has phases.*

(iii) We may use essentially the same inductive reasoning as was
used in Section 2.3 to prove

* For a similar argument in a similar situation, see [12, p. 123].
(37) \[ U_n \subseteq V_n , \quad n = 1, 2, \ldots , \]
to prove

(38) \[ L_{a_n}^a (1, c \mu) \subseteq L_{a_n}^a (c, \mu) , \quad n = 1, \ldots , \]
and thereby obtain an alternate proof of (2) for exponential service-
time distributions.* Following is a sketch of the proof of (38).

Let \( I_n = L_{a_n}^a (1, c \mu) \) = the number of customers in the \( l \)-system just
after the \( n \)th arrival, \( n = 1, 2, \ldots \); let \( J_n = L_{a_n}^a (c, \mu) \) = the number
of customers in the \( c \)-system just after the \( n \)th arrival, \( n = 1, 2, \ldots \).
For simplicity we consider arrival streams with no batch arrivals.
Assume that the first arrival occurs at time 0 and starts a \( l \)-system
busy period. We wish therefore to prove

(39) \[ I_n \subseteq J_n , \quad n = 1, 2, \ldots . \]

By the same argument as used in Section 2.3 to prove (37), it
suffices to prove

(40) \[ I_l \subseteq J_l ; \]
\[ (I_n | I_{n-1} = m) \subseteq (J_n | J_{n-1} = m) , \]

* Here we are considering the imbedded Markov chain at the arrival
points, the familiar device used to study GI|MC systems (cf. Kendall
[17]). In Section 2.3, since we were considering a fixed realization
of the arrival process (to a GI|MC system), it was not necessary to look
at an imbedded chain in order to have a Markovian process.
\[ n = 2, 3, \ldots, \text{ all feasible } m = 1, 2, \ldots; \]

(41) \quad \left( I_n | I_{n-1} = m \right) \text{ is a stochastically monotone increasing function of } m.

Relation (40) for \( n = 1 \) follows from the fact that the first arrival starts a 1-system busy period.

To prove (40) for \( n = 2, 3, \ldots \), we must show, for all \( k = 0, 1, \ldots, m+1 \),

(42) \quad \text{Prob}[I_n > k | I_{n-1} = m] \leq \text{Prob}[J_n > k | J_{n-1} = m].

Each of these probabilities equals the probability that there are more than \( m-k+1 \) departures from the system during the interarrival interval. Since the mean output rate from the 1-system is always greater than that from the c-system, there should be stochastically more departures from the 1-system than the c-system during the interarrival interval, and hence (42) should hold. The same sort of reasoning establishes (41).

Finally, to prove (2) from (38), observe that the above argument could equally well be applied to any portion of the interarrival interval.

(iv) In all our proofs, we have made extensive use of the Markovian properties of the various stochastic processes — number in the system, departure times — when the arrival stream \( a \) is fixed. Thus, for example, the distribution of \( U_n \), given any history of the system in which \( U_{n-1} = z \), is identical to the distribution of
(U_n \mid U_{n-1} = z). \ast \) Hence it appears that any attempt to extend the results of this chapter to service-time distributions that do not yield the Markovian properties will involve different, more complex methods of proof.

That the results should extend to certain more general service-time distributions seems likely from the informal argument presented in Section 2.1. There it was argued, in effect, that the fact that the \( \lambda \)-system has stochastically fewer customers than the \( \sigma \)-system was a consequence only of the fact that the \( \sigma \)-system can have some of its total capacity \((c\mu)\) idle when there are customers present, whereas in the \( \lambda \)-system this is not possible.

The service-time distributions we have been considering are all characterized by the property that decreasing the mean service time by a factor of \( \frac{1}{c} \) results in decreasing the standard deviation by the same factor, \( \frac{1}{c} \). Thus we may expect to be indifferent between one fast server and \( c \) servers each \( \frac{1}{c} \) as fast on the average, in terms of available service capacity. It is only the possibility of idle capacity in the \( \sigma \)-system that puts it at a disadvantage.

Therefore, we make the following conjecture.

CONJECTURE: The results of this chapter hold for any class of service-time distributions of the same form and characterized by the property

\* For the \( k \)-Erlang case, the \( n \)th departure time of a phase depends also on the number of busy servers at time \( z \), which is not uniquely determined by the event \((Y_{n-1} = z)\). However, it suffices, as we have seen, to know that in all cases the number of busy servers at \( z \) is \( \leq c \).
that the standard deviation is proportional to the mean.

(v) Lehman [19] presents a definition of stochastic ordering for random vectors which is stronger than the one we use (Section 1.4). Let \( X = (X_1, \ldots, X_n), \ Y = (Y_1, \ldots, Y_n) \) be random vectors.

**DEFINITION:** We say \( X \) is stochastically smaller (Lehman) than \( Y \) (denoted \( X \lesssim Y \)) if there exist functions \( g_i(\cdot) \) and \( g'_i(\cdot) \), \( i = 1, \ldots, n \), and a random variable \( Z \), such that, for \( i = 1, \ldots, n \),

\[
(43) \quad g_i(z) \leq g'_i(z), \quad \text{for all } z,
\]

\( g_i(Z) \) is distributed as \( X_i \),

\( g'_i(Z) \) is distributed as \( Y_i \).

Intuitively, this definition says that we can define a sample space \( \Omega \) such that \( X \) and \( Y \) are each defined on this space and \( X(\omega) \lesssim Y(\omega) \) for all \( \omega \in \Omega \). Clearly, this is a stronger concept than our concept of stochastic ordering.

To apply this definition to our problem, consider a \( l \)-system busy period starting at time 0 and let \( X_i \) and \( Y_i \) be the time between the \( i \)-1st departure and \( i \)-th departure from \( l \)-system and the \( c \)-system, respectively, for \( i = 1, \ldots, n \) (where \( X_0 = Y_0 = 0 \)). Then \( U_n = \sum_{i=1}^{n} X_i \) and \( V_n = \sum_{i=1}^{n} Y_i \), and to prove \( U_n \subseteq V_n \) it suffices (cf. Lehman) to prove \( X \lesssim Y \).

Let \( Z \) be exponentially distributed with mean rate \( \mu \). Let

\[
g'_i(z) = \frac{z}{c}, \quad \text{for } i = 1, \ldots, n.
\]

Then
\[
\text{Prob}\{ g_1(Z) > t \} = \text{Prob}\left\{ \frac{Z}{c} > t \right\} \\
= \text{Prob}\{ Z > ct \} \\
= e^{-ct} \\
= \text{Prob}\{ X_1 > t \}.
\]

Now if there were always \( m \) busy servers (\( m \leq c \)) in the \( c \)-system, then \( g_1'(z) = \frac{z}{m} \) would be the appropriate form for \( g_1'(\cdot) \) and we would have \( g_1(z) \leq g_1'(z) \), for all \( z \). In fact, we know that the number of busy servers in the \( c \)-system varies, so \( g_1'(\cdot) \) must be a more complicated function. However, it seems plausible that (43) would be satisfied, since the number of busy servers in the \( c \)-system is never greater than \( c \).
CHAPTER III

Single-Station Models with Non-Linear Service Costs

In this chapter some extensions of the results of Chapter II for single-station models of Type I-A are presented. In particular we investigate the ways in which the requirement that the service cost rate be linear in \( c\mu \) can be relaxed while retaining the key property that a single-server system is optimal under the conditions of Corollary 2.9.

The objective function we consider now has the form:

\[
E[C(c,\mu)] = C_s(c,\mu) + C_wE[L(c,\mu)],
\]

where \( C_s(c,\mu) \) need not be linear in \( c\mu \). It is an immediate consequence of Theorem 2.1 that a sufficient condition for \( \hat{c} = 1 \) to be optimal is

\[
C_s(1,c\mu) \leq C_s(c,\mu), \text{ for all feasible pairs } (c,\mu).
\]

In other words, given any pair \( (c,\mu) \), the \( 1 \)-system dominates the \( c \)-system in terms of service cost rate as well as expected number in the system.

In the first section of this chapter we present a particular form for the service-cost function and sufficient conditions for (2) to hold in this case. In the second section, we investigate some conditions under which \( \hat{c} = 1 \) even though (2) is violated.
3.1 Dominance by 1-System

In this section we consider service-cost functions of the form:

\[(3) \quad c_s(c, \mu) = c c_s(\mu), \]

where \( c_s(\mu) \) is a monotone increasing function of \( \mu \geq 0 \), with \( c_s(0) = 0 \) and \( \lim_{\mu \to 0^+} c_s(\mu) \geq 0 \).

Thus we retain part of the linearity assumption: linearity of the service cost rate in the number of servers. Implicit in this assumption is that servers with equal service rates have equal cost rates, regardless of how many servers are purchased (or produced) and operated. Thus the total service cost rate of \( c \) servers, each serving at mean rate \( \mu \), is \( c \) times the cost rate of a single server serving at mean rate \( \mu \). In particular, we do not allow quantity discounts for the purchase of large numbers of servers, nor do we allow the opposite sort of situation, where the cost rate per server increases as the number of servers is increased.

The cost rate of each server may include a setup-type of component \( (\lim_{\mu \to 0^+} c_s(\mu) > 0) \) which is incurred whenever a server is purchased and operated, regardless of the rate at which it is operated.

In this section we will assume that the feasible sets for \( c \) and \( \mu \) have the form:

\[
 c = 1, 2, \ldots ; \quad 0 < \mu < \infty .
\]

In this case, condition (2) becomes
(4) \[ C_0(c\mu) \leq cC_0(\mu), \quad \mu > 0, \quad c = 1, 2, \ldots \]

We present several theorems which give sufficient conditions for (4) and tests to determine if these sufficient conditions are met. Relevant references for this material are Bruckner and Ostrow [3], Bruckner [3a], [3b], and Rosenbaum [27].

We will need the following definitions and a lemma about non-negative, monotone increasing functions \( f \), with \( f(0) = 0 \), defined on the non-negative real line.

**DEFINITION:** \( f \) is called **concave** if, for all \( x, y \geq 0 \) and \( 0 \leq \alpha \leq 1 \),

\[ f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha)f(y) \]

**DEFINITION:** \( f \) is called **quasi-homogeneous** if, for all \( x \geq 0 \) and all \( \beta \geq 1 \),

\[ f(\beta x) \leq \beta f(x) \]

(Cf. Rosenbaum [27].) When the direction of the inequality is reversed, \( f \) is called **starshaped** (cf. Bruckner and Ostrow [3]).

**DEFINITION:** \( f \) is called **subadditive** if, for all \( x, y \geq 0 \),

\[ f(x+y) \leq f(x) + f(y) \]

**LEMMA 3.1:** (cf. Bruckner and Ostrow [3], Rosenbaum [27]): \( f \) is concave \( \implies \) \( f \) is quasi-homogeneous \( \implies \) \( f \) is subadditive.
Proof.
Suppose $f$ is concave and let $x \geq 0$ and $\beta \geq 1$. Then $\beta x \geq 0$, $0 \leq 1/\beta \leq 1$ and, by the definition of concavity, we have:

$$f((\frac{1}{\beta})\beta x + (1 - \frac{1}{\beta})0) \geq \frac{1}{\beta} f(\beta x) + (1 - \frac{1}{\beta})f(0).$$

Therefore, since $f(0) = 0$,

$$f(x) \geq \frac{1}{\beta} f(\beta x),$$

and hence $f$ is quasi-homogeneous.

Suppose $f$ is quasi-homogeneous and $x, y > 0$ are given. Let $\beta_1 = 1 + \frac{y}{x}$, $\beta_2 = 1 + \frac{x}{y}$. Then $\beta_1 \geq 1$ and we can apply the definition of quasi-homogeneity:

(5) \hspace{1cm} xf(x+y) = xf(1+ \frac{y}{x}x) \\
\hspace{1.5cm} = xf(\beta_1 x) \\
\hspace{1.5cm} \leq x\beta_1 f(x) \\
\hspace{1.5cm} = (x+y)f(x).$

Similarly, $\beta_2 \geq 1$ and therefore

(6) \hspace{1cm} yf(x+y) = yf((1+ \frac{x}{y})y) \\
\hspace{1.5cm} = yf(\beta_2 y) \\
\hspace{1.5cm} \leq y\beta_2 f(y) \\
\hspace{1.5cm} = (x+y)f(y).$
Therefore, adding (5) and (6) and dividing by \( x + y \),

\[ f(x+y) \leq f(x) + f(y) \]

If either \( x \) or \( y \) is 0, then the condition for subadditivity is satisfied trivially, since \( f(0) = 0 \).

This completes the proof of Lemma 3.1.

Q.E.D.

Now \( C_s(\cdot) \) is a non-negative, monotone-increasing function with \( C_s(0) = 0 \); therefore Lemma 3.1 applies to \( C_s(\cdot) \). The following theorem shows that a sufficient condition for \( C_s(\cdot) \) to satisfy (4) is that \( C_s(\cdot) \) is concave, quasi-homogeneous, or subadditive.

**THEOREM 3.2:** If \( C_s(\cdot) \) is concave, quasi-homogeneous, or subadditive, then

(4) \[ C_s(c\mu) \leq cC_s(\mu), \; \mu > 0, \; c = 1,2,\ldots \]

**Proof.**

By Lemma 3.1, it suffices to show that \( C_s(\cdot) \) satisfies (4) if \( C_s(\cdot) \) is subadditive.

For \( c = 1 \), (4) is satisfied trivially for all \( \mu > 0 \). Let \( c > 1 \) and assume (4) is true for \( c - 1 \) and all \( \mu > 0 \). Let \( \mu > 0 \). By subadditivity,

\[ C_s(c\mu) = C_s(\mu+(c-1)\mu) \]

\[ \leq C_s(\mu) + C_s((c-1)\mu) \]
By the induction assumption,

\[ C_g((c-1)\mu) \leq (c-1)C_g(\mu) . \]

Therefore,

\[ C_g(c\mu) \leq C_g(\mu) + (c-1)C_g(\mu) \]

\[ = cC_g(\mu) \]

Q.E.D.

Theorem 3.2 gives three sufficient conditions for (4). Quasi-
homogeneity is of particular interest for two reasons: (i) it is a
natural strengthening of condition (4) (i.e., we require (4) to hold
for all real \( c \geq 1 \), rather than just integral values); (ii) it has
a useful economic interpretation. The economic interpretation can best
be appreciated after we have proved the following theorem.

THEOREM 3.3: A monotone increasing function \( f \), defined for \( x \geq 0 \)
and differentiable for \( x > 0 \), with \( f(0) = 0 \), is quasi-homogeneous
if and only if one of the following equivalent conditions holds:

(7) \( \frac{f(x)}{x} \) is a monotone decreasing function of \( x > 0 \);

(8) \( f'(x) \leq \frac{f(x)}{x} \) for all \( x > 0 \).

Proof.

First we show that condition (7) is equivalent to the defining
condition for quasi-homogeneity. Suppose \( f \) is quasi-homogeneous and
let \( 0 < x < y \). Then \( y = \beta x, \beta > 1 \). Therefore,
\[ f(y) \leq \beta f(x), \quad \text{and} \]
\[ \frac{f(y)}{\beta x} \leq \frac{f(x)}{x}, \quad \text{and} \]
\[ \frac{f(y)}{y} \leq \frac{f(x)}{x}, \]

which proves condition (7). Conversely, suppose (7) holds and let \( x > 0 \) and \( \beta \geq 1 \). Then

\[ \frac{f(\beta x)}{\beta x} \leq \frac{f(x)}{x}, \quad \text{and} \]
\[ f(\beta x) \leq \beta f(x). \]

If \( x = 0 \), \( f(\beta x) \leq \beta f(x) \) trivially since \( f(0) = 0 \). Therefore \( f \) is quasi-homogeneous.

We will now show that (7) \( \iff \) (8). Since \( f \) is differentiable, \( \frac{f(x)}{x} \) is monotone decreasing in \( x > 0 \) \( \iff \) \( \frac{d}{dx} (\frac{f(x)}{x}) \leq 0 \), for all \( x > 0 \) \( \iff \) \( \frac{xf'(x) - f(x)}{x^2} \leq 0 \), for all \( x > 0 \) \( \iff \) \( \frac{f'(x)}{x} \leq \frac{f(x)}{x^2} \), for all \( x > 0 \) \( \iff \) \( f'(x) \leq \frac{f(x)}{x} \), for all \( x > 0 \), which is condition (8).

Q.E.D.

When applied to the service-cost-rate function, \( C_s(\mu) \), condition (7) may be called the criterion of decreasing average cost. In many practical applications where the servers are machines, the cost of operating a server is (at least approximately) directly proportional to the cost of manufacturing the server. In such cases decreasing average cost is a reflection of an economy of scale in the manufacturing process. If the firm manufacturing the server were in the classical
economic situation of minimizing average manufacturing cost (per unit of service rate), then under economy of scale it would produce an arbitrarily large server. Often in the economics literature economy of scale is defined by condition (8), which states that the marginal cost is never greater than the average cost.

Although condition (8) (or either of its two equivalents) is stronger than condition (4), or even subadditivity, it may be easier to verify in practice because of its economic interpretation. There may be technological reasons, known to the designer of a queueing system, why the marginal cost should always be lower than the average cost. In fact, the defining condition for concavity — monotone decreasing marginal cost — may be even easier to justify on technological grounds. By contrast, the weaker criteria — subadditivity and condition (4) — do not seem to have useful economic interpretations. Moreover, as we shall see, these criteria are difficult to verify graphically, at least for arbitrary functions.

In [3b] Bruckner presents tests for the superadditivity of functions.* He assumes continuity of the functions involved. We present below the subadditive analogues of several of his theorems, for functions \( f \) belonging to the class \( \mathcal{PC} \) defined below:

**DEFINITION:** The class \( \mathcal{PC} \) consists of all monotone increasing piece-wise continuous functions \( f \) defined on the non-negative real line with \( f(0) = 0 \). (We adopt the convention that \( f \) is left continuous

* A function \( f \) is called **superadditive** if, for all \( x, y \), \( f(x+y) \geq f(x) + f(y) \).
at points of discontinuity.)

The following definitions are based on those in Bruckner [3b].

**DEFINITION:** A function \( f \in \text{PC} \) is called \textbf{subadditive} on \([0,a]\) if, for all \( x, y \) such that \( x + y \leq a \),

\[
f(x+y) \leq f(x) + f(y) .
\]

If \( f \in \text{PC} \) is subadditive on \([0,a]\), and \( f_1 \in \text{PC} \) are subadditive on \([0,a_i]\), \( i = 1, 2, \ldots, n \), where \( a_1 + \cdots + a_n = a \), then \( f_1, \ldots, f_n \) are a \textbf{subadditive decomposition} of \( f \) (denoted \( f = f_1 \wedge f_2 \wedge \cdots \wedge f_n \)) if

\[
f(x) = f_1(x), \quad 0 \leq x \leq a_1
\]

\[
f(x) = f(a_1) + f_2(x-a_1), \quad a_1 < x \leq a_1 + a_2
\]

\[
\vdots
\]

\[
f(x) = f(\sum_{i=1}^{n-1} a_i) + f_n(x-\sum_{i=1}^{n-1} a_i), \quad \sum_{i=1}^{n-1} a_i < x \leq a .
\]

**DEFINITION:** Let \( f \) be subadditive on \([0,a]\). Then \( F \) is called a \textbf{maximal subadditive extension} of \( f \) to \([0,\infty)\) if

(i) \( F \equiv f \) on \([0,a]\),

(ii) \( F \) is subadditive on \([0,\infty)\),

(iii) if \( g \) is any other function satisfying (i) and (ii), then \( g \leq F \).

**DEFINITION:** A function \( f \in \text{PC} \) defined on \([0,a]\) is called \textbf{concavo-convex} if \( f \) is continuous on \((0,a]\) and there exists a number \( b \),
0 \leq b \leq a$, such that $f$ is concave on $[0,b]$ and convex on $[b,a]$.

THEOREM 3.4: Let $f$ be concavo-convex on $I = [0,a]$. Then a necessary and sufficient condition for $f$ to be subadditive on $[0,a]$ is that

$$\min_{x \in I} (f(x) + f(a-x)) \geq f(a).$$

Proof.

The proof is the exact analogue of the proof of Theorem 3 in Bruckner [3b] for continuous convexo-concave and superadditive functions. The continuity of $f$, which he assumes, is never required.

Let $T = \{(x,y) | x \geq 0, y \geq 0, x + y \leq a\}$ and let $g(x,y) = f(x) + f(y) - f(x+y)$. We wish to show that $g(x,y) \geq 0$ for all $(x,y) \in T$. Condition (9) insures that $g(x,y) \geq 0$ on the line $x + y = a$; $g(x,0) = g(0,y) = 0$, since $f(0) = 0$.

If we hold $x$ fixed at any value $x \in [0,a]$, then $g$ is either increasing, decreasing, or first increasing and then decreasing as $y$ increases in $[0,x-a]$. (The possibility of a jump at 0 does not affect this property, since $f$ is concave on the closed subinterval $[0,b]$.) The same is true of $x$ with $y$ held fixed. Therefore, the minimum of $g(x,y)$ in $T$ occurs on the boundary of $T$. But $g(x,y) \geq 0$ at every point on the boundary of $T$. Therefore, $g(x,y) \geq 0$ for all $(x,y) \in T$.

Q.E.D.

COROLLARY 3.5: Suppose $f \in PC$ is a single-step function defined on $[0,a]$, with the step occurring at $x = 0$; i.e., $f(0) = 0$, $f'(x) = a$,

67
for \( x \in (0,a) \). Then \( f \) is subadditive on \([0,a]\).

Proof.

Condition (9) of Theorem 3.4 is satisfied trivially, since

\[
f(0) + f(a) = f(a) + f(0) = f(a),
\]

and, for all \( x \in (0,a) \),

\[
f(x) + f(x-a) = d + d
\]

\[> d\]

\[= f(a) .\]

Q.E.D.

THEOREM 3.6: Let \( f \) be subadditive on \([0,a]\). Then the maximal subadditive extension \( F \) of \( f \) is given by

(10) \[
F(x) = \inf \sum_{i} f(u_i), \quad x \geq 0 ,
\]

where the infimum is taken over all sequences \((u_1, u_2, \ldots, u_n)\), of all lengths \( n \), such that \( \sum_{i=1}^{n} u_i = x \), and \( 0 \leq u_i \leq a, \quad i = 1, 2, \ldots, n \).

Proof.

(Cf. Bruckner [3a].)

COROLLARY 3.7: Let \( f \) be a single-step function defined on \([0,a]\) with step of height \( d \) at \( x = 0 \). Then the maximal subadditive extension \( F \) of \( f \) is given by

(11) \[
F(0) = 0 ,
\]
\[ F(x) = nd, \text{ where } (n-1)a < x \leq na. \]

That is, \( F \) is a step function with constant step height \( d \) and constant step interval \( a \).

**Proof.**

The proof is by induction on \( n \). The case \( n = 1 \) is immediate, since the maximal subadditive extension of a function \( f \) must agree with \( f \) on \([0,a]\).

Let \( n > 1 \) and assume that \( F \) has been shown to be the maximal subadditive extension of \( f \) on \([0,(n-1)a]\). Then, for an \( x \),

\[ (n-1)a < x \leq na, \inf \sum_{1} f(u_{1}) \text{ is clearly attained by the sequence } (u_{1}, u_{2}, \ldots, u_{n}), \text{ where } u_{1} = u_{2} = \cdots = u_{n-1} = a, \ u_{n} = x - (n-1)a. \]

Hence

\[ F(x) = (n-1)f(a) + f(x-(n-1)a) \]

\[ = (n-1)d + d \]

\[ = nd. \]

Q.E.D.

**THEOREM 3.8:** Let \( f_{1}, \ldots, f_{n} \) be subadditive on \([0,a_{1}], \ldots, [0,a_{n}]\), respectively, and let \( f = f_{1} \wedge \cdots \wedge f_{n} \) on \([0,a_{1} + \cdots + a_{n}]\). Let \( F_{k} \) be the maximal subadditive extension of \( f_{k}, \ k = 1, \ldots, n \). Then \( f \) is subadditive on \([0,a_{1} + \cdots + a_{n}]\) if \( f_{k} \wedge \cdots \wedge f_{n} \leq F_{k} \), \( k = 1, \ldots, n \).

**Proof.**

(Cf. Bruckner [3b].)
COROLLARY 3.9: Let $f \in \text{PC}$ be an $n$-step function defined on $[0, a_1 + \cdots + a_n]$ with steps at $0$, $a_1$, $a_1 + a_2$, ..., $\sum_{i=1}^{n} a_i$. That is, $f = f_1 \wedge \cdots \wedge f_n$, where $f_i$ is a single-step function on $[0, a_i]$ with a step of height $d_i$ at $0$, $i = 1, \ldots, n$. Then $f$ is subadditive on $[0, a_1 + \cdots + a_n]$ if the step heights $\{d_1, \ldots, d_n\}$ are a monotone decreasing sequence and the step intervals $\{a_1, \ldots, a_n\}$ are a monotone increasing sequence.

Proof.

An immediate consequence of Corollary 3.7 and Theorem 3.8.

Q.E.D.

In practice one might often expect the service-cost function $C_g(\mu)$ to be alternately concave and convex as $\mu$ increases. In such cases, each concavo-convex segment can be tested for subadditivity using Theorem 3.4 and then the function as a whole can be tested for subadditivity, at least in principle, over any finite interval containing only a finite number of concavo-convex segments, using Theorem 3.8.

For arbitrary PC functions $C_g(\cdot)$, this process may be tedious, as it involves calculating the maximal subadditive extension of each concavo-convex segment of $C_g(\cdot)$. Even if the process can be carried out, the test is only for a sufficient condition for subadditivity; a function might fail the test and still be subadditive.

For PC step functions, a much stronger and more efficient test is available. This test will determine whether or not a PC step function satisfies (4) over any finite interval which contains only a finite number of steps.
THEOREM 3.10: Let $f$ be a PC $n$-step function defined on $[0,a_1 + \cdots + a_n]$ with a step of height $d_{i1}$ at 0 and $d_i$ at $\sum_{j=1}^{i-1} a_j$, $i = 2, \ldots, n$. Let $A_0 = 0$ and $A_i = \sum_{j=1}^{i} a_j$, $D_i = \sum_{j=1}^{i} d_j$, for $i = 1, \ldots, n$. Then $f$ satisfies

$$(12) \quad f(cx) \leq cf(x) ,$$

for $c = 1, 2, \ldots$, and $cx \in [0, A_n]$ if and only if the graph of $f(x)$, for $x \in [0, A_n]$, does not intersect the set

$$(13) \quad \bigcup_{i=1}^{n} \bigcup_{k=1}^{\infty} \{(x, y) | x < kA_i , y > kD_i \} .$$

Proof.

For fixed $i$, the set

$$\bigcup_{k=1}^{\infty} \{(x, y) | x < kA_i , y > kD_i \}$$

is the region in the non-negative quadrant which lies above the maximal subadditive extension $G_i$ of the function $g_i$ defined on $[0, A_i]$ by $g_i(0) = 0$ and $g_i(x) = D_i$, $x \in (0, A_i]$. As can be easily verified, the maximal subadditive extension of $g_i$ is also the maximal extension of $g_i$ satisfying (12). Therefore, the graph of $f(x)$ must lie on or below $G_i$ for $x \in [0, A_n]$, for all $i = 1, \ldots, n$. But this is clearly

* Note that $g_i$ is not one of the one-step functions in the decomposition of $f$. 

71
a sufficient condition for (12), since it ensures that $f(cx) \leq cg_1(x) = cf(x)$, for $x \in (A_{i-\frac{1}{2}}, A_i]$, for each $i = 1, \ldots, n$.

Q.E.D.

The test based on Theorem 3.10 may be simply described as follows: For each $i$, $i = 1, \ldots, n$, draw a line from the origin of the plane through the point $(A_i, D_i)$ on the graph of $f(x)$ and mark off on the line the points $(2A_i, 2D_i), (3A_i, 3D_i), \ldots, (mA_i, mD_i)$, as long as $mA_i \leq A_n$. Connect these points to form the step function with constant step height $D_i$ and constant step interval $A_i$. The function $f(x)$ satisfies (12) if and only if its graph lies below each of these step functions, for $i = 1, \ldots, n$, and for $x \in [0, A_n]$. This method is illustrated in Figure 1. A shaded area indicates where the graph of $f(x)$ lies above one of the step functions and hence violates (12).

