MULTI-QUEUES WITH CHANGEOVER TIMES

by

MARTIN EISENBERG

S. B., Massachusetts Institute of Technology
1964

S. M., Massachusetts Institute of Technology
1964

Submitted in partial fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
September, 1967

Signature of Author ______________________
Department of Electrical Engineering, August 28, 1967

Certified by ______________________
Thesis Supervisor

Accepted by ______________________
Chairman, Departmental Committee on Graduate Students

MULTI-QUEUES WITH CHANGEOVER TIMES

by

MARTIN EISENBERG

Submitted to the Department of Electrical Engineering on August 28, 1967 in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

ABSTRACT

The problem of a queuing system with changeover times is studied to determine the effect of the queue discipline. Several specific disciplines for the two-line case are investigated and compared.

The alternating priority and strict priority disciplines are investigated for the general two-line system. A class of disciplines are analyzed for the two-line system with zero changeover times. For alternating priority, the mean waiting times are obtained and it is shown how higher moments may be derived. For each of the other disciplines explicit expressions for the Laplace-Steiltjes transforms of the waiting time distributions are obtained and the means of these distributions are computed. The non-saturation condition and several other measures of performance are found in each case. The technique used throughout is the application of generating functions to the "imbedded" process formed at the instants of service-completion.

In an appendix the mean waiting time of an arbitrary customer is obtained for a specialized K-line system. The method of solution is instructive.

The mean waiting times for various disciplines are compared for a specific system. It is observed that disciplines which increase the idle time do not necessarily decrease the waiting time, and in fact, might cause the waiting time to increase.

It is shown that disciplines which are potentially optimum can be classified in a general way. The specific disciplines studied are found to fall into this classification.

Thesis Supervisor: Alvin W. Drake
Title: Associate Professor of Electrical Engineering

ACKNOWLEDGEMENTS

The author wishes to express his sincere appreciation to Professor Alvin W. Drake for his encouragement, assistance, and concern.

Special thanks are due to Professor John D. C. Little and Professor P. M. Morse who served on the graduate committee and have contributed to the author's education through their suggestions and counsel.

The author also acknowledges the help of Professor Francis F. Lee who suggested the problem which led to this research.

Much gratitude is due the auther's wife, Esther, and child, Diane, whose companionship throughout the period of research served as a continuous inspiration.

This work was supported in part by the M. I. T. Center for Operations Research under D. S. R. Contract DA-31-124-ARO-D-209. All of the electronic computation was performed at the M. I. T. Computation Center.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I

INTRODUCTION

This report is a study of multi-queuing systems with changeover times. Our objective is to determine how the behavior of such systems may be improved by varying the queue discipline.

The basic technique employed is the application of generating functions to a Markov process which is "imbedded" within the actual process. This method was introduced by Kendall who used it to analyze the behavior of a simple queue. We discover its use to be of considerable value for the study of much more complex queuing situations.

I-a The Model

Many situations exist where customers of different types must compete for service at a single facility. Often it happens that before service can be undertaken on a customer of a new type, a certain preparation time-- which we call a changeover time -- is required by the server. An example of this is a machine used for making a variety of parts or a computer used to compile source programs written in a variety of languages.

A model that can be used to characterize such queues when there are just two customer types (1 and 2) is shown in Fig. 1. Customers of each type are assumed to arrive according to independent Poisson laws, λ_1 and λ_2 being the respective average arrival rates.

We shall refer to customers of type 1 and 2 as "1-customers" and "2-customers," respectively. The distribution functions of their service times are

$$F_{S_1}(t) = \text{Prob (service time of 1-customer} \le t)$$

$$F_{S_2}(t) = \text{Prob (service time of 2-customer} \le t).$$

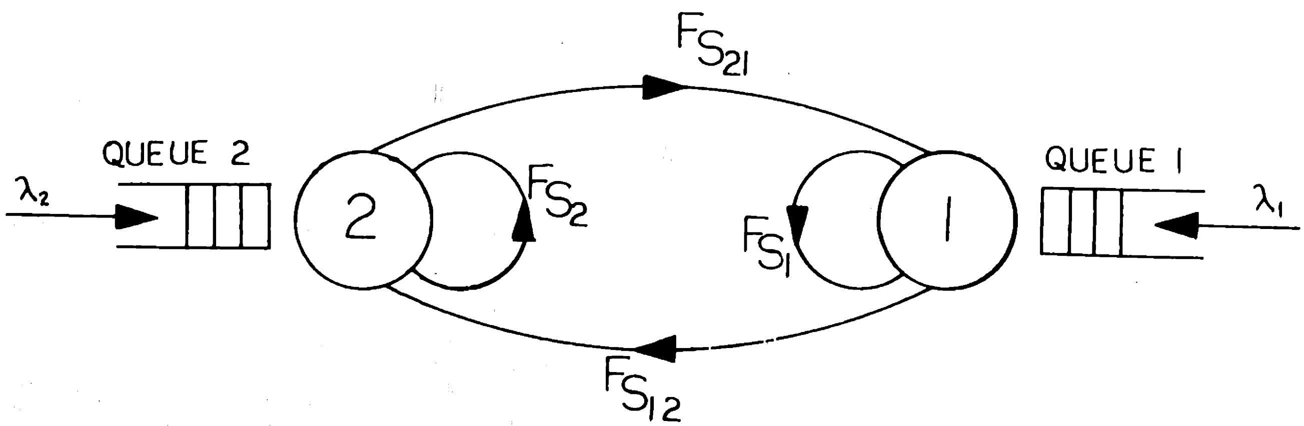
The times required for the server to cross between lines (the changeover times) have distribution functions

$$F_{S_{12}}(t) = \text{Prob (changeover time from 1 to 2} \le t)$$

$$F_{S_{21}}(t) = \text{Prob (changeover time from 2 to 1} \le t)\ .$$

One could visualize a more general model consisting of K lines. In addition to specifying the service time distribution for each of the K types of customers, one would need to specify the distribution of the changeover times between all pairs of lines.

To complete the specification of these systems, one must describe the <u>queue discipline</u> , i. e., the <u>policy</u> that is used by the server to choose the order in which customers are served.



A MODEL OF A QUEUING SYSTEM WITH CHANGEOVER TIMES

FIG. 1

We shall use the words "discipline" and "policy" synonymously throughout the remainder of this thesis.

