MATNEY, Joseph Sylvester, 1943-
SOME DESCRIPTIVE MODELS OF SINGLE FACILITY
SEQUENCING WITH INTERACTIVE SERVICE TIMES.

The Ohio State University, Ph.D., 1971
Engineering, industrial

University Microfilms, A XEROX Company, Ann Arbor, Michigan

Copyright by
Joseph Sylvester Matney
1971
SOME DESCRIPTIVE MODELS OF SINGLE FACILITY
SEQUENCING WITH INTERACTIVE SERVICE TIMES

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University

By

Joseph Sylvester Matney, B.S., M.S.

* * * * *

The Ohio State University
1971

Approved by

[Signature]

Advisor
Department of
Industrial Engineering
PLEASE NOTE:

Some Pages have indistinct print. Filmed as received.

UNIVERSITY MICROFILMS
ACKNOWLEDGMENTS

I wish to express my appreciation to The Graduate School of The Ohio State University for providing financial assistance during the writing of this dissertation through a Dissertation Year Fellowship. Gratitude is also expressed for the financial support and opportunity to serve as a teaching associate provided by the late Dr. David F. Baker. His warm, unwavering concern was a source of strength and will be long remembered.

I am grateful to the members of my committee, Dr. A. B. Bishop, Dr. W. C. Giffin, Dr. W. T. Morris, and Dr. J. B. Neuhardt, for reading and criticizing my dissertation. Their willingness to read rough drafts between quarters is certainly appreciated. A special expression of gratitude is due my adviser throughout the graduate program, Dr. W. C. Giffin. Without his support, especially during the periods of despair that so often accompany the pursuit of the doctorate, this dissertation could not have been written.

The counsel given by my parents was instrumental in my decision to undertake graduate education. Their continued interest has proven invaluable. Most important the patience and encouragement of my wife, Cris, is recognized. Her loving concern has provided inspiration and hope.
May 12, 1943 .................. Born - Grundy, Virginia

1966 .......................... B.S.I.E. (with Honors), Virginia Polytechnic Institute, Blacksburg, Virginia

1966 .......................... Industrial Engineer, Industrial Nucleonics Corporation, Columbus, Ohio (Summer)

1966-67 ...................... Tau Beta Pi Fellow, The Ohio State University

1967-68 ...................... Research Engineer, North American Rockwell Corporation, Columbus, Ohio (Summers)

1968 .......................... M.S.I.E., The Ohio State University

1967-68 ...................... Teaching Assistant, The Ohio State University

1968-70 ...................... Teaching Associate, The Ohio State University

1970-71 ...................... Dissertation Year Fellow, The Ohio State University

FIELDS OF STUDY

Major Field: Industrial Engineering

# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>Vita</td>
<td>iii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>ix</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>I. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Objectives of the Study</td>
<td></td>
</tr>
<tr>
<td>Description of the Problem</td>
<td></td>
</tr>
<tr>
<td>Description of the Methodology</td>
<td></td>
</tr>
<tr>
<td>Importance of the Topic</td>
<td></td>
</tr>
<tr>
<td>Order of Discussion</td>
<td></td>
</tr>
<tr>
<td>II. Literature Review</td>
<td>11</td>
</tr>
<tr>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td>Models with Limited Waiting Space</td>
<td></td>
</tr>
<tr>
<td>Models with Unique Service for the First Customer of a Busy Period</td>
<td></td>
</tr>
<tr>
<td>Models with Sequence Dependent Service Times</td>
<td></td>
</tr>
<tr>
<td>III. A Single Customer Class Model</td>
<td>20</td>
</tr>
<tr>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td>Problem Description</td>
<td></td>
</tr>
<tr>
<td>Model Development</td>
<td></td>
</tr>
<tr>
<td>Example One</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td></td>
</tr>
<tr>
<td>IV. A Multiple Customer Class Model</td>
<td>72</td>
</tr>
<tr>
<td>Introduction</td>
<td></td>
</tr>
<tr>
<td>Problem Description</td>
<td></td>
</tr>
<tr>
<td>Model Development</td>
<td></td>
</tr>
<tr>
<td>Example Two</td>
<td></td>
</tr>
<tr>
<td>Summary</td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
</tr>
<tr>
<td>1.</td>
<td>Arrival Probabilities for the Single Customer Class Model</td>
</tr>
<tr>
<td>2.</td>
<td>Intermediate Expressions and Arrival Probabilities for Customers that Enter the System when the Server is Idle</td>
</tr>
<tr>
<td>3.</td>
<td>Intermediate Expressions and Arrival Probabilities for Customers that Enter the System when the Server is Busy</td>
</tr>
<tr>
<td>4.</td>
<td>Steady State Probabilities for the Single Customer Class Model of Example One</td>
</tr>
<tr>
<td>5.</td>
<td>Arrival Probabilities for the Multiple Customer Class Models</td>
</tr>
<tr>
<td>6.</td>
<td>Transition Probabilities for the Non-preemptive Priority Model</td>
</tr>
<tr>
<td>7.</td>
<td>Service Time Density Functions for the ((k = 2, N = 2)) Model of Example Two</td>
</tr>
<tr>
<td>8.</td>
<td>Total Arrival Probabilities for the ((k = 2, N = 2)) Model of Example Two</td>
</tr>
<tr>
<td>9.</td>
<td>Arrival Probabilities for the ((k = 2, N = 2)) Model of Example Two</td>
</tr>
<tr>
<td>10.</td>
<td>Values of (w) for the ((k = 2, N = 2)) Non-preemptive Priority Model</td>
</tr>
<tr>
<td>11.</td>
<td>Steady State Probabilities for the ((k = 2, N = 2)) Non-preemptive Priority Model of Example Two</td>
</tr>
<tr>
<td>12.</td>
<td>Probability Mass Functions for the Number of Customers in the System and the Number of Customers in the Queue for the ((k = 2, N = 2)) Non-preemptive Priority Model of Example Two</td>
</tr>
<tr>
<td>Table</td>
<td>Measures of System Performance for the ((k = 2, N = 2)) Non-preemptive Priority Model of Example Two</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>14.</td>
<td>Values of (w) for the ((k = 2, N = 2)) Alternating Priority Model</td>
</tr>
<tr>
<td>15.</td>
<td>Values of (w) for the ((k = 2, N = 2)) Rotating Priority Model</td>
</tr>
<tr>
<td>16.</td>
<td>Steady State Probabilities for the ((k = 2, N = 2)) Alternating Priority Model of Example Three</td>
</tr>
<tr>
<td>17.</td>
<td>Probability Mass Functions for the Number of Customers in the System and the Number of Customers in the Queue for the ((k = 2, N = 2)) Alternating Priority Model of Example Three</td>
</tr>
<tr>
<td>18.</td>
<td>Measures of System Performance for the ((k = 2, N = 2)) Alternating Priority Model of Example Three</td>
</tr>
<tr>
<td>19.</td>
<td>Steady State Probabilities for the ((k = 2, N = 2)) Rotating Priority Model of Example Four</td>
</tr>
<tr>
<td>20.</td>
<td>Probability Mass Functions for the Number of Customers in the System and the Number of Customers in the Queue for the ((k = 2, N = 2)) Rotating Priority Model of Example Four</td>
</tr>
<tr>
<td>21.</td>
<td>Measures of System Performance for the ((k = 2, N = 2)) Rotating Priority Model of Example Four</td>
</tr>
<tr>
<td>22.</td>
<td>Values of (w) for the ((k = 2, N = 2)) First Come-First Served Model</td>
</tr>
<tr>
<td>23.</td>
<td>Steady State Probabilities for the ((k = 2, N = 2)) First Come-First Served Model of Example Five</td>
</tr>
<tr>
<td>24.</td>
<td>Probability Mass Functions for the Number of Customers in the System and the Number of Customers in the Queue for the ((k = 2, N = 2)) First Come-First Served Model of Example Five</td>
</tr>
<tr>
<td>Table</td>
<td>Measures of System Performance for the ((k = 2, N = 2)) First Come-First Served Model of Example Five</td>
</tr>
<tr>
<td>-------</td>
<td>---------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>25.</td>
<td>Interactive Service Time Relationships for Takeoff and Landing Operations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table</th>
<th>Measures of System Performance for the ((k = 2, N = 2)) First Come-First Served Model of Example Five</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>25.</td>
<td>Interactive Service Time Relationships for Takeoff and Landing Operations</td>
<td>239</td>
</tr>
</tbody>
</table>
# LIST OF ILLUSTRATIONS

<table>
<thead>
<tr>
<th>Figure</th>
<th>Illustration Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Imbedded Markov Chain Transition Matrix for the Single Customer Class Model</td>
<td>33</td>
</tr>
<tr>
<td>2</td>
<td>Comparison of the Augmented IMC, SMP, and RPT System Descriptions</td>
<td>41</td>
</tr>
<tr>
<td>3</td>
<td>Illustration of the Arrival Mechanism for a Service Time, $t$, and for a Random Inter­­ruption Time, $t'$</td>
<td>48</td>
</tr>
<tr>
<td>4</td>
<td>Conversion Matrix for the Single Customer Class Model</td>
<td>55</td>
</tr>
<tr>
<td>5</td>
<td>IMC Transition Matrix for the Single Customer Class Model of Example One</td>
<td>65</td>
</tr>
<tr>
<td>6</td>
<td>Conversion Matrix for the Single Customer Class Model of Example One</td>
<td>67</td>
</tr>
<tr>
<td>7</td>
<td>IMC Transition Matrix for the $(k = 2, N = 2)$ Non-preemptive Priority Model</td>
<td>118</td>
</tr>
<tr>
<td>8</td>
<td>Conversion Matrix for the $(k = 2, N = 2)$ Non-preemptive Priority Model</td>
<td>152</td>
</tr>
<tr>
<td>9</td>
<td>IMC Transition Matrix for the $(k = 2, N = 2)$ Non-preemptive Priority Model of Example Two</td>
<td>171</td>
</tr>
<tr>
<td>10</td>
<td>Conversion Matrix for the $(k = 2, N = 2)$ Non-preemptive Priority Model of Example Two</td>
<td>178</td>
</tr>
<tr>
<td>11</td>
<td>IMC Transition Matrix for the $(k = 2, N = 2)$ Alternating Priority Model</td>
<td>204</td>
</tr>
<tr>
<td>12</td>
<td>Conversion Matrix for the $(k = 2, N = 2)$ Alternating Priority Model</td>
<td>205</td>
</tr>
<tr>
<td>13</td>
<td>IMC Transition Matrix for the $(k = 2, N = 2)$ Rotating Priority Model</td>
<td>210</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>14.</td>
<td>Conversion Matrix for the ((k = 2, N = 2))</td>
<td>211</td>
</tr>
<tr>
<td></td>
<td>Rotating Priority Model</td>
<td></td>
</tr>
<tr>
<td>15.</td>
<td>IMC Transition Matrix for the ((k = 2, N = 2))</td>
<td>225</td>
</tr>
<tr>
<td></td>
<td>First Come-First Served Model</td>
<td></td>
</tr>
<tr>
<td>16.</td>
<td>Conversion Matrix for the ((k = 2, N = 2))</td>
<td>226</td>
</tr>
<tr>
<td></td>
<td>First Come-First Served Model</td>
<td></td>
</tr>
<tr>
<td>17.</td>
<td>Air Terminal System</td>
<td>233</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Objectives of the Study

The primary emphasis of this study is the development of models for a special class of single facility sequencing problems. The objectives of this study are to develop a general methodology and to provide a number of models that will aid in the investigation of sequencing phenomena. The initial modeling efforts were motivated by a desire to investigate air traffic control problems involving sequencing of landing and takeoff operations for a single runway. However, the problem context is such that the models developed in the study are appropriate for the analysis of a wide range of sequencing problems. In addition to providing complete models for a number of specific single facility sequencing rules, a general methodology is presented which can be utilized to study additional sequencing rules.

Description of the Problem

The system under consideration can be described as a M/G/1 queue operating under a specific set of assumptions. The assumptions are:
1. All customers that enter the system are classified as belonging to one of a finite number of homogeneous customer classes.

2. Customers from each customer class enter the system at random points in time with a known mean rate.

3. All customers are served by a single service facility.

4. The service time is influenced by two factors: the class of customer receiving service and the class of customer, if any, that completed service immediately before the current customer's service was initiated. These interactive service times are random variables with known general probability density functions.

5. The number of customers awaiting service at any time is required to be less than a known constant number. Customers are not allowed to enter the system when the queue is filled to capacity.

6. The order of service is determined by a fixed sequencing rule.

Models based on these assumptions are developed for a variety of sequencing rules. The effect of the sequencing rule is reflected in several measures of system performance.
As all of the systems considered attain equilibrium conditions, steady state values are examined.

Methods of finding the steady state probability mass functions of the number of each class customer in the queue (waiting for service only) and the number of each class customer in the system (waiting for service and receiving service) are given. Using this information the measures of system performance are found. These measures are effective mean arrival rate, expected number of customers in the system, expected number of customers in the queue, expected waiting time in the system, expected waiting time in the queue, and expected service time. These measures are determined for each class of customer and for the overall customer population.

**Description of the Methodology**

In order to determine the measures of system performance it is necessary to know the probability that the system is in a specified state at any random point in time after equilibrium conditions are attained. The state of the system is defined as the class of customer receiving service and the number of each class of customer waiting in the queue. These probabilities are referred to as random point in time (RPT) steady state probabilities. Because the system is extremely complex these RPT steady state probabilities cannot be determined directly. It is necessary to first consider the state of the system at only selected instants of time called
service completion epochs. A service completion epoch is defined as the instant of time that a customer completes service. Using an imbedded Markov chain approach, the probability that the system is in a specified state at any randomly selected service completion epoch after equilibrium conditions are attained is determined. These probabilities are referred to as imbedded Markov chain (IMC) steady state probabilities. Note that these IMC steady state probabilities are associated only with selected points in time. They do not indicate the state of the system at a random point in time. In other words, the IMC steady state probabilities are biased because the system is examined only at service completion epochs.

It is possible to convert the biased IMC steady state probabilities to the desired RPT steady state probabilities by performing a number of intermediate conversion steps. These conversion steps are discussed in detail in subsequent chapters. Thus the RPT steady state probabilities are determined by the circuitous route of first finding the IMC steady state probabilities and then performing a series of modifications. The exact nature of these modifications is presented in detail in Chapter III for a single customer class model and in Chapter IV for the multiple customer class models.

**Importance of the Topic**

As previously stated the study was originally motivated by a desire to investigate sequencing of landing and takeoff
operations for a single runway. The essential feature of this problem is the interactive nature of the service times. For example, the time to perform a landing operation is dependent not only on the characteristics associated with the landing aircraft, but also the characteristics of the aircraft performing the immediately preceding operation. Such factors as wake turbulence, longitudinal separation requirements, and runway occupancy times are factors in the interactive nature of service times. The value of sequencing landing and takeoff operations is that some sequencing rules provide better system performance than other sequencing rules. For example, preliminary investigation indicates that some sequencing rules provide for a larger number of operations per unit time with less mean waiting time to receive service than do others. Results of this nature have been indicated by several numerical examples.

A great many other applications require models that consider interactive service times. An industrial problem of this type is the paint mixing problem. For this problem paints in a variety of colors are mixed on a single machine. The time to mix a particular color paint depends on the amount of time required to clean the equipment between successive mixing operations. Clearly the time to mix a dark color paint followed by a light color paint requires a larger amount of time than the reverse order of processing.

Another use of the model is the analysis of a modified
form of the classical traveling-salesman problem. In this modified form the class of job currently in service indicates the city just visited, and the service time indicates the time to travel between cities as well as to conclude the visit. Note that new visits to each city are being requested at random points in time. This problem might be referred to as a dynamic, stochastic version of the traveling-salesman problem.

Regardless of the context of application the models developed in this study treat a class of problem that has received little attention in the literature. The modeling complexities introduced by explicitly considering the interactive nature of service times has deterred previous investigations. As Conway, Maxwell, and Miller (1967, p. 191) state in their discussion of models with assumptions of this type, "... one could allow the setup time to depend on the class which preceded on the machine as well as the class intended to go on the machine, in a manner analogous to the traveling-salesman problem ..., but the algebra is tedious and the work just has not been done." The literature review presented in Chapter II substantiates this statement.

Order of Discussion

The results of the investigation are presented according to the following plan. Chapter II includes a review of the literature associated with modeling efforts of the type attempted in the study. Chapter III consists of a complete
development of a sequencing model for the special case of one
customer class and a first come-first served sequencing rule.
The interactive nature of service times in this initial model
is restricted to having a unique probability density function
for the service time of a customer who enters the system to
find the server idle and a different probability density
function for the service time of a customer who enters the
system to find the server busy. The motivation for present­
ing this relatively simple case in detail is to display the
basic methodology divorced from the complexities resulting
from multiple customer class systems. Of secondary impor­
tance is the fact that all of the multiple customer class
models presented in subsequent chapters readily reduce to
this basic model when the limiting case of one customer class
is evaluated. Thus the one customer class model serves as a
suitable check for the enriched models. A small numerical
example is given at the end of the chapter to illustrate the
approach.

The basic techniques of Chapter III are extended to
multiple customer class populations in Chapter IV. The
initial sections of the chapter are concerned with mathemati­
cal expressions for transition probabilities and solution
procedures for the general multiple customer class model.
Many of the derivations utilized in these sections are
contained in Appendix A. The results contained in Appendix A
are of sufficient importance and originality to be contained
within the body of the chapter; however, the derivations interrupt the natural flow of model formulation. The logic of the derivations and end results are discussed in the body of the chapters, but the mathematical manipulations are given in Appendix A. The concluding sections of Chapter IV are devoted to an application of the multiple customer class methodology for a non-preemptive priority or head-of-the-line sequencing rule. The equations of interest are developed and illustrated by a small example problem.

Chapter V continues the application of the multi-class methodology for three additional sequencing rules. A rotating priority sequencing rule is considered in which the customer in service yields control of the server to another class of customer upon conclusion of service. Thus the service facility serves one of each class of customer according to a predetermined sequence, subject to availability of customers. This type of priority rule would seem to be advantageous when the nature of the interactive service times is complementary among various classes of customers.

An alternating priority sequencing rule is also considered in which the customer in service retains control of the server for his class of customer. When customers of the class receiving service are no longer available, the service facility receives another class of customer according to a predetermined sequence, subject to availability of customers. Thus the service facility tends to serve groups of customers
of the same class. Changes from one class of service to another class of service tend to be minimized. This sequencing rule would seem to be advantageous when the nature of the interactive service times is antagonistic among various classes of customers. Such a situation exists if a setup time is required in addition to the regular processing time when the class of customer being served changes.

The last model of Chapter V considers a multiple customer class first come-first served sequencing rule. This model is a generalization of the single customer class model developed in Chapter III. This sequencing rule would seem to be desirable when the nature of the interactive service times is neither clearly antagonistic or complementary.

In addition to developing the methodology for these particular sequencing rules small illustrative numerical examples are given to clarify the calculations. The contents of this chapter have instructional value in that the general technical approach demonstrated is appropriate for a number of other sequencing rules. Hopefully a resourceful analyst could readily develop complete models for a variety of sequencing rules with only slight modification of the given procedures.

The usefulness of the previously developed models for an analysis of sequencing of landing and takeoff operations at a single runway is presented in Chapter VI. The air terminal system of interest is defined, and data needs are
discussed. The acceptability of the assumptions on which the models are based is considered and weaknesses and strengths are emphasized. It is concluded that, given appropriate data, the use of the models would provide valuable insights into the benefits of sequencing. Chapter VII summarizes the important results of the study and indicates some fruitful areas for further research.
CHAPTER II
LITERATURE REVIEW

Introduction

The purpose of this chapter is to review previous modeling efforts that used assumptions similar to those considered by the models developed in this investigation. Additional studies that consider specific models for air traffic control application, i.e., sequencing of landing and takeoff aircraft at a single runway, are presented in Chapter VI.

It is useful to classify the models cited in this review and the models developed in this study by comparing them to the standard M/G/1 queue. The standard M/G/1 queue considers a system in which

1. a single class of customers enters the system according to a Poisson Law,
2. the service times for customers are independently distributed according to a general probability distribution,
3. a single facility provides service, and
4. the waiting space for customers in the queue is unlimited.

The order of discussion in this chapter utilizes the above method of classification with the following modifications.
First, studies of the M/G/1 queue with limited waiting space in the queue are reviewed; second, studies of the M/G/1 queue where unique service is provided to the first customer of a busy period are presented; and, third, studies of the M/G/1 queue with more than one class of customers and with service times that depend not only on the class of customer receiving service, but to some degree, on the class of customer that received the preceding service, are discussed.

Models with Limited Waiting Space

The first group of models to be discussed alters the fourth assumption of the standard M/G/1 queue by requiring the waiting space to be limited. Limited waiting space implies that the maximum number of customers allowed to wait in the queue cannot be greater than a fixed finite integer; N. When N customers wait in the queue all arriving customers are turned away. Thus when there are \((N + 1)\) customers in the system the mean arrival rate is zero. A number of analysts have explored this phenomenon.

Keilson (1966) has investigated this variation of the M/G/1 queue and obtained the steady state distribution of the number of customers in the queue at random points in time and the steady state distribution of the number of customers in the queue at service initiation epochs. A service initiation epoch is defined as the instant of time at which a customer starts to receive service. It is shown that these two distributions coincide if the waiting space is unlimited,
N = \infty but do not coincide if the waiting space is limited. This implies that if the waiting space is limited, the imbedded Markov chain approach used to develop the steady state probabilities associated with service initiation epochs provides biased estimates of the actual state of the system after equilibrium conditions are attained. The fact that the two distributions coincide when the waiting space is unlimited has also been shown by Fox and Miller (1967) using different methods.

Finch (1968) has addressed this problem and determined the steady state distribution of the number of customers in the queue at service completion epochs. Although he indicates that this steady state distribution can be obtained for any general service time distribution, only the special case of the exponential service time distribution is presented. The steady state waiting time distribution for this special case is also presented under the first come-first served assumption. Additional studies involving limited waiting space queues have been reported by Cinlar and Disney (1967) and Rao and Jaiswal (1969).

Models with Unique Service for the First Customer of a Busy Period

The second group of models to be discussed alters the second assumption of the standard M/G/1 queue. The alteration requires that a unique general distribution describe the service times for customers that enter the system to find
the server idle and a different general distribution describe the service times of customers that enter the system to find the server busy. In other words, the first customer of a busy period receives unique service. Several models of this type are discussed.

Welch (1964) has treated this problem and obtained the generating function of the steady state distribution of the number of customers in the system. The Laplace transform of the steady state distribution of waiting time in the system under the first come-first served assumption is also obtained. It is indicated that this modeling approach is appropriate for an extended model. In the extended model the service time distribution for the first customer of a busy period depends on the length of the idle period being terminated.

Odoni (1970) has addressed this problem using methods different from those used by Welch. The generating function of the steady state distribution of the number of customers in the system and the expected waiting time in the system under the first come-first served assumption are obtained. This model was specifically developed for air traffic control applications and will be cited later.

Harris (1967a), (1967b) has reported on two variations of this problem. In the (1967a) study the service times are gamma distributed for customers that receive service when there are no customers in the queue, and service times are exponentially distributed for all other customers. The
steady state distribution of the number of customers in the system is obtained for this model. In the (1967b) study the parameter of the service time distribution is itself a random variable. The service times are exponentially distributed with the mean following a uniform distribution for customers that receive service when there are no customers in the queue. The service times are exponentially distributed with the mean following a different uniform distribution for all other customers. The generating function of the steady state distribution of the number of customers in the system is obtained for this model.

Recall from the introductory chapter that the model developed in Chapter III assumes that unique service is provided to the first customer of a busy period and the waiting space is limited. All of the previous models assume that the waiting space is unlimited, so the model of Chapter III is new in this respect.

Models with Sequence Dependent Service Times

The third and final group of models to be discussed alters the first and second assumptions of the standard M/G/1 queue. One alteration requires that at least two classes of customers enter the system according to independent Poisson Laws. The other alteration requires that the service times of customers follow general probability distributions that are determined by the class of customer receiving service and the class of customer receiving the preceding
service, i.e., the service times are sequence dependent. There are several degrees of sequence dependent service times.

The simplest type of sequence dependent service times is obtained by requiring a setup time to occur in addition to the service time when the class of customer being served changes. This setup time depends only on the fact that the class of customer being served has changed and not on the specific class of customer that received the preceding service. Setup times of this type are referred to as independent setup times. The setup times are expressed in terms of general probability distributions. If there are k classes of customers, there are k service time distributions and k setup time distributions. Thus a total of $2k$ distributions would describe this phenomenon. All the models with sequence dependent service times contained in the literature make the independent setup time assumption, and the results are largely restricted to two classes of customers. Another type of sequence dependent service times is discussed after several models that utilize the independent setup time assumption are presented.

Gaver (1963) has studied a two customer class system having independent setup times for two queue disciplines: first come-first served and preemptive-resume priority. The preemptive-resume priority rule requires that when a customer whose service has been interrupted by a higher priority
customer returns to the service facility, his service time begins again from the point of interruption. The Laplace transform of the steady state distribution of waiting time in the system and the generating function for the steady state distribution of the number of customers in the system are obtained for both queue disciplines. The effect of the queue discipline on system performance is determined by comparing the probabilities of finding the server busy after equilibrium conditions are attained for several numerical examples. It is found that the assignment of preemptive-resume priorities often results in lower values of this probability as compared to a first come-first served order of service. The author is unable to draw any general conclusions regarding the conditions under which one discipline is preferable to the other.

Eisenberg (1971) has treated a two customer class system having independent setup times for two queue disciplines: alternating priority and head-of-the-line or non-preemptive priority. The Laplace transform of the steady state distribution of waiting time in the system and the generating function for the steady state distribution of the number of customers in the system are obtained for both queue disciplines. Conway, Maxwell, and Miller (1967, Chapter 9) present a general review of models having independent setup times and two customer classes. Another study of this type has been reported by Skinner (1967).
The preceding studies limited the sequence dependence of service times to the addition of an independent setup time to the service time when the class of customer being served is changed. The most general type of sequence dependent service times is accomplished by requiring the service time of a customer to explicitly depend on the specific class of customer that received the preceding service. An additional provision is made for the case where a customer enters the system during an idle period. Thus if there are \( k \) classes of customers, a total of \((k + 1)k\) general probability distributions would describe this service time phenomenon. Service times of this type are referred to as interactive service times. Note that by having completely interactive service times it is possible for the changing of the class of customer receiving service to either increase or decrease the total processing time requirement for a particular customer. Under the previous assumption it was possible only to increase the total processing time requirement. Only one study was found that utilized the interactive service time assumption.

Reinitz (1963) has considered a system that differs from the standard \( M/M/1 \) queue in three ways: first, two or more classes of customers enter the system according to independent Poisson Laws; second, service times are interactive as previously defined; and third, the waiting space for customers in the queue is limited. Note that the service times are required to be Markovian and the system under consideration
is a modified $M/M/1$ queue. The objective of the model is to determine which class of customer to serve for each possible combination of available customers in order to minimize waiting time and service costs. The costs of waiting are expressed in terms of a linear cost rate for a class $i$ customer waiting to be served, and the costs of service are expressed in terms of a linear cost rate for a class $i$ customer that is served immediately after a class $j$ customer. The optimal decisions are determined by utilizing the special methods of dynamic programming and Markov chains developed by Howard (1960). The author indicates that modified $M/M/1$ systems of this sort may be integrated into a network of queues as both the input and output of the system are Markovian.

Recall that the models developed in Chapter IV and Chapter V assume that service times are interactive and are described by general probability distributions. In addition these models assume that the waiting space is limited. The model developed by Reinitz requires Markovian service times, and the other cited models require simple independent setup times and unlimited waiting space. Therefore the models of Chapter IV and Chapter V are new.
CHAPTER III
A SINGLE CUSTOMER CLASS MODEL

Introduction

The basic features of the models developed in this study are illustrated by the development of a system description for a single customer class model. This chapter provides a detailed account of the model building steps utilized in the multiple customer class models without the complexities of extensive notation and set theory definitions used in the multiple customer class models. The arguments presented in this chapter form a foundation on which the extended multi-class arguments are based. In addition to the instructional benefits of this segment of the study, a workable model is evolved which is appropriate for a number of problem contexts involving a single service facility which gives unique service to the first customer of a busy period. Service facilities that possess this property include a vehicle that requires engine warm-up time and a machine that requires setup time.

Throughout the chapter reference is made to an imbedded Markov chain (IMC), a semi-Markov process (SMP), and a random point in time (RPT) process. It is emphasized that these three processes are simply three ways of viewing the same
underlying system. These processes are used to conveniently model a complex situation in such a way that the performance of the underlying system can be measured after equilibrium conditions are attained. The specific nature of these processes and the modeling details form the bulk of this chapter.

The order of presentation consists of a description of the problem treated and possible real world problem contexts followed by a complete discussion of the quantitative elements of the model. Intuitive arguments for the derived expressions are given within the body of the chapter, but extensive mathematical derivations are placed in Appendix A. The chapter concludes with a small numerical example.

**Problem Description**

A single service facility is considered at which customers of a single homogeneous class arrive at random points in time. This implies that the number of arrivals that enter the system during any time period is a Poisson distributed random variable with a known mean rate, r. The total waiting room is limited to a maximum of N customers in the queue at any time. This implies that the mean arrival rate to the system is zero when the system is filled to capacity. A system that is filled to capacity is referred to as a truncated system. Note that a truncated system has N customers in the queue and N + 1 customers in the system.

A customer's service time depends on the state of the
system when his service is initiated. The service time of a customer who enters the system to find the server idle differs from the service time of a customer who enters the system to find the server busy. This feature is expressed by having the service time of a customer who arrives to find the server idle follow the continuous probability density function \( f_0(t) \), and the service time of a customer who arrives to find the server busy follow the continuous probability density function \( f_1(t) \). The order of service is based on a first come-first served rule, and customers enter the service facility without delay when the previous customer completes service. In terms of queueing theory, the problem is an \( M/G/1 \) queue with finite waiting space, first come-first served queue discipline, and the first customer of any busy period receives unusual service.

This problem statement describes many intermittent processes that require start-up times. Hot ultra-sonic cleaning is an example of such a process, as the cleaning fluid must attain a specified temperature before use. This process is intermittent because the fluid is self-contaminating at operating temperatures, requiring process shutdown when not in use. Other examples involving the warm-up time for vehicles, the waiting time to gain a clerk's attention in a retail store, and machine start-up times rapidly come to mind.
Related Studies

The quantitative analysis of problems with similar assumptions has been previously undertaken although the literature does not contain a formulation corresponding to the exact set of assumptions considered in this chapter. The case of the \( M/G/1 \) queue with no distinction between service times of customers arriving when the server is busy or idle, and finite waiting space has been treated by Keilson (1966) and Finch (1958). The case of an \( M/G/1 \) queue with distinction being made between service times following another service and service time following an idle period, and with infinite waiting space, has been treated by Harris (1967a) (1967b), Odoni (1970), and Welch (1964). These investigations used approaches that are fundamentally different from the approach discussed in this chapter. It should be noted that none of these treatments considers both a truncated system and a system that gives unusual service to the first customer to be served in a busy period. The model given in this chapter is, therefore, not only a basis for the more complex multiple customer class models, but is a viable model for a particular system not previously analyzed.

Model Development

Overview

The basic technique used to model the process is an imbedded Markov chain (IMC). The use of IMC's in queueing analysis has been extensive since the original work done by
Kendall (1951). The general procedure consists of examining the system only at selected instants of time called "regeneration points". The essential feature of a regeneration point is that the state of the system at any point in time following the regeneration point can be predicted with a knowledge of the state of the system only at the regeneration point. There are a number of regeneration points in the M/G/1 queueing situation, such as the instants of the conclusion of service, the instants of the beginning of service, and any point in time when the server is idle. These regeneration points are a direct result of the random nature of arrivals.

Initially the set of regeneration points considered consists of the instants at which a customer completes service and are designated service completion epochs. After some preliminary calculations the original set of regeneration points is augmented by an additional regeneration point, namely the instant of the beginning of service of a customer that enters the system to find the server idle. This regeneration point is called an idle service initiation epoch. The state of the system at service completion epochs is the number of customers awaiting service, \( x, \ x = 0,1,2,\ldots,N \). The state of the system at an idle service initiation epoch is defined as the e state. The reason for adding the additional idle service initiation epoch is to provide modeling consistency with the multiple customer class models. In
addition, the additional epoch facilitates subsequent calculations. It should be noted that by defining the system states at only selected points in time, information about the state of the process on an interpretable time scale has been sacrificed.

Once the regeneration points have been defined a knowledge of the arrival process allows the one-step transition probabilities for the IMC to be developed. The stochastic nature of the service times requires that certain joint probability expressions be integrated in order to obtain the transition probabilities. This task is facilitated by utilizing results from the theory of Laplace transforms to avoid the actual integration. The Laplace transform identities allow concise statements of the various transition probabilities to be made. The transition probabilities are placed in the IMC to form a finite and discrete state Markov chain. The Markov chain is shown to be ergodic, thus the IMC steady state probabilities are obtained by well-known methods.

Recall that the IMC steady state probabilities describe the system at selected instants of time, namely the service completion epochs with state x as previously defined. In order to make valid statements about the underlying continuous time process, specified modifications are made to the IMC steady state probabilities. The end result of the modifications is to convert the IMC steady state probabilities to probabilities that will indicate the system state at any
random point in time after equilibrium conditions are attained. These probabilities are designated random point in time (RPT) steady state probabilities. To accomplish this task several intermediate steps are introduced.

First, the state space of the original IMC is expanded to include the idle service initiation epoch with state \( e \) as previously defined. The steady state probability for this additional state is readily obtained from the original IMC steady state probabilities. The augmented IMC probabilities are then converted to the semi-Markov process (SMP) probabilities associated with the system. The SMP retains the state descriptions as defined at the epochs (either a service completion epoch or an idle service initiation epoch) but translates these state descriptions to a time scale. The SMP stationary probabilities may be interpreted as the probability of the state of the system at the last epoch preceding a random point in time after equilibrium conditions have been attained.

The final step is to convert the state descriptions from the system state at the preceding epoch to a state description for the system state at any random point in time. These probabilities are the RPT steady state probabilities. The conversion is accomplished by modifying the SMP stationary probabilities using arguments similar to those used in developing the IMC transition probabilities. These RPT probabilities are referred to as the ergodic line length probabilities.
as they give the probability of finding exactly \( z, z = 0, 1, 2, \ldots, N + 1 \), customers in the system at any point in time after equilibrium conditions have been reached.

The RPT probabilities are used to develop measures of system performance after steady state conditions are attained. The measures developed are effective mean arrival rate, expected number of customers in the system, expected number of customers in the queue, expected waiting time in the system, expected waiting time in the queue, and expected service time. The effective arrival rate is simply the overall arrival rate, \( r \), multiplied by the probability that the system is not truncated. This gives a measure of the actual throughput of the system. The number of customer expectations are found by standard methods. The waiting time expectations utilize the previously evaluated effective mean arrival rate and number of customer expectations combined in a manner indicated by Eilon (1969). The expected overall service time is a non-trivial result as the service time distribution is dependent on the system state. Other measures of performance are available, including the higher moments of the number of customers in the system and the queue, but are not routinely calculated.

In summary, the process is modeled using an imbedded Markov chain (IMC). The IMC transition probabilities are conveniently expressed in terms of the Laplace transforms of the service time density functions. The IMC is ergodic,
and stationary probabilities are found by standard methods. These probabilities are then transformed to the desired random point in time (RPT) probabilities by first considering an augmented imbedded Markov chain and, secondly, a semi-Markov process (SMP). Using the RPT equilibrium probabilities the desired measures of system performance at equilibrium are obtained.

Notation

The notation utilized for the single customer class model is given below. Although the notation initially appears to be overly tedious, all of the subscripts are necessary to insure conformation with the modeling procedures of subsequent chapters. The instructional nature of this preliminary model is greatly enhanced by the uniform notation. One point of concern is the use of the variable z as the Laplace transform variable. This is non-standard usage but should not confuse the presentation, as the variable of transformation appears explicitly only in the derivations of Appendix A.

The following notation is used:  

\( N = \) maximum number of customers allowed to wait in the queue. \( N \) is a positive integer.  
\( r = \) mean arrival rate into the system when the number of waiting customers is less than \( N \).  
\( x = \) number of customers in the queue, \( x = 0, 1, 2, \ldots, N \).
\( f_0(t) \) = service time probability density function for a customer who arrives to find the server idle.

\( f_1(t) \) = service time probability density function for a customer who arrives to find the server busy.

Both density functions are defined for positive continuous values of \( t \) and their Laplace transforms are assumed to exist.

Additionally,

\[
m_s = \int_0^\infty t f_s(t) dt \quad \text{mean of } f_s(t),
\]

\[
F_s(t) = \int_0^t f_s(v) dv \quad \text{cumulative distribution of } f_s(t),
\]

\[
\overline{F}_s(z) = \int_0^\infty e^{-zt} f_s(t) dt = \mathcal{L}(f_s(t)) = \text{Laplace transform of } f_s(t),
\]

\[
\overline{F}_s^{(n)}(z) = \frac{d^n \overline{F}_s(z)}{dz^n} = \text{nth derivative of the Laplace transform of } f_s(t), \ n = 0,1,2,\ldots
\]

and

\[
\overline{F}_s^{(n)}(r) = \frac{d^n \overline{F}_s(z)}{dz^n} \bigg|_{z=r} = \text{nth derivative of the Laplace transform of } f_s(t), \ \text{evaluated with the variable of transformation equal to the mean arrival rate, } n = 0,1,2,\ldots
\]

Imbedded Markov Chain

The essential features of the system are easily captured by a finite state imbedded Markov chain. The states of the IMC, \( x \), are defined to be the number of customers awaiting
service at those instants of time designated as service completion epochs. Thus, a system that truncates when \( N \) customers are in the queue is described by an IMC with \( N + 1 \) states, namely \( x = 0, 1, 2, \ldots, N \).

The one-step transition probabilities are developed from a consideration of the state of the system at any service completion epoch and the state of the system at the immediately following service completion epoch, \( x' \) and \( x'' \), respectively. The immediately following epoch is referred to as the adjacent epoch. The one step transition probabilities are defined by the notation, \( \Pr(x'' \mid x') \). Two factors determine the value of \( \Pr(x'' \mid x') \). If the queue is initially empty, \( x' = 0 \), a different probability distribution governs the transition than if the queue is initially non-empty, \( x' > 0 \). Also, if at the adjacent epoch the queue has truncated, \( x'' = N \), a different mechanism governs the transition than if the queue has not truncated, \( x'' < N \). Thus a total of four cases cover all possible feasible situations. These cases are: Case a) \( x' > 0, x'' < N \); Case b) \( x' > 0, x'' = N \); Case c) \( x' = 0, x'' < N \); and Case d) \( x' = 0, x'' = N \). The transition probabilities for these four cases are determined from a consideration of the number of arrivals that enter the system between epochs. The probability of exactly \( a \) arrivals, \( a = 0, 1, 2, \ldots \) is defined under two conditions, truncation and non-truncation. Define
\[ A(a; s) = \text{marginal probability of exactly } a \text{ arrivals during a service time given that the system does not truncate and the service time density function is } f_s(t), a = 0, 1, 2, \ldots, s = 0, 1. \]

\[ \bar{A}(a; s) = \text{marginal probability of exactly } a \text{ arrivals during a service time given that the system does truncate and the service time density function is } f_g(t), a = 1, 2, \ldots, s = 0, 1. \]

It should be noted that the bar notation indicates system truncation, and the lack of the bar indicates the absence of system truncation.

Using these definitions the IMC transition probabilities are

\[
\begin{align*}
\text{Case a)} & \\
A(x'' - x' + 1; 1), & x' > 0, \\
 & x'' < N, \\
 & x'' - x' + 1 \geq 0,
\end{align*}
\]

\[
\begin{align*}
\text{Case b)} & \\
\bar{A}(N - x' + 1; 1), & x' > 0, \\
 & x'' = N,
\end{align*}
\]

\[
pr(x'' | x') = \begin{cases} 
(1) \\
\text{Case c)} & \\
A(x'; 0), & x' = 0, \\
 & x'' < N,
\end{cases}
\]

\[
\text{Case d)} \\
\bar{A}(n; 0), & x' = 0, \\
 & x'' = N,
\]

\[
\text{Case e)} & \\
0, & \text{otherwise.}
\]
It is noted that the number of customers required to cause the indicated transition is a function of the number of customers in the queue at the initial epoch, \( x' \), and the number of customers in the queue at the adjacent epoch, \( x'' \). For Case a) and Case b) the system is initially non-empty, \( x' > 0 \). For this situation the number of arrivals required to cause the transition is the difference between the number in the queue at the initial epoch and and the number in the queue at the adjacent epoch, plus one additional arrival to account for the customer receiving service. For Case c) and Case d) the system is initially empty, \( x' = 0 \). For this situation a customer must enter the system and receive service in order to trigger a service completion epoch, the adjacent epoch. Thus, the customer receiving service is not taken from the initial queue as in Case a) and Case b). For Case c) and Case d) the number of arrivals required to cause the transition is the number in the queue at the adjacent epoch.

The IMC for a general system is represented in matrix form in Figure 1. The matrix is designated the \( B \) matrix and is an \( (N + 1) \times (N + 1) \) square matrix. It should be noted that all of the elements of the first row of \( B \) depend on the service time density function \( f_0(t) \), as indicated by the variable \( s = 0 \) in the arrival probability expressions. All other non-zero elements depend on the service time density function \( f_1(t) \). Also, all the elements in the last column
Figure 1. Imbedded Markov Chain Transition Matrix for the Single Customer Class Model. The B Matrix.
of B, column \(N + 1\), are truncated arrival probabilities as indicated by the bar notation.

**Arrival Mechanism**

Before further development of the IMC, the arrival probability statements previously defined are evaluated. The statements are easily expressed in terms of Laplace transforms of the appropriate service time distributions.

The term \(A(a;s)\) is the marginal probability that exactly \(a\) arrivals enter the system in a service time which follows the probability density function \(f_s(t)\) and truncation does not occur. Consider first the joint probability that exactly \(a\) arrivals occur and the service time is between \(t\) and \(t + dt\). This probability is

\[
\frac{e^{-rt}(rt)^a}{a!} f_s(t) dt, \quad a = 0, 1, 2, \ldots, (2)
\]

\(t > 0,\)

as the mean arrival rate is \(r\) throughout the service time. Integrating over time yields the marginal probability of interest. The expression is

\[
A(a;s) = \int_0^{\infty} \frac{e^{-rt}(rt)^a}{a!} f_s(t) dt, \quad a = 0, 1, 2, \ldots (3)
\]

By Result A.1 of Appendix A this expression reduces to

\[
A(a;s) = \frac{(-r)^a}{a!} \bar{F}_s^{(a)}(r), \quad a = 0, 1, 2, \ldots
\]

where

\[
\bar{F}_s^{(a)}(r) = \frac{d^a}{dz^a} \left[ \bar{F}_s(f_s(t)) \right] \bigg|_{z=r}, a = 0, 1, 2, \ldots
\]
The importance of this finding is that possibly laborious integration can be avoided if the first $N$ derivatives of the Laplace transforms of the service time density functions are provided. As Laplace transforms are well tabulated for a large number of functions, it is not unreasonable to assert that this identity allows a sizable reduction in computational effort.