3.2 Non-Dominance by l-System

We now consider service-cost functions which do not meet the condition for dominance of the $c$-system by the $l$-system for all pairs $(c, \mu)$. That is, there exists a pair $(c, \mu)$ such that $C_s(c, \mu) < C_s(l, c\mu)$. In this case we are interested in specifying conditions under which $c = 1$ still holds, that is, where the advantage of lower waiting cost in the single-server system is sufficient to overcome the disadvantage of higher service cost.

We will consider service cost functions of the form

$$(14) \quad C_s(c, \mu) = c(C_{s1} + C_{s2} \mu^a), \quad a > 1$$

72
FIGURE 1

Test for Step Functions
It will be more convenient to express the costs in terms of $c$ and $\rho$, where $\rho = \frac{1}{c \mu}$. (We assume that $\lambda$, the mean arrival rate, is 1; equivalently, we measure time in units of mean interarrival time.) All costs will be expressed in units of waiting cost; therefore $C_w = 1$.

Let $\hat{\mu}_i$ and $\hat{\rho}_i$ denote the optimal values of $\mu$ and $\rho$, respectively, when the number of servers is fixed at $i$, $i = 1, 2, \ldots$.

Then, redefining the notation accordingly, we have the following form for the objective function:

\begin{equation}
E[C(c, \rho)] = C_s(c, \rho) + E[L(c, \rho)] \\
= c(C_{s1} + C_{s2} \left( \frac{1}{c \rho} \right)^{a}) + E[L(c, \rho)] .
\end{equation}

We will consider two cases: $C_{s1} = 0$ and $C_{s1} > 0$.

3.2.1 Case I: $C_{s1} = 0$

In this case the objective function becomes

\begin{equation}
E[C(c, \rho)] = C_{s2} \frac{c}{(c \rho)^{a}} + E[L(c, \rho)] .
\end{equation}

Rather than investigate directly the conditions under which $E[C(1, \hat{\rho}_1)] \leq E[C(c, \hat{\rho}_c)]$, and hence $\hat{c} = 1$, we will consider a sufficient condition for this to be true, namely,

\begin{equation}
E[C(1, \hat{\rho}_c)] \leq E[C(c, \hat{\rho}_c)] , \quad c = 2, 3, \ldots .
\end{equation}

Combining (16) and (17), the sufficiency criterion becomes
(18) \[ E[L(1,\widehat{\rho}_c)] - E[L(1,\widehat{\rho}_c)] \geq \frac{C_s^2}{(\widehat{\rho}_c)^a} (1 - \frac{1}{c^{a-1}}), \quad c = 2, 3, \ldots . \]

The values of \( E[L(c,\rho)] \) have been calculated and tabulated for \( M|M|c \) systems: in the appendix, for \( c = 1, 2, \ldots, 10 \), and \( \rho = 0.01, 0.02, \ldots, 0.99 \). These calculations are summarized in Figure 2 (cf. Morse [23]).

It can be observed from Figure 2 that, although the 1-system dominance proved in Chapter 2 is demonstrated, the percentage difference between \( E[L(1,\rho)] \) and \( E[L(c,\rho)] \) is quite small, especially for small \( \rho \) and \( c \). Thus there is very little leeway for a large difference in service cost and we expect that \( \widehat{c} = 1 \) will hold only for values of \( c \) close to 1.

It can also be observed from the figure and from the Appendix that, if we let \( \Delta(c,\rho) \equiv E[L(c,\rho)] - E[L(1,\rho)] \), then \( \Delta(c,\rho) \) is an increasing function of \( \rho \) for any \( c \). Also, if we denote the value of \( \widehat{\rho}_c \) as a function of \( a \) by \( \widehat{\rho}_c(a) \), then \( \widehat{\rho}_c(a) \) is an increasing function of \( a \). This is intuitively clear, since higher values of \( a \) mean higher penalties for small values of \( \rho \).

From these observations we deduce that \( \Delta(c,\widehat{\rho}_c(a)) \geq \Delta(c,\widehat{\rho}_c(1)) \).

Now \( \frac{1}{\rho^a} \) is a decreasing function of \( \rho \) for \( a > 1 \), therefore

\[
\frac{C_s^2}{(\widehat{\rho}_c(1))^a} (1 - \frac{1}{c^{a-1}}) \geq \frac{C_s^2}{(\widehat{\rho}_c(a))^a} (1 - \frac{1}{c^{a-1}}).
\]

Hence, to prove (18) for \( M|M|c \) systems, it suffices to show that

\[
(19) \quad \Delta(c,\widehat{\rho}_c(1)) \geq \frac{C_s^2}{(\widehat{\rho}_c(1))^a} (1 - \frac{1}{c^{a-1}}).
\]
FIGURE 2

Graph of \( E[L(c, \rho)] \)
We will solve for the maximum value of \( a \) such that (19) holds, in the special case: \( c = 2, \ C_{s^2} = 1 \). From the appendix, we see that \( \hat{\rho}_2(1) \approx \frac{1}{2} \), and \( \Delta(2, \frac{1}{2}) \approx 0.33 \). Therefore, (19) becomes

\[
2^a(1 - \frac{1}{2^a - 1}) \leq 0.33, \quad \text{or}
\]

\[
2^a - 2 \leq 0.33, \quad \text{or}
\]

\[
2^a \leq 2.33, \quad \text{or}
\]

\[
a \leq 1.22.
\]

This example tends to confirm the prediction that the only values of \( a \) for which \( \hat{c} = 1 \) are those which are close to \( a = 1 \).

However, (18) and (19) are only sufficient conditions for \( \hat{c} = 1 \), and therefore the inequality in (20) is not sharp. To demonstrate exactly that there are relatively low values of \( a \) for which \( \hat{c} \neq 1 \), we have numerically calculated the exact optima (to two decimal places) for \( a = 2 \). The results are \( E\{C(1, \hat{\rho}_1(2))\} = 4.23 \), and \( E\{C(2, \hat{\rho}_2(2))\} = 3.23 \), so that \( \hat{c} \neq 1 \).

3.2.2 Case II: \( C_{s^1} > 0 \)

For this case we make only the following observation.

For the last example of the previous section, if \( C_{s^1} > 0 \), then \( E\{C(1, \hat{\rho}_1(2))\} = C_{s^1} + 4.23 \), and \( E\{C(2, \hat{\rho}_2(2))\} = 2C_{s^1} + 3.23 \). Therefore in order to have \( \hat{c} = 1 \) for \( a = 2 \), it suffices that \( C_{s^1} \geq 1.00 \).
CHAPTER IV

Single-Station Models with Non-Linear Waiting Cost

4.1 Introduction

In Chapter II it was shown that, for Type I-A queueing systems with degenerate, exponential, or Erlang service-time distribution, the optimal value of \( \hat{c} \) is \( \hat{c} = 1 \) when the objective function to be minimized takes the form

\[
E(C(c, \mu)) = C_s(c, \mu) + C_w E[L(c, \mu)].
\]

In Chapter III this result was extended to all objective functions of the form

\[
E(C(c, \mu)) = C_s(c, \mu) + C_w E[L(c, \mu)],
\]

where \( C_s(\cdot, \cdot) \) satisfies

(1) \( C_s(1, c\mu) \leq C_s(c, \mu) \), for all feasible \( c \) and \( \mu \).

In this chapter we investigate the conditions under which the result, \( \hat{c} = 1 \), extends to more general functions for the steady-state waiting-cost rate, i.e., where the objective function takes the form

(2) \( E(C(c, \mu)) = C_s(c, \mu) + E[C_w(c, \mu)] \),

where \( C_s(\cdot, \cdot) \) satisfies (1). Obviously then a sufficient condition on \( C_w(\cdot, \cdot) \) for \( \hat{c} = 1 \) is

(3) \( E(C_w(1, c\mu)) \leq E(C_w(c, \mu)) \), all feasible \( c, \mu \).
Following the example of Chapter II, a natural approach to take in order to prove (3) is to try to prove the analogue of Theorem 2.1 for non-linear waiting costs:

\[ C_{w,t}^a(l, c, \mu) \subseteq C_{w,t}^a(c, \mu), \ t \geq 0, \]  
for all arrival streams \( a \) and all feasible pairs \( (c, \mu) \),

where \( C_{w,t}^a(\cdot, \cdot) \) denotes the waiting-cost rate at time \( t \), given \( a \), as a function of the decision variables. In many cases of non-linear waiting costs, the function \( C_{w,t}^a(\cdot, \cdot) \) is too complicated to permit a direct proof of (4). However, we can prove (4) directly for the special case where the waiting-cost rate is a monotone-increasing function of the number of customers in the system, as the following theorem will demonstrate:

**THEOREM 4.1:** Suppose the service-time distribution in a queueing system of Type \( \cdot-I-A \) is degenerate, exponential, or \( k \)-Erlang \( (k = 1, 2, \ldots) \). Let \( a \) be an arbitrary realization of the arrival process and let

\[ C_{w,t}^a(c, \mu) = (C_{w,t}^a(c, \mu) | A = a), \]

i.e., \( C_{w,t}^a(c, \mu) \) is the waiting-cost rate of the system at time \( t \), given \( a \), as a function of the decision variables. Suppose that

\[ C_{w,t}^a(c, \mu) = h(L_{t}^a(c, \mu)), \]

where \( h(\cdot) \) is a monotone increasing Borel function. Then, for any feasible pair \( (c, \mu) \),

\[ C_{w,t}^a(l, c, \mu) \subseteq C_{w,t}^a(c, \mu), \ t \geq 0. \]
Proof.

The proof is an immediate consequence of Theorem 2.1 and Property 2° of stochastic ordering. Q.E.D.

This sort of function for the waiting-cost rate is encountered, for example, in applications where the cost rate due to customers waiting at any given time is caused by the space that they take up. It is a natural sort of function for situations where the cost of waiting is a measure of the inconvenience suffered by the service mechanism and its environment, rather than the inconvenience suffered by the customers. The latter case, however, is probably more common, and it is to this case that we turn our attention in the remainder of this chapter.*

For the customer-inconvenience type of waiting-cost rate, the number in the system \( L_t^a(c, \mu) \) is typically not a rich enough state description for the purpose of determining the waiting-cost rate of the system at time \( t \), given arrival stream \( a \). The waiting-cost rate, \( C_w^a(t, c, \mu) \), will generally depend on the history of the system up until time \( t \). Specifically, in most cases the system waiting-cost rate will be the sum of the waiting-cost rates for each of the customers in the system, and the customer waiting-cost rate will be a function of the time a customer has been in the system.

In such cases, an augmented state space is required. A new state description sufficient for determining the system waiting-cost rate at

* Note that the linear waiting-cost model studied in Chapter II can be viewed as a special case of both the mechanism-inconvenience and the customer-inconvenience models.
time \( t \) would be the vector consisting of the number of customers in the system and the expended waiting time in the system of each of these customers. Thus, let

\[
N^a_t(c, \mu) = (m; T_1, T_2, \ldots, T_m),
\]

where \( m = L^a_t(c, \mu) \), and \( T_1 \geq T_2 \geq \ldots \geq T_m \) are the expended waiting times at \( t \) of the customers in the system. Suppose the total waiting cost of a customer who has been in the system a length of time \( T \) is \( C_w h(T) \), where \( h \) is a monotone increasing differentiable Borel function. Then the system waiting-cost rate is a well-defined function of each element of the new state vector \( N^a_t(c, \mu) \):

\[
C^a_{w,t}(c, \mu) = \sum_{i=1}^{m} C_w h'(T_i).
\]

By analogy with Theorem 2.1, we may hope to be able to prove

\[
N^a_0(1, \mu) \subseteq N^a_t(c, \mu), \text{ all feasible pairs } (c, \mu), \text{ and all } t \geq 0.
\]

If (7) is true and \( C^a_{w,t}(c, \mu) \) is a monotone increasing function of each element of \( N^a_t(c, \mu) \), then we can deduce (4) using properties of stochastic ordering. But it can be seen from (6) that \( h'(\cdot) \) must be monotone increasing (i.e., \( h(\cdot) \) must be convex) in order for

\[
C^a_{w,t}(c, \mu)
\]

to be a monotone increasing function of each element of \( N^a_t(c, \mu) \).

Even with the assumption that \( h(\cdot) \) is convex, this method of inclusion of supplementary variables in the state description does not lead readily to a proof of (4). However, we are able to prove a result analogous to (4) for the total waiting cost of each customer:
(8) \[ E(h(W_n^a(1,c,\mu))) \leq E(h(W_n^a(c,\mu))), \quad n = 1, 2, \ldots, \]
where \( W_n^a(c,\mu) \) is the waiting time of the \( n \)th arrival in the \( c \)-system with arrival stream \( a \), and \( W_n^a(1,c,\mu) \) is the waiting time of the \( n \)th arrival in the \( l \)-system with arrival stream \( a \). The relation (8) is proved in Section 4.3 under the hypotheses of Theorem 2.1 (i.e., for Type I-A systems with arbitrary arrival stream \( a \) and degenerate, exponential, or Erlang service-time distribution) when \( h(\cdot) \) is convex, monotone increasing, and satisfies a regularity condition. We prove (8) by demonstrating that, given any number \( m = 0,1,2,\ldots, \) in the system when a customer arrives, his expected waiting cost is lower in the \( l \)-system. We then use the stochastic ordering of the number in the system (proved in Theorem 2.1) and Properties 1° and 2° to prove (8).

In Section 4.2 we prove that:

(9) \[ E(C_w(c,\mu)) = \lambda E(C_w(h(W(c,\mu)))) , \]
for all systems satisfying (6) and the assumptions of Jewell [15].

This is an extension of the familiar result, \( E(L) = \lambda E(W) \), proved by Little [20] and Jewell. Relations (6), (8), and (9) together imply that \( \hat{C} = 1 \) for Type I-A systems with degenerate, exponential, or Erlang service-time distribution and objective function given by (2), where \( C_s(\cdot,\cdot) \) satisfies (1) and \( h(\cdot) \) is convex, differentiable, and monotone increasing.

While the convexity of \( h(\cdot) \) is necessary for the method of proof of (8) used in Section 4.3, it is not clear that it is necessary for (8) to be true. The proof in Section 4.3 does not take advantage of
the fact that the stochastic ordering, $L^a_t(l, c, \mu) \preceq L^a_t(c, \mu)$, is generally strict and may be sufficiently strict to compensate for any possible opposite ordering of the conditional waiting costs, even for highly non-convex monotone increasing waiting-cost functions.

The results of Section 4.4 suggest that this is true. There a proof is given that, if the arrivals are from a renewal process and the service-time distribution is exponential (GI/M/c system), then (8) holds in the steady-state for arbitrary monotone increasing $h$. The proof is algebraic, based on explicit expressions for the steady-state waiting-time distributions in the l-system and c-system. It lacks the intuitive appeal or the generality of a proof of (8) under the hypotheses of Theorem 2.1, but it suggests that such a proof may eventually be found.

4.2 Generalization of $E(L) = \lambda E(W)$

The objective of this section is to show, under very general assumptions about the arrival and waiting-time processes of a queueing system, that in the steady state

(10) $E[\text{system waiting-cost rate}] = \lambda \times E[\text{waiting cost of a customer}]$.

For the case considered in Chapter II, where the system waiting-cost rate is linear in the number in the system and each customer's total waiting cost is linear in his waiting time in the system, (10) reduces to the formula $E(L) = \lambda E(W)$, proved by Little [20] and Jewell [15].

Following Jewell but using our own notation, we define the queueing system by three stochastic processes:
\( L_t = \text{number in system at time } t \in [0, \infty) \)

\( A_i = \text{arrival time of } i^{th} \text{ customer}, \ i = 0, 1, \ldots \)

\( W_i = \text{waiting time in system of } i^{th} \text{ arrival}, \ i = 0, 1, \ldots \)

For an arbitrary busy cycle starting at time 0, let

\( v = \min \{ n > 0 | L_{A_n} = 0 \}, \ \text{the number of customers served during a busy cycle,} \)

\( \gamma = A_v, \ \text{the duration of a busy cycle,} \)

\( \beta = \text{departure time of last customer to depart from system (duration of busy period),} \)

\( \ell = \gamma - \beta, \ \text{the duration of idle period prior to arrival of } v^{th} \text{ customer.} \)

In addition we define two cost processes:

\( g(t) = \text{system waiting-cost rate at time } t \in [0, \infty) \)

\( C_w(W_i) = \text{total waiting cost for } i^{th} \text{ customer, } i = 0, 1, 2, \ldots \),

where \( h(\cdot) \) is a differentiable, monotone increasing Borel function with \( h(0) = 0. \)

These processes are assumed to be related in the following way (cf. Little's formula for the relation between \( \{ L_t \} \) and \( \{ W_i \} \)):

\( g(t) = \sum_{i=1}^{v-1} C_w'(t-A_i) \delta(t-A_i) \delta(A_i + W_i - t), \)

where

\( \delta(x) = \begin{cases} 0, & \text{if } x \leq 0, \\ 1, & \text{if } x > 0. \end{cases} \)
Relation (12) is a restatement of (6) for this (possibly more general) 
queueing system.

4.2.1 Generalization of Jewell’s proof

We will now present the generalization of Jewell’s proof, using 
his assumptions. The generalization of Little’s proof is equally 
straightforward. The generalization owes its validity essentially to 
relation (12), the basis for Lemma 4.2; this lemma corresponds to 
Jewell’s Theorem 1, which he observes is the heart of his proof.

ASSUMPTION I: The event \([L_t = 0]\) is a recurrent event, for any 
given initial condition of the system.

LEMMA 4.2: For any realization of the stochastic processes (11), if 
the busy cycle starts with \(A_0 = 0\), then

\[
(13) \quad \sum_{i=1}^{\nu-1} C_w h(W_i) = \int_0^t g(u)du, \quad \beta \leq t < \gamma.
\]

Proof.

By (12), for all \(0 \leq u \leq t\),

\[
g(u) = \sum_{i=1}^{\nu-1} C_w h'(u-A_i) \delta(u-A_i) \delta(A_i + W_i - u).
\]

Therefore,

\[
\int_0^t g(u)du = \sum_{i=1}^{\nu-1} C_w \int_0^t h'(u-A_i) \delta(u-A_i) \delta(A_i + W_i - u)du.
\]

Letting \(\tau = u-A_i\) and noting that \(t-A_i \geq W_i\) for \(i = 1, \ldots, \nu-1\),
\[
\int_0^t g(u)du = \sum_{i=1}^{v-1} C_w \int_{A_i}^{t-A_i} h'(\tau) \delta(\tau) \delta(W_i - \tau) d\tau \\
= \sum_{i=1}^{v-1} C_w \int_0^{W_i} h'(\tau) d\tau \\
= \sum_{i=1}^{v-1} C_w h(W_i).
\]

Q.E.D.

ASSUMPTION II: The joint distribution of \( \{v, A_1, A_2, \ldots, A_v; W_1, W_2, \ldots, W_{v-1}; \ell \} \) is independent from one busy cycle to the next* and is identical for each busy cycle.

We do not need Jewell's Assumption III because we are not allowing the initial busy cycle to be different from the succeeding ones.

**Lemma 4.3:** \( \lim_{t \to \infty} \frac{1}{t} \mathbb{E}\left( \int_0^t g(u)du \right) = \frac{\mathbb{E}\left( \sum_{i=1}^{v-1} C_w h(W_i) \right)}{\mathbb{E}(\ell)} \).

**Proof.**

Let

\[
\Gamma_0 = 0, \quad \Gamma_j = \sum_{i=0}^{j-1} \gamma_j, \quad \text{the starting time of } j^\text{th} \text{ busy cycle},
\]

\( j = 1, 2, \ldots \),

\( \varphi(t) = \sup\{j|\Gamma_j \leq t\}, \quad \text{the number of busy cycles that start in} \quad [0,t], \)

\( \xi_j = \int_{\Gamma_j}^{\Gamma_{j+1}} g(t)dt, \quad \text{the total system waiting cost incurred in} \quad \text{the } j^\text{th} \text{ busy cycle}. \)

* Assuming we re-index the random variables and redefine the time origin for each busy cycle.
Then \( \xi_j, \ j = 1, 2, \ldots \), are independent (from Assumption II) and identically distributed as \( \sum_{i=0}^{v-1} C_w h(W_i) \), by Lemma 4.2.

Therefore, for any \( t > 0 \),

\[
\int_0^t g(u)du = \sum_{j=0}^{t-1} \xi_j + \int_\Gamma \int_0^t g(u)du.
\]

Now \( \lim_{t \to 0} \frac{1}{t} E[\int_0^t g(u)du] = 0 \), so

\[
\lim_{t \to \infty} \frac{1}{t} E[\int_0^t g(u)du] = \lim_{t \to \infty} \frac{1}{t} E[\sum_{j=0}^{t-1} \xi_j].
\]

But

\[
\lim_{t \to \infty} \frac{1}{t} E[\sum_{j=0}^{t-1} \xi_j] = \frac{E[\sum_{i=0}^{v-1} C_w h(W_i)]}{E[\gamma]},
\]

by Assumptions I, II, and the key renewal theorem (cf. Smith [32]).

Q.E.D.

ASSUMPTION IV: Under Assumption II, the customer-average means

\[
T = \frac{E[\gamma]}{E[v]}, \quad \text{and} \quad \frac{E[\sum_{i=0}^{v-1} C_w h(W_i)]}{E[v]},
\]

are both finite.

THEOREM 4.4: Suppose a queueing system defined by (11) and (12) satisfies Assumptions I, II, and IV. Then, in the steady state,

\[
E[\text{system waiting-cost rate}] = \lambda \times E[\text{waiting cost of a customer}] .
\]
Proof.

Clearly, in the steady state,

\[
E\{\text{system waiting-cost rate}\} = \lim_{t \to \infty} \frac{1}{t} E\left[ \int_0^t g(u) du \right],
\]

(14)

\[
E\{\text{waiting cost of a customer}\} = \lim_{n \to \infty} \frac{1}{n} E\left[ \sum_{i=0}^{n-1} C_w h(W_i) \right].
\]

By the key renewal theorem applied to the discrete time process, \(\{C_w h(W_i), \ i \geq 0\}\),

\[
\lim_{n \to \infty} \frac{1}{n} E\left[ \sum_{i=0}^{n-1} C_w h(W_i) \right] = \frac{E\left[ \sum_{i=1}^{v-1} C_w h(W_i) \right]}{E(v)}.
\]

(15)

(The proof of (15) is an exact analogue of the relevant part of the proof of Lemma 4.3.)

Similarly,

\[
\frac{1}{\lambda} = \lim_{n \to \infty} \frac{1}{n} E\{A_n\}
\]

\[
= \frac{E(A_v)}{E(v)}
\]

\[
= \frac{E(\gamma)}{E(v)}.
\]

(16)

Therefore, combining (14), Lemma 4.3, (16), and (15), we obtain

\[
E\{\text{system waiting-cost rate}\} = \lim_{t \to \infty} \frac{1}{t} E\left[ \int_0^t g(u) du \right]
\]

\[
= \frac{E\{ \sum_{i=1}^{v-1} C_w h(W_i) \}}{E(\gamma)}
\]

(17)
\[
E\{v\} \times \frac{\sum_{i=1}^{v-1} C_w(h(W_i))}{E\{v\}} \\
= \lambda \times \left( \lim_{n \to \infty} \frac{1}{n} E\{\sum_{i=0}^{n-1} C_w(h(W_i))\} \right) \\
= \lambda \times E\{\text{total waiting cost of a customer}\},
\]

which is our desired result. Q.E.D.

4.2.2 Remark

In many ways the basic result \( E(L) = \lambda E(W) \) is more intuitively plausible when viewed as a special case of our result for costs. The cost result says essentially that the time at which we charge the waiting cost of a customer is not too crucial; in particular there is no change in the expected system waiting-cost rate if we charge all of a customer's waiting cost at the instant of his arrival. In many practical applications, this may in fact be a more reasonable assumption than the one we have been making: that a customer charges his waiting cost to the system continuously over the time he is in the system.

4.3 Proof of Ordering of Customer Waiting Costs: Transient Case, k-Erlang Service-Time Distribution, Convex Costs

In this section and the next, our ultimate concern is a proof of

\[(3) \quad E(C_w(1,c_\mu)) \leq E(C_w(c_\mu)), \text{ all feasible pairs } (c_\mu),\]

the sufficient condition for \( \hat{c} = 1 \) for Type I-A systems when the
objective function is

\[ EC(c, \mu) = C_s(c, \mu) + E(C_w(c, \mu)), \]

where \( C_s(\cdot, \cdot) \) satisfies

\[ C_s(1, c\mu) \leq C_s(c, \mu), \text{ all feasible pairs } (c, \mu). \]

Throughout these two sections, we assume that our queueing system satisfies (6) and Jewell's axioms, so that, for all pairs \((c, \mu)\) for which a steady-state distribution of \(W(c, \mu)\) exists,

\[ E(C_w(c, \mu)) = \lambda C_w E(h(W(c, \mu))), \]

where \(h(\cdot)\) is a monotone increasing, differentiable Borel function with \(h(0) = 0\). Thus a sufficient condition for (3) is

\[ E(h(W(1, c\mu))) \leq E(h(W(c, \mu))), \text{ all feasible } (c, \mu). \]

The immediate objective of this section and the next will be to produce proofs of (19), for various assumptions about the arrival process, service-time distribution, and the function \(h(\cdot)\).

In this section, we try to parallel the approach used in Theorem 2.1 and prove the following relation, much stronger than (19):

\[ W_n^a(1, c\mu) \subseteq W_n^a(c, \mu), \quad n = 1, 2, \ldots, \]

for arbitrary arrival stream \(a\) and all feasible pairs \((c, \mu)\). This relation would imply

\[ W(1, c\mu) \subseteq W(c, \mu), \text{ all feasible pairs } (c, \mu), \]

90
where these random variables exist. We know by Property 2.0 that (21) is a necessary and sufficient condition for (19) to hold for all monotone increasing Borel functions $h(\cdot)$.

One might try to prove (20) directly, by induction on $n$. It is clear that this is equivalent (for a fixed arrival stream $a$) to proving that $U'_n \subseteq V'_n$, where $U'_n$ denotes the departure time of the $n^{th}$ arrival from the $l$-system, and $V'_n$ denotes the departure time of the $n^{th}$ arrival from the $c$-system.

Now $U'_n$ equals $U_n$, the departure time of the $n^{th}$ departure from the $l$-system, but $V'_n$ does not always equal $V_n$, the departure time of the $n^{th}$ departure from the $c$-system, because customers do not necessarily depart from the $c$-system in the same order as they arrive. In Chapter II we did not have to be concerned with the order of departure, since we were only interested in the number in the system. Therefore, for the case of exponential service-time distribution, we could consider the system with $c$ servers each with mean service rate $\mu$ as equivalent to a single-server system with instantaneously variable mean service rate, $(n \land c)\mu$, where $n$ is the number of customers in the system. In the variable-rate single-server system, of course, departures do occur in the same order as arrivals. Thus it is not an accurate model of the $c$-server system when waiting times are the explicit concern.

If we try to prove $U'_n \subseteq V'_n$ by induction on $n$, we will need a relation of the form

$$ (U'_n | U'_{n-1} = z) \subseteq (V'_n | V'_{n-1} = z) $$

analogous to Lemma 2.11 of Chapter II. However, (22) will not always
hold, because the possible shuffling of the order of departure in the c-system implies that, for some arrival streams a, there is a positive probability of the nth arrival departing before the n-1st. That is, for some $t < z$, \( \text{Prob}(V_n' > t | V_{n-1} = z) < 1 = \text{Prob}(U_n' > t | U_{n-1} = z) \);
therefore, \( (U_n' | U_{n-1} = z) \subset (V_n' | V_{n-1} = z) \).

Hence we must give up our attempts to prove (20) in this manner.
We will instead prove non-inductively the following weaker result:

**THEOREM 4.5:** Suppose the service-time distribution in a Type I-B queueing system is k-Erlang \( (k = 1, 2, \ldots) \). Let \( a \) be an arbitrary realization of the arrival process, and let \( W_n^a(c, \mu) \) denote the waiting time in the system of the nth arrival, given \( a \), as a function of the decision variables, for \( n = 1, 2, \ldots \). Let \( (c, \mu) \) be an arbitrary pair of values for the decision variables, and let \( h(\cdot) \) be a convex monotone increasing function with \( \lim_{x \to \infty} h(x)e^{-\mu x} = 0 \). Then \( (23) \)

\[
E[h(W_n^a(1, c\mu))] \leq E[h(W_n^a(c, \mu))], \quad n = 1, 2, \ldots.
\]

For the proof of this theorem we will need the following two lemmas, which will be proved later.