I-b The Queue Discipline

As the server goes about the task of serving customers, he must from time to time make decisions which govern his actions. The queue discipline is specified by what decisions are made and when the server makes them. If the server is allowed to make decisions at any time the discipline is called "preemptive". In this thesis we shall deal solely with "non-preemptive" disciplines. That is, the server may exercise control over his behavior only at those times when neither a service nor a changeover is in progress. For simplicity we shall assume in addition that the server is not allowed to make a decision at the instant between changeover-completion and service-beginning. One justification for this assumption is that in many situations the changeover time is in reality a part of the service time and therefore cannot be divorced from it. Thus decisions by the server may take place only at instants of service-completion and at times when the server is idle.

Having specified the times at which decisions are made, we must define what the decisions are. For a system with just two lines the choice for the server is simple: either remain stationary (serving customers if any are present, otherwise being idle) or cross to the other line. For a system with K lines, the server may remain stationary or may choose to cross to any of the other K-1 lines.

Each time a decision is made, the server will take into account as much information as possible about the state of the system. For two lines the server will base his decisions on the number of customers in line 1, the number of customers in line 2, and the position of the server. (The server might also wish to take into account the order in which the customers arrived for example as in first-come-first-served. However this information is not relevant to the cost function which we shall use below.) For the cases of K lines, the decisions made by the server will depend upon the number of customers in each of the K lines and the position of the server.

Since we have defined the allowed times for server decisions it follows that the queue discipline may be completely specified by a function having the following properties: The arguments of the function are the number of customers at each line and the current position of the server. The value of the function is a number 1 , 2, ... , K, telling the server which line to "go to" next.

We shall choose a cost function which is proportional to the mean waiting time of an arbitrary customer. That is, we call "optimum" that policy which minimizes the mean waiting time. This choice has obvious drawbacks but it is one convenient measure by which to compare the effectiveness of the various disciplines.

It is desirable to be able to determine the optimum queue discipline for any situation. While we do not achieve this objective here, we do investigate several promising disciplines, and from this develop an intuitive feeling for the selection of an appropriate one.

All of the work in the main body of this thesis is concerned with the case of just two types of customers. The complexity of the analysis in the two-line case makes one view with a certain amount of apprehension the study of systems having more lines. Nevertheless, it is possible that progress in this direction could be made. Using a specialized method, some limited results for a system with K lines are obtained in Appendix VI.

I-c The Imbedded Process

The queues that we are considering are inherently non-Markovian. That is, knowledge of the present state of the system--the number of customers in each line and the position of the server-- would not be sufficient to allow an accurate prediction of future behavior. Only if additional information were incorporated into the definition of "state" would it be possible to accurately assess the probabilities of subsequent events. Such information could be "time spent so far on current service", etc. This expanded state definition would create a Markov process but would also so complicate matters that a simple analysis would become impossible.

It was recognized by Kendall[6] that by taking note of the state of the system only at service-completion instants, one can create a new process which has the Markov property. One can then use the characteristics of this so-called "imbedded" process to obtain information about the actual process.

The technique of working with the imbedded process rather than the actual process is employed throughout this thesis. To gain an understanding of the basic ideas involved, we now briefly outline the use of the method for queuing systems having one line and those having two lines .

THE IMBEDDED PROCESS FOR A QUEUING SYSTEM WITH A SINGLE LINE

We consider a single queue with Poisson arrivals (average rate = λ) and general service. The distribution function for service is $F_S(t)$ = Prob (service time $\leq$ t).

The state of the imbedded process, formed at the instants of service completions, is n, the number of customers in line.

$P(n \to n')$ is our notation for the probability that the next state will be n', given the current state is n. The calculation of these probabilities is straightforward and we omit it here. We shall have many occasions to carry out such calculations in following chapters. We assume that an equilibrium exists and let π_n be the steady-state probability that the imbedded process is in state n. By definition, the equations

$$\pi_{n'} = \sum_{n=o}^{\infty} \pi_n P(n \to n') \qquad \text{(I-1)}$$

and

$$\sum_{n=o}^{\infty} \pi_n = 1 \qquad \text{(I-2)}$$

must be satisfied. By multiplying Eq. (I-1) by $z^{n'}$ and summing over all n', and then making use of Eq. (I-2), one obtains an expression for $\pi(z)$, the generating function of the imbedded probabilities.

$$\pi(z) \equiv \sum_{n=o}^{\infty} \pi_n z^n \qquad \text{(I-3)}$$

The usual quantities of interest are the distributions of the waiting time and number-in-line (number of customers present). We now show that both distributions may be obtained from the imbedded probabilities.

Waiting Time -- Let the waiting time of a customer be defined as the length of time the customer must wait in the queue before he is taken into service. (The time a customer spends in service is not included in our definition of the waiting time.) The distribution function for the waiting time is denoted by

$$F_W(t) = \text{Prob (waiting time} \le t).$$

We observe that if the queue discipline is first-come, first-served (FCFS), the number of customers remaining when a customer is discharged is the number which arrived during his waiting time and service. Since $\frac{(\lambda t)^n}{n!} e^{-\lambda t}$ is the probability that n customers arrive in a time t, we have

$$\pi_n = \int_0^\infty \frac{(\lambda t)}{n!} e^{-\lambda t} \, dF_W \otimes F_S(t) \qquad \text{(I-4)}$$

where $dF_W \otimes F_S(t)$ is used to signify the probability that the waiting time plus the service time is between t and t +dt. Using generating functions, Eq. (1-4) becomes

$$W(s) = \frac{\pi \, (1-\frac{s}{\lambda})}{S(s)} \qquad \text{(I-5)}$$

where W(s) and S(s) are the Laplace-Steiltjes transforms of $F_W(t)$ and $F_S(t)$, respectively.

$$W(s) = \int_0^\infty e^{-st} \, dF_W(t) \qquad \text{(I-6)}$$

$$S(s) = \int_0^\infty e^{-st} \, dF_S(t) \qquad \text{(I-7)}$$

Eq. (I-5) is what we set out to demonstrate. The transform of the waiting time distribution can be expressed explicitly in terms of $\pi(z)$, the generating function of the imbedded probabilities.

Number-in-Line - We would like to find p_n, the general-time probability that the state of the system is n (the fraction of time the actual process spends in state n). In generating function form we have found π_n, the probability that a service-completion leaves the system in state n. In fact, since the arrivals are Poisson, the imbedded probabilities and the general-time probabilities are equal

$$\pi_n = p_n \qquad \text{(I-8)}$$

which we now show.