The term $\bar{A}(a;s)$ is the marginal probability that exactly $a$ arrivals enter the system, the $a$th arrival causing truncation, during a service time which follows the probability density function $f_s(t)$. Consider first the joint probability of $(a - 1)$ arrivals in a given time, $v$, before the end of service, and the arrival of the $a$th customer in time $v$ to $v + dv$. This joint probability is

$$
\int_{0}^{t} e^{-rv} \frac{(rv)^{a-1}}{(a-1)!} r dv,
$$

Integrating over $v$ yields the marginal probability of exactly $a$ arrivals occurring before a given time, $t$, as

$$
\int_{0}^{t} \frac{e^{-rv} (rv)^{a-1}}{(a-1)!} r dv,
$$

The joint probability of exactly $a$ arrivals and a service time between $t$ and $t + dt$ is

$$
\int_{0}^{t} e^{-rv} (rv)^{a-1} r dv f_s(t) dt,
$$

Integrating over $t$ yields the marginal probability of interest. The expression is
By Result A.2 of Appendix A the expression reduces to
\[ \overline{A}(a;s) = 1 - \sum_{n=0}^{a-1} \frac{(-r)^n}{n!} f_s(n)(r) , \quad a=1,2,... (9) \]
The previous comments regarding the computational efficiencies provided by identities involving Laplace transforms applies even more forcefully for this expression.

### Ergodic Nature of the Imbedded Markov Chain

To find the steady state probability of being in any state, \( x \), independent of the initial state of the system, it must first be shown that the IMC is ergodic. By the nature of the B matrix associated with the IMC it is readily proven that the IMC is ergodic. The arguments given follow those proposed by Ross (1970).

By examining the B matrix it is seen that it is possible to go from any state \( x \) to the empty state, \( x = 0 \), in a finite number of transitions. This can be accomplished, for example, by having a total of \( x \) services completed without an intervening arrival. Also, it is possible to go from the empty state, \( x = 0 \), to any non-empty state, \( x > 0 \). This can be accomplished, for example, by having a total of \( x \) arrivals during the service time of the first customer to arrive after the empty state was entered. Thus, in terms of finite Markov chains, any non-empty state and the empty state "communicate" as transitions can occur between them in a finite number of
steps, Ross (1970, p. 65). Also any non-empty state communicates with any other non-empty state. This follows from the fact that any two non-empty states communicate with the empty state. As all states communicate with each other the Markov chain is said to be "irreducible", Ross (1970, p. 66). An examination of the empty state indicates that it is possible for any system starting in the empty state to return to the empty state after any finite number of transitions. Such a state is said to be "aperiodic". As one state of the IMC is aperiodic, and the IMC is irreducible it follows that the IMC is aperiodic, Ross (1970, p. 66). Thus, the IMC is classified as a finite, irreducible, aperiodic Markov chain. The ergodic nature of the IMC is assured by a theorem due to A. A. Markov and given by Takacs (1962, p. 15) which states that all finite, irreducible, aperiodic Markov chains are ergodic.

**IMC Steady State Probabilities**

As the IMC is ergodic and has a finite number of states, the steady state probabilities exist and are independent of the initial state of the system. The equilibrium probabilities are defined as $g(x), x = 0, 1, \ldots, N$. $G$ is defined as the $N + 1$ element row vector with elements

$$G = \{g(x)\}$$

The equilibrium probabilities are found by solving the equations

$$GB = G$$

and

$$G1 = 1,$$

(10)
where \( \mathbf{l} \) is a \((N + 1)\) element column vector with all elements equal to one. As an alternative method, the method presented by Singer (1964) may be used. Briefly stated, Singer's method involves the use of the cofactors of a matrix derived from the transition matrix, \( \mathbf{B} \). Define the quantities

\[
\mathbf{D} = \mathbf{B}^T - \mathbf{I},
\]

where

\[
\mathbf{B}^T \text{ is the transpose of the } \mathbf{B} \text{ matrix,}
\]

\[
\mathbf{I} \text{ is the } (N + 1) \times (N + 1) \text{ identity matrix,}
\]

and

\[
\mathbf{D}_{xx} = \text{the cofactor of the diagonal element, } d_{xx}
\]

of the \( \mathbf{D} \) matrix.

The steady state probabilities are given by

\[
\mathbf{g}(x) = \frac{\mathbf{D}_{xx}}{\sum_{i=0}^{N} \mathbf{D}_{ii}}, \quad x = 0, 1, 2, \ldots, N. \quad (12)
\]

Augmented IMC

The first step in converting the IMC probabilities to the RPT probabilities is the addition of the idle service epoch to the state descriptions. The addition of this state to the IMC facilitates the subsequent development of the semi-Markov process. The idle service epoch, with state description \( \mathbf{e} \), is defined as the instant of time at which a customer who enters the system to find the server idle, starts service. Clearly the system can reach state \( \mathbf{e} \) only by first reaching the empty state, \( x = 0 \). A moment's reflection indicates that the probability of reaching state \( \mathbf{e} \) given that state
x = 0 is attained is unity. Thus the probability of being in state $e$ is the same as the probability of being in state $x = 0$. This fact will be expressed by writing the steady state probabilities associated with the augmented IMC in terms of the steady state probabilities of the original IMC. Define the equilibrium probabilities of the augmented IMC by

$$g'(y), \quad y = e, 0, 1, 2, \ldots, N.$$  \hfill (13)

The probabilities are

$$g'(e) = g'(0) = g(0)/c_1,$$

$$g'(x) = g(x)/c_1, \quad x = 1, 2, \ldots, N$$

where $c_1$ is a normalizing constant. The normalizing constant is needed to assure that the total of the probabilities, $g'(y)$, summed over the entire range of $y$ equals one. Clearly the events $y = e$ and $y = 0$ have the same absolute frequency of occurrence, but their relative frequency of occurrence is not $g(0)$. This is true because the additional state, state $e$, has been added to the event space, and the relative frequencies of occurrence must be reallocated on this increased event space. The normalizing constant, $c_1$, accounts for this reallocation. The normalizing constant, $c_1$, is not evaluated in the modeling procedure as it cancels in subsequent calculations. For purposes of clarification, however, the value of $c_1$ is

$$c_1 = 1 + g(0).$$  \hfill (15)

The augmented IMC is ergodic as can be shown by arguments identical to those used to show that the original IMC
is ergodic. It should also be noted that the mean time the system spends in any state is finite. The mean time spent in state $e$ is $m_0$, the mean of the service time distribution appropriate for customers that initiate service immediately after an idle period. The mean time spent in state $x$, $x = 1, 2, \ldots, N$, is $m_x$, the mean of the service time distribution appropriate for customers that initiate service during a busy period. The mean time spent in state $x = 0$ is $1/r$, the mean time required for an arrival to enter the system and thus cause a transition from state $x = 0$ to state $e$. These facts are utilized in the modification of the augmented IMC probabilities to SMP probabilities.

Semi-Markov Process (SMP)

The second step in obtaining the RPT probabilities is to convert the probabilities expressed for the augmented IMC on an epoch or event scale to a SMP which expresses the probabilities of these same states on a time scale. The SMP equilibrium probabilities consider the state of the system at the epoch just preceding some random point in time. Note that the system states that occur between adjacent epochs are not considered in the SMP. Figure 2 portrays the system states and scales for the augmented IMC, the SMP, and the RPT process system descriptions. Focusing on the augmented IMC and the SMP portrayals, it is seen that the SMP description translates the system states defined on the epoch or event
Figure 2. Comparison of Augmented IMC, SMP, and RPT System Descriptions.
scale to a time scale. From this base, the state of the system between adjacent epochs is obtained.

It is emphasized that the three processes described in Figure 2 are just three ways of viewing the same underlying system. Further, the realization of the system presented in Figure 2 is only an illustration of the way a graphical representation of the system would look when viewed as an IMC, SMP, and RPT, respectively. The calculations undertaken in this chapter provide only the equilibrium probabilities associated with the three processes. The information needed to construct such a realization is not provided.

To convert the augmented IMC probabilities to SMP probabilities a simple operation involving the mean times between adjacent epochs is performed. The validity of the operation is based on theoretical developments by Fabens (1961). Using the statement and notation of Fox (1967, p. 14), the stationary probabilities for SMP, $\rho_j$, associated with an ergodic, finite state IMC, with states $i = 1, 2, 3, \ldots, \eta$, are related to the stationary probabilities of the IMC, $\theta_i$, $i=1, 2, 3, \ldots, \eta$, by

$$
\rho_j = \frac{\theta_i v_j}{\sum_{i=1}^{\eta} \theta_i v_i}, \quad j = 1, 2, \ldots, \eta
$$

where

$$
v_j = \text{the mean time until the next epoch, given that the epoch associated with state } j \text{ has just occurred, } j = 1, 2, \ldots, \eta.
$$
Note that the summation in the denominator is merely a normalizing factor.

Applying the above result to the problem at hand provides direct evaluation of the SMP stationary probabilities defined as

\[ h(y), \quad y = e, 0, 1, 2, \ldots, N. \]  

The normalizing factor is designated as the constant \( c_2 \).

Utilizing the information concerning the mean time between epochs of the augmented IMC previously discussed, the SMP stationary probabilities are

\[ h(e) = g'(e)m_0/c_2, \]
\[ h(0) = g'(0)/(c_2r), \]

and

\[ h(x) = g'(x)m_1/c_2, \quad x = 1, 2, \ldots, N, \]  

where

\[ c_2 = g'(e)m_0 + g'(0)/r + \sum_{x=1}^{N} g'(x)m_1. \]

Substituting the expressions for \( g'(y) \) in terms of \( g(x) \) yields

\[ h(e) = g(0)m_0/(c_1c_2), \]
\[ h(0) = g(0)/(c_1c_2r), \]

and

\[ h(x) = g(x)m_1/(c_1c_2), \quad x = 1, 2, \ldots, N, \]  

where

\[ c_2 = g(0)m_0/c_1 + g(0)/(rc_1) + \sum_{x=1}^{N} g(x)m_1/c_1. \]
Defining $d = c_1 c_2$ and recalling that the sum of $g(x)$ over the entire range of $x$ is unity, yields

$$d = c_1 c_2$$  \hspace{1cm} (20)

$$= g(0)m_0 + g(0)/r + m_1 \sum_{x=1}^{N} g(x)$$

$$= g(0)m_0 + g(0)/r + m_1 (1 - g(0))$$

$$= m_1 + g(0)(m_0 + 1/r - m_1).$$

Thus the SMP stationary probabilities are

$$h(e) = g(0)m_0/d,$$

$$h(0) = g(0)/(rd),$$

and

$$h(x) = g(x)m_1/d, \quad x = 1, 2, \ldots, N.$$  \hspace{1cm} (21)

Random Point in Time (RPT) Probabilities

The RPT probabilities are developed by considering the state of the system at some random point in time following an epoch. By considering all feasible epochs, and the probabilities of making a transition from the system state at the epoch to the specified system state at a subsequent random point in time, the desired steady state probabilities are provided. To clarify this point, consider the ways in which the system with waiting room greater than two may reach a state of having two customers in the system at some random point in time. Clearly there are three mutually exclusive, collectively exhaustive cases: (1) at the preceding epoch there could have been two customers in the queue and no customer arrivals during the time between the instant of the
epoch and the random time of observation, (2) at the preceding epoch there could have been one customer in the queue and one customer arrival during the time between the instant of the epoch and the random time of observation, or (3) at the preceding epoch there could have been an initiation of service of a customer who arrived to find the system idle and one customer arrival during the time between the instant of the epoch and the random time of observation. Define the probability of exactly a arrivals during a random portion of a service time that follows the probability density function \( f_s(t) \), as \( A'(a; s) \). The probabilities of the above events are,

\[
\begin{align*}
  h(2)A'(0; 1), \\
  h(1)A'(1; 1), \\
  h(e)A'(1; 0),
\end{align*}
\]

respectively. As the events are mutually exclusive and collectively exhaustive, the sum of the above quantities yields the probability of there being exactly two customers in the system at any random point in time.

Formalizing the preceding example leads to the general method used to convert the SMP probabilities to the desired random point in time (RPT) probabilities. It is emphasized that the states of the RPT are defined by the number of customers in the system, not the number of customers in the queue as for the states of the IMC and SMP. The RPT steady
state probabilities are defined as
\[ p(z), \quad z = 0, 1, 2, \ldots, N + 1 \] (23)
the probability of exactly \( z \) customers in the system.

The change of state definition from the number of customers in the queue to the number of customers in the system is a result of viewing the system at any random point in time. When viewing the system at service completion epochs the event, zero customers in the queue, is synonymous with the event, zero customers in the system. This is based on the fact that at the cited epoch a customer has just completed service and there are no customers waiting to enter the service facility. When viewing the system at any random point in time the event, zero customers in the queue, is not synonymous with the event, zero customers in the system. Therefore, to clearly define the state of the system at any random point in time it is necessary to consider the number of customers in the system, not the number of customers in the queue. With this in mind, it is emphasized that the system truncates when \( N + 1 \) customers are in the system, and the range of the variable \( z \) is all of the integers from zero through \( N + 1 \).

**Random Interruption Arrival Mechanism**

The probability of having a specific number of customer arrivals during the time between the instant of the epoch and the subsequent random time of observation, defined as time \( t' \), is required. These probabilities are defined as
random interruption time arrival probabilities. They are evaluated before continuing the explanation of converting the SMP probabilities to RPT probabilities.

The random interruption arrival probabilities are analogous to the quantities, \( A(a;s) \) and \( \bar{A}(a;s) \), previously developed, but the mechanism is slightly modified. The previously developed arrival probabilities allow arrivals to occur at any time between the beginning of service (the instant of an epoch) and the conclusion of service. This time period is of length \( t \), which obeys the probability density function \( f_S(t) \). The random interruption arrival probabilities allow arrivals to occur at any time between the beginning of service (the instant of an epoch) and the random time of observation. This time period is of length \( t' \).

This relationship is portrayed in Figure 3. The length of time, \( t' \), is referred to in the literature as the "backward delay time" or "random interruption time" and has a probability density function related to the probability density function for the service time, \( f_S(t) \). The random interruption time probability density function as given by Fox (1967, p. 15) is

\[
f_S'(t') = \frac{(1 - F_S(t'))}{m_S}, \quad t' > 0,
\]

where

\[
F_S(t') = \int_0^{t'} f_S(t) dt.
\]

This useful result is based on renewal theory arguments. An
Number in the System vs. Time

Figure 3. Illustration of the Arrival Mechanism for a Service Time, t, and a Random Interruption Time, t'.
intuitively appealing interpretation of the density function is given by Conway, Maxwell, and Miller (1967, p. 146).

The arrival probabilities for the random interruption times are now defined with the same logic utilized in the definition of the arrival probabilities for the full service times. A prime super script indicates that the random interruption time instead of the full service time is allowed for arrivals.

\[ A'(a;s) = \text{marginal probability of exactly } a \text{ arrivals during a random interruption time given that the system does not truncate and the random interruption time density function is } f'_g(t'), a = 0,1,2,..., s = 0,1. \]

\[ \bar{A}'(a;s) = \text{marginal probability of exactly } a \text{ arrivals during a random interruption time given that the system does truncate and the random interruption time density function is } f'_s(t'), a = 1,2,..., s = 0,1. \]

The random interruption time arrival probabilities, \( A'(a;s) \) and \( \bar{A}'(a;s) \), are evaluated by the same marginal probability statements as used for the service time arrival probabilities, \( A(a;s) \) and \( \bar{A}(a;s) \), respectively, except that the time density function, \( f_s(t) \), is replaced by \( f'_g(t') \). These expressions are
and (25)

\[ A'(a;s) = \int_0^\infty e^{rt'} \frac{(rt')^a}{a!} f_s(t') dt', \quad a = 0, 1, 2, \ldots \]

By Result A.4 of Appendix A, the non-truncated arrival probability reduces to

\[ A'(a;s) = \frac{1}{m_s r} \left[ 1 - \sum_{n=0}^{a} \frac{(-r)^n}{n!} \bar{F}(n)(r) \right], \quad a = 0, 1, 2, \ldots \quad (26) \]

By Result A.5 of Appendix A, the truncated arrival probability reduces to

\[ \bar{A}'(a;s) = 1 - \frac{1}{m_s r} \left[ a - \sum_{n=0}^{a-1} \sum_{h=0}^{n} \frac{(-r)^h}{h!} \bar{F}(h)(r) \right], \quad a = 1, 2, \ldots \quad (27) \]

Therefore, the random interruption time arrival probabilities are easily expressed in terms of the derivatives of the Laplace transforms of the service time distributions. This allows possibly troublesome integration to be avoided.

Recalling the simplified expressions for the service time arrival probabilities, \( A(a;s) \) and \( \bar{A}(a;s) \), it is seen that all of the probabilities needed to develop the RFT steady state probabilities can be expressed in the same manner. The notation of these four probabilities is further simplified by defining three intermediate expressions. This simplified notation facilitates the development of the equivalent probabilities for the multiple customer class systems considered in subsequent chapters. The simplified probability statements are contained in Table 1.
<table>
<thead>
<tr>
<th>Probability of Exactly a Arrivals</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A(a;s) = P(a;s), \quad a=0,1,... )</td>
<td>Time Allowed for Arrivals</td>
</tr>
<tr>
<td></td>
<td>System Truncation</td>
</tr>
<tr>
<td>( \overline{A}(a;s) = 1 - \overline{P}(a-1;s), \quad a=1,2,... )</td>
<td>Service time</td>
</tr>
<tr>
<td></td>
<td>Truncation</td>
</tr>
<tr>
<td>( A'(a;s) = \frac{1}{m_s r}[1 - \overline{P}(a;s)], \quad a=0,1,... )</td>
<td>Random interruption portion of service time</td>
</tr>
<tr>
<td></td>
<td>No truncation</td>
</tr>
<tr>
<td>( \overline{A}'(a;s) = 1 - \frac{1}{m_s r}[a - \overline{P}(a-1;s)], )</td>
<td>Random interruption portion of service time</td>
</tr>
<tr>
<td></td>
<td>Truncation</td>
</tr>
</tbody>
</table>

Where

\[ P(a;s) = \frac{(-r)^a}{a!} \overline{P}(a)(r), \]

\[ \overline{P}(a;s) = \sum_{n=0}^{a} P(n;s), \]

and

\[ \overline{P}(a;s) = \sum_{n=0}^{a} \sum_{h=0}^{n} P(h;s). \]

Table 1. Arrival Probabilities for the Single Customer Class Model.
RPT Probability Conversion

Using the random interruption arrival probabilities and the SMP steady state probabilities, the RPT steady state probabilities can be determined. Recall the definition of \( p(z), z = 0, 1, 2, \ldots, N + 1 \), as the probability that at any randomly chosen point in time there are exactly \( z \) customers in the system (receiving service or waiting for service). The expressions for \( p(z) \), except for \( p(0) \), involve convolution-like terms involving \( h(y) \) and \( A'(a; s) \) or \( A^*(a; s) \). A general expression for \( p(z) \) is developed from a consideration of the first few probabilities, \( p(0), p(1), p(2), \ldots, \) etc.

The exceptional case of having zero in the system at a random point in time is treated first. Clearly, there can be zero in the system if and only if the preceding epoch is a service completion epoch at which zero customers remain in the queue. Thus the probability of finding zero customers in the system at a random point in time, \( p(0) \), is equal to the probability that the preceding epoch was an empty service completion epoch, \( h(0) \). Stated in equation form

\[
p(0) = h(0).
\]  

(28)

The other probabilities can be determined inductively. The case of one customer in the system at a random point in time can occur in two mutually exclusive, collectively exhaustive ways. These are: (1) at the preceding service completion epoch there was one customer in the queue and no
customers arrived in the random interruption time, or (2) the preceding epoch was an idle service initiation epoch and no customers arrived in the random interruption time. This probability, stated in equation form, is

\[ p(1) = h(1) A'(0;1) + h(e) A'(0;0). \]  

(29)

Similarly, two customers in the system at a random point in time can occur in three mutually exclusive, collectively exhaustive ways. These are (1) at the preceding service completion epoch there was one customer in the queue and one customer arrived in the random interruption time, (2) at the preceding service completion epoch there were two customers in the queue and no customers arrived in the random interruption time, or (3) the preceding epoch was an idle service initiation epoch and one customer arrived during the random interruption time. In equation form the probability is

\[ p(2) = h(1) A'(1;1) + h(2) A'(0;1) + h(e) A'(1;0). \]  

(30)

A moment's reflection indicates that the general expression for the probability of finding exactly \( z \) customers, \( z = 1, 2, \ldots, N \) in the system at a random point in time is

\[ p(z) = \sum_{y=1}^{z} h(y) A'(z-y;1) + h(e) A'(z-1;0), \quad z = 1, 2, \ldots, N. \]  

(31)

The remaining probability of interest is the probability that exactly \( N + 1 \) customers are in the system at a random point in time. This implies that the system has truncated during the random interruption time, and the arrival probabilities, \( A'(a,s) \), are appropriate for the calculation. The
expression is identical to that given for the non-truncated case except that the upper limit on the summation is reduced by one as it is impossible to have \( N + 1 \) customers in the queue. The arrival probabilities are also revised to account for the truncation phenomenon. The probability of \( N + 1 \) customers in the system at a random point in time is

\[
p(N+1) = \sum_{y=1}^{N} h(y) A'(N+1-y;1) + h(e) \overline{A}'(N;0) . \quad (32)
\]

Restating these results in concise form

\[
p(z) = \begin{cases} 
  h(0), & z = 0, \\
  \sum_{y=1}^{z} h(y) A'(z-y;1) + h(e) A'(z-1;0), & z = 1, 2, \ldots, N, \\
  \sum_{y=1}^{N} h(y) \overline{A}'(N+1-y;1) + h(e) \overline{A}'(N;0), & z = N + 1. 
\end{cases} \quad (33)
\]

The RPT probabilities can be determined by matrix operations if the conversion matrix, \( B' \), is formed. The conversion matrix is similar in composition to the IMC matrix, \( B \), and thus facilitates the calculations. The use of the \( B' \) matrix is of more importance in the multi-customer class models and is presented for the single class system in the interest of generalizing the modeling procedures. The \( B' \) matrix for the one customer system is shown in Figure 4. Note that the rows are indexed on variable \( y, y=1, 2, \ldots, N \), and the columns are indexed on variable \( z, z=1, 2, \ldots, N+1 \). The \( B' \) matrix is of dimension \( N \times (N + 1) \). Designate the elements of the \( B' \) matrix as
<table>
<thead>
<tr>
<th>from (y)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>\ldots</th>
<th>N</th>
<th>N + 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(A'(0;1))</td>
<td>(A'(1;1))</td>
<td>(A'(2;1))</td>
<td>(A'(3;1))</td>
<td>\ldots</td>
<td>(A'(N-1;1))</td>
<td>(\overline{A}'(N;1))</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>(A'(0;1))</td>
<td>(A'(1;1))</td>
<td>(A'(2;1))</td>
<td>\ldots</td>
<td>(A'(n-2;1))</td>
<td>(\overline{A}'(n-1;1))</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>(A'(0;1))</td>
<td>(A'(1;1))</td>
<td>\ldots</td>
<td>(A'(N-3;1))</td>
<td>(\overline{A}'(N-2;1))</td>
</tr>
<tr>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
<td>\ldots</td>
</tr>
<tr>
<td>N - 1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>\ldots</td>
<td>(A'(1;1))</td>
<td>(\overline{A}'(2;1))</td>
</tr>
<tr>
<td>N</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>\ldots</td>
<td>(A'(0;1))</td>
<td>(\overline{A}'(1;1))</td>
</tr>
</tbody>
</table>

Figure 4. Conversion Matrix for the Single Customer Class Model.

The \(B'\) Matrix.
These elements represent the probability of making the transition from a system with state description \( y \) at an epoch, to a system with state description \( z \) at some randomly selected point in time immediately after the epoch. These transitions occur by having the same number of arrivals as indicated by the equivalent element in the IMC transition matrix, \( B \). The only difference is that the time allowed for the arrivals is the random interruption portion of the service time, not the entire service time. Comparing the equivalent elements, \( b(x' = 1, x'' = 2) = A(1;1) \) from Figure 1, and \( b'(y = 1, z = 2) = A'(1;1) \) from Figure 4, indicates this relationship. It is emphasized that the \( B' \) matrix is not a transition matrix, but rather a matrix of transition probabilities from some of the states of the SMP, \( y, y = 1,2,\ldots, N \), to some of the states of the RPT process, \( z, z = 1,2,\ldots, N + 1 \). In short, the \( B' \) matrix provides a computationally convenient way of performing certain calculations.

Using the elements of the \( B' \) matrix, the random point in time probabilities, \( p(z) \), are

\[
p(z) = \begin{cases} 
  h(0), & z = 0, \\
  \sum_{y=1}^{N} h(y) b'(y, z) + h(e) A'(z-1;0), & z = 1,2,\ldots,N, \\
  \sum_{y=1}^{N} h(y) b'(y, N + 1) + h(e) \bar{A}'(N;0), & z = N + 1.
\end{cases}
\]

(35)
It should be noted that the last two expressions in the definition of \( p(z) \) can be partially rewritten in terms of matrix operations. The summations are equivalent to post multiplying the \( N \) element row vector.

\[
(h(1), h(2), \ldots, h(N)),
\]

by the \( z \)th column of the \( B' \) matrix. This type of operation greatly facilitates the conversion of probabilities in the multiple customer class models.

**Measures of Performance**

The previous model has provided information about the behavior of the system after a steady state condition has been attained. To utilize the information several additional calculations are performed. All measures of system performance are for systems that have reached equilibrium.

The probability of exactly \( n \) in the system, \( pr(n) \), is available directly as

\[
pr(n) = p(n), \quad n = 0,1,2,\ldots,N + 1. \quad (36)
\]

The probability of exactly \( n_q \) in the queue, \( pr(n_q) \), is obtained by

\[
pr(n_q) = \begin{cases} 
  p(0) + p(1), & n_q = 0 \\
  p(n_q + 1), & n_q = 1,2,\ldots,N.
\end{cases} \quad (37)
\]

With the above probability mass functions, it is possible to calculate the expected number of customers in the system and the expected number of customers in the queue, designated \( E(n) \) and \( E(n_q) \), respectively. The expressions are
\[ E(n) = \sum_{n=0}^{N+1} n \cdot pr(n), \]

and

\[ E(n_q) = \sum_{n_q=0}^{N} n_q \cdot pr(n_q). \]

The effective mean arrival rate, \( r_{eff} \), is the expected number of customers added to the system per unit time. Since the system truncates, not all customers desiring to enter the system are allowed to enter. To obtain the effective mean arrival rate, the mean arrival rate, \( r \), is multiplied by the probability that the system is not truncated, \( (1 - pr(n = N+1)) \). The validity of this calculation is based on the fact that customers arrive at random points in time, and the probability that the system is not truncated at any random point in time is \( (1 - pr(n = N+1)) \). The effective mean arrival rate is

\[ r_{eff} = r(1 - pr(n = N+1)). \]

The expected waiting time in the system, \( E(w) \), and the expected waiting time in the queue, \( E(w_q) \), are of interest. Expressions for these quantities are developed on the basis of a result proven by Eilon (1969). Eilon shows that under the steady state assumption, the expected waiting time in the system and the expected waiting time in the queue are simple functions of the expected number in the system and the expected number in the queue, respectively. The relationships are
\[ E(w) = E(n)/r_{\text{eff}} \]

and

\[ E(w_q) = E(n_q)/r_{\text{eff}}. \]

These relationships are very robust as a minimum number of assumptions are required. As stated by Eilon (1969, p. 916), "This proof is not dependent on any specific assumptions about arrival or service distributions, about the number of servers in the system, or about the queueing discipline."

The effective arrival rate, \( r_{\text{eff}} \), is appropriate for the calculation as Eilon considers a limiting argument based on a period of time of length \( T \) after the system has settled down to steady state. The effective mean arrival rate is defined as

\[
r_{\text{eff}} = \lim_{T \to \infty} \frac{\text{total number of arrivals during time } T}{T}. \tag{41}
\]

Using the terminology of Hadley and Whitin (1963, p. 104), \( r_{\text{eff}} \) is the time average of the arrival rate. As Hadley and Whitin state, when statistical equilibrium is reached the time average can be interpreted as an expected value. As \( r_{\text{eff}} \) has been defined as the expected mean arrival rate after equilibrium conditions are attained, it is correctly used in the calculations. The generality of the waiting time relationship as proven by Eilon is exploited in the multiple customer class models.

One additional measure of performance, the expected service time, \( E(t) \), is readily derived as
\[ E(t) = E(w) - E(w_q) \]  

(42)

This is a non-trivial result as the service time distribution is dependent on the system state, and the overall mean service time is not given as input data.

**Example One**

To illustrate the model a small numerical example is given. The data have no significance beyond computational simplicity. In general, any data complying with the restrictions stated previously are appropriate.

The small problem considered consists of a queue that contains at most two customers. The mean arrival rate is one customer per minute. Service times are also expressed in units of minutes. Service time distributions are of the gamma type with parameters \( b_s \) and \( c_s \). The parameters for the service time distribution for a customer who arrives to find the system idle are \( b_0 = 1 \), and \( c_0 = 2 \). The parameters for the service time distribution for a customer who arrives to find the system busy are \( b_1 = 1 \), and \( c_1 = 3 \). Defining the data in terms of the model notation

\[ N = 2, \text{ maximum number in the queue,} \]

\[ f_s(t) = \frac{b_s^{c_s}}{(c_s-1)!} t^{c_s-1} e^{-b_st}, \quad t > 0, \ s = 0,1, \]

= service time distribution,

and

\[ m_s = \frac{c_s}{b_s}, \text{ mean of the service time distribution.} \]
Arrival Probabilities

The arrival probabilities for the service time and the random interruption time are calculated as specified in Table 1, using the Laplace transforms of the service time density functions. The Laplace transforms of gamma density functions are particularly easy to work with. The Laplace transform of the gamma density function is

\[ \mathcal{F}_s(z) = \int_0^\infty e^{-tz} \frac{b_s^{cs}}{(c_s-1)!} t^{c_s-1} e^{-bst} \, dt, \]  

\[ = (1 + z/b_s)^{-c_s}. \]

The nth derivative of the Laplace transform can be shown by induction to be

\[ \mathcal{F}_s^{(n)}(z) = \frac{(c_s+n-1)!}{(c_s-1)!} (-b_s)^{-n} (1 + z/b_s)^{-c_s-n} \quad n=0,1,2,\ldots \]

The quantity of interest in the arrival probability calculations, \( P(a;s) \), as defined in Table 1 is

\[ P(a;s) = \frac{(-a)^a}{a!} \mathcal{F}_s^{(a)}(z) \bigg|_{z=r}, \quad a=0,1,2,\ldots \]

\[ = \frac{(-r)^a}{a!} \frac{(c_s+a-1)!}{(c_s-1)!} (-b_s)^{-a(l+r/b_s)} -c_s-a \quad a=0,1,2,\ldots \]

\[ = \frac{(c_s+a-1)!}{a!} \frac{(b_s/r)^{-a(l+r/b_s)} -c_s-a}{(b_s-1)!}, \quad a=0,1,2,\ldots \]

\[ = \frac{(c_s+a-1)}{a} \left( \frac{r}{b_s+r} \right)^a \left( \frac{b_s}{b_s+r} \right)^c, \quad a=0,1,2,\ldots \]
This negative binomial term is easily evaluated for the given values of $b_s$ and $c_s$. The other intermediate expressions are obtained by summation of the $P(a;s)$ terms as specified in Table 1. The intermediate terms and arrival probabilities are given in Table 2 and Table 3.

Calculations

The imbedded Markov chain is given in Figure 5. The stationary probabilities associated with the IMC are found by using the method given by Singer (1964). The $D$ matrix is

$$D = B^T - I$$

$$= \begin{pmatrix}
4/16 - 1 & 2/16 & 0 \\
4/16 & 3/16 - 1 & 2/16 \\
8/16 & 11/16 & 14/16 - 1
\end{pmatrix}. \quad (46)$$

The cofactors of the diagonal elements are

$$D_{00} = \begin{vmatrix}
-13/16 & 2/16 \\
11/16 & -2/16
\end{vmatrix} = 4/256,$$

$$D_{11} = \begin{vmatrix}
-12/16 & 0 \\
8/16 & -2/16
\end{vmatrix} = 24/256,$$

and

$$D_{22} = \begin{vmatrix}
-12/16 & 2/16 \\
4/16 & -13/16
\end{vmatrix} = 148/256. \quad (47)$$
<table>
<thead>
<tr>
<th>Number of Arrivals $a$</th>
<th>Intermediate Expression</th>
<th>Arrival Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P(a;0)$</td>
<td>$\overline{P}(a;0)$</td>
</tr>
<tr>
<td>0</td>
<td>4/16</td>
<td>4/16</td>
</tr>
<tr>
<td>1</td>
<td>4/16</td>
<td>8/16</td>
</tr>
<tr>
<td>2</td>
<td>3/16</td>
<td>11/16</td>
</tr>
</tbody>
</table>

Table 2. Intermediate Expressions and Arrival Probabilities for Customers that Enter the System When the Server is Idle.
<table>
<thead>
<tr>
<th>Number of Arrivals $a$</th>
<th>Intermediate Expression</th>
<th>Arrival Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$P(a;1)$</td>
<td>$\bar{P}(a;1)$</td>
</tr>
<tr>
<td>0</td>
<td>2/16</td>
<td>2/16</td>
</tr>
<tr>
<td>1</td>
<td>3/16</td>
<td>5/16</td>
</tr>
<tr>
<td>2</td>
<td>3/16</td>
<td>8/16</td>
</tr>
</tbody>
</table>

Table 3. Intermediate Expressions and Arrival Probabilities for Customers that Enter the System When the Server is Busy.
Table 2: IMC Transition Matrix for the Single Customer Class Model of Example One. The B Matrix.

<table>
<thead>
<tr>
<th>From State ( x' )</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>( A(0;0) = \frac{4}{16} )</td>
<td>( A(1;0) = \frac{4}{16} )</td>
<td>( \overline{A}(2;0) = \frac{8}{16} )</td>
</tr>
<tr>
<td>1</td>
<td>( A(0;1) = \frac{2}{16} )</td>
<td>( A(1;1) = \frac{3}{16} )</td>
<td>( \overline{A}(2;1) = \frac{11}{16} )</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>( A(0;1) = \frac{2}{16} )</td>
<td>( \overline{A}(1;1) = \frac{14}{16} )</td>
</tr>
</tbody>
</table>
The stationary probabilities are \( g(0) = 4/176 \), \( g(1) = 24/176 \), and \( g(2) = 148/176 \), where
\[
g(x) = \frac{D_{xx}}{(D_{00} + D_{11} + D_{22})}, \quad x = 0, 1, 2. \quad (48)
\]
The SMP probabilities are found by modifying the IMC probabilities. The normalizing constant, \( d \), is
\[
d = m_1 + g(0) (m_0 + 1/r - m_1) \quad (49)
\]
\[
= 3 + \left(\frac{4}{176}\right) (2 + (1/1) - 3) \\
= 3.
\]
SMP stationary probabilities, \( h(y) \), \( y = e, 0, 1, 2 \), are
\[
h(e) = g(0) \frac{m_0}{d} \\
= \left(\frac{4}{176}\right) \frac{2}{3} \\
= 0.0151,
\]
\[
h(0) = g(0) \frac{1}{rd} \\
= \left(\frac{4}{176}\right) \frac{(1) (3)}{1} \\
= 0.0076,
\]
\[
h(1) = g(1) \frac{m_1}{d} \\
= \left(\frac{24}{176}\right) \frac{3}{3} \\
= 0.1364,
\]
and
\[
h(2) = g(2) \frac{m_1}{d} \\
= \left(\frac{148}{176}\right) \frac{3}{3} \\
= 0.8409.
\]

The \( B' \) matrix used to convert the SMP stationary probabilities to the RPT stationary probabilities is displayed in Figure 6. The SMP stationary probabilities, \( p(z) \),
Figure 6. Conversion Matrix for the Single Customer Class Model of Example One. The B' Matrix.
\( z = 0, 1, 2, \ldots, N + 1 \) yield the desired probability of having exactly \( z \) customers in the system at any arbitrary point in time after equilibrium conditions have been attained. The value of \( p(0) \) is given, and then \( p(z) \), \( z = 1, 2, \ldots, N + 1 \), is found using the conversion matrix, \( B' \).

\[
p(0) = h(0) \]
\[
= 0.0076,
\]
\[
p(1) = h(1) b(1,1) + h(e) A'(0;0)
\]
\[
= h(1) A'(0;1) + h(e) A'(0;0)
\]
\[
= (0.1364) (28/96) + (0.0151) (6/16)
\]
\[
= 0.0455,
\]
\[
p(2) = h(1) b'(1,1) + h(2) b'(2,2) + h(e) A'(1;0)
\]
\[
= h(1) A'(1;1) + h(2) A'(0;1) + h(e) A'(1;0)
\]
\[
= (0.1364) (22/96) + (0.8409) (28/96)
\]
\[
+ (0.0151) (4/16)
\]
\[
= 0.2803,
\]

and

\[
p(3) = h(1) b'(1,3) + h(2) b'(2,3) + h(e) A'(2;0)
\]
\[
= h(1) A'(2;1) + h(2) A'(1;1) + h(e) A'(2;0)
\]
\[
= (0.1364) (46/96) + (0.8409) (68/96)
\]
\[
+ (0.0151) (6/16)
\]
\[
= 0.6666.
\]

All steady state probabilities are tabulated in Table 4.

**Performance Measures**

Using the probability mass function, \( p(z) \), the expected number of customers in the system and the expected number of
<table>
<thead>
<tr>
<th>State Description</th>
<th>Steady State Probability</th>
<th>Imbedded Markov Chain $g(x); x=0,1,2.$</th>
<th>Semi-Markov Process $h(y); y=e,0,1,2.$</th>
<th>Random Point in Time $p(z); z=0,1,2,3.$</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>--</td>
<td>0.0151</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>0</td>
<td>0.0227</td>
<td>0.0076</td>
<td>0.0076</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.1364</td>
<td>0.1364</td>
<td>0.0455</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.8409</td>
<td>0.8409</td>
<td>0.2803</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>0.6666</td>
</tr>
</tbody>
</table>

Table 4. Steady State Probabilities for the Single Customer Class Model of Example One.
customers in the queue are

\[ E(n) = \sum_{n=0}^{3} n p(z=n) = 2.6059 \text{ customers}, \]

and

\[ E(n_q) = \sum_{n=1}^{2} n_q p(z=n_q+1) = 1.6135 \text{ customers}, \]

respectively.

The effective mean arrival rate is

\[ r_{\text{eff}} = r(1 - p(z=3)) = 1(1 - 0.6666) = 0.3333 \text{ customers/min.} \]

Using the relationship between expected waiting time and expected number of waiting units, the expected waiting time in the system and the expected waiting time in the queue are

\[ E(w) = \frac{E(n)}{r_{\text{eff}}} = \frac{2.6059}{0.3333} = 7.8185 \text{ min.}, \]

and

\[ E(w_q) = \frac{E(n_q)}{r_{\text{eff}}} = \frac{1.6135}{0.3333} = 4.8410 \text{ min.}, \]

respectively. The expected service time is
\[ E(t) = E(w) - E(w_q) \]
\[ = 7.8185 - 4.8410 \]
\[ = 2.9775 \text{ min.} \]

**Summary**

The single customer class model presented in this chapter provides the basic modeling perspective to consider multiple customer class models. The same steps followed in the development of the single class model are appropriate for the multiple customer class models. In addition, many of the specific arguments, such as the discussion of ergodicity and the justification of the arrival probability expressions, are valid for subsequent models. Although the actual research for the multi-customer class models was performed before the one class model was proposed, the one class model provides a natural transition to the more complex extended models.
CHAPTER IV
A MULTIPLE CUSTOMER CLASS MODEL

Introduction

In this chapter the single customer class model developed in Chapter III is extended to consider a customer population that consists of a finite number of homogeneous customer classes. The modeling steps followed in the development of the single customer class model are appropriate for the multiple customer class models if certain features are modified to account for the multiple class customer population. The model developed in this chapter assumes that the service selection procedure is based on a non-preemptive priority sequencing rule; however, it is shown that the model is easily modified to treat other sequencing rules.

The order of presentation consists of a description of the problem and possible real world contexts followed by a complete discussion of the quantitative elements of the model. Intuitive arguments for the derived expressions are given within the body of the chapter, but extensive mathematical manipulations are placed in Appendix A. All elements of the multiple customer class model are formulated in such a way that they reduce to the equivalent single customer class
model when the special case of $k = 1$ is considered. The chapter concludes with a small numerical example.

**Problem Description**

The problem treated in this section of the study is based on the same assumptions considered in Chapter III except the customer population has been expanded to consist of $k$ homogeneous classes, $k = 1, 2, \ldots$. The subscript, $i$, $i = 1, 2, \ldots, k$, indicates the class of customer. All customers from the same class are indistinguishable and exhibit the same characteristics with regard to arrival rates and service times.

The mean arrival rate for the $i$th class customer is $r_i$, $i = 1, 2, \ldots, k$, and the total mean arrival rate, $r$, is the sum of the mean arrival rates for all classes of customers. The assumption of completely random arrival patterns is retained, so the Poisson distribution describes the probability of having exactly $a_i$ class $i$ customers arrive in a given time. The service time for a particular class of customer depends not only on the class of customer receiving service, but also on the class of customer, if any, that completed service immediately before the particular customer's service was initiated. Service times of this type are referred to as interactive service times. In addition, service times are stochastic. These features are expressed by the service time probability density functions, $f_s(t)$, where $s = ij$, $i = 0, 1, \ldots, k$, $j = 1, 2, \ldots, k$. The density function $f_{ij}(t)$, is
appropriate for the service time of a class $j$ customer, $j = 1, 2, \ldots, k$, that immediately follows a class $i$ customer, $i = 1, 2, \ldots, k$, on the service facility. The density function, $f_{0j}(t)$, is appropriate for the service time of a class $j$ customer, $j = 1, 2, \ldots, k$, that enters the service facility after the facility has been idle for some non-zero period of time. For a system with $k$ classes of customers, $(k + 1)k$ probability density functions describe the service time phenomena.

The sequencing rule which determines the order of service is fixed. In this section the non-preemptive priority or head-of-the-line sequencing rule is considered. This classical priority scheme assigns class indices, $i = 1, 2, \ldots, k$, in inverse order of priority. Class 1 is the highest priority class, and the customers of this class take precedence over all other customers. The non-preemptive qualification means that once a customer starts to receive service, the service is not interrupted. Within each class of customer the order of service is determined on a first come-first served basis.