**LEMMA 4.6:** Let \( F(\cdot) \) and \( G(\cdot) \) be complementary c.d.f.'s of two random variables with the same mean \( \frac{1}{\eta} \). Suppose \( F(\cdot) \) crosses \( G(\cdot) \) once from above at \( t_0^* \). Let \( \phi(\cdot) \) be a monotone increasing function.

\[\text{That is, } F(t) \geq G(t), \text{ for } t \leq t_0^*; \quad F(t_0^*) = G(t_0^*); \quad F(t) \leq G(t), \text{ for } t > t_0^*. \text{ Thus } F \text{ is spread less than } G \text{ (cf. [2]).}\]
Then

\[ \int_0^\infty \varphi(x)\overline{F}(x)dx \leq \int_0^\infty \varphi(x)\overline{G}(x)dx . \]

**Lemma 4.7:** Let \( W_n^l | m \) denote \( (W_n^a (1, c\mu) | L_n^a (1, c\mu) = m) \), the conditional waiting time of the \( n \)th customer to arrive at the \( l \)-system given that there are \( m \) phases ahead of him when he arrives. Let \( (W_n^c | m) \) be correspondingly defined for the \( c \)-system. Then

\[ (W_n^l | m) \text{ is a stochastically monotone increasing function of } m , \]

\[ E[h(W_n^l | m)] \leq E[h(W_n^c | m)], \text{ for } m = 0, 1, 2, \ldots , \]

for \( h \) satisfying the conditions of the statement of Theorem 4.5.

**Proof of Theorem 4.5.**

\[ E[h(W_n^a (1, c\mu))] = \sum_{m=0}^\infty E[h(W_n^l | m)] \cdot \text{Prob}(L_n^a (1, c\mu) = m) . \]

By (25) and Property 10 of stochastic ordering, \( E[h(W_n^l | m)] \) is a monotone increasing function of \( m \). Therefore, by (27), Properties 10 and 20 and the fact that \( L_n^a (1, c\mu) \subseteq L_n^a (c, \mu) \) (from Theorem 2.1, letting \( t = a_n \)),

\[ E[h(W_n^a (1, c\mu))] \leq \sum_{m=0}^\infty E[h(W_n^l | m)] \cdot \text{Prob}(L_n^a (c, \mu) = m) . \]

Therefore, combining (28) and (26) we obtain

\[ E[h(W_n^a (1, c\mu))] \leq \sum_{m=0}^\infty E[h(W_n^c | m)] \cdot \text{Prob}(L_n^a (c, \mu) = m) \]

\[ = E[h(W_n^c (c, \mu))] . \]

Q.E.D.
We now prove the two lemmas.

**Proof of Lemma 4.6.**

\[
\int_{0}^{\infty} \varphi(x) F(x) \, dx = \int_{0}^{\infty} \varphi(x) G(x) \, dx. \quad \text{Therefore,}
\]

\[
\int_{0}^{\infty} \varphi(x) F(x) \, dx - \int_{0}^{\infty} \varphi(x) G(x) \, dx
\]

\[
= \int_{0}^{\infty} \varphi(x) [F(x) - G(x)] \, dx
\]

\[
= \int_{0}^{\infty} \varphi(x) [F(x) - G(x)] \, dx + \varphi(t_0) \left[ \int_{0}^{\infty} F(x) \, dx - \int_{0}^{\infty} G(x) \, dx \right]
\]

\[
= \int_{0}^{\infty} \varphi(x) [F(x) - G(x)] \, dx + \int_{0}^{\infty} \varphi(t_0) [F(x) - G(x)] \, dx
\]

\[
= \int_{0}^{\infty} [\varphi(x) - \varphi(t_0)] [F(x) - G(x)] \, dx.
\]

Now for \(0 \leq x \leq t_0\), \(F(x) - G(x) \geq 0\), and since \(\varphi(\cdot)\) is non-decreasing, \(\varphi(x) \leq \varphi(t_0)\). For \(t_0 \leq x < \infty\), reverse inequalities hold. Therefore, \([\varphi(x) - \varphi(t_0)] [F(x) - G(x)] \leq 0\) for all \(0 \leq x < \infty\).

Therefore,

\[
\int_{0}^{\infty} \varphi(x) F(x) \, dx - \int_{0}^{\infty} \varphi(x) G(x) \, dx \leq 0. \quad \text{Q.E.D.}
\]

A special case of this theorem appears in Barlow and Proschan [1, p. 32].

**Proof of Lemma 4.7.**

First (25) is proved.

Given \(m\) phases ahead of him when he arrives in the 1-system, the \(n\)th arrival must wait for the \(m\) phases to be served and then
for his own $k$ phases to be served. Since the service time of phases are independent and identically distributed exponentially with mean rate $k\mu$.

\begin{equation}
W_n^1|m \text{ is distributed as } S_{m+k}(k\mu),
\end{equation}

where $S_i(\eta)$ denotes the sum of $i$ independent exponentially distributed random variables each with mean rate $\eta$. It is a simple exercise using Property 2 of stochastic ordering to prove that $S_i(\eta) \leq S_{i+1}(\eta)$, $i = 1,2,\ldots$. Therefore, $(W_n^1|m)$ is stochastically non-decreasing in $m$.

We now prove (26).

We have from (30) that $(W_n^1|m)$ is distributed as $S_{m+k}(k\mu)$.

Now if the $n$th arrival to the $c$-system finds $m$ phases ahead of him, he must wait for at least $\max(0, m-(c-1)k)$ phases to depart before he can begin service. This is because the largest number of phases which can be in the system simultaneously and still leave a server free is $(c-1)k$ (i.e., one server is free and each of the $c-1$ others has a full subqueue of $k$ phases). While the $n$th arrival is waiting to be served, all $c$ servers are busy and therefore the service mechanism is ejecting phases according to a Poisson process with parameter $ck\mu$ (each server serves phases exponentially at mean rate $k\mu$). Therefore, we have

\begin{equation}
(W_n^c|m) \Rightarrow S_{m-(c-1)k}(ck\mu) + S_1(k\mu), \quad m \geq (c-1)k,
\end{equation}

\begin{equation}
(W_n^c|m) \Rightarrow S_1(k\mu), \quad 0 \leq m \leq (c-1)k.
\end{equation}
Rewriting (30), we have

\[(W_n^1|m) \text{ distributed as } S_{m-(c-1)k}(ck\mu) + S_{ck}(ck\mu), \quad m > (c-1)k,\]
\[(W_n^1|m) \text{ distributed as } S_{m+k}(ck\mu), \quad m \leq (c-1)k.\]

Therefore,

\[(W_n^1|m) \subset S_{ck}(ck\mu), \quad m \leq (c-1)k.\]

It can be seen from (31), (32), and (33) that to prove (26), it suffices to prove

\[E(h(X)) \leq E(h(Y)),\]

where \(X = S_{m-(c-1)k}(ck\mu) + S_{ck}(ck\mu)\) and \(Y = S_{m-(c-1)k}(ck\mu) + S_{K}(k\mu)\).

Let \(F(\cdot)\) be the c.d.f. of \(X\) and \(G(\cdot)\) be the c.d.f. of \(Y\). Then \(\overline{F}(\cdot)\) and \(\overline{G}(\cdot)\) satisfy the conditions of Lemma 4.6. Therefore we have

\[\int_0^\infty \varphi(x)\overline{F}(x)dx \leq \int_0^\infty \varphi(x)\overline{G}(x)dx,\]

for any monotone increasing function \(\varphi(\cdot)\). Let \(\varphi(\cdot) = h'(\cdot)\). Since \(h(\cdot)\) is assumed convex and differentiable, \(h'(\cdot)\) is a continuous, monotone increasing function. Hence we have (integrating by parts)

\[E(h(X)) - E(h(Y)) = \int_0^\infty h(x)d\overline{F}(x) - \int_0^\infty h(x)d\overline{G}(x)\]

\[= -\int_0^\infty h(x)d\overline{F}(x) + \int_0^\infty h(x)d\overline{G}(x)\]

96
\begin{align*}
&= -h(x) \overline{F}(x) \bigg|_0^\infty + \int_0^\infty \overline{F}(x) d(h(x)) + h(x) \overline{G}(x) \bigg|_0^\infty - \int_0^\infty G(x) d(h(x)) \\
&= h(x) (\overline{G}(x) - \overline{F}(x)) \bigg|_0^\infty + \int_0^\infty h'(x) \overline{F}(x) dx - \int_0^\infty h'(x) \overline{G}(x) dx.
\end{align*}

By (24), \( \int_0^\infty h'(x) \overline{F}(x) dx - \int_0^\infty h'(x) \overline{G}(x) dx \leq 0 \), and, by the assumption of Theorem 4.5 \( \lim_{x \to \infty} h(x) e^{-\mu x} = 0 \), we have \( h(x) (\overline{G}(x) - \overline{F}(x)) \bigg|_0^\infty = 0 \).

Therefore,

\[ E[h(X)] \leq E[h(Y)]. \]

This completes the proof of Lemma 4.7. \( \quad \text{Q.E.D.} \]

We have proved Theorem 4.5 for the most difficult case: \( k \)-Erlang service-time distribution. Direct proofs for exponential or degenerate service-time distributions may be constructed along similar lines; these two cases also follow by letting \( k = 1 \) and letting \( k \to \infty \), respectively.

Thus we have the following corollary:

COROLLARY 4.8: Under the hypotheses of Theorem 4.5, if the service-time distribution is exponential or degenerate with mean rate \( \mu \), then

\[ E[h(W_n^a(1,c,\mu))] \leq E[h(W_n^a(c,\mu))], \quad n = 1, 2, \ldots. \]

The following corollary is an immediate application of Theorem 4.4 and Corollary 4.8.

COROLLARY 4.9: Consider a Type I-A queueing system in which the service-time distribution is degenerate, exponential, or \( k \)-Erlang
Suppose the arrival process is a renewal process and the system waiting-cost rate at any time \( t \) is defined by

\[
C_{w,t}(c,\mu) = \sum_{i=1}^{\infty} C_w h'(t-A_i) \delta(t-A_i) \delta(A_i + W_i - t),
\]

where the total waiting cost for the \( n \)th arrival is \( C_w(W_n) \), and \( h \) is convex, differentiable, and monotone decreasing with \( h(0) = 0 \).

Suppose the feasible sets for \( c \) and \( \mu \) have the form:

\[
c = 1, 2, \ldots, \mu_0 \leq \mu < \infty, \text{ for some } \mu_0 > 0,
\]

and suppose that \( \lim_{x \to \infty} h(x) e^{-\mu_0 x} = 0 \). Then \( \hat{c} = 1 \) is optimal for

\[
E[C(c,\mu)] = C_s(c,\mu) + E[C_w(c,\mu)],
\]

where \( C_s(1,c\mu) \leq C_s(c,\mu) \) for all feasible pairs \((c,\mu)\).

The restriction that \( \mu \geq \mu_0 \) and \( \lim_{x \to \infty} h(x) e^{-\mu_0 x} = 0 \) is probably not necessary. It is very restrictive, if \( \mu \) is allowed to assume any non-negative value, for when \( \mu = 0 \), it requires \( h(x) = 0 \), all \( x \).

However, if the feasible sets have the form (35), then we know that \( \mu \geq \mu_0 > 0 \). For fixed \( \mu_0 > 0 \), the condition that \( \lim_{x \to \infty} h(x) e^{-\mu_0 x} = 0 \) is satisfied by a wide variety of convex functions \( h(\cdot) \), in particular for \( h(x) = x^N \), where \( N \) is a positive integer. Moreover, in practical applications the asymptotic behavior of \( h(x) \) as \( x \to \infty \) is of little real concern to the analyst since the probability of very large values of \( W \) is negligible in the steady state. Therefore, we can replace

\[98\]
the tail of an arbitrary monotone increasing function \( h(\cdot) \) by a tail
which satisfies
\[
\lim_{x \to \infty} h(x)e^{-\mu x} = 0.
\]

4.4 Proof of Ordering of Customer Waiting Costs: Steady-State Case, GI/M/c System, Arbitrary Monotone Increasing Differentiable Waiting-Cost Function

In this section we exhibit a stochastic ordering of the steady-state waiting-time of a customer for GI/M/c systems:

\[(36) \quad W(1,c,\mu) \preceq W(c,\mu), \text{ all } c, \mu.\]

Throughout the section we assume that \( c \) and \( \mu \) are fixed at arbitrary feasible values such that \( \rho < 1 \) and let \( W^l = W(1,c,\mu) \) and \( W^c = W(c,\mu) \). We will derive explicit expressions for \( \text{Prob}(W^l > t) \) and \( \text{Prob}(W^c > t) \) and prove

\[(37) \quad \text{Prob}(W^l > t) \leq \text{Prob}(W^c > t), \text{ for all } t \geq 0.\]

The following notation will be used (following Karlin [16]) for the GI/M/c system:

\[H(t) = \text{c.d.f. of interarrival times},\]
\[H^*(s) = \text{Laplace-Stieltjes transform of } H(t),\]
\[\lambda = \text{mean arrival rate of customers } (\lambda > 0),\]
\[\rho = \frac{\lambda}{c\mu} = \text{traffic intensity of system},\]
\[\tau_j = \text{steady-state probability that an arrival finds } j\]
\[\text{customers ahead of him in the system, } j = 1,2,\ldots,\]
\[W^c_q = \text{steady-state waiting time in queue of a customer},\]
\[W^c = \text{steady-state waiting time in system of a customer}.\]
Kendall [17] shows that the \( \pi_j \) exist if and only if \( \rho < 1 \) and that the vector \( \pi = (\pi_0, \pi_1, \ldots) \) has the form

\[
\pi = (\beta_0, \beta_1, \ldots, \beta_{c-2}, 1, \alpha, \alpha^2, \ldots),
\]

where \( \alpha \) is the unique solution of

\[
\alpha = H^*(e\mu (1-\alpha)).
\]

The values of \( \beta_j, j = 0, 1, \ldots, c-2 \), may be found by solving the first \( c \) of the stationary equations, \( \pi = \pi P \), the Markov chain imbedded at the arrival points, where \( P \) is the transition probability matrix of the imbedded chain. The value of \( A \) is

\[
A = \left( \sum_{j=0}^{c-2} \beta_j + \frac{1}{1-\alpha} \right)^{-1},
\]

since it is used as a normalizing factor.

The various waiting-time distributions are given by the following formulae:

\[
\text{Prob}(W_q^c > t | W_q^c > 0) = e^{-c\mu (1-\alpha)t},
\]

\[
\text{Prob}(W_q^c > t) = \text{Prob}(W_q^c > 0)e^{-c\mu (1-\alpha)t}
\]

\[
\text{Prob}(W_q^c > t) = \left\{ \begin{array}{ll}
  e^{-\mu t} + \text{Prob}(W_q^c > 0)(e^{-\mu t} - e^{-c\mu (1-\alpha)t}), & \text{if } c(1-\alpha)-1 \neq 0 \\
  e^{-\mu t} + \text{Prob}(W_q^c > 0)\mu t, & \text{if } c(1-\alpha)-1 = 0.
\end{array} \right.
\]

Finally we note that
\[ \text{Prob}(W_q^c > 0) = \text{Prob}(\text{an arriving customer finds } c \text{ or more ahead of him in the system}) \]

\[ = \sum_{n=c}^{\infty} A \alpha^{n-c+1} \]

\[ = A \frac{\alpha}{1-\alpha}. \]

The value of \( A \) was derived by Takacs [34]. He used generating functions to solve the first \( c \) of the stationary equations, and thereby find the value of \( A \). Takacs proved that

\[ A^{-1} = \alpha \sum_{i=0}^{c} \binom{c}{i} \frac{D_i F_i}{1 - G_i} + \frac{\alpha}{1 - \alpha}, \]

where

\[ G_i = H^*(i\mu), \quad i = 1, 2, \ldots, c, \]

\[ D_i = \left\{ \begin{array}{ll} 1, & i = 0 \\ \frac{i}{1 - \left( \frac{l - G_k}{G_k} \right)^{i-k}}, & i = 1, 2, \ldots, c, \end{array} \right. \]

\[ F_i = \frac{c(l-G_i) - i}{c(1-\alpha) - i}, \quad i = 1, 2, \ldots, c. \]

We will prove two lemmas and a theorem, in which the relation (37) will be derived.

**Lemma 4.10:** Let \( \beta_{c-1} = 1 \). If \( c = 1, 2, \ldots, 0 < \alpha < 1 \), and \( c(1-\alpha) - 1 < 0 \), then

\[ \sum_{j=0}^{c-1} \beta_j = \frac{c \alpha}{1 - c(1-\alpha)} - \alpha \sum_{i=1}^{c-1} \frac{(c-1)!}{i!(i-c(1-\alpha))(i+1-c(1-\alpha))} D_i. \]
Proof.

From (44), \( \mathbf{A}^{-1} = \alpha \sum_{i=1}^{c} \binom{c}{i} \mathbf{D}_i (\frac{1}{1-G_i}) + \frac{\alpha}{1-\alpha} \). But by definition,

\[
\mathbf{A}^{-1} = \sum_{j=0}^{c-1} \beta_j + \frac{\alpha}{1-\alpha}.
\]

Therefore,

\[
\sum_{j=0}^{c-1} \beta_j = \alpha \sum_{i=1}^{c} \binom{c}{i} \mathbf{D}_i (\frac{1}{1-G_i}) - \frac{\alpha}{1-c(1-\alpha)}
\]

(47)

\[
= \alpha \sum_{i=1}^{c} \binom{c}{i} \mathbf{D}_i (\frac{1}{1-G_i}) \left( \frac{1-c(1-G_i)}{1-c(1-\alpha)} \right) - \alpha \sum_{i=1}^{c} \binom{c}{i} \mathbf{D}_i (\frac{c}{1-c(1-\alpha)})
\]

Now \( \mathbf{D}_1 = \left( \frac{1-G_i}{G_i} \right) \mathbf{D}_{i-1} \). Therefore \( \mathbf{D}_1 = \frac{\mathbf{D}_{i-1}}{G_i} - \mathbf{D}_{i-1} \), and hence

\[
\frac{\mathbf{D}_{i-1}}{G_i} = \mathbf{D}_1 + \mathbf{D}_{i-1}.
\]

Moreover, \( \frac{\mathbf{D}_i}{1-G_i} = \frac{\mathbf{D}_{i-1}}{G_i} \). Therefore we have:

\[
\frac{\mathbf{D}_i}{1-G_i} = \mathbf{D}_1 + \mathbf{D}_{i-1},
\]

\[
\sum_{j=0}^{c-1} \beta_j = \alpha \sum_{i=1}^{c} \binom{c}{i} \mathbf{D}_i (\frac{1}{1-c(1-\alpha)}) - \alpha \sum_{i=1}^{c} \binom{c}{i} \mathbf{D}_1 (\frac{c}{1-c(1-\alpha)})
\]

\[
= -\alpha \sum_{i=1}^{c} \binom{c}{i} \mathbf{D}_i (\frac{c-i}{1-c(1-\alpha)}) + \alpha \sum_{i=0}^{c-1} \binom{c}{i+1} \mathbf{D}_1 (\frac{i+1}{1+i-c(1-\alpha)})
\]

But

\[
\binom{c}{i+1} (i+1) = \binom{c}{i} (c-i).
\]

Therefore,
\[ \sum_{j=0}^{c-1} \beta_j = -\alpha \binom{c}{c} D_c \left( \frac{c-c}{c-c(1-\alpha)} \right) - \alpha \sum_{i=1}^{c-1} (c-i) \binom{c}{i} D_i \left( \frac{1}{1-c(1-\alpha)} - \frac{1}{i+1-c(1-\alpha)} \right) + \alpha \binom{c}{0+1} D_0 \left( \frac{0+1}{0+1-c(1-\alpha)} \right) \]

(48)

\[ = -\alpha \sum_{i=1}^{c-1} (c-i) \binom{c}{i} D_i \left( \frac{i+1-c(1-\alpha)-1+c(1-\alpha)}{(1-c(1-\alpha))(i+1-c(1-\alpha))} \right) + \frac{c \alpha}{1-c(1-\alpha)} \]

\[ = \frac{c \alpha}{1-c(1-\alpha)} - \alpha \sum_{i=1}^{c-1} (c-i) \binom{c}{i} D_i \left( \frac{1}{(1-c(1-\alpha))(i+1-c(1-\alpha))} \right) \cdot \]

Q.E.D.

**Lemma 4.11:** If \( c > 1 \) and \( c(1-\alpha) - 1 < 0 \), then

(49)

\[ c(1-\alpha) + \operatorname{Prob}[W_q^c > 0] > 1. \]

**Proof.**

Since \( c(1-\alpha) - 1 < 0 \), we have \( c(1-\alpha) - 1 < 0 \), for

\( i = 1, 2, \ldots, c \). Moreover, \( 0 < G_i < 1 \), \( i = 1, 2, \ldots, c-1 \), since

\( G_i = \int_0^\infty e^{-i\mu t} dH(t) \) and \( \mu > 0 \). Therefore, \( D_i > 0 \), \( i = 1, 2, \ldots, c-1 \).

Hence, using (46),

(50)

\[ \sum_{j=0}^{c-1} \beta_j < \frac{c \alpha}{1 - c(1-\alpha)} \cdot \]

Hence by (43) and (50) we have

\[ c(1-\alpha) + \operatorname{Prob}[W_q^c > 0] = c(1-\alpha) + A \frac{\alpha}{1 - \alpha} \]

\[ = c(1-\alpha) + \frac{\alpha}{c-1} \sum_{j=0}^{c-1} \beta_j + \frac{\alpha}{1-\alpha} \]

\[ > c(1-\alpha) + \frac{\alpha}{c \alpha} \frac{1}{1-c(1-\alpha)} + \frac{\alpha}{1-\alpha} \]

103
\[ 
\begin{align*}
&= c(1-\alpha) + \frac{\alpha}{\frac{c \alpha(1-\alpha)^2 + \alpha}{(1-\alpha)^2} - c \alpha (1-\alpha)} \\
&= c(1-\alpha) + \frac{\alpha(1-\alpha)(1-c(1-\alpha))}{(1-\alpha)\alpha} \\
&= c(1-\alpha) + 1 - c(1-\alpha) \\
&= 1. 
\end{align*}
\]

Q.E.D.

**Theorem 4.12:** For all \( t > 0 \) and \( c > 1 \),

\[ (51) \quad \text{Prob}[W^1 > t] < \text{Prob}[W^c > t]. \]

**Proof.**

From the solution of the GI/M/1 system, with mean service rate \( c\mu \),

\[ \text{Prob}[W^1 > t] = e^{-c\mu(1-\alpha)t}. \]

(This can also be derived from (42) by changing \( c \) to \( 1 \), \( \mu \) to \( c\mu \), and using (43), (44), and (45). Note that the value of \( \alpha \) is the same for the 1-system as the c-system.) Therefore, for \( c(1-\alpha) - 1 \neq 0 \),

\[ \text{Prob}[W^c > t] - \text{Prob}[W^1 > t] \\ = e^{-\mu t} + \text{Prob}[W^c_q > 0] \left( \frac{1}{c(1-\alpha) - 1} \right) (e^{-\mu t} - e^{-c\mu(1-\alpha)t}) - e^{-c\mu(1-\alpha)t} \]

\[ (52) \\
= e^{-\mu t} \left( 1 + \frac{\text{Prob}[W^c_q > 0]}{c(1-\alpha) - 1} \right) - e^{-c\mu(1-\alpha)t} \left( 1 + \frac{\text{Prob}[W^c_q > 0]}{c(1-\alpha) - 1} \right) \\
= (e^{-\mu t} - e^{-c\mu(1-\alpha)t}) \left( 1 + \frac{\text{Prob}[W^c_q > 0]}{c(1-\alpha) - 1} \right). \]

For \( c(1-\alpha) - 1 = 0 \),
\[
\text{(53) } \text{Prob}(W^c > t) - \text{Prob}(W^1 > t) = e^{-\mu t} + \text{Prob}(W^c_q > 0)\mu t - e^{-c(1-\alpha)t}.
\]

In order to show that \(\text{Prob}(W^c > t) - \text{Prob}(W^1 > t) > 0\), we distinguish three cases.

**CASE I - \(c(1-\alpha) - 1 > 0\):**

In this case, \(c\mu(1-\alpha) > \mu\). Therefore, \(e^{-\mu t} - e^{-c\mu(1-\alpha)t} > 0\).

Moreover, by (43), \(\text{Prob}(W^c_q > 0) > 0\), so that \(1 + \frac{\text{Prob}(W^c_q > 0)}{c(1-\alpha)-1} > 0\). Therefore, by (52),

\[
\text{Prob}(W^c > t) - \text{Prob}(W^1 > t) > 0.
\]

**CASE II - \(c(1-\alpha) - 1 = 0\):**

In this case, \(e^{-\mu t} = e^{-c\mu(1-\alpha)t}\). Therefore, by (53),

\[
\text{Prob}(W^c > t) - \text{Prob}(W^1 > t) = \text{Prob}(W^c_q > 0)\mu t
\]

\[
> 0, \quad \text{for } t > 0.
\]

**CASE III - \(c(1-\alpha) - 1 < 0\):**

In this case, \(c\mu(1-\alpha) < \mu\). Therefore, \(e^{-\mu t} - e^{-c\mu(1-\alpha)t} < 0\).

By (52),

\[
\text{Prob}(W^c > t) - \text{Prob}(W^1 > t)
\]

\[
= \left(e^{-\mu t} - e^{-c\mu(1-\alpha)t}\right)\left(c(1-\alpha) + \text{Prob}(W^c_q > 0) - 1\right)
\]

\[
> 0,
\]

by Lemma 4.11, which completes the proof. Q.E.D.

The following corollary is immediate.
COROLLARY 4.13: For a GI/M/c system of Type I-A with $0 < \rho < 1$, if the system waiting-cost rate at any time $t$ is defined by

$$C_{w,t}(c,\mu) = \sum_{i=1}^{\infty} C_{w} h'(t-A_i) \delta(t-A_i) \delta(A_i + W_1 - t),$$

where the total waiting cost for the $n$th arrival is $C_{w}(W_n)$, $h(\cdot)$ a monotone increasing differentiable Borel function, then

(3) \quad $E[C_{w}(1,c\mu)] \leq E[C_{w}(c,\mu)]$.

**Proof.**

Since the system is a GI/M/c system, Theorem 4.12 applies. Since arrivals occur according to a renewal process, and $0 < \rho < 1$, Jewell's assumptions are satisfied. Therefore we have

(54) \quad $E[C_{w}(c,\mu)] = \lambda E[C_{w}(W(c,\mu))], \quad c = 1, 2, \ldots; \quad 0 < \mu < \infty$.

But by Theorem 4.12 and Properties 1$^\circ$ and 2$^\circ$ of stochastic ordering,

(55) \quad $E[C_{w}(W(1,c\mu))] \leq E[C_{w}(W(c,\mu))], \quad c = 1, 2, \ldots; \quad 0 < \mu < \infty$.

Therefore, combining (54) and (55) yields (3), our desired result.

Q.E.D.
CHAPTER V

Network Queueing Systems

5.1 Introduction

In this chapter we will consider network queueing systems of Type II-A and II-B. Recall the definitions of these two systems:

TYPE II-A: A queueing system consisting of $r$ stations, $i = 1, 2, \ldots, r$, arranged in series. The output from station $i$ is the input to station $i+1$, for $i = 1, 2, \ldots, r-1$. The input to station $1$ is the input to the system. The output from station $r$ is the output from the system. At station $i$ there are $c_i$ servers whose service times are mutually independent and independent from the service times at other stations. Each service time at station $i$ is distributed with mean rate $\mu_i$ and c.d.f. $F_i(\cdot)$. Each station, considered as a queueing system by itself, satisfies Assumptions IV-VII. The queue discipline at each station is FCFS. The decision variables are $c = (c_1, \ldots, c_r)$ and $\mu = (\mu_1, \ldots, \mu_r)$, and the feasible sets for each arc:

$$c_i = 1, 2, \ldots, 0 < \mu_i < \infty, \text{ for } i = 1, 2, \ldots, r.$$}

TYPE II-B: A queueing system consisting of $r$ stations, $i = 1, 2, \ldots, r$, which are the nodes of the graph of a partially ordered set. That is, (i) a directed arc from node $j$ to node $k$ of the graph indicates that a customer may go from station $j$ to station $k$, and (ii) the nodes are numbered, $i = 1, 2, \ldots, r$, in such a way that there is a directed arc from $j$ to $k$ only if $j < k$. The route taken by a
customer through the system is independent of the state and history of
the system.

In all other respects the system is exactly like a system of Type
II-A. Examples of the two systems are illustrated in Figures 3 and 4.