Consider a particular realization of the actual process (a plot of the number of customers in the system versus t). This is a function that takes on integer values and undergoes occasional unit jumps, either upward or downward. The upward jumps are caused by arrivals and the downward jumps are caused by service-completions. We concentrate on this function for a long interval of time T. In T there are i_n downward jumps from n + 1 to n. The total number of downward jumps (service-completions) in T is $i = \sum_{n=o}^{\infty} i_n$. Thus

$$\pi_n = \lim_{T \to \infty} \frac{i_n}{i} \qquad \text{(I-9)}$$

since π_n is the fraction of service-completions that leave the system in state n.

In T there are j_n upward jumps from n to n+1. The total number of upward jumps (arrivals) in T is $j = \sum_{n=o}^{\infty} j_n$. The quantities i_n and j_n can differ by no more than one. (If a horizontal line is drawn through the function at a height n+1/2, the number of times this line is crossed from above must differ by no more than one from the number of times it is crossed from below.) As long as the system is unsaturated the ratio of i to j (number of departures to number of arrivals) must go to one as T goes to infinity. Hence it follows that

$$\lim_{T \to \infty} \frac{i_n}{i} = \lim_{T \to \infty} \frac{j_n}{j} \qquad \text{(I-10)}$$

The quantity on the right-hand side of this equation is the probability that the system is in state n at the instant of an arrival. Since arrivals occur at the points of a Poisson process operating independently of the state of the process, this probability is identical with the general-time probability p_n*. Thus Eq. (I-8) is confirmed.

Using the above results it is possible to obtain an interesting relationship. The average waiting time, $\bar{W}$ is minus the derivative of W(s) evaluated at s = 0. Using Eq. (I-5) one has

$$\bar{W} = \left.\frac{d\pi(z)}{dz}\right|_{z=1} \frac{1}{\lambda} - \pi(1)\frac{1}{\mu} \qquad \text{(I-11)}$$

where

$$\frac{1}{\mu} = \text{the average service time}$$

From Eq. (I-2) it follows that $\pi(1) = 1$. From Eq. (I-8)

$$\left.\frac{d\pi(z)}{dz}\right|_{z=1} = \sum_{n=0}^{\infty} n\pi_n = \sum_{n=0}^{\infty} np_n \qquad \text{(I-12)}$$

is recognized as $\bar{n}$, the average number of customers in line.

Thus we have

$$\bar{n} = \lambda\,(\bar{W} + \frac{1}{\mu}) \qquad \text{(I-13)}$$

---

* See, for example, Cox and Miller (p. 269)

the well-known relation between the average waiting time and the average number-in-system. (This equation is in a slightly different form than the familiar $\bar{n} = \lambda \bar{W}$ since we have defined $\bar{W}$ to be the average wait in the queue, not including service, while $\bar{n}$ is the average number in line, including service.)

THE IMBEDDED PROCESS FOR A QUEUING SYSTEM WITH TWO LINES

In this section we examine the imbedded process for a queuing system with two lines. The quantities of interest are the distributions of the waiting time for each type of customer and the distributions of the number-in-line of each type of customer. We shall find that, corresponding to the single-line case, these distributions may be related to the imbedded probabilities. In addition, we shall be able to relate the distribution of the total number of customers present to the imbedded probabilities.

Consider a queuing system having two lines with Poisson arrivals (rates λ_1 and λ_2) and general service. The distribution functions for service are

$$F_{S_1}(t) = \text{Prob (service time of 1-customer} \leq t)$$

$$F_{S_2}(t) = \text{Prob (service time of 2-customer} \leq t)$$

There may or not be changeover times. Their existence is not of importance to this discussion. The states of the imbedded process, formed at the instants of service-completion are

$\underset{m\ n}{1}$ - server is at line 1, m 1-customers and n 2-customers are present.

$\underset{m\ n}{2}$ - server is at line 2, m 1-customers and n 2-customers are present.

We let the steady-state probability (assuming it exists) that the imbedded process is in state $\underset{m\ n}{1}$ or $\underset{m\ n}{2}$ be $\pi_{m\ n}^{1}$ or $\pi_{m\ n}^{2}$, respectively. $\pi_{m\ n}^{1}$ and $\pi_{m\ n}^{2}$ satisfy a set of equilibrium equations corresponding to Eq. (I-1). (See Eqs. (II-3).) In a manner similar to what was done in the single-line case, expressions for the generating functions of the imbedded probabilities may be obtained

$$\pi^1(y,z) = \sum_{m=o}^{\infty} \sum_{n=o}^{\infty} \pi^1_{mn}\, y^m z^n \qquad \text{(I-14a)}$$

$$\pi^2(y,z) = \sum_{m=o}^{\infty} \sum_{n=0}^{\infty} \pi^2_{mn}\, y^m z^n \qquad \text{(I-14b)}$$

We now demonstrate the usefulness of these functions for finding the statistics of interest. The present discussion applies in general and does not depend upon the queue discipline or the particular form of $\pi^1(y, z)$ and $\pi^2(y, z)$.

Waiting Times -- We assume that the queue discipline in a given line is FCFS, regardless of the method the server uses to choose between lines . The waiting times have the distribution functions

$$F_{W_1}(t) = \text{Prob (waiting time of 1-customer} \leq t)$$

$$F_{W_2}(t) = \text{Prob (waiting time of 2-customer} \leq t)$$

where, as before, the service period is not included in the definition.

At the instant of service completion, with probability

$$r_1 \equiv \frac{\lambda_1}{\lambda_1+\lambda_2}$$

the customer just discharged was of type 1 and with probability

$$r_2 \equiv \frac{\lambda_2}{\lambda_1+\lambda_2}$$

he was of type 2. If he was of type 1, then the number of 1-customers remaining is the number that arrived during his waiting time and service period. Hence,

$$\sum_{n=0}^{\infty} \pi^1_{mn} = r_1 \int_0^{\infty} \frac{(\lambda_1 t)^m}{m!} e^{-\lambda_1 t} \, dF_{W_1} \otimes F_{S_1}(t) \qquad \text{(I-15)}$$

where $dF_{W_1} \otimes F_{S_1}(t)$ denotes the probability that the waiting time plus the service period for a 1-customer is between t and t + dt. Using generating functions Eq. (I-15) becomes

$$W_1(s) = \frac{\pi^1 (1 - \frac{s}{\lambda_1}, 1)}{r_1 S_1(s)} \qquad \text{(I-16a)}$$

where $W_1(s)$ and $S_1(s)$ are the Laplace - Steiltjes transforms of $F_{W_1}(t)$ and $F_{S_1}(t)$ respectively.

$$W_1(s) = \int_0^{\infty} e^{-st} dF_{W_1}(t) \qquad \text{(I-17a)}$$

$$S_1(s) = \int_0^{\infty} e^{-st} dF_{S_1}(t) \qquad \text{(I-18a)}$$

We can similarly show

$$W_2(s) = \frac{\pi^2(1, 1 - \frac{s}{\lambda_2})}{r_2 S_2(s)} \qquad \text{(I-16b)}$$

where

$$W_2(s) = \int_0^{\infty} e^{-st} \, dF_{W_2}(t) \qquad \text{(I-17b)}$$

$$S_2(s) = \int_0^{\infty} e^{-st} \, dF_{S_2}(t) . \qquad \text{(I-18b)}$$

Eqs. (I-16) give the relationship between the waiting times and the imbedded probabilities that we were seeking.

