INFORMATION TO USERS

This material was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.

2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.

4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

5. PLEASE NOTE: Some pages may have indistinct print. Filmed as received.

Xerox University Microfilms
300 North Zeeb Road
Ann Arbor, Michigan 48106

73-24,697

TAUBE-NETTO, Miguel, 1936-
TWO QUEUES IN TANDEM ATTENDED BY A SINGLE SERVER.

The University of Michigan, Ph.D., 1973
Engineering, industrial

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED.

TWO QUEUES IN TANDEM ATTENDED BY A SINGLE SERVER

by
Miguel Taube-Netto

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial Engineering)
in The University of Michigan
1973

Doctoral Committee:

Professor Herbert P. Galliher, Chairman
Professor Ralph L. Disney
Professor Arch W. Naylor
Associate Professor Stephen M. Pollock

ACKNOWLEDGMENTS

The author wishes to express his gratitude and appreciation to his doctoral committee and especially to his chairman, Professor H. P. Galliher for his support and guidance throughout the course of this work. This research would not have been completed without his encouragement during difficult periods.

Special thanks are also due to Mrs. Wanita Rasey, typist, for her excellent and prompt work in the preparation of the manuscript and to Dr. Jose Busnardo-Neto in the preparation of computer-generated plots.

The author's graduate studies were sponsored by the Brazilian government through the Conselho Nacional de Pesquisas, the Universidade de Brasilia, and the Instituto de Pesquisas Espaciais.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	ii
LIST OF ILLUSTRATIONS	v
LIST OF TABLES	vi
CHAPTER I. INTRODUCTION	1
1. The Problem	1
2. Background	5
3. Summary of Results and Organization	9
4. Terminology and Methods	12
CHAPTER II. LEAVING NO BACKLOG	16
1. Introduction	16
2. Stationary Conditions (Instantaneous Switching)	16
3. Determination of $U_1(0,y)$, $U_2(x,0)$ and $\pi(2,0,0)$	18
4. Expected Number of Customers	25
5. Waiting Times	27
6. Busy Periods	29
7. Alternative Approach for the Calculation of the State Probabilities	32
8. Comparison of the Methods of Sections 3 and 7	33
a. Computations Using the Iteration Scheme of Section 3	34
b. Computations Using the Method of Section 7	36
9. The Case of Random Duration of Switching	38
9.1 Expected Number of Customers	43
9.2 Waiting Times	45
9.3 Alternative Approach for the Calculation of the State Probabilities	45
9.4 Busy Periods	47
CHAPTER III. SIMPLE NONPREEMPTIVE PRIORITY	50
1. Introduction	50
2. The SSP Policy	50

	Page
3. The FSP Policy	54
3.1 Expected Numbers of Customers and Waiting Times	56
3.2 Busy Periods	57
3.3 The Case of Random Duration of Switching	59
CHAPTER IV. RELUCTANT SWITCHING	68
1. Introduction	68
2. The SSS ($Q_2 = N$) Policy	69
2.1 Calculation of $r_{\ell m}$, $z_{k\ell}$, β_{ck} and α_{ac}	76
2.2 Expected Number of Customers and Waiting Times	79
2.3 The SSS ($Q_2 = 2$) Policy	80
2.4 Busy Periods	82
3. The W_NFS Policy	84
3.1 Expected Number of Customers and Waiting Times	90
3.2 Busy Periods	91
4. The SFS ($Q_1 \geq N$) Policy	94
4.1 Expected Number of Customers and Waiting Times	98
4.2 Busy Periods	99
CHAPTER V. DISCUSSION	102
1. Numerical Comparison of Policies	102
2. Practical Implications of Availability of Numerical Results	110
3. History of the Work	112
4. Related Topics and Possible Extensions	115
LIST OF REFERENCES	118

LIST OF ILLUSTRATIONS

Figure		Page
I	Illustration of the Iteration Scheme for Finding $U_1(0,y)$ and $U_2(x,0)$.	22
II	Average Waiting Times in the Stages of the System, For Policies LNB, SSP and FSP, in the Case of Exponential Services, Plotted as a Function of the Traffic Intensity in the First Stage, ρ_1, for Constant Total Traffic Intensity ($\rho = 0.1$). (a) First Stage; (b) Second Stage.	104
III	Average Waiting Times in the Stages of the System, for Policies LNB, SSP and FSP, in the Case of Exponential Services, Plotted as a Function of the Traffic Intensity in the First Stage, ρ_1, for Constant Total Traffic Intensity ($\rho = 0.5$). (a) First Stage; (b) Second Stage.	105
IV	Average Waiting Times in the Stages of the System, for Policies LNB, SSP and FSP, in the Case of Exponential Services, Plotted as a Function of the Traffic Intensity in the First Stage, ρ_1, for Constant Total Traffic Intensity ($\rho = 0.9$). (a) First Stage; (b) Second Stage.	106
V	Average Busy Period of the Stages of the System, for Policies LNB, SSP and FSP, in the Case of Exponential Services, Plotted as a Function of the Traffic Intensity in the First Stage, ρ_1, for Constant Total Traffic Intensity ($\rho = 0.1$). (a) First Stage; (b) Second Stage.	107
VI	Average Busy Period of the Stages of the Syste, for Policies LNB, SSP and FSP, in the Case of Exponential Services, Plotted as a Function of the Traffic Intensity to the First Stage, ρ_1, for Constant Traffic Intensity ($\rho = 0.5$). (a) First Stage; (b) Second Stage.	108
VII	Average Busy Period of the Stages of the System, for Policies LNB, SSP and FSP, in the Case of Exponential Services, Plotted as a Function of the Traffic Intensity to the First Stage, ρ_1, for Constant Traffic Intensity ($\rho = 0.9$). (a) First Stage; (b) Second Stage.	109

LIST OF TABLES

Table		Page
I	Results.	11
II	Values of $Q_{.2}$ Obtained by Simulation and By Equation 4.9.	27
III	Illustration of the Iterative Scheme of Section 3 for $\rho_1 = \rho_2 = 0.4$.	35
IV	Values of $\bar{c}$ for Different Traffic Intensities.	36
V	Values of $\bar{c}$ Calculated by the Direct Approach, for Different Traffic Intensities and Different Truncations of the Infinite Summations.	37

CHAPTER I

INTRODUCTION

1. The Problem

In this study, we consider certain queueing situations in which customers arrive at a two-stage queueing system according to a Poisson process with constant arrival rate, λ. Each customer requires two sequential types of servicings. Both servicings are performed by a single server, who switches from one type of service (one stage of the system) to the other, according to some switching rule (policy). The server can even be temporarily removed from the system and, therefore, not available in either stage. The special case of instantaneous switching will recieve most of our attention in this thesis, however, for some switching rules, we shall consider the length of the switching times (or simply the switching times) as random variables with arbitrary distributions with finite moments.

The service time processes in each stage are assumed to constitute independent renewal processes with distributions $S_1(t)$ and $S_2(t)$. These processes are also assumed to be stochastically independent of the arrival and switching time processes.

In the analysis of this problem we shall assume infinite queue capacity in front of each service station. A customer arriving at the system will wait for service in the first stage unless the server is ready to service him in that stage, in which case, service immediately begins. When a customer completes service in the first stage, he

immediately goes to the second stage, where he will join the second queue, except that if the server moves with him (with zero switching time) to the second stage and no other customer is already waiting there, then his service in the second stage will commence without any wait. In each stage, customers are served in their order of arrival. How long customers wait depends somewhat upon the policy which the server follows in switching from stage to stage. We will consider several alternative policies, described below.

When a customer is in service in either stage the server waits until the current servicing is completed and then decides whether to remain in the present stage, switch to the other stage or remove himself from the system (the three possible states of the server). When the server is idle (removed or not busy because his present state is empty) he waits in the stage until the next arrival and then decides where to locate himself.

A criterion for these decisions by the server could be obtained if we are able to attach a cost structure to the behavior of the system. This structure would include, for instance, the cost of service in each stage, the waiting costs, idle costs and switching costs. With such a quantitative cost description we could use the expected average cost or the expected total discounted cost, as optimization criteria, and apply the techniques of Markov renewal programming (in the case of finite queue capacities) to find an optimal decision policy. However, due to the size of the state space (number of customers in each stage and the state of the server), the amount of calculation is very large. Also, as we may intuitively expect, an optimal policy might be quite complex and difficult to implement.

An alternative to this situation is to restrict our attention to "simple" decision policies for which a complete analysis of the behavior of the system is feasible. These policies may be suboptimal, but because we know better their effects on the system, we may consider them for situations where only a qualitative aspect of the costs is known. This is the approach of this study.

We consider six policies. In fact three of these policies are particular cases of one of the others, however there are advantages to presenting all of them separately. Since we are only interested in the steady-state effects of each policy, we assume that the policy becomes effective when the system is first emptied. The six policies are:

1. LNB (leaving no backlog). Switch to the other state whenever the number of customers in the present stage reaches the value zero.

2. SSP (second stage priority). Switch to the other stage whenever a service is completed in the present stage. In this case the server follows the customers through the two stages of the system. This is merely an M/G/1 queueing situation.

3. FSP (first stage priority). Customers in the first stage have nonpreemptive priority over customers in the second stage. This means that the server switches to the second stage when the first stage becomes empty, and returns to the first stage when, at a departure from the second stage, there is at least one customer in the first stage, or when no customers are left behind in the second stage.

For the next two policies, define Q_k to be the number of customers in stage k, k = 1,2.

4. SSS ($Q_2 = N$). Switch to the second stage when $Q_2 = N$ and return to the first stage when all of these N customers are attended. (If the initial value of Q_2 is greater than N, the server can, for instance, start by attending all Q_2 customers.)

5. SFS ($Q_1 \geq N$). The server switches to the second stage only when the first stage becomes empty. He stays in the second stage until it becomes empty, unless the number of customers in the first stage, at service completion epochs in the second stage, is at least N. In this case the server returns to the first stage. SFS ($Q_1 \geq N$) stands for switch to the first stage when $Q_1 \geq N$.

6. W_NFS (wait for N customers in the first stage). Switch to the second stage when the number of customers in the first stage is zero; attend all customers accumulated in the second stage (a total of at least N) and then go back to the first stage if the number of customers in it is at least N, otherwise remove the server from the system until N customers are accumulated in the first stage (this value, by the nature of the arrival process, is reached with probability one in a finite time).

Notice that under these policies, except the last one, the server is never removed from the system. Notice also that LNB $\equiv$ W_0FS (or LNB $\equiv$ W_1FS), FSP $\equiv$ SFS($Q_1 \geq 1$), SSP $\equiv$ SSS($Q_2 = 1$) and LNB $\equiv$ SSS($Q_2 = \infty$).

Justifications for considering these switching rules are as follows: If the switching costs are negligible when compared with waiting costs,

we may use the SSP policy in the case of high waiting cost in the second stage, and the FSP policy in the case of high waiting cost in the first stage. Under the LNB policy, as under the previous ones, the server tends to switch too many times when the traffic intensities are low, however this may be a good policy when the switching costs are small and the waiting cost in each stage are of the same order of magnitude. Policies $SSS(Q_2 = N)$, $SFS(Q_1 \geq N)$ and $W_N FS$, which we call "reluctant switching policies," are designed for situations of high change over costs. The policy $W_N FS$ is convenient when the server charges for his services even if he is idle (but present) in the system; under this policy the idle periods occur only when the server is removed from the system and, therefore, free of charge.

2. Background

In this study it has not been our objective to discover or develop practical applications of the above queueing model. However it seems that this model could represent various practical situations of sequential service operations. For example: a doctor who first attends his patient for preliminary examination then, in a second phase, takes care of more detailed diagnosis; a repair man who takes care of machine breakdown on a provisory basis, until there are enough machines with provisory repairs to justify a second phase of service where each machine is overhauled; a small computer installation where the compilation and execution of program takes loading and unloading of the compiler (which is stored in a deck of cards). An interesting possible application is the case of two bridges over a river where ships,

in order to pass under each bridge must have each one of them open. The automobile traffic over the bridges is interrupted when a ship is passing under it. Suppose that at least one of the bridges must be available for automobile traffic. Suppose that the arrival of ships from one direction (the ships coming from the other direction can be scheduled) is a Poisson process. Here the bridge open to ships represents an available server.

The basic feature of our model is the fact that each departure from the first stage corresponds to an arrival at the second stage. This arrival need not be the same unit departing from the first stage. Thus, it is not necessary, as in the examples above, to have a unit (patient, machine, computer program or ship) flowing through the system. We can imagine a situation where a repair man uses a special set of tools for each job. Each set of tools is used only in one job, and they have to be cleaned up by the repair man and returned to a tool crib. The cleaning operating can be performed at any time. Here each completion of a repair generates a set of dirty tools which will joint the queue of the second stage of the system.

Numerous investigators have studied sequential queueing systems, but only Nelson [13] has considered the case of a single server. He considers two service centers in tandem with Poisson arrival and exponential service times in each center. A single server switches from one center to the other according to five switching rules, for example: 1) when the server completes service, his next assignment is to perform an operation on that job which has spent the longest time

in the system; 2) when the server completes a processing operation, his next assignment is to perform an operation on the job that requires the least service time; 3) the server switches to the other center when the present center is empty. Nelson uses simulation in order to find the mean and the variance of the time in the system, for various values of the traffic intensities of each center. In another paper, Nelson [12] studies the case of assigning a single server to one of two machines in tandem, with the objective of minimizing total in-process inventory cost, over the time period $0 \leq t \leq T$. He assumes deterministic arrivals and the service times are known a priori. This is a case of what Cooper [6] calls scheduled services (this means that the service times are known when the customer or job enters the system. Nonscheduled services is the case in which the service time is drawn from a certain distribution at the time the service is initiated). Recently, Sahney [17] studied the problem of two machines and one man with setup times, where each job is processed only in one machine. He assumes that all jobs are available for processing at the time zero, and obtains an algorithm to solve the problem of finding the sequence of jobs which minimizes the mean flow time of jobs. This corresponds to the scheduled situation of two parallel service stations and one server. The nonscheduled case, in which the arrivals to each station are independent Poisson processes, is quite relevant to this thesis because the methods of analysis are somewhat related. In Section 3 of Chapter II we shall refer to Takács' [20] work, where he examines the problem of two queues in parallel, with independent Poisson arrival processes and

general renewal service time processes, attended by a single server, who switches, without delay, from one queue to the other when the present queue is empty. He obtains the Laplace-Stieltjes transform of the waiting times in each queue. The case of random length of switching times is studied by Eisenberg [8]. He considers two switching rules (switch when present queue is empty, and the case in which one of the queues has nonpreemptive priority over the other) and finds the Laplace-Stieltjes transform of the waiting times. Recently, Eisenberg [9] has extended his retults to a system of queues that are attended periodically by a single server. Avi-Itzhak, Maxwell and Miller [3] consider the problem studied by Takács (above), and, by intuitive arguments, find the expectations of queueing times as well as the first two moments of the busy periods.

As mentioned earlier, the problem of sequential queues has been studied by numerous investigators. Reich [15] proves that the lifetimes (service time included) of a customer in the two stages of a queueing system, with two exponential service stations in tandem, with infinite queue capacities and Poisson arrivals, are mutually independent. In 1963, Reich [16] generalizes this result for a sequence of m single-channel facilities. In 1964, Burke [4] shows that the waiting times (excluding the service times) in the stages of this system are stochastically dependent and in 1968 [5] he generalizes Reich's result proving the independence of lifetimes in the stages of a sequential queueing system, where the second service time is not necessarily exponential. When intermediate waiting-room capacities in a sequential queueing system are finite

the waiting times are quite dependent, since the size of queue depends on whether there is space available downstream. Avi-Itshak and Yadin [2] considered the particular case of two service stations and no intermediate queue, with Poisson arrival and general service distributions. They derived the transforms of the steady-state distributions of the queueing times and the number of customers in the system. Neuts [14] examines the case where the queue capacity in the first stage is infinite and finite in the second stage. The service time distributions in the first and second stages are, respectively, of the general type and exponential. His results include the busy periods, blocked time in the first stage, and the equilibrium conditions of the system. The effect of blocking in the utilization of exponential service stations in series is analyzed by Hunt [11] in one of the pioneering investigations of queues in series.

3. Summary of Results and Organization

This thesis is concerned with the description of stochastic aspects of the queueing system above, under each of the indicated policies. We are interested in the state probablities of the system, the waiting times until service begins in each stage, the busy period of the server, and the busy periods of each stage.

We assume steady-state conditions. For most of the cases the sufficient condition for existence of the steady state is $1 - \rho_1 - \rho_2 > 0$, where $\rho_k = \lambda E[S_1]$ is the traffic intensity in the stage k, k = 1,2. We define:

$\pi(k,i,j)$ = steady-state probability of having i customers in the first stage and j customers in the second stage, just after a departure from stage k.

$Q_{k\ell}$ = expected number of customers in stage k just after departures from stage ℓ.

W_k = time that a customer waits until he commences service in stage k (W_2 includes S_1 and W_1).

B^S = length of the busy period of the server. Time elapsed since the server started attending a customer who, at his arrival, found the system empty, until the next time the system becomes empty again.

B^I = interval of time during which the server is continuously present at the first stage.

B^{II} = interval of time during which the server is continuously present at the second stage.

In the case of random length of switching times the definitions of B^S, B^I and B^{II} are slightly different.

In the following table we indicate the number of the page where each result can be found. In the table, R and Z designate, respectively, random and zero switching time.

Except where indicated in Table I, the distributions of B^S, B^I and B^{II} are found explicitly. For the waiting times, W_k, we only obtained their Laplace-Stieltjes transform and the mean value.

Chapter V contains some numerical calculations and concluding remarks.

Table I

Results

Chapter	Policy		$\pi(k,i,j)$	$\bar{Q}_{k\ell}$	W_k	B^S	B^I	B^{II}
II	LNB	Z	32	26	27	29	30	30
	LNB	R	45	43	45	47**	47	48
III	SSP	Z	51	52	52	53	53	53
	SSP	R	51	52	52	53	53	53
	FSP	Z	56	56	57	57	57	58
	FSP	R	61	62	64	***	65	66
IV	SSS ($Q_2 = N$)		69*	79	80	***	83	82
	W_NFS		89	90	90	91	92	93
	SFS ($Q_1 \geq N$)		97	98	99	99	99	100

* Only characterized by the generating function.

** Conjecture.

*** Unsolved.

Each chapter is divided into sections which may be subdivided into subsections. The sections within a chapter are numbered by n(n = 1,2,...) and the subsections are numbered by n·k(n = 1,2,..., k = 1,2,...). The equations are numbered within sections, for example, Equation (3.11) means the 11th equation of section 3 of the present chapter. References to other chapters are made by stating explicity the number of the chapter.

4. Terminology and Methods

In addition to the symbols already defined the following designations will be helpful:

1-customer: A customer in the system who has not completed service in the first stage.

2-customer: A customer in the system who has completed service in the second stage but not in the first stage.

T_n : Time at which the nth departure occurs (departure from either stage).

E_n : Type of the customer departing at T_n.

$Q_{k,n}$: Number of k-customers in the system just after T_n.

$$P_{h,n}[(\ell,a,b) \to (k,i,j)] = P[E_{n+h} = k, Q_{1,n+h} = i, Q_{2,n+h} = j \mid E_n = \ell, Q_{1,m} = a, Q_{2,n} = b] .$$

From the assumptions on the arrival and service processes, the stochastic process $\{E_n, Q_{1,n}, Q_{2,n}\}$ is a homogeneous Markov chain, that is, the probabilities above are independent of n for all (ℓ,a,b) and (k,i,j). When h = 1 we drop the subscript h in the notation above.

The infinite system of linear equations

$$\pi(i,i,j) = \sum_{\ell=1}^{2} \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} \pi(\ell,a,b)\, P[(\ell,a,b) \to (k,i,j)]$$

$$(k = 1,2 \quad ; \quad i,j = 0,1,2,\ldots)$$

with the restriction

$$\sum_{k=1}^{2} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \pi(k,i,j) = 1 \ ,$$

will be referred to as the stationary conditions (in the text the restriction above is always implicit), and the set $\{\pi(k,i,j)\}$ as the stationary distribution of the chain.

In order to make sure that our policies are comprehensive as described, we will simply confine attention to the stationary distribution, i.e., the steady-state situation.

The analysis of each policy starts by writing down the stationary conditions and then obtaining the generating functions

$$U_k(x,y) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} x^i y^j \pi(k,i,j) \qquad (k = 1,2)$$

which are used for finding the expected values

$$\bar{Q}_{k\ell} = \frac{1}{U_\ell(1,1)} \frac{\partial}{\partial x_k} U_\ell(x_1,x_2)\Big|_{x_1=x_2=1} \qquad (k,\ell = 1,2)$$

$$\bar{Q}_{\cdot\ell} = \bar{Q}_{1\ell} + \bar{Q}_{2\ell} \qquad (\ell = 1,2) \ ,$$

$$\bar{Q}_{k\cdot} = U_1(1,1)\bar{Q}_{k1} + U_2(1,1)\bar{Q}_{k2} \qquad (k = 1,2)$$

and

$$\bar{Q} = \bar{Q}_{1\cdot} + \bar{Q}_{2\cdot} \quad .$$

The waiting time distributions, $W_k(t)$, are characterized by their Laplace-Stieltjes transforms, $\tilde{W}_k(s)$, in terms of $U_k(x,y)$, from which we can obtain the expected values $\bar{W}_k$.

The distributions of the busy periods, $B^S(t)$, $B^I(t)$ and $B^{II}(t)$, are found in terms of $\pi(k,i,j)$ and of the distribution of the busy period of an M/G/1 queueing system with appropriate service time distribution.

In general, given a random variable V, we use the symbols $V(t)$, $\bar{V}$ (or $E[V]$) and $\tilde{V}(s)$ for the distribution of V, the expected value of V, and the Laplace-Stieltjes transform of $V(t)$, respectively.

We also define

$$p_i = \int_0^\infty \frac{(\lambda t)^i e^{-\lambda t}}{i!} \, dS_1(t)$$

and

$$q_i = \int_0^\infty \frac{(\lambda t)^i e^{-\lambda t}}{i!} \, dS_2(t)$$

as the probabilities of having i arrivals during a service time of a 1-customer and 2-customer, respectively. The corresponding generating functions are

$$P(x) = \sum_{i=0}^{\infty} p_i x^i = \tilde{S}_1[\lambda(1 - x)]$$

and

$$Q(y) = \sum_{i=0}^{\infty} q_i y^i = \tilde{S}_2[\lambda(1 - x)] \quad ,$$

which converge for $|x| \leq 1$ and $|y| \leq 1$.

Additional symbols are defined in the text as they become necessary.

CHAPTER II

LEAVING NO BACKLOG

1. Introduction

In this chapter we study the LNB policy (Leaving No Backlog). Under this operating rule the server remains in one stage until it is empty (no backlog), then he switches to the other stage. We consider separately the cases of zero switching times and random switching times.

We use transform methods in order to find the expected values of pertinent random variables. We also present a direct approach for finding the state probabilities, based on properties of the busy period of an M/G/1 system.