The objective function is the expected operating cost rate of the
system as a whole:

\[ E[C(c, \mu)] = C_s(c, \mu) + E[C_w(c, \mu)] . \]

The goal of this chapter will be to determine conditions under which
\( \hat{c}_1 = 1 \) holds, where \( \hat{c}_1 \) is the optimal value of \( c_1 \) for all
\( i = 1, 2, \ldots, r \). To this end we assume that the service-cost rate is
separable, i.e.,

\[ C_s(c, \mu) = \sum_{i=1}^{r} C_{s_i}(c_i, \mu_i) , \]

and make the following assumption about the service-cost rate at each
station:

(1) \[ C_{s_i}(1, c_i, \mu_i) \leq C_{s_i}(c_i, \mu_i) , \]

for all feasible \( c_i \) and \( \mu_i \).

The waiting-cost rate at each station will be assumed to be
proportional to the number of customers in queue and in service at that
station, with the same proportionality coefficient \( (C_w) \) at each
station. Thus the system expected waiting-cost rate is proportional
to the number in the system as a whole:
FIGURE 3

Type II-A Network: Series System

FIGURE 4

Type II-B Network: Partially Ordered System
\[ E\{C_u(c, \mu)\} = C_u E\{L(c, \mu)\} \]

This simply amounts to making the same assumption that we did in Chapter II, namely, that a customer's waiting-cost rate does not vary with the length of time he has been in the system. (In Section 3.4 we discuss the case of non-linear waiting costs.)

Thus the objective function has the form:

\[ E[C(c, \mu)] = \sum_{i=1}^{r} C_{si}(c_i, \mu_i) + C_w E[L(c, \mu)] \]

Our main result will be a proof of a theorem for networks of Type II-A and II-B which is analogous to Theorem 2.1 for Type I-A systems. We will consider the case where the service-time distribution at each station is exponential. The cases of degenerate and k-Erlang service-time distribution will be discussed in Section 3.4.

For a given pair \( c = (c_1, \ldots, c_r) \) and \( \mu = (\mu_1, \ldots, \mu_r) \), let \( c*\mu \equiv (c_1 \mu_1, \ldots, c_r \mu_r) \). Let \( \overline{1} = (1, \ldots, 1) \). Let \( a = (a_1, a_2, \ldots) \) be a realization of the arrival process \( A = (A_1, A_2, \ldots) \) to a system of Type II-A or II-B. We use the following notation for the number of customers in the system as a whole at time \( t \), given \( a \), as a function of \( c \) and \( \mu \):

\[ L_u^a(c, \mu) \equiv (L_u(c, \mu)|A = a) \]

The theorem to be proved is then the following:

**THEOREM 5.1:** Suppose the service-time distribution at each station \( i \), \( i = 1, \ldots, r \), of a queueing system of Type II-A or II-B is exponential.
Then, for any arbitrary realization of the arrival process, and any feasible pair \((c, \mu)\),

\[
L^a_t(\bar{I}, c^*\mu) \leq L^a_t(c, \mu), \quad t \geq 0
\]

**Proof.**

(The proof for Type II-A networks is given in Section 5.2. The proof for Type II-B networks is given in Section 5.3.)

Just as in Chapter II, the following corollaries are immediate:

**COROLLARY 5.2:** Under the conditions of Theorem 5.1, for any \(a\) and any feasible pair \((c, \mu)\),

\[
E[L^a_t(\bar{I}, c^*\mu)] \leq E[L^a_t(c, \mu)], \quad t \geq 0
\]

**COROLLARY 5.3:** Suppose the service-time distribution at each station \(i, \ i = 1, \ldots, r\), of a queueing system of Type II-A or Type II-B is exponential. Then, for any feasible pair \((c, \mu)\),

\[
L_t(\bar{I}, c^*\mu) \subset L_t(c, \mu), \quad t \geq 0
\]

**COROLLARY 5.4:** Under the conditions of Corollary 5.3, for any feasible pair \((c, \mu)\),

\[
E[L_t(\bar{I}, c^*\mu)] \leq E[L_t(c, \mu)], \quad t \geq 0
\]

**COROLLARY 5.5:** Under the conditions of Corollary 5.3, let \((c, \mu)\) be a feasible pair for which steady state distributions of \(L(c, \mu)\) and
\( L(\bar{I},c^*\mu) \) exist. Then

(7) \[ L(1,c^*\mu) \subset L(c,\mu) . \]

COROLLARY 5.6: Under the conditions of Corollary 5.5,

(8) \[ E[L(\bar{I},c^*\mu)] \leq E[L(c,\mu)] . \]

*COROLLARY 5.7: Suppose the service-time distribution at each station \( i, ~i = 1, \ldots, r, \) of a queueing system of Type II-A or II-B is exponential. Then the optimal value of \( c_i \) for the objective function given by (2), where condition (1) holds, is

(9) \[ \hat{c}_i = 1 , ~i = 1, \ldots, r . \]

Before proceeding with the proofs of Theorem 5.1, the following remark is in order.

It might appear that Theorem 5.1 is an immediate consequence of Theorem 2.1. To minimize the number in a network, it certainly suffices to minimize simultaneously the number at each station in the network, and Theorem 2.1 tells us that, given any arrival stream to a station, the number in that station is minimized by setting the number of servers there to 1. But this observation alone is not sufficient to prove Theorem 5.1, since changing the number of servers and/or the rate at a given station has an effect, not only on the number in that station, but on the output process from that station, and hence on the input process to the other stations. Hence, it may be that setting the number
of servers at a station to one creates an input process at the next
station which results in more congestion there than there would have
been had the number of servers at the first station been set at some number
greater than 1. This could happen regardless of the number of servers at
the second station.

There is one system where interactions between the stations may
not be a problem; this is the Type II-A (series) system with Poisson
input process (mean rate \( \lambda \)) and exponential service-time distribution
at each of the stations. Burke [6] has shown that, in such a system,
the output of each station in the steady state is a Poisson process with
mean rate \( \lambda \), regardless of the values of \( c_i \) and \( \mu_i \), as long as

\[
\frac{\lambda}{c_i \mu_i} < 1, \quad \text{for } i = 1, \ldots, r.
\]

But (10) is simply the necessary and sufficient condition for the
existence of a steady-state distribution of \( L(c, \mu) \) for such a system.
In this case changing the number of servers and/or the mean service rate
at station \( i \) has no effect on the input processes to the other stations,
as long as \( \frac{\lambda}{c_i \mu_i} < 1 \). Therefore, the number of customers at each station
may be minimized separately. Thus Theorem 2.1 may be applied directly
to yield Theorem 5.1.

For general arrival processes and/or general service-time distribu-
tions, however, setting \( c_i \) to 1 at station \( i \) does affect the
arrival processes at the other stations. What can be shown in some
cases is that the effect of setting \( c_i \) to 1 is "good", that is, it
stochastically reduces the number of customers in the succeeding stations,
as well as the number in station 1. This is essentially what we prove in Sections 5.2 and 5.3.

In Section 5.2 we present a proof of Theorem 5.1 for Type II-A systems. In Section 5.3 we show how this proof can be extended to Type II-B systems.

5.2 Proof of Theorem 5.1 for Type II-A Systems

As in the proof of Theorem 2.1, we denote the system defined in the statement of Theorem 5.1 by the pair of vectors \((c, \mu)\), the \(c\)-system, and the system defined by \((\overline{1}, c^*\mu)\), the \(l\)-system. Also, let

\[
U_n \equiv n^\text{th} \text{ departure time from } l\text{-system, } n = 1, 2, \ldots \;
\]

\[
V_n \equiv n^\text{th} \text{ departure time from } c\text{-system, } n = 1, 2, \ldots .
\]

Again following the proof of Theorem 2.1, it suffices in order to prove Theorem 5.1 to prove

\[(11) \quad U_n \subset V_n, \quad n = 1, 2, \ldots ,\]

for a \(l\)-system busy period beginning at time 0.

This will be done by using Theorem 2.1 and one of the following two lemmas, which will be proved later.

LEMMA 5.8: Consider a single-station queueing system of Type I-A, but with fixed arbitrary number of servers, \(c_1\), and mean service rate, \(\mu_1\), and variable deterministic arrival stream, \(a\). Let \(Z_n(a)\) be the \(n^\text{th}\) departure time from this system, expressed as a (random) function of the infinite sequence \(a\). Then, for \(a \leq b\) (componentwise),
\[ Z_1(a) \subset Z_1(b) , \]
\[ (Z_n(a)|Z_{n-1}(a) = z) \subset (Z_n(b)|Z_{n-1}(b) = z) , \]
\[ n = 2, 3, \ldots, \] all feasible \( z \).

**Lemma 5.9:** Suppose \( 0 \leq b_1 \leq b_2 \leq \cdots \) are the fixed arrival times of customers to the first station of a Type II-A queueing system with a total of \( k \) stations, \( 1 < k \leq r \). Suppose that for \( j = 2, 3, \ldots, k \), the decision variables \( c_j \) and \( \mu_j \) are fixed at arbitrary feasible values. At the first station the decision variables \( c_1 \) and \( \mu_1 \) are allowed to vary. Suppose an arbitrary pair of values \( (c_1, \mu_1) \) is given for the decision variables at the first station. Let
\[ Y^j_n = n^{th} \text{ departure time from the } j^{th} \text{ station of this system,} \]
\[ Y^j_n = n^{th} \text{ departure time from the } j^{th} \text{ station of this system,} \]
given \((c_1, \mu_1)\), and
\[ X^j_n = n^{th} \text{ departure time from the } j^{th} \text{ station of this system,} \]
given \((1, c_1, \mu_1)\),
\[ n = 1, 2, \ldots, \] and \( j = 1, 2, \ldots, k \). Then

\[ X^j_n \leq Y^j_n , \] for \( n = 1, 2, \ldots, \) and \( j = 1, 2, \ldots, k \).

To prove (11) it suffices to show that, given arbitrary vectors \( c = (c_1, \ldots, c_r) \) and \( \mu = (\mu_1, \ldots, \mu_r) \) with \( c_i > 1 \) for some \( i, \)
\[ 1 \leq i \leq r, \] the \( n^{th} \) departure time from the system can be stochastically reduced by changing the number of servers at the \( i^{th} \) station from \( c_i \) to \( 1 \) and the mean service rate of each from \( \mu_i \) to \( c_i \mu_i \), while leaving all the other components of \( c \) and \( \mu \) unchanged. But, given
the unchanged values of \( c_1, c_2, \ldots, c_{i-1} \), and \( \mu_1, \mu_2, \ldots, \mu_{i-1} \), and the
fixed arrival stream \( a \) at the first station, the arrival times to the
\( i \)th station are a stochastic process \( B = \{ B_1, B_2, \ldots \} \) which is independ-
ent of the structure and behavior of the system beyond the \((i-1)\)st
station. Therefore, if \( i < r \), we can apply Lemma 5.9 to the reduced
network consisting of stations \( 1, i+1, \ldots, r \), by setting \( k = r - i + 1 \),
relabeling the stations, considering the arrival stream \( b \) as a
realization of \( B \), and then integrating over all realizations of \( B \) to
obtain the desired ordering of departure times. If \( i = r \), we can
apply Theorem 2.1, considering the arrival stream \( a \) defined there as
a realization of \( B \), to obtain the same result.

Hence it suffices to prove Lemmas 5.8 and 5.9, which we now do.

Proof of Lemma 5.8.

We will prove (12) for fixed, arbitrary \( n > 1 \). The proof for
\( n = 1 \) is essentially the same.

It suffices to prove that for all \( t > z \),

\[
(14) \quad \text{Prob}\{Z_n(a) > t | Z_{n-1}(a) = z\} \leq \text{Prob}\{Z_n(b) > t | Z_{n-1}(b) = z\} .
\]

Now, for \( z \leq \tau \leq t \), let

\[
m_a(\tau) = (L^{a}_{\tau}(c_1, \mu_\perp) | Z_{n-1}(a) = z, Z_n(a) > \tau) ,
\]

\[
m_b(\tau) = (L^{b}_{\tau}(c_1, \mu_\perp) | Z_{n-1}(b) = z, Z_n(b) > \tau) .
\]

That is, \( m_a(\tau) \) is the number of customers in the system at time \( \tau \)
when the arrival stream is \( a \), given that the \((n-1)\)st departure occurred at time \( z \) and another departure has not yet occurred. The corresponding interpretation is given \( m_b(\tau) \). Therefore,

\[
(15) \quad m_a(\tau) = n_a(\tau) - n + 1, \quad \text{and} \\
 m_b(\tau) = n_b(\tau) - n + 1, \quad z \leq \tau \leq t,
\]

where \( n_a(\tau) \) and \( n_b(\tau) \) are the number of arrivals in \([0, \tau]\) from arrival streams \( a \) and \( b \), respectively.

But \( a \leq b \) implies \( n_a(\tau) \geq n_b(\tau) \), for all \( \tau \geq 0 \), therefore

\[
(16) \quad m_a(\tau) \geq m_b(\tau), \quad z \leq \tau \leq t.
\]

Now if we let \( \eta_a(\tau) = (m_a(\tau) \wedge c_1) \mu_1 \), and \( \eta_b(\tau) = (m_b(\tau) \wedge c_1) \mu_1 \), for \( z \leq \tau \leq t \), then \( \eta_a(\tau) \) and \( \eta_b(\tau) \) are the mean output rates of the system at time \( \tau \), given that the \((n-1)\)st departure occurred at time \( z \) and another departure has not yet occurred, for arrival streams \( a \) and \( b \), respectively. Since under these conditions the service mechanism is acting as a Poisson output process at time \( \tau \), \( \eta_a(\tau) \) and \( \eta_b(\tau) \) satisfy equation (4) of Lemma 1.1, with \( Z \) replaced by \( Z_n(\cdot) - z \) and \( t \) replaced by \( \tau - z \). Therefore, Lemma 1.1 applies, and we have

\[
(17) \quad \text{Prob}[Z_n(a) > t | Z_{n-1}(a) = z] = \exp \left(-\int_z^t \eta_a(\tau) d\tau \right), \quad \text{and}
\]

\[
\text{Prob}[Z_n(b) > t | Z_{n-1}(b) = z] = \exp \left(-\int_z^t \eta_b(\tau) d\tau \right).
\]
Now by the definition of \( \eta_a(\tau) \) and \( \eta_b(\tau) \) and by (16),
\[
\eta_a(\tau) \geq \eta_b(\tau), \quad \text{for all} \quad z \leq \tau \leq t.
\]
Therefore,
\[
\exp \left( -\int_z^t \eta_a(\tau) d\tau \right) \leq \exp \left( -\int_z^t \eta_b(\tau) d\tau \right),
\]
and combining (17) and (18), we have (14), the desired result.

Q.E.D.

Proof of Lemma 5.9.

Instead of proving (13) directly, we will prove a slightly stronger result analogous to Lemma 2.11:

(19) \[ X_j^1 \leq Y_j^1, \]
\[(X_n^j | X_{n-1}^j = z) \subset (Y_n^j | Y_{n-1}^j = z), \quad n = 2, 3, \ldots, \quad z \geq 0, \]
for \( j = 1, 2, \ldots, k \); and

(20) \[ (X_n^j | X_{n-1}^j = z) \] is a stochastically monotone increasing function of \( z, \quad n = 2, 3, \ldots, \quad j = 1, 2, \ldots, k \).

For any fixed \( j \), the proof that (19) and (20) together imply (13) is formally equivalent to the proof of Theorem 2.1 for exponential service-time distributions, using Lemma 2.11.

Therefore it suffices to prove (19) and (20). First we prove (20). Let \( j, 1 \leq j \leq k, \quad n > 1, \quad \text{and} \quad z \geq 0 \) be given. Let \( t > z \), let \( a \) be a fixed arbitrary arrival stream to station \( j \), and let \( \eta_a(\tau) \) be the mean output rate of the service mechanism at the \( j^{th} \) station, given \( X_{n-1}^j = z \) and \( X_n^j > \tau \), for \( z \leq \tau \leq t \) and arrival stream \( a \).
Suppose \( \eta^i_a(\tau) \) is analogously defined for \( x^i_{n-1} = z + \epsilon, \epsilon > 0 \). Then clearly \( \eta^i_a(\tau) = \eta_a(\tau) \) for \( z + \epsilon \leq \tau \leq t \) (see proof of Lemma 5.8). Therefore, applying Lemma 1.1,

\[
\Pr[x^j_n > t | x^j_{n-1} = z] = \exp \left( - \int_z^t \eta_a(\tau) d\tau \right) \\
\leq \exp \left( - \int_{z+\epsilon}^t \eta_a(\tau) d\tau \right) \\
= \exp \left( - \int_{z+\epsilon}^t \eta^i_a(\tau) d\tau \right) \\
= \Pr[x^j_n > t | x^j_{n-1} = z + \epsilon],
\]

for \( t > z + \epsilon \). For \( t \leq z + \epsilon \), the right-hand side equals 1 and the relation holds trivially. Thus (20) is true.

We now prove (19) by induction on \( j, j = 1, 2, \ldots, k \).

**Case** \( j = 1 \):

The proof is an immediate application of Lemma 2.11.

**Case** \( j > 1 \):

Assume that (19) has been proved for \( j-1 \). Let \( x^{j-1,n} = (x^{j-1,1}, x^{j-1,2}, \ldots, x^{j-1,n}) \), \( y^{j-1,n} = (y^{j-1,1}, y^{j-1,2}, \ldots, y^{j-1,n}) \), and let \( x^{j-1} = (x^{j-1,1}, x^{j-1,2}, \ldots) \) and \( y^{j-1} = (y^{j-1,1}, y^{j-1,2}, \ldots) \). Then

\[
(21) \quad x^{j-1} \subseteq y^{j-1}.
\]

The proof of (21) is a straightforward induction on \( n \), similar to the proof of Theorem 2.1, together with a limiting argument. The proof will be presented shortly, after we prove (19) for \( j \) using (21).
We will prove (19) for fixed arbitrary $n > 1$. The proof for $n = 1$ is essentially the same. Denote the input process to station $j$ by $A^j = (A^j_1, A^j_2, \ldots)$; then $A^j = X^{j-1}$ or $Y^{j-1}$, depending on whether we are considering the system defined by $(1, c^1_1, \mu^1_1)$ or the system defined by $(c^1_1, \mu^1_1)$. Therefore, for any $t \geq 0$,

$$(22) \quad \text{Prob}[X^j_n > t | X^j_{n-1} = z] = \int \text{Prob}[X^j_n > t | X^j_{n-1} = z, A^j = a] \ d \text{Prob}[X^{j-1} \leq a],$$

and similarly for $\text{Prob}[Y^j_n > t | Y^j_{n-1} = z]$. But, since the decision variables $(c^j_1, \mu^j_1)$ at station $j$ are being held constant,

$$\text{Prob}[X^j_n > t | X^j_{n-1} = z, A^j = a] = \text{Prob}[Y^j_n > t | Y^j_{n-1} = z, A^j = a],$$

and by Lemma 5.8, both probabilities are monotone increasing in $a$. Therefore, using (22), Theorem 4 of Veinott [35], and (21), we obtain

$$\text{Prob}[X^j_n > t | X^j_{n-1} = z] \leq \int \text{Prob}[Y^j_n > t | Y^j_{n-1} = z, A^j = a] \ d \text{Prob}[Y^{j-1} \leq a]$$

$$= \text{Prob}[Y^j_n > t | Y^j_{n-1} = z].$$

This completes the proof of (19) for $j$, for $n > 1$.

Now it remains to prove (21). First we prove

$$(23) \quad X^{j-1,n} \subseteq Y^{j-1,n}, \quad n = 1, 2, \ldots.$$  

Case $n = 1$:

This case follows immediately from the first part of (19).
Case \( n > 1 \):

Assume that (23) has been proved for \( n-1 \). We must show that, for all \( t^n = (t_1, t_2, \ldots, t_n) \),

\[
(24) \quad \text{Prob}[X_{j-1,n} > t^n] \leq \text{Prob}[Y_{j-1,n} > t^n].
\]

Now,

\[
(25) \quad \text{Prob}[X_{j-1,n} > t^n] = \int_{\tau_{n-1} > t^n} \text{Prob}[X_{j-1,n} > t_n \mid X_{j-1,n} = \tau_{n-1}] d \text{Prob}[X_{j-1,n} \leq \tau_{n-1}]
\]

\[
= \int I(\tau_{n-1} > t^n) \text{Prob}[X_{j-1,n} > t_n \mid X_{j-1,n} = \tau_{n-1}] d \text{Prob}[X_{j-1,n-1} \leq \tau_{n-1}],
\]

where \( I(x) = 1 \) if \( x \geq 0 \), \( I(x) = 1 \) otherwise. But \( I(\tau_{n-1} > t^n) \) is a monotone increasing function of \( \tau_{n-1} \), and so is \( \text{Prob}[X_{j-1,n} > t_n \mid X_{j-1,n} = \tau_{n-1}] \) (by (20)); therefore, by (25), Theorem 4 of reference [35], and the induction assumption:

\[
(26) \quad \text{Prob}[X_{j-1,n} > t^n] \leq \int I(\tau_{n-1} > t^n) \text{Prob}[X_{j-1,n} > t_n \mid X_{j-1,n} = \tau_{n-1}] d \text{Prob}[Y_{j-1,n-1} \leq \tau_{n-1}].
\]

Therefore, by (26) and (19) for \( j-1 \),

\[
\text{Prob}[X_{j-1,n} > t^n] \leq \int I(\tau_{n-1} > t^n) \text{Prob}[Y_{j-1,n} > t_n \mid Y_{j-1,n} = \tau_{n-1}] d \text{Prob}[Y_{j-1,n-1} \leq \tau_{n-1}]
\]

121
\[ = \text{Prob} \{ Y_{j-1,n} > t^n \} , \]

which proves (23) for \( n \).

This completes the proof of (23). Let \( t = (t_1, t_2, \ldots) \) and let \( t^n = (t_1^n, t_2^n, \ldots, t_n^n) \). By the monotone sequential continuity of probability,

\[ \text{Prob} \{ X_{j-1} > t \} = \lim_{n \to \infty} \text{Prob} \{ X_{j-1,n} > t^n \} \quad \text{and} \quad \text{Prob} \{ Y_{j-1} > t \} \]

\[ = \lim_{n \to \infty} \text{Prob} \{ Y_{j-1,n} > t^n \}. \]

But by (23),

\[ \text{Prob} \{ X_{j-1,n} > t^n \} \leq \text{Prob} \{ Y_{j-1,n} > t^n \} . \]

Therefore,

\[ \text{Prob} \{ X_{j-1} > t \} \leq \text{Prob} \{ Y_{j-1} > t \} , \]

i.e., (21) holds.

This completes the proof of Lemma 5.9.

Q.E.D.

5.3 Proof of Theorem 5.1 for Type II-B Networks

In this section we give an informal argument that the results of the previous section extend to networks of Type II-B.

Property (ii) in the definition of Type II-B networks implies that there are no directed cycles in the graph of the system. Thus a customer cannot return to a station at which he has already been served. More important for our purposes: changing the design parameters of a station can have no effect on the input to that station. This "no-feedback" condition is essential to our method of analysis of this system.
In a network of Type II-B the input to a station will perhaps be composed of the output from many other stations. To prove Theorem 5.1 for a Type II-B network, we must show that if \( c_i > 1 \) for some station \( i, 1 \leq i \leq r \), setting \( c_i = 1 \) stochastically decreases the departure times from the last station (i.e., from the network). But we know from Theorem 2.1 that setting \( c_i = 1 \) (keeping the product of \( c_i \) and \( \mu_i \) constant) will stochastically decrease the departure times from station \( i \). Now each departure from station \( i \) becomes an arrival at some station \( j, j > i \). Hence, for any realization, some of the arrival times at station \( j \) will be reduced and none will be increased. Let the arrival stream at station \( j \) be \( a = (a_1, a_2, \ldots) \). If one of these numbers is made smaller, say \( a_n \rightarrow b_n < a_n \), then the new arrival stream \( a' \) formed by replacing \( a_n \) by \( b_n \) will satisfy

\[
a' \leq a.
\]

This is obvious as long as \( b_n \geq a_{n-1} \). In this case \( a'_n = b_n \) and all other \( a_m, m \neq n \), remain the same. If \( a_{n-2} \leq b_n < a_{n-1} \), then \( a'_n = a_{n-1}, a'_{n-1} = b_n \), and all other \( a_m, m \neq n \), remain the same. But in this case, too, \( a'_n \leq a_n \) and \( a'_{n-1} \leq a_{n-1} \). This argument can be continued to establish that \( a'_i \leq a_i \) for all \( i \).

Thus, by Lemma 5.8, the immediate effect of setting \( c_i = 1 \) is to stochastically reduce the departure times at one or more stations \( j \), \( 1 \leq j \), perhaps including station \( r \). But departures from these stations become arrivals to higher numbered stations and hence, applying Lemma 5.8 again, we obtain more stochastic reduction of the departure times from these stations. This inductive reasoning could be formalized, but
it should be clear that the ultimate effect of setting $c_1 = 1$ must be
to stochastically lower the departure times from the system as a whole.

5.4 Remarks

We have proved that $\hat{c} = 1$ for the simplest case: linear waiting
costs and exponential service-time distribution. The proofs we have
given do not extend immediately to more general waiting-cost functions
or service-time distributions; however, it seems reasonable that the
results of this chapter would be valid for monotone increasing waiting-
cost functions and degenerate and k-Erlang service-time distributions,
as there do not seem to be any properties of network systems that would
render the intuitive arguments of Section 2.1 invalid.

We therefore make the following conjectures:

CONJECTURE I: Theorem 5.1 remains valid if the service-time distribution
is degenerate or k-Erlang. That is, for an arbitrary realization $a$ of
the arrival process, and any feasible pair $(c, \mu)$,

$$L_t^a(\bar{c}, c^* \mu) \subset L_t^a(c, \mu), \quad t \geq 0.$$

CONJECTURE II: Corollary 5.7 remains valid if the service-time distri-
bution is degenerate or k-Erlang and the objective function has the
form:

$$\mathbb{E}[C(c, \mu)] = \sum_{i=1}^{r} C_{s_i}(c_i, \mu_i) + \lambda C_w \mathbb{E}[h(W(c, \mu))] ,$$

where $C_{s_i}('', '')$ satisfies (1) and $h$ is a monotone increasing Borel
function. That is,

\[ \hat{c}_i = 1, \quad i = 1, 2, \ldots, r \]

The results of this chapter may also be valid in networks containing directed cycles. However, for such systems, the proofs would be considerably more complicated.
CHAPTER VI

Single-Station Models: Some Generalizations of Operating Policy, Decision Space, and Cost Structure; Summary and Conclusions

6.1 Introduction and Review of Previous Chapters

In Section 2.2, we proved the stochastic ordering of the number of customers in the system

\[(1) \quad L_{t}^{a}(l, c, \mu) \preceq L_{t}^{a}(c, \mu), \text{ for all } t \geq 0 ,\]

for all feasible \((c, \mu)\) and all arrival streams \(a\) to a queueing system of Type I-A with exponential service-time distribution. A corollary of this theorem was that \(\hat{c} = 1\) minimizes objective functions of the form,

\[(2) \quad E[C(c, \mu)] = C_{s} c \mu + C_{w} E[L(c, \mu)] .\]

These are the key results of this paper. In the remainder of Chapter II and in Chapters III, IV, and V, respectively, these results were extended to:

(i) more general service-time distributions (degenerate and k-Erlang);

(ii) more general service-cost functions (those satisfying \(C_{a}(l, c, \mu) \preceq C_{a}(c, \mu)\), for all feasible \(c, \mu\));

(iii) more general waiting-cost functions (monotone increasing functions of number of customers in system, convex monotone increasing functions of waiting time of customer).
In this chapter some further extensions of the results, (1) and \( \hat{c} = 1 \) for (2), are presented. Here, however, the generalizations are primarily to models with different operating policies, basic cost structure, or set of decision variables and their feasible values. Specifically, the following generalizations are considered:

(i) Type I-B systems: single-station systems with non-FCFS queue disciplines (Section 6.2);

(ii) Type I-A systems where the service cost for a server is charged only when he is busy (Section 6.3);

(iii) Type I-C systems: single-station systems where each server may work at a different mean rate (Section 6.4);

(iv) Type I-D systems: single-station systems where the service mechanism is not characterized by the number of servers and the mean rate of each, but rather by an instantaneous mean output rate \( \mu_n \), dependent on the number of customers \( n \) in the system (Section 6.5);

(v) Type I-E systems: single-station systems with an upper bound on the feasible values of \( \mu \) (Section 6.6);

(vi) a single-station system in which servers may work separately or together in a team, and in which the form of the service-time distribution may vary with the size of the team (Section 6.7).