Numbers-in-Line -- One wonders whether the general-time probability of the state $\overset{1}{m\;n}$ is equal to the imbedded probability $\pi\overset{1}{m\;n}$. In fact these probabilities are not equal. In order to demonstrate their equality, one would need to show that the imbedded probabilities at service-completion instants are equal to those at arrival instants.

In the single-line case, the state space had but one dimension and it was impossible to avoid entering a state due to a service-completion the same number of times it was left due to an arrival (which is the statement that i_n and j_n, defined earlier, differ by at most one). In the present case, the state space has two dimensions. Now it is possible for the system to "circle around" a particular state. State $\overset{1}{m\;n}$, for example , may be entered many times due to a service-completion without ever being left due to arrival and vice versa. We therefore cannot equate the imbedded probability $\pi\overset{1}{m\;n}$ to a corresponding general-time probability.

Let us, however, consider only m, the number of 1-customers in line. This variable increases with 1-arrivals and decreases with 1-services. Since m has but one dimension, we may argue exactly as we did for the single-line case to show the equality between the imbedded probabilities at 1-service-completion instants and those at 1-arrival instants. Since arrivals of type 1 are Poisson, the imbedded probabilities and the general-time probabilities are equal. Hence,

$$\text{Prob (m 1-customers present)} = \frac{1}{r_1} \sum_{n=o}^{\infty} \pi^1_{mn} \tag{I-19a}$$

(general-time probability)

where the expression on the right-hand side is the probability m 1-customers are present at a service-completion conditional on the service being of type 1. Likewise

$$\text{Prob (n 2-customers present)} = \frac{1}{r_2} \sum_{m=o}^{\infty} \pi^2_{mn} \tag{I-19b}$$

(general-time probability)

Eqs. (I-19) give the desired relationship between the distributions of numbers-in-line and the imbedded probabilities.

We see that when two types of customers exist, the imbedded process can be very useful, as it was for the single-line case. In either case the waiting time and number-in-line statistics may be found from the imbedded probabilities. The corresponding equations are Eq. (I-5) and Eqs. (I-16) for the waiting times and Eq. (I-8) and Eqs. (I-19) for the numbers-in-line. Note that a result more general than Eqs. (I-19) would have been the joint distribution of the numbers in each line (m and n). However, from an earlier discussion, we saw that this could not be obtained. Nevertheless, we shall determine below the distribution for the total number of customers in the system (m+n).

Before doing this, let us find the relationship between the average waiting times and numbers-in-line corresponding to Eq. (I-13). Taking the derivatives of Eqs. (I-16) one obtains

$$\overline{W}_1 = \frac{1}{r_1} \frac{d}{dy} \pi^1 (y, 1) \Big|_{y=1} \frac{1}{\lambda_1} - \frac{1}{\mu_1} \tag{I-20a}$$

$$\overline{W}_2 = \frac{1}{r_2} \frac{d}{dz} \pi^2 (1, z) \Big|_{z=1} \frac{1}{\lambda_2} - \frac{1}{\mu_2} \tag{I-20b}$$

where

$1/\mu_1$ = mean service time of 1-customer

$1/\mu_2$ = mean service time of 2-customer

In deriving Eqs. (I-20) we have used the fact that $\pi^1(1, 1) = r_1$ and $\pi^2(1, 1) = r_2$, since the fraction of 1-customers and 2-customers is r_1 and r_2 respectively. From Eqs. (I-19) we see that

$$\frac{1}{r_1} \frac{d}{dy} \pi^1(y, 1) \Big|_{y=1} = \overline{m} \tag{I-21a}$$

$$\frac{1}{r_2} \frac{d}{dz} \pi^2(1, z) \Big|_{z=1} = \overline{n} \tag{I-21b}$$

the average numbers of customers in lines 1 and 2, so that Eqs. (I-20) become

$$\bar{m} = \lambda_1 (\bar{W}_1 + \frac{1}{\mu_1}) \qquad (I\text{-}22a)$$

$$\bar{n} = \lambda_2 (\bar{W}_2 + \frac{1}{\mu_2}) \qquad (I\text{-}22b)$$

as expected.

We now find one further use for the imbedded probabilities π^1_{mn} and π^2_{mn}. The total number of customers present $k = m+n$ (number-in-system) is a one-dimensional variable that increases when arrivals occur and decreases when services occur. Thus for this variable the imbedded probabilities at service-completion instants and those at arrival instants are identical. This follows from the same reasoning as before. Since arrivals are Poisson, the imbedded probabilities and the general-time probabilities are equal. We can therefore write

$$\text{Prob (system empty)} = \pi^1_{oo} + \pi^2_{oo} \qquad (I\text{-}23a)$$

$$\text{Prob(k customers in system)} = \sum_{m}\sum_{\substack{n \\ m+n=k}} (\pi^1_{mn} + \pi^2_{mn}) \qquad (I\text{-}23b)$$

(general time probabilities)

Eq. (I-23a) is of particular value and we shall have several occasions to make use of it in succeeding chapters.

I-d Work of Other Authors

Gaver -- As previously noted, the queuing process of Fig. 1 is non-Markovian if the state description is specified only by the position of the server and the number in each line. When the queue discipline is strictly FCFS (and only in this case) a new definition of state is possible which has the Markov property. This is the virtual waiting time introduced by Takács[12]. Using this concept, Gaver[4] was able to find the non-saturation condition and mean waiting times for the two-line system of Fig. 1 under the FCFS discipline. The results are

Non-Saturation Condition:

$$1-\rho_1 - \rho_2 - \frac{\lambda_1\lambda_2}{\lambda_1+\lambda_2}\left(\frac{1}{\mu_{12}} + \frac{1}{\mu_{21}}\right) > 0 \qquad \text{(I-24)}$$

where

$1/\mu_{12}$ = mean time of a 1 to 2 changeover

$1/\mu_{21}$ = mean time of a 2 to 1 changeover

$\rho_1 = \dfrac{\lambda_1}{\mu_1}$

$\rho_2 = \dfrac{\lambda_2}{\mu_2}$

Mean Waiting Times:

$$\overline{W}_1 = \overline{W}_o + \frac{r_2}{\mu_{21}} \qquad \text{(I-25a)}$$

$$\overline{W}_2 = \overline{W}_o + \frac{r_1}{\mu_{12}} \qquad \text{(I-25b)}$$

$\overline{W}_o$ = the average time to clear the system. The expression is complicated and relatively uninformative. Therefore it is not included here.