2. Stationary Conditions (Instantaneous Switching)

In this case the server becomes immediately available in the second stage after he completes the service of the last customer present in the first stage, and vice-versa. We have

$$(2.1)\quad (a)\quad \pi(1,i,1) = \sum_{a=1}^{i+1} p_{i-a+1}\, \pi(2,a,0) + p_i\, \pi(2,0,0) \qquad (i \geq 0)\ ,$$

$$(b)\quad \pi(1,i,j) = \sum_{a=1}^{i+1} p_{i-a+1}\, \pi(1,a,j-1) \qquad (i \geq 0,\ j \geq 2)\ ,$$

$$(c)\quad \pi(2,i,j) = \sum_{a=0}^{i} q_{i-a}\, \pi(2,a,j+1) + q_i\, \pi(1,0,j+1) \qquad (i \geq 0,\ j \geq 0)\ .$$

Multiplying both sides of these equations by $x^i y^i$ and adding separately

$$(2.2) \qquad U_1(x,y) = \sum_{i=0}^{\infty} \sum_{a=1}^{i+1} p_{i-a+1}\, \pi(2,a,0) x^i y + \sum_{i=0}^{\infty} p_i\, \pi(2,0,0) x^i y$$

$$+ \sum_{j=2}^{\infty} \sum_{i=0}^{\infty} \sum_{a=1}^{i+1} p_{i-a+1}\, \pi(1,a,j-1) x^i y^j$$

and

$$(2.3) \qquad U_2(x,y) = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} q_i\, \pi(1,0,j+1) x^i y^j$$

$$+ \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \sum_{a=0}^{i} q_{i-a}\, \pi(2,a,j+1) x^i y^j \quad ,$$

which, in view of the property

$$(2.4) \qquad \sum_{i=0}^{\infty} \sum_{a=1}^{i+1} f_{a,i} = \sum_{i=0}^{\infty} \sum_{a=0}^{\infty} f_{a+1,i+a} \quad ,$$

and the fact that $\pi(1,i,0) = 0$, $(i \geq 0)$, give

$$(2.5) \qquad U_1(x,y) = \frac{(x-1)\pi(2,0,0) + U_2(x,0) - U_1(0,y)}{x - yP(x)}\, yP(x)$$

and

$$(2.6) \qquad U_2(x,y) = \frac{U_1(0,y) - U_2(x,0)}{y - Q(x)}\, Q(x) \quad .$$

In order to characterize completely the distribution $\{\pi(k,i,j)\}$ by the transforms $U_k(x,y)$, we need to determine $U_1(0,y)$, $U_2(x,0)$ and $\pi(2,0,0)$.

3. Determination of $U_1(0,y)$, $U_2(x,0)$ and $\pi(2,0,0)$

Since $|Q(x)| \leq 1$ for $|x| \leq 1$, it is clear that the denominator of the right-hand side of Equation 2.6, for each x in the domain $|x| \leq 1$, has only one root in the domain $|y| \leq 1$, namely $y = Q(x)$. The numerator of Equation 2.6 must vanish when $y = Q(x)$ because $|U_2(x,y)| \leq 1$ if $|x| \leq 1$ and $|y| \leq 1$. Thus, letting $Q = Q(x)$, we get

$$(3.1) \qquad U_2(x,0) = U_1(0,Q) \quad .$$

If $|y| < 1$, the inequality $|x| > |yP(x)|$ holds on the boundary of the domain $|x| \leq 1$. Therefore, by Rouche's theorem, the denominator of the right-hand side of Equation 2.5 has just one root in the domain $|x| \leq 1$. If $|y| = 1$ we prove by the Taylor expansion of $P(x)$ about $x = 1$, that the inequality $|x| > |yP(x)| = |P(x)|$ holds on the boundary of the domain $|x| \leq 1 + \epsilon$, for small ϵ, under the restriction $\rho_1 = (d/dx)P(x)\big|_{x=1} < 1$, which we shall assume. We have

$$P(1+\epsilon) = 1 + \rho_1\epsilon + \sum_{n=2}^{\infty} \frac{1}{n!} P^{(n)}(1)\epsilon^n$$

$$< 1 + \rho_1\epsilon + (1 - \rho_1)\epsilon = 1 + \epsilon \quad ,$$

for some $\epsilon > 0$. Hence, by Rouche's theorem, the equation $x = P(x)$ has the same number of roots in the domain $|x| \leq 1 + \epsilon$ as the equation $x = 0$, namely one root. Since $x = 1$ is a root of $x = P(x)$ we conclude that $x = yP(x)$ has just one root in the domain $|x| \leq 1$, given $|y| \leq 1$. Denote this root by $\delta(y)$.

When $x = \delta(y)$, that is $x = yP(x)$, the numerator of the right-hand side of Equation 2.5 must be zero, since $|U_1(x,y)| \leq 1$ in the domain $|x| \leq 1$ and $|y| \leq 1$. Thus, letting $\delta = \delta(y)$, we get

$$U_1(0,y) = U_2(\delta,0) + (\delta - 1)\pi(2,0,0) \quad . \tag{3.2}$$

Next we present an iterative scheme for determining $U_1(x,y)$ and $U_2(x,y)$. A similar scheme is used by Takacs [20] in the case of two queues in parallel attended by a single server.

Let $x_0 = x$ and $y_0 = y$ with $|x| \leq 1$ and $|y| \leq 1$. Define x_i and y_i, $i \geq 0$, by

$$x_{i+1} = \delta(y_i) \tag{3.3}$$

and

$$y_{i+1} = Q(x_i) \quad . \tag{3.4}$$

Substituting Equations 3.3 and 3.4 into Equations 3.1 and 3.2 we get

$$U_1(0,y_i) = U_2(x_{i+1},0) + \pi(2,0,0)(x_{i+1} - 1) \tag{3.5}$$

and

$$(3.6) \qquad U_2(x_i,0) = U_1(0,y_{i+1}) \quad .$$

Adding Equations 3.5 for i = 0,2,...,2N-2, to Equations 3.6 for i = 1,3,...,2N-1, we obtain

$$(3.7) \qquad U_1(0,y) = \pi(2,0,0) \sum_{n=1}^{N} (x_{2n-1} - 1) + U_1(0,y_{2N}) \quad .$$

Similarly, let i = 1,3,...,2N-1 in Equation 3.5 and i = 0,2,...,2N-2 in Equation 3.6. Adding up these equations we get

$$(3.8) \qquad U_2(x,0) = \pi(2,0,0) \sum_{n=1}^{N} (x_{2n} - 1) + U_2(x_{2N},0) \quad .$$

We shall now show that x_{2N} and y_{2N} tend to one as N tends to infinity, and that the series

$$\sum_{n=1}^{N} (x_{2n-1} - 1) \qquad \text{and} \qquad \sum_{n=1}^{N} (x_{2n} - 1)$$

are convergent. We will first study the behavior of the functions

$$y_{(1)}(x) = \frac{x}{P(x)} \qquad \text{and} \qquad y_{(2)}(x) = Q(x)$$

which have been used in definitions (3.3) and (3.4). We have the following properties:

$$y_{(1)}(0) = 0 \; ; \; y_{(1)}(1) = 1 \; ; \; y'_{(1)}(0) = \frac{1}{p_o} \; ;$$

$$y'_{(1)}(1) = 1 - \rho_1 \; ; \; y_{(2)}(0) = q_o \; ; \; y_{(2)}(1) = 1 \; ;$$

$$y'_{(2)}(0) = q_1 \; ; \; y'_{(2)}(1) = \rho_2 \; .$$

Also notice that $y_{(1)}(x) = y_{(2)}(x)$ implies

$$x - P(x)Q(x) = 0 \; ,$$

which, by Rouche's theorem and the condition $1 - \rho_1 - \rho_2 > 0$, has just one root in the domain $|x| \leq 1$ (see arguments preceding Equation 3.2). Since $x = 1$ is a root, we conclude that the graphs of $y_{(1)}(x)$ and $y_{(2)}(x)$ have no intersection in the domain $|x| \leq 1$, except at $x = 1$.

From the preceding properties and the fact that $y_1(x)$ is convex and $y_{(2)}(x)$ is concave (since $P(x)$ is convex), we can sketch the graphs of these functions as in Figure 1. It is clear from the graphs above that the sequences $\{x_n\}$ and $\{y_n\}$ converge to one.

In order to prove that the series

$$\sum_{n=1}^{N} (1 - x_{2n-1})$$

is convergent we obtain

$$y^*(x) = (1 - \rho_1)x + \rho_1 \quad \text{and} \quad y^{**}(x) = \rho_2 x + (1 - \rho_2) \; ,$$

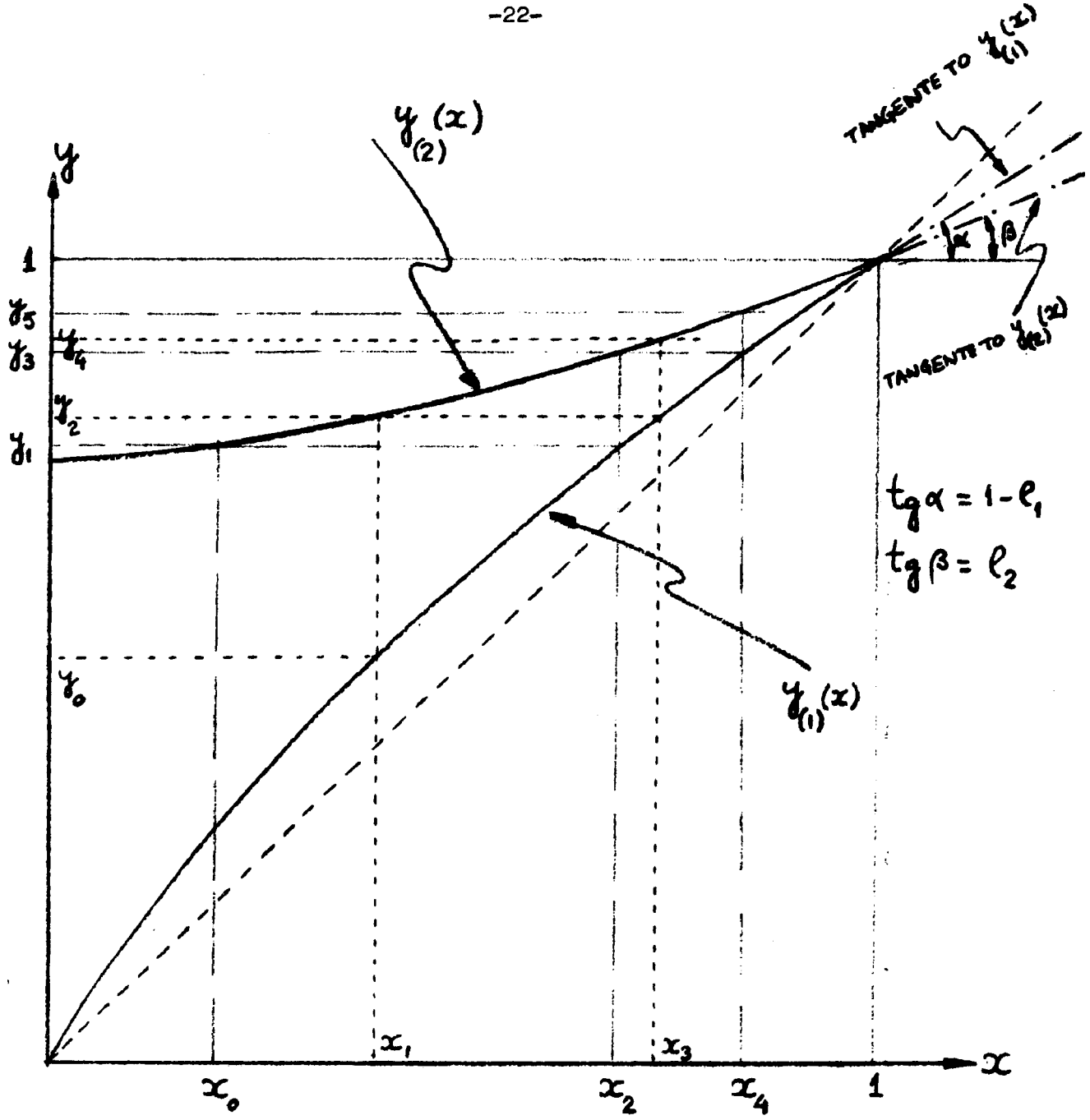


Figure I. Illustration of the Iteration Scheme for Finding $U_1(0,y)$ and $U_2(x,0)$.

which are the equations of the tangents to $y_{(1)}(x)$ and $y_{(2)}(x)$, respectively, at $x = 1$. Define

$$\bar{x}_{i+1} = \frac{1}{1 - \rho_1} \bar{y}_i - \frac{\rho_1}{1 - \rho_1}$$

and

$$\bar{y}_{i+1} = \rho_2 \bar{x}_i + (1 - \rho_2)$$

for $i = 0,1,\ldots,$ with $\bar{x}_0 = x_0$ and $\bar{y}_0 = y_0$. By the convexity of $y_{(1)}(x)$ and the concavity of $y_{(2)}(x)$ we have $y_{(1)}(x) \geq y^*(x)$ and $y_{(2)}(x) \leq y^{**}(x)$. This implies $1 - x_0 = 1 - \bar{x}_0$, $1 - x_2 \leq 1 - \bar{x}_2, \ldots$. Therefore it suffices to prove that

$$\sum_{n=1}^{N} (1 - \bar{x}_{2n-1})$$

converges. This can be easily seen from the definitions of $\bar{x}_i$ and $\bar{y}_i$, from which we get

$$1 - \bar{x}_{2n-1} = \left(\frac{\rho_2}{1 - \rho_1} \right)^{n-1} (1 - y_0) \quad .$$

This proves the convergence of $\sum_{n=1}^{N} (1 - \bar{x}_{2n-1})$, since $1 - \rho_1 - \rho_2 > 0$. In a similar way we can prove the convergence of $\sum_{n=1}^{N} (1 - \bar{x}_{2n})$.

We may now write

$$A_1(y) = \sum_{n=1}^{\infty} (1 - x_{2n-1}) \tag{3.9}$$

and

$$(3.10) \qquad A_2(x) = \sum_{n=1}^{\infty} (1 - x_{2n})$$

which are positive and finite for $|x| \leq 1$ and $|y| \leq 1$.

Hence, Equations 3.7 and 3.8 become

$$(3.11) \qquad U_1(0,y) = -\pi(2,0,0)A_1(y) + U_1(0,1)$$

and

$$(3.12) \qquad U_2(x,0) = -\pi(2,0,0)A_2(x) + U_2(1,0) \quad .$$

Setting $x_o = y_o = 0$ in Equations 3.3 and 3.4 we get $x_{2n} = x_{2n+1}$, $n = 0,1,\ldots,$ which implies $A_1(0) = 1 + A_2(0)$. Noting that $U_1(0,0) = 0$ and $U_2(0,0) = \pi(2,0,0)$ we obtain

$$(3.13) \qquad U_1(0,1) = U_2(1,0) = A_1(0)\pi(2,0,0)$$

which yields

$$(3.14) \qquad U_1(0,y) = [A_1(0) - A_1(y)]\pi(2,0,0)$$

and

$$(3.15) \qquad U_2(x,0) = [A_1(0) - A_2(x)]\pi(2,0,0). \ *$$

---

* See remark following Equation 6.6 for the interpretation of $A_1(0)$.

The value of $\pi(2,0,0)$ is found from the condition $U_1(1,1) + U_2(1,1) = 1$, where $U_1(1,1)$ and $U_2(1,1)$ are obtained from Equations 2.5 and 2.6 by applying L'Hopital's rule. We get

$$(3.16) \qquad \pi(2,0,0) = \frac{1}{2}(1 - \rho_1 - \rho_2) \quad .$$

We see that the condition for nonsaturation is $1 - \rho_1 - \rho_2 > 0$. The factor $1/2$ in Equation 3.16 is due to the fact that the same customer is considered twice in the embedded chain.

At this point we have completed the characterization of the steady-state joint distribution of the number of customers in each stage of the system and the type of departing customer, $\pi(k,i,j)$, which is done by Equations 2.5 and 2.6 where the values of $U_1(0,y)$, $U_2(x,0)$ and $\pi(2,0,0)$ are given by Equations 3.14 through 3.16. In Section 7 we shall present a preferable method for finding $\pi(k,i,j)$. We shall also give an example to show that the approach of this section is very efficient for calculating $\bar{B}^I$ and $\bar{B}^{II}$.

4. Expected Number of Customers

In this section we shall calculate the expected values $\bar{Q}_{k\ell}$, $\bar{Q}_{k\cdot}$, $\bar{Q}_{\cdot\ell}$ and $\bar{Q}$, which are defined in Chapter I.

The partial derivatives of $U_k(x,y)$ are obtained from Equations 2.5 and 2.6. The limits of these derivatives, as x and y tend to one, are indeterminate; however L'Hopital's rule can be applied in order to resolve the indeterminacy. The first and second derivatives of $U_2(x,0)$ and $U_1(0,y)$, which are necessary in the calculations, can be

obtained by using Equations 3.1 and 3.2. The derivatives of $\delta(y)$ are found, by implicit differentiation, from $\delta(y) = yP[\delta(y)]$.

Let

$$(4.1) \qquad Q_o = \frac{\lambda^2 E[(S_1 + S_2)^2]}{2(1 - \rho_1 - \rho_2)} \ .$$

This is the expression for the average number of customers waiting in an M/G/1 queueing system whose service time distribution is the convolution $S_1 * S_2(t)$ (Cooper [6], page 164). We get the following results:

$$(4.2) \qquad \bar{Q}_{11} = (\rho_1 + Q_o) \frac{1 - \rho_1}{1 - \rho_1 + \rho_2} \ ,$$

$$(4.3) \qquad \bar{Q}_{12} = (1 + \rho_2 + Q_o) \frac{\rho_2}{1 - \rho_1 + \rho_2} \ ,$$

$$(4.4) \qquad \bar{Q}_{21} = (1 + \rho_2 + Q_o) \frac{1}{1 - \rho_1 + \rho_2} \ ,$$

$$(4.5) \qquad \bar{Q}_{22} = (\rho_1 + Q_o) \frac{1}{1 - \rho_1 + \rho_2} \ ,$$

$$(4.6) \qquad \bar{Q}_{1.} = \frac{1}{2} (\rho_1 + \rho_2 + Q_o) \ ,$$

$$(4.7) \qquad \bar{Q}_{2.} = \frac{1}{2} (1 + \rho_1 + \rho_2 + 2Q_o) \frac{1}{1 - \rho_1 + \rho_2} \ ,$$

$$(4.8) \qquad \bar{Q}_{.1} = [1 + \rho_2 + \rho_1(1 - \rho_1) + (2 - \rho_1)Q_o] \frac{1}{1 - \rho_1 + \rho_2} \ ,$$

$$(4.9) \qquad \bar{Q}_{.2} = [\rho_1 + (1 + \rho_2)(\rho_2 + Q_o)] \frac{1}{1 - \rho_1 + \rho_2} \ ,$$

$$(4.10) \qquad \bar{Q} = [1 + (\rho_1 + \rho_2)(2 - \rho_1 + \rho_2) + (3 - \rho_1 + \rho_2)Q_o] \cdot \frac{1}{2(1 - \rho_1 + \rho_2)} .$$

In the following table we compare (when S_1 and S_2 are exponential) some values of $\bar{Q}_{.2}$ obtained by Nelson [13] through simulation (32,000 arrivals), with the values calculated by Equation 4.9, for various values of ρ_1 and ρ_2.

Table II

Values of $Q_{.2}$ Obtained by Simulation and by Equation 4.9

	ρ_1=0.1	ρ_1=0.2	ρ_1=0.3	ρ_1=0.45	ρ_1=0.6	ρ_1=0.7	ρ_1=0.8
	ρ_2=0.8	ρ_2=0.7	ρ_2=0.6	ρ_2=0.45	ρ_2=0.3	ρ_2=0.2	ρ_2=0.1
Simulation (Nelson [13])	8.51	8.50	8.86	10.14	13.13	18.04	19.96
Equation 4.9	8.54	8.38	8.73	10.90	13.11	17.96	29.80

5. Waiting Times

Let W_k denote the time a customer waits until he commences service in stage k (W_2 includes W_1 and S_1). We can find the Laplace-Stieltzes transform of W_k by an argument that is also available for the M/G/1 queueing system. A customer departing from the first stage will leave behind him i customers who arrived during his waiting time W_1 and his service time S_1. Thus

$$\frac{1}{U_1(1,1)} \sum_{j=o}^{\infty} \pi(1,i,j) = \int_0^{\infty} \frac{(\lambda t)^i e^{-\lambda t}}{i!} \, dW_1 * S_1(t)$$

which gives

$$\frac{U_1(x,1)}{U_1(1,1)} = \tilde{W}_1[\lambda(1-x)] \cdot \tilde{S}_1[\lambda(1-x)]$$

or

$$(5.1) \qquad \tilde{W}_1(s) = \frac{2U_1(1 - S/\lambda, 1)}{\tilde{S}_1(s)} \quad ,$$

where $RE(s) > 0$ and $|\lambda - s| \leq \lambda$.

Similarly, a customer departing from the second stage will leave behind him (in both stages) i customers who arrived during his waiting time W_2 (which includes his waiting and service time in the first stage) and his service time S_2. Thus

$$\frac{1}{U_2(1,1)} \sum_{a=0}^{i} \pi(2,a,i-a) = \int_0^\infty \frac{(\lambda t)^i e^{-\lambda t}}{i!} \, dW_2 * S_2(t) \quad ,$$

which implies

$$(5.2) \qquad \tilde{W}_2(s) = \frac{2U_2(1 - s/\lambda, 1)}{\tilde{S}_1(s)}$$

where $RE(s) > 0$ and $|\lambda - s| \leq \lambda$.

The expected values of W_1 and W_2 are obtained by differentiation of Equations 5.1 and 5.2. We get

$$(5.3) \qquad \overline{W}_1 = -\frac{d}{ds}\tilde{W}_1(s)\Big|_{s=0} = \frac{\overline{Q}_{11}}{\lambda} - \overline{S}_1 \text{ or } \overline{Q}_{11} = \lambda(\overline{W}_1 + \overline{S}_1) \quad ,$$

where $\overline{Q}_{11}$ is given by Equation 4.2.

$$(5.4) \qquad \bar{W}_2 = -\frac{d}{ds} W_2(s)\Big|_{s=0} = \frac{2}{\lambda} \frac{d}{dx} U_2(x,x)\Big|_{x=1} - \bar{S}_2$$

$$= \frac{1}{\lambda} \bar{Q}_{\cdot 2} - \bar{S}_2 \quad \text{or} \quad \bar{Q}_{\cdot 2} = \lambda(\bar{W}_2 + \bar{S}_2) \ ,$$

where $\bar{Q}_{\cdot 2}$ is given by Equation 4.9.

6. Busy Periods

Clearly there are two different categories of busy periods. First, the busy period of the server, B^S, is the time elapsed since the arrival of a customer who finds the system empty, until the departure of a customer who leaves an empty system behind him. Notice that at the completion of a busy period of the server, regardless of the order in which the customers are serviced, all of the customers have completed their two-stage service, which has the distribution $S_1 * S_2(t)$. Therefore, this busy period has the distribution of the busy period of an M/G/1 queueing system whose service time distribution is $S_1 * S_2(t)$. Its expected value is $\bar{B}^S = (\bar{S}_1 + \bar{S}_2)/(1 - \rho_1 - \rho_2)$ (Cooper [6], page 176).

We define B^I and B^{II} as the time the server is continuously busy in the first and second stages, respectively.

The busy period B^I starts immediately when the second stage is emptied and $i \geq 1$ customers are present in the first stage. In this case the distribution of B^I will be $B_i(t)$, the distribution of the busy period initiated by i customers in an M/G/1 queueing system with service distribution $S_1(t)$. When i = 0 the busy period starts at the time of the next arrival and it has distribution $B_1(t)$. Thus

$$(6.1) \qquad B^{I}(t) = \frac{1}{U_2(1,0)} \left[\pi(2,0,0)B_1(t) + \sum_{i=0}^{\infty} \pi(2,i,0)B_i(t) \right] .$$

The Laplace-Stieltzes transform of $B^{I}(t)$ is

$$(6.2) \qquad \tilde{B}^{I}(s) = \frac{1}{U_2(1,0)} \left\{ \pi(2,0,0)\tilde{B}(s) + \sum_{i=1}^{\infty} \pi(2,i,0)[\tilde{B}(s)]^{i} \right\}$$

$$= \frac{1}{U_2(1,0)} \left\{ \frac{1}{2} (1 - \rho_1 - \rho_2)[\tilde{B}(s) - 1] + U_2[\tilde{B}(s),0] \right\} ,$$

where $\tilde{B}(s) = \tilde{B}_1(s)$ and $\tilde{B}_i(s) = [\tilde{B}(s)]^{i}$, since B_i is the sum of i independent initial busy periods (that is a busy period initiated by just one customer). The value of $U_2[\tilde{B}(s),0]$ is given by Equation 3.15.