In many of these cases, after making certain assumptions about the generalized model, we are able to prove an appropriate generalization of (1). Often this suffices to prove that \( \hat{c} = 1 \) for (2); in the cases where it does not, we use the generalization of (1) to draw some conclusions about the new optimal value of \( c \).

In many cases, for simplicity of presentation, we consider only
single-station models with exponential service-time distribution and linear costs. Often it is obvious that our results and method of proof will extend to other service-time distributions, non-linear costs, or networks, under the same assumptions and using the same sort of reasoning as was employed to extend Theorem 2.1 to these cases. We will indicate such extensions in corollaries. Cases where the extensions are not obvious but where intuition indicates that they are probably valid will be indicated by conjectures. We will also indicate where an extension is not valid and why.

The work in this chapter is primarily of a tentative nature. There has been no attempt to push generalizations as far as possible. Rather, in each case the simplest generalization is presented and directions for future research are indicated.

6.2 Effect of Queue Discipline

In this section we show that the relation (1) is independent of the queue discipline under Assumptions IV-VII, as long as the rule for the selection of the next customer is independent of the length of his service time.

Recall the definition of Type I-B queueing systems:

TYPE I-B: A single-station queueing system with $c$ servers whose service times are mutually independent, with mean rate $\mu$ and c.d.f. $F(\cdot)$. The queue discipline is strict, but not necessarily FCFS. The decision variables are $c$ and $\mu$, and the feasible sets for each are:

$$c \leq \bar{c}, \quad \bar{c} \text{ a positive integer or } \infty;$$
\[ \mu \in R, \ R \text{ a } \bar{c}\text{-complete set}. \]

**THEOREM 6.1:** Suppose the service-time distribution in a queueing system of Type I-B is degenerate, exponential, or k-Erlang \((k = 1, 2, \ldots)\).

Suppose the rule for the selection of the next customer to be served is independent of the length of his service time. Then, for any realization \(a\) of the arrival process, and any feasible pair \((c, \mu)\),

\[ L^a_t(1, c \mu) \subseteq L^a_t(c, \mu), \quad t \geq 0. \]

**Proof.**

We give the proof for exponential service-time distribution. The proofs for degenerate and k-Erlang service-time distributions are analogous.

Because the queue discipline is strict, it suffices to prove that Lemma 2.11 still holds for this system, i.e.,

\[ U_n \subseteq V_n; \]

\[ (U_n | U_{n-1} = z) \subseteq (V_n | V_{n-1} = z), \quad n = 2, 3, \ldots; \quad \text{and all } z; \]

\[ (U_n | U_{n-1} = z) \text{ is a stochastically monotone increasing function of } z, \quad n = 2, 3, \ldots, \]

where \(U_n\) and \(V_n\) are defined as in the proof of Theorem 2.1.

Clearly, to prove (4) and (5) it suffices to prove that the distributions of \(U_1, V_1, (U_n | U_{n-1} = z)\), and \((V_n | V_{n-1} = z)\), for \(n = 2, 3, \ldots, \) and all feasible \(z\), are not affected by the queue discipline, as long
as Assumptions IV-VII hold and the rule for selection of the next customer to be served does not depend on the length of his service time.

But these assumptions together imply that all of the customers in service at time $z$ have the same distribution of remaining service-time, namely exponential with mean rate $c\mu$ for the l-system and $\mu$ for the c-system. Therefore

$$\text{Prob}\{U_n > t | U_{n-1} = z\} = \exp(-c\mu(t-z)),$$

$$\text{Prob}\{V_n > t | V_{n-1} = z\} = \exp\left(-\int_z^t \eta(\tau) d\tau\right),$$

where $\eta(\tau) = ((n_\alpha(\tau) - (n-1))/c)\mu$, just as in the proof of Lemma 2.1. Hence the distributions of $U_1$, $V_1$, $(U_n | U_{n-1} = z)$, and $(V_n | V_{n-1} = z)$ are the same as in the FCFS system of Theorem 2.1, and hence (4) and (5) still hold.

Q.E.D.

Note that, if the rule for selection of the next customer to be served does depend on the length of his service-time, then the total service times (and hence the remaining service times) of the customers in service in the l-system (c-system) at time $z$ might not all be exponential with mean rate $c\mu$ (\mu) and (4) and (5) might not hold. For example, if the service-times of all customers in the queue are known in advance and if the rule is to select for service the customer in the queue with the shortest service-time, then the remaining service times of the customers in service at time $z$ would tend to be shorter than in the FCFS case.

All the corollaries of Theorem 2.1 remain valid for Type I-B systems.
In particular, we have:

**COROLLARY 6.2:** Under the conditions of Theorem 6.1, \( \hat{c} = 1 \) minimizes objective functions of the form:

\[
E[C(c,\mu)] = C_s(c,\mu) + C_w E[L(c,\mu)]
\]

where \( C_s(1,\mu) \leq C_s(c,\mu) \), for all feasible \( c \) and \( \mu \).

The following corollary is also immediate:

**COROLLARY 6.3:** Suppose the service-time distribution at each station \( i, \ i = 1, 2, \ldots, r \), of a queueing system of Type II-A or II-B is exponential. Suppose the queue discipline at each station is strict, but not necessarily FCFS, and the rule for selection of the next customer to be served does not depend on the length of his service time. Then, for any realization \( a \) of the arrival process, and any feasible pair \( (c,\mu) \),

\[
L_t^a(I, c^*\mu) \leq L_t^a(c,\mu), \quad t \geq 0
\]

Moreover, if the objective function is

\[
E[C(c,\mu)] = \sum_{i=1}^{r} C_s(c_i,\mu_i) + C_w E[L(c,\mu)]
\]

where \( C_s(c_i,\mu_{i_1}) \leq C_s(c_i,\mu_i) \), for all feasible \( c_i, \mu_i, i = 1, 2, \ldots, r \), then \( \hat{c} = (1, 1, \ldots, 1) \) is optimal.

The optimality of \( \hat{c} = 1 \) does not seem to extend easily to non-linear waiting cost rates which are of the customer-inconvenience type
(see Section 4.1). For example, if the service-time distribution is exponential and the queue discipline is last-come, first-served (LCFS) (cf. [23], [29]), then a customer's waiting time, given the number of customers \( n \) in the system when he arrives, is:

(a) stochastically smaller in the l-system than the c-system, when \( n = 0 \) (since it is equal to his service time in both systems);

(b) stochastically larger (at least for large \( \lambda \)) in the l-system when \( 0 < n < c \) (since it is equal to a busy-period plus a service time in the l-system and to just a service time in the c-system);

(c) stochastically smaller in the l-system, when \( n \geq c \), since it is equal to a busy period plus a service time in each system and the busy periods for the l-system and the c-system have the same distribution.

Note the similarity to the conditional waiting-time distributions in the FCFS situation. In that case, the departure times of customers in the c-system were not necessarily in the same order as their arrival times. For the LCFS case, departure times are not necessarily in the same order as arrival times in both l-system and c-system. Convexity of the customer waiting-cost function does not seem to help here, however, because of the reverse stochastic ordering in case (b) above.

6.3 Models Where Service-Cost is Charged Only for Server's Busy Time

In this case we assume that each server has a service-cost rate, charged only when he is busy, which is a monotone increasing function (the same for all servers) of his mean service rate \( \mu \).

We consider Type I-A queueing systems with objective function
(8) \[ E[C(c, \mu)] = E[C_s(c, \mu)] + E[C_w(c, \mu)] \]

We make the following assumptions about the system:

(i) the service-cost rate at time \( t \) has the form

(9) \[ C_{s,t}(c, \mu) = N_t(c, \mu)C_s(\mu) \]

where \( N_t(c, \mu) \equiv \) number of busy servers at time \( t \), and \( C_s(\mu) \) is a monotone increasing function of \( \mu \);

(ii) the waiting-cost rate at time \( t \) has the form

(10) \[ C_{w,t}(c, \mu) = \sum_{i=1}^{\infty} C_w h'(t-A_i) \delta(t-A_i) \delta(A_i + W_i - t) \]

where \( h \) is a monotone increasing differentiable Borel function, and \( \delta(\cdot) \), \( \{A_i\} \), and \( \{W_i\} \) are as defined in Section 4.2;

(iii) the assumptions of Jewell [15] are satisfied;

(iv) the feasible sets for \( c \) and \( \mu \) have the form:

\[ \mu \leq \bar{\mu}; \quad c \leq \bar{c}; \]

where \( 0 < \bar{\mu} \leq \infty \), \( 1 \leq c \leq \infty \);

(v) \( W(c, \mu) \) is a stochastically monotone decreasing function of \( c \) and \( \mu \).

**Theorem 6.4:** Consider a queueing system of Type I-A with arbitrary service-time distribution and objective function given by (8), for which Assumptions (i)-(iv) hold. Then the optimal value of \( c \) is \( \hat{c} = \bar{c} \).

133
If $C_s(\cdot)$ is quasi-homogeneous (see Chapter III), then the optimal value of $\mu$ is $\hat{\mu} = \bar{\mu}$.

**Proof.**

By the argument of Section 4.2, Assumptions (ii) and (iii) imply that

$$E[C_w(c,\mu)] = \lambda C_w E[h(W(c,\mu))].$$

Under Assumptions (iii) and (i), if we consider the service mechanism alone as a queueing system, then this queueing system satisfies Jewell's assumptions and its mean arrival rate is $\lambda$. Therefore,

$$E[C_s(c,\mu)] = N(c,\mu)C_s(\mu)$$

$$= \lambda E[\text{service time}|\mu]C_s(\mu)$$

$$= \lambda \frac{C_s(\mu)}{\mu}.$$

Hence, combining (11) and (12)

$$E[C(c,\mu)] = \lambda \left( \frac{C_s(\mu)}{\mu} + C_w E[h(W(c,\mu))] \right).$$

Since $W(c,\mu)$ is a stochastically monotone decreasing function of $c$ and $\mu$, $C_w E[h(W(c,\mu))]$ is a monotone decreasing function of $c$ and $\mu$. Since $C_s(\mu)/\mu$ does not depend on $c$, for any value of $\mu$, (13) is minimized by $c = \bar{c}$. Therefore $\hat{c} = \bar{c}$.

For $c$ fixed at $\bar{c}$, $C_w E[h(W(c,\mu))]$ is a monotone decreasing
function of $\mu$. Since $C_s(\cdot)$ is quasi-homogeneous, $C_s(\mu)/\mu$ is monotone decreasing in $\mu$. Therefore $\hat{\mu} = \bar{\mu}$.

Q.E.D.

Note that, unlike the theorems of the other sections of this chapter, this theorem does not depend on Theorem 2.1.

6.4 Models where Each Server Works at a Different Mean Rate

Recall the definition of Type I-C queueing systems:

TYPE I-C: A single-station queueing system with $c$ servers whose service times are mutually independent. The $i^{th}$ server serves all customers according to a service-time distribution with mean rate $\mu_i$ and c.d.f. $F_i(\cdot)$, where all the $F_i(\cdot)$ have the same form, $i = 1, 2, \ldots, c$. The queue discipline is FCFS. The decision variables are $c$ and $\mu_1, \mu_2, \ldots, \mu_c$, and the feasible sets for each are:

$$c = 1, 2, \ldots; \quad 0 < \mu_i < \infty, \quad i = 1, 2, \ldots, c$$

We will prove the following theorem for Type I-C systems:

THEOREM 6.5: Suppose the service-time distribution in a Type I-C queueing system is exponential. Let $a$ be an arbitrary realization of the arrival process and let $L_t^a(c; \mu_1, \mu_2, \ldots, \mu_c)$ denote the number of customers in the system at time $t \geq 0$, given $a$, as a function of the decision variables. Let $(c; \mu_1, \mu_2, \ldots, \mu_c)$ be an arbitrary feasible set of values for the decision variables and let $\mu = \frac{1}{c} \left( \sum_{i=1}^{c} \mu_i \right)$. Then
\[(\ref{14}) \quad L_t^a(1,c\mu) \subseteq L_t^a(c;\mu_1,\mu_2,\ldots,\mu_c), \text{ for all } t \geq 0. \]

**Proof.**

It suffices to prove that Lemma 2.11 holds in such a system, i.e.,

\[(\ref{15}) \quad U_n \subseteq V_1, \]

\[(\ref{16}) \quad (U_n\mid U_{n-1} = z) \subseteq (V_n\mid V_{n-1} = z), \]

\[n = 2,3,\ldots, \text{ and all } z \geq 0, \]

\[(\ref{17}) \quad (U_n\mid U_{n-1} = z) \text{ is a stochastically monotone increasing function of } z, \]

\[n = 2,3,\ldots, \]

where \(U_n\) and \(V_n\) are defined as in the proof of Theorem 2.1.

We prove (15) for \(n = 2,3,\ldots\); the case \(n = 1\) is proved in the same manner. For any \(t > z,\)

\[(\ref{18}) \quad \text{Prob}\{U_n > t\mid U_{n-1} = z\} = \exp(-c\mu(t-z)), \]

just as in the proof of Lemma 2.11, since the server in the 1-system serves exponentially at the constant rate \(c\mu\) during a 1-system busy period. Similarly,

\[(\ref{19}) \quad \text{Prob}\{V_n > t\mid V_{n-1} = z\} = \exp\left(-\int_z^t \eta(\tau)\,d\tau\right), \]

where \(\eta(\tau) = \text{mean output rate of c-system at time } \tau, \text{ if no departures have occurred since the } n\text{-lst occurred at time } z \text{ for } z \leq \tau \leq t. \) But,
at any time \( t \), the output process of the \( c \)-system is instantaneously Poisson with a mean rate equal to the sum of the rates for all the servers busy at \( t \), and this sum must be less than or equal to \( \sum_{i=1}^{c} \mu_i = c \mu \). Therefore, in particular, \( \eta(t) \leq c \mu \), for all \( z \leq t \leq t \). Therefore, from (18) and (17),

\[
\text{Prob}\{V_n > t | V_{n-1} = z\} \geq \exp\left( -\int_{z}^{t} c \mu d\tau \right) = \exp(-c \mu (t-z)) = \text{Prob}\{U_n > t | U_{n-1} = z\}.
\]

Condition (16) follows immediately from (17), thus completing the proof.

Q.E.D.

The following corollary is immediate.

COROLLARY 6.6: Suppose the service-time distributions in a queueing system of Type I-C are exponential. Then the optimal value of \( c \) for the objective function

\[(19) \quad E\{C(c; \mu_1, \ldots, \mu_c)\} = c C_0 \sum_{i=1}^{c} \mu_i + C_c E\{L(c; \mu_1, \ldots, \mu_c)\}\]

is \( \hat{c} = 1 \).

Extensions to non-linear waiting costs and networks are not immediate. However, the following conjecture seems reasonable:

CONJECTURE: All the results of Chapter IV and V extend to systems in which each server may serve at a different mean rate.

6.5 Models with State-Dependent Service Rate

In Chapter II we noted the equivalence, in terms of number in the
system, of (i) a queueing system with \( c \) independent servers each with exponential service-time distribution with mean rate \( \mu \), and (ii) a queueing system with a single server who serves with an exponential service-time distribution with instantaneously variable service rate \( \mu_n = (n \wedge c)\mu \). This equivalence yielded an intuitive argument for the \( 1 \)-system having stochastically fewer customers than the \( c \)-system: it is "obvious" that the exponential server who is always serving at a higher mean rate has less congestion.

The equivalence also suggests a generalization. Suppose the service mechanism (we do not need to specify whether it is one server or many) is a variable-rate Poisson output process with mean output rate \( \mu_n \), \( n = 1, 2, \ldots \), where \( n \) is the number of customers in the system. We are interested in specifying a cost model for such a system which will make it possible to generalize Theorem 2.1 in a meaningful way. The following model suggests itself:

Let the expected steady-state operating cost, as a function of the infinite sequence \( \{\mu_1, \mu_2, \ldots\} \), have the form

\[
(20) \quad E[C(\mu_1, \mu_2, \ldots)] = C_s \sup_{1 \leq i < \infty} \{\mu_i\} + C_w E[L(\mu_1, \mu_2, \ldots)],
\]

where \( L(\mu_1, \mu_2, \ldots) \) is the number in the system in steady-state, as a function of \( \{\mu_1, \mu_2, \ldots\} \). The problem is to choose the sequence \( \{\mu_1, \mu_2, \ldots\} \) which minimizes (20). Obviously, then, the choice will be made among sequences for which \( \sup_{1 \leq i < \infty} \{\mu_i\} < \infty \) and the steady-state distribution of \( L(\mu_1, \mu_2, \ldots) \) exists.

For a given pair \( (c, \mu) \), the \( c \)-system defined in Theorem 2.1 is
the special case of the system described above in which $\mu_n = (n \land c) \mu$, $n = 1, 2, \ldots$; thus in this case $\sup_{1 \leq i < \infty} \{\mu_i\} = c \mu$ and the service-cost rate is $c \sum \mu_i$, as specified in Chapter II.

We now state and prove the appropriate generalization of Theorem 2.1.

THEOREM 6.7: Let $a$ be an arbitrary realization of the arrival process to a queueing system of Type I-D, and let $L_t^a(\mu_1, \mu_2, \ldots)$ denote the number of customers in the system at time $t \geq 0$, given $a$, as a function of the decision variables $\mu_1, \mu_2, \ldots$. Let $\{\mu_1, \mu_2, \ldots\}$ be an arbitrary feasible sequence and let $\tilde{\mu} = \sup_{1 \leq i < \infty} \{\mu_i\}$.

\begin{equation}
L_t^a(\tilde{\mu}, \tilde{\mu}, \ldots) \leq L_t^a(\mu_1, \mu_2, \ldots), \quad t \geq 0.
\end{equation}

Proof.

Again paralleling the proof of Theorem 2.1 for exponential service-time distribution, it suffices to prove

\begin{equation}
U_n \subseteq V_n, \quad n = 1, 2, \ldots,
\end{equation}

where now $U_n \equiv n^{th}$ departure time from system defined by sequence $\{\tilde{\mu}, \tilde{\mu}, \ldots\}$, and $V_n \equiv n^{th}$ departure time from system defined by sequence $\{\mu_1, \mu_2, \ldots\}$.

Hence it suffices to prove the analogue of Lemma 2.11:

\begin{equation}
U_1 \subseteq V_1,
\end{equation}

\begin{equation}
(U_n|U_{n-1} = z) \subseteq (V_n|V_{n-1} = z),
\end{equation}

139
\[ n = 2, 3, \ldots, \text{ and all } z \geq 0, \]

\[(24) \quad (U_n|U_{n-1} = z) \text{ is a stochastically monotone increasing function of } z, \]

\[n = 2, 3, \ldots \]

We prove (23) for \( n = 2, 3, \ldots \); the case \( n = 1 \) is proved in the same manner. By Lemma 1.1, for any \( t > z \),

\[(25) \quad \text{Prob}\{U_n > t|U_{n-1} = z\} = \exp(-\tilde{\mu}(t-z)), \]

\[\text{Prob}\{V_n > t|V_{n-1} = z\} = \exp\left(-\int_z^t \eta(\tau)d\tau\right), \]

where \( \eta(\tau) = \text{mean output rate at time } \tau \text{ from the system defined by} \]

\( \{\mu_1, \mu_2, \ldots\} \), given \( V_{n-1} = z \) and no departures since time \( z \), for all \( z \leq \tau \leq t \). Now, for any \( \tau \), the mean output rate at \( \tau \) is one of the elements of the sequence \( \{\mu_n\} \) and thus is less than or equal to \( \tilde{\mu} = \sup_{1 \leq n \leq \infty} \{\mu_n\} \). Therefore, in particular, \( \eta(\tau) \leq \tilde{\mu} \)

\( z \leq \tau \leq t \).

and hence, using (25), we conclude immediately that (23) holds. Condition (24) is an immediate consequence of (25).

This completes the proof of Theorem 6.7.

Q.E.D.

The following corollary gives the conclusion of Theorem 6.7 for a special case which is just a slight generalization of the multi-server exponential system.

COROLLARY 6.8: Let \( a \) be an arbitrary realization of the arrival
process to a Type I-D queueing system in which \( \mu_n = \mu_c \), for some \( c \) and all \( n \geq c \). Let \( L_t^a(c; \mu_1, \ldots, \mu_c) \) denote the number in such a system at time \( t \geq 0 \), given \( a \), as a function of the decision variables \( (c; \mu_1, \ldots, \mu_c) \). Let \( (c; \mu_1, \ldots, \mu_c) \) be an arbitrary feasible set of values of the decision variables, and let \( \mu = \frac{\mu_c}{c} \). Then

\[
L_t^a(1; c\mu) \leq L_t^a(c; \mu_1, \ldots, \mu_c), \quad t \geq 0
\]

(26)

**Proof.**

Observe that \( \mu_c = \sup_{1 \leq n < \infty} \{\mu_n\} = \tilde{\mu} = c\mu \) and apply Theorem 6.7.

Q.E.D.

In the case of the system defined by Corollary 6.8, the condition that \( \mu_n \leq \mu_c \), \( n = 1, 2, \ldots, c-1 \), is obviously not necessary for (26) to hold or for \( \hat{c} = 1 \) to be optimal for objective functions of the form:

\[
E[C(c; \mu_1, \ldots, \mu_c)] = C_s \mu_c + C_w E[L(c; \mu_1, \ldots, \mu_c)]
\]

All that is required is that the advantage gained for the \( c \)-system over the \( 1 \)-system by those \( \mu_n \) which exceed \( \mu_c \) is offset by those \( \mu_n \) which are less than \( \mu_c \). To determine, for a given sequence \( \{\mu_1, \mu_2, \ldots, \mu_c\} \), whether this is the case, we must be able to express \( E[L(c; \mu_1, \mu_2, \ldots, \mu_c)] \) explicitly as a function of \( (c; \mu_1, \mu_2, \ldots, \mu_c) \). Following Kendall [17] (cf. Section 4.4), the imbedded-chain steady-state equations \( \pi = \pi P \) can be written for systems with renewal input processes. However, without the special structure of the GI|M|c system, these equations do not seem to have a simple closed-form solution.

There does not seem to be any easy generalization of Theorem 6.7 to
other service-time distributions (i.e., non-Poisson output processes); however, the following corollaries of Theorem 6.7 are immediate:

**Corollary 6.9:** If the objective function for a Type I-D queueing system has the form (20), then the optimal sequence \( \{\mu_1, \mu_2, \ldots\} \) is a constant function of \( n \), \( \{\tilde{\mu}, \tilde{\mu}, \ldots\} \), for some \( 0 < \tilde{\mu} < \infty \).

**Corollary 6.10:** Let \( W_n^a(\mu_1, \mu_2, \ldots) \) be the waiting time in a system of Type I-D for the \( n \)th customer \( n = 1, 2, \ldots \), given arrival stream \( a \), as a function of the decision variables. Suppose customers depart from the system in the order of their arrival (FIFO). Then, for any sequence \( \{\mu_1, \mu_2, \ldots\} \) such that \( \tilde{\mu} = \sup_{1 \leq i < \infty} \{\mu_i\} < \infty \)

\[
W_n^a(\tilde{\mu}, \tilde{\mu}, \ldots) \leq W_n^a(\mu_1, \mu_2, \ldots), \quad n = 1, 2, \ldots
\]

**Proof.**

Under the FIFO discipline, the \( n \)th departure time is the departure time of the \( n \)th arrival. Therefore, for fixed arrival stream \( a \), to prove (27) it suffices to prove that \( U_n \subseteq V_n \), \( n = 1, 2, \ldots \), which was done in the proof of Theorem 6.7.

**Q.E.D.**

**Corollary 6.11:** Consider a queueing system of Type I-D. Suppose the assumptions of Jewell [15] are met and that the objective function has the form

\[
E[C(\mu_1, \mu_2, \ldots)] = C_s \sup_{1 \leq i < \infty} \{\mu_i\} + \lambda C_w E[h(W(\mu_1, \mu_2, \ldots))]
\]

142
where \( h(\cdot) \) is a monotone increasing differentiable Borel function.
Then the optimal sequence \( \{\mu_1, \mu_2, \ldots\} \) is a constant function of \( \tilde{\mu} \),
\( \{\tilde{\mu}, \tilde{\mu}, \ldots\} \), for some \( 0 < \tilde{\mu} < \infty \).

\textbf{Proof.}

Apply Corollary 6.10.

Q.E.D.

6.6 Models with an Upper Bound on the Mean Service Rate

We consider Type I-E queueing systems with exponential service-time distribution in which

\[
E[C(c, \mu)] = C_s c \mu + C_w E[L(c, \mu)] ,
\]

subject to the constraint

\[
\mu \leq \bar{\mu} ,
\]

where the upper bound \( \bar{\mu} \) is a specified positive number.

We denote by \( \hat{c} \) the optimal value of \( c \), and by \( \hat{\mu}_c \) the optimal value of \( \mu \) for fixed \( c \), for objective function (29) without the constraint on \( \mu \) (i.e., for Type I-A systems with feasible sets \( c = 1, 2, \ldots, 0 < \mu < \infty \)). We denote by \( \hat{c}' \) the optimal value of \( c \), and by \( \hat{\mu}'_c \) the optimal value of \( \mu \) for fixed \( c \), for objective function (29) with the constraint (30) on \( \mu \) (i.e., for Type I-D systems). We know from Corollary 2.9 that \( \hat{c} = 1 \). Clearly, if \( \hat{\mu}_c \leq \bar{\mu} \) then \( \hat{c}' = 1 \). However, if \( \hat{\mu}_c > \bar{\mu} \) then the optimal constrained pair \( (\hat{c}', \hat{\mu}') \) could be \( (1, \bar{\mu}) \) or \( (1, \tilde{\mu}) \), where \( \tilde{\mu} \) is a local optimum of \( E[C(1, \mu)] \),
\( \hat{\mu} < \bar{\mu} \). Or the optimal constrained pair could be \((c, \hat{\mu}_c)\), \((c, \bar{\mu})\), or \((c, \tilde{\mu})\) for some \( c > 1 \), where \( \tilde{\mu} \) is a local optimum of \( E[C(c, \mu)] \), \( \tilde{\mu} < \bar{\mu} \).

It would be valuable to have some way of efficiently searching the set of candidates for the optimal constrained pair. In this section we give some indications of how this might be done. The following theorem implies that, if we can find an integer \( i \) such that \( \hat{\mu}_1 \leq \bar{\mu} \), then the search for the optimal constrained pair may be restricted to values of \( c \leq i \).

**Theorem 6.12:** Suppose the service-time distribution in a Type I-E queueing system is exponential. Let \( a \) be an arbitrary realization of the arrival process and let \((c, \mu)\) be an arbitrary pair. Let \( \mu' = \left( \frac{c}{c+1} \right) \mu \). Then

\[
L_t^a(c, \mu) \subset L_t^a(c+1, \mu'), \quad t \geq 0 .
\]

**Proof.**

Following the method of proof of Theorem 2.1, it suffices to prove

\[
U_1 \subset V_1 ,
\]

\[
(U_{n-1} | U_{n-1} = z) \subset (V_{n-1} | V_{n-1} = z) ,
\]

\[ n = 2, 3, \ldots, \text{ and all } z \geq 0 , \]

\[
(U_n | U_{n-1} = z) \text{ is a stochastically monotone increasing function of } z ,
\]

for \( n = 2, 3, \ldots \),

144
where $U_n \equiv n^{th}$ departure time from $c$-system, and $V_n \equiv n^{th}$ departure time from $(c+1)$-system.

By Lemma 1.1, for any $t > z$, and $n = 2, 3, \ldots$,

$$\text{Prob}\{U_n > t | U_{n-1} = z\} = \exp\left(-\int_z^t \eta_c(\tau) d\tau \right),$$

$$\text{Prob}\{V_n > t | V_{n-1} = z\} = \exp\left(-\int_z^t \eta_{c+1}(\tau) d\tau \right),$$

where $\eta_c(\tau) = \text{mean output rate of the } c\text{-system at time } \tau$, given $U_{n-1} = z$ and no departures in $(z, \tau]$, for $z \leq \tau \leq t$, and $\eta_{c+1}(\tau)$ is defined similarly. Now $\eta_c(\tau) = ((n_c(\tau) - n + 1)\mu) \wedge (c\mu)$, and $\eta_{c+1}(\tau) = ((n_c(\tau) - n + 1)\mu') \wedge ((c+1)\mu')$. Therefore, $\eta_c(\tau) \geq \eta_{c+1}(\tau)$, for all $z \leq \tau \leq t$, since $\mu > \mu'$ and $c\mu = (c+1)\mu'$. Therefore, using (34), we obtain, for all $t > z$,

$$\text{Prob}\{U_n > t | U_{n-1} = z\} \leq \text{Prob}\{V_n > t | V_{n-1} = z\}.$$  

Condition (32) for $n = 1$ is proved in the same way.