Gaver also obtained results for the case of a preemptive strict priority discipline. Since preemption is prohibited in our work, we do not repeat the results here.

R. G. Miller -- Using the method of the imbedded process, Miller[10] was able to obtain expressions for the mean waiting times in the case of two queues with no changeover times under a strict priority discipline. The absence of changeover times allowed the position of the server to be omitted from the specification of state. The results are

Non-Saturation Condition:

$$1 - \rho_1 - \rho_2 > 0 \qquad \text{(I-26)}$$

Mean Waiting Times:

$$\overline{W}_1 = \frac{\lambda_1 E(S_1^2) + \lambda_2 E(S_2^2)}{2(1-\rho_1)} \qquad \text{(I-27a)}$$

$$\overline{W}_2 = \frac{\lambda_1 E(S_1^2) + \lambda_2 E(S_2^2)}{2(1-\rho_1)(1-\rho_1-\rho_2)} \qquad \text{(I-27b)}$$

where:

$E(S_1^2)$ = second moment of the service time for a 1-customer

$$= \int_0^\infty t^2 \, dF_{S_1}(t)$$

$E(S_2^2)$ = second moment of the service time for a 2-customer

$$= \int_0^\infty t^2 \, dF_{S_2}(t)$$

Avi-Itzhak, Maxwell, and L. W. Miller -- The alternating priority discipline (which we discuss in Chapter II) was investigated for the two-line case with no changeover times. Through a specialized mathematical technique involving conditional expectations, the following results were obtained[1]

Non-Saturation Condition:

$$1 - \rho_1 - \rho_2 > 0 \qquad \text{(I-28)}$$

Mean Waiting Times:

$$\overline{W}_1 = \frac{[(1-\rho_1)(1-\rho_2)^2 + \rho_2^2(1+\rho_1)] \lambda_1 E(S_1^2) + (1-\rho_1)\lambda_2 E(S_2^2)}{2(1-\rho_1-\rho_2)(1-\rho_1-\rho_2+2\rho_1\rho_2)} \qquad \text{(I-29)}$$

with a similar expression for $\overline{W}_2$.

L. W. Miller -- In his Ph. D. thesis, Miller[9] examined the use of Alternating Priority when changeover times are present. He determined the non-saturation condition to be

$$1- \rho_1 - \rho_2 > 0 \qquad (I-30)$$

independent of the changeover time distribution. He also obtained an expression for the mean waiting time which is left in terms of infinite summations. The assumptions of his model pertaining to changeover times are slightly different than those we make so that our results are not directly comparable.

I-2 Preview of Work

In the next three chapters, we study the system of Fig. 1 under some specific queue disciplines. In Chapter II we consider the alternating priority discipline and in Chapter III we consider strict priority. With the help of a simplifying assumption, we are able to treat a wide class of disciplines in Chapter IV. Then in Chapter V we compare the various disciplines and attempt to assess their relative advantages.

(A fairly complete list of notation is given in Appendix I. The reader may find it helpful to review this before continuing.)

CHAPTER II

ALTERNATING PRIORITY

The alternating priority rule for the server is: 1) empty the current line before moving, 2) if both lines are empty, remain stationary. This discipline was first discussed by Avi-Itzhak, Maxwell and Miller[1] and is so named because the line which is receiving service has priority for the moment.

We choose this discipline for study because it is simple and because it has the desirable property that the amount of switching by the server is kept small. Although it is surely not the best policy to employ in every situation, there are occasions when its performance is superior to that of other simple disciplines, and it may well be the optimum discipline to use in some circumstances. The queuing systems that stand to benefit the most from its use are those having large changeover times relative to service times, and those where the service times of each customer type are nearly equal.

Our primary objective in this chapter is to obtain expressions for the mean and distribution of the waiting time of each type of customer. We shall use the method of the imbedded process outlined in Chapter I. We saw there that the key to the method was the determination of the generating function of the state probabilities. Once this is known, the waiting times may easily be found.

We first write the transition probabilities for the imbedded process and from them we obtain expressions for the generating functions of the state probabilities. We note in these expressions the appearance of undetermined boundary conditions. Careful inspection of the generating functions, however, reveals the necessity for the boundary conditions to satisfy a pair of simultaneous functional equations. Although the functional equations can not be solved in general, enough information can be extracted from them to enable us to obtain the moments of the waiting times. As a by-product, we also obtain several other quantities which are of value in judging the performance of the system.

II-a The Imbedded Process

We consider the imbedded process formed by observing the system at the instants of service completion. The state of the system is $\overset{1}{m\ n}$ (or $\overset{2}{m\ n}$) , which denotes "server is at line 1 (or 2), m customers are waiting at line 1, and n customers are waiting at line 2." The transition probabilities for the process may be expressed in terms of p_{ij}, q_{ij}, u_{ij}, and v_{ij}, which are defined below.

p_{ij} = Prob (i 1-customers and j 2-customers arrive during the service time of a 1-customer)

$$= \int_0^\infty \frac{(\lambda_1 t)^i}{i!} \frac{(\lambda_2 t)^j}{j!} e^{-(\lambda_1+\lambda_2) t} \, dF_{S_1}(t) \qquad \text{(II-1a)}$$

This can be explained as follows. The probability that i customers of type 1 arrive in a interval of duration t is $\frac{(\lambda_1 t)^i}{i!} e^{-\lambda_1 t}$