The remaining assumptions are identical to those previously used and will not be restated at length. In queueing theory terms the problem is an $M/G/1$ queue that serves $k$ classes of customers according to a non-preemptive priority service discipline. The waiting space is truncated and service times are interactive.
Related Studies

Problems with similar assumptions have been analyzed previously. Conway, Maxwell, and Miller (1967, Chapter 9) discuss a number of queueing type models in which a setup time is required between jobs of different classes. The setup time is assumed to be independent of the class of job previously served, and there are no limitations on waiting space.

Reinitz (1963) has considered the $M/M/1$ queue with finite waiting space, multiple customer classes and interactive service times. His approach utilizes the dynamic programming and Markov chain optimization techniques developed by Howard (1960). The decision is to determine the class of customer to serve for each possible combination of available customers with the objective of minimizing the steady state waiting time costs. Other studies that have less features in common with the models in this chapter are discussed in detail in Chapter II.

Model Development

Overview

Two objectives are served by this section. First, the model components previously applied to a single customer class model are extended to the case of a multiple customer class population for use with any general sequencing rule. Second, the extended model components are applied to the specific case of a non-preemptive priority sequencing rule.
The explanation of this specific application is given in detail as the same procedures are again utilized for the models of the three additional sequencing rules given in Chapter V. Concisely stated this section is the documented justification for the multiple customer class model formulations.

Extension of the model components includes the development of arrival probability expressions for an exact number of each class of customer, \( a_i, i = 1, 2, \ldots, k \). The arrival probabilities for a service time and a random interruption time under truncated and non-truncated conditions are defined, and simplified expressions are derived. These expressions are used in the same sequence of steps as for the one customer class model.

The first step consists of forming an imbedded Markov chain (IMC) which views the system only at service completion epochs. The states of the system at these points are given by the vector, \( X = (x_0, x_1, \ldots, x_k) \), where \( x_0 = 1, 2, \ldots, k \), is the class of customer just completing service, and \( x_i = 0, 1, 2, \ldots, i = 1, 2, \ldots, k \), is the number of class \( i \) customers awaiting service at the epoch. Using the sequencing rule and the service time arrival probabilities, the transition matrix for the IMC, \( B \), is given. The IMC is ergodic, and steady state probabilities are found by the standard procedure. The IMC steady state probabilities are designated \( g(X) \), where \( X = (x_0, x_1, \ldots, x_k) \). The IMC
probability, $g(X)$, is interpreted as the probability that at any randomly selected epoch after equilibrium conditions have been attained, the customer just completing service is of class $x_0$, and the number of class $i$ customers awaiting service is $x_i$, $i = 1, 2, \ldots, k$. The remaining steps in the procedure are devoted to converting the IMC steady state probabilities to the random point in time (RPT) steady state probabilities.

The augmented IMC is obtained from the IMC by adding the idle service initiation epochs to the state descriptions of the IMC. The idle service initiation epoch, with state description $e_i$, $i = 1, 2, \ldots, k$, is defined as the instant of time when a class $i$ customer enters the system to find the server idle. The steady state probabilities of the augmented IMC are designated $g'(Y)$, $Y = e_i$, $i = 1, 2, \ldots, k$, or $X$. The augmented IMC steady state probability, $g'(Y)$, is interpreted as the probability that at any randomly selected epoch (either a service completion epoch or an idle service initiation epoch) after equilibrium conditions have been attained, the state description is given by $Y$. The state description, $Y$, is $e_i$ if the randomly selected epoch is a class $i$ customer idle service initiation epoch. The state description, $Y$, is $X$ if the randomly selected epoch is a service completion epoch. The augmented IMC probabilities are obtained directly from the IMC probabilities.

The semi-Markov process (SMP) equilibrium probabilities
are obtained from the augmented IMC equilibrium probabilities by the procedure given by Fox (1967). The SMP equilibrium probabilities are designated \( h(Y) \), \( Y = e_i, \ i = 1,2, \ldots, k \), or \( X \). The SMP steady state probability, \( h(Y) \), is interpreted as the probability that given any randomly selected point in time after equilibrium conditions have been attained, at the immediately preceding epoch (either a service completion epoch or an idle service initiation epoch) the state description was \( Y \). The state description for the randomly selected point in time, \( Y \), is \( e_i \) if the immediately preceding epoch was a class \( i \) customer idle service initiation epoch. The state description for the randomly selected point in time, \( Y \), is \( X \) if the immediately preceding epoch was a service completion epoch.

The final step in obtaining the steady state probabilities of ultimate interest is the conversion of the SMP equilibrium probabilities to the random point in time (RPT) equilibrium probabilities. The RPT equilibrium probabilities are designated \( p(Z) \), \( Z = 0 \), or \( X \). The RPT steady state probability, \( p(Z) \), is interpreted as the probability that at any randomly selected point in time after equilibrium conditions have been attained, the state description is \( Z \). The state description for the randomly selected point in time, \( Z \), is \( 0 \) if at the randomly selected point in time the system is completely empty. The state description for the randomly selected point in time, \( Z \), is \( X = (x_0, x_1, \ldots, x_k) \) if at the
randomly selected point in time a class \( x_0 \) customer is receiving service and there are exactly \( x_i \), \( i = 1, 2, \ldots, k \), class \( i \) customers awaiting service.

Converting the SMP equilibrium probabilities to the RPT probabilities is facilitated by using a conversion matrix, \( B' \). The conversion matrix, \( B' \), is closely related to the IMC transition matrix, \( B \), and is developed from a consideration of the non-zero elements of the IMC transition matrix. The relationship between these two matrices is valid for all of the sequencing rules considered in the study and provides an efficient method of making the final steady state probability conversion.

The same measures of system performance considered in the single class model are evaluated for the multiple class models. In addition to the overall performance measures, performance measures for each class of customer are calculated. Thus the effect of the sequencing rule on the performance of each class of customer, as well as on the overall customer population, is obtained. All the performance measures calculated are for equilibrium conditions. These measures are effective mean arrival rate, expected number of customers in the system, expected number of customers in the queue, expected waiting time in the system, expected waiting time in the queue, and expected service time.

The multiple customer class models developed in this study have many features that do not depend on the particular
sequencing rule used to determine the order of service. The sensitivity of a particular model feature is noted at the time of development. If the feature is not affected by the sequencing rule, an intuitive justification of this independence is given. If the feature is affected by the sequencing rule, the modifications needed to capture this dependence are stated and justified. As it is difficult to discuss specific modeling features without extensive notation, no attempt is made to classify the features on the basis of sequence rule dependence in this preliminary review.

Notation

The following notation in addition to that given in Chapter III is used.

\[ k = \text{total number of customer classes, } k = 1,2, \ldots \]  
\[ r_i = \text{mean arrival rate for the } i\text{th class of customers, } i = 1,2, \ldots , k \]  
\[ r = \sum_{i=1}^{k} r_i = \text{total mean arrival rate.} \]  
\[ f_s(t) = \text{service time probability density function.} \]  
Case a) \( s = ij, i = 1,2, \ldots , k, j = 1,2, \ldots , k \). The service time is for a class \( j \) customer receiving service immediately after a class \( i \) customer,

Case b) \( s = 0j, j=1,2, \ldots , k \). The service time is for a class \( j \) customer receiving service immediately after the facility has been idle for a non-zero period of time.
\[ X = (x_0, x_1, \ldots, x_k) = \text{state of the system where a} \]
\[ \text{class } x_0, x_0 = 1,2,\ldots,k, \text{ customer has just received} \]
\[ \text{service (a service completion epoch), and there are} \]
\[ \text{exactly } x_i, i = 1,2,\ldots,k, \text{ class } i \text{ customers} \]
\[ \text{awaiting service.} \]
\[ X' = (x'_0, x'_1, \ldots, x'_k) = \text{state of the system at an arbitrary} \]
\[ \text{service completion epoch referred to as the initial epoch.} \]
\[ X'' = (x''_0, x''_1, \ldots, x''_k) = \text{state of the system at the} \]
\[ \text{service completion epoch immediately following the arbitrary initial epoch.} \]
\[ \text{This epoch is referred to as the adjacent or final epoch.} \]
\[ \overline{x}' = \sum_{i=1}^{k} x'_i = \text{total number of customers awaiting service} \]
\[ \text{at the initial epoch.} \]
\[ \overline{x}'' = \sum_{i=1}^{k} x''_i = \text{total number of customers awaiting service} \]
\[ \text{at the adjacent or final epoch.} \]

Arrival Mechanism

Before discussing the sequence of steps followed in the development of steady state probabilities it is convenient to derive expressions for the joint probability of exactly \( a_i \)
\[ i = 1,2,\ldots,k, \text{ class } i \text{ customers entering the system under various conditions.} \]
\[ \text{These expressions are the basic elements of the IMC transition matrix, } B, \text{ as well as the conversion matrix, } B'. \]
\[ \text{The derivations are given as a preliminary step} \]
to avoid an unnecessarily lengthy deviation from model formulation at a later point.

The notation utilized for the probability expressions is a slightly modified version of that used for the one customer class model. As in the single customer class model, the symbol $A$ indicates an arrival probability, a bar notation, $\bar{A}$, indicates that the arrival time was truncated. Also, a single prime notation, $A'$, indicates that the random interruption portion of the service time instead of the entire service time is allowed for the arrivals to take place. Reference to Table 1 of Chapter III substantiates the uniformity of notation.

The arrival phenomena of interest concerns not the probability of exactly a customers arriving, but rather the probability of exactly $a_i$ customers of class $i$, $i = 1, 2, \ldots, k$, arriving. To concisely express this quantity the set notation, $\{a_i\}$, is adopted, where $\{a_i\} = a_1, a_2, \ldots, a_k$, indicates the $k$ components of interest. As usual the variable $s$ indicates the service time density function, where

$$s = ij, \quad i = 0, 1, \ldots, k$$

$$j = 1, 2, \ldots, k$$

As a reminder a system is said to be truncated if the total number of customers awaiting service is $N$, the upper limit on waiting space. It is emphasized that when a total of $N$ customers await service, all mean arrival rates are zero. Also, as defined in Chapter III, the random
interruption time is the time between the initiation of service and some randomly selected point in time before the conclusion of service. If the service time, \( t \), follows probability density function \( f_g(t) \), then the associated random interruption time, \( t' \), follows the probability density function \( f'_s(t') \), which is given by Fox (1967, p. 15) as

\[
f'_s(t') = \frac{(1 - F_s(t'))}{m_s}, \quad t > 0
\]

where

\[
m_s = \int_0^\infty t f_s(t) \, dt, \tag{58}
\]

and

\[
F_s(t') = \int_0^{t'} f_s(t) \, dt.
\]

Using this notation, the four arrival probabilities of interest are

\[
A([a_i]; s) = \text{marginal probability of exactly } a_i \text{ class } i \text{ arrivals, } i = 1, 2, \ldots, k, \text{ during a service time given that the system does not truncate and the service time density function is } f_g(t), s = i j, i = 0, 1, \ldots, k, j = 1, 2, \ldots, k,
\]

\[
\bar{A}([a_i]; s) = \text{marginal probability of exactly } a_i \text{ class } i \text{ arrivals, } i = 1, 2, \ldots, k, \text{ during a service time given that the system does truncate and the service time density function is } f_g(t), s = i j, i = 0, 1, \ldots, k, j = 1, 2, \ldots, k,
\]

\[
A'([a_i]; s) = \text{marginal probability of exactly } a_i \text{ class } i \text{ arrivals, } i = 1, 2, \ldots, k, \text{ during a random
interruption time given that the system does not truncate and the random interruption time density function is \( f_s'(t') \), \( s = ij, \)
\( i = 0, 1, \ldots, k, \ j = 1, 2, \ldots, k. \)

and

\[ \overline{A}'([a_i]; s) = \text{marginal probability of exactly } a_i \text{ class } i \]
arrivals, \( i = 1, 2, \ldots, k, \) during a random interruption time given that the system does truncate and the random interruption time density function is \( f_s'(t') \), \( s = ij, \)
\( i = 0, 1, \ldots, k, \ j = 1, 2, \ldots, k. \)

The probabilities defined are all directly related to the equivalent probabilities for the single customer class system, as subsequent derivations prove. It is seen that, although the multiple class arrival mechanism is orders of magnitude more complex than the single class arrival mechanism, simple expressions involving the derivatives of the Laplace transforms of the service time distributions are again used to avoid difficult integration. Each of the probability expressions is derived from basic probability theory considerations. The initial expressions are rather complex, involving integrals of undefined functions. The mathematical manipulations required to reduce the initial equations to a more reasonable form are given in Appendix A. The end result of the manipulations are interpreted within the body of the chapter. The service time arrival
probabilities are considered first, then the random interruption time arrival probabilities are addressed.

The expression for \( A([a_i]; s) \) involves the product of \( k \) Poisson probability expressions. The general Poisson expression for exactly \( a_i \) class \( i \) arrivals with a mean arrival rate, \( r_i \), and a given service time, \( t \), is

\[
p(a_i; r_i t), \quad a_i = 0,1,2,\ldots, i = 1,2,\ldots,k.
\]  

(59)

Define the total number of arrivals to be the quantity \( a \), where

\[
a = \sum_{i=1}^{k} a_i.
\]  

(60)

As the arrivals are independent, the probability of the set of arrivals, \([a_i]\), given a service time, \( t \), is

\[
\prod_{i=1}^{k} p(a_i; r_i t), \quad a = 0,1,\ldots
\]  

(61)

The joint probability of the set of arrivals, \([a_i]\), and a service time between \( t \) and \( t + dt \), is

\[
\int_{t}^{t+dt} \prod_{i=1}^{k} p(a_i; r_i t) f_s(t) dt, \quad a = 0,1,\ldots
\]  

(62)

The marginal probability of the set of arrivals, \([a_i]\), is the probability of interest and is yielded by integrating over the range of \( t \) as

\[
A([a_i]; s) = \int_{0}^{\infty} \prod_{i=1}^{k} p(a_i; r_i t) f_s(t) dt, \quad a = 0,1,\ldots
\]  

(63)
By Result A.6 of Appendix A this expression reduces to

$$A([a_i];s) = \left[ \left( \frac{a_i^0}{a_1^0! \cdots a_k^0!} \right) \frac{r_i}{r} \right]^k A(a_i;s), \quad a = 0,1,\ldots$$

where

$$\left( \frac{a_i^0}{a_1^0! \cdots a_k^0!} \right) = \frac{a_i^0}{a_1^0 \cdot a_2^0 \cdots a_k^0}$$

and

$$A(a_i;s) = \frac{(-r)^a}{a!} f_s(a)(r) ,$$

where

$$f_s(a)(r) = \frac{d^a}{dz^a} \left( \mathcal{L}(f_s(t)) \right) \bigg|_{z=r}$$

The interpretation of this expression is straightforward in terms of probability theory. The expression

$$\left( \frac{a_i^0}{a_1^0! \cdots a_k^0!} \right) \frac{r_i}{r} \quad (65)$$

is the conditional multinomial probability of having exactly $a_i^0$ arrivals of class $i$ out of a given total of $a$ arrivals.

The probability that any arrival is a class $i$ customer is $r_i/r$. The expression, $A(a_i;s)$, is the probability of a total of exactly $a$ arrivals. Thus, the product of these two terms yields the probability of the event defined for $A([a_i];s)$.

In addition to the interpretation, it should be noted that possibly troublesome integration is avoided by expressing the probability in terms of Laplace transforms.

The expression for $\overline{A}([a_i];s)$ considers a truncated system. Recall that a system truncates when the maximum number of customers, $N$, await service. The effect of truncation is to cause the mean arrival rates of all classes
of customers to become zero until the total number of customers in the queue is again less than \( N \). Consider the set of arrivals, \( \{a_i\} \), which caused truncation. The total number of arrivals is the quantity, \( a \), and clearly the system truncated at the instant the \( a \)th customer entered the system. Designate the time of truncation as time \( v \), \( v \leq t \), where \( t \) is the length of the service time. Given that at least one class \( j \) customer arrived, \( a_j > 0 \), it is possible that the last class \( j \) customer to arrive caused truncation. In other words, the \( a \)th customer to arrive was a class \( j \) customer. The probability that the system truncated with the set of arrivals \( \{a_i\} \), given that the last customer is of class \( j \), is derived from basic probability considerations.

If time \( v \), \( v \leq t \), is the time of truncation, then all of the non-class \( j \) customers arrived before time \( v \). The probability of this event is given by the product of \((k-1)\) Poisson probabilities as

\[
\prod_{i=1, i \neq j}^{k} p(a_i; r_i, v), \quad a_i = 0, 1, 2, \ldots \quad (66)
\]

The probability that \( a_j - 1 \) class \( j \) customers arrived before time \( v \), and the last class \( j \) customer arrived between time \( v \) and \( v + dv \) is

\[
p(a_j - 1; r_j, v) r_j \, dv, \quad a_j = 1, 2, \ldots \quad (67)
\]

The joint probability that the set of customers \( \{a_i\} \) arrived before truncation time \( v \) and the last customer to arrive was of class \( j \) is
Defining the Kronecker delta, the probability is rewritten

\[
\sum_{i=1}^{k} \prod_{i \neq j} p(a_i - d_{ij}, r_i v) r_j dv, \quad a_i = 0, 1, 2, \ldots, i \neq j
\]

where

\[
d_{ij} = \begin{cases} 1, & i = j \\ 0, & \text{otherwise} \end{cases}
\]

Integrating the above probability over the range of the truncation time, \( v \), yields the marginal probability that the last customer to arrive was of class \( j \), and the set of customers \( \{a_i\} \) arrived. This probability is

\[
\int_0^t \sum_{i=1}^{k} \prod_{i \neq j} p(a_i - d_{ij}, r_i v) r_j dv, \quad a_i = 0, 1, 2, \ldots, i \neq j
\]

It is clear that the set of customers \( \{a_i\} \) can arrive before truncation if any of the classes of customers which had a non-zero number of arrivals, \( a_i > 0, i = 1, 2, \ldots, k \), was the last class of customer to arrive. To incorporate this fact the set \( J \) is defined as

\[
J = \{i; a_i > 0, i=1,2,\ldots,k\}.
\]

By summing the probability expression over the subscript \( j \), \( j \in J \), the probabilities of the mutually exclusive, collectively exhaustive events of interest are summed. Performing this operation, the probability of the set of customers \( \{a_i\} \) arriving before a service time \( t \) is
The joint probability that the set of customers \( \{a_i\} \) arrives and the service time is between time, \( t \), and \( t + dt \), is

\[
\sum_{j \in J} \int_0^t \left[ \prod_{i=1}^k p(a_i - d_{ij}; r_i v) \right] r_j dv, \quad a_i = 0,1,2,\ldots \tag{72}
\]

Integrating over the range of \( t \) yields the desired marginal probability of interest which is

\[
\bar{A}(\{a_j\}; s) = \int_0^\infty \sum_{j \in J} \int_0^t \left[ \prod_{i=1}^k p(a_i - d_{ij}; r_i v) \right] r_j dv f_s(t) dt
\]

\[a = 1,2,\ldots \tag{73}\]

Note that the quantity \( a \) is required to be non-zero as the set \( J \) is empty when \( a = 0 \), i.e., if \( a_i = 0 \), \( i = 1,2,\ldots,k \), then the set \( J \) is an empty set. By Result A.7, the above expression reduces to

\[
\bar{A}(\{a_j\}; s) = \left[ \sum_{j \in J} \left[ \prod_{i=1}^k \frac{r_i^{a_i}}{a_i!} \right] \bar{A}(a_i; s) \right], \quad a = 1,2,\ldots \tag{75}
\]

where

\[
\left[ \prod_{i=1}^k \frac{r_i^{a_i}}{a_i!} \right] \quad a_i \in \{a_i \}\quad a = 1,2,\ldots
\]

and

\[
\bar{A}(a_i; s) = 1 - \sum_{n=0}^{a_i-1} \frac{(-r)^n}{n!} \frac{\bar{F}(n)}{F_s(r)}, \quad a = 1,2,\ldots
\]

The interpretation of this expression is straightforward in terms of probability theory. The expression
\[ \left( \frac{a}{a_i - d_{ij}} \right)^k \prod_{i=1}^{k} \left( \frac{r_i}{r} \right)^{a_i}, \quad a = 1, 2, \ldots \quad (76) \]

is the conditional negative multinomial probability of having exactly \((a_i - d_{ij})\) arrivals of class \(i\), \(i = 1, 2, \ldots, k\), out of a given total of \(a\) arrivals enter the system before the \(a_j\)th class \(j\) customer arrives. As before, \(r_i/r\) is the probability that any arrival is a class \(i\) customer. Summing these negative multinomial probabilities over all feasible values of \(j\) gives the conditional probability of the set of customers \([a_i]\) arriving, given a total of \(a\) arrivals. The expression \(\overline{A}(a_i; s)\) is the probability of a total of exactly \(a\) arrivals. Therefore, the product of these two quantities yields the probability of the event defined for \(\overline{A}(\{a_i\}; s)\). The computational advantage of the simplified expression is of considerable importance. A comparison of the initial expression and the expression yielded by the manipulations of Result A.7 indicates the importance of the Laplace transform identity.

The random interruption time arrival probability expressions, \(A'([a_i]; s)\) and \(\overline{A}'([a_i]; s)\), are derived from a comparison with the analogous service time probabilities, \(A([a_i]; s)\) and \(\overline{A}([a_i]; s)\), respectively. A review of the definitions of these quantities indicates only one point of difference.

This point is that the probability density function \(f_{s}'(t')\) controls the time allowed for the arrivals during the random interruption time, \(t'\), and the probability density function \(f_{s}(t)\) controls the time allowed for the arrivals during the
service time, \( t \). Replacing the density function \( f_s(t) \) with the density function, \( f'_s(t') \), and modifying the associated variable of integration, provides expressions for the random interruption time arrival probabilities in terms of the analogous service time arrival probabilities. This plan is followed in order to avoid repetition of the logical arguments previously given.

Making the indicated revisions in the expression for \( A([a_i]; s) \) yields the expression for \( A'(a_i); s) \) as

\[
A'(a_i; s) = \int_0^\infty \prod_{i=1}^{k} p(a_i; r_i t') f_s(t') dt', \quad a = 0, 1, \ldots
\]

By Result A.8 of Appendix A this expression reduces to

\[
A'(a_i; s) = \left( \left( \prod_{i=1}^{a} \frac{r_i}{r} \right)^{a_i} \right) A'(a; s), \quad a = 0, 1, \ldots
\]

where

\[
\prod_{i=1}^{a} a_i = \frac{a!}{a_1! a_2! \ldots a_k!}
\]

and

\[
A'(a; s) = \frac{1}{m_s} \left[ 1 - \sum_{n=0}^{a} \frac{(-r)^n}{n!} \frac{r^n}{s^n} \right], \quad a = 0, 1, \ldots
\]

The interpretation of this simplified expression is identical to that of the simplified service time arrival probability expression. The only change is the duration of time allowed for the arrival phenomenon to operate.

The Laplace transform quantities are again utilized to avoid possibly troublesome integration. The computational savings are emphasized if the Poisson expressions,
p_i(a; r_i t), and the expression for the random interruption
time density function, f_s'(t'), are written in expanded form. This unwieldy expanded equation is

\[ A'(\{a_i\}; s) = \int_0^\infty \left[ \sum_{i=1}^k \frac{(r_i t)^{a_i} e^{-r_i t}}{a_i!} \right] \frac{1}{m_s} \left[ 1 - \int_0^{t'} f_s(t) dt \right] dt', \]

\[ a = 0, 1, \ldots \quad (79) \]

The final arrival probability expression, \( \bar{A}'(\{a_i\}; s) \), is now found. Making the previously indicated revisions in the expression for \( \bar{A}(\{a_i\}; s) \) yields the expression for \( \bar{A}'(\{a_i\}; s) \) as

\[ \bar{A}'(\{a_i\}; s) = \int_0^\infty \sum_{j \in J} \int_0^{t'} \left[ \sum_{i=1}^k \left( \frac{r_i}{r} \right)^{a_i} \right] \frac{1}{r} f_s'(t') dt', \]

\[ a = 1, 2, \ldots \quad (80) \]

By Result A.9 of Appendix A this expression reduces to

\[ \bar{A}'(\{a_i\}; s) = \sum_{j \in J} \left[ \sum_{a_i-d_{ij}}^{a} \left( \frac{r_i}{r} \right)^{a_i} \right] \bar{A}'(a; s), a = 1, 2, \ldots \quad (81) \]

where

\[ \left( [a_i - d_{ij}] \right) = a_1! : a_2! : \ldots : a_j! : (a_j-1)! : a_{j+1}! : \ldots : a_k! : j \in J \]

and

\[ \bar{A}'(a; s) = 1 - \frac{1}{m_s} \left[ a - \sum_{n=0}^{a-1} \sum_{h=0}^{(n-r)h} \frac{(-r)^h f_s(h)}{h!} \right], a = 1, 2, \ldots \]

The comment regarding the revised interpretation of the nontruncated random interruption time arrival probability is also applicable for this expression.

The computational advantages of the simplified expression are apparent if the basic expression is written in
expanded form. The expanded expression is
\[ A'([a_i], s) = \int_0^\infty \sum_{j \in J} \int_0^{t'} \left[ \sum_{i=1}^{k} \frac{(r_i v)^a_i - d_{ij} e^{-r_i v}}{(a_i - d_{ij})!} \right] r_j dv \frac{1}{m_r} \]
\[ x[1 - \int_0^{t'} f_s(t) dt] dt', \ a=1,2,\ldots \] (82)

It is indeed fortunate that this bewildering maze of integrals reduces to a relatively simple expression in terms of the Laplace transforms of the service time distributions.

The simplified expressions for all of the arrival probabilities are restated in Table 5. The equivalent expressions for the single customer class model are given in Table 1 of Chapter III. Comparing these expressions produces an expected result; the multiple customer class expressions reduce to the single customer class expressions for the special case of \( k = 1 \). An additional point should be noted. These arrival probabilities have been developed independent of any sequencing rule assumptions, and as a result, they are appropriate components for all of the multiple customer class models.

**IMC State Descriptions**

The states of the IMC are defined in terms of the number of each class customer awaiting service and the class of customer just completing service. Recall that the IMC considers the state of the system only at selected instants of time called service completion epochs. The state of the system at the service completion epoch is \( X, X = (x_0, x_1, \ldots) \).
### Probability of the Exact Set of Arrivals \( \{a_i\} \)

\[ \{a_i\} = a_1, a_2, \ldots, a_k \]

<table>
<thead>
<tr>
<th>Conditions</th>
<th>( A([a_i];s) )</th>
<th>( \bar{A}([a_i];s) )</th>
<th>( A'({a_i};s) )</th>
<th>( \bar{A}'({a_i};s) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>time allowed for arrivals</td>
<td>( A([a_i];s) = \left[ \left( \sum_{j=1}^{a} \left( \sum_{i=1}^{a} \left( \frac{r_i}{r} \right)^{a_i} \right) \right] A(a;s), \right. )</td>
<td>( \bar{A}([a_i];s) = \left[ \sum_{j \in J} \left( \sum_{i=1}^{a} \left( \frac{r_i}{r} \right)^{a_i} \right) \left( \frac{1}{r} \right)^{a_i} \bar{A}(a;s), \right. )</td>
<td>( A'({a_i};s) = \left[ \left( \sum_{j=1}^{a} \left( \sum_{i=1}^{a} \left( \frac{r_i}{r} \right)^{a_i} \right) \right] A'(a;s), \right. )</td>
<td>( \bar{A}'({a_i};s) = \left[ \sum_{j \in J} \left( \sum_{i=1}^{a} \left( \frac{r_i}{r} \right)^{a_i} \right) \left( \frac{1}{r} \right)^{a_i} \bar{A}'(a;s), \right. )</td>
</tr>
<tr>
<td>service time ( a = 0, 1, \ldots )</td>
<td>service time ( a = 0, 1, \ldots )</td>
<td>random interruption portion of service time ( a = 0, 1, \ldots )</td>
<td>random interruption portion of service time ( a = 0, 1, \ldots )</td>
<td></td>
</tr>
<tr>
<td>no truncation</td>
<td>no truncation</td>
<td>no truncation</td>
<td>no truncation</td>
<td></td>
</tr>
</tbody>
</table>

\( A(a;s), \bar{A}(a;s), A'(a;s), \) and \( \bar{A}'(a;s) \) are defined in Table 1.

Table 5. Arrival Probabilities for the Multiple Customer Class Models
... $x_k$). The element $x_0$ is the class of customer just completing service, and the element $x_i$, $i = 1, 2, \ldots, k$, is the number of class $i$ customers awaiting service. From a consideration of the probable changes in state between adjacent epochs the IMC transition matrix, $B$, is obtained.

It is of interest to determine the total number of states needed to describe a particular multiple customer class model. The two pertinent parameters are the maximum number of customers allowed to wait, $N$, and the number of distinct classes of customers, $k$. This model is referred to as the general $(k, N)$ model. From Chapter III it is known that for a single customer class model, $k = 1$, with a truncation point of $N$ customers, a total of $N + 1$ states are needed for the IMC. In terms of the previous notation the $(1, N)$ model requires a total of $N + 1$ states to describe the IMC. The total number of states needed to describe the IMC for the general $(k, N)$ model is now determined by the use of some results from the study of classical occupancy problems as discussed by Parzen (1960).

To assist the discussion, the notation,

$$
\overline{x} = \sum_{i=1}^{k} x_i ,
$$

is adopted. Verbally, $\overline{x}$ is the total number of customers of all classes awaiting service at a service completion epoch. For the general $(k, N)$ model, $\overline{x}$ takes on all the integer values zero through $N$, $\overline{x} = 0, 1, 2, \ldots, N$. The approach used to
develop the total number of states required to describe the IMC is to first develop an expression for the number of states required to describe the general \((k, N)\) model for an arbitrary value of \(\bar{x}\). Summing this expression for the values of \(\bar{x}\), zero through \(N\), yields the total number of states required.

The number of states needed to describe the general \((k, N)\) model for a specified value of \(\bar{x}\) is determined by considering the number of ways a total of \(\bar{x}\) customers can be distributed among \(k\) distinct customer classes. This value is given by the Bose-Einstein statistic as discussed by Parzen (1960, p. 70). The equivalent occupancy problem is to determine the number of ways \(\bar{x}\) indistinguishable balls can be distributed to \(k\) distinguishable urns when there is no limit on the number of balls that can be in any urn. The number of ways is

\[
\binom{k + \bar{x} - 1}{\bar{x}}, \quad \bar{x} = 0, 1, \ldots \quad (84)
\]

\[
k = 1, 2, \ldots
\]

which is referred to as the Bose-Einstein statistic. Therefore, if there are a total of \(\bar{x}\) customers in the system, then the number of ways the \(\bar{x}\) customers can be allocated to the \(k\) classes of customers is given by the above expression. This accounts for the various values \(x_i, i=1, 2, \ldots, k\), may take on for a given value of \(\bar{x}\), but not the value of \(x_0\). Regardless of the value of \(\bar{x}\), \(x_0\) may take on the values \(x_0 = 1, 2, \ldots, k\), or a total of \(k\) values. Therefore the number
of states, \( X = (x_0, x_1, \ldots, x_k) \), needed to describe a system that has a total of exactly \( \bar{x} \) customers awaiting service is

\[
k^* \left( \frac{k + \bar{x} - 1}{x} \right), \quad \bar{x} = 0, 1, 2, \ldots, N \quad (85)\]

For any model the number of distinct classes, \( k \), is fixed but the total number of customers awaiting service takes on the integer values zero through \( N \). The total number of states needed to describe the IMC for the general \((k, N)\) model is

\[
\sum_{x=0}^{N} k^* \left( \frac{k + \bar{x} - 1}{x} \right), \quad N = 1, 2, \ldots \quad (86)
\]

This summation is expressed in closed form by using an identity from Jolley (1961, p. 36).

\[
\sum_{x=0}^{N} k^* \left( \frac{k + \bar{x} - 1}{x} \right) = k^* \left( \frac{k + N}{N} \right), \quad N = 1, 2, \ldots \quad (87)
\]

The validity of the expression is verified for the special case of one class of customers, \( k = 1 \), as the total number of states is

\[
1^* \left( \frac{1 + N}{N} \right) = N + 1, \quad N = 1, 2, \ldots,
\]

as previously stated.

Thus the IMC transition matrix, \( B \), is a square matrix of dimension \([k^*(N + k)] \times [k^*(N + k)]\). The entries of the IMC transition matrix are determined by a subsequently developed procedure and are evaluated using the service time arrival
probability expressions. The above expression is also useful in the description of a matrix closely associated with the IMC transition matrix, $B'$. Subsequent development shows that the conversion matrix, $B'$, is related to the IMC transition matrix, $B$. The $B'$ matrix consists of the $B$ matrix with the first $k$ rows removed, and the service time arrival probabilities replaced with the equivalent random interruption time arrival probabilities. The $B'$ matrix is a non-square matrix of dimension

$$[k(N + k) - k] \times [k(N + k)].$$

One additional point should be noted. The number of states required to describe the IMC transition matrix, $B$, and the conversion matrix, $B'$, have been determined independent of any sequencing rule. As a result these conclusions are appropriate for all of the multiple customer class models. The remaining sections of this chapter are concerned with a specific model for the non-preemptive priority sequencing rule. Careful attention is given to the details of the formulation as the same techniques are utilized in the sequencing models of Chapter V.

Non-Preemptive Priority Sequencing Rule

The IMC transition matrix gives the probability of making the transition from state $X' = (x'_0, x'_1, \ldots, x'_k)$ at a given service completion epoch to state $X'' = (x''_0, x''_1, \ldots, x''_k)$ at the immediately following service completion epoch. The
transition probability is controlled by a number of factors, including the sequencing rule. The non-preemptive priority sequencing rule considered in this section assigns indices to customer classes, with the customer class with the lowest index, class 1 having the highest priority, class 2 the next highest priority, etc. The non-preemptive feature means that once a customer has initiated service, his service time is not interrupted even if a higher priority customer should arrive during his service time. Within each customer class, service is on a first come-first served basis. Therefore for a given initial state, \( X' \), the class of customer that is selected to receive service, say class \( w \), is the class of customer that has a non-zero number of customers waiting and has the lowest class index. Stating this procedure in equation form,

\[
  w = \min \left( i; x_i > 0, i = 1, 2, \ldots, k \right). \tag{89}
\]

This procedure is appropriate when the total number of customers initially in the queue is non-zero, i.e., when \( \bar{x}' > 0 \), where

\[
  \bar{x}' = \sum_{i=1}^{k} x_i'. \tag{90}
\]

This procedure is not appropriate when the initial state is empty, i.e., when \( \bar{x}' = 0 \). In this case the first customer to enter the system during the idle period immediately initiates service regardless of his priority class. The probability that a class i customer is the first to arrive is \( r_i/r \).
i = 1, 2, ..., k. Thus when the system is initially empty, the class of customer that is selected to receive service, say class w, is a class i customer with probability \( r_i / r \), \( i = 1, 2, ..., k \). Summarizing this non-preemptive priority sequencing rule in equation form:

\[
    w = \begin{cases} 
        \min \left( i; x_i^t > 0, i = 1, 2, ..., k \right), & x' > 0 \\
        i, \text{ with probability } r_i / r, & x' = 0,
    \end{cases}
\]

where

\[
    x' = \sum_{i=1}^{k} x_i^t.
\]

It should be noted that (1) \( w \) is a function of the initial state \( X' \), (2) \( w \) is not a function of the final state, \( X'' \), (3) \( w \) is not a function of the class of customer that received the immediately preceding service, class \( x_0^t \), and (4) \( w \) is known with certainty when the initial state is non-empty but is known probabilistically when the initial state is empty. The notation, \( x' \), is interpreted as the total number of customers awaiting service at the initial epoch and is defined to conveniently express whether the initial state is empty or non-empty.

**IMC Transition Probabilities**

The IMC transition probabilities may be written as \( \text{pr}(X''|X') \). This is the probability of having state description \( X'' \) at a service completion epoch, given that at the immediately preceding service completion epoch the state description was \( X' \). As in the one customer class model
these probabilities are given in terms of the service time arrival probabilities, $A([a_i];s)$ and $\bar{A}([a_i];s)$, and are the elements of the IMC transition matrix.

The determination of the specific entries for the IMC transition matrix, $B$, is based on a number of factors. The service time arrival probability notation provides a key to these factors. One factor is the set of arrivals, $[a_i]$, required to cause the indicated transition. A second factor is the density function that controls the duration of the service time occurring between adjacent epochs. This density function is indicated by the quantity $s = ij$. A third factor is whether or not the system truncates as indicated by the presence or absence of the bar notation. In addition to the factors suggested by the notation, it is of overriding importance to determine if a transition is feasible. All infeasible transitions are assigned a zero probability of occurrence. Needless to say, the first three factors are considered only if the indicated transition is feasible. For this reason the question of feasibility is discussed first and the remaining factors addressed in turn. It should be noted that there is considerable interaction between these factors, but they are presented independently in the interest of clarifying the discussion. The net effect of all the factors is treated in subsequent sections. The general one step transition from state $X'$ to the adjacent state $X''$ is considered.
A transition can be infeasible for two reasons: (1) the sequencing rule is violated or (2) more than one customer leaves the system between adjacent epochs. The sequencing rule indicates that for a given initial state a specified class of customer receives service. Thus, the class of customer completing service at the adjacent epoch, class $x''_0$, must be of class $w$. As the sequencing rules are followed without fail, any variance of the sequencing rule is assumed to be impossible and a zero probability is assigned to the indicated transition. If at the initial epoch the system is non-empty, $x' > 0$, the non-preemptive priority rule designates a class $w$ to receive service with certainty. If at the initial epoch the system is empty, $x' = 0$, the non-preemptive priority rule specifies that any class of customer can receive service. This feasibility constraint is stated as

$$x''_0 = \begin{cases} w, & x' > 0, \\ 1, 2, \ldots, k, & x' = 0, \end{cases} \quad (92)$$

where $w$ is specified by the non-preemptive priority sequencing rule.

Between adjacent epochs only one customer leaves the system, namely the customer receiving service, which is of class $x''_0$. When the initial epoch is non-empty, $x' > 0$, the class $x''_0$ customer being served is taken from the class $x''_0$ customers awaiting service at the initial epoch. Thus the minimum feasible number of class $i$ customers awaiting service at the adjacent epoch is given by
\[ x_i'' \geq \begin{cases} x_i' & i = 1,2,\ldots,k, i \neq x_0''; \\ x_i' - 1 & i = x_0'' \end{cases} \]

when \( \bar{x}' > 0 \).

When the initial epoch is empty, \( \bar{x}' = 0 \), all of the customers that are awaiting service at the adjacent epoch arrived during the service time of the first customer to enter the system between epochs. Thus the minimum feasible number of class \( i \) customers awaiting service at the adjacent epoch is given by

\[ x_0'' > 0, \quad i = 1,2,\ldots,k \]

when \( \bar{x}' = 0 \).

Restating these feasibility constraints using the previously defined Kronecker delta symbol, \( \delta_{ij} \), as

\[ \{x_i'' - x_i' + \delta_{ix_0''}\} \geq 0, \quad \bar{x}' > 0, \]

and

\[ \{x_i''\} \geq 0, \quad \bar{x}' = 0, \]

where the notation, \( \{a_i\} \geq 0 \), indicates that all \( k \) elements of the set are non-zero. The remaining factors are presented with the assumption the indicated transitions are feasible.

The set of arrivals, \( \{a_i\} \), required to cause the indicated transition depends on the number of each class of customer in the initial state, \( x_i' \), \( i = 1,2,\ldots,k \), and the number of each class of customer in the adjacent state, \( x_i'' \), \( i = 1,2,\ldots,k \). In addition, if the initial state is non-empty, \( \bar{x}' > 0 \), the class of the customer completing service
at the adjacent epoch, class $x''_0$, is pertinent. This served customer is taken from the class $x''_0$, customers awaiting service at the initial epoch. The number of class $i$ arrivals required is

$$a_i = \begin{cases} x_i' - x_i, & i = 1, 2, \ldots, k, \ i \neq x''_0 \\ x''_i - x_i' + 1, & i = x''_0, \end{cases}$$

when

$$\bar{x}' > 0.$$ (96)

If the initial state is empty, $\bar{x}' = 0$, the class of the served customer is immaterial, as all of the customers awaiting service at the adjacent epoch must have arrived during the intervening service time. The number of class $i$ arrivals needed is

$$a_i = x_i' , \quad i = 1, 2, \ldots, k,$$

when

$$\bar{x}' = 0.$$ (97)

Restating these conclusions in terms of the set of arrivals required to cause the indicated transition, yields

$$\{a_i\} = \begin{cases} [x_i'' - x_i' + d_i x_0''] \ , & \bar{x}' > 0 \\ [x_i'] \ , & \bar{x}' = 0. \end{cases}$$ (98)

The two remaining factors are presented with the assumption that the indicated transition is feasible, and the set of arrivals required to cause the transition is as given above.

The density function that controls the duration of the service time occurring between adjacent epochs depends on the class of customer completing service at the adjacent
epoch, class $x_0'$. In addition, if the initial state is non-empty, $x'>0$, the class of customer completing service at the initial epoch, class $x_0'$, is pertinent. If the initial state is non-empty, $x'>0$, the service time is for a class $x_0'$ customer immediately preceded by a class $x_0'$ customer. The density function for the service time occurring between adjacent epochs is

$$f_s(t) = \frac{1}{x_0''},$$

when

$$x'>0.$$  \hspace{1cm} (99)

If the initial state is empty, $x'=0$, an idle period occurs before the next epoch. The service time is for a class $x_0''$ customer immediately preceded by an idle period. The density function for the service time occurring between adjacent epochs is

$$f_s(t), \quad s = 0x_0'',$$

when

$$x' = 0.$$  \hspace{1cm} (100)

Restating these conclusions for all cases, the density function for the service time occurring between adjacent epochs is

$$f_s(t) = \begin{cases} \frac{1}{x_0''}(t), & x > 0, \\ \frac{1}{0x_0''}(t), & x' = 0. \end{cases}$$  \hspace{1cm} (101)

The determination of system truncation is based on the total number of customers awaiting service at the adjacent
epoch, $\bar{x}''$. The system is truncated if the total number waiting is equal to the upper limit on waiting space, $N$. Simply stated, the system is truncated if $\bar{x}'' = N$, and is non-truncated if $\bar{x}'' < N$.

All of the previous factors can be incorporated in the transition probability conditions by considering the situations where at the initial epoch the system is empty or non-empty, and situations where at the adjacent epoch the system is truncated or non-truncated. There is a total of four combinations of $\bar{x}'$ and $\bar{x}''$ if the stated conditions on $\bar{x}'$ and $\bar{x}''$ are considered. Thus there is a total of four mutually exclusive, collectively exhaustive cases for the feasible transitions. A fifth case is added to assign zero probabilities to infeasible transitions. The transition probability, $pr(X'|X'')$, for all four feasible cases is given, and the influencing factors referenced.