The distribution of the time the server is continuously servicing 2-customers is

$$(6.3) \qquad B^{II}(t) = \frac{1}{U_1(0,1)} \sum_{j=1}^{\infty} \pi(1,0,j)\, S_2^{*j}(t) ,$$

where $S^{*j}(t)$ is the j-fold convolution of the service distribution $S_2(t)$ with itself. The Laplace-Stieltzes transform of $B^{II}(t)$ is

$$(6.4) \qquad \tilde{B}^{II}(s) = \frac{1}{U_1(0,1)} \sum_{j=1}^{\infty} \pi(1,0,j)[\tilde{S}_2(s)]^{j}$$

$$= \frac{1}{U_1(0,1)} U_1[0,\tilde{S}_2(s)] .$$

The expected values of B^I and B^{II} are

$$(6.5) \qquad \bar{B}^I = -\frac{d}{ds}\tilde{B}^I(s)\Big|_{s=0} = \frac{1}{U_2(1,0)}\left[\frac{1}{2}(1-\rho_1-\rho_2)\frac{\bar{S}_1}{1-\rho_1} + \frac{\bar{S}_1}{1-\rho_1}\frac{\bar{S}_1}{2}\right] = \frac{1}{U_2(1,0)}\frac{\bar{S}_1}{2}$$

$$= \frac{1}{A_1(0)}\frac{\bar{S}_1}{1-\rho_1-\rho_2}$$

and

$$(6.6) \qquad \bar{B}^{II} = -\frac{d}{ds}\tilde{B}^{II}(s)\Big|_{s=0} = \frac{1}{U_1(0,1)}\frac{\bar{S}_2}{2}$$

$$= \frac{1}{A_1(0)}\frac{\bar{S}_2}{1-\rho_1-\rho_2},$$

where $U_1(0,1)$ and $U_2(1,0)$ were obtained from Equation 3.13 and $A_1(0)$ is given by Equation 3.9.

Remark

We notice that

$$\bar{B}^I + \bar{B}^{II} = \frac{1}{A_1(0)}\frac{\bar{S}_1+\bar{S}_2}{1-\rho_1-\rho_2} = \frac{1}{A_1(0)}\bar{B}^S .$$

This implies that $A_1(0) = \bar{B}^S/(\bar{B}^I + \bar{B}^{II})$ is the average number of cycles in the busy period of the server (a cycle starts when the server commences service in the first stage and terminates when all the customers, who were served during the busy period in the first stage, have completed service in the second stage).

7. Alternative Approach for the Calculation of the State Probabilities

In this section we present a direct method for obtaining the state probabilities $\pi(k,i,j)$. It is based on known properties of the busy period of an M/G/1 system.

Define

$b_m(i)$ = probability that i departures occur during a busy period of the first stage that starts with m customers ($i \geq m \geq 1$).

$a_i(n)$ = probability that n arrivals occur during the service time of i 2-customers ($i \geq 1$, $n \geq 0$).

$c_k(n)$ = probability that, starting with an empty system, the termination of the kth cycle occurs when there are n customers in the first stage ($k \geq 1$, $n \geq 0$).

We have

$$(7.1) \qquad c_1(n) = \sum_{i=1}^{\infty} b_1(i)a_i(n) \qquad (n \geq 0)$$

$$c_k(n) = \sum_{m=1}^{\infty} \sum_{i=m}^{\infty} c_{k-1}(m)b_m(i)a_i(n) \qquad (k > 1,\ n \geq 0) \ ,$$

where

$$(7.2) \qquad b_m(i) = \frac{m}{i} \int_0^{\infty} \frac{(\lambda t)^{i-m}}{(i-m)!} e^{-\lambda t} \, dS_1^{*i}(t) \qquad (i \geq m \geq 1)$$

and

$$(7.3) \qquad a_i(n) = \int_0^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} \, dS_2^{*i}(t) \ .$$

The state probabilities $\pi(2,n,0)$ can be written as

$$(7.4) \qquad \pi(2,n,0) = \pi(2,0,0) \sum_{k=1}^{\infty} c_k(n) \qquad (n > 0) \quad ,$$

where $\pi(2,0,0) = 1/2(1 - \rho_1 - \rho_2)$. The probabilities $\pi(1,i,j)$, and $\pi(2,i,j)$ for $j \neq 0$, can be calculated by Equations 2.1.

The average number of cycles during the busy period of the server is

$$(7.5) \qquad \bar{c} = \sum_{k=1}^{\infty} k c_k(0) \quad .$$

8. Comparison of the Methods of Sections 3 and 7

With the objective of comparing the iterative approach of Section 3 with the alternative approach of Section 7, we have made a few experimental computations for the case of exponential service time.

Evidently, if we need to calculate each value of the state probabilities, the method of the previous section is preferable. However, when we need only $U_1(0,1) = U_2(1,0) = \sum_{i=0}^{\infty} \pi(2,i,0)$, as in the cases of the expected values $\bar{B}^{I}$, $\bar{B}^{II}$ and $\bar{c}$ (see Section 6), the iterative method of Section 3 has some advantages. We shall illustrate this point by the following computations.

Let $\lambda = 1$ and, consequently, $\rho_1 = \bar{S}_1$ and $\rho_2 = \bar{S}_2$. We assume that S_1 and S_2 are exponentially distributed and calculate the values of $\bar{c}$ (average number of cycles during the busy period of the server) by each of the methods mentioned above, for various values of ρ_1 and ρ_2.

a. Computations Using the Iteration Scheme of Section 3

We first obtain

$$P(x) = \tilde{S}_1[\lambda(1 - x)] = \frac{1}{1 + (1 - x)\rho_1}$$

and

$$Q(x) = \tilde{S}_2[\lambda(1 - x)] = \frac{1}{1 + (1 - x)\rho_2} .$$

Using Equations 3.3 and 3.4 we get

$$y_i = \frac{x_{i+1}}{P(x_{i+1})} = (1 + \rho_1)x_{i+1} - \rho_1 x_{i+1}^2$$

and

$$(8.1) \qquad y_{i+1} = \frac{1}{1 + (1 - x_i)\rho_2} .$$

Thus

$$(8.2) \qquad x_{i+1} = \delta(y_i) = \frac{1 + \rho_1 - \sqrt{(1 + \rho_1)^2 - 4\rho_1 y_i}}{2\rho_1}$$

(since $|x_{i+1}| \leq 1$ for $|y_i| \leq 1$). In order to calculate the quantity $\bar{c} = A_1(0)$ (see remark following Equation 6.6) we construct Table III from Equations 8.1 and 8.2, where $y_0 = 0$. This table corresponds to $\rho_1 = \rho_2 = 0.4$.

Table III

Illustration of the Iterative Scheme of Section 3 for $\rho_1 = \rho_2 = 0.4$

i	y_i	x_i
0	0	--
1	--	0
2	0.71428	--
3	--	0.62005
4	0.86807	--
5	--	0.80537
6	0.92777	--
7	--	0.88798
.	.	.
.	.	.
.	.	.
30	0.99959	
31	--	0.99932

The value of $\bar{c} = A_1(0)$ is $\sum_{i=0}^{\infty} (1 - x^{o}_{2i-1}) \cong 1.8787$. For other values of ρ_1 and ρ_2 we get the values of Table IV, where $i_{(0.0001)}$ is the number of terms in the series of $A_1(0)$ for having an accuracy of 0.0001.

Table IV

Values of $\bar{c}$ for Different Traffic Intensities

$\rho = \rho_1 + \rho_2$	ρ_1	ρ_2	$i_{(0.0001)}$	$\bar{c}$
0.900	0.800	0.100	9	1.390
"	0.450	0.450	26	2.295
"	0.100	0.800	41	2.895
0.500	0.400	0.100	5	1.165
"	0.250	0.250	7	1.358
"	0.100	0.400	9	1.516
0.100	0.090	0.010	3	1.010
"	0.050	0.050	4	1.052
"	0.010	0.090	4	1.091

The values of Table IV were calculated by an IBM 360/67 computer (MTS system of The University of Michigan). The CPU time for calculating all the cases was about six seconds (including compilation of the Fortran program).

b. Computations Using the Method of Section 7

We first find

$$b_m(i) = \frac{m}{i}\binom{2i - m - 1}{i - m}\left[\frac{\rho_1}{1 + \rho_1}\right]^{i-m}\left[\frac{1}{1 + \rho_1}\right]^{i} \quad (i \geq m \geq 1)$$

from Equation 7.2, and

$$a_i(n) = \binom{n + i - 1}{n}\left[\frac{\rho_2}{1 + \rho_2}\right]^{n}\left[\frac{1}{1 + \rho_2}\right]^{i}$$

from Equation 7.3. Next, we use Equations 7.1 for calculating $c_k(n)$. Equation 7.5 is then used for obtaining $\bar{c}$. Table V gives the values of $\bar{c}$ (for the same values of the traffic intensities of Table IV), where N is the number of terms of the summations of Equations 7.1 through 7.5, and t is the approximated CPU time (in seconds) for compiling the Fortran program and calculating the values of $\bar{c}$, for all traffic intensities.

Table V

Values of $\bar{c}$ Calculated by the Direct Approach, for Different Traffic Intensities and Different Truncations of the Infinite Summations

			$\bar{c}$			
$\rho = \rho_1 + \rho_2$	ρ_1	ρ_2	N = 5 t = 7	N = 10 t = 27	N = 15 t = 122	N = 20 t = 200
0.900	0.800	0.100	0.904	1.033	1.032	1.119
"	0.450	0.450	1.261	1.640	1.850	1.970
"	0.100	0.800	1.427	1.991	2.309	2.500
0.500	0.400	0.100	1.092	1.150	1.161	1.163
"	0.250	0.250	1.295	1.354	1.358	--
"	0.100	0.400	1.426	1.513	1.516	--
0.100	0.090	0.010	1.010	1.010	1.010	--
"	0.050	0.050	1.052	1.052	1.052	--
"	0.010	0.090	1.091	1.091	1.091	--

Comparing the values of $\bar{c}$ in Tables IV and V we see that this second approach is quite inefficient for high values of the total traffic intensity $\rho = \rho_1 + \rho_2$. In the first method the convergence is quite satisfactory for all values of ρ.

In using the direct approach for more general cases, we expected to have difficulties with the integrations of Equations 7.2 and 7.3, in addition to the large amount of calculations involved in the recursive relationship (7.1). In using the iterative scheme of Section 3 we, in general, expect no difficulties in obtaining $P(x)$ and $Q(x)$, and the solution of $x_{i+1} = \delta(y_i)$ can be found by simple search techniques, because of concavity of $y_i = x_{i+1}/P(x_{i+1})$.

Hence we conclude that, in general, the iteration scheme of Section 3 is a better method for calculating the expected values of B^I, B^{II} and $\bar{c}$.

9. The Case of Random Duration of Switching

In this section we shall review the previous sections under the assumption that there is a random time $T_{k\ell}$ with general distribution $T_{k\ell}(\cdot)$ for switching the server from stage k to stage ℓ ($k,\ell = 1,2$ and $k \neq \ell$). Define

$$(9.1) \qquad t_i^{k\ell} = \int_0^\infty \frac{(\lambda t)^i e^{-\lambda t}}{i!} dT_{k\ell}(t) \quad ,$$

the probability of having i arrivals during the switching duration $T_{k\ell}$. Also define

$$(9.2) \qquad R_{k\ell}(x) = \sum_{i=o}^{\infty} x^i t_i^{k\ell} = \tilde{T}_{k\ell}[\lambda(x-1)] \ .$$

In this case the stationary conditions are

$$(9.3)\ (a)\quad \pi(1,i,1) = \sum_{a=1}^{i+1} \sum_{b=o}^{i-a+1} t_b^{21} p_{i-a-b+1}\, \pi(2,a,0)$$

$$+ \sum_{b=1}^{i+1} t_b^{21} p_{i-b+1}\, \pi(2,0,0) + t_o^{21}\, p_i \pi(2,0,0) \qquad (i \geq 0)$$

$$(b)\quad \pi(1,i,j) = \sum_{a=1}^{i+1} p_{i-a+1} \pi(1,a,j-1) \qquad (i \geq 0,\ j \geq 2)$$

$$(c)\quad \pi(2,i,j) = \sum_{b=o}^{i} t_b^{12}\, q_{i-b}\, \pi(1,0,j+1)$$

$$+ \sum_{a=o}^{i} q_{i-a}\ \pi(2,a,j+1) \qquad (i \geq 0,\ j \geq 0) \ ,$$

from which we get

$$(9.4) \qquad U_1(x,y) = \frac{(x-1)t_o^{21}\, \pi(2,0,0) + R_{21}(x)U_2(x,0) - U_1(0,y)}{x - yP(x)}\, yP(x)$$

and

$$(9.5) \qquad U_2(x,y) = \frac{R_{12}(x)U_1(0,y) - U_2(x,0)}{y - Q(x)}\, Q(x) \ .$$

Notice that in the case of instantaneous switching we have $t_o^{21} = R_{12}(x) = R_{21}(x) = 1$, in which case Equations 9.4 and 9.5 reduce to Equations 2.5 and 2.6.

By the arguments used to derive Equations 3.1 and 3.2 we get

$$U_2(x,0) = R_{12}(x)\, U_1(0,Q) \tag{9.6}$$

and

$$U_1(0,y) = R_{21}(\delta)U_2(\delta,0) + (\delta - 1)\, t_o^{21}\, \pi(2,0,0) \quad , \tag{9.7}$$

where $\delta = \delta(y)$ is the inverse function of $y = x/P(x)$.

Using Equations 3.3 and 3.4 we get

$$U_2(x_i,0) = R_{12}(x_i)U_1(0,y_{i+1}) \tag{9.8}$$

and

$$U_1(0,y_i) = R_{21}(x_{i+1})U_2(x_{i+1},0) + (x_{i+1} - 1)t_o^{21}\, \pi(2,0,0) \quad . \tag{9.9}$$

Set $i = 0,2,4,\ldots$ in Equation 9.9 and multiply each resulting equation respectively, by 1, $R_{12}(x_1)R_{21}(x_1)$, $R_{12}(x_1)R_{21}(x_1)R_{12}(x_3)R_{21}(x_3)$, ..., set $i = 1,3,5, \ldots,$ in Equation 9.8 and multiply each resulting equation by $R_{21}(x_1)$, $R_{21}(x_1)R_{12}(x_1)R_{21}(x_2)$, $R_{21}(x_1)R_{12}(x_1)R_{21}(x_3)$ $R_{12}(x_2)R_{21}(x_5)$, ...; adding all together we get

$$U_1(0,y) = -\pi(2,0,0)A_1(y) + B_1(y)U_1(0,1) \quad , \tag{9.10}$$

where

$$B_1(y) = \prod_{n=0}^{\infty} R_{12}(x_{2n+1})R_{21}(x_{2n+1}) \tag{9.11}$$

and

$$A_1(y) = t_o^{21} \sum_{n=0}^{\infty} (1 - x_{2n-1}) \prod_{i=0}^{n-1} R_{12}(x_{2i+1})R_{21}(x_{2i+1}) \quad . \tag{9.12}$$

Similarly, set i = 0,2,4,..., in Equation 9.8 and multiply each resulting equation, respectively, by 1, $R_{12}(x_o)R_{21}(x_2)$, $R_{12}(x_o)R_{12}(x_2)$ $R_{12}(x_2)R_{21}(x_4)$,...; set i = 1,3,5,..., in Equation 9.9 and multiply each resulting equation, respectively, by $R_{12}(x_o)$, $R_{12}(x_o)R_{21}(x_2)R_{12}(x_2)$, $R_{12}(x_o)R_{21}(x_2)R_{12}(x_2)R_{21}(x_4)R_{12}(x_4)$, ...; adding all together we get

$$U_2(x,0) = -\pi(2,0,0)A_2(x) + B_2(x)\, U_2(1,0) \quad , \tag{9.13}$$

where

$$B_2(x) = \prod_{n=0}^{\infty} R_{12}(x_{2n})R_{21}(x_{2n+2}) \tag{9.14}$$

and

$$A_2(x) = t_o^{21}R_{12}(x) \sum_{n=1}^{\infty} (1 - x_{2n}) \prod_{i=1}^{n-1} R_{12}(x_{2i})R_{21}(x_{2i}) \quad , \tag{9.15}$$

where

$$\prod_{i=a}^{b} p_i = 1$$

when $a > b$.

The values of $U_1(0,1)$ and $U_2(1,0)$ can be found from Equations 9.10 and 9.13 by setting $x = y = 0$, and noticing that $U_1(0,0) = 0$ and $U_2(0,0) = \pi(2,0,0)$. We have

$$(9.16) \qquad U_1(0,1) = \frac{A_1(0)}{B_1(0)}\pi(2,0,0) \quad \text{and} \quad U_2(1,0) = \frac{1 + A_2(0)}{B_2(0)}\pi(2,0,0) .$$

But when $x = y = 0$ ($x_o = y_o = 0$) we have $x_n = x_{2n+1}$, $n = 0,1,2,\ldots$, which implies $A_1(0) = t_o^{21}[1 + A_2(0)]$ and $B_1(0) = t_o^{21}B_2(0)$. Hence

$$(9.17) \qquad U_1(0,1) = U_2(1,0) = \frac{A_1(0)}{B_1(0)}\pi(2,0,0)$$

which substituted in Equations 9.10 and 9.13 yields

$$(9.18) \qquad U_1(0,y) = \left[\frac{A_1(0)}{B_1(0)} B_1(y) - A_1(y)\right]\pi(2,0,0)$$

and

$$(9.19) \qquad U_2(x,0) = \left[\frac{A_1(0)}{B_1(0)} B_2(x) - A_2(x)\right]\pi(2,0,0) .$$

For the case of instantaneous switching we have $t_o^{21} = R_{12}(x) = R_{21}(x) = 1$ and, consequently, $B_1(y) = B_2(x) = 1$, $A_1(y) = \sum_{n=o}^{\infty} (1 - x_{2n+1})$ and $A_2(x) = \sum_{n=1}^{\infty} (1 - x_{2n})$. Therefore, Equations 9.18 and 9.19 reduce to Equations 3.14 and 3.15, respectively.

In order to determine $\pi(2,0,0)$ we use the fact that $U(1,1) + U(1,1) = 1$ and apply L'Hopital's rule to find

$$\lim_{\substack{x \to 1 \\ y \to 1}} U_k(x,y) \qquad k = 1,2 \ ;$$

we get

$$(9.20) \qquad \pi(2,0,0) = \frac{1 - \rho_1 - \rho_2}{2\left[t_o^{21} + \lambda(\overline{T}_{12} + \overline{T}_{21}) \frac{A_1(0)}{B_1(0)}\right]} \ .$$

Notice that, independently of the switching times, $1 - \rho_1 - \rho_2 > 0$ is the condition for the system to be unsaturated.

9.1 Expected Number of Customers. Defining $\theta_{k\ell} = \lambda E[T_{k\ell}]$ and using Equations 9.6 and 8.7 we obtain

$$(9.21) \qquad U_1''(0,1) = \frac{d^2}{dy^2} U_1(0,y)\Big|_{y=1} = \frac{1}{(1 - \rho_1 - \rho_2)(1 - \rho_1 + \rho_2)}$$

$$\left[\lambda^2 E[(T_{12} + T_{21})^2]\, U_1(0,1) + (\theta_{12} + \theta_{21})\left(\rho_2 + 2\rho_1 U_1(0,1) + \frac{1}{1 - \rho_1} \lambda^2 E[S_1^2] U_1(0,1)\right) + \rho_1\rho_2 + 2\rho_1 t_o^{21}\, \pi(2,0,0) + \frac{1}{1 - \rho_1}\left(\frac{\rho_2}{2} + t_o^{21}\, \pi(2,0,0) + \frac{1}{2}\lambda^2 E[S_2^2]\right)\right]$$

and

$$(9.22)\qquad U_2''(1,0) = \frac{d^2}{dx^2} U_2(x,0)\Big|_{x=1} = \lambda^2 E[T_{12}^2]U_1(0,1) + \theta_{12}\rho_2$$

$$+ \frac{1}{2}\lambda^2 E[S_2^2] + \frac{\rho_2}{(1-\rho_1-\rho_2)(1-\rho_1+\rho_2)}\Bigg[\lambda^2 E[(T_{12}+T_{21})^2]U_1(0,1)$$

$$+ (\theta_{12}+\theta_{21})\left(\rho_2 + 2\rho_1 U_1(0,1) + \frac{1}{1-\rho_1}\lambda^2 E[S_1^2]U_1(0,1)\right)$$

$$+ \rho_1\rho_2 + 2\rho_1 t_o^{21}\,\pi(2,0,0) + \frac{1}{1-\rho_1}\left(\frac{1}{2}\rho_2 + t_o^{21}\,\pi(2,0,0)\right)$$

$$\cdot\ \lambda^2 E[S_1^2] + \frac{1}{2}\lambda^2 E[S_2^2]\Bigg] ,$$

where

$$U_1(0,1) = \frac{A_1(0)}{B_1(0)}\,\pi(2,0,0)$$

and $\pi(2,0,0)$ is given by Equation 9.20.

From Equations 9.4 and 9.5 we get

$$(9.23)\qquad \frac{\partial}{\partial x} U_1(x,y)\Big|_{\substack{x=1\\y=1}} = \frac{(1-\rho_1)A'' + \lambda^2 E[S_1^2]A'}{2(1-\rho_1)^2} ,$$

$$(9.24)\qquad \frac{\partial}{\partial y} U_1(x,y)\Big|_{\substack{x=1\\y=1}} = \frac{1}{2} + \frac{1}{2} U_1''(0,1) ,$$

$$(9.25)\qquad \frac{\partial}{\partial y} U_2(x,y)\Big|_{\substack{x=1\\y=1}} = -\frac{\rho_2 B'' - \frac{1}{2}\lambda^2 E[S_2^2]\rho_2}{2\rho_2^2} ,$$

$$(9.26) \qquad \frac{\partial}{\partial x} U_2(x,y)\Big|_{\substack{x=1\\y=1}} = \frac{1}{2} U_1''(0,1) \quad ,$$

where

$$A' = (\theta_{12} + \theta_{21})U_1(0,1) + \frac{1}{2}\rho_2 + t_o^{21}\,\pi(2,0,0) \quad ,$$

$$A'' = [\lambda^2 E[T_{21}^2] + 2\rho_1(\theta_{12} + \theta_{21}) + 2\theta_{12}\theta_{21}]\,U_1(0,1) + \theta_{21}\rho_2$$

$$+ 2\rho_1 t_o^{21}\,\pi(2,0,0) + \rho_1\rho_2 + U_2''(1,0) \quad ,$$

$$B'' = \lambda^2 E[T_{12}^2]U_1(0,1) - \rho_2^2 - U_2''(1,0) \quad .$$

The expected values $\bar{Q}_{k\ell}$, $\bar{Q}_{k.}$, $\bar{Q}_{.\ell}$ and $\bar{Q}$ can be obtained by using Equations 9.23 through 9.26.

9.2 Waiting Times. Equations 5.1 through 5.4 still hold for this case of random length of switching. The quantities $\bar{Q}_{11}$ and $\bar{Q}_{.2}$ of Equations 5.3 and 5.4 are obtained by the procedure indicated above.

9.3 Alternate Approach for the Calculation of the State Probabilities. Let $b_i(n)$, $a_i(n)$ and $c_k(n)$ be as defined in Section 7, and let $t_i^{k\ell}$ be defined as in Equation 9.1. We consider the server busy during switching times. We have

$$(9.27) \qquad c_1(n) = \sum_{i=0}^{\infty} \sum_{j=i}^{\infty} \sum_{\ell=0}^{\infty} t_i^{21} b_i(j) t_\ell^{12} a_j(n-\ell) \qquad (n \geq 0)$$

and

$$c_k(n) = \sum_{m=0}^{\infty} \sum_{i=0}^{\infty} \sum_{j=i+m}^{\infty} \sum_{\ell=0}^{\infty} c_{k-1}(m) t_i^{21} b_{i+m}(j) t_\ell^{12} a_j(n-\ell)$$

$$(k > 1,\ m \geq 0) \quad .$$

These equations represent the fact that new arrivals may occur during the switching times T_{12} and T_{21}. As in Equation 7.4, the state probabilities $\pi(2,n,0)$ are given by

$$(9.28) \qquad \pi(2,n,0) = \pi(2,0,0) \sum_{k=1}^{\infty} c_k(n) \quad ,$$

where $\pi(2,0,0)$ is given by Equation 9.20.