The probability that j customers of type 2 arrive in this interval is $\frac{(\lambda_2 t)^j}{j!} e^{-\lambda_2 t}$. $dF_{S_1}(t)$ is the probability that the service time of a 1-customer lasts a time between t and t +dt. Thus the product of these probabilities, integrated over all t, produces the desired result for p_{ij}. Likewise

q_{ij} = Prob (i 1-customers and j 2-customers arrive during the service time of a 2-customer)

$$= \int_0^\infty \frac{(\lambda_1 t)^i}{i!} \frac{(\lambda_2 t)^j}{j!} e^{-(\lambda_1+\lambda_2)t} \, dF_{S_2}(t) \qquad \text{(II-1b)}$$

u_{ij} = Prob (i 1-customers and j 2-customers arrive during a changeover from 2 to 1)

$$= \int_0^\infty \frac{(\lambda_1 t)^i}{i!} \frac{(\lambda_2 t)^j}{j!} e^{-(\lambda_1+\lambda_2)t} \, dF_{S_{21}}(t) \qquad \text{(II-1c)}$$

v_{ij} = Prob (i 1-customers and j 2-customers arrive during a changeover from 1 to 2)

$$= \int_0^\infty \frac{(\lambda_1 t)^i}{i!} \frac{(\lambda_2 t)^j}{j!} e^{-(\lambda_1+\lambda_2)t} \, dF_{S_{12}}(t) \qquad \text{(II-1d)}$$

The generating functions for the above quantities are

$$P(y,z) = \sum_{i=o}^{\infty} \sum_{j=o}^{\infty} p_{ij}\, y^i z^j = \sum_{i=o}^{\infty} \sum_{j=o}^{\infty} \int_o^{\infty} \frac{(\lambda_1 t y)^i}{i!} \frac{(\lambda_2 t z)^j}{j!} e^{-(\lambda_1+\lambda_2)t} dF_{S_1}(t)$$

$$= \int_o^{\infty} e^{-(\lambda_1-\lambda_1 y + \lambda_2-\lambda_2 z)t} dF_{S_1}(t)$$

$$= S_1(\lambda_1-\lambda_1 y + \lambda_2-\lambda_2 z) \qquad \text{(II-2a)}$$

where we recall that $S_1(s)$ is the Laplace-Steiltjes transform of a type-1 service time. And

$$Q(y,z) = \sum_{i=o}^{\infty} \sum_{j=o}^{\infty} q_{ij}\, y^i z^j = S_2(\lambda_1-\lambda_1 y + \lambda_2 - \lambda_2 z) \qquad \text{(II-2b)}$$

$$U(y,z) = \sum_{i=o}^{\infty} \sum_{j=o}^{\infty} u_{ij} y^i z^j = S_{21}(\lambda_1-\lambda_2 y+\lambda_2-\lambda_2 z) \qquad \text{(II-2c)}$$

$$V(y,z) = \sum_{i=o}^{\infty} \sum_{j=o}^{\infty} v_{ij} y^i z^j = S_{12}(\lambda_1-\lambda_1 y + \lambda_2-\lambda_2 z) \qquad \text{(II-2d)}$$

We denote the transition probabilities from state $\genfrac{}{}{0pt}{}{1}{m\ \ n}$ to state $\genfrac{}{}{0pt}{}{1}{m'\ n'}$ as $P\left(\genfrac{}{}{0pt}{}{1}{m\ \ n} \rightarrow \genfrac{}{}{0pt}{}{1}{m'\ n'}\right)$ and the other transition probabilities correspondingly. We have for alternating priority

1 → 1 TRANSITIONS

m≥1, all n

$$P(\overset{1}{mn} \to \overset{1}{m'n'}) = \begin{cases} 0 & , m' < m-1 \\ 0 & , n' < n \\ p_{m'-m+1,\, n'-n}, & \text{otherwise} \end{cases}$$

m=0, n≥1

$$P(\overset{1}{on} \to \overset{1}{m'n'}) = 0$$

m=o, n=o

$$P(\overset{1}{oo} \to \overset{1}{m'n'}) = r_1\, p_{m'n'}$$

1 → 2 TRANSITIONS

m≥1, all n

$$P(\overset{1}{mn} \to \overset{2}{m'n'}) = 0$$

m=o, n≥1

$$P(\overset{1}{on} \to \overset{2}{m'n'}) = \begin{cases} 0 & , n' < n-1 \\ \sum_{i=o}^{m'} \sum_{j=o}^{n'-n+1} v_{ij}\, q_{m'-i,\, n'-n+i-j}, & \text{otherwise} \end{cases}$$

m=o, n=o

$$P(\overset{1}{oo} \to \overset{2}{m'n'}) = r_2 \sum_{i=o}^{m'} \sum_{j=o}^{n'} v_{ij}\, q_{m'-i,\, n'-j}$$

2 → 2 TRANSITIONS

n≥1, all m

$$P(\overset{2}{mn} \rightarrow \overset{2}{m'n'}) = \begin{cases} 0 & , \quad n' < n-1 \\ 0 & , \quad m' < m \\ q_{m',n'-n+1} & , \quad \text{otherwise} \end{cases}$$

n=o, m≥1

$$P(\overset{2}{mo} \rightarrow \overset{2}{m'n'}) = 0$$

n=o, m=o

$$P(\overset{2}{oo} \rightarrow \overset{2}{m'n'}) = r_2 \; q_{m'n'}$$

2 → 1 TRANSITIONS

n≥1, all m

$$P(\overset{2}{mn} \rightarrow \overset{1}{m'n'}) = \begin{cases} 0 & , \quad m' < m-1 \\ \sum_{i=o}^{m'-m+1} \sum_{j=o}^{n'} u_{ij} \, p_{m'-m+1-i,\, n'-j} & , \quad \text{otherwise} \end{cases}$$

n=o, m=o

$$P(\overset{2}{oo} \rightarrow \overset{1}{m'n'}) = r_1 \sum_{i=o}^{m'} \sum_{j=o}^{n'} u_{ij} \; p_{m'-i,\, n'-j}$$

Let us look at some of the above expressions.

$$P(\overset{1}{0\ n} \to \overset{1}{m'\ n'}) = 0 \text{ for } n \geq 1 \text{ and}$$

$P(\overset{2}{m\ 0} \to \overset{2}{m'n'}) = 0$ for $m \geq 1$ since, whenever the server empties a line and finds the other line occupied, he will move. A transition from $\overset{1}{0\ 0}$ to $\overset{1}{m'n'}$ will occur only if the next customer to arrive is of type 1 and m' 1-customers and n' 2-customers arrive during his service. Thus $P(\overset{1}{0\ 0} \to \overset{1}{m'\ n'}) = r_1 p_{m'n'}$. Finally the expression for $P(\overset{1}{0\ n} \to \overset{2}{m'\ n'})$ for $n \geq 1$, is a statement of the fact that, in order for this event to occur, it must happen that during the time that it takes the server to cross to line 2 and serve a 2-customer, a total of m' 1-arrivals and n'-n+1 2-arrivals enter the system.