For the case where the initial state is non-empty and the adjacent state is non-truncated, the transition probability is

Case A) $\bar{x}' > 0$, $\bar{x}'' < N$

$$pr(X''|X') = A([a_i]; s),$$

where

$$[a_i] = [x_i'' - x_i' + d_{iw}],$$

$$s = x_0', \quad w,$$

$$x_0'' = w,$$

and

factor

arrival set

service time

sequence rule

feasibility
Verbally, the probability that the system makes a transition from an initial service completion epoch with state description \( X' \) where at least one customer awaits service and a class \( w \) customer initiates service, to an adjacent service completion epoch with state description \( X'' \) where less than \( N \) customers await service, is equivalent to the given arrival probability. The arrival probability is for the event that the set of customers \( \{x''_i - x'_i + d_{iw}\} \) arrive during a service time that follows the probability density function \( f_{x''_0w}(t) \), and the mean arrival rates are non-zero throughout the service time.

For the case where the initial state is non-empty and the adjacent state is truncated, the transition probability is

\[
\text{Case B) } \bar{x}' > 0, \bar{x}'' < N
\]

\[
\text{pr}(X''|X') = A(\{a_i\}; s),
\]

where

\[
\{a_i\} = \{x''_i - x'_i + d_{iw}\},
\]

\[
s = x'_0 w,
\]

\[
x''_0 = w,
\]

and

\[
\{x''_i - x'_i + d_{iw}\} \geq 0.
\]

The verbal description of this case is identical to Case A) except at the adjacent epoch there are exactly \( N \) customers.
awaiting service. The equivalent arrival probability is for the same event described in Case A) with one exception. The exception is that immediately after the arrival of \((N - x' + 1)\) customers during the service time, all mean arrival rates are zero.

For the case where the initial state is empty and the adjacent state is non-truncated, the transition probability is

\[
\text{Case C)} \quad x' = 0, \quad x'' < N \\
pr(X''|X') = A([a_i], s) \frac{r_{x''}}{r},
\]

where

\[
[a_i] = \{x_i''\}, \quad \text{arrival set} \\
s = 0x''_0, \quad \text{service time} \\
x''_0 = 1, 2, \ldots, k, \quad \text{sequence rule} \\
\text{feasibility}
\]

and

\[
\{ x_i'' \} \geq 0. \quad \text{arrivals} \quad \text{feasibility}
\]

The element \((r_{x''_0}/r)\) is the probability that the first customer to arrive during the idle period is a class \(x''_0\) customer. Verbally, the probability that the system makes a transition from an initial service completion epoch with state description \(X'\) where no customers await service, to an adjacent service completion epoch with state description \(X''\), where less than \(N\) customers await service and a class \(x''_0\) has just completed service, is equivalent to the given arrival probability. The arrival probability is for the joint event that
a class \( x_0 \) customer is the first to arrive and the additional set of customers, \( \{x_i\} \), arrive during his service time. The service time follows the probability density function \( f_{0x_0}(t) \), and the mean arrival rates are non-zero throughout the service time.

For the final feasible case where the initial state is empty and the adjacent state is truncated, the transition probability is

**Case D)** \( x' = 0, x'' = N \)

\[
pr(X''|X') = A(\{a_i\}; \Delta) \frac{r_{x''}}{r}
\]

where

\[
\{a_i\} = \{x_i''\}, \quad \text{arrival set}
\]

\[
\Delta = 0x_0', \quad \text{service time}
\]

\[
x''_0 = 1, 2, \ldots, k, \quad \text{sequence rule}
\]

and

\[
\{x_i''\} \geq 0, \quad \text{arrival feasibility}
\]

The verbal description of this case is identical to that of Case C) except that at the adjacent epoch there are exactly \( N \) customers awaiting service. The equivalent arrival probability is for the same event described in Case C) with one exception. The exception is that immediately after the arrival of \( N \) customers during the service time, all mean arrival rates are zero.

All of the previously defined cases are restated in Table 6. It is emphasized that any transition is defined by
A Case Transition Probability Conditions
\[
\text{pr}(X'' | X') = A(\{x''_i - x'_i + d_{iw}\}, x'_0, w)
\]
- \(x' > 0,\)
- \(x'' < N,\)
- \(x''_0 = w,\)
- \(\{x''_i - x'_i + d_{iw}\} \geq 0.\)

B Case Transition Probability Conditions
\[
\text{pr}(X'' | X') = A(\{x''_i - x'_i + d_{iw}\}, x'_0, w)
\]
- \(x' > 0,\)
- \(x'' = N,\)
- \(x''_0 = w,\)
- \(\{x''_i - x'_i + d_{iw}\} \geq 0.\)

C Case Transition Probability Conditions
\[
\text{pr}(X'' | X') = A(\{x''_i\}, 0, x''_0) \frac{r_{x''}}{r}
\]
- \(x' = 0,\)
- \(x'' < N.\)

D Case Transition Probability Conditions
\[
\text{pr}(X'' | X') = A(\{x''_i\}, 0, x''_0) \frac{r_{x''}}{r}
\]
- \(x' = 0,\)
- \(x'' = N.\)

E Case Transition Probability Conditions
\[
\text{pr}(X'' | X') = \text{Zero}
\]
- Otherwise.

Table 6. Transition Probabilities for the Non-preemptive Priority Model.
one, and only one, of the five cases listed in Table 6. Using this information it is possible to assign appropriate probabilities to all the transitions for a given system. These probabilities are displayed in the IMC transition matrix, B. Before discussing the actual form of the IMC transition matrix a procedure for assigning a probability to an arbitrary transition is developed.

The procedure is based on the fact that the five cases listed in Table 6 are mutually exclusive and collectively exhaustive. This means that an arbitrary transition is defined by one, and only one, of the five cases. The first four cases describe all the feasible transitions, and the last case describes all infeasible transitions. One procedure of assigning a probability to an arbitrary transition is to check the conditions stated for each case in the order given in Table 6. If the arbitrary transition is defined by one of the first four cases, Case A) through Case D), then the transition is feasible. The non-zero probability associated with the selected case is assigned to the transition. If the arbitrary transition is not defined by one of the first four cases, then the transition is infeasible. A zero probability is assigned to this infeasible transition. The use of this procedure assures that all of the factors discussed in previous sections are considered when the transition probabilities are entered in the IMC transition matrix.
As an example of this procedure, a small system with two classes of customers, \( k = 2 \), and with a maximum of two customers allowed to wait for service, \( N = 2 \), is considered. The general transition is from state \( X' = (x'_0, x'_1, x'_2) \) to state \( X'' = (x''_0, x''_1, x''_2) \). The specific transition from state \( X' = (x'_0 = 2, x'_1 = 1, x'_2 = 1) \) to state \( X'' = (x''_0 = 1, x''_1 = 0, x''_2 = 1) \) is analyzed. This transition probability is written as

\[
pr((1, 0, 1)|(2, 1, 1)).
\]

To determine the case which defines this transition, the quantities \( \overline{x}' \) and \( \overline{x}'' \) are evaluated. These quantities are

\[
\overline{x}' = x'_1 + x'_2 = 1 + 1 = 2,
\]

and

\[
\overline{x}'' = x''_1 + x''_2 = 0 + 1 = 1,
\]

respectively. As \( \overline{x}' > 0 \) and \( \overline{x}'' < N \), the transition is a Case A transition if the additional conditions are met. The third condition for Case A requires

\[
x''_0 = w,
\]

where

\[
w = \min \left( i; x'_i > 0, i = 1, 2 \right)
\]

\[
= \min \left( i = 1, i = 2 \right)
\]

\[
= 1.
\]
For the indicated transition $x_0^{n} = 1$ as required, so the transition is feasible from the standpoint of following the sequencing rule. The fourth condition for Case A) requires

$$\{x_i^{n} - x_i^{1} + d_i^{lw}\} \geq 0.$$  

This is equivalent to requiring

$$x_i^{n} - x_i^{1} + d_{11} \geq 0,$$

and

$$x_2^{n} - x_2^{1} + d_{21} \geq 0.$$  

For the indicated transition these quantities are

$$x_1^{n} - x_1^{1} + d_{11} = 0 - 1 + 1$$

$$= 0,$$

and

$$x_2^{n} - x_2^{1} + d_{21} = 1 - 1 + 0$$

$$= 0,$$

respectively. As both quantities are non-negative, the transition is feasible from the standpoint of requiring a non-negative set of arrivals. All of the conditions that define a Case A) transition are satisfied, so the transition probability is

$$pr[(x_0^{n}, x_1^{n}, x_2^{n}) | (x_0^{1}, x_1^{1}, x_2^{1})]$$

(109)

$$= A[(x_1^{n} - x_1^{1} + d_1^{lw}), (x_2^{n} - x_2^{1} + d_2^{lw}), x_0^{w}]$$

$$= A[(0 - 1 + 1), (1 - 1 + 0), 21]$$

$$= A(0, 0, 21).$$

A verbal description of the indicated transition substantiates the results of the above procedure. Initially
the system has one class 1 customer and one class 2 customer awaiting service. A class 2 customer has just completed service. According to the sequencing rule, the top priority class 1 customer receives service next. His random service time is the time for a class 1 customer to receive service following a class 2 customer. In order to go to an adjacent state that has no class 1 customers, one class 2 customer, and a class 1 customer just completing service, it must be true that there were no arrivals during the service time. This probability of no arrivals during the indicated service time is \( A(0, 0; 21) \), the probability given by the previous procedure.

The results given in Table 6 were discussed in terms of the non-preemptive priority sequencing rule. It should be noted that these results are also appropriate for other sequencing rules. A review of the conditions given in Table 6 reveals that the only effect of the sequencing rule is to uniquely define the variable \( w \) for Case A) and Case B). This definition is given by the equation

\[
    w = \min_i (i; x_i > 0, \ i = 1, 2, \ldots, k).
\]

Thus when \( \bar{x}' > 0 \) the value of \( w \) is determined with certainty. Without loss of generality the relationship which defines \( w \) can be of any form as long as the value of \( w \) is uniquely determined for the situation where the initial state is non-empty, \( \bar{x}' > 0 \). Verbally, this means that the sequencing rule absolutely determines the class of customer to receive
service next, when there is at least one customer waiting. Many of the sequencing rules considered in Chapter V exhibit this property. Only a slight modification of the results of Table 6 is required when the sequencing rule determines the class of customer to receive service next on a probabilistic basis.

**IMC Transition Matrix**

The IMC transition matrix, \( B \), is determined by placing entries in a square matrix according to the conditions given in Table 6. The vertical axis of the matrix is labeled \( X' \), and the horizontal axis is labeled \( X'' \). The elements of the matrix are

\[
b(X', X'') = \Pr(X''|X').
\]

The particular arrangement of the state descriptions is immaterial as long as all feasible states are listed and the horizontal and vertical arrangement of the descriptions is identical. In other words, the assignment of the vector state descriptions, \( X' \) and \( X'' \), to particular rows and columns of the \( B' \) matrix, respectively, can be made arbitrarily as long as the same order of assignment is used for both the rows and columns. The convention adopted in this study is to assign state descriptions based on the values of \( \bar{X}' \) and \( \bar{X}'' \). In the following discussion of the assignment procedure the prime notation is dropped, as the assignment of state descriptions to the rows and columns is identical.

The first \( k \) state descriptions to be assigned are for
the set of empty states, \( \bar{X} = 0 \). These state descriptions are \( X = (i, 0, 0, ..., 0) \), \( i = 1,2,...,k \). Thus the first \( k \) rows of the IMC transition matrix are labeled \( X' = (i, 0, 0, ..., 0) \), \( i = 1,2,...,k \), respectively. Similarly the first \( k \) columns of the IMC transition matrix are labeled \( X'' = (i, 0, 0, ..., 0) \), \( i = 1,2,...,k \), respectively. The next state descriptions to be assigned are for the set of states that describe a system with a total of one customer awaiting service at a service completion epoch, \( \bar{X} = 1 \). From the Bose-Einstein statistic previously cited, there are a total of

\[
\frac{k(k + \bar{X} - 1)}{\bar{X}} = k(1 + \frac{1}{1} - 1)
\]

\( = k^2 \)

states with \( x = 1 \). These \( k^2 \) state descriptions are assigned to the next \( k^2 \) rows and columns of the IMC transition matrix. The particular ordering is arbitrary. All feasible states with \( \bar{X} = 1 \) are listed once, and the ordering of the state descriptions is identical for the rows and columns. This procedure is continued for \( \bar{X} = 2, \bar{X} = 3, ..., \bar{X} = N \). To illustrate this procedure an example is given for the particular case of two customer classes, \( k = 2 \), and a maximum waiting space of two customers, \( N = 2 \).

The total number of states for this \((k = 2, N = 2)\) model is

\[
k(\frac{k + N}{N}) = 2(2 + 2)
\]

\( = 12 \).

The first two state descriptions to be assigned are for the
two empty states. These are \( X = (1,0,0) \) and \( X = (2,0,0) \). The next four state descriptions to be assigned are for the states of systems with a total of one customer awaiting service, \( \bar{x} = 1 \). These are \( X = (1,1,0) \), \( X = (2,1,0) \), \( X = (1,0,1) \), and \( X = (2,0,1) \), respectively. The final state descriptions to be assigned are for the states of systems with a total of two customers awaiting service, \( \bar{x} = 2 \). From the Bose-Einstein statistic, the total number of states with \( \bar{x} = 2 \) is

\[
k(k + \bar{x} - 1) = \binom{2 + 2 - 1}{2} \quad (113)
\]

\[
= 6.
\]

The assigned states are \( X = (1,2,0) \), \( X = (2,2,0) \), \( X = (1,0,2) \), \( X = (2,0,2) \), \( X = (1,1,1) \), and \( X = (2,1,1) \), respectively. The IMC transition matrix given in Figure 7 is labeled with the ordering given above.

Subsequent discussion of the properties of the IMC transition matrix is aided by the given labeling convention. This convention is adopted for all multiple customer class models. It should be noted that this labeling procedure reduces to the straightforward labeling procedure used in the single customer class model when the special case of \( k = 1 \) is considered.

The IMC transition matrix for the \((k = 2, N = 2)\) model is displayed in Figure 7. The horizontal and vertical axes are labeled according to the convention previously discussed. The elements of the matrix are determined from the five
To $(x_0^n, x_1^n, x_2^n)$
From $(x_0'^n, x_1'^n, x_2'^n)$

<table>
<thead>
<tr>
<th>From $(x_0'^n, x_1'^n, x_2'^n)$</th>
<th>$(1,0,0)$</th>
<th>$(2,0,0)$</th>
<th>$(1,1,0)$</th>
<th>$(2,1,0)$</th>
<th>$(1,0,1)$</th>
<th>$(2,0,1)$</th>
<th>$(1,2,0)$</th>
<th>$(2,2,0)$</th>
<th>$(1,0,2)$</th>
<th>$(2,0,2)$</th>
<th>$(1,1,1)$</th>
<th>$(2,1,1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(1,0,0)$</td>
<td>$f_1^r x A(0,0;01)$</td>
<td>$f_2^r x A(0,0;02)$</td>
<td>$f_1^r x A(1,0;01)$</td>
<td>$f_2^r x A(1,0;02)$</td>
<td>$f_1^r x A(0,1;01)$</td>
<td>$f_2^r x A(0,1;02)$</td>
<td>$f_1^r x A(2,0;01)$</td>
<td>$f_2^r x A(2,0;02)$</td>
<td>$f_1^r x A(0,2;01)$</td>
<td>$f_2^r x A(0,2;02)$</td>
<td>$f_1^r x A(1,1;01)$</td>
<td>$f_2^r x A(1,1;02)$</td>
</tr>
<tr>
<td>$(2,0,0)$</td>
<td>$f_1^r x A(0,0;01)$</td>
<td>$f_2^r x A(0,0;02)$</td>
<td>$f_1^r x A(1,0;01)$</td>
<td>$f_2^r x A(1,0;02)$</td>
<td>$f_1^r x A(0,1;01)$</td>
<td>$f_2^r x A(0,1;02)$</td>
<td>$f_1^r x A(2,0;01)$</td>
<td>$f_2^r x A(2,0;02)$</td>
<td>$f_1^r x A(0,2;01)$</td>
<td>$f_2^r x A(0,2;02)$</td>
<td>$f_1^r x A(1,1;01)$</td>
<td>$f_2^r x A(1,1;02)$</td>
</tr>
<tr>
<td>$(1,1,0)$</td>
<td>$A(0,0;01)$</td>
<td>$0$</td>
<td>$A(1,0;01)$</td>
<td>$0$</td>
<td>$A(0,1;01)$</td>
<td>$0$</td>
<td>$A(0,2;01)$</td>
<td>$0$</td>
<td>$A(1,2;01)$</td>
<td>$0$</td>
<td>$A(0,0;02)$</td>
<td>$0$</td>
</tr>
<tr>
<td>$(2,1,0)$</td>
<td>$A(0,0;01)$</td>
<td>$0$</td>
<td>$A(1,0;01)$</td>
<td>$0$</td>
<td>$A(0,1;01)$</td>
<td>$0$</td>
<td>$A(0,2;01)$</td>
<td>$0$</td>
<td>$A(1,2;01)$</td>
<td>$0$</td>
<td>$A(0,0;02)$</td>
<td>$0$</td>
</tr>
<tr>
<td>$(1,0,1)$</td>
<td>$0$</td>
<td>$A(0,0;02)$</td>
<td>$0$</td>
<td>$A(0,1;02)$</td>
<td>$0$</td>
<td>$A(0,2;02)$</td>
<td>$0$</td>
<td>$A(1,2;02)$</td>
<td>$0$</td>
<td>$A(0,0;02)$</td>
<td>$0$</td>
<td></td>
</tr>
<tr>
<td>$(2,0,1)$</td>
<td>$0$</td>
<td>$A(0,0;02)$</td>
<td>$0$</td>
<td>$A(0,1;02)$</td>
<td>$0$</td>
<td>$A(0,2;02)$</td>
<td>$0$</td>
<td>$A(1,2;02)$</td>
<td>$0$</td>
<td>$A(0,0;02)$</td>
<td>$0$</td>
<td></td>
</tr>
<tr>
<td>$(1,2,0)$</td>
<td>$0$</td>
<td>$0$</td>
<td>$A(0,0;01)$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$A(0,2;01)$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$A(0,0;01)$</td>
<td>$0$</td>
</tr>
<tr>
<td>$(2,2,0)$</td>
<td>$0$</td>
<td>$0$</td>
<td>$A(0,0;01)$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$A(0,2;01)$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$A(0,0;01)$</td>
<td>$0$</td>
</tr>
<tr>
<td>$(1,0,2)$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$A(0,0;02)$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$A(0,1;02)$</td>
<td>$0$</td>
<td>$A(0,0;02)$</td>
<td>$0$</td>
</tr>
<tr>
<td>$(2,0,2)$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$A(0,0;02)$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$A(0,1;02)$</td>
<td>$0$</td>
<td>$A(0,0;02)$</td>
<td>$0$</td>
</tr>
<tr>
<td>$(1,1,1)$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$A(0,0;01)$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$A(0,1;01)$</td>
<td>$0$</td>
<td>$A(0,0;01)$</td>
<td>$0$</td>
</tr>
<tr>
<td>$(2,1,1)$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$A(0,0;01)$</td>
<td>$0$</td>
<td>$0$</td>
<td>$0$</td>
<td>$A(0,1;01)$</td>
<td>$0$</td>
<td>$A(0,0;01)$</td>
<td>$0$</td>
</tr>
</tbody>
</table>

Figure 7. IMC Transition Matrix for the $(k=2, N=2)$ Non-preemptive Priority Model.
transition probability cases listed in Table 6. Several examples are given to illustrate the procedure.

The element \( b((1,1,1), (1,0,0)) \) describes a candidate for a Case A) transition, as \( \bar{x}' = 2 > 0 \), and \( \bar{x}'' = 0 < N \). The third condition for Case A) is satisfied as \( x_0'' = w = 1 \). This condition is simply a requirement that a class 1 customer be served if available. The third condition is not satisfied as the quantity

\[
x_2'' = x_2' - d_{21} = 0 - 1 - 0
\]

\( = -1 \),

is negative. This condition is simply a requirement that the transition be caused by a non-negative set of arrivals. As it is clearly impossible to have a negative number of arrivals, a probability of zero is assigned to this transition, i.e., \( b((1,1,1), (1,0,0)) = 0 \).

The element \( b((1,0,0), (2,1,1)) \) describes a Case D) transition, as \( \bar{x}' = 0 \) and \( \bar{x}'' = 2 = N \). Note that for this case there are no additional conditions. This is because it is possible to go from any of the empty initial states, \( X' \) where \( \bar{x}' = 0 \), to any adjacent state. The probability associated with Case D) is assigned to this transition, i.e.,

\[
b((1,0,0), (2,1,1)) = \frac{r_{x_2''}}{r} \bar{A}(x_2'', x_2'; 0x_0')
\]

\( = \frac{r_2}{r} \bar{A}(1,1; 01). \)

The interpretation of this probability affirms the validity of the assignment. The ratio \( (r_2/r) \) is the probability that
a class 2 customer is the first customer to arrive during the period the server is idle. The quantity, \( \bar{A}(1,1; 02) \), is the probability that exactly one class 1 customer, \( a_1 = 1 \), and one class 2 customer, \( a_2 = 1 \), arrive during the service time. The service time is for a class 2 customer that enters the system to find the server idle, \( s = 02 \). The bar notation indicates that the system is truncated, as the maximum number of customers are waiting for service at the adjacent state. Other examples could be given, but it is felt that the preceding illustrations have sufficiently illustrated the probability assignment procedure. By considering each of the transitions indicated by the IMC matrix, a zero or non-zero probability is assigned to each corresponding element of the matrix.

An examination of Figure 7 indicates several significant features of the IMC transition matrix, \( B \), for the general \((k, N)\) model, even though the \( B \) matrix displayed in Figure 7 is for the specific \((k = 2, N = 2)\) model. The features of the IMC transition matrix are first identified and discussed in terms of the \((k = 2, N = 2)\) model. This discussion is then extended to the general \((k, N)\) model.

An examination of Figure 7 reveals that all elements of the first two rows of the \( B \) matrix for the \((k = 2, N = 2)\) model are non-zero. These rows correspond to the initial states that are empty, \( X' \), where \( x' = 0 \). This implies that in the \((k = 2, N = 2)\) model it is possible to make a
transition from any of the empty initial states to any other state in one step. For the general \((k, N)\) model the first \(k\) rows of the \(B\) matrix correspond to the initial states that are empty. All the elements of the first rows of the \(B\) matrix for the \((k, N)\) model are also non-zero. The justification for this result is based on the fact that the system is empty immediately after an empty initial state, \(x'\), where \(\bar{x}' = 0\). Before the time the adjacent state occurs a customer enters the system and completes service. The customer that enters the empty system can be of any class, and any set of arrivals can enter the system during this customer’s service time. Thus it is possible for the adjacent state to take on any state description. This point is stated formally by the Case C) and Case D) transitions given in Table 6.

Further examination of Figure 7 reveals a characteristic of all non-zero elements of the last six columns of the \(B\) matrix for the \((k = 2, N = 2)\) model. All of the non-zero elements in the last six columns are expressed in terms of the truncated arrival probabilities. The notation \(\bar{A}(a_1, a_2, s)\) indicates a truncated arrival probability. These six columns correspond to the adjacent states that are truncated, \(x''\), where \(\bar{x}'' = N = 2\). This characteristic is a result of the fact that any feasible transition from an initial state, \(x'\), to a truncated adjacent state, \(x''\), where \(\bar{x}' = N\), must account for the fact that the waiting space is filled at some point in time before the adjacent state occurs. For the general
(k, N) model the last columns also correspond to the truncated adjacent states, $X''$, where $\bar{x}'' = N$. From the Bose-Einstein statistic the total number of adjacent states with $\bar{x}'' = N$ is

$$k^{(k + \bar{x}'' - 1)} = \binom{k + N - 1}{N}. \quad (116)$$

Any transition with a truncated adjacent state is defined as either a Case B), Case D), or Case E) transition as given in Table 6. If the transition is feasible, it is defined as a Case B) or Case D) transition and is expressed in terms of the truncated arrival probability indicated for Case B) or Case D). If the transition is infeasible it is defined as a Case E) transition and is assigned a zero probability of occurrence. Thus the probability of any transition with a truncated adjacent state is expressed in terms of a truncated arrival probability if the probability is non-zero. This implies that all non-zero elements of the last $k^{(k + N - 1)}$ columns (the columns corresponding to the truncated adjacent epoch) of the B matrix for the (k, N) model are expressed in terms of the truncated arrival probabilities.

An additional insight into the structure of the IMC transition matrix is gained from the example given in Figure 7. The first ten initial states listed in the B matrix of Figure 7 are $X' = (1,0,0), (2,0,0), \ldots, (2,0,2)$. The transitions from these first ten initial states are not influenced by the sequencing rule. This is due to the fact that at most one class of customer is awaiting service in
these first ten initial states. As only one class of customer is available for service, no decision is made regarding the class of customer to serve next. The class of customer that is available is automatically served next. For example, for the initial state (1,0,2) only two class 2 customers await service, so a class 2 customer is automatically served next. For this initial state the set of feasible adjacent states is determined from Table 6 to be \( X'' = (2,0,1), (2,0,2), \) and \( (2,1,1) \). The appropriate non-zero probabilities of making these three transitions are determined by the use of Table 6 and are listed in the IMC transition matrix as \( A(0,0; 12), \overline{A}(0,1; 12) \) and \( \overline{A}(1,0; 12) \), respectively. It is emphasized that the set of feasible transitions and the associated probabilities are determined without regard to the sequencing rule. For this reason the entries in the first ten rows of the B matrix for the \((k = 2, N = 2)\) model, as displayed in Figure 7, do not vary with the sequencing rule.

The last two initial states listed in the B matrix of Figure 7 are \( X' = (1,1,1) \) and \( X' = (2,1,1) \). The transitions from these last two initial states are influenced by the sequencing rule. This is due to the fact that more than one class of customer is waiting for service, and a decision must be made as to which class of customer to serve next. This decision is made by the use of the sequencing rule. For example, for the initial state \((1,1,1)\) there is one class 1 customer and one class 2 customer awaiting service. The
non-preemptive priority sequencing rule always selects a class 1 customer for service if available. Thus the class 1 customer is the next customer to receive service. For the initial state (1,1,1) the set of feasible adjacent states is determined from Table 6 to be $X'' = (1,0,1), (1,1,1), \text{and} (1,0,2)$. The appropriate non-zero probabilities of making these three transitions are determined by the use of Table 6, and are listed in the IMC transition matrix as $A(0,0; 12), A(1,0; 12) \text{and} A(0,1; 12)$, respectively. The non-zero elements of the last row of the matrix are determined by an identical operation which depends on the sequence rule.

These results have major importance only for the $(k=2, N=2)$ model and are not extended to the general $(k, N)$ model. The $(k=2, N=2)$ model is used to illustrate all of the models developed in this study. As the first ten rows of the B matrix associated with this model are not dependent on the sequencing rule, attention can be focused on the last two rows of the B matrix. By considering these last two rows of the IMC transition matrix the effect of the sequencing rule is readily demonstrated.

Intermediate Summary

Before continuing with the model formulation it is advisable to review the developments to this point. The most difficult portions of the multiple customer class model formulation have been treated in the previous sections. It is hoped that a brief summary will clarify the procedures
obtained, as well as indicate the relationship of these procedures to the subsequent model development.

The primary objective of the preceding discussion was to provide an unambiguous method of determining the imbedded Markov chain transition matrix, B, for the general \((k, N)\) model. This objective was attained by considering five related tasks. The tasks were: (1) develop an expression for the marginal probability of an arbitrary set of customers arriving during a service time, (2) develop an expression to indicate the class of customer to be served next, given an arbitrary initial state and a non-preemptive priority sequencing rule, (3) develop a method of assigning a probability to a one step transition from an arbitrary initial state to an arbitrary final state, (4) develop a convention for labeling the states of the IMC transition matrix that conforms to the procedure given in Chapter III, and (5) combine the results of the preceding tasks to yield the IMC transition matrix of ultimate interest and indicate some of the general structural properties of this matrix.

The objective of defining the imbedded Markov chain transition matrix, \(B\), has been attained. Although this objective was addressed in terms of a non-preemptive priority sequencing rule, the results are more general. Only slight modification is required for the development of the IMC transition matrix for other sequencing rules. This modification consists of redefining the variable \(w\) which indicates
the class of customer that is served next for any arbitrary non-empty initial state.

The remaining steps in the model development follow the pattern established in Chapter III. After establishing the ergodic nature of the imbedded Markov chain the IMC steady state probabilities, \( g(X) \), are determined. These probabilities are modified to yield the augmented IMC steady state probabilities, \( g'(Y) \). The semi-Markov process is then discussed, and the augmented IMC steady state probabilities are converted to the SMP steady state probabilities, \( h(Y) \). Finally the steady state probabilities of ultimate interest, the random point in time (RPT) steady state probabilities, \( p(Z) \), are determined by modifying the SMP steady state probabilities. These probabilities are used to develop expressions for the measures of system performance. The model procedures are then illustrated by a small example problem.

The preceding steps are based on the same logic used in Chapter III. Thus the extensive discussion given to motivate and intuitively justify the procedures of Chapter III are appropriate for the procedures of this chapter. One other point should be emphasized. The procedures developed in subsequent sections are appropriate for any sequencing rule which uniquely determines the class of customer to be served next, given that at least one customer is waiting for service. In terms of the model notation this means that the value of \( w \) must be uniquely determined for all possible
initial states that are non-empty, $X'$ where $\overline{x'} > 0$. If $w$ is determined probabilistically, only slight modifications are required to assure the validity of the procedures. These conclusions are essential to the development of the models given in Chapter V.

**Ergodic Nature of the IMC**

The IMC steady state probabilities are designated $g(X)$, where $X = (x_0, x_1, \ldots, x_k)$. The IMC probability, $g(X)$, is interpreted as the probability that at any randomly selected service completion epoch after equilibrium conditions have been attained, the customer completing service is of class $x_0$, and the number of class $i$ customers awaiting service is $x_i$, $i = 1, 2, \ldots, k$. To find the steady state probability of being in any state, $X$, independent of the starting state of the system it must first be shown that the IMC is ergodic. From the nature of the system which the IMC describes, it is readily proven that the IMC is ergodic. The arguments given follow those proposed by Ross (1970).

To assist the discussion an arbitrary empty state, $X^0 = (j, 0, 0, \ldots, 0)$, $j = 1, 2, \ldots, k$, is considered. The zero superscript is used to indicate the system state where a class $j$ customer has just completed service and there are no customers of any class awaiting service. An additional arbitrary state, $X^+ = (x_0^+, x_1^+, \ldots, x_k^+)$, is also considered. The plus superscript is used to indicate the system state where a class $x_0^+$ customer has just received service, and the
number of class $i$ customers awaiting service is $x_i^+$, $i = 1, 2, \ldots, k$. The total number of customers awaiting service is

$$\bar{x}^+ = \sum_{i=1}^{k} x_i^+ .$$

(117)

It is possible to go from any arbitrary state, $X^+$, to any arbitrary empty state, $X^0$, in a finite number of transitions. This can be accomplished, for example, in a total of $(\bar{x}^+ + 1)$ transitions. The first $\bar{x}^+$ transitions consist of having a total of $\bar{x}^+$ customers receive service (according to any sequence rule ordering of service) without an intervening arrival. The last transition consists of having a class $j$ customer enter the idle system and receive service with no arrivals during this class $j$ customer's service time. It is also possible to go from any arbitrary empty state, $X^0$, to any arbitrary state, $X^+$, in a finite number of transitions. This can be accomplished, for example, in one transition. The transition consists of having a class $x_0^+$ customer enter the idle system and receive service, with $x_i^+$ class $i$ customers, $i = 1, 2, \ldots, k$, arriving during this class $x_0^+$ customer's service time. Expressing this conclusion in terms of finite Markov chains, the empty state and any other state "communicate", as transitions can occur between them in a finite number of steps, Ross (1970, p. 65). Any two arbitrary states also communicate as they both communicate with the empty states. As all the states of the Markov chain
communicate with each other the IMC is said to be "irreducible", Ross (1970, p. 66).

An examination of the empty state $x^0$ indicates that it is possible for any system starting in this state to return to state $x^0$ after any finite number of transitions. Such a state is said to be "aperiodic". As one state of the IMC is aperiodic and the IMC is irreducible, it follows that the IMC is aperiodic, Ross (1970, p. 66). Therefore the IMC is classified as a finite, irreducible, aperiodic Markov chain. The ergodic nature of the IMC is assured by a theorem due to A.A. Markov and given by Takacs (1962, p. 15). The theorem states that all finite, irreducible, aperiodic Markov chains are ergodic. It should be noted that the preceding arguments were independent of any sequencing rule considerations. Thus it has been proven that the IMC is ergodic for all the multiple customer class models, regardless of the sequencing rule.

**IMC Steady State Probabilities**

As the IMC is ergodic and has a finite number of states, the steady state probabilities, $g(x)$, exist and are independent of the starting state of the system. The equilibrium probabilities can be expressed as a $k^{(N+k)}$ element row vector, $G$, with elements

$$G = \{g(x)\}.$$  \hspace{1cm} (118)

The ordering of the elements in the $G$ vector is the same ordering used for the IMC transition matrix rows and columns.
The equilibrium probabilities are found by solving the well-known equations

\[ GB = G \]  

and \[ G_l = l \]  

where \( l \) is a \( k(\binom{N+k}{k}) \) element column vector with all elements equal to one. As an alternative method, the procedure developed by Singer (1964), and discussed in Chapter III may be used to determine the steady state probabilities. The remaining steps in the procedure consist of modifying the IMC steady state probabilities in order to obtain the probabilities of ultimate interest, the RPT steady state probabilities.

**Augmented IMC**

The first modification of the IMC steady state probabilities is undertaken in order to obtain information about the state of the system at some points in time that do not correspond to service completion epochs. The device used to accomplish this modification is an augmented imbedded Markov chain. It is shown that the steady state probabilities associated with this augmented IMC are readily obtained from the original IMC steady state probabilities, \( g(x) \).

The original IMC considers the state of the system at selected instants of time called service completion epochs, with state descriptions \( X = (x_0, x_1, \ldots, x_k) \). The augmented IMC considers the state of the system at service completion
epochs and at other selected instants of time called idle service initiation epochs. An idle service initiation epoch is defined as the instant of time that service is initiated for a customer that has entered the system to find the server idle. The state description for an idle service initiation epoch is \( e_i \) if the customer starting service at the idle service initiation epoch is a class \( i \) customer, \( i = 1, 2, \ldots, k \). This specific epoch is referred to as a class \( i \) idle service initiation epoch. The state description for the augmented IMC is designated \( Y \), where \( Y = e_i, i = 1, 2, \ldots, k \), or \( X \). If an epoch is a class \( i \) idle service initiation epoch, then the augmented IMC state description is \( Y = e_i \). If the epoch is a service completion epoch, then the augmented IMC state description is \( Y = X \). It is emphasized that both the augmented IMC and the original IMC describe the same underlying systems, but the augmented IMC considers more events than the original IMC. The service completion epoch events, with state descriptions \( X \), are considered by both the augmented IMC and the original IMC. The idle service initiation epoch events, with state descriptions \( e_i, i = 1, 2, \ldots, k \), are considered only by the augmented IMC.

The steady state probabilities associated with the augmented IMC are designated \( g'(Y) \), where \( Y = e_i, i = 1, 2, \ldots, k \), or \( Y = X \). These probabilities can be determined directly from the original IMC steady state probabilities, \( g(X) \). The augmented IMC steady state probabilities are determined by
considering two cases. The first case considers the steady state probabilities associated with service completion epochs, i.e., $g'(Y)$ where $Y = X$. The second case considers the steady state probabilities associated with the idle service initiation epochs, i.e., $g'(Y)$ where $Y = e_i$, $i = 1, 2, ..., k$.

The augmented IMC steady state probabilities associated with the service completion epochs, $g'(Y)$ where $Y = X$, are obtained by showing that they are proportional to the original IMC steady state probabilities, $g(X)$. As both the original IMC and the augmented IMC describe the same system, the absolute frequency of occurrence of the service completion epochs is the same for both the augmented IMC and the original IMC. However, the relative frequency occurrence of the service completion epochs is not the same for the augmented IMC and the original IMC. This is a result of the fact that the augmented IMC has a larger event space than the original IMC. The augmented IMC relative frequencies of occurrence are therefore equivalent to the original IMC relative frequencies, adjusted for the increased event space. This means that the relative frequency of occurrence of an arbitrary service completion epoch with state description, $X$, as given by the original IMC is directly proportional to the relative frequency of occurrence of the same event given by the augmented IMC. This conclusion is restated in terms of steady state probabilities. The steady state probabilities
associated with service completion epochs as given by the original IMC are directly proportional to the steady state probabilities for the equivalent events as given by the augmented IMC. Stating this result in equation form

\[ g'(Y) = g(X)/c_1, \quad Y = X, \]  

(121)

where \( c_1 \) is a normalizing constant. A method of evaluating \( c_1 \) is provided by a subsequent procedure.

The augmented IMC steady state probabilities associated with the service initiation epochs, \( g'(Y) \), where \( Y = e_i \), \( i = 1, 2, \ldots, k \), are obtained by showing that they are related to the augmented IMC steady state probabilities of the empty service completion epochs, \( g'(Y) \), where \( Y = X \) and \( \bar{X} = 0 \). This relationship is based on the fact that any idle period is always preceded by an empty service completion epoch, and any idle period is always concluded by an idle service initiation epoch. Thus for each idle period there is exactly one occurrence of an empty service completion epoch with state description, \( X \), where \( \bar{X} = 0 \), and exactly one occurrence of an idle service initiation epoch with state description \( e_i \), where \( i = 1, 2, \ldots, k \). This means that the set of empty service completion epochs has the same steady state probability as the set of idle service initiation epochs. As all of the events corresponding to empty service completion epochs are mutually exclusive and collectively exhaustive, the steady state probability of an empty service completion epoch is
The expression above is also the steady state probability that an idle service initiation epoch occurs. This probability does not indicate the class of the customer starting service at the idle service initiation epoch. The probability of interest is the steady state probability of a class i customer idle service initiation epoch. The probability of this event is found by first determining the conditional probability that a class i customer causes the idle service initiation epoch, given that an idle service initiation epoch occurs. Recall that an idle service initiation epoch occurs when the first customer to enter the system during an idle period starts to receive service. If this customer is of class i, then a class i idle service initiation epoch occurs. The probability that a class i customer is the first to arrive is \( r_i / r \), where \( r_i \) is the mean arrival rate of class i customers and \( r \) is the total mean arrival rate. Multiplying this conditional probability by the previously defined probability of occurrence of an idle service initiation epoch yields the desired joint probability. The steady state probability of a class i idle service initiation epoch is

\[
\Sigma_{X} g'(Y=X) = \sum_{j=1}^{k} g'(j,0,0,\ldots,0) \quad \text{s.t. } X=0
\]

\[
= \sum_{j=1}^{k} g(j,0,0,\ldots,0)/c_1.
\]
\[ g'(Y) = (r_i/r) \sum_{j=1}^{k} g(j,0,0,\ldots,0)/c_i, \quad Y = e_i, \quad i = 1,2,\ldots,k. \] (123)

The complete set of augmented IMC steady state probabilities is

\[ g'(Y) = \begin{cases} \frac{g(X)}{c_1}, & Y = X, \\ \frac{(r_i/r) \sum_{j=1}^{k} g(j,0,0,\ldots,0)/c_i}{c_i}, & Y = e_i, \quad i = 1,2,\ldots,k, \end{cases} \] (124)

where \( c_1 \) is a normalizing constant. The normalizing constant is readily evaluated, as the total of the probabilities must be one. The value of \( c_1 \) is not needed as it cancels in subsequent calculations; however, for purposes of clarification the value is

\[ c_1 = 1 + \sum_{j=1}^{k} g(j,0,0,\ldots,0). \] (125)

Two additional points should be emphasized. First, the method of converting IMC steady state probabilities to augmented IMC steady state probabilities is independent of any sequencing rule considerations. Therefore the method is suitable for all the multiple customer class models. Second, the augmented IMC is ergodic. This conclusion can be justified by the same logic used to prove that the original IMC is ergodic.

**Semi-Markov Process (SMP)**

The augmented IMC provides information about the state of the system only at selected instants of time called
service completion epochs and idle service initiation epochs. In this section the augmented IMC steady state probabilities are modified to indicate a limited amount of information about the state of the system at any random point in time after equilibrium conditions have been attained. The device used to accomplish this modification is a semi-Markov process (SMP). It is shown that the steady state probabilities associated with the SMP are readily obtained from the augmented IMC steady state probabilities, \( g'(Y) \).

The augmented IMC considers the state of the system only at selected instants of time called idle service initiation epochs and service completion epochs. The state description for an epoch is \( Y \). The state description, \( Y \), is \( e_i \) if the selected epoch is a class \( i \) idle service initiation epoch, \( i = 1,2,\ldots,k \). The state description, \( Y \), is \( X = (x_0, x_1, \ldots, x_k) \) if the selected epoch is a service completion epoch. The semi-Markov process retains the state description of the augmented IMC, \( Y \), but considers the process at random points in time. The state description for the SMP, \( Y \), is based on the state of the system at the epoch immediately preceding a randomly selected point in time. The epoch can be either a service completion epoch or an idle service initiation epoch. For example, consider an observation of the \((k = 2, N = 2)\) system at some random point in time. If the immediately preceding epoch was a service completion epoch with a class 2 customer leaving service, and one class 1 and zero class 2
customers awaiting service, then the state of the SMP assigned to the random point in time is \( Y = (x_0 = 2, x_1 = 1, x_2 = 0) \). It should be noted that this determination is made without any consideration given to the actual state of the system at the randomly selected point in time. The state description assigned to the random observation is determined only on the basis of the state of the system at the immediately preceding epoch. The SMP steady state probabilities are designated \( h(Y) \), \( Y = e_i, i = 1,2,\ldots,k \), or \( X \). The SMP equilibrium probability, \( h(Y) \), is interpreted as the probability that given any randomly selected point in time after equilibrium conditions have been attained, at the immediately preceding epoch the state description was \( Y \). The state description for the randomly selected point in time, \( Y \), is \( e_i \), if the immediately preceding epoch was a class \( i \) idle service initiation epoch. The state description for the randomly selected point in time, \( Y \), is \( X \) if the immediately preceding epoch was a service completion epoch. These steady state probabilities can be expressed in terms of the augmented IMC steady state probabilities, \( g'(Y) \).

To convert the augmented IMC steady state probabilities to the SMP steady state probabilities a simple operation involving the mean times between adjacent epochs is performed. The validity of the operation is based on theoretical developments given by Fabens (1961). Using the terminology and notation of Fox (1967, p. 14), the stationary probabilities
for a SMP, $\rho_i$, associated with an ergodic, finite state IMM, with states $i = 1, 2, \ldots, \eta$, are related to the stationary probabilities of the IMM, $\theta_i$, $i = 1, 2, \ldots, \eta$, by

$$\rho_j = \frac{\theta_j v_j}{\sum_{i=1}^{\eta} \theta_i v_i}, \quad j = 1, 2, \ldots, \eta$$

where

$$v_j = \text{the mean time until the next epoch, given that the epoch associated with state } j \text{ has just occurred, } j = 1, 2, \ldots, \eta.$$ 

Note that the summation in the denominator is merely a normalizing factor. This normalizing factor is designated as the constant $c_2$.