The quantity

$$(9.29) \qquad \bar{c} = \sum_{k=1}^{\infty} k c_k(0)$$

is the average number of cycles during a busy period of the server

The remaining state probabilities are obtained from Equations 9.3.

9.4 Busy Periods. A busy period of the server, B^S, commences when the server is idle in the first stage and a new customer arrives, it terminates when, at the end of a switching interval T_{21}, the system is empty. B^I is the time the server is continuously present in the first stage including the switching interval T_{12}, B^{II} is defined similarly.

Let B_o^S be the busy period of the server when the switching is instantaneous. In the case of random duration of switching, in each cycle of the server, additional arrivals have to be served and each of them will add a random time to B_o^S which has the distribution of B_o^S. We conjecture that on the average we have

$$\overline{B}^S = \overline{B}_o^S + \lambda(\overline{T}_{12} + \overline{T}_{21})\overline{C}\,\overline{B}_o^S + (\overline{T}_{12} + \overline{T}_{21})\overline{C} \quad , \tag{9.30}$$

where $\overline{B}_o^S = (\overline{S}_1 + \overline{S}_2)/(1 - \rho_1 - \rho_2)$ and $\overline{C}$ is given by Equation 9.29. However, we do not utilize this conjecture.

The distribution of B^I and B^{II} are given by

$$B^I(t) = \frac{1}{U_2(1,0)}\left[\sum_{i=1}^{\infty}\sum_{j=o}^{\infty}\pi(2,i,0)t_j^{21}B_{i+j}{}^*T_{12}(t) + \sum_{j=1}^{\infty}\pi(2,0,0)t_j^{21}B_j{}^*T_{12}(t) + \pi(2,0,0)t_o^{21}B_1{}^*T_{12}(t)\right] \tag{9.31}$$

and

$$(9.32) \qquad B^{II}(t) = \frac{1}{U_1(o,1)} \sum_{j=1}^{\infty} \pi(1,0,j)\, S_2^{*j} * T_{21}(t)$$

where $B_i * T_{12}(t)$ is the convolution of the busy period distribution of the first stage, initiated by i customers, with the switching time distribution $T_{12}(t)$. $S_2^{*j} * T_{21}(t)$ is the convolution of the distribution $T_{21}(t)$ with the j-fold convolution of the service time distribution $S_2(t)$ with itself.

The Laplace-Stieltjes transform of these distributions are

$$(9.33) \qquad \tilde{B}^{I}(s) = \frac{1}{U_2(1,0)} \Big[\tilde{T}_{12}(s) R_{21}[\tilde{B}(s)] \Big[U_2[\tilde{B}(s),0] - \pi(2,0,0) \Big] + \tilde{T}_{12}(s) \Big[R_{21}[\tilde{B}(s)] - t_o^{21} \Big] \pi(2,0,0) + \tilde{T}_{12}(s) t_o^{21} \tilde{B}(s) \pi(2,0,0) \Big]$$

and

$$(9.34 \qquad \tilde{B}^{II}(s) = \frac{1}{U_1(0,1)} \tilde{T}_{21}(s)\, U_1[0,\tilde{S}_2(s)]$$

where $U_2(x,0)$ and $U_1(0,y)$, $R_{21}(x)$, t_o^{21} and $\pi(2,0,0)$ are given by Equations 9.16, 9.2, 9.1 and 9.20, respectively.

The expected values of B^I and B^{II} are

$$(9.35) \qquad \bar{B}^{I} = -\frac{d}{ds} \tilde{B}^{I}(s)\Big|_{s=o}$$

$$= \bar{T}_{12} + \frac{\bar{S}_2}{1-\rho_1} \left[\lambda(\bar{T}_{12} + \bar{T}_{21}) + \frac{B_1(0)}{A_1(0)} \left\{ t_o^{21} + \frac{\rho_2}{2\pi(2,0,0)} \right\} \right]$$

and

$$(9.36) \qquad \bar{B}^{II} = -\frac{d}{ds}\tilde{B}^{II}(s)\Big|_{s=o}$$

$$= \bar{T}_{21} + \frac{B_1(0)}{A_1(0)}\,\frac{\bar{S}_2}{2\pi(2,0,0)} .$$

where $A_1(0)$, $B_1(0)$ and $\pi(2,0,0)$ are given by Equations 9.12, 9.11 and 9.20, respectively.

Note: Do not confuse $S_1(0)$ of Equation 9.11 with the distribution of the initial busy of the first stage at $t = 0$, which is zero.

Notice that when $\bar{T}_{21} = \bar{T}_{12} = 0$ we have $B_i(0) = t_o^{21} = 1$ and $\pi(2,0,0) = 1/2(1 - \rho_1 - \rho_2)$; therefore Equations 9.35 and 9.36 reduce to Equations 6.5 and 6.6, respectively.

CHAPTER III

SIMPLE NONPREEMPTIVE PRIORITY

1. Introduction

In this chapter we determine the operating characteristics of the SSP policy (second stage priority) and of the FSP policy (first stage priority). Under the SSP policy the analysis of the behavior of the system is quite simple since each phase of each customer's service is performed without interruption. This is the case of an M/G/1 queueing system. However, in order to be consistent with the analysis of the other policies, we shall choose the Markov chain embedded at departures from either stage. The switching durations can be assumed to be part of the service times, therefore we present only the case of instantaneous switching. This policy could also be called "follow the customer," since the server switches to the second stage when the customer is served in the first stage. By contrast, in the FSP policy the server switches to the first stage at the service completion of 2-customers if one or more 1-customers are present.

2. The SSP Policy

Since there is no queue in the second stage, the state of the system is characterized by the pair (k,i) where $k = 1,2$ indicates the stage from which the customer is departing, and $i = 0,1,\ldots$ is the number of customers left in the system by the departing customer. We denote by $\pi(k,i)$ the steady-state probabilities. We have

$$(2.1)\ (a) \qquad \pi(1,i) = \sum_{a=1}^{i+1} \pi(2,a)p_{i-a} + \pi(2,0)p_{i-1} \qquad (i \geq 1) \ ,$$

$$(b) \qquad \pi(2,i) = \sum_{a=1}^{i+1} \pi(1,a)q_{i-a+1} \qquad (i \geq 0) \ .$$

Notice that a customer departing from the first stage is considered to be in the system. The generating functions $U_k(x) = \sum_{i=0}^{\infty} \pi(i,i)x^i$, $k = 1,2$, are

$$(2.2) \qquad U_1(x) = P(x)[U_2(x) - \pi(2,0)] + xP(x)\pi(2,0)$$

and

$$(2.3) \qquad U_2(x) = \frac{1}{x} Q(x)U_1(x) \ ,$$

from which we get

$$(2.4) \qquad U_1(x) = \frac{x(x-1)\pi(2,0)P(x)}{x - P(x)Q(x)} \ .$$

The value of $\pi(2,0)$ is obtained from Equations 2.3, 2.4 and the condition $U_1(1) + U_2(1) = 1$. We get

$$(2.5) \qquad \pi(2,0) = \frac{1}{2}(1 - \rho_1 - \rho_2) \ .$$

The expected number of customers in the system at departure epochs, $\overline{Q}$, is given by

$$\bar{Q} = U_1(1)\bar{Q}_1 + U_2(1)\bar{Q}_2 \quad , \tag{2.6}$$

where

$$\bar{Q}_k = \frac{1}{U_k(1)} \frac{d}{dx} U_k(x)\Big|_{x=1}$$

is the expected number of customers in the system just after the completion of service in the kth stage. We have

$$\bar{Q}_1 = 1 + \rho_1 + Q_o \quad , \tag{2.7}$$

$$\bar{Q}_2 = \rho_1 + \rho_2 + Q_o \quad , \tag{2.8}$$

$$\bar{Q} = \frac{1}{2}(1 + 2\rho_1 + \rho_2) + Q_o \quad , \tag{2.9}$$

where

$$Q_o = \frac{\lambda^2 E[(S_1 + S_2)^2]}{2(1 - \rho_1 - \rho_2)} \quad .$$

Notice that Equation 2.8 is the Pollaczek-Khintchine formula for an M/G/1 queueing system where the service time distribution is the convolution of $S_1(t)$ with $S_2(t)$.

The time a customer waits until he commences service in the first stage, W_1, and the time he waits until he commences service in the second stage, W_2, have distributions whose Laplace-Stieltjes transform are given by

$$(2.10) \qquad \tilde{W}_1(s) = \frac{1}{U_1(1)} \frac{U_1\left(\frac{\lambda - s}{\lambda}\right)}{\frac{\lambda - s}{\lambda} \tilde{S}_1(s)}$$

and

$$(2.11) \qquad \tilde{W}_2(s) = \frac{1}{U_2(1)} \frac{U_2\left(\frac{\lambda - s}{\lambda}\right)}{\tilde{S}_2(s)} \quad ,$$

where $RE(s) > 0$ and $|\lambda - s| \leq \lambda$. The expected values are

$$(2.12) \qquad \bar{W}_1 = \frac{1}{\lambda} (\bar{Q}_1 - 1) - \bar{S}_1$$

and

$$(2.13) \qquad \bar{W}_2 = \frac{1}{\lambda} \bar{Q}_2 - \bar{S}_2 \quad .$$

The busy period of the server, B^S, has distribution

$$B^S(t) = \sum_{n=1}^{\infty} \frac{1}{n} \int_0^t \frac{(\lambda u)^{n-1}}{(n-1)!} e^{-\lambda u} \, dS_1^{*n} * S_2^{*n}(u) \quad ,$$

which is the distribution of the busy period of an M/G/1 system with service time distribution $S_1 * S_2(t)$. The busy period of the stages, B^I and B^{II}, have distributions $S_1(t)$ and $S_2(t)$, respectively.

3. The FSP Policy

Under this policy the server goes to the second stage only when he has finished attending all of the customers in the first stage. Once in the second stage, the server keeps attending 2-customers who are present as long as no 1-customers have arrived. If an arrival occurs during the service time of a 2-customer, the server completes serving that customer and then switches to the first stage. If he succeeds in finishing all the present 2-customers and no 1-customers are yet present, he goes to the first stage and waits for the next arrival.

We first study the instantaneous switching case. The stationary conditions are

$$(3.1)\quad (a)\quad \pi(1,i,1) = p_i\,\pi(2,0,0) + \sum_{a=1}^{i+1} p_{i-a+1}\,\pi(2,a,0) \qquad (i \geq 0)$$

$$(b)\quad \pi(1,i,j) = \sum_{a=1}^{i+1} p_{i-a+1}\,\pi(1,a,j-1) + \sum_{a=1}^{i+1} p_{i-a+1}\,\pi(2,a,j-1) \qquad (i \geq 0,\ j \geq 2)$$

$$(c)\quad \pi(2,i,j) = q_i\,\pi(1,0,j+1) + q_i\,\pi(2,0,j+1) \qquad (i \geq 0,\ j \geq 0)\ ,$$

from which we obtain

$$(3.2) \qquad U_1(x,y) = \frac{x\pi(2,0,0) + U_2(x,y) - U_1(0,y) - U_2(0,y)}{x - yP(x)}\, yP(x)$$

and

$$(3.3) \qquad U_2(x,y) = \frac{1}{y} Q(x)[U_1(0,y) + U_2(0,y) - \pi(2,0,0)] \quad .$$

From the arguments preceding Equation 3.2 of Chapter II, and Equation 3.2 we get

$$(3.4) \qquad U_1(0,y) + U_2(0,y) = \delta\pi(2,0,0) + U_2(\delta,y) \quad ,$$

where $\delta = \delta(y)$, as before, is the inverse function of $y = x/P(x)$. The value of $\pi(2,0,0)$ is obtained from Equation 3.2 by applying L'Hopital's rule whne x and y tend to one) and the condition $U_1(1,1) = U_2(1,1) = 1/2$, knowing that $\delta(1) = 1$ and $\delta'(1) = (1 - \rho_1)^{-1}$. We get

$$(3.5) \qquad \pi(2,0,0) = \frac{1}{2}(1 - \rho_1 - \rho_2) \quad .$$

Remarks

This result and the results (3.16) of Chapter II and (2.5) of the previous section are expected, since, under the corresponding policies, the server is active whenever a customer is present in the system. This does not happen when the system is controlled by the policies SSP($Q_2 \geq N$) and W_NFS (Chapter IV). Under a SSP($Q_2 \geq N$) policy the server is inactive when he finishes serving all 1-customers and there are less than N 2-customers; under a W_NFS policy he is inactive

when he finishes all 2-customers and there are less than N 1-customers.

From Equation 3.3 and 3.4, when $x = \delta = \delta(y)$, we get

$$(3.6) \qquad U_1(0,y) + U_2(0,y) = \frac{y\delta - Q(\delta)}{y - Q(\delta)} \pi(2,0,0)$$

which substituted into Equations 3.2 and 3.3 gives

$$(3.7) \qquad U_1(x,y) = \frac{[y - Q(\delta)]x - [y - Q(x)] + Q(\delta) - Q(x)}{[x - yP(x)][y - Q(\delta)]} yP(x)\pi(2,0,0)$$

and

$$(3.8) \qquad U_2(x,y) = \frac{(\delta - 1)Q(x)}{y - Q(\delta)} \frac{1 - \rho_1 - \rho_2}{2} .$$

This last equation, for $y = 0$, gives

$$(3.9) \qquad \pi(2,i,0) = \frac{1}{2}(1 - \rho_1 - \rho_2)\frac{q_i}{q_o} .$$

Using Equations 3.1 we can obtain the remaining state probabilities.

3.1 Expected Number of Customers and Waiting Times. The expected values of the number of customers in each stage at departure epochs are found by the same procedure of Section 4 of Chapter II: from Equations 3.7 and 3.8 we find the derivatives of $U_k(x,y)$, as x and y tend to unity, by applying L'Hopital's rule. The derivatives of $\delta = \delta(y)$ are calculated from $\delta(y) = yP[\delta(y)]$ by implicit differentiation. Letting

$$Q_o = \frac{\lambda^2 E[(S_1 + S_2)^2]}{2(1 - \rho_1 - \rho_2)}$$

we get

$$(3.10) \qquad \bar{Q}_{11} = (\rho_1 + Q_o) \frac{1 - \rho_1 - \rho_2}{1 - \rho_1} ,$$

$$(3.11) \qquad \bar{Q}_{21} = (1 + Q_o) \frac{1}{1 - \rho_1} ,$$

$$(3.12) \qquad \bar{Q}_{12} = \rho_2 ,$$

and

$$(3.13) \qquad \bar{Q}_{22} = (\rho_1 + Q_o) \frac{1}{1 - \rho_1} .$$

The quantities $\bar{Q}_{k.}$, $\bar{Q}_{.k}$ and $\bar{Q}$ can readily be obtained from the above results.

The waiting times W_1 and W_2 are as in Section 5 of Chapter II, where $U_k(x,y)$, $k = 1,2$, is given by Equations 3.7 and 3.8.

3.2 Busy Periods

With instantaneous switching the busy period of the server, B^S, has the distribution of the busy period of an M/G/1 system with service time distribution $S_1 * S_2(t)$. The reasoning for this is as in Section 6 of Chapter II. The distribution of B^I is

$$(3.14) \qquad B^I(t) = \left[\sum_{i=1}^{\infty} \sum_{j=o}^{\infty} \pi(2,i,j) B_i(t) + \pi(2,0,0) B(t) \right] \cdot \frac{1}{U_2(1,1) - U_2(0,1) + \pi(2,0,0)}$$

and its Laplace-Stieltjes transform is

$$(3.15) \qquad \tilde{B}^{I}(s) = \left[U_2[\tilde{B}(s),1] - U_2(0,1) + \pi(2,0,0)\tilde{B}(s) \right] \cdot \frac{2}{2 - \rho_1 - \rho_2 - q_o} ,$$

where $\tilde{B}(s)$ is the Laplace-Stieltjes transform of $B(t)$, the busy period of an M/B/1 system with service time distribution $S_1(t)$. $U_2(x,1)$ is given by Equation 3.8. The expected value of B^I is

$$(3.16) \qquad \bar{B}^{I} = \frac{1}{2 - \rho_1 - \rho_2 - q_o} \bar{S}_1 .$$

In order to find the distribution of B^{II}, we see that, when this busy period starts, the state of system is $(1,0,j)$ with probability $\pi(1,0,j)/U_1(0,1)$ (conditional to the start of B^{II}). The server will complete service of just one 2-customer with probability $(1 - q_o)$ and in a time with the distribution $S_2(t)$; of just two 2-customers with probability $(1 - q_o)q_o$ and distribution $S_2^{*2}(t)$; of just k 2-customers with probability $(1 - q_o)q_o^{k-1}$ and distribution $S_2^{*k}(t)$, $1 \leq k < j$, $j \geq 2$; of just j 2-customers with probability q_o^j and distribution $S_2^{*j}(t)$. Therefore we get

$$(3.17) \qquad B^{II}(t) = \frac{1}{U_1(0,1)} \left[\sum_{j=2}^{\infty} (1 - q_o)\pi(1,0,j) \sum_{k=1}^{j-1} q_o^{k-1} S_2^{*k}(t) + \sum_{j=1}^{\infty} \pi(1,0,j) q_o^{j-1} S_2^{*j}(t) \right] ,$$

whose Laplace-Stieltjes transform is

$$(3.18)\qquad \tilde{B}^{II}(s) = \frac{1}{1 - q_o \tilde{S}_2(s)} \left[(1 - q_o)\tilde{S}_2(s) + \frac{1}{U_1(0,1)} \cdot (1 - \tilde{S}_2(s)) U_1[0, q_o \tilde{S}_2(s)] \right]$$

and whose expected value is

$$(3.19)\qquad \bar{B}^{II} = \frac{1}{2 - \rho_1 - \rho_2 - q_o} \bar{S} \quad .$$

The average number of cycles during the busy period of the server is

$$(3.20)\qquad \bar{C} = \frac{\bar{B}^S}{\bar{B}^I + \bar{B}^{II}} = 1 + \frac{1 - q_o}{1 - \rho_1 - \rho_2} \quad .$$

3.3 The Case of Random Duration of Switching. In this subsection we shall use the notation of Section 8 of Chapter II.

The server, after finishing all 1-customers, goes to stage 2. The switching duration is a random variable T_{12} with distribution $T_{12}(t)$. During T_{12} some customers may arrive, but the server will attend to the next 2-customer before switching back to the first stage.

The stationary conditions are

$$(3.21)\quad (a)\quad \pi(1,i,1) = t_o^{21} p_i \pi(2,0,0) + \pi(2,0,0) \sum_{a=1}^{i+1} t_a^{21} p_{i-a+1} + \sum_{a=1}^{i+1} \sum_{b=o}^{i-a+1} t_b^{21} p_{i-a-b+1} \pi(2,a,0) \qquad (i \geq 0) \quad ,$$

$$(b) \qquad \pi(1,i,j) = \sum_{a=1}^{i+1} p_{i-a+1}\pi(1,a,j-1) + \sum_{a=1}^{i+1}\sum_{b=0}^{i-a+1} \cdot t_b^{21} p_{i-a-b+1}\pi(2,x,j-1) \qquad (i \geq 0,\ j \geq 2)$$

$$(c) \qquad \pi(2,i,j) = \sum_{a=0}^{i} t_a^{12} q_{i-a} \pi(1,0,j+1) + q_i\pi(2,0,j+1) \qquad (i \geq 0,\ j \geq 0)$$

which yield

$$(3.22) \qquad U_1(x,y) = \frac{1}{x - yP(x)}\Big[t_0^{21} xyP(x)\pi(2,0,0) + yP(x) \cdot [(R_{21}(x) - t_0^{21})\pi(2,0,0) + R_{21}(x) \cdot U_2(x,y) - U_1(0,y) - R_{21}(x)U_2(0,y)]\Big]$$

and

$$(3.23) \qquad U_2(x,y) = \frac{Q(x)}{y}[R_{12}(x)U_1(0,y) + U_2(0,y) - \pi(2,0,0)] \ .$$

Applying the arguments preceding Equation 3.2 of Chapter II to Equation 3.22 we obtain

$$(3.24) \qquad U_1(0,y) = -R_{21}(\delta)U_2(0,y) + [(\delta - 1)t_0^{21} + R_{21}(\delta)]\pi(2,0,0) + R_{21}(\delta)U_2(\delta,y) \ ,$$

$$(3.25)\qquad U_1(0,y) = \frac{R_{21}(\delta)q_o+(y-q_o)(\delta-1)t_o^{21}+(y-q_o)R_{21}(\delta)-R_{21}(\delta)Q(\delta)}{y(y-q_o)+yR_{21}(\delta)q_o t_o^{12}-R_{21}(\delta)Q(\delta)[(y-q_o)R_{12}(\delta)+q_o t_o^{12}]}$$

$$\cdot\, y\pi(2,0,0) \quad .$$

Equation 3.23, as $y \to 0$, implies

$$(3.26)\qquad U_2(x,0) = \alpha Q(x)R_{12}(x) + \beta Q(x) \quad ,$$

where

$$\alpha = \frac{d}{dy} U_1(0,y)\Big|_{y=o} \quad \text{and} \quad \beta = \frac{d}{dy} U_2(0,y)\Big|_{y=o} \quad .$$

This means that

$$(3.27)\qquad \pi(2,i,0) = \alpha \sum_{a=o}^{i} q_a t_{i-a}^{12} + \beta q_i \quad .$$

The quantities α and β are given by

$$(3.28)\qquad \alpha = \frac{p_o q_o (t_o^{21} + t_i^{21}) + p_o q_1 t_o^{21}}{q_o(1 - p_o q_o t_o^{21} t_o^{12})}\ \pi(2,0,0)$$

and

$$(3.29)\qquad \beta = \frac{1}{q_o}\ \pi(2,0,0) - t_o^{12}\ \alpha \quad .$$

The remaining state probabilities are given by Equations 3.21.

The quantity $\pi(2,0,0)$ is obtained from Equations 3.23, 3.25 and the fact that $U_2(1,1) = 1/2$ (because each customer corresponds to two departures). Letting $\theta_{k\ell} = \lambda E[T_{k\ell}] = \lambda \bar{T}_{k\ell}$ we have

$$(3.30) \qquad \pi(2,0,0) = \frac{1}{2} \frac{(1-q_o)(1-\rho_1-\rho_2-\theta_{12}-\theta_{21})+q_o t_o^{12}(1-\rho_1-\rho_2)}{(1-q_o)t_o^{21}+q_o t_o^{21} t_o^{12}}$$

from which we see that the condition for nonsaturation is

$$(3.31) \qquad \rho_1 + \rho_2 + (\theta_{12} + \theta_{21}) \cdot \left[1 + \frac{q_o t_o^{12}}{1 - q_o}\right]^{-1} < 1 \ .$$

We notice that $(\theta_{12} + \theta_{21})[1 + q_o t_o^{12}/1 - q_o]^{-1}$ represents the portion of the total utilization of the system corresponding to the switching times. The quantity $[1 + q_o t_o^{12}/1 - q_o]^{-1}$ is the "share" of each customer in that portion.