To prove (33) we note that raising the value of $z$ to $z' > z$ in (34) doesn't affect $\eta_c(\tau)$, for $z' \leq \tau \leq t$. Therefore, since $\eta_c(\tau) \geq 0$ for all $\tau$, $\int_z^t \eta_c(\tau) d\tau \geq \int_z^t \eta_c(\tau) d\tau$, and therefore

$$\text{Prob}\{U_n > t | U_{n-1} = z\} \leq \text{Prob}\{U_n > t | U_{n-1} = z'\} \text{ for } z' > z.$$  

Q.E.D.

The following corollary yields the property which we need in order to limit the search for the optimal constrained pair $(\hat{c}', \hat{\mu}')$. 

145
COROLLARY 6.13: Suppose the service-time distribution in a Type I-A queueing system is exponential. Then, for objective functions of the form (31),

\[(35) \quad E[C(c, \mu_c)] \leq E[C(c+1, \mu_{c+1})], \quad c = 1, 2, \ldots \]

It is useful to know that the search for the optimal constrained pair \((c', \mu')\) can be limited to a finite number of values of \(c\) if we have \(\mu_1 \leq \mu\) for some \(i\). However, the usefulness of this property is somewhat limited if we cannot be reasonably assured of finding a value \(c = i\) such that \(\mu_i \leq \mu\) without searching a large number of values of \(c\).

The following conjecture about general queueing systems, together with some numerical evidence about a particular system, suggests that we can find a value \(c = i\) such that \(\mu_i \leq \mu\) relatively quickly in most cases.

CONJECTURE I: Consider a Type I-A queueing system in which the objective is to choose \((c, \mu)\) to minimize

\[(36) \quad E[C(c, \mu)] = c_s(c, \mu) + \lambda c_w E[h(W(c, \mu))] \]

among those pairs \((c, \mu)\) such that \(c = 1, 2, \ldots\), and \(0 < \mu < \infty\).

Suppose that \(c_s(c, \mu)\) is monotone increasing in \(c\) and \(\mu\), and that \(c_s(1, \mu c) \leq c_s(c, \mu)\) for all feasible pairs \((c, \mu)\). Suppose that \(W(c, \mu)\) is stochastically monotone increasing in \(c\) and \(\mu\) and \(h(\cdot)\) is a monotone increasing Borel function. Then, for such a system,
\[(37) \quad h_1 \geq h_2 \geq \ldots , \]

\[(38) \quad \lim_{c \to \infty} \hat{\mu}_c = 0 . \]

In Theorem 6.14 below we will exhibit a simple necessary and sufficient condition for (37) to be valid for linear service-cost under some plausible regularity assumptions about the form of the expected waiting-cost function and the function \( h(\cdot) \). For convenience, in the remainder of this section, we will assume that \( \lambda = 1 \) and \( C_w = l \) (i.e., time is measured in units of mean inter-arrival time and cost rate for the system is measured in units of the waiting-cost coefficient). The objective function will now be written

\[(39) \quad E[C(c, \mu)] = cK\mu + f_c(\mu) , \]

where \( K = C_s \) in the new cost-rate units, and \( f_c(\mu) = E[h(W(c, \mu))] \).

**ASSUMPTION I:** For any \( n = 1, 2, \ldots \), \( f_n(\mu) \) is a differentiable convex monotone decreasing function of \( \mu \), with

\[\lim_{\mu \downarrow 1/n} f_n(\mu) = +\infty , \]

\[\lim_{\mu \uparrow +\infty} f_n(\mu) = 0 . \]

This assumption is a concise statement of a sufficient condition for the theorem we are about to prove. It would be useful to specify directly some sufficient conditions on \( h(\cdot) \) and \( W(n, \mu) \) for the assumption to hold, but we will not consider this problem here. The
conditions do not seem severe. The two limits indicate that the expected waiting-cost rate approaches \( +\infty \) as the traffic intensity \( \rho \) approaches 1, and the rate approaches 0 as \( \rho \) approaches 0; both of these are reasonable requirements. The convexity requirement also seems reasonable; intuition tells us that, for \( \mu \) only slightly larger than \( \frac{1}{n} \) (\( \rho \) close to 1), small increases in \( \mu \) will bring about large decreases in congestion and hence large decreases in the waiting-cost rate, whereas when \( \mu \) gets large (\( \rho \) approaches 0) it will take relatively large increases in \( \mu \) to produce minor decreases in congestion.

Let \( \hat{\mu}_c(K) \) denote the optimal unconstrained value of objective function (39) as a function of the cost coefficient \( K \).

**Theorem 6.11:** Under Assumption I, a necessary and sufficient condition for (37) to hold for all values of \( K, 0 < K < \infty \), in objective function (39) is:

\[
(40) \quad (c+1)f_c^*(\mu) \leq cf_{c+1}^*(\mu), \quad \text{for all } \frac{1}{c} < \mu < \infty, \quad c = 1, 2, \ldots.
\]

**Proof.**

Under Assumption I, for any \( K, 0 < K < \infty \), and \( n = 1, 2, \ldots \),

\( E[C(n, \mu)] \) is a differentiable, convex function of \( \mu, \frac{1}{n} < \mu < \infty \), and

\[
\lim_{\mu \downarrow 1/n} E[C(n, \mu)] = +\infty,
\]

\[
\lim_{\mu \uparrow +\infty} E[C(n, \mu)] = +\infty.
\]

Therefore, \( \hat{\mu}_n \) is the unique solution of \( \frac{\partial}{\partial \mu} E[C(n, \mu)] = 0 \). But from
(39) we have:

$$\frac{\partial}{\partial \mu} E\{C(n, \mu)\} = nK + f'_n(\mu).$$

Therefore, \( \hat{\mu}_n(K) \) is the unique solution of

\[(41) \quad f'_n(\mu) = -nk.\]

Assume we are given a \( K, \ 0 < K < \infty \). We will show that (40) holds for \( \mu = \hat{\mu}_c(K) \) if and only if \( \hat{\mu}_c(K) \geq \hat{\mu}_{c+1}(K) \). Substituting \( \mu = \hat{\mu}_c(K) \) in (40) and using (41) with \( n = c \) and \( n = c+1 \):

\[(42) \quad (c+1)f'_c(\hat{\mu}_c(K)) \leq cf'_{c+1}(\hat{\mu}_c(K)) \iff (c+1)(-cK) \leq cf'_{c+1}(\hat{\mu}_c(K)) \]

\[\iff f'_c(\hat{\mu}_c(K)) \geq -(c+1)K\]

\[\iff f'_{c+1}(\hat{\mu}_c(K)) \geq f'_c(\hat{\mu}_{c+1}(K))\]

But \( f'_{c+1}(\mu) \) is a monotone increasing function of \( \mu \) (by the convexity of \( f_{c+1}(\mu) \)). Therefore, from (42),

\[(c+1)f'_c(\hat{\mu}_c(K)) \leq cf'_{c+1}(\hat{\mu}_c(K)) \iff \hat{\mu}_c(K) \geq \hat{\mu}_{c+1}(K),\]

for all \( 0 < K < \infty, \ c = 1, 2, \ldots \).

But, given \( c = 1, 2, \ldots \), and any \( \mu, \ \frac{1}{c} < \mu < \infty \), there exists a \( K, \ 0 < K < \infty \), such that \( \mu = \hat{\mu}_c(K) \). For let \( K = -\frac{1}{c} f'_c(\mu) \); then \( \mu \) satisfies (41) for that value of \( K \) and \( n = c \), and hence \( \mu = \hat{\mu}_c(K) \). Therefore, (40) is a necessary and sufficient condition for (37) to hold for all \( K, \ 0 < K < \infty \).

Q.E.D.
COROLLARY 6.15: Under Assumption I, a sufficient condition for

\[(43) \quad \hat{\mu}_1(K) \geq \hat{\mu}_2(K) \geq \cdots, \text{ for all } 0 < K < \infty, \]

is

\[(44) \quad f_c'(\mu) \leq f_{c+1}'(\mu), \text{ all } \frac{1}{c} < \mu < \infty, \quad c = 1, 2, \ldots. \]

Proof.

Since both \(f_c'(\mu) \leq 0\) and \(f_{c+1}'(\mu) \leq 0\) for all \(\mu > \frac{1}{c}\), this follows immediately from Theorem 6.14, since \((44)\) implies \((40)\).

Q.E.D.

Again, our intuition tells us that condition \((44)\) is a reasonable one. For \(\mu\) close to \(\frac{1}{c}\) at least, \((44)\) must hold, since \(f_{c+1}'(\frac{1}{c}) > -\infty\) and \(\lim_{\mu \downarrow 1/c} f_c'(\mu) = -\infty\) (by the convexity of \(f_c(\mu)\) and the fact that \(\lim_{\mu \downarrow 1/c} f_c(\mu) = +\infty\)).

In any case, numerical analysis of the \(M|M|c\) system yields evidence for the conjecture that \((37)\) holds for all \(0 < K < \infty\) and for fairly general non-decreasing waiting-cost functions \(h(\cdot)\). (See the appendix for a detailed description of this analysis.) In the numerical analysis we considered waiting-cost functions of the form

\[(45) \quad h(x) = x^A, \quad A > 0.\]

Note that \(h(\cdot)\) is convex for \(A \geq 1\) and concave for \(A \leq 1\).

In Figure 5 below we have considered the case \(A = 1\) and graphed \(\hat{\mu}_c\) as a function of \(K\) for \(c = 1, 2, \ldots, 10\). In Figure 6, \(K = 1\),

150
FIGURE 5

Optimal $\hat{\mu}_c$ as Functions of $K$, $c = 1,2,3,4,5$

($A = 1$)
Optimal $\hat{\theta}_c$ as Functions of $A$, $c = 1, 2, 3, 4, 5$, ($K = 1$)
and \( \hat{\mu}_c \) is plotted as a function of \( A \), for \( c = 1, 2, \ldots, 10 \). Note that in Fig. 5 the curve for \( \hat{\mu}_c \) lies uniformly above that for \( \hat{\mu}_{c+1} \), for all \( c \). In Fig. 6 the curve for \( \hat{\mu}_c \) as a function of \( A \) lies above that for \( \hat{\mu}_{c+1} \), with minor exceptions for large values of \( A \). Thus we have exhibited cases where Conjecture I is not valid. However, we have also demonstrated that it is valid for a wide range of values of \( K \) and a broad class of expected waiting-cost functions.

Theorem 6.12 and Conjecture I yield the following algorithm for determining the optimal pair \( (\hat{e}', \hat{\mu}') \) for objective functions of the form (39), subject to \( \mu \leq \bar{\mu} \):

**ALGORITHM:**

**Step 0:** Calculate \( c_o = \min\{c | \bar{\mu} > \frac{1}{c} \} \) and \( \hat{\mu}_c \). If \( \hat{\mu}_c \leq \bar{\mu} \), let \( M = E[C(c_o, \bar{\mu})] \) and \( c(M) = c_o, \mu(M) = \hat{\mu}_c \), and stop. If \( \hat{\mu}_c > \bar{\mu} \), let \( M = E[C(c_o, \bar{\mu})] \), \( c(M) = c_o, \mu(M) = \bar{\mu} \), and go to **Step c**, with \( c = c_o + 1 \).

**Step c:** Calculate \( \hat{\mu}_c \).

(a) If \( \hat{\mu}_c \leq \bar{\mu} \), compare \( M \) and \( E[C(c, \hat{\mu}_c)] \). If \( M \leq E[C(c, \hat{\mu}_c)] \), then stop. If \( M > E[C(c, \hat{\mu}_c)] \) let \( M = E[C(c, \hat{\mu}_c)] \), \( c(M) = c, \mu(M) = \hat{\mu}_c \), and stop.

(b) If \( \hat{\mu}_c > \bar{\mu} \), calculate \( E[C(c, \bar{\mu})] \). Compare \( M \) and \( E[C(c, \bar{\mu})] \). If \( M \leq E[C(c, \bar{\mu})] \), then let \( c = c+1 \) and return to the beginning of **Step c**. If \( M > E[C(c, \bar{\mu})] \), let \( M = E[C(c, \bar{\mu})] \), \( c(m) = c, \mu(m) = \bar{\mu} \), and return to the beginning of **Step c**, with \( c = c+1 \).
Thus the algorithm terminates when and only when a value \( c = i \) is found such that \( \hat{\mu}_i \leq \bar{\mu} \). It calculates the constrained optimum for each \( c < i \) (since \( E[C(c, \mu)] \) is convex in \( \mu \) and \( \hat{\mu}_c > \bar{\mu} \) for \( c < i \), this optimum always occurs at \( \mu = \bar{\mu} \)); the lower of these optima is carried forth for comparison with \( E[C(i, \hat{\mu}_i)] \) and the lower of these two must be the overall constrained optimum. When the algorithm terminates, \( c(M) = \hat{\sigma}' \), \( \mu(M) = \hat{\mu}' \), and \( M = \text{minimal total expected cost} \).

We state the following theorem, the proof of which is immediate.

**Theorem 6.16:** Consider a Type I-E queueing system with objective function (39) and the constraint, \( \mu \leq \bar{\mu} \), and suppose that the conditions of Corollary 6.13, Assumption I, and (40) hold. Then the above algorithm will yield the optimal constrained pair \( (\hat{\sigma}', \hat{\mu}') \) for (39).

It would be helpful if \( E[C(c, \bar{\mu})] \) were a monotonic function of \( c \), for \( \frac{1}{\mu} < c < \min \{ i \mid \hat{\mu}_i \leq \bar{\mu} \} \). In this case, it would not be necessary to calculate and compare the values of \( E[C(c, \bar{\mu})] \) for all \( c \) in this interval: we would only have to calculate the first or the last and compare it with \( E[C(i, \hat{\mu}_i)] \). However, the numerical analysis of the \( M|M|c \) system has demonstrated that \( E[C(c, \bar{\mu})] \) is not in general a monotonic function of \( c \) over this range. In Figure 7, for example, the values of \( E[C(c, \bar{\mu})] \) are plotted for \( A = 1 \), \( K = 0.25 \) and \( \bar{\mu} = 0.75 \).

6.7 Team Model

In all the models described so far in this paper, we have assumed that it is possible to vary the mean service rate of a server without changing the form of the service-time distribution. Moreover, for each of the service-time distributions which have been considered —
FIGURE 7

$E[C(c, \bar{\mu})]$ as Function of $c = 2, 3, 4, 5, 6, 7$

($A = 1, \ K = 0.25, \ \bar{\mu} = 0.75$)
degenerate, exponential, and k-Erlang — the higher moments are specified automatically when the mean service time (or rate) is specified; for each of these distributions, multiplying the mean service time by a factor \( k \) results in the standard deviation being multiplied by the same factor \( k \). Therefore, if the service-cost rate is linear, i.e., 
\[ C_s(c, \mu) = C_s \mu, \]
then it is possible simultaneously (i) to reduce the mean service time of a server by a factor of \( \frac{1}{c} \), (ii) reduce the standard deviation of service time by a factor of \( \frac{1}{c} \), and (iii) increase the service-cost rate by a factor of \( c \). Because of these properties, we could expect to be indifferent between \( c \) slow servers and one server working \( c \) times as fast on the average, as long as the \( c \) slow servers are always busy when the first server is busy. As we have seen, it is the fact that some of the \( c \) slow servers are sometimes idle while the fast server is busy that allows the system with the single fast server to dominate the system with the \( c \) slow servers.

In many practical applications, however, it may not be possible to do (i), (ii), and (iii) simultaneously. In Section 3.2 we considered briefly some cases where (iii) is violated whenever we do (i) and (ii), that is, doing (i) and (ii) necessarily increases the service-cost rate by more than a factor of \( c \). In this section we take a different approach; we consider a model in which the service-cost rate is linear, but (i) and (ii) are violated whenever we replace \( c \) slow servers by one fast server whose service-cost rate is \( c \) times that of each of the slow servers.

The model considered is one in which a server may be a single man or a team of several men. In such a case there may be inefficiencies
when several men work together in a team on a single customer. The result of these inefficiencies may be that $c$ men work together on a team less than $c$ times as fast on the average than each of them works alone.

TEAM MODEL: Consider a single-station queueing system fed by an arrival process, $0 \leq A_1 \leq A_2 \leq \cdots$, with a strict, FCFS queue discipline. The service mechanism consists of $c$ servers, where $c$ is fixed. Each arriving customer's service consists of $c$ separate tasks which must be completed. The time required for a server to complete a task has the same exponential distribution, with mean rate $\mu$, for all servers and all tasks, where $\mu$ is a decision variable. The other decision to be made is how to assign tasks to servers (i.e., how to structure the service mechanism). There are two options:

1-system: All $c$ servers form a team and work on a single customer. Each server performs one of the tasks. The times required for the servers to complete the tasks are mutually independent. The service time of a customer is $\max(S_1, \ldots, S_c)$, where $S_i$ is the time to complete the $i$th task (distributed exponentially with mean rate $\mu$);

$c$-system: All the tasks of a customer are performed sequentially by the same server. The times to complete the tasks are mutually independent. The service time of a customer is $\sum_{i=1}^{c} S_i$, where $S_i$ is the time to complete the $i$th task, $i = 1, 2, \ldots, c$.

An example of a practical application for which this model might be useful is the operation of a maintenance shop for aircraft. The assumption that each customer has the same number of tasks is appropriate
in situations where it is economical to perform a complete maintenance on an aircraft whenever any repair is needed.

We will consider objective functions of the form

\[(46) \quad E[C(n, \mu)] = cK\mu + f_n(\mu),\]

where \(K\) and \(f_n(\mu)\) are as defined in Section 6.6 and where for this model \(n = 1\) or \(n = c\). If we could easily calculate \(\hat{\mu}_n\) and \(\hat{\mu}_c\) (the optimal values of \(\mu\) for the 1-system and c-system, respectively), then we could find the optimal pair \((\hat{n}, \hat{\mu}_n)\) for (46) simply by comparing \(E[C(1, \hat{\mu}_1)]\) and \(E[C(c, \hat{\mu}_c)]\). Barring this possibility, we would at least like to determine if either the 1-system or the c-system is optimal for all \(K\) and a general class of expected waiting-cost rates \(f_n(\mu)\), \(n = 1, c\). To this end we make the following assumption about \(f_n(\mu)\):

**ASSUMPTION I:** \(f_1(\mu)\) and \(f_c(\mu)\) are both convex, differentiable, monotone decreasing functions of \(\mu\), \(\mu > \frac{1}{c}\). Moreover,

\[
\lim_{\mu \downarrow 1/c} f_c(\mu) = +\infty,
\]

\[
\lim_{\mu \uparrow +\infty} f_1(\mu) = 0.
\]

The following conjecture, if valid, implies that there will not be clear dominance of one of the systems by the other for all \(K\) and all waiting-cost functions.

**CONJECTURE I:** If a queueing system defined by the team model satisfies Assumption I then there exists a point \(\mu^*\) such that
\[ f_1(\mu) \leq f_c(\mu), \quad \mu \leq \mu^* \]
\[ f_1(\mu) \geq f_c(\mu), \quad \mu \geq \mu^* \]

The following intuitive argument shows why this conjecture is reasonable:

From the definition of the Team Model, we note that the service-time distribution of a customer in the c-system has a c-Erlang distribution with mean rate \( \mu \). The service time of a customer in the l-system is distributed as the maximum of \( c \) independent random variables, each exponentially distributed with mean rate \( c\mu \). However, it is more useful to consider the time between completions of tasks in the l-system and express the service time of a customer in terms of these. Thus the service time of a customer in the l-system is distributed as \( \sum_{i=1}^{c} Y_i \), where \( Y_i \) is exponentially distributed with mean rate \( (c-i+1)c\mu \), \( i = 1, 2, \ldots, c \), and \( Y_1, Y_2, \ldots, Y_c \) are mutually independent.

Recall that in the proof of Theorem 2.1 for k-Erlang service-time distributions, the crucial condition used to prove that the l-system had stochastically fewer customers was that, for any number of phases in the system \( (n) \), the mean output rate of phases \( (\mu_n) \) was larger in the l-system. Now, in the l-system as here defined, the mean output rate of tasks ranges cyclically through the values \( c(c\mu), (c-1)c\mu, \ldots, c\mu \), whereas the mean output rate of tasks in the c-system here is the same as in Chapter II (with \( k = c \)), namely, the number of busy servers times \( c\mu \). Therefore, if the number in the system is large, the mean output rate of tasks in the c-system will be \( c(c\mu) \), whereas the mean
output rate of tasks in the $l$-system is usually less than $c(c \mu)$ and 
is never larger. However, if there is only one task in the system, 
then the mean output rate of tasks in the $c$-system will be $c \mu$, whereas 
the mean output rate in the $l$-system is usually greater than $c \mu$ and 
is never smaller.

Now, for large enough values of $\mu$, the number of customers in 
either system will be stochastically small and thus, by the above rea-
soning, the mean output rate will be higher most of the time in the 
$l$-system. Therefore, we can expect $f'_l(\mu) \leq f'_c(\mu)$, for large values 
of $\mu$. Similarly, for small values of $\mu$, the number in the system 
will be stochastically large and thus the mean output rate will be higher 
most of the time in the $c$-system. Therefore, we can expect $f'_c(\mu) \leq f'_l(\mu)$, 
for small values of $\mu$. Moreover, it seems reasonable that the curves 
of $f'_l(\mu)$ and $f'_c(\mu)$ will cross at only one point. Call that point 
$\mu^*$.

This completes the heuristic argument for Conjecture I.

We are now in a position to prove the following theorem.

**Theorem 6.17:** Consider a queueing system defined by the Team Model.

If Assumption I holds and Conjecture I is valid, then there exist numbers 
$K \leq \bar{K}$ such that for $K \leq \bar{K}$ the pair $(1, \hat{\mu}_l(K))$ is optimal for (46), 
and for $K \geq \bar{K}$ the pair $(c, \hat{\mu}_c(K))$ is optimal for (46).

**Proof.**

For any given $K$, $\hat{\mu}_n(K)$ is the unique solution of

$$f'_n(\hat{\mu}_n(K)) = -cK$$

160
for \( n = 1, c \). Therefore, as \( K \) increases, \( f_1'(\hat{\mu}_n(K)) \) decreases. But, by Assumption I, \( f_1'(\cdot) \) is a monotone increasing function. Therefore, \( \hat{\mu}_n(K) \) must be a monotone decreasing function of \( K \).

Let \( \mu^* \) be as defined in Conjecture I. Let \( \overline{K} = \frac{1}{c} f_1'(\mu^*) \) and let \( \underline{K} = \frac{1}{c} f'_c(\mu^*) \). Clearly, \( f'_1(\mu^*) \leq f'_c(\mu^*) \), since \( f_1(\mu) \) crosses \( f_c(\mu) \) from above at \( \mu^* \). Therefore, \( \overline{K} \geq \underline{K} \). Moreover, \( \mu^* = \hat{\mu}_1(\overline{K}) \) and \( \mu^* = \hat{\mu}_c(\underline{K}) \), since \( \mu^* \) satisfies (47) for \( n = 1, K = \overline{K} \) and \( n = c, K = \underline{K} \), respectively. Therefore, since \( \hat{\mu}_1(K) \) is monotone decreasing in \( K \), \( \hat{\mu}_1(K) \leq \mu^* \) for \( K \geq \overline{K} \). Similarly, since \( \hat{\mu}_c(K) \) is monotone decreasing in \( K \), \( \hat{\mu}_c(K) \geq \mu^* \) for \( K \leq \underline{K} \).

Suppose \( K \geq \overline{K} \). Then, as we have proved above, \( \hat{\mu}_1(K) \leq \mu^* \).

Therefore, by Conjecture I, \( f_1(\hat{\mu}_1(K)) \geq f_c(\hat{\mu}_1(K)) \). Substituting in the objective function (46), we obtain

\[
(48) \quad E[C(1, \hat{\mu}_1(K))] \geq E[C(c, \hat{\mu}_1(K))] .
\]

But, by the optimality of \( \hat{\mu}_c(K) \) for \( K \),

\[
(49) \quad E[C(c, \hat{\mu}_1(K))] \geq E[C(c, \hat{\mu}_c(K))] .
\]

Therefore, combining (40) and (49),

\[
E[C(1, \hat{\mu}_1(K))] \geq E[C(c, \hat{\mu}_c(K))] ,
\]

for all \( K \geq \overline{K} \). Thus we have proved that the pair \( (c, \hat{\mu}_c(K)) \) is optimal for (46) for \( K \geq \overline{K} \).

In exactly the same manner, it can be proved that \( (1, \hat{\mu}_1(K)) \) is optimal for (46) for \( K \leq \underline{K} \).
This completes the proof of Theorem 6.17.

Q.E.D.

Note that Theorem 6.17 tells us nothing about which system is optimal in the interval \( K < K < \bar{K} \). In order to draw conclusions about the behavior of the objective functions for \( K \) in this interval, it would be necessary to make assumptions about the higher order derivatives of \( f_1(\mu) \) and \( f_c(\mu) \).

The most important implications of Theorem 6.17 are qualitative:

1° - For small fixed values of \( \mu \), both systems tend to be highly saturated and therefore the c-system minimizes (46), because in saturated conditions the expected waiting cost is lower in the c-system than in the l-system. For large fixed values of \( \mu \), the systems tend to be highly unsaturated and therefore the l-system minimizes (46), because in unsaturated conditions the expected waiting cost is lower in the l-system.

2° - Suppose the value of \( \mu \) may be varied by the user of the system. Then for large values of \( K \), the optimal values of \( \mu \) in both systems will tend to be small, since service cost outweighs waiting cost; therefore, by implication 1° above, the c-system (with \( \mu = \hat{\mu}_c(K) \)) will tend to minimize (46) in such situations. For small values of \( K \), the optimal values of \( \mu \) in both systems will tend to be large, since waiting cost outweighs service cost; therefore, in such situations, the l-system (with \( \mu = \hat{\mu}_l(K) \)) will tend to minimize the objective function.
6.8 Summary and Conclusions

We have considered the optimal design of queueing systems in which the decision variable $\delta$ is a measure of the service capacity and the objective is to minimize the steady-state expected operating cost rate of the system. The objective functions studied were of the form

$$E[C(\delta)] = E[C_s(\delta)] + E[C_w(\delta)],$$

where $C_s(\delta)$ is the cost of operating the servers and $C_w(\delta)$ is the cost due to waiting by customers, when the designed service capacity is $\delta$.

In most of the models considered, the decision variables were the number of servers ($c$) and the mean service rate ($\mu$). The main result of the paper was that, for a wide class of arrival processes and service-time distributions, a wide variety of service and waiting cost functions, and a wide variety of system structures and operating policies, a single-server system is optimal.

In Chapter II this result was demonstrated for systems with general arrival processes and degenerate, exponential, or $k$-Erlang service-time distributions, with service costs proportional to both $c$ and $\mu$ and waiting costs proportional to the number of customers in the system, and consisting of a single station under a strict queue discipline and certain regularity conditions. The result was proved as a corollary to a theorem which stated that, for any realization of the arrival process, the number of customers in the system at any time is stochastically minimized, among systems with equal total service capacity, $c\mu$, by the
system with \( c = 1 \).

This basic theorem was extended in Chapter III, to single-station systems with non-linear service costs which satisfy \( C_s(1, c\mu) \leq C_s(c, \mu) \) for all feasible pairs \((c, \mu)\). The special case where

\[
C_s(c, \mu) = cC_s(\mu)
\]

was studied. It was shown that the above inequality holds if \( C_s(\mu) \) is concave, quasi-homogeneous, or subadditive. Several tests for subadditivity of a function were given.

In Chapter IV the optimality of the single-server system was proved for single-station systems with certain types of non-linear waiting-cost functions. It was shown that, under very general assumptions (essentially those of Little [20] or Jewell [15]),

\[
E\{\text{system waiting-cost rate}\} = \lambda \cdot E\{\text{customer waiting cost}\}.
\]

We proved that the expected customer waiting cost is minimized, among systems with equal total capacity \( c\mu \), by a single-server system if either (i) the service-time distribution is \( k \)-Erlang and the waiting cost of a customer is a certain type of convex monotone increasing function of his waiting time, or (ii) the service-time distribution is exponential and the customer waiting-cost function is monotone increasing (but not necessarily convex) in the waiting time. Combining these results yields the optimality of a single-server system for the total cost function.

In Chapter V the results of Chapter II were extended to series
and partially-ordered networks of queues with exponential service-time distributions.