Before applying the above result it is necessary to know the mean time until the next epoch, given that the epoch associated with state $Y$ has just occurred, designated $v_Y$ for $Y = e_i$, $i = 1, 2, \ldots, k$, and $Y = X$. The values of $v_Y$ are determined by considering three mutually exclusive, collectively exhaustive subsets of $Y$. These subsets are (1) $Y = X$, where $x = 0$, (2) $Y = X$, where $x > 0$, and (3) $Y = e_i$, where $i = 1, 2, \ldots, k$.

**Subset (1)**

Consider the case where a service completion epoch has just occurred and no customers are awaiting service. At this epoch the facility becomes idle. The next epoch occurs when a customer of any class arrives and immediately starts to
receive service. The time between epochs is therefore the time until the arrival of the next customer. This period of time is a random variable following the exponential distribution with a mean of $1/r$, where $r$ is the total mean arrival rate. This distribution is appropriate because arrivals occur at random points in time. Thus the mean time until the next epoch is $1/r$ given that a service completion epoch has occurred and no customers await service.

Stating this result in equation form

$$v_Y = 1/r, \quad Y = X, \text{ where } \overline{X} = 0. \quad (127)$$

Subset (2)

Consider the case where a service completion epoch has just occurred and a non-zero number of customers await service. At this epoch a class $x_0$ customer has just completed service and a class $w$ customer starts to receive service.

The value of $w$ is determined with certainty by the sequencing rule. For the non-preemptive priority sequencing rule $w$ is determined with certainty by the equation

$$w = \min_{i} (i; x_i > 0, \ i = 1,2,\ldots,k). \quad (128)$$

The next epoch occurs when this class $w$ customer completes service. The time between epochs is therefore the time required to serve a class $w$ customer that was immediately preceded on the service facility by a class $x_0$ customer. This service time is a random variable following the probability density function $f_{x_0w}(t)$ which has a mean of $m_{x_0w}$.

Thus the mean time until the next epoch is $m_{x_0w}$, given that
a service completion epoch has just occurred and a non-zero number of customers await service. Stating this result in equation form

\[ v_Y = m_{X_0} w, \quad Y = X, \text{ where } \bar{x} > 0. \]  

(129)

where \( w \) is determined with certainty by the sequencing rule.

**Subset (3)**

Consider the case where a class \( i \) idle service initiation epoch has just occurred. At this epoch a class \( i \) customer starts to receive service. The next epoch occurs when this class \( i \) customer completes service. The time between epochs is therefore the time required to serve a class \( i \) customer that enters the system to find the server idle. This service time is a random variable following the probability density function \( f_{0i}(t) \) which has a mean of \( m_{0i} \). Thus the mean time until the next epoch is \( m_{0i} \) given that a class \( i \) idle service initiation epoch has just occurred. Stating this result in equation form

\[ v_Y = m_{0i}, \quad Y = e_i, \quad i = 1, 2, \ldots, k. \]  

(130)

Collecting the previous results for all three subsets of \( Y \) yields expressions for the mean time until the next epoch given that the epoch with state description \( Y \) has just occurred as

\[ v_Y = \begin{cases} 
  m_{0i}, & Y = e_i, \quad i = 1, 2, \ldots, k, \\
  1/r, & Y = X, \text{ where } \bar{x} = 0, \\
  m_{X_0} w, & Y = X, \text{ where } \bar{x} > 0.
\end{cases} \]  

(131)
The only restriction on $w$ is that for any given state $x$, where $x > 0$, $w$ be determined with certainty.

The SMP equilibrium probabilities, $h(Y)$, can now be determined by a straightforward application of the previously stated formula. The SMP steady state probabilities are

$$h(Y) = \begin{cases} 
g'(e_i)m_{0i}/c_2, & Y = e_i, \; i = 1, 2, \ldots, k, 
g(X)/(rc_2), & Y = X, \; \text{where } x = 0, 
g'(X)m_{x0w}/c_2, & Y = X, \; \text{where } x > 0, \end{cases}$$

where

$$c_2 = \sum_{i=1}^{k} g'(e_i)m_{0i} + \sum_{X} g(X)/r + \sum_{X} g'(X)m_{x0w}.$$

Substituting the expressions for $g'(Y)$ in terms of $g(X)$ yields

$$h(Y) = \begin{cases} 
(\sum_{X} g(X))r_{i}m_{0i}/(rc_1 c_2), & Y = e_i, \; i = 1, 2, \ldots, k, \text{ s.t. } x = 0 
g(X)/(rc_1 c_2), & Y = X, \; \text{where } x = 0, 
g'(X)m_{x0w}/(c_1 c_2), & Y = X, \; \text{where } x > 0, \end{cases}$$

where

$$c_2 = \sum_{i=1}^{k} \left( \sum_{X} g(X))r_{i}m_{0i}/(rc_1) \right) + \sum_{X} g(X)/(rc_1) \text{ s.t. } x = 0$$

$$+ \sum_{X} g(X)m_{x0w}/c_1 \text{ s.t. } x > 0.$$

This expression can be simplified by defining the quantity $d = c_1 c_2$. Making this substitution yields
\[ h(Y) = \begin{cases} 
\left( \sum_{X} g(x) \right) r_i m_{0i}/(rd), & Y = e_i, \ i = 1, 2, \ldots, k, \\
\sum_{X} g(x)/(rd), & Y = X, \text{ where } \overline{x} = 0, \\
g(x)m_{X0}/d, & Y = X, \text{ where } \overline{x} > 0, 
\end{cases} \]

where

\[ d = \left( \sum_{i=1}^{k} \left( r_i m_{0i}/r + 1/r \right) \right) \sum_{X} g(x) + \sum_{X} g(x)m_{X0}/d. \]

The above expressions depend only on the IMC steady state probabilities, \( g(x) \). This implies that it is not necessary to calculate the augmented IMC steady state probabilities, \( g'(Y) \), in order to find the SMP steady state probabilities, \( h(Y) \).

It should be noted that this method of converting IMC steady state probabilities to SMP steady state probabilities depends on the sequencing rule. This dependence is based only on the fact that the calculation of \( v_y, Y = X, \text{ where } \overline{x} > 0 \), involves the variable \( w \). As long as \( w \) is defined with certainty the above procedure is valid. In other words, if the sequencing rule determines the next class of customer to serve with certainty when there is at least one customer waiting for service, the above procedure is valid. The procedure is easily modified to account for the case where the next class of customer to serve is determined probabilistically.
Random Point in Time (RPT)

The SMP steady state probabilities provide information about the state of the system at the epoch immediately preceding a random point in time after equilibrium conditions have been attained. In order to calculate measures of system performance it is necessary to have steady state probabilities that provide information about the state of the system at a random point in time after equilibrium conditions have been attained. These steady state probabilities are designated random point in time (RPT) steady state probabilities. It is shown that the RPT steady state probabilities are readily obtained from the SMP steady state probabilities, $h(Y)$.

The state of the system at any random point in time is assigned the state description $Z$, $Z = 0$ or $X$. If at the random point in time the system is completely empty, i.e., there are no customers in the system, then the RPT state description is $Z = 0$. If at the random point in time the system is non-empty, i.e., there is at least one customer in the system, then the RPT state description is $Z = X = (x_0, x_1, \ldots, x_k)$. The vector $X = (x_0, x_1, \ldots, x_k)$ indicates that at the random point in time the customer receiving service is a class $x_0$ customer and that $x_i$ class $i$ customers are awaiting service, $i = 1, 2, \ldots, k$. It is emphasized that at the random time the system is observed the class $x_0$ customer is receiving service, not completing service.
In subsequent discussion the state of the system at a random point in time, \( Z \), is related to the state of the system at the immediately preceding epoch, \( Y \). To facilitate this discussion it is necessary to differentiate between the state of system at these two points in time with additional notation. The state of the system at the preceding epoch is \( Y = e_i, \; i = 1, 2, \ldots, k \) or \( Y = X' \), where \( X' = (x'_0, x'_1, \ldots, x'_k) \). The state of the system at the random point in time is \( Z = 0 \) or \( Z = X \) where \( X = (x_0, x_1, \ldots, x_k) \). Note that the single prime superscript on the state description symbol indicates the state of the system at the preceding epoch. The absence of a superscript on the state description symbol indicates the state of the system at the random point in time. This notation may seem awkward, but subsequent developments will demonstrate its usefulness.

It is clear that the state of the system at a random observation, \( Z = 0 \) or \( X \), is related to the state of the system at the immediately preceding epoch, \( Y = X' \) or \( e_i, \; i = 1, 2, \ldots, k \). This relationship is clarified by adopting some previously defined notation. The random observation occurs at some point in time between adjacent epochs. Designate the time of occurrence of the preceding epoch as time zero and the time of occurrence of the following epoch as time \( t \). The time of occurrence of the random observation is designated \( t' \), where \( 0 \leq t' \leq t \). The duration of time from the occurrence of the preceding epoch to the random observation
is $t'$ and is called the random interruption portion of the time between epochs. This implies that if the system was in state $Y$ at the preceding epoch and is in state $Z$ at the random observation, then the change in system state occurred in the random interruption time, $t'$. The conditional probability of being in state $Z$ at a random observation, given that the system was in state $Y$ at the preceding epoch, is designated $\text{pr}(Z|Y)$. Recall that the steady state probability of being in state $Y$ at any epoch is $h(Y)$. Using these quantities the unconditional probability of being in state $Z$ at some random point in time after equilibrium conditions are attained is

$$p(Z) = \sum_Y h(Y) \text{pr}(Z|Y), \ Z = 0 \text{ or } X \quad (135)$$

This probability can be rewritten by performing the summation over three mutually exclusive, collectively exhaustive subsets of $Y$. The subsets are (1) $Y = X'$, where $x' = 0$, the set of preceding epochs that are empty service completion epochs; (2) $Y = X'$, where $x' > 0$, the set of preceding epochs that are non-empty service completion epochs; and (3) $Y = e_i$, $i = 1, 2, \ldots, k$, the set of preceding epochs that are idle service initiation epochs. The probability of being in state $Z$ at a random point in time after equilibrium conditions are attained is

$$p(Z) = \sum_{Y = X'} h(Y) \text{pr}(Z|Y) + \sum_{Y = x'} h(Y) \text{pr}(Z|Y) + \sum_{Y = e_i} h(Y) \text{pr}(Z|Y),$$

$$s.t. x' = 0 \quad s.t. x' > 0 \quad s.t. x = e_i \quad Z = 0 \text{ or } X \quad (136)$$
The RPT steady state probabilities, \( p(Z) \), can therefore be determined from a knowledge of the SMP steady state probabilities, \( h(Y) \), and the conversion probabilities, \( \text{pr}(Z|Y) \).

These conversion probabilities are determined by considering the events that must occur to cause the system to go from a state description of \( Y \) at the preceding epoch to a state description of \( Z \) at a randomly selected time, \( t' \), before the next epoch occurs. The conversion probabilities for each of the three subsets of \( Y \) are stated, and a brief justification given. The justification draws on previously developed conclusions.

**Subset (1)**

The conversion probabilities for the set of preceding epochs that are empty service completion epochs are

\[
\text{pr}(Z|Y) = \begin{cases} 
1, & \text{if } Z = 0, \\
0, & \text{if } Z = X; 
\end{cases}
\]

where

\[
Y = X' \text{ where } X' = 0. \tag{137}
\]

Recall that a system idle period always begins with an empty service completion epoch and always ends with an idle service initiation epoch. Thus the state of the system at any random point in time, \( t' \), after an empty service completion epoch and before the next epoch, is known with certainty to be the empty state, \( Z = 0 \).

**Subset (2)**

The conversion probabilities for the set of preceding
epochs that are non-empty service completion epochs are

\[
\begin{align*}
\Pr(Z|Y) &= \begin{cases} 
0, & Z = 0, \\
A'(\{x_i - x_i' + d_{iw}\}; x_{i0}^w), & Z = x, \\
\bar{x} < N, & x_0 = w, \\
[x_i - x_i' + d_{iw}] \geq 0, & Z = x, \\
\bar{x} = N, & x_0 = w, \\
[x_i - x_i' + d_{iw}] \geq 0, & Z = x, \\
0, & \text{otherwise.}
\end{cases}
\end{align*}
\]

where

\[Y = x', \text{ where } \bar{x}' > 0.\]

From the justification of the first subset of conversion probabilities it is known that the system can be empty if and only if the preceding epoch is an empty service completion epoch. This implies that it is impossible for the system to be empty if the preceding epoch is a non-empty service completion epoch. In probability terms,

\[\Pr(Z = 0 | Y = x' \text{ where } \bar{x}' > 0) = 0.\]

The remaining conversion probabilities, \(\Pr(Z = x | Y = x', \text{ where } \bar{x}' > 0)\), are determined by considering the events that must occur during the random interruption portion of the time between epochs, \(t'\). The events are identical to those required for the Case A) and Case B) IMC transition
probabilities as described in Table 6. The only difference is that the events for the IMC transition probabilities occur during the entire time between epochs, and the events for the conversion probabilities occur during the random interruption portion of the time between epochs. Recall that the time between epochs is just the service time of the customer that started service at the preceding epoch. This customer must be of class $w$, as determined by the sequencing rule, in order for the change of state to be feasible. This same class $w$ customer must be receiving service at the time of the random observation in order for the change of state to be feasible. Stating this feasibility condition in terms of the state description notation, it must be true that $x_0 = w$ for the conversion probability to be non-zero.

For the change of state to occur the non-negative set of arrivals $\{x_{-1} - x_{-1} + d_{iw}\}$ must enter the system in the random interruption time, $t'$. Note that this set of arrivals is just the difference in the contents of the queue at the preceding epoch and the contents of the queue at the random point in time adjusted for the class $w$ customer that entered service at the preceding epoch. This change of state is feasible only if the number of arrivals of each class of customer is non-negative. Stating this feasibility condition in terms of the state description notation, it must be true that $\{x_{-1} - x_{-1} + d_{iw}\} \geq 0$ for the conversion probability to be non-zero. The time between epochs is the service time of a
class w customer that was immediately preceded on the service facility by a class x₀ customer. The density function of the service time is \( f_{X_0^w}(t) \). The random interruption portion of this service time, \( t' \), follows the density function \( f'_{X_0^w}(t') \) which has been previously defined in terms of the service time density function. Thus if both feasibility conditions are met, the change of state occurs by having the non-negative set of arrivals \( \{x_i - x_i' + d_{iw}\} \) enter the system during the random interruption portion of the service time for a class w customer that is served after a class x₀ customer is served. If the system has not truncated at time \( t' \), i.e. \( \bar{x} < N \), the probability of this event is \( A'(\{x_i - x_i' + d_{iw}\};x_0^w) \). If the system has truncated at time \( t' \), i.e., \( \bar{x} = N \), the probability of the event is \( \bar{A}'(\{x_i - x_i' + d_{iw}\};x_0^w) \). If both feasibility conditions are not met the probability that the change of state occurs is zero. These results are restated in equation form in order to relate them to the IMC transition probabilities. The equations are
Case A')
\[ A'(\{x_i - x_i' + d_{iw}\}; x_0'w), \quad \bar{x}' > 0, \]
\[ \bar{x} < N, \]
\[ x_0 = w, \]
\[ \{x_i - x_i' + d_{iw}\} \geq 0 \]

Case B')
\[ A'(\{x_i - x_i' + d_{iw}\}; x_0'w), \quad \bar{x}' > 0, \]
\[ \bar{x} = N, \]
\[ x_0 = w, \]
\[ \{x_i - x_i' + d_{iw}\} \geq 0, \]

Case E')
\[ 0 \]
\[ \bar{x}' > 0, \]
otherwise.

A comparison of the Case A'), Case B'), and Case E') probability statements above, with the probability statements for Case A), Case B), and Case E) of Table 6, respectively, indicates the relationship between the IMC transition probabilities associated with non-empty initial epochs and the conversion probabilities associated with non-empty preceding epochs. The notation differs only by the prime superscript on the arrival probabilities. This superscript indicates that the set of arrivals enters the system during the random interruption portion of the service time.

It is desirable to express the above probabilities in matrix form in order to expedite calculations. For this purpose the conversion matrix, \( B' \), is formed. The horizontal
axis is labeled $Z = X$, and the ordering of the labels is identical to that of the horizontal axis of the IMC transition matrix, $B$. The vertical axis is labeled $Y = X'$, where the state descriptions $X'$ with $x' = 0$ are deleted. The ordering of the labels is identical to that of the vertical axis of the IMC transition matrix, $B$, with the first $k$ rows (the rows corresponding to the state descriptions $X'$ with $x' = 0$) deleted. An example of this labeling procedure for the conversion matrix for the $(k = 2, N = 2)$ model is displayed in Figure 8. For comparison purposes the IMC transition matrix for the $(k = 2, N = 2)$ model is displayed in Figure 7.

The elements of the conversion matrix are assigned the values

$$b'(Y = x', Z = X) = \text{pr}(Z = X | Y = X'),$$

(140)

where $Y = X'$ is not defined for $x' = 0$. Values are assigned to the elements of the conversion matrix, $B'$, using Case $A'$), Case $B'$), and Case $E'$) as defined in Table 6. There are only two differences between the resulting matrices, $B$ and $B'$. First, the first $k$ rows of the $B$ matrix have no corresponding elements in the $B'$ matrix. Second, the service time arrival probabilities assigned to the non-zero elements of the $B$ matrix, $A(\{a_i\};s)$ and $\overline{A}(\{a_i\};s)$, are replaced by the equivalent random interruption time arrival probabilities, $A'(\{a_i\};s)$ and $\overline{A}'(\{a_i\};s)$, respectively. The zero elements are identical in both matrices. This similarity is verified
To Z
\( \{x_0, x_1, x_2\} \)

From Y
\( \{x_0', x_1', x_2'\} \)

<table>
<thead>
<tr>
<th></th>
<th>(1,0,0)</th>
<th>(2,0,0)</th>
<th>(1,1,0)</th>
<th>(2,1,0)</th>
<th>(1,0,1)</th>
<th>(2,0,1)</th>
<th>(1,2,0)</th>
<th>(2,2,0)</th>
<th>(1,0,2)</th>
<th>(2,0,2)</th>
<th>(1,1,1)</th>
<th>(2,1,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,1,0)</td>
<td>0</td>
<td>( A'(0,0;11) )</td>
<td>0</td>
<td>( A'(1,0;11) )</td>
<td>0</td>
<td>( \tilde{A}(2,0;11) )</td>
<td>0</td>
<td>( \tilde{A}'(0,2;11) )</td>
<td>0</td>
<td>( \tilde{A}'(1,1;11) )</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>(2,1,0)</td>
<td>0</td>
<td>( A'(0,0;21) )</td>
<td>0</td>
<td>( A'(1,0;21) )</td>
<td>0</td>
<td>( \tilde{A}'(2,0;21) )</td>
<td>0</td>
<td>( \tilde{A}'(0,2;21) )</td>
<td>0</td>
<td>( \tilde{A}'(1,1;21) )</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>(1,0,1)</td>
<td>0</td>
<td>0</td>
<td>( A'(0,0;12) )</td>
<td>0</td>
<td>( A'(1,0;12) )</td>
<td>0</td>
<td>( \tilde{A}(2,0;12) )</td>
<td>0</td>
<td>( \tilde{A}(0,2;12) )</td>
<td>0</td>
<td>( \tilde{A}(1,1;12) )</td>
<td></td>
</tr>
<tr>
<td>(2,0,1)</td>
<td>0</td>
<td>0</td>
<td>( A'(0,0;22) )</td>
<td>0</td>
<td>( A'(1,0;22) )</td>
<td>0</td>
<td>( \tilde{A}(2,0;22) )</td>
<td>0</td>
<td>( \tilde{A}(0,2;22) )</td>
<td>0</td>
<td>( \tilde{A}(1,1;22) )</td>
<td></td>
</tr>
<tr>
<td>(1,2,0)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( A'(0,0;11) )</td>
<td>0</td>
<td>0</td>
<td>( \tilde{A}(1,0;11) )</td>
<td>0</td>
<td>0</td>
<td>( \tilde{A}(0,1;11) )</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>(2,2,0)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( A'(0,0;21) )</td>
<td>0</td>
<td>0</td>
<td>( \tilde{A}(1,0;21) )</td>
<td>0</td>
<td>0</td>
<td>( \tilde{A}(0,1;21) )</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>(1,0,2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( A'(0,0;12) )</td>
<td>0</td>
<td>0</td>
<td>( \tilde{A}(1,0;12) )</td>
<td>0</td>
<td>( \tilde{A}(0,1;12) )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2,0,2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( A'(0,0;22) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \tilde{A}(1,0;22) )</td>
<td>0</td>
<td>( \tilde{A}(0,1;22) )</td>
<td></td>
</tr>
<tr>
<td>(1,1,1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( A'(0,0;11) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \tilde{A}(0,1;11) )</td>
<td>0</td>
<td>( \tilde{A}(0,1;11) )</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>(2,1,1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( A'(0,0;21) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \tilde{A}(0,1;21) )</td>
<td>0</td>
<td>( \tilde{A}(0,1;21) )</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Figure 8. Conversion Matrix for the \((k=2, N=2)\) Non-preemptive Priority Model.
by comparing the IMC transition matrix and the conversion matrix for the \((k = 2, N = 2)\) non-preemptive priority model displayed in Figure 7 and Figure 8, respectively. It is seen that the \(B'\) matrix can be obtained from the \(B\) matrix by deleting the first \(k\) rows of the \(B\) matrix and replacing the non-zero elements of the \(B\) matrix with the equivalent random interruption time probabilities.

The conversion matrix, \(B'\), provides a convenient method of performing certain summations by matrix operations. This feature is demonstrated when the conversion probabilities are utilized to calculate the RPT steady state probabilities. Before calculating the RPT steady state probabilities the third and final subset of conversion probabilities are developed.

**Subset (3)**

The conversion probabilities for the set of preceding epochs that are idle service initiation epochs are

\[
\text{pr}(Z|Y) = \begin{cases} 
0, & Z = 0, \\
A'([x_i];0x_0) & Z = x, \\
\bar{A}'([x_i];0x_0) & Z = x, \\
0, & x_0 \neq i, \\
\end{cases} 
\]

where

\(x \leq N, \bar{x} < N, x_0 = i\)
\[ y = e_i, \ i = 1, 2, \ldots, k. \]

From the justification of the first subset of conversion probabilities it is known that the system can be empty if and only if the preceding epoch is an empty service completion epoch. Thus it is impossible for the system to be empty if the preceding epoch is an idle service initiation epoch. In probability terms, \( \Pr(Z = 0|Y = e_i) = 0 \), where \( i = 1, 2, \ldots, k \).

The remaining conversion probabilities, \( \Pr(Z = X|Y = e_i) \), where \( i = 1, 2, \ldots, k \), are determined by considering the events that must occur during the random interruption portion of the time between epochs, \( t' \). The time between epochs is just the service time of the class \( i \) customer that started service at the preceding idle service initiation epoch. This same class \( i \) customer must still be receiving service at the time of the random observation in order for the change of state to be feasible. Stating this feasibility condition in terms of the state description notation, it must be true that \( x_0 = i \) for the conversion probability to be non-zero. The time between epochs is the service time of a class \( x_0 \) customer that receives service after the server has been idle for a non-zero period of time. The density function of the service time is \( f_{0x_0}(t) \). The random interruption portion of this service time, \( t' \), follows the density function \( f_{0x_0}(t') \) which has been previously defined in terms of the service time density function.

For the change of state to occur the set of arrivals
[$x_i$] must enter the system in time $t'$. Note that this set of arrivals is just the contents of the queue at the random point in time, as the queue was empty at the preceding service initiation epoch. Thus if the feasibility condition is met, the change of state occurs by having the set of arrivals [$x_i$] enter the system during the random interruption portion of a service time for a class $x_0$ customer that is served after the system is idle. If the system has not truncated at time $t'$, i.e., $x < N$, the probability of this event is $A'([x_i]; 0x_0)$. If the system has truncated at time $t'$, i.e., $x = N$, the probability of this event is $A'([x_i]; 0x_0)$. If the feasibility condition is not met the probability that the change of state occurs is zero.

The three subsets of conversion probabilities are used to determine the RPT steady state probabilities. The probability that the system is empty at a random point in time after equilibrium conditions are attained is

$$p(Z=0) = \sum_{Y=X'} h(Y) pr(Z=0|Y) + \sum_{Y=X'} h(Y) pr(Z=0|Y) + \sum_{Y=E_i} h(Y) pr(Z=0|Y)$$

$$s.t. X'=0 \quad s.t. X'>0 \quad i=1,2,\ldots,k$$

$$= \sum_{Y=X'} h(Y) 1 + \sum_{Y=X'} h(Y) 0 + \sum_{Y=E_i} h(Y)0$$

$$s.t. X'=0 \quad s.t. X'>0 \quad i=1,2,\ldots,k$$

$$= \sum_{Y=X'} h(Y).$$

$$s.t. X'=0$$

The probability that a class $x_0$ customer is receiving service and $x_i$ class $i$ customers are awaiting service,
i=1,2,...,k, at a random point in time after equilibrium conditions are attained is given by two expressions. The first expression is appropriate for the situation where a class $x_0$ customer is receiving service and the waiting room is not filled to capacity, i.e. $\bar{x}<N$. This probability is

$$p(Z=X) = \sum_{Y=X'} h(Y)pr(Z=X|Y) + \sum_{Y=X'} h(Y)pr(Z=X|Y) + \sum_{Y=e_i} h(Y)pr(Z=X|Y),$$

s.t. $\bar{x}'=0$ s.t. $\bar{x}>0$ i=1,2,...,k ($143$)

$$= \sum_{Y=X'} h(Y) 0 + \sum_{Y=X'} h(Y)b'(Y,X) + h(e_{x_0})A'([x_i];0x_0),$$

s.t. $\bar{x}=0$ s.t. $\bar{x}'>0$

The second expression is appropriate for the situation where a class $x_0$ customer is receiving service and the waiting room is filled to capacity, i.e., $\bar{x}=N$. This probability is

$$p(Z=X) = \sum_{Y=X'} h(Y)pr(Z=X|Y) + \sum_{Y=X'} h(Y)pr(Z=X|Y) + \sum_{Y=e_i} h(Y)pr(Z=X|Y),$$

s.t. $\bar{x}'=0$ s.t. $\bar{x}>0$ i=1,2,...,k ($144$)

$$= \sum_{Y=X'} h(Y) 0 + \sum_{Y=X'} h(Y)b'(Y,X) + h(e_{x_0})A'([x_i];0x_0),$$

s.t. $\bar{x}=0$ s.t. $\bar{x}'>0$

+ $\sum_{Y=e_i} h(Y) 0$, i$\neq x_0$

$$= \sum_{Y=X'} h(Y)b'(Y,X) + h(e_{x_0})A'([x_i];0x_0),$$

s.t. $\bar{x}=N$.
Collecting the results yields the RPT steady state probabilities

$$p(Z) = \begin{cases} 
\sum_{Y=X'} h(Y) & Z=0, \\
\text{s.t. } \sum_{X'} = 0 \\
\sum_{Y=X'} h(Y)b'(Y,X) + h(x_0)A'(x_i;0x_0), & Z=X, \\
\text{s.t. } x'>0 & \bar{x}=0,1,...,N-1, \\
\sum_{Y=X'} h(Y)b'(Y,X) + h(x_0)\bar{A}'(x_i;0x_0), & Z=X, \\
\text{s.t. } x'>0 & \bar{x}=N. 
\end{cases} \tag{145}$$

Note that the summations involving elements of the conversion matrix, $B'$, can be performed by an equivalent matrix operation. The summation can be accomplished by multiplying the row vector $H(X')$ with elements $h(Y = X')$, where $x'>0$, by the column of the conversion matrix with state description $Z = X$.

One additional point is emphasized, namely the above expressions are valid for any sequencing rule. This result is based on the fact that the effect of the sequencing rule is imbedded in the elements of the conversion matrix.

Queue Length Probabilities

The RPT steady state probabilities provide information about the exact state of the system in terms of the queue contents and the class of customer receiving service at any random point in time after equilibrium conditions are attained. In order to calculate measures of system performance the marginal probabilities of having exactly $j$ class $i$ customers in the queue.
customers in the system (in the queue) after equilibrium conditions are attained are required. These marginal probabilities are readily obtained from the RPT steady state probabilities, $p(Z)$.

The steady state probability that there are exactly $j$ class $i$ customers in the system is

$$
\text{pr}(n_i = j) = \begin{cases} 
  p(Z=0) + \sum_{Z=X \atop s.t. x_i \neq i} p(Z), & j=0, \\
  \sum_{Z=X \atop s.t. x_i=i} p(Z) + \sum_{Z=X \atop s.t. x_i \neq i} p(Z), & j=1,2,\ldots,N \\
  \text{and } x_i=j \\
  p(Z = (x_0 = i, x_1 = 0, \ldots, x_i = N, \ldots, x_k = 0)), & j=N+1.
\end{cases}
$$

The steady state probability that there are exactly $j$ class $i$ customers in the queue is

$$
\text{pr}(n_{qi} = j) = \begin{cases} 
  p(Z=0) + \sum_{Z=X \atop s.t. x_i = 0} p(Z), & j=0, \\
  \sum_{Z=X \atop s.t. x_i = j} p(Z), & j=1,2,\ldots,N. 
\end{cases}
$$

The steady state probability that there is a total of exactly $j$ customers of all classes in the system is

$$
\text{pr}(n = j) = \begin{cases} 
  p(Z=0), & j=0, \\
  \sum_{Z=X \atop s.t. x = j-1} p(Z), & j=1,2,\ldots,N+1.
\end{cases}
$$
The steady state probability that there is a total of exactly \( j \) customers of all classes in the queue waiting for service is

\[
\Pr(n_q = j) = \begin{cases} 
\sum_{Z=X}^{j=0} p(Z) & \text{for } j=0, \\
\sum_{Z=X}^{j=1,2,\ldots,N} p(Z) & \text{for } j=1,2,\ldots,N. 
\end{cases}
\]

Measures of Performance

Using the previously defined probability mass functions the expected number of each class of customer in the system and in the queue can be found. The expected waiting time for each class of customer in the system and in the queue is also determined. As the waiting space is limited to a maximum of \( N \) customers, not all customers that desire to enter the system can enter. This feature is reflected in the mean effective arrival rate for each class of customer. The unconditional expected service time of each class of customer is also of interest as the service times are interactive. In addition to the measures of performance for each class of customer, the measures for the overall customer population are provided. All measures of system performance are for systems that have reached equilibrium conditions.

The expected number of class \( i \) customers in the system and the expected number of class \( i \) customers in the queue, designated \( \mathbb{E}(n_i) \) and \( \mathbb{E}(n_{q_i}) \), respectively, are
The expected total number of customers in the system and the expected total number of customers in the queue regardless of customer class, designated $E(n)$ and $E(n_q)$, respectively are

$$E(n) = \sum_{j=0}^{N+1} j \, p_r(n=j),$$  
and

$$E(n_q) = \sum_{j=0}^{N} j \, p_r(n_q=j),$$  

The effective mean arrival rate for class $i$ customers, $r_{eff_i}$, is the expected number of customers added to the system per unit time. To obtain the effective mean arrival rate, the mean arrival rate, $r_i$, is multiplied by the probability that the system is not truncated, $(1 - p_r(n = N + 1))$. The validity of this calculation is based on the fact that customers arrive at random points in time, and the probability that the system is not truncated at a random point in time is $(1 - p_r(n = N + 1))$. The effective mean arrival rate for class $i$ customers is

$$r_{eff_i} = r_i(1 - p_r(n = N + 1)), \quad i=1,2,\ldots,k. \quad (152)$$

The effective mean arrival rate for the overall system is

$$r_{eff} = r(1 - p_r(n = N + 1)). \quad (153)$$

The expected waiting time in the system for a class $i$ customer is

$$E(n_i) = \sum_{j=0}^{N+1} j \, p_r(n_i=j), \quad i=1,2,\ldots,k, \quad (150)$$

and

$$E(n_{q_i}) = \sum_{j=0}^{N} j \, p_r(n_{q_i}=j), \quad i=1,2,\ldots,k. \quad (151)$$
customer, \( E(w_i) \), and the expected waiting time in the queue for a class \( i \) customer, \( E(w_{qi}) \), are calculated by using a result proven by Eilon (1969). Eilon indicates that these quantities are

\[
E(w_i) = E(n_i)r_{eff}, \quad i=1,2,\ldots,k,
\]

and

\[
E(w_{qi}) = E(n_{qi})/r_{eff}, \quad i=1,2,\ldots,k.
\]  

(154)

The expected waiting time of customers in the system and the expected waiting time of customers in the queue regardless of customer class are

\[
E(w) = E(n)/r_{eff},
\]

and

\[
E(w_q) = E(n_q)/r_{eff},
\]

(155)

respectively. As stated in Chapter III these relationships involving the mean waiting time, the mean effective arrival rate and the mean number of customers are not dependent on any specific assumptions about the arrival distributions, the service distributions, or the sequencing rule.

The remaining measure of system performance, the expected service time for class \( i \) customers is

\[
E(t_i) = E(w_i) - E(w_{qi}), \quad i=1,2,\ldots,k.
\]  

(156)

The expected service time for a customer regardless of his class is

\[
E(t) = E(w) - E(w_q).
\]  

(157)

It should be noted that all of the preceding calculations
are appropriate for a system that uses any sequencing rule to
determine the order of service.

**Example Two**

To illustrate the model a small numerical example is
given. The data used for this example have no significance
beyond computational simplicity. In general, any data
complying with previously stated restrictions are appropri­
ate.

The small system under consideration serves two distinct
classes of customers, class 1 and class 2, i.e., \( k = 2 \). The
class 1 customers take non-preemptive priority over the class
2 customers. The mean arrival rate of both classes of custom­
ers is one-half customer per minute, i.e., \( r_1 = r_2 = 1/2 \).
The total arrival rate is one customer per minute, i.e.,
\( r = 1 \). A maximum of two customers are allowed to wait in the
queue at any time, i.e., \( N = 2 \). This implies that a maximum
of three customers are in the system at any point in time.
The model is therefore a \( (k = 2, N = 2) \) non-preemptive priori­
ty model.

The service times are interactive and follow gamma
distributions with known parameters, \( c_s \) and \( b_s \). The gamma
service time density function is

\[
f_s(t) = \frac{b_s^{c_s}}{(c_s-1)!} t^{c_s-1} e^{-b_s t}, \quad t > 0, \quad \text{s=ij, where } i = 0,1,2, \\
\text{and } j = 1,2.
\]
The parameters of the service time density functions for the six possible types of interactive service are given in Table 7. Recall that the variable \( s = 0j, j = 1, 2 \), indicates a service time for a class \( j \) customer that receives service after the server has been idle for a non-zero period of time. The variable \( s = ij, i = 1, 2, \text{ and } j = 1, 2 \), indicates a service time for a class \( j \) customer that receives service immediately after a class \( i \) customer receives service. For example, a class 1 customer that receives service after a class 2 customer \((s = 21)\) has a service time that follows the gamma density function with parameters \( b_{21} = 1 \) and \( c_{21} = 3 \). The mean service time is \( m_{21} = 3 \) minutes.

The state descriptions for the imbedded Markov chain are \( X = (x_0, x_1, x_2) \) where \( x_0 = 1, 2, \text{ and } x_1 \text{ and } x_2 \) take on non-negative integer values with the restriction that \( x \leq 2 \). For the \((k = 2, N = 2)\) model \( \bar{x} = x_1 + x_2 \) and indicates the total number of customers waiting for service. To determine the IMC transition matrix for this model it is necessary to find the arrival probabilities.

**Arrival Probabilities**

Before further development of the model, the service time arrival probabilities, \( A(a_1, a_2; s) \) and \( \bar{A}(a_1, a_2; s) \), and the random interruption probabilities, \( A'(a_1, a_2; s) \) and \( \bar{A}'(a_1, a_2; s) \), are calculated. These probabilities are determined by first calculating the total arrival probabilities \( A(a; s), \bar{A}(a; s), A'(a; s), \text{ and } \bar{A}'(a; s) \), where \( a = a_1 + a_2 \).
<table>
<thead>
<tr>
<th>Type of Service $s = ij$</th>
<th>Service Time Density Function Parameters</th>
<th>Mean Service Time (min.) $m_s = c_s/b_s$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b_s$</td>
<td>$c_s$</td>
</tr>
<tr>
<td>01</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>21</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>02</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>22</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 7. Service Time Density Functions for the $(k = 2, N = 2)$ Model of Example Two.
The procedure used to calculate these total arrival probabilities is given in Table 1 and was demonstrated in Example One. The total arrival probabilities associated with a gamma service time density function with parameters \( b_s = 1 \) and \( c_s = 2 \) are given in Table 2. Reference to Table 7 indicates that these probabilities are appropriate for the \( s = 11 \) and \( s = 02 \) types of service as the service time density function parameters and the total arrival rate are the same. The total arrival probabilities associated with a gamma service time density function with parameters \( b_s = 1 \) and \( c_s = 3 \) are given in Table 3. These probabilities are appropriate for the \( s = 11 \) and \( s = 02 \) types of service as the service time density function parameters and the total arrival rate are the same. The total arrival probabilities for the \( s = 01 \) and \( s = 12 \) types of service are found by the same methods demonstrated in Example One and are not displayed. The total arrival probabilities for the six types of service for the \((k = 2, N = 2)\) model are given in Table 8.

The total arrival probabilities describe the event that a total of exactly a customers enter the service facility during the service time or the random interruption portion of the service time of a type \( s \) service. For example, consider a class 1 customer that receives service immediately after a class 2 customer, \( s = 21 \). The probability that exactly one customer enters the system during this service time, given that the system does not truncate during the
<table>
<thead>
<tr>
<th>Type of Service</th>
<th>Number of Arrivals</th>
<th>Total Arrival Probability</th>
<th>Random Interruption Time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Service Time</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$A(a; s)$</td>
<td>$\overline{A}(a; s)$</td>
</tr>
<tr>
<td>$s = 01$</td>
<td>0</td>
<td>4/8</td>
<td>----</td>
</tr>
<tr>
<td>$s = 12$</td>
<td>1</td>
<td>2/8</td>
<td>4/8</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>----</td>
<td>2/8</td>
</tr>
<tr>
<td>$s = 11$</td>
<td>0</td>
<td>4/16</td>
<td>----</td>
</tr>
<tr>
<td>$s = 02$</td>
<td>1</td>
<td>4/16</td>
<td>12/16</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>----</td>
<td>8/16</td>
</tr>
<tr>
<td>$s = 21$</td>
<td>0</td>
<td>2/16</td>
<td>----</td>
</tr>
<tr>
<td>$s = 22$</td>
<td>1</td>
<td>3/16</td>
<td>14/16</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>----</td>
<td>11/16</td>
</tr>
</tbody>
</table>

Table 8. Total Arrival Probabilities for the ($k = 2$, $N = 2$) Model of Example Two.
service time is $A(l; 21) = 3/16$. Given that the system truncates during the service time, the probability of the event is $A(l; 21) = 14/16$. The probability that exactly one customer enters the system during the random interruption portion of the service time, given that the system does not truncate during this period, is $A'(l; 21) = 22/96$. Given that the system truncates during the random interruption portion of the service time, the probability of the event is $A'(l; 21) = 68/96$.

The total arrival probabilities are modified in order to determine the joint probability that exactly $a_1$ class 1 customers and exactly $a_2$ class 2 customers enter the system under the conditions stated in Table 5. For example, consider a class 1 customer that receives service immediately after a class 2 customer, $s = 21$. The probability that exactly one class 1 customer, $a_1 = 1$, and zero class two customers, $a_2 = 0$, enter the system during the service time, given that the system does not truncate during the service time, is $A(1, 0; 21)$. From Table 5 the non-truncated service time arrival probability is

$$A(a_1, a_2; s) = (a_1 a_2)(\frac{r_1}{r})^{a_1}(\frac{r_2}{r})^{a_2}A(a; s), a = 0, 1. \quad (159)$$

For this example $a = a_1 + a_2 = 1$, and the desired probability is

$$A(1, 0; 21) = (1/10)(1/2)^1 (1/2)^0 A(1; 21) \quad (160)$$

$$= \frac{1}{10} \cdot (1/2) (1) (3/16)$$
Given that the system truncates during the service time, the probability of the event is \( \bar{A}(1, 0; 21) \). From Table 5 the truncated service time arrival probability is

\[
\bar{A}(a_1, a_2; s) = \sum_{j \in J} (a_1 - d_1 j, a_2 - d_2 j) \left( \frac{r_1}{r} \right)^{a_1} \left( \frac{r_2}{r} \right)^{a_2 - 1} \bar{A}(a; s), \quad a = 1, 2.
\]  

(161)

For this example, the set \( J \) is

\[
J = \{ i; a_i > 0, i = 1, 2 \}
\]  

(162)

\[
J = \{ i = 1 \}.
\]

The probability of the event is

\[
\bar{A}(1, 0; 21) = \left( \frac{1}{1-1} \right) \left( \frac{1}{0-0} \right) \left( \frac{1}{1} \right) \left( \frac{1}{0} \right) \bar{A}(1; 21)
\]  

(163)

\[
= \frac{1}{0} \cdot \frac{1}{0} \cdot (1/2) \cdot (1) \cdot (14/16)
\]

\[
= 14/32.
\]

The other arrival probabilities are determined by these same methods. All the probabilities of interest are given in Table 9. The imbedded Markov chain transition matrix, \( B \), and the conversion matrix, \( B' \), are evaluated using these probabilities.

**IMC Steady State Probabilities**

Using the service time arrival probabilities, \( A(a_1, a_2; s) \) and \( \bar{A}(a_1, a_2; s) \) the IMC transition matrix, \( B \), is evaluated. The elements of the matrix are assigned values according to the five mutually exclusive, collectively exhaustive transition cases defined in Table 6. The procedures of Table 6
| Type of Service | Number of Arrivals $a_1, a_2$ | | Arrival Probability | | Random Interruption Time |
|---|---|---|---|---|
| | | Service Time | | Interruption Time |
| | $A(a_1,a_2; s)$ | $A'(a_1,a_2; s)$ | $A'(a_1,a_2; s)$ |
| $s = 01$ | 0,0 | 4/8 | ---- | 4/8 | ---- |
| | 1,0 | 1/8 | 2/8 | 1/8 | 2/8 |
| | 0,1 | 1/8 | 2/8 | 1/8 | 2/8 |
| | 2,0 | ---- | 1/16 | ---- | 1/16 |
| | 0,2 | ---- | 1/16 | ---- | 1/16 |
| | 1,1 | ---- | 1/8 | ---- | 1/8 |
| $s = 12$ | 0,0 | 4/16 | ---- | 6/16 | ---- |
| | 1,0 | 2/16 | 6/16 | 7/16 | 5/16 |
| | 0,1 | 2/16 | 6/16 | 7/16 | 5/16 |
| | 2,0 | ---- | 2/16 | ---- | 3/32 |
| | 0,2 | ---- | 7/16 | ---- | 3/32 |
| | 1,1 | ---- | 4/16 | ---- | 3/16 |
| $s = 11$ | 0,0 | 4/32 | ---- | 28/96 | ---- |
| | 1,0 | 3/32 | 14/32 | 11/96 | 34/96 |
| | 0,1 | 3/32 | 14/32 | 11/96 | 34/96 |
| | 2,0 | ---- | 11/64 | ---- | 23/192 |
| | 0,2 | ---- | 11/64 | ---- | 23/192 |
| | 1,1 | ---- | 11/32 | ---- | 23/96 |

Table 9. Arrival Probabilities for the $(k=2, N=2)$ Model of Example Two.
have been previously demonstrated for the \((k = 2, N = 2)\) non-preemptive priority model, and the resultant IMC transition matrix is displayed in Figure 7. Evaluating the non-zero elements of this matrix with the service time arrival probabilities of Table 9 yields the IMC transition matrix for this example problem. This transition matrix is displayed in Figure 9.