Using Equations 3.22, 3.23 and 3.25 we get

$$(3.32) \qquad \frac{\partial}{\partial x} U_2(x,y)\Big|_{\substack{x=1\\y=1}} = \left[\theta_{12} + \rho_2 \left[1 + \frac{q_o t_o^{12}}{1 - q_o}\right]\right] U_1 - \frac{\rho_2}{1 - q_o} \pi(2,0,0)$$

$$(3.33) \qquad \frac{\partial}{\partial y} U_2(x,y)\Big|_{\substack{x=1\\y=1}} = \left[1 + \frac{q_o t_o^{12}}{1 - q_o}\right] (U_1' - U_1) + \frac{1}{1 - q_o} \pi(2,0,0)$$

$$(3.34) \qquad \frac{\partial}{\partial y} U_1(x,y)\Big|_{\substack{x=1\\y=1}} = \left[1 + \frac{q_o t_o^{12}}{1 - q_o}\right] U_1'$$

and

$$(3.35)\qquad \frac{\partial}{\partial x}\, U_1(x,y)\Big|_{\substack{x=1\\y=1}} = \frac{\rho_1}{1-\rho_1}\, A' + \frac{1}{2(1-\rho_1)^2}$$

$$\cdot\, [(1-\rho_1)A'' + \lambda^2 E[S_1^2]A'] \quad ,$$

where

$$A' = \frac{1}{2}\,\theta_{21} + \left[t_o^{21} + \frac{1}{1-q_o}\,(\theta_{21} - \rho_2)\right]\pi(2,0,0)$$

$$+\left[\theta_{12} + \rho_2 + (\rho_2 - \theta_{21})\,\frac{q_o t_o^{12}}{1-q_o}\right] U_1$$

$$A'' = \frac{1}{2}\,\lambda^2 E[T_{21}^2] + \left[2\lambda^2 E[T_{21}^2] - \lambda^2 E[(T_{21} + T_{12})^2]\right]\pi(2,0,0)$$

$$+\left[\lambda^2 E[(T_{21} + T_{12} + S_2)^2] - \left[2\lambda^2 E[T_{21}^2]\right.\right.$$

$$\left.\left. - \lambda^2 E[(T_{21} + S_2)^2]\right]\frac{q_o t_o^{12}}{1-q_o} - \lambda^2 E[S_2^2]\right] U_1$$

and

$$U_1 = U_1(0,1) = \frac{B'}{C'}\,\pi(2,0,0)$$

$$U_1' = \frac{d}{dy}\, U_1(0,y)\Big|_{y=1} = \left[\frac{B''C' - B'C''}{2C'^2} + \frac{B'}{C'}\right]\pi(2,0,0)$$

with

$$B' = \frac{1}{1-\rho_1}\,[1 - \rho_1 - \rho_2 + (1-q_o)t_o^{21}]$$

$$B'' = \theta_{21} + \frac{1}{1-\rho_1}(2t_o^{12} + \theta_{21}) + \frac{1 - q_o + \theta_{21}}{(1-\rho_1)^2}$$

$$\cdot \left[2\rho_1 + \frac{1}{1-\rho_1}\lambda^2 E[S_1^2]\right] + \frac{1}{(1-\rho_1)^2}\lambda^2 E[T_{21}^2]$$

$$- \lambda^2 E[(T_{21} + S_2)^2]$$

$$C' = \frac{1-q_o}{1-\rho_1}(1 - \rho_1 - \rho_2 - \theta_{21} - \theta_{12}) + \frac{q_o t_o^{12}}{1-\rho_1}(1 - \rho_1 - \rho_2)$$

$$C'' = 2 + \frac{1}{1-\rho_1} q_o t_o^{12} - \frac{1}{(1-\rho_1)^2}\left[2 + \frac{\lambda^2 E[S_1^2]}{1-\rho_1}\right]$$

$$\cdot [q_o t_o^{12}\rho_2 - (1-q_o)(\rho_2 + \theta_{21} + \theta_{12})]$$

$$- \frac{\lambda^2}{(1-\rho_1)^2}\left[(1-q_o)E[(T_{21} + T_{12} + S_2)^2]\right.$$

$$\left. + q_o t_o^{12} E[(T_{21} + S_2)^2] - q_o t_o^{12} E[T_{21}^2]\right]$$

$$- \frac{2}{1-\rho_1}(\rho_2 + \theta_{21} + \theta_{12}) \quad .$$

The expected values of the number of customers at departure epochs are easily found by using Equations 3.32 through 3.35.

The waiting times W_1 and W_2 are as in Section 5 of Chapter II, where $U_k(x,y)$, $k = 1,2$, is given by Equations 3.22 and 3.23.

The busy period of the first stage, B^I, starts at the completion of the switching time T_{21}, when the state of the system is $(2,i,j)$, with $i \geq 1$ and $j \geq 0$, or $i = j = 0$. The probability of the set formed by all of these states is

$$(3.36) \qquad P_o = U_2(1,1) - U_2(0,1) + \pi(2,0,0)$$

$$= \frac{1}{2} + \frac{1 - \rho_1 - \rho_2 - \theta_{12} - \theta_{21} - q_o(1 - q_o)(t_o^{12})^2}{(1 - q_o)t_o^{21} + q_o t_o^{21} t_o^{12}} \quad .$$

During the switching time T_{21} new customers may arrive, say k customers. The server will then be busy in the first stage for a time T_{12} plus a busy period initiated by $i + k$ customers, B_{i+k}, where i is the number of 1-customers present at the beginning of T_{21}. By definition T_{12} is part of B^I. We have

$$(3.37) \qquad B^I(t) = \frac{1}{P_o}\left[\sum_{i=}^{\infty} \sum_{j=o}^{\infty} \sum_{k=o}^{\infty} \pi(2,i,j) t_k^{21} B_{i+k} * T_{12}(t) \right.$$

$$\left. + \sum_{k=1}^{\infty} \pi(2,0,0) t_k^{21} B_k * T_{12}(t) + \pi(2,0,0) t_o^{21} B_1 * T_{12}(t) \right] \quad ,$$

where P_o is given above. The Laplace-Stieltjes transform of $B^I(t)$ is

$$(3.38)\qquad \tilde{B}^{I}(s) = \frac{\tilde{T}_{12}(s)}{P_o}\left[R_{21}[\tilde{B}(s)]\left[U_2[\tilde{B}(s),1] - U_2(0,1)\right] + \pi(2,0,0)\left[R_{21}[\tilde{B}(s)] - t_o^{21} + \tilde{B}(s)\right]\right]$$

and the expected value is

$$(3.39)\qquad \overline{B}^{I} = \frac{\overline{S}_1}{(1-\rho_1)P_o}\left[P_o\theta_{21} + \frac{\partial}{\partial x} U_2(x,y)\Big|_{\substack{x=1\\y=1}} + \pi(2,0,0)\right] + \overline{T}_{12}\left[1 + \frac{1 - t_o^{21}}{P_o}\,\pi(2,0,0)\right]$$

where $\partial/\partial x\ U_2(x,y)\big|_{\substack{x=1\\y=1}}$, $\pi(2,0,0)$ and P_o are given by Equations 3.32, 3.30 and 3.36, respectively.

The distribution of the busy period of the second stage (which includes T_{21}), $B^{II}(t)$, is

$$(3.40)\qquad B^{II}(t) = \sum_{j=1}^{\infty} (1 - t_o^{12})\,\pi(1,0,j)T_{21}{}^{*}S_2(t) + \sum_{j=2}^{\infty}\sum_{k=1}^{j-1} t_o^{12}\,\pi(1,0,j)(1-q_o)q_o^{k-1}\,S_2^{*k}{}^{*}T_{21}(t) + \sum_{j=1}^{\infty} t_o^{12}\,\pi(1,0,j)q_o^{j-1}S^{*j}{}^{*}T_{21}(t)\ ,$$

where the first term of the right-hand side corresponds to the case when new customers arrive during the switching time T_{12}, that is, $B^{II} = S_2 + T_{21}$ (T_{21} is considered part of the busy period of the second stage); the second term corresponds to the cases when the server succeeds in completing service of k 2-customers, that is, $B^{II} = S_2^{(1)} + \ldots + S_2^{(k)} + T_{21}$ where $S_2^{(n)}$ $(n = 1,\ldots,k)$ have distribution $S_2(t)$; the third term corresponds to the cases when all of the present 2-customers are serviced.

The Laplace-Stieltjes transform of $B^{II}(t)$ is

$$(3.41) \qquad \tilde{B}^{II}(s) = (1 - t_o^{12})\tilde{T}_{21}(s)\tilde{S}_2(t) + \frac{t_o^{12}}{1 - q_o\tilde{S}_2(s)} \cdot \left[(1 - q_o)\tilde{S}_2(s) + \frac{1}{U_1(0,1)}[1 - \tilde{S}_2(s)] \cdot U_1[0, q_o\tilde{S}_2(s)]\right]$$

and the expected value is

$$(3.42) \qquad \bar{B}^{II}(s) = (1 - t_o^{12})(\bar{T}_{21} + \bar{S}_2) + \frac{t_o^{12}\bar{S}_2}{1 - q_o}\left(2 - q_o - \frac{C'}{B'}\,\frac{1}{t_o^{12}}\right),$$

where C' and B' are given following Equation 3.30.

CHAPTER IV.

RELUCTANT SWITCHING

1. Introduction

In the two previous chapters we studied three simple policies which are quite likely to be considered for the control of our queueing system. In general we have to balance the behavior of the system, at least qualitatively, in terms of switching costs, waiting costs and service costs. We mentioned that the SSP-policy is good when the waiting cost in the second stage is high, and the FSP-policy if preferable when the opposite occurs; the LNB-policy is better when the switching cost is the critical factor.

It is desirable to have some intermediate alternatives, for example, how to avoid high queues in the second stage without incurring high switching costs. To this end we shall now consider more general control rules which can still be qualified as "simple" in the sense of feasibility of implementation. Under these policies we shall see that the server is comparatively reluctant to switching. We will examine three policies, namely:

SSS($Q_2 = N$). The server continues available in the first stage until the number of 2-customers reaches the value N. When this happens the server goes to the second stage and serves all the present 2-customers before switching back to the first stage. The label of this policy intends the rule "Switch to the Second Stage when $Q_2 = N$." Notice that the SSP policy is the SSS ($Q_2 = 1$) policy.

SFS ($Q_1 \geq N$). After finishing all of the 1-customers (end of the busy period of the first stage), the server goes to the second stage. He stays there until all of the present 2-customers are served, unless the number of 1-customers reaches the value N. In this case the server completes the service of the present 2-customer and then goes back to the first stage. The label of this policy intends to rule "Switch to the First Stage when $Q_1 \geq N$." The FSP policy is the SFS ($Q_1 \geq 1$) policy.

W_NFS. After finishing all present 2-customers the server switches immediately to the first stage only if the number of 1-customers is equal to or greater than N, otherwise he is removed from the system (not available in either stage) and comes back to the first stage only when the number of 1-customers reaches the value N. The server stays in the first stage until all of the 1-customers complete service (end of the busy period of the first stage), then goes to the second stage. The LNB policy is the W_1FS policy except for the fact that the server is never removed under the LNB policy. (It would be better to define W_0FS as the LNB policy.)

We shall discuss the effect of these policies on the queue sizes, waiting times and busy periods as we did in Chapters II and III, but only the case of instantaneous switching will be considered.

2. The SSS ($Q_2 = N$) Policy

The stationary conditions are

$$(2.1)\ (a) \qquad \pi(1,i,1) = \sum_{a=1}^{i+1} \pi(2,a,0)p_{i-a+1} + \pi(2,0,0)p_i \qquad (i \geq 0) \ ,$$

(Eq. 2.1 cont.)

$$(b)\qquad \pi(1,i,j) = \sum_{a=1}^{i+1} \pi(1,a,j-1)p_{i-a+1} + \pi(1,0,j-1)p_i$$

$$(i \geq 0,\ 2 \leq j \leq N)\ ,$$

$$(c)\qquad \pi(1,i,j) = 0 \qquad (i \geq 0,\ j \geq N+1)\ .$$

$$(d)\qquad \pi(2,i,j) = \sum_{a=0}^{i} \pi(2,a,j+1)q_{i-a} \qquad (i \geq 0,\ 0 \leq j \leq N-2)\ .$$

$$(e)\qquad \pi(2,i,N-1) = \sum_{a=0}^{i} \pi(1,a,N)q_{i-a} \qquad (i \geq 0)\ ,$$

$$(f)\qquad \pi(2,i,j) = 0 \qquad (i \geq 0,\ j \geq N)\ .$$

When $G(x) = \sum_{a=0}^{\infty} x^a \pi(1,a,N)$ is defined, we obtain the following generating functions:

$$(2.2)\qquad U_1(x,y) = yP(x)\,\frac{\left[\begin{array}{l} U_2(x,0) - U_1(0,y) - y^N G(x) + y^N \pi(1,0,N) \\ - \pi(2,0,0) + x[U_1(0,y) + \pi(2,0,0) \\ - y^N \pi(1,0,N)] \end{array}\right]}{x - yP(x)}$$

and

$$(2.3)\qquad U_2(x,y) = \frac{y^N G(x) - U_2(x,0)}{y - Q(x)}\,Q(x)\ .$$

In order to completely characterize the state probabilities $\pi(k,i,j)$ we now need to determine $\pi(2,0,0)$, $U_1(0,y)$, $U_2(x,0)$ and $G(x)$:

Since $|U_2(0,y)| < 1$ for $|y| \leq 1$, in particular when $y \to q_o$, Equation 2.3 yields

$$(2.4) \qquad \pi(2,0,0) = q_o^N \pi(1,0,N) = q_o^N G(0)$$

and then, by taking the derivatives of $U_2(0,y)$ at $y = 0$, we get

$$(2.5) \qquad \pi(2,0,j) = \frac{1}{q_o^j} \pi(2,0,0) \quad .$$

From Equations 2.2 and 2.3, using the arguments preceding Equation 3.2 of Chapter II we get

$$(2.6) \qquad U_1(0,y) = \frac{U_2(\delta,0) - y^N G(\delta) + \left[\left(\frac{y}{q_o}\right)^N - 1\right] (1-\delta)\pi(2,0,0)}{1-\delta}$$

$$= \frac{U_2(\delta,0) - y^N G(\delta)}{1-\delta} + \left[\left(\frac{y}{q_o}\right)^N - 1\right] \pi(2,0,0)$$

and

$$(2.7) \qquad U_2(x,0) = [Q(x)]^N G(x)$$

where $\delta = \delta(y)$ is, as before, the inverse function of $y = x/P(x)$.

At this point it only remains to obtain $G(x)$, since the quantity $\pi(2,0,0)$ is, from Equation 2.4, equal to $q_o^N G(0)$.

We have

$$G(x) = \frac{1}{N!} \frac{\partial^N}{\partial y^N} U_1(x,y)\Big|_{y=0} , \tag{2.8}$$

where, from Equation 2.2, we can write

$$\frac{\partial^N}{\partial y^N} U_1(x,y) = P(x) \sum_{k=0}^{N} \binom{N}{k} \frac{\partial^k}{\partial y^k} A(x,y) \frac{\partial^{N-k}}{\partial y^{N-k}} B(x,y) \tag{2.9}$$

with

$$A(x,y) = y[U_2(x,0) - (1 - x)\pi(2,0,0)]$$

$$+ [(1 - x)\pi(1,0,N) - G(x)]y^{N+1} - y(1 - x)U_1(0,y)$$

$$\frac{\partial}{\partial y} A(x,y)\Big|_{y=0} = U_2(x,0) + (1 - x)[G(0) - q_0^N] - G(x)$$

$$- (1 - x)U_1(0,y)$$

$$= -\{1 - [Q(x)]^N\} G(x) + (1 - x)[G(0) - q_0^N]$$

$$- (1 - x)U_1(0,y)$$

$$\frac{\partial^k}{\partial y^k} A(x,y)\Big|_{y=0} = -(1 - x) \frac{d^k}{dy^k} U_1(0,y)\Big|_{y=0} = -(1 - x)kU_1^{(k-1)}$$

$$(2 \leq k \leq N)$$

and

$$B(x,y) = \frac{1}{x - yP(x)}$$

$$\frac{\partial^{N-k}}{\partial y^{N-k}} B(x,y)\Big|_{y=o} = \frac{(N-k)!\,[P(x)]^{N-k}}{x^{N-k+1}} \qquad (0 \le k \le N) \quad .$$

Hence

$$(2.10) \qquad G(x) = \frac{(1-x)[P(x)]^N}{[P(x)Q(x)]^N - x^N} \left[q_o^N G(0) + \sum_{k=1}^{N-1} \frac{1}{k!} \left(\frac{x}{P(x)} \right)^k U_1^{(k)} \right] \quad .$$

The quantities $U_1^{(k)} = (d^k/d^{yk})U_1(0,y)\big|_{y=o}$ are obtained from Equations 2.6 and 2.7 in terms of $G^{(k)} = (d/dx)G(x)\big|_{x=o}$, $0 \le k \le N-1$. Then, by taking the derivatives $G^{(k)}$ in Equation 2.10, we will have a system of N linear equations and N unknowns, namely $G^{(o)}, G^{(1)},\ldots,G^{(N-1)}$.

We have

$$(2.11) \qquad U^{(k)} = \frac{d^k}{dy^k} \frac{[Q(\delta)]^N G(\delta)}{1-\delta}\Big|_{y=o} \qquad (1 \le k \le N-1) \quad .$$

The formula for differentiation of a function of a function is

$$(2.12) \qquad \frac{d^k}{dx^k} g(g) = \sum_{m=o}^{k} f^{(m)}(g) \sum (k;a_1,a_2,\ldots,a_k)' \cdot [g^{(1)}]^{a_1}[g^{(2)}]^{a_2}\ldots [g^{(k)}]^{a_k}$$

summed over $a_1 + 2a_2 + \ldots + ka_k = k$ and $a_1 + a_2 + \ldots + a_k = m$. $f^{(m)}$ and $g^{(k)}$ denote the derivatives $(d^m/dg^m)f$ and $(d^k/dx^k)g$, respectively, and

$$(k;a_1,a_2,\ldots,a_k)' = \frac{k!}{a_1!(1!)^{a_1}a_2!(2!)^{a_2}\ldots a_k!(k!)^{a_k}} .$$

The coefficients $a_1,a_2,\ldots,a_k$ and $(k;a_1,a_2,\ldots,a_k)'$, for each term of the inner sum of Equation 2.2, are given by Table 24.2 or reference 1, for $1 \leq k \leq 10$.

With the above result we rewrite Equation 2.11 as

$$U_1^{(k)} = \sum_{\ell=0}^{k} z_{k\ell} \frac{d^\ell}{dy^\ell} G(\delta)\Big|_{y=0} \qquad (1 \leq k \leq N-1) \tag{2.13}$$

where

$$z_{k\ell} = \binom{k}{\ell} \frac{d^{k-\ell}}{dy^{k-\ell}} \frac{[Q(\delta)]^N}{1-\delta}\Big|_{y=0} \qquad \begin{matrix} 1 \leq k \leq N-1 \\ 0 \leq \ell \leq N-1 \end{matrix} \tag{2.14}$$

and

$$\frac{d^\ell}{dy^\ell} G(\delta)\Big|_{y=0} = \sum_{m=0}^{\ell} r_{m\ell}\, G^{(m)} \qquad (0 \leq \ell \leq N-1) \tag{2.15}$$

with

$$r_{\ell m} = \sum_{\ell} (\ell;a_1,a_2,\ldots,a_\ell)' [\delta^{(1)}]^{a_1}[\delta^{(2)}]^{a_2}\ldots[\delta^{(\ell)}]^{a_\ell} \tag{2.16}$$

$$(0 \leq \ell,m \leq N-1)$$

summed over $a_1 + 2a_1 + \ldots + \ell a_\ell = \ell$ and $a_1 + a_2 + \ldots + a_\ell = m$, and where

$$\delta^{(\ell)} = \frac{d^\ell}{dy^\ell} \delta(y)\Big|_{y=0} .$$

Hence

$$U_1^{(k)} = \sum_{\ell=0}^{k} \sum_{m=0}^{\ell} z_{k\ell} r_{\ell m} G^{(m)} \qquad (1 \le k \le N-1) \quad . \tag{2.17}$$

$$\alpha_{ac} = \binom{c}{a} \frac{d^a}{dx^a} \left. \frac{(1-x)[P(x)]^N}{[P(x)Q(x)]^N - x^N} \right|_{x=0} \tag{2.18}$$

and

$$\beta_{ck} = \frac{1}{k!} \frac{d^c}{dx^c} \left. \left(\frac{x}{P(x)} \right)^k \right|_{x=0} \quad . \tag{2.19}$$

Then, using Equation 2.10, we obtain the system of N linear equations with unknowns $G^{(0)}, G^{(1)}, \ldots, G^{(N-1)}$,

$$G^{(c)} = \sum_{a=1}^{c} \sum_{k=1}^{N-1} \sum_{\ell=0}^{k} \sum_{m=0}^{\ell} \alpha_{ac} \beta_{ck} g_{k\ell} r_{\ell m} G^{(m)} + \alpha_{oc} q_o^N G^{(o)} \qquad (1 \le c \le N-1) \quad . \tag{2.20}$$

This is an indeterminate linear system. One additional equation is obtained by using the fact that $G(1) = 1/2N$ (which can be seen from Equations 2.2 and 2.3 and the relatiionship $U_1(1,1) + U_2(1,1) = 1$). Equation 2.10 and L'Hopital's rule give

$$(2.21) \qquad q_o^N G(o) = \pi(2,0,0) = \frac{1}{2}(1 - \rho_1 - \rho_2) - \sum_{k=1}^{N-1} \frac{1}{k!} U_1^{(k)}$$

$$= \frac{1}{2}(1 - \rho_1 - \rho_2) - \sum_{k=1}^{N-1} \sum_{\ell=o}^{k} \sum_{m=o}^{\ell} \frac{1}{k!} z_k \, r_{\ell m} G^{(m)} \quad .$$

2.1 Calculation of $r_{\ell m}$, $z_{k\ell}$, β_{ck} and α_{ac}. Our first step is to find the derivatives $\delta^{(\ell)} = (d^\ell/dy^\ell)\delta(y)\big|_{y=o}$. We have $\delta = yP(\delta)$ which gives

$$(2.22) \qquad \delta^{(\ell)} = \ell \left. \frac{d^{\ell-1}}{dy^{\ell-1}} P(\delta) \right|_{y=o} ,$$

where the derivative of $P(\delta)$ are found by using Equation 2.12. For example, using Table 24.2 of Reference 1, and denoting $(d^k/dx^k)P(x)\big|_{x=o} = P^{(k)} = k!p_k$ we get

$$\delta^{(1)} = P^{(o)}$$

$$\frac{1}{2}\delta^{(2)} = \delta^{(1)}P^{(1)}$$

$$\frac{1}{3}\delta^{(3)} = \delta^{(2)}P^{(1)} + [\delta^{(1)}]^2 P^{(1)}$$

$$\frac{1}{4}\delta^{(4)} = \delta^{(3)}P^{(1)} + 3\delta^{(1)}\delta^{(2)}P^{(2)} + [\delta^{(1)}]^3 P^{(3)}$$

$$\frac{1}{5}\delta^{(5)} = \delta^{(4)}P^{(1)} + \{3[\delta^{(2)}]^2 + 4\delta^{(1)}\delta^{(3)}\}P^{(2)} + 6[\delta^{(1)}]^2 P^{(3)} + [\delta^{(1)}]^4 P^{(4)}$$

. etc.

Thus it is possible to determine $\delta^{(\ell)}$ recursively. We can now calculate $r_{\ell m}$ by Equation 2.16.

The derivative of the expression for $z_{k\ell}$ (Equation 2.14) is obtained as follows:

$$(2.23) \qquad \frac{d^k}{dy^k} \left. \frac{[Q(\delta)]^N}{1-\delta} \right|_{y=0} = \sum_{i=0}^{k} a_{ki} b_i \ ,$$

where

$$(2.24) \qquad a_{ki} = \binom{k}{i} \frac{d^{k-i}}{dy^{k-i}} \left. [Q(\delta)]^N \right|_{y=0} = \binom{k}{i} \left. \frac{d^{k-i}}{dy^{k-i}} Q_N(\delta) \right|_{y=0}$$

$$= \binom{k}{i} \sum_{m=0}^{k-i} Q_N^{(m)} \sum (k-i, a_1, a_2, \ldots, a_{k-i}) ,$$

$$\cdot [\delta^{(1)}]^{a_1} [\delta^{(2)}]^{a_2} \ldots [\delta^{(k-i)}]^{a_{k-i}}$$

with $Q_N^{(m)} = m!\, q_m^{*N}$, where $\{q_m^{*N}\}$ is the Nth convolution of $\{q_m\}$ with itself, and

$$(2.25) \qquad b_i = \left. \frac{d^i}{dy^i} \frac{1}{1-\delta} \right|_{y=0} = \sum_{m=0}^{i} \sum (i; a_1, a_2, \ldots, a_i) ,$$

$$\cdot [\delta^{(1)}]^{a_1} \ldots [\delta^{(i)}]^{a_i} \ .$$

Hence

$$(2.26) \qquad z_{k\ell} = \binom{k}{\ell} \sum_{i=0}^{k-\ell} a_{k-\ell, i} \, b_i \ .$$

The quantity β_{ck} is

$$\beta_{ck} = \begin{cases} 0 \ , & \text{if} \quad c < k \\ \dfrac{d^{c-k}}{dx^{c-k}} [1/P(x)]^k \Big|_{x=0} \ , & \text{if} \quad c \geq k \end{cases} \ , \tag{2.27}$$

where

$$\frac{d^{c-k}}{dx^{c-k}} \left(\frac{1}{P(x)} \right)^k = \sum_{m=0}^{c-k} \frac{1}{P_k^{(m)}} \sum (c-k; a_1, a_2, \ldots, a_{c-k}) \cdot [p_1]^{a_1} [p_2]^{a_2} \ldots [p_{c-k}]^{a_{c-k}}$$

with $P_k^{(m)} = m! p_m^{*k}$, where $\{p_m^{*k}\}$ is the k-fold convolution of $\{p_m\}$ with itself.