We shall assume that an equilibrium exists and let the steady-state probabilities be π^1_{mn} and π^2_{mn}. The equilibrium equations are

$$\pi^1_{m'\,n'} = \sum_{m=o}^{\infty} \sum_{n=o}^{\infty} \left[\pi^1_{mn}\ P(\overset{1}{mn} \to \overset{1}{m'n'}) + \pi^2_{mn}\ P(\overset{2}{mn} \to \overset{1}{m'n'})\right] \tag{II-3a}$$

$$\pi^2_{m'\,n'} = \sum_{m=o}^{\infty} \sum_{n=o}^{\infty} \left[\pi^1_{mn}\ P(\overset{1}{mn} \to \overset{2}{m'n'}) + \pi^2_{mn}\ P(\overset{2}{mn} \to \overset{2}{m'n'})\right] \tag{II-3b}$$

These equations are homogeneous and serve to determine the ratios of all the probabilities, leaving one to be arbitrarily specified. The ambiguity may be removed by using the condition that

$$\sum_{m=o}^{\infty} \sum_{n=o}^{\infty} (\pi^1_{mn} + \pi^2_{mn}) = 1 \qquad \text{(II-4)}$$

or equally as well by using either of the two conditions

$$\sum_{m=o}^{\infty} \sum_{n=o}^{\infty} \pi^1_{mn} = r_1 \qquad \text{(II-5a)}$$

$$\sum_{m=o}^{\infty} \sum_{n=o}^{\infty} \pi^2_{mn} = r_2 \qquad \text{(II-5b)}$$

The above conditions are valid since the fraction of customers which are of type 1 is r_1 and the fraction which are of type 2 is r_2.

Let $\pi^1(y, z)$ and $\pi^2(y, z)$ be the generating functions of $\pi^1 mn$ and $\pi^2 mn$ respectively.

$$\pi^1(y, z) = \sum_{m=o}^{\infty} \sum_{n=o}^{\infty} \pi^1_{mn} \; y^m z^n \qquad \text{(II-6a)}$$

$$\pi^2(y, z) = \sum_{m=o}^{\infty} \sum_{n=0}^{\infty} \pi^2_{mn} \; y^m z^n \qquad \text{(II-6b)}$$

Thus,

$$\pi^1(y,z) = \sum_{m=o}^{\infty} \sum_{n=o}^{\infty} [\pi^1_{mn} \sum_{m'=o}^{\infty} \sum_{n'=o}^{\infty} P(mn \overset{1}{\rightarrow} \overset{1}{m'n'})\, y^{m'} z^{n'}$$

$$+\pi^2_{mn} \sum_{m'=o}^{\infty} \sum_{n'=o}^{\infty} P(mn \overset{2}{\rightarrow} \overset{1}{m'n'})\, y^{m'} z^{n'}\ ]$$

(II-7a)

$$\pi^2(y,z) = \sum_{m=o}^{\infty} \sum_{n=o}^{\infty} [\pi^1_{mn} \sum_{m'=o}^{\infty} \sum_{n'=o}^{\infty} P(mn \overset{1}{\rightarrow} \overset{2}{m'n'})\, y^{m'} z^{n'}$$

$$+\pi^2_{mn} \sum_{m'=o}^{\infty} \sum_{n'=o}^{\infty} P(mn \overset{2}{\rightarrow} \overset{2}{m'n'})\, y^{m'} z^{n'}]$$

(II-7b

Upon substitution of the expressions for the transition probabilities the above equations become

$$\pi^1(y,z) = \frac{P(y,z)}{y-P(y,z)} [\pi^1_{oo} r_1 y + \pi^2_{oo}(r_1 y-1)U(y,z) - \pi^1(0,z) + \pi^2(y,0)U(y,z)]$$

(II-8a)

$$\pi^2(y,z) = \frac{Q(y,z)}{z-Q(y,z)} [\pi^2_{oo} r_2 z + \pi^1_{oo}(r_2 z-1)V(y,z) - \pi^2(y,0) + \pi^1(0,z)V(y,z)]$$

(II-8b)

where P(y, z), Q(y, z), and U(y, z) and V(y, z), are defined in Eqs. (II-a)

The functions $\pi^1(0, z)$ and $\pi^2(y, 0)$ are boundary conditions of a type which often arises when solving equilibrium equations of this kind. We shall later need to normalize $\pi^1(y, z)$ and $\pi^2(y, z)$ as indicated in Eqs. (II-4) and (II-5). Until then, we shall assume π^1_{oo} to be arbitrary and we will deal entirely with the ratios of all the state probabilities to π^1_{oo}. We therefore make the following definitions.

$$b \equiv \frac{\pi^2_{oo}}{\pi^1_{oo}} \tag{II-9}$$

$$\sigma'(z) \equiv \frac{\pi^1(0, z)}{\pi^1_{oo}} \tag{II-10a}$$

$$b\sigma^2(y) \equiv \frac{\pi^2(y, 0)}{\pi^1_{oo}} \tag{II-10b}$$

The constant b is the ratio of the number of times the system is emptied by completing service on a 2-customer to the number of times the system is emptied by completing service on a 1-customer. $\sigma^1(z)$ and $b\sigma^2(y)$ are the generating functions for π^1_{on}/π^1_{oo} and π^2_{mo}/π^1_{oo} respectively. These quantities are not probabilities but are ratios of probabilities, and as a result remain fixed, independent of the normalization. Notice that

$\sigma^1(0) = \sigma^2(0) = 1$ since $\pi^1(0,0) = \pi^1_{oo}$ and $\pi^2(0,0) = b\pi^1_{oo}$. With the above definitions the equations become

$$\pi^1(y,z) = \frac{\pi^1_{oo} P(y,z)}{y-P(y,z)} \left[r_1 y + b(r_1 y - 1)U(y,z) - \sigma^1(z) + b\sigma^2(y)U(y,z) \right] \qquad \text{(II-11a)}$$

$$\pi^2(y,z) = \frac{\pi^1_{oo} Q(y,z)}{z-Q(y,z)} \left[b\, r_2 z + (r_2 z - 1)V(y,z) - b\sigma^2(y) + \sigma^1(z)\, V(y,z) \right] \qquad \text{(II-11b)}$$

Were it not for the fact that $\sigma^1(z)$, $\sigma^2(y)$, and b are unknown, Eqs. (II-11) would essentially be the solution to our problem. Most of the remainder of this chapter deals with the determination of these quantities.