The IMC steady state probabilities, \(g(X)\), are found by solving the set of linear equations

\[
GB = G,
\]

and

\[
G1 = 1,
\]

where

\[
G = \{g(X)\},
\]

and \(1\) is a column vector with all elements equal to one. The solution of these equations is accomplished by standard methods and is not demonstrated. The resultant imbedded Markov chain steady state probabilities are given in Table 11 on page 181. The interpretation of the IMC steady state probabilities is straightforward. For example, \(g(2,0,1) = 0.1129\). This means that after equilibrium conditions are attained, the probability is 0.1129 that, at any randomly selected service completion epoch a class 2 customer is completing service, zero class 1 customers are awaiting service, and one class 2 customer is awaiting service.
To
\( (x_0'', x_1'', x_2'') \)

\[ x_0, x_1, x_2 \]

From
\( (x_0', x_1', x_2') \)

<table>
<thead>
<tr>
<th></th>
<th>(1, 0, 0)</th>
<th>(2, 0, 0)</th>
<th>(1, 1, 0)</th>
<th>(2, 1, 0)</th>
<th>(1, 0, 1)</th>
<th>(2, 0, 1)</th>
<th>(1, 2, 0)</th>
<th>(2, 2, 0)</th>
<th>(1, 0, 2)</th>
<th>(2, 0, 2)</th>
<th>(1, 1, 1)</th>
<th>(2, 1, 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 0, 0)</td>
<td>( \frac{1}{2} ) () \frac{4}{8} )</td>
<td>( \frac{1}{2} ) () \frac{4}{16} )</td>
<td>( \frac{1}{2} ) () \frac{1}{8} )</td>
<td>( \frac{1}{2} ) () \frac{1}{16} )</td>
<td>( \frac{1}{2} ) () \frac{1}{8} )</td>
<td>( \frac{1}{2} ) () \frac{1}{16} )</td>
<td>( \frac{1}{2} ) () \frac{1}{16} )</td>
<td>( \frac{1}{2} ) () \frac{1}{16} )</td>
<td>( \frac{1}{2} ) () \frac{1}{8} )</td>
<td>( \frac{1}{2} ) () \frac{1}{16} )</td>
<td>( \frac{1}{2} ) () \frac{1}{8} )</td>
<td>( \frac{1}{2} ) () \frac{1}{16} )</td>
</tr>
<tr>
<td>(2, 0, 0)</td>
<td>( \frac{1}{2} ) () \frac{4}{8} )</td>
<td>( \frac{1}{2} ) () \frac{4}{16} )</td>
<td>( \frac{1}{2} ) () \frac{1}{8} )</td>
<td>( \frac{1}{2} ) () \frac{1}{16} )</td>
<td>( \frac{1}{2} ) () \frac{1}{8} )</td>
<td>( \frac{1}{2} ) () \frac{1}{16} )</td>
<td>( \frac{1}{2} ) () \frac{1}{16} )</td>
<td>( \frac{1}{2} ) () \frac{1}{16} )</td>
<td>( \frac{1}{2} ) () \frac{1}{8} )</td>
<td>( \frac{1}{2} ) () \frac{1}{16} )</td>
<td>( \frac{1}{2} ) () \frac{1}{8} )</td>
<td>( \frac{1}{2} ) () \frac{1}{16} )</td>
</tr>
<tr>
<td>(1, 1, 0)</td>
<td>( \frac{4}{16} )</td>
<td>0</td>
<td>( \frac{2}{16} )</td>
<td>0</td>
<td>( \frac{2}{16} )</td>
<td>0</td>
<td>( \frac{2}{16} )</td>
<td>0</td>
<td>( \frac{4}{16} )</td>
<td>0</td>
<td>( \frac{4}{16} )</td>
<td>0</td>
</tr>
<tr>
<td>(2, 1, 0)</td>
<td>( \frac{4}{32} )</td>
<td>0</td>
<td>( \frac{3}{32} )</td>
<td>0</td>
<td>( \frac{3}{32} )</td>
<td>0</td>
<td>( \frac{11}{64} )</td>
<td>0</td>
<td>( \frac{11}{64} )</td>
<td>0</td>
<td>( \frac{11}{32} )</td>
<td>0</td>
</tr>
<tr>
<td>(1, 0, 1)</td>
<td>0</td>
<td>( \frac{4}{8} )</td>
<td>0</td>
<td>( \frac{1}{8} )</td>
<td>0</td>
<td>( \frac{1}{8} )</td>
<td>0</td>
<td>( \frac{1}{16} )</td>
<td>0</td>
<td>( \frac{1}{16} )</td>
<td>0</td>
<td>( \frac{1}{8} )</td>
</tr>
<tr>
<td>(2, 0, 1)</td>
<td>0</td>
<td>( \frac{4}{32} )</td>
<td>0</td>
<td>( \frac{3}{32} )</td>
<td>0</td>
<td>( \frac{3}{32} )</td>
<td>0</td>
<td>( \frac{11}{64} )</td>
<td>0</td>
<td>( \frac{11}{64} )</td>
<td>0</td>
<td>( \frac{11}{32} )</td>
</tr>
<tr>
<td>(1, 2, 0)</td>
<td>0</td>
<td>0</td>
<td>( \frac{4}{16} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \frac{6}{16} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \frac{6}{16} )</td>
<td>0</td>
</tr>
<tr>
<td>(2, 2, 0)</td>
<td>0</td>
<td>0</td>
<td>( \frac{4}{32} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \frac{14}{32} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \frac{14}{32} )</td>
<td>0</td>
</tr>
<tr>
<td>(1, 0, 2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \frac{4}{8} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \frac{2}{8} )</td>
<td>0</td>
<td>( \frac{2}{8} )</td>
<td>0</td>
</tr>
<tr>
<td>(2, 0, 2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \frac{4}{32} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \frac{14}{32} )</td>
<td>0</td>
<td>( \frac{14}{32} )</td>
<td>0</td>
</tr>
<tr>
<td>(1, 1, 1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \frac{4}{16} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \frac{6}{16} )</td>
<td>0</td>
<td>( \frac{6}{16} )</td>
<td>0</td>
<td>( \frac{6}{16} )</td>
</tr>
<tr>
<td>(2, 1, 1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \frac{4}{32} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \frac{14}{32} )</td>
<td>0</td>
<td>( \frac{14}{32} )</td>
<td>0</td>
<td>( \frac{14}{32} )</td>
</tr>
</tbody>
</table>

Figure 9. IMC Transition Matrix for the \((k=2, N=2)\) Non-preemptive Priority Model of Example Two.
SMP Steady State Probabilities

The SMP steady state probabilities, \( h(Y) \), can be derived directly from the IMC steady state probabilities, \( g(X) \). The simplified expressions for the SMP steady state probabilities depend only on the IMC steady state probabilities, therefore it is not necessary to explicitly calculate the augmented IMC steady state probabilities. The simplified expressions for the SMP steady state probabilities are

\[
\begin{align*}
    h(Y) = \begin{cases} 
    \left( \sum_{X} g(X) \right) \frac{r_{i} m_{01}}{(rd)}, & Y = e_i, i = 1,2, \\
    \frac{g(X)}{(rd)}, & Y = X, \text{ where } \overline{x} = 0, \\
    \frac{g(X)}{d \sum_{X} m_{0w}}, & Y = X, \text{ where } \overline{x} > 0,
    \end{cases}
\end{align*}
\]

where

\[
d = \left( \frac{1}{2} \sum_{i=1}^{2} \frac{r_{i} m_{01}}{r} + \frac{1}{r} \right) \sum_{X} g(X) + \sum_{X} g(X) m_{0w},
\]

\[
s.t. \quad \overline{x} = 0 \quad \text{s.t.} \quad \overline{x} > 0
\]

Before calculating these probabilities the values of \( w \) are found.

Recall that the class of customer to serve next, given that a non-empty service completion epoch has just occurred, is determined by the variable \( w \) as

\[
w = \min_{i} (i; x_i > 0, i = 1,2), \quad \overline{x} > 0.
\]

The values of \( w \) for the non-empty service completion epochs of the \((k = 2, N=2)\) non-preemptive priority model are given in Table 10. Reference to Table 10 indicates that for the first eight non-empty service completion epochs listed
State Description for Non-empty Service Completion Epoch

\[ x = (x_0, x_1, x_2) \]

| \((1, 1, 0)\) | 1 |
| \((2, 1, 0)\) | 1 |
| \((1, 0, 1)\) | 2 |
| \((2, 0, 1)\) | 2 |
| \((1, 2, 0)\) | 1 |
| \((2, 2, 0)\) | 1 |
| \((1, 0, 2)\) | 2 |
| \((2, 0, 2)\) | 2 |
| \((1, 1, 1)\) | 1 |
| \((2, 1, 1)\) | 1 |

Table 10. Class of Customer to Serve Next After a Non-empty Service Completion Epoch has Occurred.

The values of \(w\) for the \((k = 2, N = 2)\) Non-preemptive Priority Model.
(X = (1,1,0) to X = (2,0,2)), there is no choice as to the class of customer to serve next. Only one class of customer is available, and this class of customer is served. For example, for the system state description X = (2,0,2) only class 2 customers are available, so a class 2 customer receives service, i.e., w = 2. For the last two non-empty service completion epochs listed (X = (1,1,1) and X = (2,1,1)), there is a choice as to the class of customer to serve next. For both of these states the non-preemptive priority sequencing rule requires that a class 1 customer receive service next, i.e., w = 1.

Using the values of g(X) given in Table 11, the values of w given in Table 10, and the values of m_s given in Table 7, the SMP steady state probabilities can be determined. The value of the normalizing constant d is determined from the fact that the sum of the SMP steady state probabilities must equal one. The calculations are illustrated with several examples.

The SMP steady state probability associated with the class 1 customer idle service initiation epoch is

\[
h(Y= e_1) = \sum_{X}^{} g(X) \frac{r_1 m_{01}}{(rd)}
\]

\[
\text{s.t. } X = 0
\]

\[
= \sum_{X_0=0}^{2} g(x_0, 0, 0) \frac{r_1 m_{01}}{(rd)}
\]

\[
= (g(1,0,0) + g(2,0,0)) \frac{r_1 m_{01}}{(rd)}
\]

\[
= (0.0342 + 0.0644) \frac{1}{2} \frac{1}{(1)d}
\]

\[
= 0.0493/d.
\]
This means that after equilibrium conditions are attained, the probability is \( \frac{0.0493}{d} \) that at the epoch immediately preceding a randomly selected point in time a class 1 customer started to receive service after the server had been idle for a non-zero length of time.

The SMP steady state probability associated with the empty service completion epoch having state description \( X = (1,0,0) \) is

\[
  h(Y = (1,0,0)) = \frac{g(1,0,0)}{rd}
  = \frac{0.0342}{(1)d}
  = \frac{0.0342}{d}.
\]

This means that after equilibrium conditions are attained, the probability is \( \frac{0.0342}{d} \) that at the epoch immediately preceding a randomly selected point in time a class 1 customer has just completed service and there are no customers awaiting service.

The SMP steady state probability associated with the non-empty service completion epoch having state description \( X = (2,0,1) \) is

\[
  h(Y = (2,0,1)) = \frac{g(2,0,1)}{m_{2w}/d}.
\]

From Table 10, \( w = 2 \) when \( X = (2,0,1) \). The SMP steady state probability is

\[
  h(Y = (2,0,1)) = \frac{g(2,0,1)}{m_{2w}/d}
  = \frac{(0.1129)(3)}{d}
  = \frac{0.3387}{d}.
\]

This means that after equilibrium conditions are attained,
the probability is \(0.3387/d\) that at the epoch immediately preceding a randomly selected point in time a class 2 customer is completing service, zero class 1 customers are awaiting service, and one class 2 customer is awaiting service.

The other SMP steady state probabilities are found by using these same procedures. The value of the normalizing constant is \(d = 2.2649\). The SMP steady state probabilities are given in Table 11.

**RPT Steady State Probabilities**

The SMP steady state probabilities are converted to the random point in time steady state probabilities, \(p(z)\), by using the random interruption time arrival probabilities, \(A'(a_1, a_2; s)\) and \(\overline{A}'(a_1, a_2; s)\). The appropriate equations are

\[
p(z) = \begin{cases} \sum_{Y=X'} \sum_{s.t. x'=0} h(Y) & z = 0, \\ \sum_{Y=X'} \sum_{s.t. x'>0} h(Y)b'(Y,X) + h(e_{x_0})A'(x_1, x_2; 0x_0), & z = X, \end{cases} \quad (171)
\]

The conversion matrix, \(B'\), for the \((k = 2, N = 2)\) non-preemptive priority model was previously developed and is displayed in Figure 8. Evaluating the non-zero elements of this matrix with the random interruption time arrival probabilities of Table 9 yields the conversion matrix for this
example problem. This conversion matrix is displayed in
Figure 10. Using the values of \( h(Y) \) given in Table 11, the
values of \( b^\prime(Y,X) \) given by the \( B^\prime \) matrix of Figure 10, and
the values of \( A^\prime(a_1, a_2, s) \) and \( A^\prime(a_1, a_2, s) \) given in Table
9 the RPT steady state probabilities can be determined. The
calculations are illustrated with several examples.

The RPT steady state probability that the system is
completely empty is

\[
p(Z = 0) = \sum_{Y=X'} h(Y)
\]
\[
\text{s.t.} \bar{X}' = 0
\]
\[
= \sum_{x_0'} h(x_0', 0, 0)
\]
\[
= h(1,0,0) + h(2,0,0)
\]
\[
= 0.0151 + 0.0284
\]
\[
= 0.0435.
\]

This means that after equilibrium conditions are attained,
the probability is 0.0435 that at a randomly selected point
in time the system is empty.

The RPT steady state probability that a class 2 customer
is receiving service, one class 1 customer is awaiting
service and zero class 2 customers are awaiting service is
given by

\[
p(Z = X = (2,1,0)) = \sum_{Y=X'} h(Y)b^\prime(Y,X) + h(e_{x_0})A^\prime(x_1, x_2, 0; x_0).
\]
\[
\text{s.t.} \bar{X} > 0
\]

(173)

Note that the total number of customers awaiting service,
\( \bar{x} = x_1 + x_2 = 1 \), is less than the capacity of the waiting
Figure 10. Conversion Matrix for the (k=2, N=2) Non-preemptive Priority Model of Example Two.
room, N = 2. Thus the system is not truncated. The probability calculation is continued by substituting the appropriate values of \( b'(Y, (2,1,0)) \) from the conversion matrix of Figure 8. Only the non-zero values of \( b'(Y, (2,1,0)) \) are considered. The RPT steady state probability is

\[
p(Z = (2,1,0)) = \sum_{Y=X'} h(Y) b'(Y, (2,1,0)) + h(e_2) A'(1,0;02) \quad (174)
\]

Before evaluating this probability the three terms in the above expression are interpreted.

The first term, \( h(1,0,1)A'(1,0;12) \), is the joint probability that at the preceding epoch a class 1 customer completed service and the queue contained zero class 1 customers and one class 2 customer, together with the event that one class 1 customer and zero class 2 customers entered the system during the random interruption portion of the service time. The service time is for a class 2 customer preceded by a class 1 customer. The second term, \( h(2,0,1)A'(1,0;12) \), is the joint probability that at the preceding epoch a class 2 customer completed service and the queue contained zero class one customers and one class 2 customer, together with the event that one class 1 customer and zero class 2 customers entered the system during the random interruption portion of
the service time. The service time is for a class 2 customer preceded by a class 2 customer. The third term, \( h(e_2) A'(1, 0; 02) \), is the joint probability that at the preceding epoch a class 2 customer initiated service after the server had been idle for some non-zero period of time, together with the event that one class 1 customer and zero class 2 customers entered the system during the random interruption portion of the service time. The service time is for a class 2 customer that receives service after the server has been idle for some non-zero period of time. These three mutually exclusive events are the only events that allow the system to be in state \( X = (2,1,0) \) at a random point in time. Therefore the sum of the probabilities for these three events yields the probability that the system is in state \( X = (2,1,0) \) at a random point in time.

Evaluating this RPT steady state probability yields

\[
p(Z = (2,1,0)) = (0.0335)(1/8) + (0.1495)(11/96) + (0.0435)(1/8)
\]

\[= 0.0267.\]

This means that after equilibrium conditions are attained, the probability is 0.0267 that at a randomly selected point in time a class 2 customer is receiving service, one class 1 customer is awaiting service, and zero class 2 customers are awaiting service. The other RPT steady state probabilities are found by using these same procedures. The RPT steady state probabilities are given in Table 11.
<table>
<thead>
<tr>
<th>State Description</th>
<th>Imbedded Markov Chain $g(X); X = (x_0, x_1, x_2)$</th>
<th>Semi-Markov Process $h(Y); Y = e_1, e_2, X$</th>
<th>Random Point in Time $p(Z); Z = 0, X$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$</td>
<td>--</td>
<td>0.0218</td>
<td>--</td>
</tr>
<tr>
<td>$e_2$</td>
<td>--</td>
<td>0.0435</td>
<td>--</td>
</tr>
<tr>
<td>$0$</td>
<td>--</td>
<td>--</td>
<td>0.0435</td>
</tr>
<tr>
<td>$(x_0, x_1, x_2)$</td>
<td>0.0342</td>
<td>0.0151</td>
<td>0.0294</td>
</tr>
<tr>
<td>$(1, 0, 0)$</td>
<td>0.0644</td>
<td>0.0284</td>
<td>0.0767</td>
</tr>
<tr>
<td>$(2, 0, 0)$</td>
<td>0.0252</td>
<td>0.0223</td>
<td>0.0339</td>
</tr>
<tr>
<td>$(1, 1, 0)$</td>
<td>0.0262</td>
<td>0.0347</td>
<td>0.0267</td>
</tr>
<tr>
<td>$(2, 1, 0)$</td>
<td>0.0759</td>
<td>0.0335</td>
<td>0.1271</td>
</tr>
<tr>
<td>$(1, 0, 1)$</td>
<td>0.1129</td>
<td>0.1495</td>
<td>0.1041</td>
</tr>
<tr>
<td>$(2, 0, 1)$</td>
<td>0.0384</td>
<td>0.0339</td>
<td>0.0324</td>
</tr>
<tr>
<td>$(1, 2, 0)$</td>
<td>0.0303</td>
<td>0.0401</td>
<td>0.0241</td>
</tr>
<tr>
<td>$(2, 2, 0)$</td>
<td>0.1440</td>
<td>0.0636</td>
<td>0.1275</td>
</tr>
<tr>
<td>$(1, 0, 2)$</td>
<td>0.1179</td>
<td>0.1562</td>
<td>0.0953</td>
</tr>
<tr>
<td>$(2, 0, 2)$</td>
<td>0.1824</td>
<td>0.1611</td>
<td>0.1599</td>
</tr>
<tr>
<td>$(1, 1, 1)$</td>
<td>0.1482</td>
<td>0.1963</td>
<td>0.1194</td>
</tr>
<tr>
<td>$(2, 1, 1)$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 11. Steady State Probabilities for the $(k = 2, N = 2)$ Non-preemptive Priority Model of Example Two.
Queue Length Probabilities

The marginal probabilities of having exactly \( j \) class 1 customers in the system (in the queue) after equilibrium conditions are attained, are readily obtained from the RPT steady state probabilities, \( p(Z) \). The calculations are illustrated by several examples.

The probability mass function for the number of class 1 customers in the system is \( \text{pr}(n_1 = j), \ j = 0, 1, 2, 3 \). The probabilities are

\[
\text{pr}(n_1 = 0) = p(Z = 0) + \sum_{Z=X} p(Z) \\
\quad \text{s.t. } x_0 \neq 1 \text{ and } x_1 = 0
\]

\[
= p(0) + p(2,0,0) + p(2,0,1) + p(2,0,2) \\
= 0.0435 + 0.0767 + 0.1041 + 0.0953 \\
= 0.3196,
\]

\[
\text{pr}(n_1 = 1) = \sum_{Z=X} p(Z) + \sum_{Z=X} p(Z) \\
\quad \text{s.t. } x_0 = 1 \text{ s.t. } x_0 \neq 1 \\
\quad \text{and } x_1 = 0 \text{ and } x_1 = 0
\]

\[
= p(1,0,0) + p(1,0,1) + p(1,0,2) + p(2,1,0) \\
\quad + p(2,1,1) \\
= 0.0294 + 0.1271 + 0.1275 + 0.0267 + 0.1194 \\
= 0.4301,
\]

\[
\text{pr}(n_1 = 2) = \sum_{Z=X} p(Z) + \sum_{Z=X} p(Z) \\
\quad \text{s.t. } x_0 = 1 \text{ s.t. } x_0 \neq 1 \\
\quad \text{and } x_1 = 1 \text{ and } x_1 = 2
\]

\[
= p(1,1,0) + p(1,1,1) + p(2,2,0) \\
= 0.0339 + 0.1599 + 0.0241
\]
\[ \text{pr}(n_1 = 2) = 0.2179, \]
\[ \text{pr}(n_1 = 3) = p(1,2,0), \quad (179) \]
\[ = 0.0324. \]

The probability mass function for the number of class 1 customers in the queue is \( \text{pr}(n_{q_1} = j), j = 0, 1, 2. \) The probabilities are
\[
\text{pr}(n_{q_1} = 0) = p(Z=0) + \Sigma_{Z=X} p(Z) \quad \text{s.t. } x_1=0 \\
= p(0) + p(1,0,0) + p(2,0,0) + p(1,0,1) + p(2,0,1) \\
+ p(1,0,2) + p(2,0,2) \\
= 0.0435 + 0.0294 + 0.0767 + 0.1271 + 0.1041 \\
+ 0.1275 + 0.0953 \\
= 0.6036.
\]
\[
\text{pr}(n_{q_1} = 1) = \Sigma_{Z=X} p(Z) \quad \text{s.t. } x_1=1 \\
= p(1,1,0) + p(2,1,0) + p(1,1,1) + p(2,1,1) \\
= 0.0339 + 0.0267 + 0.1599 + 0.1194 \\
= 0.3399,
\]
\[
\text{pr}(n_{q_1} = 2) = \Sigma_{Z=X} p(Z) \quad \text{s.t. } x_1=2 \\
= p(1,2,0) + p(2,2,0) \\
= 0.0324 + 0.0241 \\
= 0.0565.
\]

The equivalent probabilities for class 2 customers, \( \text{pr}(n_2=j) \) and \( \text{pr}(n_{q_2}=j), \) are found by using the above methods. The equivalent probabilities for the overall customer
population, \( pr(n=j) \) and \( pr(n_q=j) \), are found by using different methods, but the calculations are straightforward and are not demonstrated. The probability mass functions for the number of customers in the system and the number of customers in the queue for class 1 customers, class 2 customers, and the overall customer population are given in Table 12.

**Measures of Performance**

Using the probability mass functions given in Table 12 the measures of system performance are evaluated. The procedures are illustrated by calculating the measures of system performance for class 1 customers. The effective mean arrival rate of class 1 customers is

\[
\lambda_{eff1} = \lambda_1 (1 - pr(n = N + 1)) \quad (183)
\]

\[
= \lambda_1 (1 - pr(n = 3))
\]

\[
= (1/2) (1 - 0.5586)
\]

\[
= 0.2207.
\]

The expected number of class 1 customers in the system is

\[
E(n_1) = \sum_{j=0}^{\infty} j \cdot pr(n_1 = j) \quad (184)
\]

\[
= 0.9631
\]

The expected number of class 1 customers in the queue is

\[
E(n_{q1}) = \sum_{j=0}^{2} j \cdot pr(n_{q1} = j) \quad (185)
\]

\[
= 0.4529
\]
### Table 12. Probability Mass Functions for the Number of Customers in the System and the Number of Customers in the Queue for the \( k = 2, N = 2 \) Non-preemptive Priority Model of Example Two.

<table>
<thead>
<tr>
<th>( j )</th>
<th>( n_1 = j )</th>
<th>( n_2 = j )</th>
<th>Overall</th>
<th>( nq_1 = j )</th>
<th>( nq_2 = j )</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.3196</td>
<td>0.1392</td>
<td>0.0435</td>
<td>0.6036</td>
<td>0.2667</td>
<td>0.1496</td>
</tr>
<tr>
<td>1</td>
<td>0.4301</td>
<td>0.4145</td>
<td>0.1061</td>
<td>0.3399</td>
<td>0.5105</td>
<td>0.2918</td>
</tr>
<tr>
<td>2</td>
<td>0.2179</td>
<td>0.3510</td>
<td>0.2918</td>
<td>0.0565</td>
<td>0.2228</td>
<td>0.5586</td>
</tr>
<tr>
<td>3</td>
<td>0.0324</td>
<td>0.0953</td>
<td>0.5586</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
</tbody>
</table>
The expected waiting time in the system for class 1 customers is
\[ E(w_1) = E(n_1)/r_{eff1} \]
\[ = 0.9631/0.2207 \]
\[ = 4.3638 \text{ minutes.} \]

The expected waiting time in the queue for class 1 customers is
\[ E(w_q) = E(n_{q1})/r_{eff1} \]
\[ = 0.4529/0.2207 \]
\[ = 2.0521 \text{ minutes.} \]

The expected service time for class 1 customers is
\[ E(t_1) = E(w_1) - E(w_{q1}) \]
\[ = 4.3638 - 2.0521 \]
\[ = 2.3117 \text{ minutes.} \]

The measures of system performance for class 2 customers and the overall customer population are calculated by the same procedures demonstrated above. The measures of system performance are given in Table 13.

Summary

This chapter has provided a complete methodology for determining the steady state measures of system performance associated with a special type of M/G/1 queueing system. The modeling assumptions for this system are given below.

Model Assumptions

1. All customers that enter the system are classified as belonging to one of \( k \) homogeneous customer classes, where \( k \) is a known finite integer.
<table>
<thead>
<tr>
<th></th>
<th>Effective Mean Arrival Rate $r_{eff}$</th>
<th>Expected Number Customers in System $E(n_i)$</th>
<th>Expected Number Customers in Queue $E(n_{qi})$</th>
<th>Expected Waiting Time in System $E(w_i)(\text{min})$</th>
<th>Expected Waiting Time in Queue $E(w_{qi})(\text{min})$</th>
<th>Expected Service Time $E(t_i)(\text{min})$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Class 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customers $i = 1$</td>
<td>0.2207</td>
<td>0.9631</td>
<td>0.4529</td>
<td>4.3638</td>
<td>2.0521</td>
<td>2.3117</td>
</tr>
<tr>
<td><strong>Class 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customers $i = 2$</td>
<td>0.2207</td>
<td>1.4024</td>
<td>0.9561</td>
<td>6.3543</td>
<td>4.3321</td>
<td>2.0222</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td>0.4414</td>
<td>2.3655</td>
<td>1.4090</td>
<td>5.3591</td>
<td>3.1921</td>
<td>2.1670</td>
</tr>
</tbody>
</table>

**Table 13.** Measures of System Performance for the $(k = 2, N = 2)$ Non-preemptive Priority Model of Example Two.
2. Class \( i \) customers enter the system at random points in time with known mean rate \( r_i, i=1,2,\ldots,k \). The total mean arrival rate is \( r, \) where \( r = \sum_{i=1}^{k} r_i. \)

3. All customers are served by a single service facility.

4. The service time is influenced by two factors:
   a) the class of customer receiving service (a class \( j \) customer) and b) the class of customer, if any, that completed service immediately before the class \( j \) customer's service was initiated (a class \( i \) customer). These interactive service times are random variables with known general probability density functions \( f_s(t), \) where \( s = ij, i=0,1,\ldots,k, \) and \( j=1,2,\ldots,k. \) The quantity \( s = 0j \) indicates that the server was idle for a non-zero period of time before the class \( j \) customer received service.

5. The total number of customers awaiting service at any time is required to be less than or equal to a known constant number \( N. \) Customers are not allowed to enter the system when the queue is filled to capacity, i.e., when the total number of customers in the queue is equal to \( N. \)

6. The order of service is determined by a non-preemptive priority sequencing rule. The model developed to describe this system is referred to as a general \((k, N)\) non-preemptive priority model.
In order to determine the steady state measures of system performance, it is necessary to have a knowledge of the steady state probabilities associated with the system. The steady state probabilities of interest are determined by using four distinct procedures. The objective of these procedures is to determine the probability that the system is in state $Z$ at any randomly selected point in time after equilibrium conditions are attained. These probabilities are designated $p(Z)$ and are referred to as random point in time (RPT) steady state probabilities. The state is defined to be $Z = X = (x_0, x_1, ..., x_k)$, if at the selected point in time a class $x_0$ customer is receiving service and $x_1$ class $i$ customers are in the queue, $i = 1, 2, ..., k$. The state is defined to be $Z = 0$ if at the selected point in time the system is completely empty. Methods of determining these steady state probabilities directly are not available. Thus a series of intermediate procedures are undertaken in order to obtain the desired RPT steady state probabilities. The objective of each procedure is given below.

Model Procedures

1. The first procedure determines the probability that the system is in state $X$ at any randomly selected service completion epoch after equilibrium conditions are attained. These probabilities are designated $g(X)$ and are referred as imbedded Markov chain (IMC) steady state probabilities. The state is
defined to be \( X = (x_0, x_1, \ldots, x_k) \), if at the service completion epoch a class \( x_0 \) customer is just completing service and \( x_i \) class \( i \) customers are in the queue, \( i = 1, 2, \ldots, k \).

2. The second procedure modifies the IMC steady state probabilities in order to determine the probability that the system is in state \( Y \) at any randomly selected service completion epoch or idle service initiation epoch after equilibrium conditions are attained. These probabilities are designated \( g'(Y) \) and are referred to as augmented IMC steady state probabilities. The state is defined to be \( Y = e_i \) if the selected epoch is an idle service initiation epoch and a class \( i \) customer initiated service, \( i = 1, 2, \ldots, k \). The state is defined to be \( Y = X = (x_0, x_1, \ldots, x_k) \) if the selected epoch is a service completion epoch.

3. The third procedure modifies the augmented IMC steady state probabilities in order to determine the probability that the system is in state \( Y \) at the epoch immediately preceding a randomly selected point in time after equilibrium conditions are attained. These probabilities are designated \( h(Y) \) and are referred to as semi-Markov process (SMP) steady state probabilities. The state assigned to the randomly selected point in time is \( Y = e_i \) if the
preceding epoch is an idle service initiation epoch and a class i customer initiated service, \( i = 1, 2, ..., k \). The state assigned to the randomly selected point in time is \( Y = X = (x_0, x_1, ..., x_k) \) if the preceding epoch is a service completion epoch.

4. The fourth procedure modifies the SMP steady state probabilities in order to determine the probability that the system is in state \( Z \) at a randomly selected point in time after equilibrium conditions are attained. These probabilities are designated \( p(Z) \) and are the random point in time (RPT) steady state probabilities of ultimate interest.

After the RPT steady state probabilities are obtained the measures of system performance are readily calculated. The measures of system performance are given below.

**Measures of System Performance**

1. Effective mean arrival rate of class \( i \) customers, \( r_{eff_i}, \ i = 1, 2, ..., k \).

2. Effective total mean arrival rate, \( r_{eff} \).

3. Expected number of class \( i \) customers in the system, \( E(n_i), \ i = 1, 2, ..., k \).

4. Expected total number of customers in the system, \( E(n) \).

5. Expected number of class \( i \) customers in the queue, \( E(n_{qi}), \ i = 1, 2, ..., k \).
6. Expected total number of customers in the queue, \( E(n_q) \).

7. Expected waiting time in the system for class \( i \) customers, \( E(w_i) \), \( i = 1, 2, \ldots, k \).

8. Expected waiting time in the system without regard to customer class, \( E(w) \).

9. Expected waiting time in the queue for class \( i \) customers, \( E(w_{qi}) \), \( i = 1, 2, \ldots, k \).

10. Expected waiting time in the queue without regard to customer class, \( E(w_q) \).

11. Expected service time for class \( i \) customers, \( E(t_i) \), \( i = 1, 2, \ldots, k \).

12. Expected service time without regard to customer class, \( E(t) \).

The multiple customer class non-preemptive priority model developed in this chapter is modified in Chapter V to allow the order of service to be determined by a number of additional sequencing rules. All of the model assumptions listed in this summary are retained except one, namely, the sixth assumption which states that the order of service is determined by a non-preemptive priority sequencing rule. It is shown that the model procedures developed in this chapter are appropriate for the additional models if minor changes are made.
CHAPTER V
ADDITIONAL MULTIPLE CUSTOMER CLASS MODELS

Introduction

The non-preemptive priority model developed in the preceding chapter is modified in this chapter to allow the order of service to be determined by a number of other sequencing rules. The model assumptions, model procedures, and measures of system performance listed in the summary of the preceding chapter are appropriate for the extended models. The outline given in the summary is used to guide the discussion in this chapter, but the detailed lists are not restated.

The three sequencing rules considered in this chapter are alternating priority, rotating priority, and first come-first served. The alternating priority sequencing rule operates on the basis that the customer receiving service retains control of the server for his class of customer. When the class of customer receiving service is no longer available, the server receives another class of customer according to a predetermined sequence, subject to availability of customers. The rotating priority rule operates on the basis that the customer receiving service yields control of
the server to another class of customer when his service time is completed. The class of customer receiving control is based on a predetermined sequence, subject to availability of customers. The first come-first served sequencing rule serves customers according to their time of entry into the system, regardless of their customer class.

It should be noted that all of these sequencing rules have features in common with the non-preemptive priority sequencing rule. First, the decision as to which class of customer to serve next is made only when a customer completes service. This implies that a customer's service time is not interrupted once he enters the service facility. Second, the decision as to which class of customer to serve after the system has been idle for a non-zero period of time does not depend on the sequencing rule. This implies that the first customer to enter the system during a system idle period immediately starts to receive service. Third, the order of service within each customer class is first come-first served. The net effect of these common features is that it is necessary to change only one equation of the non-preemptive priority model procedures in order to provide models for the alternating priority and rotating priority sequencing rules. Three equations of the non-preemptive priority model procedures must be changed in order to provide a model for the first-come-first served sequencing rule.
Alternating Priority Sequencing Rule

The alternating priority sequencing rule represents an attempt to minimize the number of changes in the class of customer receiving service. This is accomplished by requiring the customer receiving service to retain control of the server for his class of customer. This sequencing rule allows customers of a particular class to be served until no further customers of this class are waiting for service. When this point is reached, the customer of another class is selected to receive service according to a predetermined sequence, subject to availability of customers. The predetermined sequence is denoted by the assignment of class indices. For example, for a three customer class system the indices are assigned to indicate that the desired sequence: class 1 customers, class 2 customers, class 3 customers, class 1 customers, class 2 customers, class 3 customers, etc. It should be noted that the relative order of the class index, not the absolute value of the index, is the relevant factor in the assignment of indices. For example, if the customers are classified as type a, type b, and type c jobs, then the assignment; type a = class 1, type b = class 2, and type c = class 3; is equivalent to the assignment; type b = class 1, type c = class 2, and type a = class 3. The next task consists of expressing the sequencing rule in terms of the model notation.

The class of customer to serve next, given that a
non-empty service completion epoch has just occurred is indicated by the variable \( w \). The variable \( w \) is defined for the non-preemptive priority rule by Equation (89). The variable \( w \) is defined for the alternating priority rule by an expression involving terms similar to those used in Equation (89).

The state of the system when a customer is completing service is given by the vector \( X' = (x'_0, x'_1, \ldots, x'_k) \). The class of the customer just completing service is \( x'_0 \), and the number of class \( i \) customers waiting in the queue is \( x'_i \), \( i = 1,2,\ldots,k \).

The sequencing rule is appropriate only if there is at least one customer in the queue, \( x' > 0 \). The alternating priority rule requires that a class \( x'_0 \) customer be served next if there is at least one such customer in the queue. If no class \( x'_0 \) customers are available, \( x'_0 = 0 \), then the class of customer to serve next is determined by examining the class indices of the customers in the queue. The class of customer to serve next is a) the class of customer with the lowest class index greater than \( x'_0 \), if there is at least one such customer in the queue, or b) the class of customer with the lowest class index, if there are no customers in the queue with a class index greater than \( x'_0 \).

All of these factors are stated in equation form as

\[
w = \begin{cases} 
\min_i \left( i; x'_i > 0, i \geq x'_0 \right), x'_i > 0, \text{for some } i \geq x'_0, \\
\min_i \left( i; x'_i > 0, i \geq 1 \right), x'_i = 0, \text{for all } i \geq x'_0.
\end{cases}
\]
It should be noted that the above expression uniquely defines the class of customer to serve next for non-empty service completion epochs.

The values of \( w \) for the non-empty service completion epochs of the \((k = 2, N = 2)\) alternating priority model are given in Table 14. As noted in the preceding chapter the sequencing rule has no bearing on the class of customer to serve next when there is only one class of customer waiting to be served, as in the case for the first eight states listed in Table 14. The last two states, \( X' = (1,1,1) \) and \( X' = (2,1,1) \), are subject to the sequencing rule. For the state \( X' = (1,1,1) \), a class 1 customer is completing service, \( x'_0 = 1 \), and a class 1 customer is available, \( x'_1 = 1 \), so the class of customer to serve next is class 1, \( w = 1 \). For the state \( X' = (2,1,1) \), a class 2 customer is completing service, \( x'_0 = 2 \) and a class 2 customer is available, \( x'_2 = 1 \), so the class of customer to serve next is class 2, \( w = 2 \). Before continuing the development of the alternating priority model the essential features of the rotating priority model are presented. The procedures for determining the steady state probabilities and measures of system performance for both models are then stated.

**Rotating Priority Sequencing Rule**

The rotating priority sequencing rule represents an attempt to maximize the number of changes in the class of customer receiving service. This is accomplished by
Table 14. Class of Customer to Serve Next After a Non-empty Service Completion Epoch has Occurred. The values of $w$ for the $(k = 2, N = 2)$ Alternating Priority Model.
requiring the customer receiving service to yield control of the server to a class of customer that is different from his class. In operation this sequencing rule requires that the class of customer to be served next be selected according to a predetermined sequence. The predetermined sequence is denoted by the assignment of class indices in the same manner indicated for the alternating priority rule. Thus the assignment of class indices indicates that the desired order of service is one class 1 customer, one class 2 customer, ..., one class k customer, one class 1 customer, one class 2 customer, etc. It should be noted that the relative order of the index, not the absolute value of the index, is the relevant factor in the assignment of indices.

The variable w is defined for the rotating priority sequencing rule by an equation similar to Equation (189). The state of the system at a non-empty service completion epoch is $X' = (x_0', x_1', ..., x_k')$. The class of customer completing service is $x_0'$ and the rotating priority sequencing rule requires that a class $(x_0' + 1)$ customer be served next if there is at least one such customer in the queue, $x'(x_0' + 1) > 0$. If no class $(x_0' + 1)$ customers are available, then the class of customer to serve next is determined by examining the class indices of the customers in the queue. The class of customer to serve next is a) the class of customer with the lowest class index greater than $(x_0' + 1)$, if there is at least one such customer in the queue, or
b) the class of customer with the lowest class index, if there are no customers in the queue with a class index greater than \( x_0' + 1 \). All of these factors are stated in equation form as

\[
\begin{align*}
  w &= \begin{cases} 
  \min (i; x_i > 0, i > x_0'), & x_i > 0, \text{ for some } i > x_0', \\
  \min (i; x_i' > 0, i > 1), & x_i' = 0, \text{ for all } i > x_0' \end{cases} 
\end{align*}
\]

(190)

It should be noted that the above expression uniquely defines the class of customer to serve next for non-empty service completion epochs.

The values of \( w \) for the non-empty service completion epochs of the \((k = 2, N = 2)\) rotating priority model are given in Table 15. Examining the last two states listed in Table 15 reveals the effect of the sequencing rule. For the state \( X' = (1,1,1) \), a class 1 customer is receiving service, \( x_0' = 1 \), and a class 2 customer is available, \( x_2' = 1 \), so the class of customer to serve next is class 2, \( w = 2 \). For the state \( X' = (2,1,1) \), a class 2 customer is completing service, \( x_0' = 2 \), and a class 1 customer is available, \( x_1' = 1 \), so the class of customer to serve next is class 1, \( w = 1 \). The procedures needed to determine the steady state probabilities and the measures of system performance for the two previous models are now given.

Model Procedures and Measures of Performance

Both the alternating priority sequencing rule and the rotating priority sequencing rule determine the value of \( w \).
<table>
<thead>
<tr>
<th>State Description for Non-empty Service Completion Epoch</th>
<th>Class of Customer to serve next</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X' = (x'_0, x'_1, x'_2)$ where $x' &gt; 0$</td>
<td>$w$</td>
</tr>
<tr>
<td>$(1, 1, 0)$</td>
<td>1</td>
</tr>
<tr>
<td>$(2, 1, 0)$</td>
<td>1</td>
</tr>
<tr>
<td>$(1, 0, 1)$</td>
<td>2</td>
</tr>
<tr>
<td>$(2, 0, 1)$</td>
<td>2</td>
</tr>
<tr>
<td>$(1, 2, 0)$</td>
<td>1</td>
</tr>
<tr>
<td>$(2, 2, 0)$</td>
<td>1</td>
</tr>
<tr>
<td>$(1, 0, 2)$</td>
<td>2</td>
</tr>
<tr>
<td>$(2, 0, 2)$</td>
<td>2</td>
</tr>
<tr>
<td>$(1, 1, 1)$</td>
<td>2</td>
</tr>
<tr>
<td>$(2, 1, 1)$</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 15. Class of Customer to Serve Next After a Non-empty Service Completion Epoch has Occurred. The values of $w$ for the $(k = 2, N = 2)$ Rotating Priority Model.
with certainty for all non-empty service completion epochs. As a result, the equations developed in the preceding chapter are appropriate for determining the steady state probabilities and measures of system performance. This result is valid because the explicit values of the variable \( w \) were not required in the development of the equations. The equations needed to determine the quantities of interest are listed without comment. The values of \( w \) used in the equations are given by Equation (189) for the alternating priority model and by Equation (190) for the rotating priority model.

1. The imbedded Markov chain steady state probabilities, \( g(X) \), are determined by solving the set of simultaneous linear equations given by Equation (120), or, alternatively, using the matrix method due to Singer and given by Equation (12).