The quantity α_{ac} is

$$\alpha_{ac} = \binom{c}{a} \frac{d^a}{dx^a} \frac{1}{[Q(x)]^N - [x/P(x)]^N} \Big|_{x=0} - \binom{c}{a} \frac{d^{a-1}}{dx^{a-1}} \frac{1}{[Q(x)]^N - [x/P(x)]^N} \Big|_{x=0} \ , \tag{2.28}$$

where

$$(2.29) \qquad \frac{d^a}{dx^a} \frac{1}{[Q(x)]^N - [x/P(x)]^N}\Big|_{x=0}$$

$$= \sum_{m=0}^{a} \frac{1}{q_o^N} \sum^{m} (a;a_1,a_2,\ldots,a_a)' [q_1^{*N} - \beta_{1N}]^{a_1}$$

$$\cdot [q_2^{*N} - \beta_{2N}]^{a_2} \cdots [q_a^{*N} - \beta_{aN}]^{a_a}$$

$$= \sum_{m=0}^{a} \left(\frac{1}{q_o^N}\right)^m \sum (a;a_1,a_2,\ldots,a_a)'$$

$$\cdot [q_1^{*N}]^{a_1} [q_2^{*N}]^{a_2} \cdots [q_a^{*N}]^{a_a}$$

since $\beta_{aN} = 0$ because $a < N$.

2.2 Expected Number of Customers and Waiting Times. Using Equations 2.2, 2.6, 2.7 and the relationship $U_1(1,1) + U_2(1,1) = 1$ we, first get

$$G(1) = \frac{1}{2N} ,$$

$$\frac{d}{dx} U_2(x,0)\Big|_{x=1} = N\rho_2 G(1) + G'(1) = \frac{1}{2}\rho_2 + G'(1) ,$$

and

$$\frac{d}{dy} U_1(0,y)\Big|_{y=1} = \frac{1}{2}(1 - \rho_1 - \rho_2) + \pi(1,0,N) - \pi(2,0,0) .$$

Then, applying L'Hopital's rule in order to evaluate the partial derivatives of $U_k(x,y)$ at $x = y = 1$, we obtain

$$(2.30) \qquad \bar{Q}_{11} = \frac{1 - \rho_1 - \rho_2}{1 - \rho_1}(\rho_1 + Q_o) + \frac{\rho_2}{1 - \rho_1}[1/2(N - 1)\rho_2 + 2NG'(1)] ,$$

$$(2.31) \qquad \bar{Q}_{21} = \frac{N + 1}{2} ,$$

$$(2.32) \qquad \bar{Q}_{12} = \frac{1}{2}(N + 1)\rho_2 + 2G'(1) ,$$

$$(2.33) \qquad \bar{Q}_{22} = \frac{N - 1}{2} ,$$

where

$$Q_o = \frac{\lambda^2 E[(S_1 + S_2)^2]}{2(1 - \rho_1 - \rho_2)} .$$

The expected values $\bar{Q}_{k.}$, $\bar{Q}_{.k}$ and $\bar{Q}$ can be found easily from the above results.

The waiting times, W_1 and W_2, are given as in Section 5 of Chapter II.

2.3 The SSS ($Q_2 = 2$) Policy

The long differentiation procedure used in the determination of $G(x)$, for an arbitrary N, is very simple for $N = 2$. In this subsection we shall study this case.

Equation 2.10 reduces to

$$(2.34) \qquad G(x) = \frac{(1 - x)[P(x)]^2}{[P(x)Q(x)]^2 - x^2}[q_o^2 G^{(o)} + x/P(x)\, U_1^{(1)}] ,$$

where the quantity $U_1^{(1)} = d/dy\ U_1(0,y)\big|_{y=0}$ is obtained from Equation 2.6 in terms of $G^{(0)} = G(0)$ and $G^{(1)} = G'(0)$. We have

$$U_1^{(1)} = \delta^{(1)} \frac{d}{dx} U_2(x,0)\Big|_{x=0} + U_2(1,0) \tag{2.35}$$

which, combined with Equation 2.7, yields

$$U_1^{(1)} = 2p_o q_o q_1 G^{(o)} + G^{(1)} \quad . \tag{2.36}$$

The quantity $G^{(1)}$ is calculated from Equation 2.38. We get

$$G^{(1)} = -\frac{q_o + 2q_1}{q_o} G^{(o)} + \frac{1}{p_o q_o^2} U_1^{(1)} \quad , \tag{2.37}$$

which, combined with Equations 2.36 and 2.21, determines the quantities $G^{(o)}$ and $U_1^{(1)}$ of Equation 2.34. We get

$$G(x) = \frac{1}{2}(1 - \rho_1 - \rho_2) \frac{1 - x}{[Q(x)]^2 - \left[\frac{x}{P(x)}\right]^2} \cdot \frac{q_o + p_o(q_o + 2q_1)\frac{x}{P(x)}}{q_o + p_o(q_o + 2q_1)} \quad , \tag{2.38}$$

and the probability of an empty system,

$$(2.39) \qquad \pi(2,0,0) = \frac{1}{2}(1 - \rho_1 - \rho_2)\frac{q_o}{q_o + p_o(q_o + 2q_1)} \quad .$$

Notice that the nonsaturation condition is still $1 - \rho_1 - \rho_2 > 0$.

The expected number of customers given by Equations 2.30 through 2.33 are

$$(2.40) \qquad \bar{Q}_{11} = \frac{1 - \rho_1 - \rho_2}{1 - \rho_1}(\rho_1 + Q_o) + \frac{\rho_2}{1 - \rho_1}\left[\frac{1}{2}\rho_2 + 2G'(1)\right] ,$$

$$(2.41) \qquad \bar{Q}_{21} = \frac{3}{2} ,$$

$$(2.42) \qquad \bar{Q}_{12} = \frac{3}{2}\rho_2 + 2G'(1) ,$$

$$(2.43) \qquad \bar{Q}_{22} = \frac{1}{2} ,$$

where, from Equation 2.38,

$$(2.44) \qquad G'(1) = \frac{p_o(q_o + 2q_1)(1 - \rho_1)}{2[q_o + p_o(q_o + 2q_1)]} + \frac{1}{2}(3\rho_1 - \rho_2 - 1) + Q_o \quad .$$

2.4 Busy Periods. The busy period B^{II} during which the server is continuously working in the second stage has distribution

$$(2.45) \qquad B^{II}(t) = S_2^{*N}(t) \quad .$$

The busy period B^I is defined as the time necessary to accumulate N customers in the second stage. During this time the server may be inactive in the first stage. This happens when the first stage becomes empty and the number of customers in the second stage is less than N. We see that if B^I starts with i 1-customers, its duration has the distribution of the time until the Nth departure from an M/G/1 queue having service time distribution $S_1(t)$, given that the system starts with i customers.

Let α_N be the number of times the first stage becomes empty until just after the N - 1 first departures. Given $\alpha_N = k$, we see that B^I has distribution $S_1^{*N} * A^{*k}(t)$, where $A(t) = 1 - e^{-\lambda t}$ is the interarrival distribution and $S_1(t)$ is the service time distribution of the first stage. If Q_{12} is the number of customers in the first stage at the time the server arrives from the second stage, we have

$$(2.46) \qquad B^I(t) = \frac{1}{U_2(1,0)} \sum_{k=0}^{N-1} \sum_{i=0}^{\infty} \pi(2,i,0) \cdot P_r[\alpha_N = k | Q_{12} = i] S_1^{*N} * A^{*k}(t) \quad .$$

In Takacs ([21], page 98), we find

$$(2.47) \qquad P_r[\alpha_N > k | Q_{12} = i] = P_r[\eta < N | Q_{12} = i + k] \quad ,$$

where η is the number of customers served in a busy period. Thus we have

$$(2.48)\qquad B^{I}(t) = \frac{1}{U_2(1,0)} \sum_{i=0}^{N-1} \sum_{k=0}^{N-i} \pi(2,i,0)$$

$$\cdot\Big[\{P[\eta < N|Q_{12} = i + k - 1] - P[\eta < N|Q_{12} = i + k]\}\Big]$$

$$\cdot\, S_1^{*} * A^{*k}(t) + \frac{1}{U_2(1,0)} \sum_{i=N}^{\infty} \pi(2,i,0)\; S_1^{*N}(t) \quad ,$$

where

$$P[\eta < N|Q_{12} = j] = \sum_{n=0}^{N-1} \frac{j}{n} \int_0^{\infty} \frac{(\lambda\mathcal{T})^{n-j}}{(n-j)!}\, e^{-\lambda\mathcal{T}}\, dS_1^{*n}(\mathcal{T}) \qquad (N > j)$$

$$= 0 \quad . \qquad (N \leq j) \quad .$$

Simplification of the right-hand side of the above equation is not apparent.

The busy period of the server, B^S, is defined as the time elapsed since the server starts the service of a 1-customer that terminates an empty period of the system, until the system becomes empty again for the first time. This busy period may contain, as for the busy period B^I, intervals corresponding to situations when the first stage becomes empty and the number of customers in the second stage is less than N.

We have not succeeded in finding the distribution $B^S(t)$, nor its Laplace-Stieltjes transform.

3. The W_NFS Policy

Under this policy the server switches from the first to the second stage when the first stage is empty and returns to the first stage

when all of the second stage is empty and if at least N 1-customers are present. Otherwise the server removes himself from the system, and returns to the first stage when N 1-customers are present. As indicated in Section 1, the W_NFS policy is a generalization of the LNB policy discussed in Chapter II. For this reason we expect to have a solution procedure similar to the iterative scheme presented in Section 3 of Chapter II. Some difficulties will arise with this type of solution, but as in Chapter II, we are able to find the state probabilities by a direct approach.

The stationary conditions are

(3.1) (a) $$\pi(1,i,1) = 0 \qquad (i < N-1)$$

(b) $$\pi(1,i,1) = \sum_{a=N}^{i+1} \pi(2,a,0)p_{i-a+1} + p_{i-N+1} \sum_{a=0}^{N-1} \pi(2,a,0) \qquad (i \geq N-1)$$

(c) $$\pi(1,i,j) = \sum_{a=1}^{i+1} \pi(1,a,j-1)p_{i-a+1} \qquad (i \geq 0, j \geq 2)$$

(d) $$\pi(2,i,j) = \pi(1,0,j+1)q_i + \sum_{a=0}^{i} \pi(2,a,j+1)q_{i-a}$$

$$(i \geq 0, j \geq 0) \quad .$$

The generating functions of $\{\pi(1,i,j)\}$ and $\{\pi(2,i,j)\}$ are

$$(3.2) \qquad U_1(x,y) = yP(x)\,\frac{U_2(x,0) - U_1(0,y) + x^N H(1) - H(x)}{x - yP(x)}$$

and

$$(3.3) \qquad U_2(x,y) = \frac{Q(x)[U_1(0,y) - U_2(x,0)]}{y - Q(x)} ,$$

where $H(x) = \sum_{a=0}^{N-1} \pi(2,a,0)x^a$.

Using the arguments preceding Equation 3.2 of Chapter II, we obtain

$$(3.4) \qquad U_1(0,y) = U_2(\delta,0) + \delta^N H(1) - H(\delta)$$

and

$$(3.5) \qquad U_2(x,0) = U_1(0,Q) ,$$

where, once again $\delta = \delta(y)$ is the inverse function of $y = x/P(x)$, and $Q = Q(x)$.

Let $x_o = x$ and $y_o = y$ with $|x| \leq 1$ and $|y| < 1$. Define x_i and y_i, $i \geq 0$, by

$$(3.6) \qquad x_{i+1} = \delta(y_i)$$

and

$$(3.7) \qquad y_{i+1} = Q(x_i) .$$

Substituting Equations 3.6 and 3.7 into Equations 3.4 and 3.5 we get

$$(3.8) \qquad U_1(0,y_i) = U_2(x_{i+1},0) + x_{i+1}^N H(1) - H(x_{i+1})$$

and

$$U_2(x_i,0) = U_1(0,y_{i+1}) \quad . \tag{3.9}$$

Adding Equations 3.8 for i = 0,2,...,2N-2, to Equations 3.9 for i = 1,3,...,2N-1, we obtain

$$U_1(0,y) = U_1(0,y_{2N}) + \sum_{n=0}^{N-1} [x_{2n+1}^N H(1) - H(x_{2n+1})] \quad . \tag{3.10}$$

Similarly, let i = 1,3,...,2N-1 in Equation 3.8 and i = 0,2,...,2N-2 in Equation 3.9. Summing these equations we get

$$U_2(x,0) = U_2(x_{2N},0) + \sum_{n=1}^{N} [x_{2n}^N H(1) - H(x_{2n})] \quad . \tag{3.11}$$

We proved in Section 3 of Chapter II that the sequences $\{x_{2N}\}$ and $\{y_{2N}\}$ converge to unity. The sums in Equations 3.10 and 3.11 are convergent, since $|H(x_i)| \leq 1$ for $|x_i| \leq 1$ and the sums $\sum_{n=1}^{N} (x_{2n} - 1)$ and $\sum_{n=0}^{N-1} (x_{2n+1} - 1)$ are convergent (see Chapter II).

Thus we may write

$$U_1(0,y) = U_1(0,1) - A_1(y) \tag{3.12}$$

and

$$U_2(x,0) = U_2(1,0) - A_2(x) \quad , \tag{3.13}$$

where

$$A_1(y) = \sum_{n=0}^{\infty} [H(x_{2n+1}) - x_{2n+1}^N H(1)] \tag{3.14}$$

and

$$(3.15) \qquad A_2(x) = \sum_{n=1}^{\infty} [H(x_{2n}) - x_{2n}^{N} H(1)] \quad .$$

Notice that when N = 1 we have $H(x) = \pi(2,0,0)$, consequently Equations 3.12 and 3.13 reduce to Equations 3.11 and 3.12 of Chapter III, respectively.

We now discuss the difficulties of determining H(x). First we notice that

$$(3.16) \qquad H^{(k)}(0) = \frac{d^k}{dx^k} H(x)\Big|_{x=o} = \frac{d}{dx^k} U_2(x,0)\Big|_{x=o}$$

$$= U_2^{(k)}(0,0) \qquad (0 \leq k \leq N-1) \quad .$$

One method to be attempted is to use Equations 3.5, 3.4 and 3.16 in order to relate the quantities $H^{(k)}(0)$. We have

$$(3.17) \qquad U_2'(x,0) = Q'(x) U_1'[0,Q(x)]$$

and

$$(3.18) \qquad U_1'[0,Q(x)] = \delta'[Q(x)] \; U_2'[\delta[Q(x)],0] + N[\delta[Q(x)]]^{N-1}$$

$$- H'[\delta(Q(x))] \quad .$$

Unfortunately $Q(0) = q_o$; this means that, in general, $H^{(k)}(0)$ will appear as functions of unknown quantities, $H^{(k)}(q_o)$.

Fortunately, however, the state probabilities $\pi(k,i,j)$ can be determined by the direct approach of Section 7 of Chapter II, which is based on the properties of the busy period of an M/G/1 system. Let $b_m(i)$, $a_i(n)$ and $c_k(n)$ be as defined in Section 7 of Chapter II. We have

$$(3.19) \qquad c_1(n) = \sum_{i=N}^{\infty} b_N(i)a_i(n) \quad ,$$

$$c_k(n) = \left[\sum_{m=1}^{N-1} c_{k-1}(m)\right] \sum_{i=N}^{\infty} b_N(i)a_i(n) + \sum_{m=N}^{\infty} \sum_{i=m}^{\infty} c_{k-1}(m)b_m(i)a_i(n) \qquad (k > 1) \quad ,$$

where $b_m(i)$ and $a_i(n)$ are given by Equations 7.2 and 7.3 of Chapter II, respectively. The state probabilities $\pi(2,n,0)$ can be written as

$$(3.20) \qquad \pi(2,n,0) = \pi(2,0,0) \sum_{k=1}^{\infty} c_k(n) \qquad (n > 0) \quad ,$$

where $\pi(2,0,0)$ is determined by the relationship

$$(3.21) \qquad NH(1) - H'(1) = N \sum_{n=0}^{N-1} \pi(2,n,0) - \sum_{n=1}^{N-1} n\pi(2,n,0)$$

$$= \frac{1}{2}(1 - \rho_1 - \rho_2)$$

which is derived from Equations 3.3 through 3.5 combined with the fact that $U_2(1,1) = 1/2$. The remaining state probabilities are found by Equations 3.1.

3.1 Expected Number of Customers and Waiting Times. Using Equations 3.4, 3.5, 3.21 and the fact that $\delta'(1) = (1 - \rho_1)^{-1}$, we find that $U_1'(0,1) = 1/2$ and $U_2'(1,0) = 1/2\ \rho_2$.* Next, knowing that $\delta''(1) = (1 - \rho_1)^{-2}\ [2\rho_1 + \lambda^2 E[S_1^2](1 - \rho_1)^{-1}]$, we obtain

$$U_1''(0,1) = \frac{1}{1 - \rho_1 + \rho_2}(\rho_1 + Q_o) + \frac{1}{(1 - \rho_1)^2 - \rho_2^2} \cdot [N(N - 1)H(1) - H''(1)]$$

and

$$U_2''(1,0) = \frac{1}{2}\lambda^2 E[S_2^2] + \rho_2^2 U_1''(0,1) \quad ,$$

where

$$Q_o = \frac{\lambda^2 E[(S_1 + S_2)^2]}{2(1 - \rho_1 - \rho_2)} \quad .$$

These derivatives are then used to obtain the expected values $\bar{Q}_{k\ell}$ from Equations 3.2 and 3.3. It results that formulas 4.1 through 4.10 of Chapter II are also valid for this case if for Q_o we substitute

$$Q_o + (1 - \rho_1 - \rho_2)^{-1}\ [N(N - 1)H(1) - H''(1)] \quad .$$

The probability distributions and expected values of the waiting times W_1 and W_2 are given as in Section 5 of Chapter II. The values of $U_k(x,y)$ are given by Equations 3.2 and 3.3; the quantities $\bar{Q}_{11}$ and $\bar{Q}_{.2}$ are obtained as indicated above.

*

$$U_1'(0,1) = \frac{\partial}{\partial y} U_1(x,y)\Big|_{\substack{x=0 \\ y=1}} \quad \text{and} \quad U_2'(1,0) = \frac{\partial}{\partial x} U_2(x,y)\Big|_{\substack{x=1 \\ y=0}}$$

3.2 Busy Periods. For this $W_N FS$ policy the busy period of the server, B^S, is the time elapsed from the moment the server returns to the system (N 1-customers and no 2-customers are present) until the departure of a 2-customer who leaves behind him less than N 1-customers and no 2-customers. In order to find the distribution of B^S we notice that, if the same realizations of arrivals and services are simulated in an M/G/1 system, with service time equal to the sum of the service times in both stages of our sequential system, and if at the termination of B^S the sequential system has i customers (1-customers), then the M/G/1 system will also have i customers (assuming N customers were present at the beginning of the simulation), $0 \leq i < N$. This occurs with probability $\pi(2,i,0)$ (unconditionally) and distribution $B_o^{*(N-i)}(t)$, where $B_o(t)$ is the distribution of the busy period of an M/G/1 system with service time distribution $S_1 * S_2(t)$. Hence, conditional to the set of all possible ways of starting the busy period B^S, which has probability $H(1) = \sum_{i=0}^{N-1} \pi(2,i,0)$, we have

$$B^S(t) = \frac{1}{H(1)} \sum_{i=0}^{N-1} \pi(2,i,0)\, B_o^{*(N-i)}(t) \quad . \tag{3.22}$$

The Laplace-Stieltjes transform is

$$\tilde{B}^{(S)}(s) = \frac{1}{H(1)} [\tilde{B}(s)]^N H\left[\frac{1}{\tilde{B}(s)}\right] \tag{3.23}$$

$$\begin{aligned} \bar{B}^S &= \frac{\bar{B}_o}{H(1)} [NH(1) - H'(1)] \\ &= \frac{\bar{S}_1 + \bar{S}_2}{2H(1)} \quad , \end{aligned} \tag{3.24}$$

because of Equation 3.21 and the fact that $\bar{B}_o = (\bar{S}_1 + \bar{S}_2)/(1 - \rho_1 - \rho_2)$.

A similar situation occurs with the distribution of the time during which the server is removed, I(t). We have

$$(3.25) \qquad I(t) = \frac{1}{H(1)} \sum_{i=o}^{N-1} \pi(2,i,0)\, A^{*(N-i)}(t) \quad ,$$

where $A(t) = 1 - e^{-\lambda t}$ is the distribution of the interarrival times. We get

$$(3.26) \qquad \tilde{I}(s) = \frac{1}{H(1)} \left[\frac{\lambda}{\lambda + s} \right]^N H\left[\frac{\lambda + s}{\lambda} \right]$$

and

$$(3.27) \qquad \bar{I} = \frac{1}{2\lambda H(1)} (1 - \rho_1 - \rho_2) \quad .$$

The busy periods of the first and second stages, B^I and B^{II}, are defined as before.

For the first stage we have

$$(3.28) \qquad B^I(t) = \frac{1}{U_2(1,0)} \left[H(1) B_N(t) + \sum_{i=N}^{\infty} \pi(2,i,0) B_i(t) \right] \quad ,$$

where, as in Equation 6.1 of Chapter II, $B_i(t)$ is the distribution of the busy period initiated by i customers in an M/G/1 system with service time distribution $S_1(t)$. We get

$$(3.29) \qquad \tilde{B}^{I}(s) = \frac{1}{U_2(1,0)}\left[H(1)[\tilde{B}(s)]^N - H[\tilde{B}(s)] + U_2[\tilde{B}(s),0]\right] ,$$

where $U_2(x,0)$ is given by Equation 3.13. Equations 3.2 and 3.21 and the fact that $U_2'(1,0) = 1/2\ \rho_2$, yield

$$(3.30) \qquad \bar{B}^{I} = \frac{\bar{S}_1}{2U_2(1,0)} = \frac{\bar{S}_1}{2A_1(0)} ,$$

where $A_1(0)$ is given by Equation 3.14.

For the second stage we have

$$(3.31) \qquad B^{II}(t) = \frac{1}{U_1(0,1)} \sum_{j=1}^{\infty} \pi(1,0,j) S_2^{*j}(t) ,$$

$$(3.32) \qquad \tilde{B}^{II}(s) = \frac{1}{U_1(0,1)} U_1[0,\tilde{S}_2(s)]$$

and

$$(3.33) \qquad \bar{B}^{II} = \frac{\bar{S}_2}{2U_1(0,1)} = \frac{\bar{S}_2}{2U_2(1,0)} = \frac{\bar{S}_2}{2A_1(0)} .$$

The average number of cycles in the busy period of the server is

$$(3.34) \qquad \bar{c} = \frac{\bar{B}^{S}}{\bar{B}^{I} + \bar{B}^{II}} = \frac{A_1(0)}{H(1)} = \sum_{n=0}^{\infty} \left[\frac{H(x_{2n+1}^{o})}{H(1)} - x_{2n+1}^{o}\right]$$

$$= \sum_{k=1}^{\infty} k c_k(0) ,$$

4. The SFS ($Q_i \geq N$) Policy

Under this policy the server switches to the other state leaving no backlog in the present state, except when, at departure of a 2-customer, at least N 1-customers are present in which case the server switches to the first stage. The analysis of this policy is similar to the analysis of the W_NFS policy. In this case the function H is two-dimensional.