In Chapter VI we studied single-server systems with different queue disciplines, different rules for charging the service cost, generalized sets of decision variables, and constraints on the decision variables. The results of Chapter II extended to several of these systems. A team model was also studied.

There are many directions which future research in this area might take. An obvious one is extension of the optimality of the single-server system to even more general systems than those considered in this paper. At the end of several of the chapters we have indicated more specific possibilities for future work.
APPENDIX

Numerical Analysis of M|M|c System

In this appendix some of the results of a numerical analysis of the M|M|c system are listed in tables. The objective function considered is

\[ E[G(c, \rho)] = \frac{K}{\rho} + E[(W(c, \rho))^A] \]

where \( c \) = number of servers,
\( \rho \) = traffic intensity
\( K \) = service-cost rate,
\( W(c, \rho) \) = steady-state waiting time,

and \( A \) is a positive constant. We assume that the mean arrival rate \( \lambda \) is 1 and that the waiting-cost rate \( (c')_w \) is 1 (or, equivalently, that time is measured in units of mean interarrival time and cost is measured in units of waiting-cost rate).

For M|M|c systems we have:

\[ E[(W(c, \rho))^A] = 1(A+1) \left[ (c\rho)^A + \bar{F}(c, \rho) \left( \frac{\rho}{1-\rho} \right)^A \left( \frac{(c(1-\rho))A-1}{c(1-\rho)-1} \right) \right] \]

where

\[ \bar{F}(c, \rho) = \text{Prob\{an arrival must wait\}} \]

\[ = \frac{\sum_{i=0}^{c-1} \frac{(c\rho)^i}{i!}}{\sum_{i=0}^{\infty} \frac{(c\rho)^i}{i!} + \frac{\rho^c}{c!(1-\rho)}} \]
(For the analysis of the M|M|c system upon which these results are based, see Saaty [29].)

On page 169 the values of \( \overline{F}(c, \rho) \) are tabulated for \( c = 1, 2, \ldots, 10 \), and \( \rho = 0.01, 0.02, \ldots, 0.99 \).

On pages 170 through 180 the values of \( E[(W(c, \rho))^A] \) are tabulated for representative values of \( A \) from 0.1 to 10.0 (the successive values are in constant ratio to one another) and \( c = 1, 2, \ldots, 10 \), and \( \rho = 0.01, 0.02, \ldots, 0.99 \). The Case \( A = 1.00 \) (linear waiting cost) is summarized in Figure 2 on page 76 of the text.

On pages 181 through 191 the values of \( E[C(c, \rho)] \) are tabulated for \( A \) fixed at 1.00, for representative values of \( K \) from 0.1 to 10.0 (with successive values in constant ratio), and \( c = 1, 2, \ldots, 10 \), and \( \rho = 0.01, 0.02, \ldots, 0.99 \). At the bottom of each table are listed the minimum value of the objective function, \( E[C(c, \mu)] \), the optimal traffic intensity, \( \hat{\rho}_c \), and the optimal mean service rate, \( \hat{\mu}_c \), for \( c = 1, 2, \ldots, 10 \). These tables are summarized in Figure 5 in Chapter VI.

On page 192 through 201 the values of \( E[C(c, \rho)] \) are tabulated again, this time for \( K \) fixed at 1.00 and \( A \) ranging from 0.1 to 10.0 (with successive values in constant ratio). Again the optimal values of the objective function, the traffic intensity, and the mean service rate are listed at the bottom of the table. These tables are summarized in Figure 6.

The final table, on page 202, lists the values of the objective function as a function of \( c \) and \( \overline{\mu} \) for \( A = 1.00 \) and \( K = 0.25 \), where \( \overline{\mu} \) is assumed to be an upper bound on \( \mu \) (cf. Section 6.6) and the objective function values are tabulated only for \( \frac{1}{\mu} < c < \min\{i | \hat{\mu}_i \leq \overline{\mu}\} \).
that is, for those values of $c$ for which the objective function is finite (i.e., the steady state exists) and the optimal unconstrained value of $\mu$ lies above the upper bound.
<table>
<thead>
<tr>
<th>x</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.6099</td>
<td>0.6349</td>
<td>0.6690</td>
<td>0.0895</td>
<td>0.7051</td>
<td>0.7181</td>
<td>0.7287</td>
<td>0.7390</td>
<td>0.7476</td>
<td>0.7557</td>
</tr>
<tr>
<td>0.02</td>
<td>0.6263</td>
<td>0.6546</td>
<td>0.6890</td>
<td>0.0902</td>
<td>0.7178</td>
<td>0.7307</td>
<td>0.7409</td>
<td>0.7504</td>
<td>0.7582</td>
<td>0.7657</td>
</tr>
<tr>
<td>0.03</td>
<td>0.6472</td>
<td>0.6781</td>
<td>0.7148</td>
<td>0.0909</td>
<td>0.7370</td>
<td>0.7495</td>
<td>0.7591</td>
<td>0.7680</td>
<td>0.7756</td>
<td>0.7828</td>
</tr>
<tr>
<td>0.04</td>
<td>0.6690</td>
<td>0.7015</td>
<td>0.7374</td>
<td>0.0916</td>
<td>0.7557</td>
<td>0.7678</td>
<td>0.7771</td>
<td>0.7858</td>
<td>0.7933</td>
<td>0.7999</td>
</tr>
<tr>
<td>0.05</td>
<td>0.6919</td>
<td>0.7246</td>
<td>0.7600</td>
<td>0.0922</td>
<td>0.7791</td>
<td>0.7907</td>
<td>0.7996</td>
<td>0.8078</td>
<td>0.8150</td>
<td>0.8216</td>
</tr>
<tr>
<td>0.06</td>
<td>0.7152</td>
<td>0.7431</td>
<td>0.7933</td>
<td>0.0927</td>
<td>0.8052</td>
<td>0.8167</td>
<td>0.8253</td>
<td>0.8333</td>
<td>0.8406</td>
<td>0.8471</td>
</tr>
<tr>
<td>0.07</td>
<td>0.7389</td>
<td>0.7690</td>
<td>0.8310</td>
<td>0.0931</td>
<td>0.8240</td>
<td>0.8352</td>
<td>0.8434</td>
<td>0.8511</td>
<td>0.8580</td>
<td>0.8643</td>
</tr>
<tr>
<td>0.08</td>
<td>0.7629</td>
<td>0.7973</td>
<td>0.8774</td>
<td>0.0935</td>
<td>0.8472</td>
<td>0.8583</td>
<td>0.8665</td>
<td>0.8740</td>
<td>0.8816</td>
<td>0.8882</td>
</tr>
<tr>
<td>0.09</td>
<td>0.7875</td>
<td>0.8250</td>
<td>0.9287</td>
<td>0.0938</td>
<td>0.8722</td>
<td>0.8832</td>
<td>0.8910</td>
<td>0.8980</td>
<td>0.9051</td>
<td>0.9116</td>
</tr>
<tr>
<td>0.10</td>
<td>0.8128</td>
<td>0.8530</td>
<td>0.9755</td>
<td>0.0941</td>
<td>0.8941</td>
<td>0.9049</td>
<td>0.9123</td>
<td>0.9192</td>
<td>0.9259</td>
<td>0.9322</td>
</tr>
</tbody>
</table>
### Expected Waiting Cost

<table>
<thead>
<tr>
<th>C</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.4691</td>
<td>0.5154</td>
<td>0.5716</td>
<td>0.7797</td>
<td>0.9378</td>
<td>0.9790</td>
<td>0.6803</td>
<td>0.6876</td>
<td>0.6924</td>
<td>0.6999</td>
</tr>
<tr>
<td>0.02</td>
<td>0.5207</td>
<td>0.5594</td>
<td>0.6096</td>
<td>0.7797</td>
<td>0.9378</td>
<td>0.9790</td>
<td>0.6803</td>
<td>0.6876</td>
<td>0.6924</td>
<td>0.6999</td>
</tr>
<tr>
<td>0.03</td>
<td>0.5662</td>
<td>0.5957</td>
<td>0.6352</td>
<td>0.7668</td>
<td>0.9098</td>
<td>0.9490</td>
<td>0.6708</td>
<td>0.6671</td>
<td>0.6718</td>
<td>0.6791</td>
</tr>
<tr>
<td>0.04</td>
<td>0.6133</td>
<td>0.6407</td>
<td>0.6673</td>
<td>0.7435</td>
<td>0.9129</td>
<td>0.9541</td>
<td>0.6560</td>
<td>0.6507</td>
<td>0.6544</td>
<td>0.6619</td>
</tr>
<tr>
<td>0.05</td>
<td>0.6612</td>
<td>0.6819</td>
<td>0.7035</td>
<td>0.7177</td>
<td>0.8783</td>
<td>0.9281</td>
<td>0.6388</td>
<td>0.6311</td>
<td>0.6337</td>
<td>0.6402</td>
</tr>
<tr>
<td>0.06</td>
<td>0.7114</td>
<td>0.7221</td>
<td>0.7321</td>
<td>0.7297</td>
<td>0.8531</td>
<td>0.9030</td>
<td>0.6223</td>
<td>0.6134</td>
<td>0.6150</td>
<td>0.6213</td>
</tr>
<tr>
<td>0.07</td>
<td>0.7635</td>
<td>0.7735</td>
<td>0.7791</td>
<td>0.7612</td>
<td>0.8184</td>
<td>0.8677</td>
<td>0.6053</td>
<td>0.5952</td>
<td>0.5964</td>
<td>0.6022</td>
</tr>
<tr>
<td>0.08</td>
<td>0.8183</td>
<td>0.8276</td>
<td>0.8291</td>
<td>0.8060</td>
<td>0.7759</td>
<td>0.8250</td>
<td>0.5875</td>
<td>0.5766</td>
<td>0.5775</td>
<td>0.5828</td>
</tr>
<tr>
<td>0.09</td>
<td>0.8758</td>
<td>0.8844</td>
<td>0.8825</td>
<td>0.8535</td>
<td>0.8174</td>
<td>0.8657</td>
<td>0.5695</td>
<td>0.5577</td>
<td>0.5581</td>
<td>0.5627</td>
</tr>
<tr>
<td>0.1</td>
<td>0.9358</td>
<td>0.9443</td>
<td>0.9408</td>
<td>0.9060</td>
<td>0.8653</td>
<td>0.9132</td>
<td>0.5515</td>
<td>0.5389</td>
<td>0.5393</td>
<td>0.5438</td>
</tr>
<tr>
<td>0.11</td>
<td>0.9979</td>
<td>1.0062</td>
<td>1.0014</td>
<td>0.9577</td>
<td>0.9150</td>
<td>0.9621</td>
<td>0.5335</td>
<td>0.5204</td>
<td>0.5208</td>
<td>0.5254</td>
</tr>
<tr>
<td>0.12</td>
<td>1.0525</td>
<td>1.0604</td>
<td>1.0552</td>
<td>0.9984</td>
<td>0.9555</td>
<td>0.9943</td>
<td>0.5154</td>
<td>0.5025</td>
<td>0.5029</td>
<td>0.5076</td>
</tr>
<tr>
<td>0.13</td>
<td>1.1097</td>
<td>1.1171</td>
<td>1.1117</td>
<td>1.0543</td>
<td>1.0116</td>
<td>1.0482</td>
<td>0.4974</td>
<td>0.4847</td>
<td>0.4851</td>
<td>0.4898</td>
</tr>
<tr>
<td>0.14</td>
<td>1.1698</td>
<td>1.1767</td>
<td>1.1708</td>
<td>1.0907</td>
<td>1.0482</td>
<td>1.0835</td>
<td>0.4805</td>
<td>0.4681</td>
<td>0.4684</td>
<td>0.4732</td>
</tr>
<tr>
<td>0.15</td>
<td>1.2324</td>
<td>1.2387</td>
<td>1.2325</td>
<td>1.1187</td>
<td>1.0774</td>
<td>1.1125</td>
<td>0.4642</td>
<td>0.4520</td>
<td>0.4523</td>
<td>0.4572</td>
</tr>
<tr>
<td>0.16</td>
<td>1.2974</td>
<td>1.3032</td>
<td>1.2969</td>
<td>1.0722</td>
<td>1.0313</td>
<td>1.0664</td>
<td>0.4485</td>
<td>0.4365</td>
<td>0.4368</td>
<td>0.4418</td>
</tr>
<tr>
<td>0.17</td>
<td>1.3641</td>
<td>1.3694</td>
<td>1.3628</td>
<td>1.0268</td>
<td>0.9864</td>
<td>1.0215</td>
<td>0.4331</td>
<td>0.4213</td>
<td>0.4216</td>
<td>0.4268</td>
</tr>
<tr>
<td>0.18</td>
<td>1.4325</td>
<td>1.4373</td>
<td>1.4297</td>
<td>0.9805</td>
<td>0.9411</td>
<td>0.9762</td>
<td>0.4180</td>
<td>0.4064</td>
<td>0.4067</td>
<td>0.4121</td>
</tr>
<tr>
<td>0.19</td>
<td>1.5026</td>
<td>1.5070</td>
<td>1.4989</td>
<td>0.9342</td>
<td>0.8956</td>
<td>0.9303</td>
<td>0.4032</td>
<td>0.3917</td>
<td>0.3920</td>
<td>0.3975</td>
</tr>
<tr>
<td>0.2</td>
<td>1.5744</td>
<td>1.5784</td>
<td>1.5690</td>
<td>0.8873</td>
<td>0.8495</td>
<td>0.8843</td>
<td>0.3886</td>
<td>0.3772</td>
<td>0.3775</td>
<td>0.3833</td>
</tr>
<tr>
<td>0.21</td>
<td>1.6466</td>
<td>1.6502</td>
<td>1.6403</td>
<td>0.8396</td>
<td>0.7927</td>
<td>0.8273</td>
<td>0.3741</td>
<td>0.3629</td>
<td>0.3632</td>
<td>0.3692</td>
</tr>
<tr>
<td>0.22</td>
<td>1.7191</td>
<td>1.7223</td>
<td>1.7119</td>
<td>0.7910</td>
<td>0.7441</td>
<td>0.7784</td>
<td>0.3598</td>
<td>0.3488</td>
<td>0.3490</td>
<td>0.3552</td>
</tr>
<tr>
<td>0.23</td>
<td>1.7919</td>
<td>1.7947</td>
<td>1.7837</td>
<td>0.7413</td>
<td>0.6944</td>
<td>0.7284</td>
<td>0.3457</td>
<td>0.3348</td>
<td>0.3350</td>
<td>0.3415</td>
</tr>
<tr>
<td>0.24</td>
<td>1.8645</td>
<td>1.8669</td>
<td>1.8553</td>
<td>0.6906</td>
<td>0.6437</td>
<td>0.6774</td>
<td>0.3319</td>
<td>0.3212</td>
<td>0.3214</td>
<td>0.3281</td>
</tr>
<tr>
<td>0.25</td>
<td>1.9372</td>
<td>1.9393</td>
<td>1.9271</td>
<td>0.6390</td>
<td>0.5920</td>
<td>0.6257</td>
<td>0.3185</td>
<td>0.3081</td>
<td>0.3083</td>
<td>0.3154</td>
</tr>
<tr>
<td>0.26</td>
<td>2.0099</td>
<td>2.0119</td>
<td>1.9994</td>
<td>0.5864</td>
<td>0.5392</td>
<td>0.5726</td>
<td>0.3055</td>
<td>0.2954</td>
<td>0.2956</td>
<td>0.3029</td>
</tr>
<tr>
<td>0.27</td>
<td>2.0825</td>
<td>2.0842</td>
<td>2.0704</td>
<td>0.5328</td>
<td>0.4850</td>
<td>0.5182</td>
<td>0.2929</td>
<td>0.2831</td>
<td>0.2833</td>
<td>0.2908</td>
</tr>
<tr>
<td>0.28</td>
<td>2.1550</td>
<td>2.1565</td>
<td>2.1423</td>
<td>0.4783</td>
<td>0.4305</td>
<td>0.4635</td>
<td>0.2809</td>
<td>0.2713</td>
<td>0.2715</td>
<td>0.2793</td>
</tr>
<tr>
<td>0.29</td>
<td>2.2272</td>
<td>2.2284</td>
<td>2.2138</td>
<td>0.4230</td>
<td>0.3743</td>
<td>0.4069</td>
<td>0.2694</td>
<td>0.2600</td>
<td>0.2602</td>
<td>0.2683</td>
</tr>
<tr>
<td>0.3</td>
<td>2.3000</td>
<td>2.3008</td>
<td>2.2859</td>
<td>0.3670</td>
<td>0.3179</td>
<td>0.3492</td>
<td>0.2585</td>
<td>0.2494</td>
<td>0.2496</td>
<td>0.2579</td>
</tr>
<tr>
<td>0.31</td>
<td>2.3712</td>
<td>2.3717</td>
<td>2.3568</td>
<td>0.3102</td>
<td>0.2607</td>
<td>0.2916</td>
<td>0.2481</td>
<td>0.2393</td>
<td>0.2395</td>
<td>0.2480</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>C</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.01</td>
<td>0.297</td>
<td>0.339</td>
<td>0.1756</td>
<td>0.4037</td>
<td>0.4270</td>
<td>0.4670</td>
<td>0.4646</td>
<td>0.4805</td>
<td>0.4949</td>
</tr>
<tr>
<td>0.01</td>
<td>0.3696</td>
<td>0.405</td>
<td>0.4058</td>
<td>0.4532</td>
<td>0.4930</td>
<td>0.5350</td>
<td>0.5775</td>
<td>0.6123</td>
<td>0.6332</td>
<td>0.6522</td>
</tr>
<tr>
<td>0.04</td>
<td>0.3738</td>
<td>0.3637</td>
<td>0.4045</td>
<td>0.4080</td>
<td>0.4074</td>
<td>0.4047</td>
<td>0.4021</td>
<td>0.3971</td>
<td>0.3929</td>
<td>0.3876</td>
</tr>
<tr>
<td>0.05</td>
<td>0.3708</td>
<td>0.3604</td>
<td>0.4026</td>
<td>0.4052</td>
<td>0.4039</td>
<td>0.4016</td>
<td>0.3987</td>
<td>0.3958</td>
<td>0.3925</td>
<td>0.3882</td>
</tr>
<tr>
<td>0.06</td>
<td>0.3678</td>
<td>0.3580</td>
<td>0.4007</td>
<td>0.4024</td>
<td>0.4003</td>
<td>0.3974</td>
<td>0.3946</td>
<td>0.3916</td>
<td>0.3883</td>
<td>0.3842</td>
</tr>
<tr>
<td>0.07</td>
<td>0.3658</td>
<td>0.3565</td>
<td>0.3989</td>
<td>0.4007</td>
<td>0.3987</td>
<td>0.3959</td>
<td>0.3929</td>
<td>0.3899</td>
<td>0.3865</td>
<td>0.3825</td>
</tr>
<tr>
<td>0.08</td>
<td>0.3645</td>
<td>0.3553</td>
<td>0.3971</td>
<td>0.3989</td>
<td>0.3970</td>
<td>0.3942</td>
<td>0.3912</td>
<td>0.3882</td>
<td>0.3847</td>
<td>0.3807</td>
</tr>
<tr>
<td>0.09</td>
<td>0.3635</td>
<td>0.3542</td>
<td>0.3959</td>
<td>0.3978</td>
<td>0.3959</td>
<td>0.3932</td>
<td>0.3902</td>
<td>0.3872</td>
<td>0.3836</td>
<td>0.3796</td>
</tr>
<tr>
<td>0.10</td>
<td>0.3625</td>
<td>0.3531</td>
<td>0.3948</td>
<td>0.3967</td>
<td>0.3948</td>
<td>0.3922</td>
<td>0.3892</td>
<td>0.3862</td>
<td>0.3826</td>
<td>0.3786</td>
</tr>
<tr>
<td>0.11</td>
<td>0.3615</td>
<td>0.3520</td>
<td>0.3938</td>
<td>0.3958</td>
<td>0.3939</td>
<td>0.3914</td>
<td>0.3884</td>
<td>0.3854</td>
<td>0.3818</td>
<td>0.3778</td>
</tr>
<tr>
<td>0.12</td>
<td>0.3605</td>
<td>0.3509</td>
<td>0.3928</td>
<td>0.3948</td>
<td>0.3930</td>
<td>0.3905</td>
<td>0.3876</td>
<td>0.3846</td>
<td>0.3810</td>
<td>0.3770</td>
</tr>
<tr>
<td>0.13</td>
<td>0.3596</td>
<td>0.3498</td>
<td>0.3918</td>
<td>0.3939</td>
<td>0.3921</td>
<td>0.3897</td>
<td>0.3868</td>
<td>0.3838</td>
<td>0.3802</td>
<td>0.3762</td>
</tr>
<tr>
<td>0.14</td>
<td>0.3586</td>
<td>0.3487</td>
<td>0.3909</td>
<td>0.3930</td>
<td>0.3912</td>
<td>0.3888</td>
<td>0.3859</td>
<td>0.3829</td>
<td>0.3793</td>
<td>0.3753</td>
</tr>
<tr>
<td>0.15</td>
<td>0.3577</td>
<td>0.3476</td>
<td>0.3899</td>
<td>0.3921</td>
<td>0.3903</td>
<td>0.3879</td>
<td>0.3850</td>
<td>0.3820</td>
<td>0.3784</td>
<td>0.3744</td>
</tr>
</tbody>
</table>

**EXPECTED WAITING COST**
<table>
<thead>
<tr>
<th>C</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.000005836</td>
<td>0.000006718</td>
<td>0.000016855</td>
<td>0.000017740</td>
<td>0.000020777</td>
<td>0.000021663</td>
<td>0.000024738</td>
<td>0.000025624</td>
<td>0.000028691</td>
<td>0.000030578</td>
</tr>
<tr>
<td>0.02</td>
<td>0.000006674</td>
<td>0.000007607</td>
<td>0.000018742</td>
<td>0.000020111</td>
<td>0.000023247</td>
<td>0.000024132</td>
<td>0.000027177</td>
<td>0.000028062</td>
<td>0.000031128</td>
<td>0.000033014</td>
</tr>
<tr>
<td>0.03</td>
<td>0.000007511</td>
<td>0.000008472</td>
<td>0.000021666</td>
<td>0.000022888</td>
<td>0.000026100</td>
<td>0.000027005</td>
<td>0.000030068</td>
<td>0.000029963</td>
<td>0.000033049</td>
<td>0.000034935</td>
</tr>
<tr>
<td>0.04</td>
<td>0.000008348</td>
<td>0.000009246</td>
<td>0.000024897</td>
<td>0.000026119</td>
<td>0.000029333</td>
<td>0.000030240</td>
<td>0.000033271</td>
<td>0.000033176</td>
<td>0.000036262</td>
<td>0.000038149</td>
</tr>
<tr>
<td>0.05</td>
<td>0.000009185</td>
<td>0.000010013</td>
<td>0.000028151</td>
<td>0.000029374</td>
<td>0.000032606</td>
<td>0.000033516</td>
<td>0.000036564</td>
<td>0.000036470</td>
<td>0.000039556</td>
<td>0.000041442</td>
</tr>
<tr>
<td>0.06</td>
<td>0.000009922</td>
<td>0.000010764</td>
<td>0.000031429</td>
<td>0.000032668</td>
<td>0.000035917</td>
<td>0.000036829</td>
<td>0.000039893</td>
<td>0.000039800</td>
<td>0.000042886</td>
<td>0.000044772</td>
</tr>
<tr>
<td>0.07</td>
<td>0.000010659</td>
<td>0.000011414</td>
<td>0.000034723</td>
<td>0.000035975</td>
<td>0.000039239</td>
<td>0.000040156</td>
<td>0.000043230</td>
<td>0.000043137</td>
<td>0.000046223</td>
<td>0.000048109</td>
</tr>
<tr>
<td>0.08</td>
<td>0.000011395</td>
<td>0.000012150</td>
<td>0.000038034</td>
<td>0.000039297</td>
<td>0.000042575</td>
<td>0.000043497</td>
<td>0.000046584</td>
<td>0.000046492</td>
<td>0.000049578</td>
<td>0.000051464</td>
</tr>
<tr>
<td>0.09</td>
<td>0.000012131</td>
<td>0.000012884</td>
<td>0.000041361</td>
<td>0.000042633</td>
<td>0.000045928</td>
<td>0.000046851</td>
<td>0.000049947</td>
<td>0.000049855</td>
<td>0.000052942</td>
<td>0.000054828</td>
</tr>
<tr>
<td>0.10</td>
<td>0.000012867</td>
<td>0.000013608</td>
<td>0.000044707</td>
<td>0.000045994</td>
<td>0.000049316</td>
<td>0.000050247</td>
<td>0.000053352</td>
<td>0.000053260</td>
<td>0.000056346</td>
<td>0.000058232</td>
</tr>
<tr>
<td>0.11</td>
<td>0.000013502</td>
<td>0.000014231</td>
<td>0.000048072</td>
<td>0.000049372</td>
<td>0.000052715</td>
<td>0.000053653</td>
<td>0.000056760</td>
<td>0.000056668</td>
<td>0.000059754</td>
<td>0.000061640</td>
</tr>
<tr>
<td>0.12</td>
<td>0.000014138</td>
<td>0.000014852</td>
<td>0.000051457</td>
<td>0.000052765</td>
<td>0.000056135</td>
<td>0.000057081</td>
<td>0.000060195</td>
<td>0.000059993</td>
<td>0.000063080</td>
<td>0.000064966</td>
</tr>
<tr>
<td>0.13</td>
<td>0.000014774</td>
<td>0.000015469</td>
<td>0.000054857</td>
<td>0.000056177</td>
<td>0.000059572</td>
<td>0.000060528</td>
<td>0.000063642</td>
<td>0.000063439</td>
<td>0.000066526</td>
<td>0.000068412</td>
</tr>
<tr>
<td>0.14</td>
<td>0.000015410</td>
<td>0.000016075</td>
<td>0.000058271</td>
<td>0.000059605</td>
<td>0.000062927</td>
<td>0.000063891</td>
<td>0.000066994</td>
<td>0.000066789</td>
<td>0.000069882</td>
<td>0.000071768</td>
</tr>
<tr>
<td>0.15</td>
<td>0.000016046</td>
<td>0.000016679</td>
<td>0.000061699</td>
<td>0.000063040</td>
<td>0.000066383</td>
<td>0.000067355</td>
<td>0.000070458</td>
<td>0.000069853</td>
<td>0.000072949</td>
<td>0.000074834</td>
</tr>
<tr>
<td>0.16</td>
<td>0.000016682</td>
<td>0.000017278</td>
<td>0.000065130</td>
<td>0.000066481</td>
<td>0.000069845</td>
<td>0.000070825</td>
<td>0.000073926</td>
<td>0.000073321</td>
<td>0.000076416</td>
<td>0.000078301</td>
</tr>
<tr>
<td>0.17</td>
<td>0.000017317</td>
<td>0.000017972</td>
<td>0.000068568</td>
<td>0.000069931</td>
<td>0.000073316</td>
<td>0.000074302</td>
<td>0.000077403</td>
<td>0.000076807</td>
<td>0.000079891</td>
<td>0.000081776</td>
</tr>
<tr>
<td>0.18</td>
<td>0.000017953</td>
<td>0.000018582</td>
<td>0.000071911</td>
<td>0.000073287</td>
<td>0.000076691</td>
<td>0.000077687</td>
<td>0.000080788</td>
<td>0.000079992</td>
<td>0.000083075</td>
<td>0.000084960</td>
</tr>
<tr>
<td>0.19</td>
<td>0.000018589</td>
<td>0.000019196</td>
<td>0.000075262</td>
<td>0.000076650</td>
<td>0.000080067</td>
<td>0.000081073</td>
<td>0.000084173</td>
<td>0.000083277</td>
<td>0.000086359</td>
<td>0.000088244</td>
</tr>
<tr>
<td>0.20</td>
<td>0.000019224</td>
<td>0.000019820</td>
<td>0.000078619</td>
<td>0.000080022</td>
<td>0.000083452</td>
<td>0.000084463</td>
<td>0.000087562</td>
<td>0.000086666</td>
<td>0.000089747</td>
<td>0.000091631</td>
</tr>
</tbody>
</table>

**expected waiting cost**
<table>
<thead>
<tr>
<th>C</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.000</td>
<td>0.002</td>
<td>0.000</td>
<td>0.002</td>
<td>0.000</td>
<td>0.002</td>
<td>0.000</td>
<td>0.002</td>
<td>0.000</td>
<td>0.002</td>
</tr>
<tr>
<td>0.2</td>
<td>0.002</td>
<td>0.001</td>
<td>0.000</td>
<td>0.002</td>
<td>0.001</td>
<td>0.002</td>
<td>0.001</td>
<td>0.002</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td>0.4</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>0.6</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>0.8</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>1.0</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