2. The semi-Markov process steady state probabilities, \( h(Y) \), are determined by Equation (134).

3. The random point in time steady state probabilities, \( p(Z) \), are determined by Equation (145).

4. The probability mass functions of the number of customers in the system and the number of customers in the queue are given by Equations (146) through (149).

5. The measures of system performance are given by Equations (150) through (157).
Example Three

The alternating priority model is illustrated by reconsidering the \((k = 2, N = 2)\) system described in Example Two. The data given for this system are treated using the alternating priority sequencing rule. As all of the equations have been demonstrated in Example Two, only the end results of the calculations are presented.

The IMC transition matrix, \(B\), is developed using the procedures of Table 6 and the values of \(w\) given by Equation (189). The \(B\) matrix is displayed in Figure 11. The effect of the sequencing rule is limited to the last two rows of the matrix, as these rows correspond to the only states where more than one class of customer is awaiting service. The non-zero entries of the \(B\) matrix are evaluated using the service time arrival probabilities given in Table 9. The IMC steady state probabilities, \(g(X)\), are determined by Equation (120) and are displayed in Table 16. The SMP steady state probabilities, \(h(Y)\), are determined by Equation (134) and are also displayed in Table 16.

The conversion matrix, \(B'\), used to transform the SMP steady state probabilities to the RPT steady state probabilities is developed by using Equations (139) and (140). The \(B'\) matrix is displayed in Figure 12. Comparing this conversion matrix with the IMC transition matrix displayed in Figure 11 verifies the previously stated result that the \(B'\) matrix can be readily derived from the \(B\) matrix. This is accomplished
Figure 11. IMC Transition Matrix for the (k=2, N=2) Alternating Priority Model.
Figure 12. Conversion Matrix for the \((k=2, N=2)\) Alternating Priority Model.
by deleting the first two rows of the B matrix and replacing
the non-zero entries of the B matrix (the service time
arrival probabilities \( A(a_1, a_2; s) \) or \( \overline{A}(a_1, a_2; s) \)) with the
equivalent random interruption time arrival probabilities,
\( A'(a_1, a_2; s) \) or \( \overline{A}'(a_1, a_2; s) \), respectively. The basis for
this conversion was justified in the preceding chapter. The
non-zero elements of the \( B' \) matrix are evaluated using the
random interruption time arrival probabilities given in
Table 9. The RPT steady state probabilities, \( p(Z) \), are
determined by utilizing this conversion matrix in the manner
indicated by Equation (145). The RPT steady state probabili­
ties are displayed in Table 16.

The probability mass functions of the number of custom­
ers in the queue and the number of customers in the system
are given in Table 17. The measures of system performance
are given in Table 18. The equations used to determine these
quantities were cited earlier.

Example Four

The rotating priority model is illustrated by treating
the data given for the \((k = 2, N = 2)\) system described in
Example Two. The results of the calculations are stated
without comment as the only difference between this model and
the alternating priority model of Example Three is that the
values of \( w \) are given by Equation (190). The IMC transition
matrix is displayed in Figure 13, and the conversion matrix
is displayed in Figure 14. The steady state probabilities
## Table 16. Steady State Probabilities for the \((k = 2, N = 2)\) Alternating Priority Model of Example Three.

<table>
<thead>
<tr>
<th>State Description</th>
<th>Imbedded Markov Chain (g(X); X = (x_0, x_1, x_2))</th>
<th>Semi-Markov Process (h(Y); Y = e_1, e_2, X)</th>
<th>Random Point in Time (p(Z); Z = 0, X)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(e_1)</td>
<td>--</td>
<td>0.0191</td>
<td>--</td>
</tr>
<tr>
<td>(e_2)</td>
<td>--</td>
<td>0.0383</td>
<td>--</td>
</tr>
<tr>
<td>0</td>
<td>--</td>
<td>--</td>
<td>0.0383</td>
</tr>
<tr>
<td>((x_0, x_1, x_2))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>((1, 0, 0))</td>
<td>0.0404</td>
<td>0.0172</td>
<td>0.0409</td>
</tr>
<tr>
<td>((2, 0, 0))</td>
<td>0.0497</td>
<td>0.0211</td>
<td>0.0529</td>
</tr>
<tr>
<td>((1, 1, 0))</td>
<td>0.0510</td>
<td>0.0433</td>
<td>0.0789</td>
</tr>
<tr>
<td>((2, 1, 0))</td>
<td>0.0408</td>
<td>0.0520</td>
<td>0.0813</td>
</tr>
<tr>
<td>((1, 0, 1))</td>
<td>0.0595</td>
<td>0.0253</td>
<td>0.0694</td>
</tr>
<tr>
<td>((2, 0, 1))</td>
<td>0.0697</td>
<td>0.0889</td>
<td>0.0631</td>
</tr>
<tr>
<td>((1, 1, 0))</td>
<td>0.0929</td>
<td>0.0789</td>
<td>0.0793</td>
</tr>
<tr>
<td>((2, 1, 0))</td>
<td>0.0956</td>
<td>0.1218</td>
<td>0.0925</td>
</tr>
<tr>
<td>((1, 0, 2))</td>
<td>0.0817</td>
<td>0.0347</td>
<td>0.0578</td>
</tr>
<tr>
<td>((2, 0, 2))</td>
<td>0.0742</td>
<td>0.0946</td>
<td>0.0580</td>
</tr>
<tr>
<td>((1, 1, 1))</td>
<td>0.1746</td>
<td>0.1483</td>
<td>0.1371</td>
</tr>
<tr>
<td>((2, 1, 1))</td>
<td>0.1699</td>
<td>0.2165</td>
<td>0.1505</td>
</tr>
</tbody>
</table>
## Table 17. Probability Mass Functions for the Number of Customers in the System and the Number of Customers in the Queue for the \((k = 2, N = 2)\) Alternating Priority Model of Example Three.

<table>
<thead>
<tr>
<th>j</th>
<th>Customers in System</th>
<th></th>
<th>Customers in Queue</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Class 1 (pr(n_1 = j))</td>
<td>Class 2 (pr(n_2 = j))</td>
<td>Overall (pr(n = j))</td>
<td>Class 1 (pr(nq_1 = j))</td>
<td>Class 2 (pr(nq_2 = j))</td>
<td>Overall (pr(n_q = j))</td>
</tr>
<tr>
<td>0</td>
<td>0.2123</td>
<td>0.2374</td>
<td>0.0383</td>
<td>0.3804</td>
<td>0.4641</td>
<td>0.1321</td>
</tr>
<tr>
<td>1</td>
<td>0.3999</td>
<td>0.4332</td>
<td>0.0938</td>
<td>0.4478</td>
<td>0.4201</td>
<td>0.2927</td>
</tr>
<tr>
<td>2</td>
<td>0.3085</td>
<td>0.2714</td>
<td>0.2927</td>
<td>0.1718</td>
<td>0.1158</td>
<td>0.5752</td>
</tr>
<tr>
<td>3</td>
<td>0.0793</td>
<td>0.0580</td>
<td>0.5752</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Class 1 Customers</td>
<td>Effective Mean Arrival Rate $r_{eff}$</td>
<td>Expected Number Customers in System $E(n_i)$</td>
<td>Expected Number Customers in Queue $E(n_{qi})$</td>
<td>Expected Waiting Time in System $E(w_i)(min)$</td>
<td>Expected Waiting Time in Queue $E(w_{qi})(min)$</td>
<td>Expected Service Time $E(t_i)(min)$</td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------------------------------</td>
<td>---------------------------------------------</td>
<td>---------------------------------------------</td>
<td>---------------------------------------------</td>
<td>---------------------------------------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>$i = 1$</td>
<td>0.2124</td>
<td>1.2548</td>
<td>0.7914</td>
<td>5.9077</td>
<td>3.7260</td>
<td>2.1817</td>
</tr>
<tr>
<td>Class 2 Customers</td>
<td>0.2124</td>
<td>1.1500</td>
<td>0.6517</td>
<td>5.4143</td>
<td>3.0683</td>
<td>2.3460</td>
</tr>
<tr>
<td>$i = 2$</td>
<td>0.4248</td>
<td>2.4048</td>
<td>1.4431</td>
<td>5.6610</td>
<td>3.3971</td>
<td>2.2639</td>
</tr>
<tr>
<td>Overall Customer Population</td>
<td>0.4248</td>
<td>2.4048</td>
<td>1.4431</td>
<td>5.6610</td>
<td>3.3971</td>
<td>2.2639</td>
</tr>
</tbody>
</table>

Table 18. Measures of System Performance for the $(k = 2, N = 2)$ Alternating Priority Model of Example Three.
To $(x_0^n, x_1^n, x_2^n)$

<table>
<thead>
<tr>
<th>From $(x_0^l, x_1^l, x_2^l)$</th>
<th>$(1,0,0)$</th>
<th>$(2,0,0)$</th>
<th>$(1,1,0)$</th>
<th>$(2,1,0)$</th>
<th>$(1,0,1)$</th>
<th>$(2,0,1)$</th>
<th>$(1,2,0)$</th>
<th>$(2,2,0)$</th>
<th>$(1,0,2)$</th>
<th>$(2,0,2)$</th>
<th>$(1,1,1)$</th>
<th>$(2,1,1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(1,0,0)$</td>
<td>$r_1 A(0,0,01)$</td>
<td>$r_2 A(0,0,02)$</td>
<td>$r_1 A(1,0,01)$</td>
<td>$r_2 A(1,0,02)$</td>
<td>$r_1 A(2,0,01)$</td>
<td>$r_2 A(2,0,02)$</td>
<td>$r_1 A(0,0,01)$</td>
<td>$r_2 A(0,0,02)$</td>
<td>$r_1 A(2,0,01)$</td>
<td>$r_2 A(2,0,02)$</td>
<td>$r_1 A(1,1,01)$</td>
<td>$r_2 A(1,1,02)$</td>
</tr>
<tr>
<td>$(2,0,0)$</td>
<td>$r_1 A(0,0,01)$</td>
<td>$r_2 A(0,0,02)$</td>
<td>$r_1 A(1,0,01)$</td>
<td>$r_2 A(1,0,02)$</td>
<td>$r_1 A(2,0,01)$</td>
<td>$r_2 A(2,0,02)$</td>
<td>$r_1 A(0,0,01)$</td>
<td>$r_2 A(0,0,02)$</td>
<td>$r_1 A(2,0,01)$</td>
<td>$r_2 A(2,0,02)$</td>
<td>$r_1 A(1,1,01)$</td>
<td>$r_2 A(1,1,02)$</td>
</tr>
<tr>
<td>$(1,1,0)$</td>
<td>$A(0,0,11)$</td>
<td>$A(0,0,12)$</td>
<td>$A(1,0,11)$</td>
<td>$A(1,0,12)$</td>
<td>$A(0,1,11)$</td>
<td>$A(0,1,12)$</td>
<td>$A(1,2,11)$</td>
<td>$A(1,2,12)$</td>
<td>$A(0,2,11)$</td>
<td>$A(0,2,12)$</td>
<td>$A(1,1,11)$</td>
<td>$A(1,1,12)$</td>
</tr>
<tr>
<td>$(2,1,0)$</td>
<td>$A(0,0,21)$</td>
<td>$A(0,0,22)$</td>
<td>$A(1,0,21)$</td>
<td>$A(1,0,22)$</td>
<td>$A(0,1,21)$</td>
<td>$A(0,1,22)$</td>
<td>$A(1,2,21)$</td>
<td>$A(1,2,22)$</td>
<td>$A(0,2,21)$</td>
<td>$A(0,2,22)$</td>
<td>$A(1,1,21)$</td>
<td>$A(1,1,22)$</td>
</tr>
<tr>
<td>$(1,2,0)$</td>
<td>$A(0,0,11)$</td>
<td>$A(0,0,12)$</td>
<td>$A(1,0,11)$</td>
<td>$A(1,0,12)$</td>
<td>$A(0,1,11)$</td>
<td>$A(0,1,12)$</td>
<td>$A(1,2,11)$</td>
<td>$A(1,2,12)$</td>
<td>$A(0,2,11)$</td>
<td>$A(0,2,12)$</td>
<td>$A(1,1,11)$</td>
<td>$A(1,1,12)$</td>
</tr>
<tr>
<td>$(2,2,0)$</td>
<td>$A(0,0,21)$</td>
<td>$A(0,0,22)$</td>
<td>$A(1,0,21)$</td>
<td>$A(1,0,22)$</td>
<td>$A(0,1,21)$</td>
<td>$A(0,1,22)$</td>
<td>$A(1,2,21)$</td>
<td>$A(1,2,22)$</td>
<td>$A(0,2,21)$</td>
<td>$A(0,2,22)$</td>
<td>$A(1,1,21)$</td>
<td>$A(1,1,22)$</td>
</tr>
<tr>
<td>$(1,0,2)$</td>
<td>$A(0,0,11)$</td>
<td>$A(0,0,12)$</td>
<td>$A(1,0,11)$</td>
<td>$A(1,0,12)$</td>
<td>$A(0,1,11)$</td>
<td>$A(0,1,12)$</td>
<td>$A(1,2,11)$</td>
<td>$A(1,2,12)$</td>
<td>$A(0,2,11)$</td>
<td>$A(0,2,12)$</td>
<td>$A(1,1,11)$</td>
<td>$A(1,1,12)$</td>
</tr>
<tr>
<td>$(2,0,2)$</td>
<td>$A(0,0,21)$</td>
<td>$A(0,0,22)$</td>
<td>$A(1,0,21)$</td>
<td>$A(1,0,22)$</td>
<td>$A(0,1,21)$</td>
<td>$A(0,1,22)$</td>
<td>$A(1,2,21)$</td>
<td>$A(1,2,22)$</td>
<td>$A(0,2,21)$</td>
<td>$A(0,2,22)$</td>
<td>$A(1,1,21)$</td>
<td>$A(1,1,22)$</td>
</tr>
<tr>
<td>$(1,1,1)$</td>
<td>$A(0,0,11)$</td>
<td>$A(0,0,12)$</td>
<td>$A(1,0,11)$</td>
<td>$A(1,0,12)$</td>
<td>$A(0,1,11)$</td>
<td>$A(0,1,12)$</td>
<td>$A(1,2,11)$</td>
<td>$A(1,2,12)$</td>
<td>$A(0,2,11)$</td>
<td>$A(0,2,12)$</td>
<td>$A(1,1,11)$</td>
<td>$A(1,1,12)$</td>
</tr>
<tr>
<td>$(2,1,1)$</td>
<td>$A(0,0,21)$</td>
<td>$A(0,0,22)$</td>
<td>$A(1,0,21)$</td>
<td>$A(1,0,22)$</td>
<td>$A(0,1,21)$</td>
<td>$A(0,1,22)$</td>
<td>$A(1,2,21)$</td>
<td>$A(1,2,22)$</td>
<td>$A(0,2,21)$</td>
<td>$A(0,2,22)$</td>
<td>$A(1,1,21)$</td>
<td>$A(1,1,22)$</td>
</tr>
</tbody>
</table>

Figure 13. IMC Transition Matrix for the $(k=2, N=2)$ Rotating Priority Model.
To Z
\( \{ x_0, x_1, x_2 \} \)

From Y
\( \{ y_0, y_1, y_2 \} \)

<table>
<thead>
<tr>
<th></th>
<th>(1, 0, 0)</th>
<th>(2, 0, 0)</th>
<th>(1, 1, 0)</th>
<th>(2, 1, 0)</th>
<th>(1, 0, 1)</th>
<th>(2, 0, 1)</th>
<th>(1, 2, 0)</th>
<th>(2, 2, 0)</th>
<th>(1, 0, 2)</th>
<th>(2, 0, 2)</th>
<th>(1, 1, 1)</th>
<th>(2, 1, 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1, 0)</td>
<td>( A'(0,0;11) )</td>
<td>0</td>
<td>( A'(1,0;11) )</td>
<td>0</td>
<td>( A'(0,1;11) )</td>
<td>0</td>
<td>( \tilde{A}(2,0;11) )</td>
<td>0</td>
<td>( \tilde{A}(0,2;11) )</td>
<td>0</td>
<td>( \tilde{A}(1,1;11) )</td>
<td>0</td>
</tr>
<tr>
<td>(2, 1, 0)</td>
<td>( A'(0,0;21) )</td>
<td>0</td>
<td>( A'(1,0;21) )</td>
<td>0</td>
<td>( A'(0,1;21) )</td>
<td>0</td>
<td>( \tilde{A}(2,0;21) )</td>
<td>0</td>
<td>( \tilde{A}(0,2;21) )</td>
<td>0</td>
<td>( \tilde{A}(1,1;21) )</td>
<td>0</td>
</tr>
<tr>
<td>(1, 0, 1)</td>
<td>0</td>
<td>( A'(0,0;12) )</td>
<td>0</td>
<td>( A'(1,0;12) )</td>
<td>0</td>
<td>( A'(0,1;12) )</td>
<td>0</td>
<td>( \tilde{A}(2,0;12) )</td>
<td>0</td>
<td>( \tilde{A}(0,2;12) )</td>
<td>0</td>
<td>( \tilde{A}(1,1;12) )</td>
</tr>
<tr>
<td>(2, 0, 1)</td>
<td>0</td>
<td>( A'(0,0;22) )</td>
<td>0</td>
<td>( A'(1,0;22) )</td>
<td>0</td>
<td>( A'(0,1;22) )</td>
<td>0</td>
<td>( \tilde{A}(2,0;22) )</td>
<td>0</td>
<td>( \tilde{A}(0,2;22) )</td>
<td>0</td>
<td>( \tilde{A}(1,1;22) )</td>
</tr>
<tr>
<td>(1, 2, 0)</td>
<td>0</td>
<td>0</td>
<td>( A'(0,0;11) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \tilde{A}(1,0;11) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \tilde{A}(0,1;11) )</td>
<td>0</td>
</tr>
<tr>
<td>(2, 2, 0)</td>
<td>0</td>
<td>0</td>
<td>( A'(0,0;21) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \tilde{A}(1,0;21) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \tilde{A}(0,1;21) )</td>
<td>0</td>
</tr>
<tr>
<td>(1, 0, 2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( A'(0,0;12) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \tilde{A}(0,1;12) )</td>
<td>0</td>
<td>( \tilde{A}(1,0;12) )</td>
</tr>
<tr>
<td>(2, 0, 2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( A'(0,0;22) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \tilde{A}(0,1;22) )</td>
<td>0</td>
<td>( \tilde{A}(1,0;22) )</td>
<td></td>
</tr>
<tr>
<td>(1, 1, 1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( A'(0,0;12) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \tilde{A}(1,0;12) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \tilde{A}(0,1;12) )</td>
</tr>
<tr>
<td>(2, 1, 1)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( A'(0,0;21) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \tilde{A}(1,0;21) )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \tilde{A}(0,1;21) )</td>
</tr>
</tbody>
</table>

Figure 14. Conversion Matrix for the (k=2, N=2) Rotating Priority Model.
are given in Table 19, the probability mass functions are given in Table 20, and the measures of system performance are given in Table 21.

First Come-First Served Sequencing Rule

The first come-first served sequencing rule represents an attempt to provide service on an equitable basis according to the order of arrival. The class of customer that is served next is determined solely on the basis of the order of entry into the system; the assignment of class indices in no way affects the order of service. It should be noted that the class of customer to serve next is determined with certainty if the time of arrival of each customer in the queue is known, but this information is not provided by the state descriptions of the models developed in this study. As a result, the class of customer to serve next can only be determined by indirect means. The indirect method indicates the possible classes of customers to be served next with the associated probability that a particular class of customer will be selected. In terms of the model notation, the variable $w$ is determined probabilistically. The equations used to define $w$ are motivated by considering a small example problem.

Suppose that a non-empty service completion epoch has just occurred in a three customer class system and the state of the system is given by the vector $X' = (x'_0 = 2, x'_1 = 3, x'_2 = 0, x'_3 = 1)$. The total number of customers in the queue is $\bar{x}' = 4$. Clearly the class of customer to serve next is
<table>
<thead>
<tr>
<th>State Description</th>
<th>Imbedded Markov Chain ( g(X); X = (x_0, x_1, x_2) )</th>
<th>Semi-Markov Process ( h(Y); Y = e_1, e_2, X )</th>
<th>Random Point in Time ( p(Z); Z = 0, X )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( e_1 )</td>
<td>(--)</td>
<td>0.0202</td>
<td>(--)</td>
</tr>
<tr>
<td>( e_2 )</td>
<td>(--)</td>
<td>0.0403</td>
<td>(--)</td>
</tr>
<tr>
<td>( 0 )</td>
<td>(--)</td>
<td>(--)</td>
<td>0.0404</td>
</tr>
<tr>
<td>( (x_0, x_1, x_2) )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( (1, 0, 0) )</td>
<td>0.0487</td>
<td>0.0222</td>
<td>0.0707</td>
</tr>
<tr>
<td>( (2, 0, 0) )</td>
<td>0.0399</td>
<td>0.0182</td>
<td>0.0526</td>
</tr>
<tr>
<td>( (1, 1, 0) )</td>
<td>0.0528</td>
<td>0.0481</td>
<td>0.0820</td>
</tr>
<tr>
<td>( (2, 1, 0) )</td>
<td>0.1069</td>
<td>0.1460</td>
<td>0.0593</td>
</tr>
<tr>
<td>( (1, 0, 1) )</td>
<td>0.0398</td>
<td>0.0181</td>
<td>0.0814</td>
</tr>
<tr>
<td>( (2, 0, 1) )</td>
<td>0.0714</td>
<td>0.0975</td>
<td>0.0690</td>
</tr>
<tr>
<td>( (1, 2, 0) )</td>
<td>0.0900</td>
<td>0.0820</td>
<td>0.0804</td>
</tr>
<tr>
<td>( (2, 2, 0) )</td>
<td>0.0651</td>
<td>0.0889</td>
<td>0.0370</td>
</tr>
<tr>
<td>( (1, 0, 2) )</td>
<td>0.0894</td>
<td>0.0407</td>
<td>0.0914</td>
</tr>
<tr>
<td>( (2, 0, 2) )</td>
<td>0.0758</td>
<td>0.1036</td>
<td>0.0635</td>
</tr>
<tr>
<td>( (1, 1, 1) )</td>
<td>0.1793</td>
<td>0.0817</td>
<td>0.1718</td>
</tr>
<tr>
<td>( (2, 1, 1) )</td>
<td>0.1409</td>
<td>0.1925</td>
<td>0.1005</td>
</tr>
</tbody>
</table>

Table 19. Steady State Probabilities for the \((k = 2, N = 2)\) Rotating Priority Model of Example Four.
| j | Customers in System | | Customers in Queue | | | | | |
|---|-----------------|---|-----------------|---|---|---|---|
|   | Class 1 \( pr(n_1 = j) \) | Class 2 \( pr(n_2 = j) \) | Overall \( pr(n = j) \) | Class 1 \( pr(nq_1 = j) \) | Class 2 \( pr(nq_2 = j) \) | Overall \( pr(nq = j) \) |
| 0 | 0.2255 | 0.2735 | 0.0404 | 0.4690 | 0.4224 | 0.1637 |
| 1 | 0.4033 | 0.4021 | 0.1233 | 0.4136 | 0.4227 | 0.2917 |
| 2 | 0.2908 | 0.2609 | 0.2917 | 0.1174 | 0.1549 | 0.5446 |
| 3 | 0.0804 | 0.0635 | 0.5446 | ---- | ---- | ---- |

Table 20. Probability Mass Functions for the Number of Customers in the System and the Number of Customers in the Queue for the \((k = 2, N = 2)\) Rotating Priority Model of Example Four.
<table>
<thead>
<tr>
<th></th>
<th>Effective Mean Arrival Rate $r_{effi}$</th>
<th>Expected Number Customers in System $E(n_i)$</th>
<th>Expected Number Customers in Queue $E(n_{qi})$</th>
<th>Expected Waiting Time in System $E(w_i)(min)$</th>
<th>Expected Waiting Time in Queue $E(w_{qi})(min)$</th>
<th>Expected Service Time $E(t_i)(min)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1 Customers $i = 1$</td>
<td>0.2277</td>
<td>1.2261</td>
<td>0.6484</td>
<td>5.3847</td>
<td>2.8476</td>
<td>2.5371</td>
</tr>
<tr>
<td>Class 2 Customers $i = 2$</td>
<td>0.2277</td>
<td>1.1144</td>
<td>0.7325</td>
<td>4.8942</td>
<td>3.2170</td>
<td>1.6772</td>
</tr>
<tr>
<td>Overall Customer Population</td>
<td>0.4554</td>
<td>2.3405</td>
<td>1.3809</td>
<td>5.1394</td>
<td>3.0323</td>
<td>2.1071</td>
</tr>
</tbody>
</table>

Table 21. Measures of System Performance for the $(k = 2, N = 2)$ Rotating Priority Model of Example Four.
determined by indicating which of the four customers waiting in the queue arrived immediately after the customer that has just completed service. Because of the random nature of the arrival process, any of the four customers waiting in the queue is equally likely to have arrived first. As three of the customers are of class 1, \( x'_1 = 3 \), the probability is \( \left( x'_1 / x' \right) = \left( 3 / 4 \right) \), that a class 1 customer is served next. As none of the customers are of class 2, \( x'_2 = 0 \), the probability is \( \left( x'_2 / x' \right) = \left( 0 / 4 \right) = 0 \), that a class 2 customer is served next. Finally, as one of the customers waiting for service is of class 3, \( x'_3 = 1 \), the probability is \( \left( x'_3 / x' \right) = \left( 1 / 4 \right) \), that a class 3 customer is served next. The generalization of this logic for the \( (k, N) \) model is straightforward. The class of customer to serve next, given that a non-empty service completion epoch has occurred with state description \( X' = (x'_0, x'_1, ..., x'_k) \), is

\[
w = i, \quad \text{with probability} \quad \left( x'_i / x' \right), \quad i = 1, 2, ..., k (191)
\]

Note that the above equation assigns a zero probability to the event that the class of customer to serve next is class \( i \) if no class \( i \) customers are waiting in the queue. The values of \( w \) for the \( (k = 2, N = 2) \) first come-first served sequencing rule are given in Table 22. An examination of the last two states listed in Table 15, \( X' = (1,1,1) \) and \( X' = (2,1,1) \), reveals the effect of the sequencing rule.

Model Modifications

As the variable \( w \) is determined probabilistically for
### Table 22. Class of Customer to Serve Next After a Non-empty Service Completion Epoch has Occurred.

The values of w for the \((k = 2, N = 2)\) First come-First served Model.

<table>
<thead>
<tr>
<th>State Description for Non-empty Service Completion Epoch</th>
<th>Class of Customer to serve next</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X' = (x'_0, x'_1, x'_2)) where (x' &gt; 0)</td>
<td>(w)</td>
</tr>
<tr>
<td>(1, 1, 0)</td>
<td>1</td>
</tr>
<tr>
<td>(2, 1, 0)</td>
<td>1</td>
</tr>
<tr>
<td>(1, 0, 1)</td>
<td>2</td>
</tr>
<tr>
<td>(2, 0, 1)</td>
<td>2</td>
</tr>
<tr>
<td>(1, 2, 0)</td>
<td>1</td>
</tr>
<tr>
<td>(2, 2, 0)</td>
<td>1</td>
</tr>
<tr>
<td>(1, 0, 2)</td>
<td>2</td>
</tr>
<tr>
<td>(2, 0, 2)</td>
<td>2</td>
</tr>
<tr>
<td>(1, 1, 1)</td>
<td>1 with probability 1/2</td>
</tr>
<tr>
<td>(2, 1, 1)</td>
<td>2 with probability 1/2</td>
</tr>
<tr>
<td></td>
<td>1 with probability 1/2</td>
</tr>
<tr>
<td></td>
<td>2 with probability 1/2</td>
</tr>
</tbody>
</table>

\(T_1\) with probability 1/2

\(T_2\) with probability 1/2
non-empty service completion epochs, several modifications must be made to the equations developed in the preceding chapter in order to calculate the steady state probabilities of interest. These modifications are minor, and the basic structure of the model remains intact. The three changes are [1] redefinition of the IMG transition probabilities for Case A) and Case B) transitions as given in Table 6; [2] substitution of a redefined mean time until the next epoch given that a non-empty service completion epoch has just occurred, \((v_Y, Y = X' \text{ where } X'>0)\) for use in the equations that determine the SMP steady state probabilities; and [3] redefinition of the conversion probabilities for Case A') and Case B') as given by Equation (139).

**Modification [1]**

The Case A) and Case B) transitions given in Table 6 correspond to non-empty service completion epochs. As the class of customer to be served next is determined probabilistically for this situation, the transition probabilities must be modified. The Case A) transition probability given in Table 6 is \(A([x''_i - x'_i + d_{iw}]; x'_w)\). This is the probability that the set of arrivals \([x''_i - x'_i + d_{iw}]\) enters the system during the service time of a class \(w\) customer preceded by a class \(x'_0\) customer. For the first come-first served sequencing rule, \(w\) is defined by Equation (191) to be class \(i\) with probability \((x'_i/x')\). Therefore the joint probability that a class \(w\) customer is selected to receive service and the set
of arrivals \([x''_i - x'_i + d_{iw}]\) enters the system during the service time of the class \(w\) customer preceded by a class \(x'_0\) customer is \((x''_w / x') A([x''_i - x'_i + d_{iw}], x'_0w)\). Stating this result with all of the conditions associated with the Case A) transition yields

\[
\text{pr}(X''|X') = \frac{x'_w}{x'} \bar{A}([x''_i - x'_i + d_{iw}], x'_0w), \quad \bar{x'} > 0, \quad \bar{x''} < N, \quad x''_0 = w, \quad \{x''_i - x'_i + d_{iw}\} \geq 0.
\]

Using the same logic the Case B) transition is

\[
\text{pr}(X''|X') = \frac{x'_w}{x'} \bar{A}([x''_i - x'_i + d_{iw}], x'_0w), \quad \bar{x'} > 0, \quad \bar{x''} = N, \quad x''_0 = w, \quad \{x''_i - x'_i + d_{iw}\} \geq 0.
\]

The value of \(w\) is defined by Equation (191). The case C), Case D), and Case E) transitions given by Table 6 are unchanged. Using this modified set of equations the IMC transition matrix for the first come-first served model are determined.

**Modification [2]**

The unconditional mean time until the next epoch occurs given that a non-empty service completion epoch has just occurred, \(v_Y, Y = X'\) where \(\bar{x'} > 0\), is redefined to account for the probabilistically determined value of \(w\). If the
customer to serve next is known to be a class i customer, then the mean time until the next epoch is \( m_{x_i} \). This is the mean time to serve a class i customer preceded by a class \( x_0 \) customer. For the first come-first served sequencing rule the probability that a class i customer is served next is \( \frac{x_i}{\overline{x}_i} \). Therefore the unconditional mean time between epochs is determined by weighting the mean time to serve a class i customer by the probability that a class i customer is served and summing over all values of i. Using the notation of the preceding chapter the unconditional mean time is

\[
v_Y = \sum_{i=1}^{k} m_{x_i} \frac{x_i}{\overline{x}_i}, \quad Y = \bar{X}, \text{ where } \overline{x} > 0. \tag{194}
\]

This modified expression is an element of the equation used to define the SMP steady state probabilities, \( h(Y) \), associated with non-empty service completion epochs. These probabilities are given by the last expression of Equation (134) for the previous models. For the first come-first served model the SMP steady state probabilities associated with the non-empty service completion epochs are

\[
h(Y) = g(Y)\frac{v_Y}{d}, \quad Y = \bar{X}, \text{ where } \overline{x} > 0, \tag{195}
\]

\[
g(\bar{x}')\left( \sum_{i=1}^{k} m_{x_i} \frac{x_i}{\overline{x}_i} \right)/d, \quad Y = \bar{X}', \text{ where } x' > 0,
\]

where \( d \) is a normalizing constant. The complete set of SMP steady state probabilities associated with state \( Y \) are stated in equation form for the first come-first served model. The prime notation is deleted in order to conform
with the notation of Equation (134). The probabilities are

\[
\begin{align*}
\text{h}(Y) = \begin{cases} 
\frac{(\Sigma_{X} g(X))r_{i}m_{0i}/(rd),}{s.t. X=0} & Y = e_{i}, \ i = 1,2,\ldots,k, \\
\frac{g(X)/(rd),}{Y = X, \text{ where } X = 0,} & Y = X, \text{ where } X > 0, \\
\frac{k}{g(X)} \left( \Sigma_{i=1}^{k} m_{x0i}(x_{i}/X) \right) / d, & Y = X, \text{ where } X = 0,
\end{cases}
\end{align*}
\]

where

\[
d = \left( \Sigma_{i=1}^{k} \left( r_{i}m_{0i}/r \right) + 1/r \right) \Sigma_{X} g(X) \quad \text{s.t. } X = 0
\]

\[
+ \Sigma_{X}^{k} g(X) \left( \Sigma_{i=1}^{k} m_{x0i}(x_{i}/X) \right). \\
\text{s.t. } X > 0
\]

(196)

Modification [3]

The Case A') and Case B') conversion probabilities given by Equation (139) are used to determine the B' matrix that is, in turn, utilized in the transformation of the SMP steady state probabilities to RPT steady state probabilities. These probabilities are redefined to reflect the fact that the first come-first served rule determines the class of customer to serve next probabilistically. The Case A') and Case B') conversion probabilities are modified by using the same logic used to modify the Case A) and Case B) IMC transition probabilities. The Case A') conversion probability given by Equation (139) is A'(\{x_{i} - x_{i}^{\uparrow} + d_{iw}\}; x_{0}^{\uparrow w}). This is the probability that the set of arrivals \{x_{i} - x_{i}^{\uparrow} + d_{iw}\} enters the system during the random interruption portion of the
service time of a class \( w \) customer preceded by a class \( x \) customer. For the first come-first served sequencing rule, the probability that a class \( w \) customer is served is \( \left( \frac{x^w}{\bar{x}'} \right)^w, w = 1, 2, \ldots, k \). Therefore the joint probability of interest is \( \left( \frac{x^w}{\bar{x}'} \right)^w \mathcal{A}'(\{ x_i - x_i' + d_{iw} \}; x_0^w) \). The case \( B' \) conversion probability for the first come-first served sequencing rule is found by using the same logic to be \( \left( \frac{x^w}{\bar{x}'} \right)^w \mathcal{A}'(\{ x_i - x_i' + d_{iw} \}; x_0^w) \). Stating these results in equation form, the conversion probabilities are

\[
\text{Case A'}
\]

\[
(x^w/\bar{x}')^w \mathcal{A}'(\{ x_i - x_i' + d_{iw} \}; x_0^w), \quad \bar{x} > 0,
\]

\[
\text{Case B'}
\]

\[
(x^w/\bar{x}')^w \mathcal{A}'(\{ x_i - x_i' + d_{iw} \}; x_0^w), \quad \bar{x} = N,
\]

\[
\text{Case E'}
\]

\[
0 \quad \text{otherwise.}
\]

A comparison of the above equations with Equations (192) and (193) indicates that the relationship between the IMC transition matrix, \( B \), and the conversion matrix, \( B' \), discussed in preceding sections is maintained for this model.
Model Procedures and Measures of Performance

The equations needed to determine the steady state probabilities and measures of system performance for the first come-first served model are given below. The values of \( w \) are given by Equation (191).

1. The imbedded Markov chain transition matrix, \( B \), is determined by using the procedures of Table 6 with Case A) replaced by Equation (192) and Case B) replaced by Equation (193).

2. The imbedded Markov chain steady state probabilities, \( g(X) \), are determined by either Equation (120) or Equation (12).

3. The semi-Markov process steady state probabilities, \( h(Y) \), are determined by the modified procedures given by Equation (196).

4. The conversion matrix, \( B' \), is determined by the modified procedures given by Equation (197).

5. The random point in time steady state probabilities, \( p(Z) \), are determined by using Equation (145) with the modified conversion matrix.

6. The probability mass functions of the number of customers in the system and the number of customers in the queue are given by Equations (146) through (149).

7. The measures of systems performance are given by Equations (150) through (157).
Example Five

The first come-first served model is illustrated by treating the data given for the \((k = 2, N = 2)\) system described in Example Two. The steady state probabilities and measures of system performance are determined by the equations cited in the previous section. The IMC transition matrix is displayed in Figure 15, and the conversion matrix is displayed in Figure 16. The steady state probabilities are given in Table 23, the probability mass functions are given in Table 24, and the measures of system performance are given in Table 25.

Summary

It has been shown that the multiple customer class model developed in the preceding chapter for the non-preemptive priority sequencing rule can readily modified to treat additional sequencing rules. If the sequencing rule uniquely determines the class of customer to serve next, given that a non-empty service completion epoch has just occurred, only the variable \(w\) must be redefined. The alternating priority and rotating priority sequencing rules are of this type. If the sequencing rule determines the class of customer to serve next probabilistically, given that a non-empty service completion epoch has just occurred, three equations must be modified in addition to redefining the variable \(w\). The first come-first served sequencing rule is an example of this type. It is hoped that the previous extended models have provided
Figure 15. IMC Transition Matrix for the (k=2, N=2) First Come-First Served Model.
Figure 16. Conversion Matrix for the (k=2, N=2) First Come-First Served Model.
<table>
<thead>
<tr>
<th>State Description</th>
<th>Imbedded Markov Chain $g(X); X = (x_0, x_1, x_2)$</th>
<th>Semi-Markov Process $h(Y); Y = e_1, e_2, X$</th>
<th>Random Point in Time $p(Z); Z = 0, X$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$</td>
<td>--</td>
<td>0.0198</td>
<td>--</td>
</tr>
<tr>
<td>$e_2$</td>
<td>--</td>
<td>0.0395</td>
<td>--</td>
</tr>
<tr>
<td>$0$</td>
<td>--</td>
<td>--</td>
<td>0.0396</td>
</tr>
<tr>
<td>$(x_0, x_1, x_2)$</td>
<td>0.0442</td>
<td>0.0195</td>
<td>0.0545</td>
</tr>
<tr>
<td>$(1, 0, 0)$</td>
<td>0.0456</td>
<td>0.0201</td>
<td>0.0540</td>
</tr>
<tr>
<td>$(2, 0, 0)$</td>
<td>0.0506</td>
<td>0.0445</td>
<td>0.0781</td>
</tr>
<tr>
<td>$(1, 1, 0)$</td>
<td>0.0727</td>
<td>0.0960</td>
<td>0.0778</td>
</tr>
<tr>
<td>$(2, 1, 0)$</td>
<td>0.0506</td>
<td>0.0223</td>
<td>0.0708</td>
</tr>
<tr>
<td>$(1, 0, 1)$</td>
<td>0.0727</td>
<td>0.0960</td>
<td>0.0680</td>
</tr>
<tr>
<td>$(2, 0, 1)$</td>
<td>0.0887</td>
<td>0.0781</td>
<td>0.0774</td>
</tr>
<tr>
<td>$(1, 2, 0)$</td>
<td>0.0772</td>
<td>0.1020</td>
<td>0.0674</td>
</tr>
<tr>
<td>$(2, 2, 0)$</td>
<td>0.0887</td>
<td>0.0390</td>
<td>0.0713</td>
</tr>
<tr>
<td>$(1, 0, 2)$</td>
<td>0.0772</td>
<td>0.1020</td>
<td>0.0625</td>
</tr>
<tr>
<td>$(2, 0, 2)$</td>
<td>0.1773</td>
<td>0.1171</td>
<td>0.1488</td>
</tr>
<tr>
<td>$(2, 1, 1)$</td>
<td>0.1545</td>
<td>0.2041</td>
<td>0.1298</td>
</tr>
</tbody>
</table>

Table 23. Steady State Probabilities for the $(k = 2, N = 2)$ First Come-First Served Model of Example Five.
Table 24. Probability Mass Functions for the Number of Customers in the System and the Number of Customers in the Queue for the \((k = 2, N = 2)\) First Come-First Served Model of Example Five.
<table>
<thead>
<tr>
<th></th>
<th>Effective Mean Arrival Rate $r_{effi}$</th>
<th>Expected Number Customers in System $E(n_i)$</th>
<th>Expected Number Customers in Queue $E(n_{qi})$</th>
<th>Expected Waiting Time in System $E(w_i)(\text{min})$</th>
<th>Expected Waiting Time in Queue $E(w_{qi})(\text{min})$</th>
<th>Expected Service Time $E(t_i)(\text{min})$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Class 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$i = 1$</td>
<td>0.2214</td>
<td>1.2250</td>
<td>0.7241</td>
<td>5.5330</td>
<td>3.2706</td>
<td>2.2626</td>
</tr>
<tr>
<td><strong>Class 2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$i = 2$</td>
<td>0.2214</td>
<td>1.1445</td>
<td>0.6850</td>
<td>5.1694</td>
<td>3.0939</td>
<td>2.0755</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>0.4428</td>
<td>2.3695</td>
<td>1.4091</td>
<td>5.3512</td>
<td>3.1822</td>
<td>2.1690</td>
</tr>
</tbody>
</table>

Table 25. Measures of System Performance for the $(k = 2, N = 2)$ First Come-First Served Model of Example Five.
sufficient insight into the methodology so that models for other sequencing rules can be readily developed. Several sequencing rules of interest are the shortest mean processing time rule and the longest mean processing time rule. These sequencing rules provide service on the basis of the expected duration of the customer's service time. Without interactive service times these rules reduce to the non-preemptive priority sequencing rule of the preceding chapter.

The numerical examples for the three models developed in this chapter and the numerical example of the preceding chapter provide an interesting result. Comparing the measures of performance for the overall system indicates that the rotating priority sequencing rule provides the most desirable levels of performance. In other words, the rotating priority rule provides the largest effective mean arrival rate, the smallest expected number of customers in the system, the smallest expected number of customers in the queue, the smallest expected waiting time in the system, the smallest expected waiting time in the queue, and the smallest expected service time. This result implies that for the system described in Example Two, the rotating priority sequencing rule processes more customers per unit time with less average congestion and waiting time than any other sequencing rule considered. This is a surprising result as one would expect that there would be some tradeoffs involving waiting time and effective arrival rate. The nature of this finding
indicates that computational experience with the models could be very rewarding if similar results are obtained for other specific systems.
CHAPTER VI
AIR TRAFFIC CONTROL MODEL APPLICATIONS

Introduction

The models developed in this study were originally motivated by a desire to investigate the sequencing of landing and takeoff operations at a single runway. The purpose of this chapter is to indicate how the models of Chapter IV and Chapter V might be used for such an investigation. The following order of discussion is adopted: first, the air traffic control system under consideration is described, the boundaries of the system are defined, and operating assumptions are stated; second, related studies are reviewed; and third, specific uses of the models contained in this study are discussed.