The steady-state probabilities are related by

$$(4.1)\quad (a)\qquad \pi(1,i,1) = \pi(2,0,0)p_i + \sum_{a=1}^{i+1} \pi(2,a,0)p_{i-a+1} \qquad (i \geq 0)$$

$$(b)\qquad \pi(1,i,j) = \sum_{a=1}^{i+1} \pi(1,a,j-1)p_{i-a+1} \qquad (0 \leq i \leq N-2,\ j \geq 2)$$

$$(c)\qquad \pi(1,i,j) = \sum_{a=1}^{i+1} \pi(1,a,j-1)p_{i-a+1} + \sum_{a=N}^{i+1} \pi(2,a,j-1) \qquad (i \geq N-1,\ j \geq 2)$$

$$(d)\qquad \pi(2,i,j) = q_i \pi(1,0,j+1) + \sum_{a=0}^{i} q_{i-a}\, \pi(2,a,j+1) \qquad (9 \leq i \leq N-2,\ j \geq 0)$$

$$(e)\qquad \pi(2,i,j) = q_i \pi(1,0,j+1) + \sum_{a=0}^{N-1} q_{i-a}\, \pi(2,a,j+1) \qquad (i \geq N-1,\ j \geq 0)\ .$$

Letting

$$H(x,y) = \sum_{j=1}^{\infty} \sum_{a=0}^{\infty} x^{N+a} y^{j} \pi(2, N + a, j)$$

we get

$$(4.2) \qquad U_1(x,y) = \frac{(x - 1)\pi(2,0,0) + U_2(x,0) - U_1(0,y) + H(x,y)}{x - yP(x)} yP(x)$$

and

$$(4.3) \qquad U_2(x,y) = \frac{U_1(0,y) - U_2(x,0) - H(x,y)}{y - Q(x)} Q(x) \quad .$$

Notice that the two last equations have the same structure as Equations 2.5 and 2.6 of Chapter II. This is expected since the present policy, for $N \to \infty$, is the policy studied in Chapter II.

Applying once more the arguments of Chapter II we get

$$(4.4) \qquad U_1(0,y) = U_2(\delta,0) + (\delta - 1)\pi(2,0,0) + H(\delta,y)$$

and

$$(4.5) \qquad U_2(x,0) = U_1(0,Q) - H(x,Q) \ ,$$

where $Q = Q(x)$ and $\delta = \delta(y)$ is the inverse function of $y = x/P(x)$.

Using the same iteration scheme of Chapter II, we find

$$(4.6) \qquad U_1(0,y) = U_1(0,1) - \pi(2,0,0)A_1(y) + B_1(x,y)$$

and

$$U_2(x,0) = U_2(1,0) - \pi(2,0,0)A_2(x) + B_2(x,y) \quad , \tag{4.7}$$

where

$$A_1(y) = \sum_{n=0}^{\infty} (1 - x_{2n+1}) \quad , \tag{4.8}$$

$$A_2(x) = \sum_{n=1}^{\infty} (1 - x_{2n}) \quad , \tag{4.9}$$

$$B_1(x,y) = \sum_{n=0}^{\infty} \left[H(x_{2n+1},y_{2n}) - H(x_{2n+1},y_{2n+2})\right] \tag{4.10}$$

and

$$B_2(x,y) = \sum_{n=1}^{\infty} \left[H(x_{2n},y_{2n-1}) - H(x_{2n-2},y_{2n-1})\right] \quad . \tag{4.11}$$

The value of $\pi(2,0,0)$ can be determined analytically from the equations above and the fact that $U_k(1,1) = 1/2$. However we can see that its value is $1/2(1 - \rho_1 - \rho_2)$ since, under this policy, the server is active whenever customers are present in the system (see comments following Equation 3.5 of Chapter III).

If we let $x = y = 0$ in Equations 4.6 through 4.11, we get $A_1(0) = 1 + A_2(0)$ and $B_1(0,0) = B_2(0,0) - H(1,1)$. Thus

$$U_1(0,y) = \frac{1}{2}(1 - \rho_1 - \rho_2)A_1(0) - B_1(0,0) \tag{4.12}$$

and

$$(4.13) \qquad U_2(1,0) = \frac{1}{2}(1 - \rho_1 - \rho_2)A_1(0) - B_1(0,0) - H(1,1) \quad .$$

We shall now obtain the function $H(x,y)$, by finding the state probabilities $\pi(2,i,j)$, using an approach similar to the determination of the function $H(x)$ of the previous section.

Let $b_m(i)$ and $a_i(n)$ be as defined in Section 7 of Chapter II, and define

$c_k(m,n)$ = probability that, starting with an empty system, the termination of the kth cycle occurs when there are m 1-customers and n 2-customers in the system.

We have

$$(4.14) \quad (a) \qquad c_1(m,n) = \sum_{r=0}^{N-1} \sum_{i=m+1}^{\infty} b_1(i)a_{i-n-1}(r)a_1(m - r)$$

$$(b) \qquad c_k(m,n) = \sum_{u=1}^{\infty} \sum_{i=I_0}^{\infty} \sum_{r=0}^{N-1} c_{k-1}(u,0)b_u(i)a_{i-n-1}(r)a_1(m - r)$$

$$+ \sum_{n=N}^{\infty} \sum_{v=1}^{\infty} \sum_{i=I_v}^{\infty} \sum_{r=0}^{N-1} c_{k-1}(u,v)b_u(i)a_{i+v-n-1}$$

$$\cdot a_1(m - r) \qquad (k > 1) \quad ,$$

where $I_v = \max\{u + 1, n - v + 1\}$. From this we get

$$(4.15) \qquad \pi(2,m,n) = \frac{1}{2}(1 - \rho_1 - \rho_2) \sum_{k=1}^{\infty} c_k(m,n) \quad .$$

With this we can find $H(x,y)$ from its definition.

The remaining state probability, if needed, can be found from Equations 4.1.

4.1 Expected Number of Customers and Waiting Times. Let

$$H_x = \frac{\partial}{\partial x} H(x,y)\Big|_{\substack{x=1 \\ y=1}} ,$$

and let H_y, H_{xx}, H_{yy} and H_{xy} be similarly defined.

Using Equations 4.4 and 4.5 and the fact that $\delta'(1) = (1 - \rho_1)^{-1}$, we find $U_1'(0,1) = 1/2 + H_y$ and $U_2'(1,0) = 1/2\ \rho_2 - H_x$. Next, knowing that $\delta''(1) = (1 - \rho_1)^{-2}\,[2\rho_1 + \lambda^2 E[S_1^2](1 - \rho_1)^{-1}]$, we obtain

$$U_1''(0,1) = \frac{1}{1 - \rho_1 + \rho_2}(\rho_1 + Q_0) + \frac{2}{1 - \rho_1 + \rho_2} + H_{yy}$$

and

$$U_2''(1,0) = \frac{1}{2}\lambda^2 E[S_2^2] + \rho_2^2 U_1''(0,1) - \frac{2\rho_2}{1 - \rho_1 - \rho_2} H_{xy} - H_{xx} \quad .$$

These derivatives are then used to obtain the expected values $\bar{Q}_{k\ell}$ from Equations 4.2 and 4.3. We get

$$(4.16) \qquad \bar{Q}_{11} = (\rho_1 + Q_0 + 2\rho_2 H_{xy}) \frac{1 - \rho_1}{1 - \rho_1 + \rho_2} ,$$

$$(4.17) \qquad \bar{Q}_{21} = (1 + \rho_2 + Q_0 + 2H_{xy}) \frac{1}{1 - \rho_1 + \rho_2} ,$$

$$(4.18) \qquad \bar{Q}_{12} = \left[1 + \rho_2 + Q_o - \frac{2}{\rho_2} H_{xy}\right] \frac{\rho_2}{1 - \rho_1 + \rho_2}$$

and

$$(4.19) \qquad \bar{Q}_{22} = (\rho_1 + Q_o + 2H_{xy}) \frac{1}{1 - \rho_1 + \rho_2} .$$

The probability distributions and expected values of the waiting times W_1 and W_2 are as in Section 5 of Chapter II.

4.2 Busy Periods. By the arguments of Section 6 of Chapter II, the busy period of the server, B^S, has the distribution of the busy period of an M/G/1 system with service time distribution $S_1 * S_2(t)$. Consequently its expected value is $\bar{B}^S = (\bar{S}_1 + \bar{S}_2)/(1 - \rho_1 - \rho_2)$.

The distribution of the time the server is continuously serving 1-customers is

$$(4.20) \qquad B^I(t) = \frac{1}{H(1,1) + U_2(1,0)} \left[\sum_{i=N}^{\infty} \sum_{j=1}^{\infty} \pi(2,i,j) B_i(t) \right.$$

$$\left. + \sum_{i=1}^{\infty} \pi(2,i,0) B_i(t) + \pi(2,0,0) B_1(t) \right] ,$$

where $B_i(t)$, as in Section 6 of Chapter II, is the distribution of the busy period of the first stage initiated by i customers.

Knowing that $U_1(0,1) = U_2(1,0) + H(1,1)$ and $\bar{B}_1 = \bar{B} = \bar{S}_1/(1 - \rho_1)$ we get

$$(4.21)\qquad \tilde{B}^{I}(s) = \frac{1}{U_1(0,1)}\Big[H[\tilde{B}(s),1] + U_2[\tilde{B}(s),0] + \frac{1}{2}(1-\rho_1-\rho_2)[\tilde{B}(s)-1]\Big] ,$$

which implies

$$(4.22)\qquad \bar{B}^{I} = \frac{1}{2U_1(0,1)}\bar{S}_1 = \frac{1}{(1-\rho_1-\rho_2)A_1(0) - 2B_1(0,0)}\bar{S}_1 .$$

In order to derive the distribution of the busy period of the second station, we first get

$$(4.23)\qquad \bar{A}(n) = \sum_{i=n}^{\infty} a_i(i) = 1 - \lambda \sum_{i=0}^{n-1} \bar{S}_2^{(i)}$$

where

$$\bar{S}_2^{(i)} = \int_0^{\infty} \frac{t^i}{i!}\, dS_2(t) .$$

Now, noticing that B^{II} starts when the state of the system is $(1,0,j)$ with conditional probability $\pi(1,0,j)/U_1(0,1)$, and that the server can complete service of k 2-customers $(1 \le k \le j-1)$ with probability

$$(4.24)\qquad \bar{a}_{kj} = \sum_{r=0}^{N-1} a_{k-1}(r)\bar{A}(N-r) ,$$

and of j 2-customers with probability

$$(4.25)\qquad \bar{a}_{jj} = \sum_{r=0}^{N-1} a_{j-1}(r) ,$$

we get

$$(4.26) \qquad B^{II}(t) = \frac{1}{U_1(0,1)} \sum_{j=1}^{\infty} \pi(1,0,j) \sum_{k=1}^{j} \bar{a}_{kj} S^{*k}(t) \quad .$$

Instead of obtaining $\bar{B}^{II}$ from the equation above we can use the relationship

$$(4.27) \qquad \bar{c} = \sum_{k=1}^{\infty} k c_k(0,0) = \frac{\bar{B}^S}{\bar{B}^I + \bar{B}^{II}} \quad ,$$

where $c_k(0,0)$ is found by Equation 4.14. We get

$$(4.28) \qquad \bar{B}^{II} = \frac{\bar{S}_1 + \bar{S}_2}{(1 - \rho_1 - \rho_2)\bar{c}} - \bar{B}^I \quad .$$

CHAPTER V

DISCUSSION

1. Numerical Comparison of Policies

In this section we compare the effect of policies LNB, SSP and FSP, for the case of exponential service times, on the average values of the waiting time in the first stage ($\overline{W}_1$), the waiting time in the second stage ($\overline{W}'_2 = \overline{W}_2 - \overline{W}_1 - \overline{S}_1$), the busy period of the first stage ($\overline{B}^{I}$) and the busy period of the second stage ($\overline{B}^{II}$). We consider three situations of total traffic intensity, $\rho = \rho_1 + \rho_2$, representing heavy traffic ($\rho = 0.9$), normal traffic ($\rho = 0.5$) and light traffic ($\rho = 0.1$). In each case we obtain curves of the pertinent average value as a function of the traffic intensity in the first stage (ρ_1) which varies from zero to ρ (ρ_2 varies from ρ to zero).

Assuming $\lambda = 1$ and using the formulas of Chapters II and III (see Table I) we get

$$Q_o = [\rho^2 - \rho_1(\rho - \rho_1)]/(1 - \rho) \quad .$$

LNB-Policy

$$\overline{W}_1 = (\rho_1 + Q_o)(1 - \rho_1)/(1 - 2\rho_1 + \rho) - \rho_1$$

$$\overline{W}'_2 = [\rho_1^2 + (\rho - \rho_1)(1 + \rho - \rho_1) + Q_o]/(1 - 2\rho_1 + \rho)$$

$$\overline{B}^{I} = \rho_1/(1 - \rho)/A_1(0)$$

$$\overline{B}^{II} = \rho_2/(1 - \rho)/A_1(0)$$

where $A_1(0)$ is obtained as in the example of Section 8 of Chapter II.

SSP-Policy

$$\bar{W}_1 = Q_o$$

$$\bar{W}'_2 = 0$$

$$\bar{B}^I = \rho_1$$

$$\bar{B}^{II} = \rho_2$$

FSP-Policy

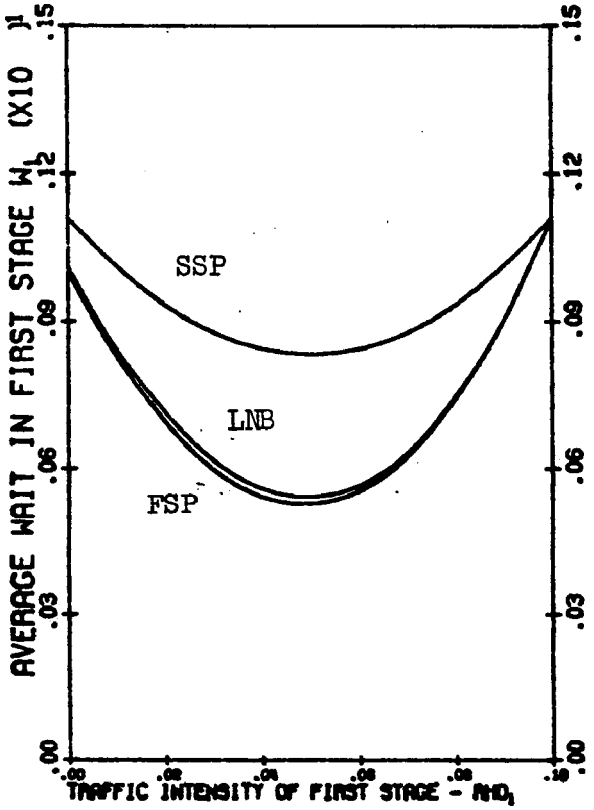
$$\bar{W}_1 = (\rho_1 + Q_o)(1 - \rho)/(1 - \rho_1) - \rho_1$$

$$\bar{W}'_2 = \rho(\rho_1 + Q_o)/(1 - \rho_1)$$

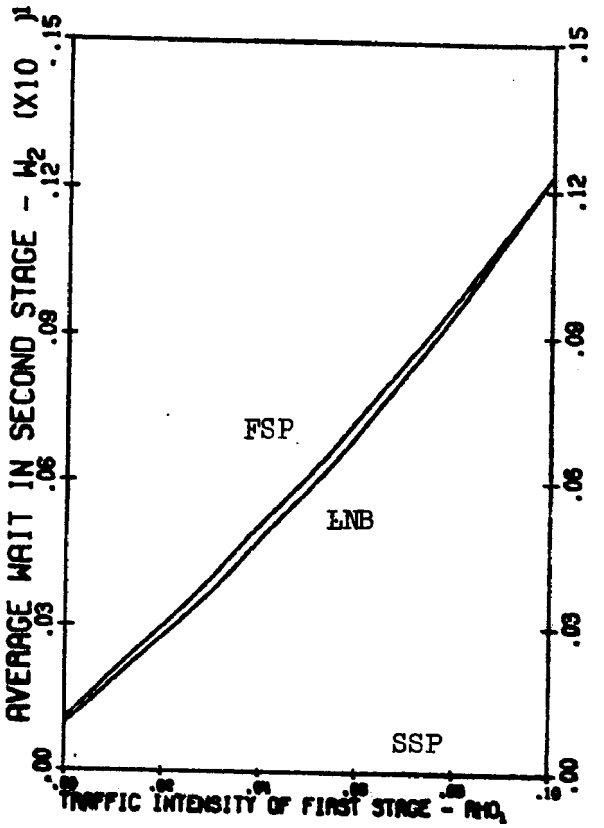
$$\bar{B}^I = \rho_1 /[2 - \rho - (1 + \rho - \rho_1)^{-1}]$$

$$\bar{B}^{II} = \rho_2 /[2 - \rho - (1 + \rho - \rho_1)^{-1}]$$

The quantities above are plotted in Figures II through VII. From these figures we clearly notice that the effect of different policies on the average busy period is negligible for low values of ρ. For all policies, the average wait in the first stage is minimized near $\rho_1 = \rho/2$ and in the case of policies LNB and FSP we see that the average wait in the second stage increases surprisingly fast as the traffic intensity in that stage decreases (ρ_1 increases).

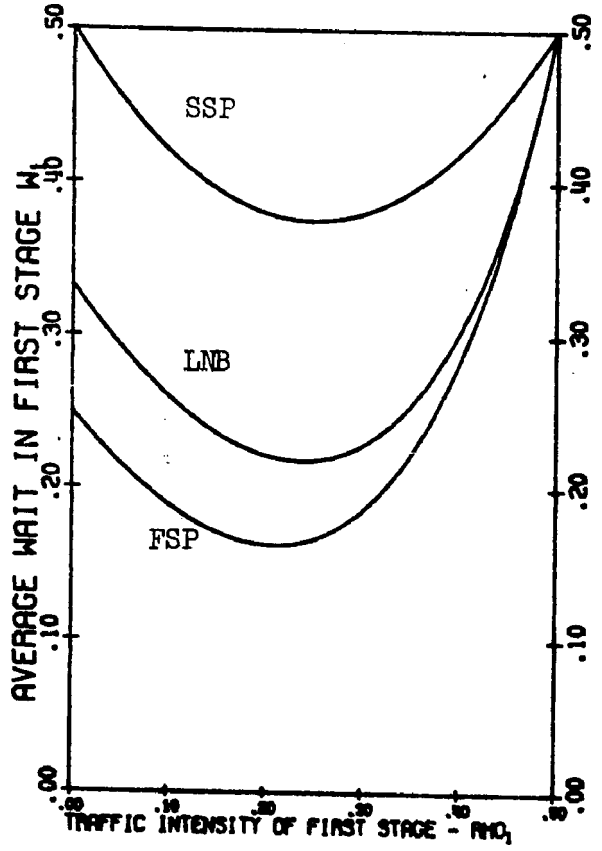


(a)

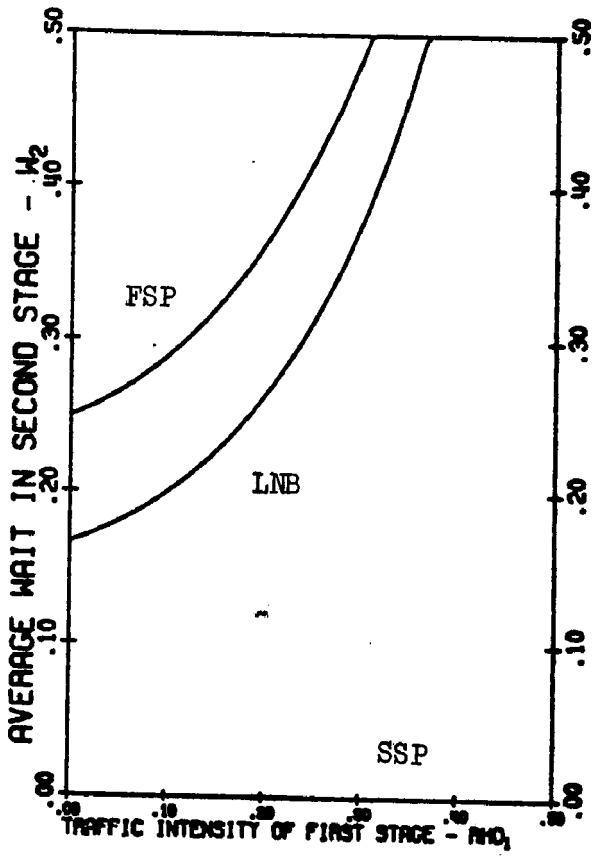


(b)

Figure II. Average Waiting Times in the Stages of the System, For Policies LNB, SSP and FSP, in the Case of Exponential Services, Plotted as a Function of the Traffic Intensity in the First Stage, ρ_1, for Constant Total Traffic Intensity ($\rho = 0.1$). (a) First Stage; (b) Second Stage.

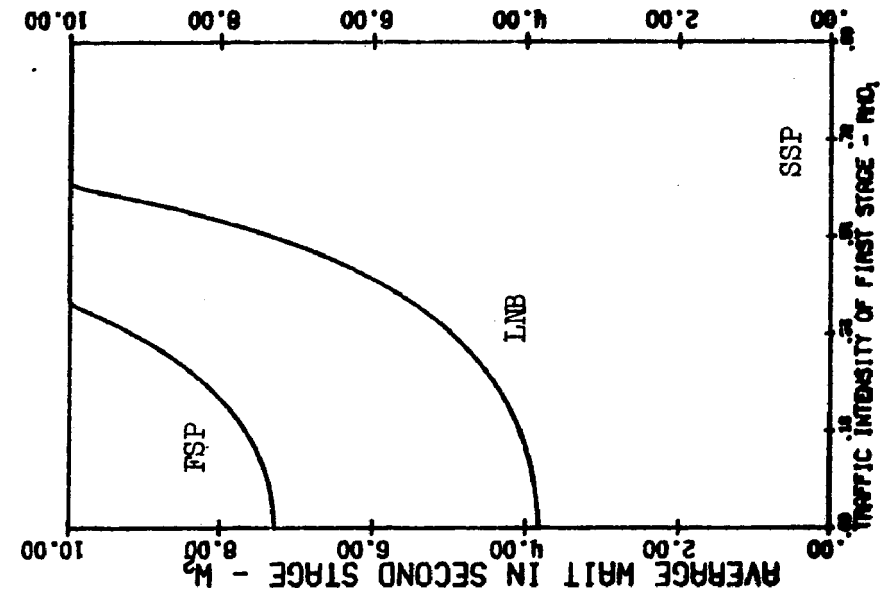


(a)

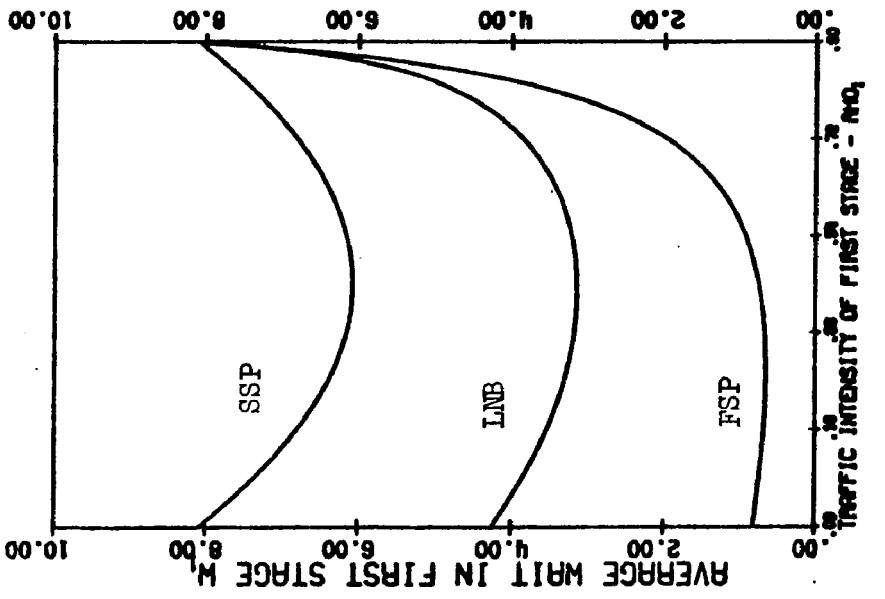


(b)

Figure III. Average Waiting Times in the Stages of the System, for Policies LNB, SSP and FSP, in the Case of Exponential Services Plotted as a Function of the Traffic Intensity in the First Stage, ρ_1, for Constant Total Traffic Intensity ($\rho = 0.5$). (a) First Stage; (b) Second Stage.