**EXPECTED WAITING COST**
### Waiting Cost

<table>
<thead>
<tr>
<th>C</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.02</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.03</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.04</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>0.05</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

**Note:**
- The table above illustrates the expected waiting cost for different values of C. Each row represents a different value of C, with the subsequent columns showing the waiting cost for each column value.
- The data is presented in a structured format, typical of a table found in a report or a textbook section discussing cost analysis in a technical or financial context.
<table>
<thead>
<tr>
<th>C</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
<td>0.06</td>
<td>0.07</td>
<td>0.08</td>
<td>0.09</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
<td>0.06</td>
<td>0.07</td>
<td>0.08</td>
<td>0.09</td>
<td>0.10</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
<td>0.06</td>
<td>0.07</td>
<td>0.08</td>
<td>0.09</td>
<td>0.10</td>
<td>0.11</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.05</td>
<td>0.06</td>
<td>0.07</td>
<td>0.08</td>
<td>0.09</td>
<td>0.10</td>
<td>0.11</td>
<td>0.12</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>0.06</td>
<td>0.07</td>
<td>0.08</td>
<td>0.09</td>
<td>0.10</td>
<td>0.11</td>
<td>0.12</td>
<td>0.13</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>0.07</td>
<td>0.08</td>
<td>0.09</td>
<td>0.10</td>
<td>0.11</td>
<td>0.12</td>
<td>0.13</td>
<td>0.14</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>0.08</td>
<td>0.09</td>
<td>0.10</td>
<td>0.11</td>
<td>0.12</td>
<td>0.13</td>
<td>0.14</td>
<td>0.15</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td>0.09</td>
<td>0.10</td>
<td>0.11</td>
<td>0.12</td>
<td>0.13</td>
<td>0.14</td>
<td>0.15</td>
<td>0.16</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>0.09</td>
<td>0.10</td>
<td>0.11</td>
<td>0.12</td>
<td>0.13</td>
<td>0.14</td>
<td>0.15</td>
<td>0.16</td>
<td>0.17</td>
<td>0.18</td>
</tr>
</tbody>
</table>

**EXCEPTED TOTAL COST**

**MINIMIZING AX**

**MINIMIZING BX**

181
<table>
<thead>
<tr>
<th>C</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.03</td>
<td>8.4030 8.4330 8.4630 8.4830 8.5030 8.5230 8.5430 8.5630 8.5830 8.6030</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.04</td>
<td>5.4820 5.5020 5.5220 5.5420 5.5620 5.5820 5.6020 5.6220 5.6420 5.6620</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>3.8790 3.9090 3.9290 3.9490 3.9690 3.9890 4.0090 4.0290 4.0490 4.0690</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.06</td>
<td>2.6260 2.6460 2.6660 2.6860 2.7060 2.7260 2.7460 2.7660 2.7860 2.8060</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.07</td>
<td>1.7820 1.8020 1.8220 1.8420 1.8620 1.8820 1.9020 1.9220 1.9420 1.9620</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>1.2650 1.2850 1.3050 1.3250 1.3450 1.3650 1.3850 1.4050 1.4250 1.4450</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.09</td>
<td>0.9500 0.9700 0.9900 1.0100 1.0300 1.0500 1.0700 1.0900 1.1100 1.1300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>0.7150 0.7350 0.7550 0.7750 0.7950 0.8150 0.8350 0.8550 0.8750 0.8950</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**EXPECTED TOTAL COST**

**MIN EXP COST**

<table>
<thead>
<tr>
<th>C</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.03</td>
<td>8.4030 8.4330 8.4630 8.4830 8.5030 8.5230 8.5430 8.5630 8.5830 8.6030</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.04</td>
<td>5.4820 5.5020 5.5220 5.5420 5.5620 5.5820 5.6020 5.6220 5.6420 5.6620</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>3.8790 3.9090 3.9290 3.9490 3.9690 3.9890 4.0090 4.0290 4.0490 4.0690</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.06</td>
<td>2.6260 2.6460 2.6660 2.6860 2.7060 2.7260 2.7460 2.7660 2.7860 2.8060</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.07</td>
<td>1.7820 1.8020 1.8220 1.8420 1.8620 1.8820 1.9020 1.9220 1.9420 1.9620</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>1.2650 1.2850 1.3050 1.3250 1.3450 1.3650 1.3850 1.4050 1.4250 1.4450</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.09</td>
<td>0.9500 0.9700 0.9900 1.0100 1.0300 1.0500 1.0700 1.0900 1.1100 1.1300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>0.7150 0.7350 0.7550 0.7750 0.7950 0.8150 0.8350 0.8550 0.8750 0.8950</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>----</td>
</tr>
</tbody>
</table>

**EXPECTED TOTAL COST**

- **MIN FIN COST**: 1.6001
- **MINIMIZING RH**: 0.0020 0.0020 0.0020 0.0020 0.0020 0.0020 0.0020 0.0020 0.0020 0.0020
- **MINIMIZING NH**: 2.6041 1.3889 0.9904 0.8065 0.7143 0.9410 0.5952 0.5868 0.5241 0.5000

![Image](image-url)
<table>
<thead>
<tr>
<th>C</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>72.5261</td>
<td>72.5261</td>
<td>72.5261</td>
<td>72.5261</td>
<td>72.5261</td>
<td>72.5261</td>
<td>72.5261</td>
<td>72.5261</td>
<td>72.5261</td>
<td>72.5261</td>
</tr>
<tr>
<td>0.06</td>
<td>123.5608</td>
<td>123.5608</td>
<td>123.5608</td>
<td>123.5608</td>
<td>123.5608</td>
<td>123.5608</td>
<td>123.5608</td>
<td>123.5608</td>
<td>123.5608</td>
<td>123.5608</td>
</tr>
<tr>
<td>0.08</td>
<td>316.4127</td>
<td>316.4127</td>
<td>316.4127</td>
<td>316.4127</td>
<td>316.4127</td>
<td>316.4127</td>
<td>316.4127</td>
<td>316.4127</td>
<td>316.4127</td>
<td>316.4127</td>
</tr>
<tr>
<td>0.09</td>
<td>466.4036</td>
<td>466.4036</td>
<td>466.4036</td>
<td>466.4036</td>
<td>466.4036</td>
<td>466.4036</td>
<td>466.4036</td>
<td>466.4036</td>
<td>466.4036</td>
<td>466.4036</td>
</tr>
<tr>
<td>0.10</td>
<td>657.8840</td>
<td>657.8840</td>
<td>657.8840</td>
<td>657.8840</td>
<td>657.8840</td>
<td>657.8840</td>
<td>657.8840</td>
<td>657.8840</td>
<td>657.8840</td>
<td>657.8840</td>
</tr>
</tbody>
</table>

**Expected Total Cost**

**Maximum and Minimum**

- Maximum: 2.3272
- Minimum: 2.2272
$A = 1.00$

$\times = 1.58$

**EXPECTED TOTAL COST**

<table>
<thead>
<tr>
<th>C</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>155.44</td>
<td>158.50</td>
<td>160.51</td>
<td>162.52</td>
<td>164.53</td>
<td>166.54</td>
<td>168.55</td>
<td>170.56</td>
<td>172.57</td>
<td>174.58</td>
</tr>
<tr>
<td>0.02</td>
<td>168.25</td>
<td>170.26</td>
<td>172.27</td>
<td>174.28</td>
<td>176.29</td>
<td>178.30</td>
<td>180.31</td>
<td>182.32</td>
<td>184.33</td>
<td>186.34</td>
</tr>
<tr>
<td>0.03</td>
<td>187.26</td>
<td>187.27</td>
<td>187.28</td>
<td>187.29</td>
<td>187.30</td>
<td>187.31</td>
<td>187.32</td>
<td>187.33</td>
<td>187.34</td>
<td>187.35</td>
</tr>
<tr>
<td>0.04</td>
<td>199.57</td>
<td>199.58</td>
<td>199.59</td>
<td>199.60</td>
<td>199.61</td>
<td>199.62</td>
<td>199.63</td>
<td>199.64</td>
<td>199.65</td>
<td>199.66</td>
</tr>
<tr>
<td>0.05</td>
<td>216.95</td>
<td>216.96</td>
<td>216.97</td>
<td>216.98</td>
<td>216.99</td>
<td>217.00</td>
<td>217.01</td>
<td>217.02</td>
<td>217.03</td>
<td>217.04</td>
</tr>
<tr>
<td>0.06</td>
<td>233.65</td>
<td>233.66</td>
<td>233.67</td>
<td>233.68</td>
<td>233.69</td>
<td>233.70</td>
<td>233.71</td>
<td>233.72</td>
<td>233.73</td>
<td>233.74</td>
</tr>
<tr>
<td>0.07</td>
<td>250.76</td>
<td>250.77</td>
<td>250.78</td>
<td>250.79</td>
<td>250.80</td>
<td>250.81</td>
<td>250.82</td>
<td>250.83</td>
<td>250.84</td>
<td>250.85</td>
</tr>
<tr>
<td>0.08</td>
<td>268.46</td>
<td>268.47</td>
<td>268.48</td>
<td>268.49</td>
<td>268.50</td>
<td>268.51</td>
<td>268.52</td>
<td>268.53</td>
<td>268.54</td>
<td>268.55</td>
</tr>
<tr>
<td>0.09</td>
<td>282.93</td>
<td>282.94</td>
<td>282.95</td>
<td>282.96</td>
<td>282.97</td>
<td>282.98</td>
<td>282.99</td>
<td>283.00</td>
<td>283.01</td>
<td>283.02</td>
</tr>
<tr>
<td>0.10</td>
<td>297.21</td>
<td>297.22</td>
<td>297.23</td>
<td>297.24</td>
<td>297.25</td>
<td>297.26</td>
<td>297.27</td>
<td>297.28</td>
<td>297.29</td>
<td>297.30</td>
</tr>
</tbody>
</table>

**MIN EXP COST**

4.1029

**MINIMIZING PPM**


**MINIMIZING MM**

1.7807 0.3091 0.6289 0.4880 0.4200 0.3977 0.3186 0.2413 0.1701 0.1286

187
<table>
<thead>
<tr>
<th>C</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

EXPECTED TOTAL COST

<table>
<thead>
<tr>
<th>C</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MIN EXP COST 5.6819 0.0608 6.5166 7.0213 7.5228 6.2398 8.5566 9.0678 10.0750 0.0700

MINIMIZING WTC 0.6109 0.0150 0.6000 0.5900 0.5700 0.5600 0.5500 0.5400 0.5300 0.5200

MINIMIZING MS 1.6939 0.0197 0.6556 0.4237 0.5909 0.2976 0.2940 0.2358 0.2179 0.2041
<table>
<thead>
<tr>
<th>C</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>630.9459</td>
<td>630.9459</td>
<td>630.9459</td>
<td>630.9459</td>
<td>630.9459</td>
<td>630.9459</td>
<td>630.9459</td>
<td>630.9459</td>
<td>630.9459</td>
<td>630.9459</td>
</tr>
<tr>
<td>0.02</td>
<td>630.9459</td>
<td>355.1500</td>
<td>177.5750</td>
<td>88.7875</td>
<td>44.3938</td>
<td>22.1969</td>
<td>11.0985</td>
<td>5.5493</td>
<td>2.7747</td>
<td>1.3874</td>
</tr>
<tr>
<td>0.03</td>
<td>630.9459</td>
<td>55.1500</td>
<td>28.5875</td>
<td>14.2938</td>
<td>7.1969</td>
<td>3.5985</td>
<td>1.7747</td>
<td>0.8874</td>
<td>0.4438</td>
<td>0.2219</td>
</tr>
<tr>
<td>0.04</td>
<td>630.9459</td>
<td>5.1500</td>
<td>2.8587</td>
<td>1.4294</td>
<td>0.7197</td>
<td>0.3598</td>
<td>0.1775</td>
<td>0.0887</td>
<td>0.0444</td>
<td>0.0222</td>
</tr>
<tr>
<td>0.05</td>
<td>630.9459</td>
<td>0.1500</td>
<td>0.2859</td>
<td>0.1429</td>
<td>0.0719</td>
<td>0.0359</td>
<td>0.0177</td>
<td>0.0089</td>
<td>0.0044</td>
<td>0.0022</td>
</tr>
<tr>
<td>0.06</td>
<td>630.9459</td>
<td>0.0150</td>
<td>0.0286</td>
<td>0.0143</td>
<td>0.0072</td>
<td>0.0036</td>
<td>0.0018</td>
<td>0.0010</td>
<td>0.0005</td>
<td>0.0002</td>
</tr>
<tr>
<td>0.10</td>
<td>63.0946</td>
<td>63.0946</td>
<td>63.0946</td>
<td>63.0946</td>
<td>63.0946</td>
<td>63.0946</td>
<td>63.0946</td>
<td>63.0946</td>
<td>63.0946</td>
<td>63.0946</td>
</tr>
<tr>
<td>0.20</td>
<td>63.0946</td>
<td>31.5476</td>
<td>15.7738</td>
<td>7.8889</td>
<td>3.9445</td>
<td>1.9722</td>
<td>0.9861</td>
<td>0.4931</td>
<td>0.2465</td>
<td>0.1231</td>
</tr>
<tr>
<td>0.30</td>
<td>63.0946</td>
<td>5.1476</td>
<td>2.5738</td>
<td>1.2889</td>
<td>0.6445</td>
<td>0.3222</td>
<td>0.1611</td>
<td>0.0806</td>
<td>0.0403</td>
<td>0.0201</td>
</tr>
<tr>
<td>0.40</td>
<td>63.0946</td>
<td>0.1476</td>
<td>0.0738</td>
<td>0.0389</td>
<td>0.0164</td>
<td>0.0082</td>
<td>0.0041</td>
<td>0.0020</td>
<td>0.0010</td>
<td>0.0005</td>
</tr>
<tr>
<td>0.50</td>
<td>63.0946</td>
<td>0.0147</td>
<td>0.0074</td>
<td>0.0039</td>
<td>0.0016</td>
<td>0.0008</td>
<td>0.0004</td>
<td>0.0002</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>0.60</td>
<td>63.0946</td>
<td>0.0015</td>
<td>0.0008</td>
<td>0.0004</td>
<td>0.0002</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

**EXPECTED TOTAL COST**

A = 1.00

K = 6.31
<table>
<thead>
<tr>
<th>X</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>1.024</td>
<td>1.172</td>
<td>2.115</td>
<td>2.2492</td>
<td>2.2490</td>
<td>2.2894</td>
<td>2.3048</td>
<td>2.3058</td>
<td>2.3250</td>
<td>2.3385</td>
<td>2.3453</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.02</td>
<td>1.004</td>
<td>1.0695</td>
<td>1.1275</td>
<td>1.1550</td>
<td>1.1638</td>
<td>1.1873</td>
<td>1.2004</td>
<td>1.2126</td>
<td>1.2170</td>
<td>1.2262</td>
<td>1.2343</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.03</td>
<td>1.000</td>
<td>0.9944</td>
<td>0.9860</td>
<td>0.9781</td>
<td>0.9773</td>
<td>0.9768</td>
<td>0.9760</td>
<td>0.9755</td>
<td>0.9755</td>
<td>0.9754</td>
<td>0.9753</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.04</td>
<td>1.000</td>
<td>0.9800</td>
<td>0.9600</td>
<td>0.9400</td>
<td>0.9300</td>
<td>0.9200</td>
<td>0.9100</td>
<td>0.9000</td>
<td>0.8900</td>
<td>0.8800</td>
<td>0.8700</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>1.000</td>
<td>0.9600</td>
<td>0.9400</td>
<td>0.9200</td>
<td>0.9100</td>
<td>0.9000</td>
<td>0.8900</td>
<td>0.8800</td>
<td>0.8700</td>
<td>0.8600</td>
<td>0.8500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.06</td>
<td>1.000</td>
<td>0.9400</td>
<td>0.9200</td>
<td>0.9000</td>
<td>0.8900</td>
<td>0.8800</td>
<td>0.8700</td>
<td>0.8600</td>
<td>0.8500</td>
<td>0.8400</td>
<td>0.8300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.07</td>
<td>1.000</td>
<td>0.9200</td>
<td>0.9000</td>
<td>0.8800</td>
<td>0.8700</td>
<td>0.8600</td>
<td>0.8500</td>
<td>0.8400</td>
<td>0.8300</td>
<td>0.8200</td>
<td>0.8100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>1.000</td>
<td>0.9000</td>
<td>0.8800</td>
<td>0.8600</td>
<td>0.8500</td>
<td>0.8400</td>
<td>0.8300</td>
<td>0.8200</td>
<td>0.8100</td>
<td>0.8000</td>
<td>0.7900</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.09</td>
<td>1.000</td>
<td>0.8800</td>
<td>0.8600</td>
<td>0.8400</td>
<td>0.8300</td>
<td>0.8200</td>
<td>0.8100</td>
<td>0.8000</td>
<td>0.7900</td>
<td>0.7800</td>
<td>0.7700</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td>1.000</td>
<td>0.8600</td>
<td>0.8400</td>
<td>0.8200</td>
<td>0.8100</td>
<td>0.8000</td>
<td>0.7900</td>
<td>0.7800</td>
<td>0.7700</td>
<td>0.7600</td>
<td>0.7500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**EXPECTED TOTAL COST**

**X** = 1.000

**A** = 0.40
<table>
<thead>
<tr>
<th>C</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.02</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.04</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.06</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.08</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.10</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.12</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.14</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.16</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.18</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.20</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

**EXP COST**

| 2.1417 | 2.1467 | 2.1517 | 2.1567 | 2.1617 | 2.1667 | 2.1717 | 2.1767 | 2.1817 | 2.1867 |

**Simplifying Pho**

| 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |

**Simplifying W**

<p>| 0.1194 | 0.2387 | 0.3580 | 0.4773 | 0.5966 | 0.7159 | 0.8352 | 0.9545 | 1.0738 | 1.1931 |</p>
<table>
<thead>
<tr>
<th>C</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>1.037</td>
<td>1.006</td>
<td>1.052</td>
<td>1.017</td>
<td>1.063</td>
<td>1.022</td>
<td>1.069</td>
<td>1.028</td>
<td>1.075</td>
<td>1.032</td>
</tr>
<tr>
<td>Y</td>
<td>0.244</td>
<td>0.715</td>
<td>0.324</td>
<td>0.831</td>
<td>0.412</td>
<td>0.924</td>
<td>0.523</td>
<td>0.635</td>
<td>0.746</td>
<td>0.857</td>
</tr>
</tbody>
</table>

**EXPECTED TOTAL COST**

- **MIN EXP COST**
  - 2.7543, 0.0025, 2.1519
  - 3.5149, 0.6207, 1.7933
- **MINIMIZING X**
  - 0.6104, 0.6600, 0.6603
  - 0.5800, 0.5800, 0.5800
- **MINIMIZING Y**
  - 1.0843, 0.8333, 0.5650
  - 0.4310, 0.3468, 0.7262

---

**MIN EXP COST**

- 2.7543, 0.0025, 2.1519
  - 3.5149, 0.6207, 1.7933

**MINIMIZING X**

- 0.6104, 0.6600, 0.6603
  - 0.5800, 0.5800, 0.5800

**MINIMIZING Y**

- 1.0843, 0.8333, 0.5650
  - 0.4310, 0.3468, 0.7262

---

**EXPECTED TOTAL COST**

- **C 1  2  3  4  5  6  7  8  9  10**
- **X   Y**
- 1.037 0.244
  - 1.006 0.715
  - 1.052 0.324
  - 1.017 0.831
  - 1.063 0.412
  - 1.022 0.924
  - 1.069 0.523
  - 1.028 0.635
  - 1.075 0.746
  - 1.032 0.857

---

**MIN EXP COST**

- 2.7543, 0.0025, 2.1519
  - 3.5149, 0.6207, 1.7933

**MINIMIZING X**

- 0.6104, 0.6600, 0.6603
  - 0.5800, 0.5800, 0.5800

**MINIMIZING Y**

- 1.0843, 0.8333, 0.5650
  - 0.4310, 0.3468, 0.7262

---

**EXPECTED TOTAL COST**

- **C 1  2  3  4  5  6  7  8  9  10**
- **X   Y**
- 1.037 0.244
  - 1.006 0.715
  - 1.052 0.324
  - 1.017 0.831
  - 1.063 0.412
  - 1.022 0.924
  - 1.069 0.523
  - 1.028 0.635
  - 1.075 0.746
  - 1.032 0.857
<table>
<thead>
<tr>
<th>C</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>G</td>
<td>100.0010</td>
<td>100.0029</td>
<td>100.0054</td>
<td>100.0086</td>
<td>100.0123</td>
<td>100.0146</td>
<td>100.0179</td>
<td>100.0258</td>
<td>100.0311</td>
<td>100.0378</td>
</tr>
<tr>
<td>6.67</td>
<td>0.05030</td>
<td>0.05086</td>
<td>0.05141</td>
<td>0.05196</td>
<td>0.05252</td>
<td>0.05308</td>
<td>0.05364</td>
<td>0.05420</td>
<td>0.05477</td>
<td>0.05533</td>
</tr>
<tr>
<td>16.07</td>
<td>0.07592</td>
<td>0.07649</td>
<td>0.07706</td>
<td>0.07762</td>
<td>0.07818</td>
<td>0.07874</td>
<td>0.07930</td>
<td>0.07986</td>
<td>0.08042</td>
<td>0.08098</td>
</tr>
<tr>
<td>1.05</td>
<td>5.088</td>
<td>5.171</td>
<td>5.254</td>
<td>5.337</td>
<td>5.420</td>
<td>5.503</td>
<td>5.586</td>
<td>5.669</td>
<td>5.752</td>
<td>5.835</td>
</tr>
<tr>
<td>0.46</td>
<td>2.939</td>
<td>2.920</td>
<td>2.879</td>
<td>2.869</td>
<td>2.859</td>
<td>2.849</td>
<td>2.839</td>
<td>2.829</td>
<td>2.819</td>
<td>2.809</td>
</tr>
<tr>
<td>0.33</td>
<td>2.547</td>
<td>2.555</td>
<td>2.563</td>
<td>2.571</td>
<td>2.579</td>
<td>2.587</td>
<td>2.595</td>
<td>2.603</td>
<td>2.611</td>
<td>2.619</td>
</tr>
<tr>
<td>0.27</td>
<td>2.342</td>
<td>2.342</td>
<td>2.342</td>
<td>2.342</td>
<td>2.342</td>
<td>2.342</td>
<td>2.342</td>
<td>2.342</td>
<td>2.342</td>
<td>2.342</td>
</tr>
<tr>
<td>0.23</td>
<td>2.254</td>
<td>2.254</td>
<td>2.254</td>
<td>2.254</td>
<td>2.254</td>
<td>2.254</td>
<td>2.254</td>
<td>2.254</td>
<td>2.254</td>
<td>2.254</td>
</tr>
<tr>
<td>0.21</td>
<td>2.235</td>
<td>2.235</td>
<td>2.235</td>
<td>2.235</td>
<td>2.235</td>
<td>2.235</td>
<td>2.235</td>
<td>2.235</td>
<td>2.235</td>
<td>2.235</td>
</tr>
<tr>
<td>0.19</td>
<td>2.224</td>
<td>2.224</td>
<td>2.224</td>
<td>2.224</td>
<td>2.224</td>
<td>2.224</td>
<td>2.224</td>
<td>2.224</td>
<td>2.224</td>
<td>2.224</td>
</tr>
<tr>
<td>0.17</td>
<td>2.214</td>
<td>2.214</td>
<td>2.214</td>
<td>2.214</td>
<td>2.214</td>
<td>2.214</td>
<td>2.214</td>
<td>2.214</td>
<td>2.214</td>
<td>2.214</td>
</tr>
<tr>
<td>0.16</td>
<td>2.205</td>
<td>2.205</td>
<td>2.205</td>
<td>2.205</td>
<td>2.205</td>
<td>2.205</td>
<td>2.205</td>
<td>2.205</td>
<td>2.205</td>
<td>2.205</td>
</tr>
<tr>
<td>0.15</td>
<td>2.196</td>
<td>2.196</td>
<td>2.196</td>
<td>2.196</td>
<td>2.196</td>
<td>2.196</td>
<td>2.196</td>
<td>2.196</td>
<td>2.196</td>
<td>2.196</td>
</tr>
<tr>
<td>0.13</td>
<td>2.178</td>
<td>2.178</td>
<td>2.178</td>
<td>2.178</td>
<td>2.178</td>
<td>2.178</td>
<td>2.178</td>
<td>2.178</td>
<td>2.178</td>
<td>2.178</td>
</tr>
<tr>
<td>0.12</td>
<td>2.169</td>
<td>2.169</td>
<td>2.169</td>
<td>2.169</td>
<td>2.169</td>
<td>2.169</td>
<td>2.169</td>
<td>2.169</td>
<td>2.169</td>
<td>2.169</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td>100.0000</td>
<td>100.0000</td>
<td>100.0000</td>
<td>100.0000</td>
<td>100.0000</td>
<td>100.0000</td>
<td>100.0000</td>
<td>100.0000</td>
<td>100.0000</td>
<td></td>
</tr>
<tr>
<td>0.02</td>
<td>50.0000</td>
<td>33.3333</td>
<td>33.3333</td>
<td>33.3333</td>
<td>33.3333</td>
<td>33.3333</td>
<td>33.3333</td>
<td>33.3333</td>
<td>33.3333</td>
<td></td>
</tr>
<tr>
<td>0.03</td>
<td>33.3333</td>
<td>33.3333</td>
<td>33.3333</td>
<td>33.3333</td>
<td>33.3333</td>
<td>33.3333</td>
<td>33.3333</td>
<td>33.3333</td>
<td>33.3333</td>
<td></td>
</tr>
<tr>
<td>0.04</td>
<td>25.0000</td>
<td>25.0000</td>
<td>25.0000</td>
<td>25.0000</td>
<td>25.0000</td>
<td>25.0000</td>
<td>25.0000</td>
<td>25.0000</td>
<td>25.0000</td>
<td></td>
</tr>
<tr>
<td>0.05</td>
<td>20.0000</td>
<td>20.0000</td>
<td>20.0000</td>
<td>20.0000</td>
<td>20.0000</td>
<td>20.0000</td>
<td>20.0000</td>
<td>20.0000</td>
<td>20.0000</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td></td>
</tr>
</tbody>
</table>

**EXPECTED TOTAL COST**

**MIN EXP COST**

**MINIMIZING PHO**

**MINIMIZING MU**
BIBLIOGRAPHY


(1965), pp. 181 ff.


Static Decision Models for Queueing Systems with Non-Linear Waiting Costs

Some models for the optimal design of queueing systems are presented. In most models studied, the decision variables are the number of servers \( (c) \) and the mean rate \( (\mu) \) at which each serves. The objective function is the steady-state total expected cost rate of operating the system, which is assumed to be the sum of a cost of operating the service mechanism and a cost due to customers waiting in the system. It is shown that a single-server system is optimal for a wide class of arrival processes and service-time distributions, a wide variety of service and waiting cost functions, and a wide variety of system structures and operating policies. The optimality of the single-server system is first demonstrated for single-station models with general arrival process and degenerate, exponential, or Erlang service-time distribution, where the service-cost rate is proportional to both \( c \) and \( \mu \) and the waiting-cost rate is proportional to the number of customers in the system. Several generalizations of this model are presented: (1) single-station models with service-cost rate proportional to \( c \) and concave, quasi-homogeneous, or subadditive in \( \mu \); (2) single-station models with customer waiting costs which are monotone increasing in the time spent by a customer in the system; (3) multi-station models in which the service times at each station are exponentially distributed and the stations are arranged in a series or as the nodes of a partially ordered graph; (4) single-station models with various generalizations of operating policy, cost structure, and the set of decision variables.
Security Classification

14. KEY WORDS

queueing theory
operations research
design of queueing systems
optimization of queueing systems
stochastic orderings
queueing networks
non-linear waiting costs
stochastic service systems
decision models for queueing systems

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization (corporate author) issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capital letters in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers (either by the originator or by the sponsor), also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

   (1) "Qualified requestors may obtain copies of this report from DDC.

   (2) "Foreign announcement and dissemination of this report by DDC is not authorized.

   (3) "U.S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through

   (4) "U.S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through

   (5) "All distribution of this report is controlled. Qualified DDC users shall request through

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring (paying for) the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report. Even though it may also appear elsewhere in the body of the technical report, if additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as (TS), (S), (C), or (U).

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

UNCLASSIFIED
Security Classification