System Description

The system under consideration is portrayed by the sketch shown in Figure 17. It is seen that a landing aircraft enters the system several miles outside the outer marker and is placed in a holding pattern to await landing clearance. In some air terminal systems which experience moderate traffic loads, the function of the holding pattern is filled by issuing a series of radar vectors corresponding
Figure 17. Air Terminal System.
to a variable flight path in the terminal area. If the system is idle the entering aircraft immediately receives landing clearance and is not involved with the holding pattern or radar vector procedures. When landing clearance has been given the aircraft departs the holding pattern and flies to the final approach path by a fixed route defined by electronic landing aids or by radar vectors. The final approach path is a straight-line route defined both horizontally and vertically by electronic landing aids. The final approach path is referred to as the common path as all landing aircraft are required to follow this same route. The landing aircraft proceeds down the common path to the runway where the aircraft touches down and rapidly decelerates. When the velocity of the landing aircraft has sufficiently decreased the aircraft departs the active runway by the nearest exist taxiway. For modeling purposes a landing aircraft is assumed to enter the system at the holding pattern and to leave the system at the exit taxiway. Assumptions regarding the behavior of landing aircraft are subsequently stated.

Reference to Figure 17 indicates that a takeoff aircraft enters the system at the entry taxiway to the active runway. A waiting line forms at this point as aircraft await takeoff clearance. When takeoff clearance is given the aircraft enters the active runway and begins to accelerate. When the velocity of the takeoff aircraft has sufficiently increased
the aircraft becomes airborne and departs the active runway. For modeling purposes a takeoff aircraft is assumed to enter the system at the entry taxiway and to leave the system at the point the aircraft becomes airborne. A number of assumptions are made about the behavior of landing and takeoff aircraft while in the system. The assumptions are stated explicitly.

1. **Poisson Arrivals**

   Both landing and takeoff aircraft enter the system at random points in time with known mean rates. This Poisson arrival assumption is fairly standard in air traffic control studies as evidenced by the models developed by Odoni (1970) and Pestalozzi (1964). The reasoning for this assumption seems to be well-founded. Although aircraft are scheduled to takeoff and land at certain times throughout the day, random fluctuations due to weather, human error, and other sources tend to distort the orderliness of the arrival processes. Arguments of this sort are advanced and substantiating data presented in a report by Airborne Instruments Laboratory (1960).

2. **Interactive Service Times**

   The time required to conduct a landing or takeoff operation is a random variable following a general distribution which depends on a) the type of operation being conducted and b) the type of operation previously conducted. This inter-
active service time assumption is based on the many factors that influence the time required to perform an operation. The time required to perform a landing or takeoff is certainly more than just the time the aircraft is on the active runway. Under ideal saturated conditions an aircraft (landing or takeoff) is always in physical contact with the active runway. In this situation the service time for a landing or takeoff operation is simply the time the landing or takeoff aircraft is in physical contact with the runway. This time is referred to as runway occupancy time. However, this ideal service time is rarely obtainable due to a number of factors.

One factor affecting service times is wake turbulence. A trail of very turbulent air is left behind an aircraft during the time the lifting surfaces are effective. The turbulence of this air is most severe for large aircraft with efficient lifting surfaces. Thus before a landing aircraft can follow a preceding landing aircraft down the common path, or an aircraft can follow a preceding aircraft on the active runway, a brief delay is often incurred in order to allow the wake turbulence of the preceding aircraft to dissipate. For example, if a light plane is given takeoff clearance immediately after a large jet the small plane must wait for some time after the large jet becomes airborne to start takeoff roll. In other words, a period of time due to wake turbu-
ence is often incurred in addition to runway occupancy time when two successive operations are performed.

Another factor affecting the time to conduct landing and takeoff operations is a random error due to such variable elements as the air traffic controller, the pilot, and air and ground based equipment. It is reasonable to assume that these errors are also interactive due to the fact that the thought processes utilized by air traffic controllers and pilots depend on the sequence of operations. For example, a different set of procedures and thought processes governs a landing operation that follows another landing than governs a landing operation that follows a takeoff.

The net effect of these factors on the service time for a landing and takeoff operation can be stated in concise form by defining several quantities. The service time is defined in terms of the operation being performed and the preceding operation. The service time is the time the runway is committed to an operation and includes the runway occupancy time as well as the wake turbulence time, minimum longitudinal separation time, and random error time. It should be noted that wake turbulence time and minimum longitudinal separation time are defined as times in addition to the runway occupancy time. The notation is

\[ l = \text{landing operation}, \]
\[ t = \text{takeoff operation}, \]
\[ 0 = \text{no operation} \]
Runway occupancy times = \( \text{occ}(1), \text{occ}(t) \),
Wake turbulence times = \( \text{wake}(t, t), \text{wake}(1, 1), \text{wake}(1, t), \text{wake}(t, 1) \),
Minimum longitudinal separation time = \( \text{sep}(1, 1) \),
and
Random error times = \( e(0, 1), e(t, 1), e(1, 1), e(0, t), e(t, t), e(1, t) \).

In the above factors the first element in parentheses indicates the type of preceding operation and the second element in parentheses indicates the type of operation being conducted. The service time relationships are given in Table 26.

3. Limited Waiting Space

The air terminal system portrayed in Figure 17 can accommodate, at most, a fixed number of aircraft. This assumption is based on the physical limitation of the terminal ground facilities and the danger associated with a large number of aircraft flying in the relatively small terminal area airspace. This limitation has long been recognized by Federal Aviation Administration authorities and treated by Advanced Flow Control Procedures. Briefly stated, the procedures limit the flow of air traffic into a terminal area to a rate that can be served without incurring excessive delays in holding patterns. This effect is achieved by holding inbound air traffic on the ground at another terminal. The effect of limited waiting space on takeoff aircraft is to deny clearance for aircraft to taxi to the active runway when
<table>
<thead>
<tr>
<th>Preceding Operation</th>
<th>Operation</th>
<th>Service Time for Operation</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>landing</td>
<td>(\text{occ}(1) + e(0,1))</td>
</tr>
<tr>
<td>takeoff</td>
<td>landing</td>
<td>(\text{occ}(1) + \text{wake}(t,1) + e(t,1))</td>
</tr>
<tr>
<td>landing</td>
<td>landing</td>
<td>(\text{occ}(1) + \max[\text{wake}(1,1) + \text{sep}(1,1)] + e(1,1))</td>
</tr>
<tr>
<td>none</td>
<td>takeoff</td>
<td>(\text{occ}(t) + e(0,t))</td>
</tr>
<tr>
<td>takeoff</td>
<td>takeoff</td>
<td>(\text{occ}(t) + \text{wake}(t,t) + e(t,t))</td>
</tr>
<tr>
<td>landing</td>
<td>takeoff</td>
<td>(\text{occ}(t) + \text{wake}(1,t) + e(1,t))</td>
</tr>
</tbody>
</table>

the system is filled to capacity. The assumption of limited waiting space in the terminal area for landing aircraft has also been made by Raisbeck, Koopman, et al (1970).

4. Other Assumptions

In addition to the three major assumptions previously stated, several minor assumptions are made. It is assumed that all aircraft that enter the system receive service, i.e., no aircraft depart the holding pattern except to land and no aircraft depart the entry taxiway except to takeoff. Also, the order of service for landing and takeoff operations is based on a known sequencing rule, and this rule is followed without exception. Current Federal Aviation Administration regulations require landing operations to take non-preemptive priority over takeoff operations with a first come-first served order of service within the takeoff queue and the landing queue.

In summary, the air terminal system is viewed as a single service facility serving two types of customers, landing aircraft and takeoff aircraft. Both types of customers enter the system according to independent Poisson Laws. The service times are interactive due to the factors given in Table 26. In addition, the total number of aircraft waiting to takeoff and land is required to be less than a known constant. Accepting these basic assumptions, it is clear that the models developed in the previous chapters can be used to determine the effect of various sequencing rules on
the steady state performance of the air terminal system. Before discussing the use of the models, a brief review of similar analytical studies is given.

Related Studies

Air traffic control has received attention from operations researchers for a great many years as evidenced by the early report of Bowen and Pearcy (1948) and the bibliography of modeling efforts reported by Bell (1949). This attention has continued to the present day. A recent paper by Odoni (1971) lists 149 separate quantitative studies that treat air traffic control problems. The review given in this chapter is restricted to models that are similar to those proposed in the preceding section.

The analytic models of landing and takeoff operations can be roughly divided into two classes: saturation condition models and queueing models. The saturation condition models assume that landing and takeoff aircraft are always available to receive service and make no assumptions about the arrival process. As a result these models indicate only the throughput of the system in terms of the number of operations per unit time. Waiting time information is not obtainable. Service times are assumed to be highly interactive. Models of this type have been formulated by Blumstein (1960) and the National Bureau of Standards (1969).

The queueing models assume Poisson arrival distributions and attempt to provide information about waiting times. All
but two of the studies in this group are restricted to steady state results. Pestalozzi (1964) has used the standard M/G/1 queue with non-preemptive priorities as formulated by Cox and Smith (1961) to determine the effect of the assignment of strict priorities on the steady state waiting time of landing and takeoff aircraft. In this study service times are not interactive and are based strictly on runway occupancy time. Different types of aircraft are grouped according to the operation being performed and runway occupancy time into identifiable homogeneous customer classes. Each class is assigned a priority. Using data for runway occupancy time and waiting time cost from a study by Airborne Instruments Laboratory (1960), it was found that the assignment of priorities has little effect on average waiting time but average delay cost can be reduced appreciably.

Odoni (1970) has formulated a steady state queueing model for landing operations in which an aircraft that enters the system when the runway is idle receives service that is different from other landing aircraft. In all other respects, the model is identical to the standard M/G/1 queueing model. It was found that this model provides steady state mean waiting times and queue length distributions that are only slightly different from those provided by the standard M/G/1 queueing model with a single service time distribution.

Galliher and Wheeler (1958) have developed a queueing model for landing operations which treats the case of Poisson
arrivals with mean arrival rates that vary with time. Service times are assumed to be constant. The model provides the queue length distribution and waiting time distribution at the conclusion of each service time. The model was exercised with hypothetical mean demand rates of 0.7 to 1.1 arrivals per minute that changed at one hour intervals and with constant service times of one and two minutes. Results were inconclusive.

Raisbeck, Koopman, et al (1970) have investigated a queueing model for landing operations which treats the case of Poisson arrivals with mean arrival rates that vary with time. Service times are assumed to be exponentially distributed. In addition, the queue length is limited to a known finite capacity. The model provides the queue length distribution and the mean waiting time as a function of time. The model was exercised using arrival rate data for John F. Kennedy and LaGuardia Airports and mean service rates of 45, 55, and 75 aircraft per hour. It was found that the uncertainty in queue length as measured by the standard deviations of queue length exhibits some interesting properties. The standard deviation varies directly with the queue length when the queue is very short, varies inversely with the queue length when the queue is near capacity, and obtains its maximum value when the population of the queue is in a period of rapid transition. The authors indicate that the model can be extended to treat takeoff aircraft as well as landing
operations if the exponential service time assumption is retained. The equations for such models are given but not demonstrated.

It should be noted that none of the queueing models discussed in this review treat the case of interactive service times that follow general probability distributions. It would seem that the inclusion of such an assumption would improve the realism of models used to investigate the sequencing of landing and takeoff operations.

Model Application

The application of the models contained in this study for the investigation of sequencing of landing and takeoff operations is straightforward. The data needed for the application includes mean arrival rates for landing and takeoff aircraft, waiting room capacity, and the six interactive service time distributions indicated by Table 26. Recent private conversations with William Koch of the National Aviation Facilities Experimental Center, Atlantic City, New Jersey, revealed that such data were probably available from his agency. However, these data were not routinely used in their studies and a search of records would be necessary. Using real world data from this or any other source, the models could be exercised for the sequencing rules considered in Chapter IV and Chapter V. If additional sequencing rules are of interest, other models can be developed as indicated in Chapter V. Sensitivity analyses are also of interest as
the data are likely to represent only one airport performing under one set of operating conditions.

It should be noted that the steady state measures of system performance provide a wide range of information about the operation of the system. The expected number in the system and in the queue provides estimates of congestion on the takeoff entry ramp and estimates of congestion in the holding pattern. The expected waiting time in the queue and in the system indicates the delay costs incurred by landing and takeoff aircraft. Finally, the effective mean arrival rate indicates the actual throughput of the system in terms of expected number of operations per unit time. This piece of information is similar to the operations rate provided by the saturation condition models of Blumstein (1960). Thus the models of this study provide information about system throughput as well as congestion and delay. Queuing models with unlimited waiting space provide no useful information about system throughput as all customers desiring service are served and the output of the system equals the input. Saturation condition models provide no useful information about congestion as the arrival process is not considered.

The interpretation of the model results must be tempered by the fact that several assumptions simplify reality. The assumption of constant mean arrival rates is open to question as the arrival patterns at a number of airports have been shown to be highly time dependent, Raisbeck, Koopman, et al
The models also assume that the system has attained steady state conditions, which seems unlikely because of the time dependence of the mean arrival rates. In addition, in the model, the effect of limited waiting space is treated by requiring arriving aircraft to be forever lost to the system when the system is filled to capacity. In reality the aircraft would rejoin the arrival pattern at some future time.

In light of the above reservations the value of such steady state models might be questioned. Perhaps the best response to this inquiry has been given by Odoni (1971, p. 11). He states, "The steady state models are helpful in two respects: first, they can demonstrate graphically to airport administrators the drastic increase in expected delays at high utilization rates and the severe consequences of even minor degradations in service; secondly, for planning purposes, they are useful in identifying potentially troublesome situations that can result from chronic over-scheduling". It is hoped that these statements would be verified by the actual use of the models contained in this study.
CHAPTER VII
SUMMARY AND RECOMMENDATIONS

Introduction

The purpose of this chapter is to summarize the results contained in this study and to recommend areas for continued research. The introductory and summary sections of each chapter have provided a detailed account of the research contained therein and will not be repeated at length. The interested reader is referred to these sections for a thorough discussion of the research efforts reported in this dissertation.

Summary

This study consisted of developing models for a type of M/G/1 queueing system characterized by the following assumptions.

1. All customers that enter the system are classified as belonging to one of a finite number of homogeneous customer classes.

2. Customers from each customer class arrive at the system according to independent Poisson Laws.

3. All customers are served by a single service facility.
4. The service time is influenced by two factors: the class of customer receiving service and the class of customer, if any, that completed service immediately before the current customers service was initiated. These interactive service times are random variables with known general probability density functions.

5. The number of customers awaiting service at any time is required to be less than a known constant. Customers are not allowed to enter the system when the queue is filled to capacity.

6. The order of service is determined by a fixed sequencing rule.

The methodology used in the development of the models consisted of four distinct procedures. The net result of the procedures was to provide steady state probabilities of the number of each class of customer in the system. The four procedures are stated below.

1. The first procedure determines the steady state probability of the number of each class of customer in the queue at service completion epochs.

2. The second procedure determines the steady state probability of the number of each class of customer in the queue at service completion epochs and idle service initiation epochs.

3. The third procedure determines the steady state
probability of the number of each class of customer in the queue at the service completion epoch or idle service initiation epoch immediately preceding a randomly selected point in time.

4. The fourth procedure determines the steady state probability of the number of each class of customer in the system at a random point in time.

Using the steady state probabilities from the fourth procedure, the measures of system performance were calculated. The measures of system performance were effective mean arrival rate, expected number of customers in the system, expected number of customers in the queue, expected waiting time in the system, expected waiting time in the queue, and expected service time. These measures were calculated for each customer class as well as for the overall customer population.

The preceding procedures were demonstrated for a single customer class model in Chapter III. In this model the interactive nature of the service time was restricted to requiring the first customer of a busy period to receive unusual service. The procedures were demonstrated for a multiple customer class model in Chapter IV. The order of service was determined on the basis of a non-preemptive or head-of-the-line sequencing rule. The flexibility of the procedures was demonstrated in Chapter V by developing models for alternating priority, rotating priority, and first come-
first served sequencing rules. It was concluded that the modeling procedures were readily adaptable to a wide variety of sequencing rules. Chapter VI indicated the possible application of the models for an investigation of the sequencing of landing and takeoff operations at a single runway. It was concluded that such an application would demonstrate the effects of degradation of service and assist in identifying troublesome situations that result from over-scheduling.

**Recommendations**

The objectives of this study were stated in Chapter I to be the development of a general model building methodology and the provision of a number of specific models that would aid in the investigation of sequencing phenomena. It is concluded that these objectives have been met. The general technical procedures developed in Chapter IV have been shown to be readily adaptable to the formulation of models for a variety of sequencing rules. In addition, the formulations contained in Chapter IV and Chapter V have provided complete models for specific sequencing rules.

The findings of this research are by no means complete, however. A number of areas are worthy of continued research. A partial listing of possible fruitful research tasks is given below.

1. Develop an efficient computer code for the models
developed in the study in order to conduct sensitivity analyses and investigate computational aspects of the problems. A special purpose matrix language, such as MATLAN, International Business Machines Corporation (1969), might be useful in this effort.

2. Formulate models for additional sequencing rules, such as shortest mean processing time priority, using the model procedures given in Chapter IV.

3. Investigate the use of transform techniques to determine the probability generating functions of the steady state probabilities for the imbedded Markov chain, augmented imbedded Markov chain, semi-Markov process, and random point in time process. Initial efforts in this direction were abandoned because of the complexity of the transforms obtained.

4. Investigate model procedures for a modified system with infinite waiting space in the queue. Preliminary work in this area was undertaken using probability generating functions, but the approach was unsuccessful.

5. Conduct the investigation of takeoff and landing operations at a single runway as outlined in Chapter VI.

Two of the above suggestions are worthy of additional comment.
The development of an efficient computer code to assist in the calculations is essential to the study of large problems. From Equation (87) the total number of states required to describe a system with $k$ customer classes and waiting space for a total of $N$ customers is $k\binom{k+N}{N}$. Substitution of several values of $k$ and $N$ into this expression indicates the effect of a large number of customer classes and a large waiting space. For example, a system with waiting space for 10 customers requires a total of 11, 132, 858, 4004, and 15015 states for systems with 1, 2, 3, 4, and 5 customer classes, respectively. Similarly, a system with 2 customer classes requires a total of 132, 462, 870, 1560, and 2652 states for systems with waiting space for 10, 20, 30, 40, and 50 customers, respectively. To manipulate the matrices and perform the required operations in a reasonable amount of time, it is clear that computer assistance is necessary. Recall that many of the operations can be expressed in terms of matrix and vector operations. This indicates that a special purpose language such as MATLAB, International Business Machines Corporation (1969), would be appropriate. MATLAB is capable of efficiently dealing with large matrices as the language automatically provides storage management control over core, disk, and tape storage. Preliminary discussions with J. M. Snowden of The Ohio State University Instructional and Research Computing Center indicated that a $(k = 2, N = 20)$ system with 462 states could be
solved efficiently using MATLAN in conjunction with the IBM System 360/75 computer. This investigation was not pursued as the MATLAN program was not operational. It would seem that computer resources can be readily applied to the problem at hand.

Also of considerable interest is the application of the models in the investigation of sequencing of landing and takeoff operations. It would seem that the ultimate engineering payoff of this dissertation research would result from such a study. The tasks associated with this undertaking include data collection, sensitivity analyses, and model verification. Particular features of these efforts have been stated in Chapter VI. Another effort of interest includes a comparison of the effect of the rotating priority, alternating priority, and first come-first served priority sequencing rules with the current non-preemptive priority sequencing rule. This analysis should include an estimate of the acceptability of the sequencing procedure to air traffic controllers and users. For example, there would probably be considerable dissatisfaction with the alternating priority sequencing rule because of the long delays often incurred by aircraft that enter the system when their type of operation does not have top priority. On the other hand, the rotating priority sequencing rule might be more acceptable as it tends to provide a more equitable allocation of runway use. It is
hoped that an opportunity to conduct a complete applied study becomes available in the near future.

This research cannot be presented as providing the final word in the formulation of sequencing models that consider interactive service times. It is hoped, however, that a framework and basic methodology have been provided which will assist in future investigations of a class of problems that has heretofore received little attention.
APPENDIX A
ARRIVAL PROBABILITY DERIVATIONS

The following notation will be used in the derivations.

\[ p(a;rt) = \frac{(rt)^ae^{-rt}}{a!}, \quad a = 0,1,2,... \]

= Poisson probability mass function.

\[ f_s(t) = \text{probability density function of the service time t, defined for continuous positive values of } t. \]

\[ F_s(t) = \int_0^t f_s(v)dv = \text{cumulative distribution of } f_s(t). \]

\[ m_s = \int_0^\infty t f_s(t)dt = \text{mean of } f_s(t). \]

\[ \overline{f_s}(z) = \int_0^\infty e^{-zt}f_s(t)dt = \text{Laplace transform of } f_s(t). \]

\[ \overline{f_s}(z) = \frac{d^n\overline{f}(z)}{dz^n} = \text{nth derivative of the Laplace transform of } f_s(t), n = 0,1,2,... \]

\[ \overline{f_s}(r) = \overline{f_s}(z) \bigg| z = r = \text{nth derivative of the Laplace transform of } f_s(t) \text{ evaluated at } z = r, n = 0,1,2,... \]

\[ g(v;a,r) = \frac{r^a}{(a-1)!}v^{a-1}e^{-rv}, \quad v > 0 \]

= gamma probability density function.
\( g_c(t; a, r) = \int_0^t \frac{r^a}{(a-1)!} v^{a-1} e^{-rv} dv, \quad t > 0 \)

= gamma cumulative distribution function.

\( f'_s(t') = \) probability density function of the random interruption time, \( t' \), associated with the density function \( f_s(t) \).

\[ f'_s(t') = \frac{1 - F_s(t')}{m_s}, \quad t > 0. \]

\( \bar{F}'_s(z) = \) Laplace transform of the probability density function for the random interruption time associated with the density function, \( f_s(t) \).

\( \bar{F}'_s(z) = \) nth derivative of the Laplace transform of \( f'_s(t') \), \( n = 0, 1, 2, ... \)

\( \bar{F}'_s(r) = \) nth derivative of the Laplace transform of \( f'_s(t') \) evaluated at \( z = r \), \( n = 0, 1, 2, ... \)
Result A.1

\[ A(a; s) = \frac{(-r)^a}{a!} \bar{f}_s(r) , \quad a = 0, 1, 2, \ldots \]

Derivation:

\[ A(a; s) = \int_0^\infty p(a; rt) f_s(t) dt \]

\[ = \int_0^\infty \frac{e^{-rt} (rt)^a}{a!} f_s(t) dt \]

\[ = \frac{r^a}{a!} \int_0^\infty t^a e^{-rt} f_s(t) dt \]

\[ = \frac{r^a}{a!} \int_0^\infty e^{-rt} t^a f_s(t) dt . \]

By definition of the Laplace transform

\[ \int_0^\infty e^{-rt} t^a f_s(t) dt = \mathcal{L}(t^a f_s(t)) \bigg|_{z = r} , \]

and from Holl, Maple and Vinograde (1959, p. 52)

\[ \mathcal{L}(t^a f_s(t)) = (-1)^a \bar{f}_s^{(a)}(z) \bigg|_{z = r} . \]

where \( \bar{f}(z) \) is the \( a \)th derivative of the Laplace transform of \( f_s(t) \).

Therefore

\[ A(a; s) = (-1)^a \frac{\bar{f}_s^{(a)}(a)}{a!} . \]

Substituting the result into the expression

\[ A(a; s) = \frac{r^a}{a!} (-1)^a \frac{\bar{f}_s^{(a)}(a)}{a!} \]

\[ = \frac{(-r)^a}{a!} \bar{f}_s(r) . \]
Result A.2
\[ \bar{A}(a;s) = 1 - \sum_{n=0}^{a-1} \frac{(-r)^n}{n!} f_s(r) , \quad a = 1,2,... \]

Derivation
\[ \bar{A}(a;s) = \int_0^\infty \int_0^t \frac{p(a-1;rv)}{v!} dv \int_0^v f_s(t)dt \]
\[ = \int_0^\infty \int_0^t \frac{rv^{a-1}e^{-rv}}{(a-1)!} \frac{1}{v!} dv \int_0^v f_s(t)dt \]
\[ = \int_0^\infty \int_0^t \frac{r^a}{(a-1)!} v^{a-1} e^{-rv} dv \int_0^v f_s(t)dt. \]

The inner integral is the gamma cumulative distribution and is related to the Poisson distribution by the well-known identity, Parzen (1960, p. 261),
\[ \int_0^t \frac{r^a}{(a-1)!} v^{a-1} e^{-rv} dt = 1 - \sum_{n=0}^{a-1} p(n;rt). \]

Substituting this identity into the expression
\[ \bar{A}(a;s) = \int_0^\infty \left( 1 - \sum_{n=0}^{a-1} p(n;rt) \right) f_s(t)dt \]
\[ = \int_0^\infty f_s(t)dt - \int_0^\infty \sum_{n=0}^{a-1} p(n;rt)f_s(t)dt \]
\[ = 1 - \sum_{n=0}^{a-1} \int_0^\infty p(n;rt) f_s(t)dt. \]

From Result A.1
\[ \int_0^\infty p(n;rt) f_s(t)dt = \frac{(-r)^n}{n!} f_s(r). \]

Substituting this identity yields the desired identity
\[ \bar{A}(a;s) = 1 - \sum_{n=0}^{a-1} \frac{(-r)^n}{n!} f_s(r) . \]
Result A.3

\[ \bar{f}'_s(r) = \frac{a!}{(-r)^a m_s r} \left[ 1 - \sum_{n=0}^{\infty} \frac{(-r)^n}{n!} \bar{f}_s^{(n)}(r) \right], \quad a = 0, 1, 2, \ldots \]

Derivation:

\[ \bar{f}'_s(r) = \frac{d}{dz} \left[ \frac{1 - F_s(t)}{m_s} \right] \bigg|_{z = r} \cdot \]

Examining the Laplace transform and using an identity from Holl, et al (1959, p. 47)

\[ \mathcal{L} \left[ \frac{1 - F_s(t)}{m_s} \right] = \frac{1}{m_s} \left[ \mathcal{L}(1) - \mathcal{L}(F_s(t)) \right] \]

\[ = \frac{1}{m_s} \left[ \frac{1}{z} - \frac{1}{z} \bar{F}_s(z) \right] \]

\[ = \frac{1}{m_s z} \left[ 1 - \bar{F}_s(z) \right]. \]

Substituting this result

\[ \bar{f}'_s(r) = \frac{d}{dz} \left[ \frac{1}{m_s z} \left[ 1 - \bar{F}_s(z) \right] \right] \bigg|_{z = r} \cdot \]

From the CRC Standard Mathematical Tables (1965, p. 304) the identity for the ath derivative of a product is

\[ \frac{d^a}{dz^a} [uv] = \sum_{n=0}^{a} \binom{a}{n} \frac{d^n u}{dz^n} \frac{d^{(a-n)} v}{dz^{(a-n)}}. \]

Applying this identity

\[ \bar{f}'_s(r) = \frac{1}{m_s} \sum_{n=0}^{a} \binom{a}{n} \frac{d^n (1/z)}{dz^n} \frac{d^{(a-n)} (1 - \bar{F}_s(z))}{dz^{(a-n)}} \bigg|_{z = r} , \]

where

\[ \frac{d^n (1/z)}{dz^n} = n! \frac{(-1)^n}{z^{n+1}}, \quad n = 0, 1, 2, \ldots \]
\[
\frac{d^{(a-n)}(1 - \bar{f}(z))}{dz^{(a-n)}} = d^{(a-n)} - \bar{f}(z), \quad n = 0, 1, 2, \ldots, a
\]

and the Kronecker delta is defined as
\[
d^{(a)} = \begin{cases} 
1, & n = a \\
0, & \text{otherwise}.
\end{cases}
\]

The \(a\)th derivative of the Laplace transform is
\[
\bar{f}^{(a)}(s) = \frac{1}{ms} \sum_{n=0}^{a} \frac{(a)_n}{n!} (-1)^n \frac{d^{(a-n)} - \bar{f}(z)}{z^{n+1}} \bigg|_{z=r}
\]
\[
= \frac{1}{ms} \sum_{n=0}^{a} \frac{(a)_n}{n!} (-1)^n \frac{d^{(a-n)} - \bar{f}(z)}{z^{n+1}} \bigg|_{z=r}
\]
\[
= \frac{1}{ms} \left[ (a) \sum_{n=0}^{a} \frac{(-1)^n}{n!} \frac{d^{(a-n)} - \bar{f}(z)}{z^{n+1}} \bigg|_{z=r}
\right.
\]
\[
= \frac{1}{ms} \left[ (a) \sum_{n=0}^{a} \frac{(-1)^n}{n!} \frac{d^{(a-n)} - \bar{f}(z)}{z^{n+1}} \bigg|_{z=r}
\right.
\]
\[
= \frac{a}{(-z)^{a} ms z^a} \left[ 1 - \frac{a}{\sum_{n=0}^{a} \frac{(-z)^{a-n}}{n!} \bar{f}(z) } \bigg|_{z=r}
\right.
\]
\[
= \frac{a}{(-z)^{a} ms z^a} \left[ 1 - \frac{a}{\sum_{n=0}^{a} \frac{(-z)^{a-n}}{n!} \bar{f}(z) } \bigg|_{z=r}
\right.
\]
\[
= \frac{a}{(-z)^{a} ms z^a} \left[ 1 - \frac{a}{\sum_{n=0}^{a} \frac{(-z)^{a-n}}{n!} \bar{f}(z) } \bigg|_{z=r}
\right.
\]
\[
= \frac{a}{(-z)^{a} ms z^a} \left[ 1 - \frac{a}{\sum_{n=0}^{a} \frac{(-z)^{a-n}}{n!} \bar{f}(z) } \bigg|_{z=r}
\right.
\]
\[
= \frac{a}{(-z)^{a} ms z^a} \left[ 1 - \frac{a}{\sum_{n=0}^{a} \frac{(-z)^{a-n}}{n!} \bar{f}(z) } \bigg|_{z=r}
\right.
\]
\[
= \frac{a}{(-z)^{a} ms z^a} \left[ 1 - \frac{a}{\sum_{n=0}^{a} \frac{(-z)^{a-n}}{n!} \bar{f}(z) } \bigg|_{z=r}
\right.
\]
\[
= \frac{a}{(-z)^{a} ms z^a} \left[ 1 - \frac{a}{\sum_{n=0}^{a} \frac{(-z)^{a-n}}{n!} \bar{f}(z) } \bigg|_{z=r}
\right.
\]
\[
= \frac{a}{(-z)^{a} ms z^a} \left[ 1 - \frac{a}{\sum_{n=0}^{a} \frac{(-z)^{a-n}}{n!} \bar{f}(z) } \bigg|_{z=r}
\right.
\]
\[
= \frac{a}{(-z)^{a} ms z^a} \left[ 1 - \frac{a}{\sum_{n=0}^{a} \frac{(-z)^{a-n}}{n!} \bar{f}(z) } \bigg|_{z=r}
\right.
\]
\[
= \frac{a}{(-z)^{a} ms z^a} \left[ 1 - \frac{a}{\sum_{n=0}^{a} \frac{(-z)^{a-n}}{n!} \bar{f}(z) } \bigg|_{z=r}
\right.
\]
\[
= \frac{a}{(-z)^{a} ms z^a} \left[ 1 - \frac{a}{\sum_{n=0}^{a} \frac{(-z)^{a-n}}{n!} \bar{f}(z) } \bigg|_{z=r}
\right.
\]
\[
= \frac{a}{(-z)^{a} ms z^a} \left[ 1 - \frac{a}{\sum_{n=0}^{a} \frac{(-z)^{a-n}}{n!} \bar{f}(z) } \bigg|_{z=r}
\right.
\]
Result A.4

\[ A^*(a;s) = \frac{1}{m^r_s} \left[ 1 - \sum_{n=0}^{a} \frac{(-r)^n}{n!} f_s^{(n)}(r) \right], \quad a = 0, 1, 2, \ldots \]

Derivation:

\[ A^*(a;s) = \int_0^\infty p(a;r't') f'_s(t') \, dt'. \]

From Result A.1

\[ \int_0^\infty p(a;r't') f'_s(t') \, dt' = \frac{(-r)^a}{a!} f_s^{(a)}(r). \]

Substituting the expression for \( f_s^{(a)}(r) \) as given by Result A.3

\[ A^*(a;s) = \frac{(-r)^a}{a!} \left[ \frac{a!}{(-r)^a m^r_s} \left[ 1 - \sum_{n=0}^{a} \frac{(-r)^n}{n!} f_s^{(n)}(r) \right] \right] \]

\[ = \frac{1}{m^r_s} \left[ 1 - \sum_{n=0}^{a} \frac{(-r)^n}{n!} f_s^{(n)}(r) \right]. \]
Result A.5

\[ A'(a;s) = 1 - \frac{1}{m_{sr}} \left[ a - \sum_{n=0}^{a-1} \sum_{h=0}^{n} (-r)^{h} \frac{F_{s}(r)}{h!} \right] , \quad a = 1, 2, \ldots \]

Derivation:

\[ A'(a;s) = \int_{0}^{\infty} \int_{0}^{t'} p(a-1;rv) rdv f_{s}'(t')dt' . \]

From Result A.2

\[ \int_{0}^{\infty} \int_{0}^{t'} p(a-1;rt') rdv f_{s}'(t')dt' = \left( -r \right)^{n} \frac{\Gamma(n)}{n!} \frac{F_{s}'(r)}{s} . \]

Substituting the expression for \( \frac{\Gamma(n)}{n!} \frac{F_{s}'(r)}{s} \) as given by Result A.3

\[ A'(a;s) = 1 - \sum_{n=0}^{a-1} \frac{(-r)^{n}}{n!} \frac{F_{s}'(r)}{s} \left[ 1 - \sum_{h=0}^{n} (-r)^{h} \frac{F_{s}(r)}{h!} \right] . \]

\[ = 1 - \sum_{n=0}^{a-1} \frac{1}{m_{sr}} \left[ 1 - \sum_{h=0}^{n} (-r)^{h} \frac{F_{s}(r)}{h!} \right] . \]
Result A.6

\[ A([a_i];s) = \left[ \left( \frac{r_i}{r} \right)^{a_i} \right] A(a; s), \quad a = 0, 1, 2, \ldots \]

where

\[ [a_i] = a_1, a_2, \ldots, a_k \]

\[ ([a_i]^a) = \frac{a!}{a_1! a_2! \ldots a_k!} \]

\[ a = \sum_{i=1}^{k} a_i, * \]

\[ r = \sum_{i=1}^{k} r_i, * \]

and

\[ A(a; s) \text{ is defined by Result A.1.} \]

* For convenience the limits on the summation are deleted within the derivation.

Derivation:

\[ A([a_i];s) = \int_0^\infty \left[ \prod_{i=1}^{k} p(a_i; r_i t) \right] f_s(t) dt \]

\[ = \int_0^\infty \left[ \prod_{i=1}^{k} \frac{(r_i t)^{a_i} e^{-r_i t}}{a_i!} \right] f_s(t) dt \]

\[ = \left[ \prod_{i=1}^{k} \frac{r_i^{a_i}}{a_i!} \right] \int_0^\infty t^{a_i} e^{-(\Sigma r_i)t} f_s(t) dt \]

\[ = \left[ \prod_{i=1}^{k} \frac{r_i^{a_i}}{a_i!} \right] \int_0^\infty t^a e^{-rt} f_s(t) dt \]

\[ = \left[ \prod_{i=1}^{k} \frac{r_i^{a_i}}{a_i!} \right] \int_0^\infty \frac{r^a e^{-rt}}{a!} f_s(t) dt \]
By rearranging terms and recognizing the expression for $A(a;s)$ as given in Result A.1, the expression reduces to

$$A([a_i],s) = \left[ \frac{\prod_{i=1}^{k} \frac{r_i}{a_i^s}}{\prod_{i=1}^{k} \frac{r_i}{a_i}} \right] \frac{\prod_{i=1}^{k} \frac{r_i}{a_i}}{\prod_{i=1}^{k} \frac{r_i}{a_i}} A(a;s), \quad a = 0,1,2,\ldots$$

$$= \left[ \frac{\prod_{i=1}^{k} \frac{r_i}{a_i}}{\prod_{i=1}^{k} \frac{r_i}{a_i}} \right] A(a;s), \quad a = 0,1,2,\ldots$$
Result A.7

\[ \bar{A}(\{a_i\}; s) = \left[ \sum_{j \in J} \left( a_i - d_{ij} \right) \prod_{i=1}^{k} \frac{r_i}{x} \right] \bar{A}(a_i; s), \quad a = 1, 2, \ldots \]

where

\[ J = \{i; a_i > 0, i = 1, 2, \ldots, k\}, \]

\[ d_{ij} = \begin{cases} 1, & i = j, \\ 0, & \text{otherwise}. \end{cases} \]

the Kronecker delta,

\[ \left( a_i - d_{ij} \right) = a_i - a_1 - a_2 - \ldots - a_{j-1} - (a_j - 1) - a_{j+1} - \ldots - a_k. \]

\[ \bar{A}(a_i; s) \] is defined by Result A.2,

and the remaining notation is defined as for Result A.6.

Derivation: \[ \bar{A}(\{a_i\}; s) \]

\[ = \int_0^\infty \left[ \sum_{j \in J} \int_0^t \left[ \prod_{i=1}^{k} \frac{r_i}{x} \right] \bar{A}(a_i - d_{ij}; r_i v) \right] r_j dv \int_0^t f_s(t) dt \]

\[ = \int_0^\infty \left[ \sum_{j \in J} \int_0^t \frac{k}{\prod_{i=1}^{k} \frac{r_i}{x}} \bar{A}(a_i - d_{ij}; r_i v) \right] r_j dv \int_0^t f_s(t) dt \]

\[ = \int_0^\infty \left[ \sum_{j \in J} \left[ \frac{k}{\prod_{i=1}^{k} \frac{r_i}{x}} \right] r_j \int_0^t \bar{A}(a_i - d_{ij}; r_i v) \right] \int_0^t f_s(t) dt \]

\[ = \int_0^\infty \left[ \sum_{j \in J} \left[ \frac{k}{\prod_{i=1}^{k} \frac{r_i}{x}} \right] r_j \int_0^t \bar{A}(a_i - d_{ij}; r_i v) \right] \int_0^t f_s(t) dt \]

\[ = \int_0^\infty \left[ \sum_{j \in J} \left[ \frac{k}{\prod_{i=1}^{k} \frac{r_i}{x}} \right] r_j \int_0^t \bar{A}(a_i - d_{ij}; r_i v) \right] \int_0^t f_s(t) dt \]
\[ \bar{A}(\{a_i\}; s) \]

\[ = \int_0^\infty \left[ \sum_{j \in J} \left( \frac{k}{\pi (a_i - d_{ij})^2} \right) \frac{(a-1)\alpha}{r^\alpha} \int_0^t \frac{r^\alpha}{(a-1)^\alpha} v^{a-1} e^{-rv} dv \right] f_s(t) dt \]

\[ = \left[ \sum_{j \in J} \left( \frac{k}{\pi (a_i - d_{ij})^2} \right) \frac{(a-1)\alpha}{r^\alpha} \int_0^\infty t^{n} r^\alpha v^{a-1} e^{-rv} dv \right] f_s(t) dt \]

\[ = \left[ \sum_{j \in J} \left( \frac{k}{\pi (a_i - d_{ij})^2} \right) \frac{(a-1)\alpha}{r^\alpha} \int_0^\infty \frac{t}{(a-1)^\alpha} e^{-rv} rdv \right] f_s(t) dt \]

\[ = \left[ \sum_{j \in J} \left( \frac{k}{\pi (a_i - d_{ij})^2} \right) \frac{(a-1)\alpha}{r^\alpha} \int_0^\infty \frac{p(a-1; rv)}{r^\alpha} rdv \right] f_s(t) dt \]

Recognizing the expression for \( \bar{A}(a; s) \) as given in Result A.2 and utilizing the initial definitions, the terms reduce to

\[ \bar{A}(\{a_i\}; s) = \left[ \sum_{j \in J} \left( \frac{k}{\pi (a_i - d_{ij})^2} \right) \frac{(a-1)\alpha}{r^\alpha} \right] \bar{A}(a; s), \quad a = 1, 2, \ldots \]
Result A.8

\[ A'(\{a_i\};s) = \left( \prod_{i=1}^{a} \left( \frac{r_i}{r} \right)^{a_i} \right) A'(a;s), \quad a = 0,1,2,\ldots \]

where

\[ A'(a;s) \] is defined by Result A.4, and the remaining notation is defined as for Result A.6.

Derivation:

\[ A'(\{a_i\};s) = \int_{0}^{\infty} \left( \prod_{i=1}^{a} p(a_i; r_it) \right) f_s(t') dt'. \]

Performing the same sequence of manipulations as used in Result A.6 leads to the intermediate result

\[ A'(\{a_i\};s) = \left( \prod_{i=1}^{a} \left( \frac{r_i}{r} \right)^{a_i} \right) \int_{0}^{\infty} p(a;rt') f_s(t') dt'. \]

By rearranging terms and recognizing the expression for \( A'(a;s) \) as given in Result A.4, the expression reduces to

\[ A'(\{a_i\};s) = \left( \frac{a!}{a_1! \cdot a_2! \cdot \ldots \cdot a_k!} \right) \left( \prod_{i=1}^{a} \left( \frac{r_i}{r} \right)^{a_i} \right) A'(a;s) \]

\[ = \left[ \left( \prod_{i=1}^{a} \frac{r_i}{r} \right)^{a_i} \right] A'(a;s), \quad a = 0,1,2,\ldots \]
Result A.9

\[ \bar{A}'([a_i]; s) = \left[ \sum_{j \in J} \left( \frac{a_i - d_{ij}}{r_i} \right)^{a_i} \right] ^k \prod_{i=1}^k \left( \frac{r_i}{r} \right) \bar{A}'(a; s), \quad a = 1, 2, \ldots \]

where

\[ \bar{A}'(a; s) \text{ is defined by Result A.5,} \]

and the remaining notation is defined as for Result A.7.

Derivation:

\[ \bar{A}'([a_i]; s) = \int_0^\infty \left[ \sum_{j \in J} \int_0^{t'} \left[ \prod_{i=1}^k p(a_i - d_{ij}; r_i v) \right] r_j dv \right] f_s(t') dt'. \]

Performing the same sequence of manipulations as used in Result A.7 leads to the intermediate result

\[ \bar{A}'([a_i]; s) = \left[ \sum_{j \in J} \left( \frac{a_1 - a_2 - \ldots - a_j - \ldots - a_k \cdot \ldots \cdot a_l}{a - 1} \right) \right] ^k \prod_{i=1}^k \left( \frac{r_i}{r} \right) \cdot \int_0^\infty \int_0^{t'} p(a-1; rv) \, rdv \, f_s(t') dt'. \]

Recognizing the expression for \( \bar{A}'(a; s) \) as given in Result A.5 and utilizing the definitions of Result A.6, the terms reduce to

\[ \bar{A}'([a_i]; s) = \left[ \sum_{j \in J} \left( \frac{a_i - d_{ij}}{r_i} \right)^{a_i} \right] ^k \prod_{i=1}^k \left( \frac{r_i}{r} \right) \bar{A}'(a; s), \quad a = 1, 2, \ldots \]
BIBLIOGRAPHY


17. Fox, B. L. and Miller, L. W., Equivalent Stationary Measures for the $M/G/1$ Queue, RAND Corp. Report No. P-3590, Santa Monica, California, 1967.


48. Singer, A., "The Steady State Probabilities of a Markov Chain as a Function of the Transition Probabili-