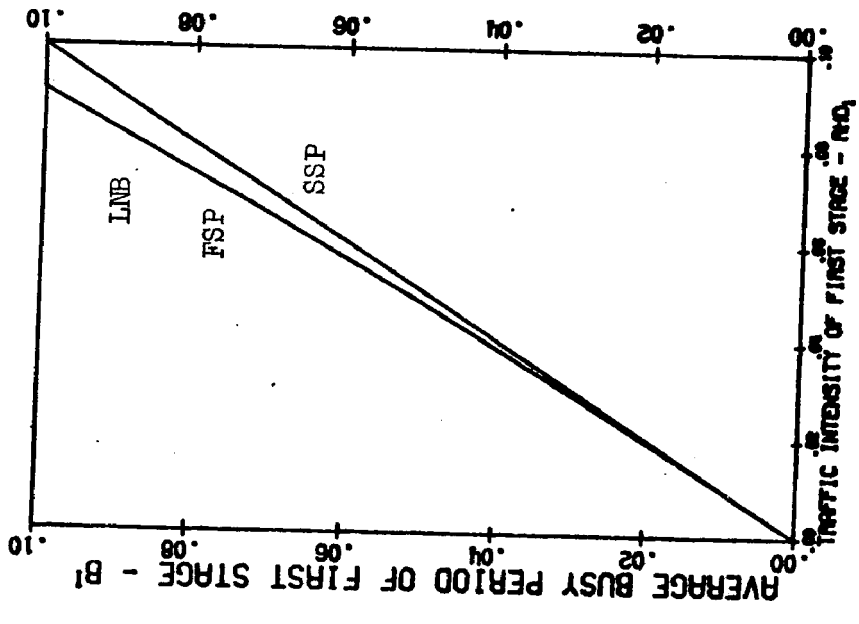


(b)

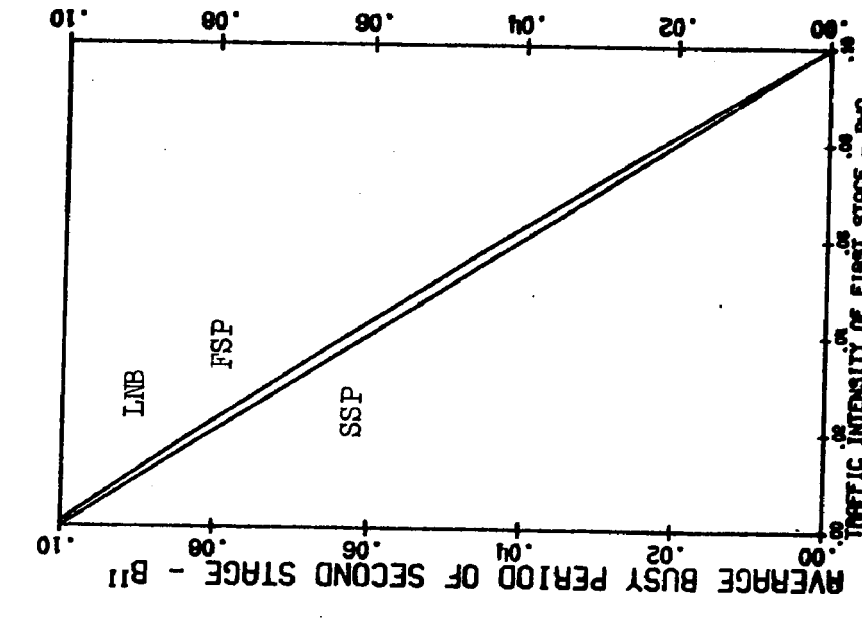


(a)

Figure IV. Average Waiting Times in the Stages of the System, for Policies LNB, SSP and FSP, in the Case of Exponential Services, Plotted as a Function of the Traffic Intensity in the First Stage, ρ_1, for Constant Total Traffic Intensity ($\rho = 0.9$). (a) First Stage; (b) Second Stage.

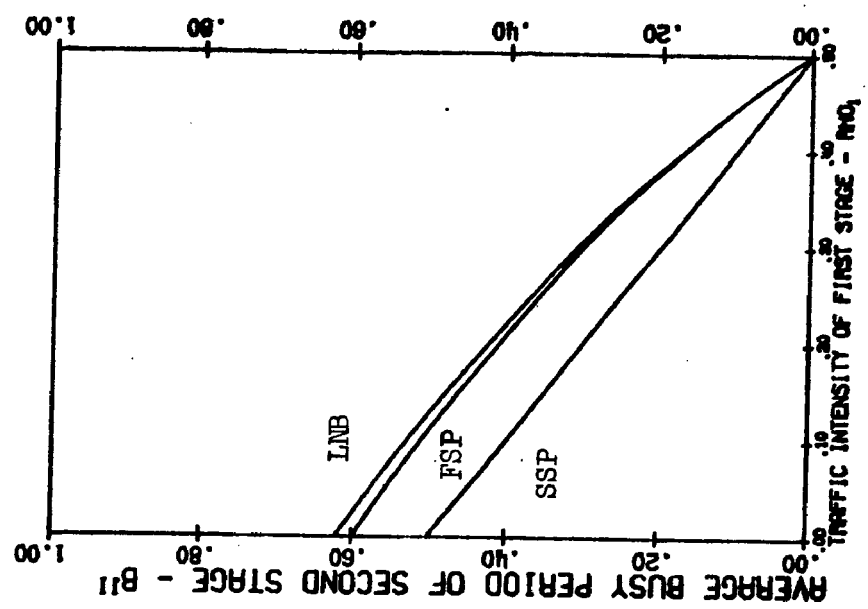


(a)

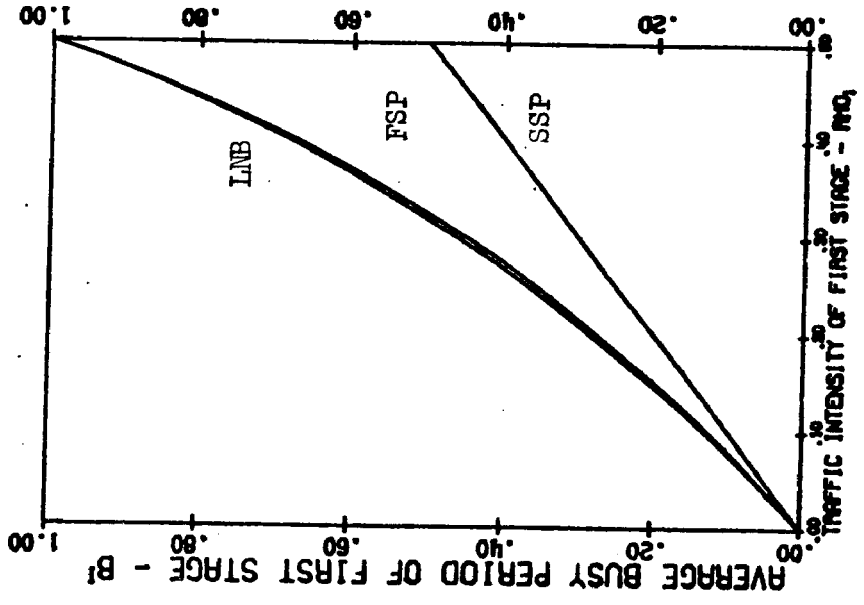


(b)

Figure V. Average Busy Period of the Stages of the System, for Policies LNB, SSP and FSP, in the Case of Exponential Services, Plotted as a Function of the Traffic Intensity to the First Stage, ρ_1, for Constant Total Traffic Intensity ($\rho = 0.1$). (a) First Stage; (b) Second Stage.

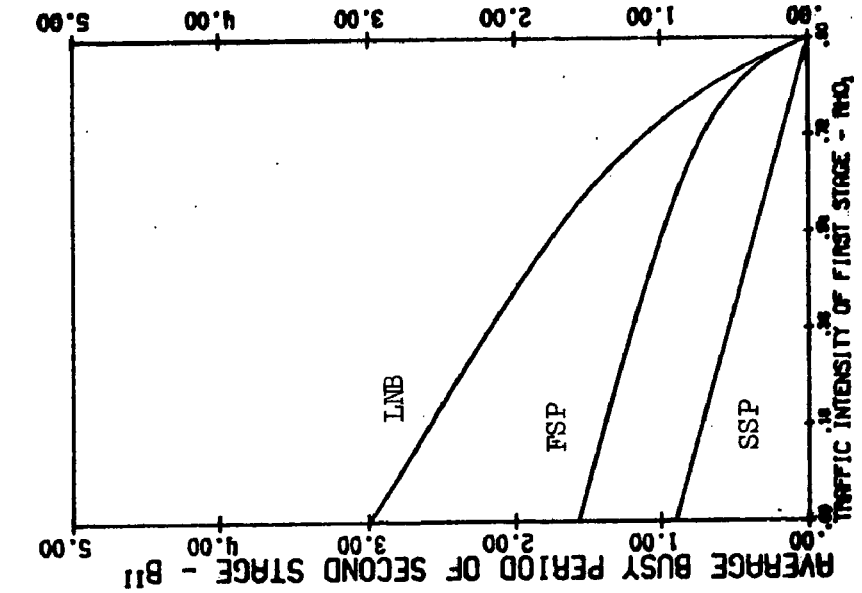


(b)

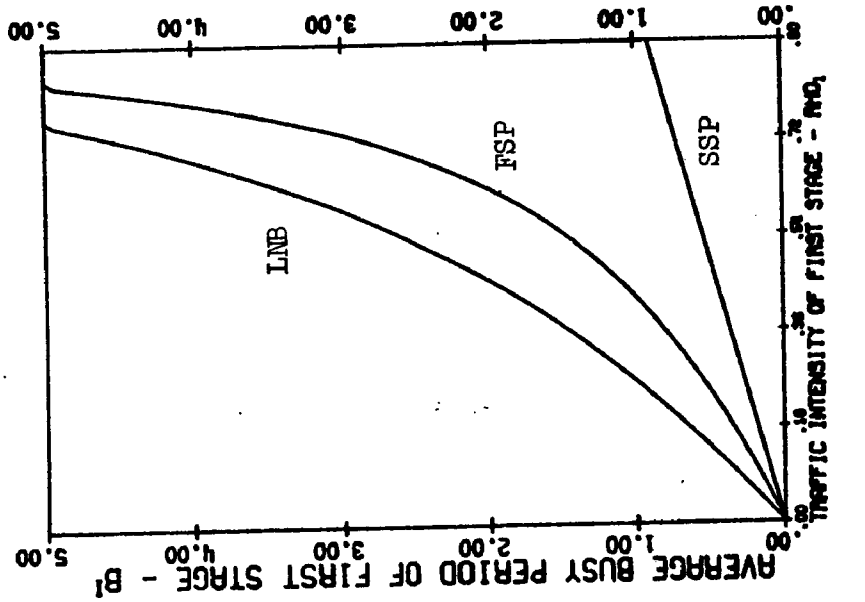


(a)

Figure VI. Average Busy Period of the Stages of the System, for Policies LNB, SSP and FSP, in the Case of Exponential Services, Plotted as a Function of the Traffic Intensity to the First Stage, ρ_1, for Constant Traffic Intensity ($\rho = 0.5$). (a) First Stage; (b) Second Stage.



(b)



(a)

Figure VII. Average Busy Period of the Stages of the System, for Policies LNB, SSP and FSP, in the Case of Exponential Services, Plotted as a Function of the Traffic Intensity to the First Stage, ρ_1, for Constant Traffic Intensity ($\rho = 0.9$). (a) First Stage; (b) Second Stage.

2. Practical Implications of Availability of Numerical Results

For practical employment and engineering consideration the first question that may be asked is: Are the average values numerically available? We shall now have a few remarks on this question and more general aspects.

In the case of policies LNB, SSP and FSP, with instantaneous or with random duration of switching, the average values of all considered queueing properties can be easily calculated by the formulas of Chapters II and III, which do not require calculation of the state probabilities. The determination of the values of $\bar{B}^{I}$ and $\bar{B}^{II}$, in the case of the LNB policy, takes a few more steps since they depend upon the quantity $A_1(0) = U_2(1,0)/\pi(2,0,0) = 2 \sum_{i=0}^{\infty} \pi(2,i,0)/(1 - \rho_1 - \rho_2)$, but, as illustrated in Section 7 of Chapter II, the iterative scheme developed in Section 3 of that chapter is very efficient for finding $A_1(0)$ without explicit calculation of $\pi(2,i,0)$.

For policies W_NFS and SFS ($Q_1 \geq N$) we need first to calculate some of the state probabilities, $\pi(2,i,0)$, for which we provided methods of solution. These methods depend upon the values of the probabilities $b_m(i)$ of having i departures from the first stage during the busy period of that stage started by m customers. These values, in the case of high utilization of the first stage, may be not negligible for large i, therefore the nested summations involved must have large upper limits, which implies a great amount of calculation, as, for example, in the case of Equations 4.14 and 4.15 of Chapter IV, where we have five nested summations. Although a straightforward computer program can be written for these policies, we have not made any actual calculations.

For the SSS ($Q_2 = N$) policy, even the calculations of average values, for which we provided solutions depending on the function $G(.)$, may be expected to encounter numerical difficulties. In this case the use of special computer languages with capabilities of symbolic differentiation might be helpful.

Next we shall comment on the feasibility of the calculation of the distribution of the pertinent random variables: for the state probabilities, except in the case of policy SSS ($Q_2 = N$), we would first find $\pi(2,i,0)$ ($0 \leq i \leq i_T$, where i_T is some truncating value which depends on desired accuracy) for which we provided recursive schemes of solution, then we use the corresponding stationary conditions for calculating the remaining state probabilities. The distributions of the waiting times, $W_1(t)$ and $W_2(t)$, can be found by numerical inversion of their Laplace transforms (an efficient algorithm is given by Stehfest [18]) which are given in terms of the functions $U_k(x,y)$. The distributions $B^I(t)$ and $B^{II}(t)$ can also be found by the same approach, although explicit expressions for these distributions are given in terms of the distribution of the busy period of the M/G/1 system started by i customers, which may take great amounts of calculation in the case of high utilizations. Evidently, in the case where $B^I(t)$ and $B^{II}(t)$ are given by simple expressions like, for example, $B^{II}(t) = S_2^{*N}(t)$ or $B^I(t) = S_1(t)$, we would not use this approach.

As shown in Table I, we were not able to obtain the distribution of the busy period of the server for the LNB and FSP policies for the case of random duration of switching, and for the SSS ($Q_2 = N$) policy.

The difficulties in these cases are due to the fact that during these busy periods the cycles are not homogeneous in the sense that they are formed, possibly, by the sum of distinct random variables. For example, for the LNB policy with random duration of switching, the switching intervals T_{12} are combined with the intervals during which the server is working in each stage. We can say that each customer served during B^S has his average service time increased to $\overline{S}_1 + \overline{S}_2 + S_o$ where S_o represents the average "share" of each customer in the total duration of switching. It is not clear how S_o can be obtained. In the case of the SSP policy, since the server follows the customer through the system, this "share" is simply $\overline{T}_{12} + \overline{T}_{21}$ and, therefore, we can consider that the service time distributions in the first and second stages are $S_1 * T_{12}(t)$ and $S_2 * T_{21}(t)$, respectively. Thus, the busy period B^S has distribution equal to the distribution of the busy period of an M/G/1 queueing system whose service time distribution is the convolution $S_1 * S_2 * T_{12} * T_{21}(t)$.

3. History of the Work

At the beginning of this research we were concerned with two sequential queues with exponential services and interarrival times, where the service rates in each service facility could be controlled, at any time, depending upon the number of customers in each stage of the system. We assumed at most two alternatives for the service rate of each facility (which were allowed to operate simultaneously), each alternative having a corresponding cost rate. Given linear waiting costs and no setup costs, we intended to characterize the optimal

policy of changing the service rates, so as to minimize the average cost rate. Our first step was to consider special cases as, for example, the case of "local" optimization, in which the decision of changing the service rate of one facility depends only on the number of customers in the queue in front of that facility, independently of the number of customers in the other (optimization by stages). The answer to this problem turned out to be trivial, since, according to Reich [15] the departure process of a state dependent M/M/1 queueing system is an independent Poisson process with the same rate of the arrival process. With this result we see that the two stages of the system are stochastically independent, therefore, the optimal policy is as characterized by Crabill [7], in each stage.

Another case is when the service rate in one of the facilities is constant and the service rate in the other one may vary, depending on the number of customers in both stages of the system. We did some numerical experimentation, using Markov decision programming techniques (Howard's and White's algorithms) in the case of finite queue capacities. These indicated that the optimal policy is approximately the same as the optimization by stages. Some theoretical characterizations were attempted mainly because the algebraic structure of this problem seemed to suggest some simplifications (the matrix of rate of transitions is tridiagonal by blocks, each block being sparse and of simple structure). Fitting to this same algebraic structure is the case of an $M/E_k/1$ queueing system where the rate in each phase of the Erlang type of service is controllable. For this problem we were able to

generalize one of Crabill's [7] results of the controllable M/M/1 system. In the case of monotone increasing holding costs, under some restrictions, with two alternatives for service rate, it appears that the optimal policy is characterized by the values ℓ^* and N^* such that, if the number of customers in the system is $n < N^*$, then use the slow service, if $n > N^*$, then use of the fast service, if $n = N^*$, then use of the slow service when the service is in phase $\ell < \ell^*$ and use the fast service when $\ell \geq \ell^*$. However, we do not offer a proof to this effect.

Still another interesting problem, which later led us to the problem of this thesis, is the case where the chosen service rate alternatives at a given time, say μ_1 and μ_2, must be such that $a_1\mu_1 + a_2\mu_2 \leq c$, where a_1 and a_2 are constants and c represents a certain total service capacity.

Analytical results of the cases mentioned above were quite limited and the numerical experimentations did not lead us to any discovery of efficient methods of reducing the cost of the computations to the point of usefulness. However, a simpler form of the problem with limited total service capacity, with the additional restriction of μ_k being equal to c/a_k or zero, $k = 1,2$, opened up new possibilities, since in this case simultaneous service in both stages was not allowed (this is the case of a single server). It was then possible to assume more general service time distributions keeping the analytical considerations at a tractable level. Using departure epochs, this problem is a Markov renewal programming problem and, in the case of finite queue capacities,

the optimal policy can, at least, in principle, be obtained by Howard's algorithm in any given case. However, in the meantime we became more interested in the behavior of special policies that by their simplicity promise to be of practical significance. This thesis has been concerned with such policies.

4. Related Topics and Possible Extensions

Evidently, the six policies studied in this thesis do not constitute all possible "simple" switching policies. We can, for example, imagine the case where the server, upon completing the serving of a customer in one stage, remains in that stage only if the number of customers (Q_k) in it is less than the number of customers (Q_ℓ) in the other stage, otherwise he switches. The "reluctant case" for this policy would be: The server remains in the present stage if $Q_k \geq Q_\ell + N$, otherwise he switches. Other policies could be obtained from the combination of the previously defined policies, for example, policies $SSS(Q_2 = M)$ and $W_N FS$ can be followed simultaneously. These new rules can be analyzed by the methods used in this thesis.

As mentioned in Section 3 of Chapter II, the methods of analysis for parallel queues [Takacs [20] and Eisenberg [8,9]) are similar to the methods of this thesis and, therefore, we may expect that some more general models, results of the combination of parallel and series queues attended by one server, can be analyzed by similar methodology. An interesting problem to consider would be the case where the second stage of the sequential model receives, besides the customers from the first stage, an independent Poisson stream of customers. We could also consider

a parallel system with two service facilities (F_1 and F_2), where a customer departing from one service facility, say F_1, is fed into the queue of facility F_2, provided that the customer has not received service in facility F_2. These two models are special cases of a more general situation where a customer, upon completing service in facility F_k, leaves the system with probability p_k and goes to facility F_ℓ ($\ell \neq k$; $k = 1,2$; $\ell = 1,2$) with probability $1 - p_k$, provided that the customer has not received service in F_ℓ. The dimension of the state space for this model would be quite large, since we have to keep track of all types of customers in the queues and in service. A simple case of this model is when the arrival rate (of customers coming from outside the system) to one of the facilities is zero. This corresponds to the model of this thesis with the additional condition that a customer departing from the first stage leaves the system with probability p_1 and goes to the second stage with probability $1 - p_1$. There should be no difficulties in extending the results of this thesis to this model, for all of the six policies.

An important aspect to be considered is the numerical implementation of the solutions, mainly the ones corresponding to the reluctant policies, in particular policy $SSS(Q_2 = N)$. The assessment of the accuracies in the calculation of the state probabilities is very important since all of the pertinent quantities are expressed in terms of these probabilities. This is a task that we expect to undertake in the near future. At the time this thesis was concluded only a few numerical experimentations were made and no conclusive answers were found. This is why we did not include the case of reluctant policies in the numerical examples of Section 1.

The method of analysis of the policy SSS($Q_2 = N$) seems to be weak. An alternative approach is to eliminate from the original chain the departure epochs of the second stage, except the last departure. However with this approach we would have a weaker description of the behavior of the system.

An intriguing question is whether or not there is any relationship between the iterative scheme of Section 3 of Chapter II and the direct approach shown in Section 7 of the same chapter.

LIST OF REFERENCES

1. M. Abramowitz and I. A. Stegun (editors), Handbook of Mathematical Functions, Dover.

2. B. Avi-itzhak and M. Yadin, "A Sequence of Two Servers with No Intermediate Queue," Management Science, vol. 11, 553-564 (1965).

3. B. Avi-itzhak, W. L. Maxwell and L. W. Miller, "Queueing with Alternating Priorities," Operations Research, vol. 13, 306-318 (1965).

4. P. J. Burke, "The Dependence of Delays in Tandem Queues," Annals of Mathematical Statistics, vol. 35, 874-875 (1964).

5. _____, "The Output Process of a Stationary M/M/s Queueing System," Annals of Mathematical Statistics, vol. 39, 1144-1152, (1968).

6. R. B. Cooper, Introduction to Queueing Theory, The Macmillan Company, New York, 1972.

7. T. B. Crabill, "Optimal Control of a Queue with Variable Service Rate," Tech. Rep. No. 75 (1969), Dept. of Operations Research, Cornell University.

8. M. Eisenberg, "Two Queues with Change Over Times," Operations Research, vol. 19, 386-401 (1971).

9. _____, "Queues with Periodic Service and Changeover Time," Operations Research, vol. 20, 440-451 (1972).

10. R. A. Howard, Dynamic Probabilistic Systems, vol. II, Chapter 15, Wiley (1971).

11. G. C. Hunt, "Sequential Arrays of Waiting Lines," Operations Research, vol. 4, 674-683 (1956).

12. R. T. Nelson, "Labor Assignment as a Dynamic Control Problem," Operations Research, vol. 14, 369-376 (1966).

13. _____, "Dual-Resource Constrained Series Service System," Operations Research, vol. 16, 324-341 (1968).

14. M. F. Neuts, "Two Queues in Series with a Finite, Intermediate Waiting Room," Journal of Applied Probability, vol. 5, 123-142 (1968).

15. E. Reich, "Waiting Times When Queues Are in Tandem," Annals of Mathematical Statistics, vol. 28, 768-773 (1957).

16. _____,"Note on Queues in Tandem," Annals of Mathematical Statistics, vol. 34, 338-341, (1963).

17. V. K. Sahney, "Single-Server Two-Machine Sequencing with Switching Time," Operations Research, vol. 20, 24-36 (1972).

18. H. Stehfest, "Numerical Inversion of Laplace Transforms," Communications of the ACM, vol. 13, No. 1, (1970).

19. J. S. Sykes, "Simplified Analysis of an Alternating-Priority Queueing Model with Setup Times," Operations Research, vol. 18, 1188-1192 (1970).

20. L. Takacs, "Two Queues Attended by a Single Server," Operations Research, vol. 16, 639-650 (1968).

21. _____, Combinatorial Methods in the Theory of Stochastic Processes, Wiley (1967).