II-b Determination of Boundary Conditions

In this section we attempt to determine the functions $\sigma^1(z)$ and $\sigma^2(y)$ defined above. We do not get explicit expressions for them but we are able to specify them as the unique solution to a pair of simultaneous functional equations. Although the equations are too difficult to solve, we shall find in subsequent sections that much valuable information can be obtained from them.

In general there are two ways to determine boundary conditions of the kind that we are dealing with here. One is by examination of a more deeply imbedded process. The other is by the use of consistency arguments. Very often one of these methods is simpler to carry out than the other. In this case the later method is simpler and is the one we now discuss. In Chapter IV we shall have occasion to make use of the former method.

Inspection of Eqs. (II-11) reveals that the denominator of one of the expressions goes to zero if $y = P(y, z)$ or if $z = Q(y, z)$. In order that $\pi^1(y, z)$ and $\pi^2(y, z)$ be true generating functions of probability distributions, it is necessary that the numerators of the expressions be zero at the same time.

The equation

$$y = P(y, z) = S_1(\lambda_1 - \lambda_1 y + \lambda_2 - \lambda_2 z) \tag{II-12}$$

defines a functional relationship between y and z. If it can be solved for, say, $y=f(z)$, then when $f(z)$ is substituted for y in the numerator of Eq. (II-11a), the resulting expression must be equal to zero.

In order to solve Eq. (II-12) we shall need some results originally derived by Takács[12]. We consider a single service facility with Poisson arrivals (average rate = λ) and arbitrary service (distribution function $F_S(t)$.) We define the busy period to be the time during which the system is continuously busy, successive busy periods being separated by empty periods. We define the distribution function for the length of a busy period to be $F_B(t)$. Its Laplace-Steiltjes transform is B(s). Takács has shown that B(s) satisfies the functional equation

$$B(s) = S\ (s+\lambda - \lambda\ B(s)\ ) \qquad \text{(II-13)}$$

where S(s) is the Laplace-Steiltjes transform of the service time. In addition he showed that the solution to Eq. (II-13) is unique.

Returning to our problem, we see that the Laplace-Steiltjes transform of a type 1 busy period in isolation, $B_1(s)$, satisfies

$$B_1(s) = S_1(s+\lambda_1-\lambda_1 B(s)\ ) \qquad \text{(II-14)}$$

Upon setting $s = \lambda_2-\lambda_2 z$ there results

$$B_1(\lambda_2-\lambda_2 z) = S_1(\lambda_1-\lambda_1 B_1(\lambda_2-\lambda_2 z) + \lambda_2-\lambda_2 z) \qquad \text{(II-15)}$$

Now if we let

$$y = B_1\ (\lambda_2 - \lambda_2 z) \qquad \text{(II-16a)}$$

we have

$$y\ = S_1(\lambda_1-\lambda_1\ y + \lambda_2-\lambda_2 z) \qquad \text{(II-17)}$$

We have thus shown that the solution of Eq. (II-12) is Eq. (II-16a). Similarly the solution of z= Q(y, z) is

$$z = B_2\ (\lambda_1-\lambda_1 y) \qquad \text{(II-16b)}$$

where $B_2(s)$ is the Laplace Steiltjes transform of a type 2 busy period in isolation. Making the appropriate substitutions of Eqs. (II-16) into the numerators of Eqs. (II-11) and setting the resulting expressions to zero yields

$$\sigma^1(z) = r_1B_1(\lambda_2-\lambda_2z)+b(r_1B_1(\lambda_2-\lambda_2z)-1)\ U(B_1(\lambda_2-\lambda_2z), z)$$

$$+b\sigma^2(B_1(\lambda_2-\lambda_2z))\ U\ (B_1(\lambda_2-\lambda_2z), z) \qquad \text{(II-18a)}$$

$$b\sigma^2(y) = br_2B_2(\lambda_1-\lambda_1y) +(r_2B_2(\lambda_1-\lambda_1y)-1)\ V\ (y, B_2(\lambda_1-\lambda_1y))$$

$$+ \sigma^1(B_2(\lambda_1-\lambda_1y))\ V\ (y, B_2(\lambda_1-\lambda_1y)) \qquad \text{(II-18b)}$$

$$\sigma^1(0) = \sigma^2(0) = 1$$

We have almost completely solved our problem, in principle at least. The above functional equations specify $\sigma^1(z)$ and $\sigma^2(y)$ uniquely. The constant b is still unknown and we now turn our attention to its determination.

II-c Determination of b

We note the appearance of the quantity b in Eqs. (II-18). In this section we discover that only one value of b leads to a consistent solution of these functional equations. We use this condition to determine the proper value of b.

Let us examine Eqs. (II-18) carefully and make the assumption for the moment that b has been correctly determined. Since $\sigma^1(0)$ is known, ($\sigma^1(0) = 1$) substitution of $z = 0$ into Eq. (II-18a) enables us to determine $\sigma^2(B_1(\lambda_2))$. Now if $y = B_1(\lambda_2)$ is substituted into Eq. (II-11b) the value of $\sigma^1(B_2(\lambda_1 - \lambda_1 B_1(\lambda_2)))$ may be found. Continuing in this manner, by alternating the equations in which substitutions are made, it is possible to evaluate the functions $\sigma^1(z)$ and $\sigma^2(y)$ for many different sizes of their arguments.

On the other hand this process could also have been initiated by first letting $y = 0$ in Eq. (II-18b), revealing the value of $\sigma^1(B_2(\lambda_1))$. Now setting $z = B_2(\lambda_1)$ in Eq. (II-18a) allows us to find $\sigma^2(B_1(\lambda_2 - \lambda_2 B_2(\lambda_1)))$. Again this procedure may be repeated many times so that $\sigma^1(z)$ and $\sigma^2(y)$ become known for some additional values of their arguments.

The procedure we have just considered may be described formally by rewriting Eqs. (II-18) in the following manner:

$$z_o = 0 \quad \underline{\text{or}} \quad y_o = 0$$

$$y_i = B_1(\lambda_2 - \lambda_2 z_i) \tag{II-19a}$$

$$z_{i+1} = B_2(\lambda_1 - \lambda_1 y_i) \tag{II-19b}$$

$$\sigma^2(y_i) = 1 - r_1 y_i + \frac{\sigma^1(z_i) - r_1 y_i}{bU(y_i, z_i)} \tag{II-20a}$$

$$\sigma^1(z_{i+1}) = 1 - r_2 z_{i+1} + b \frac{\sigma^2(y_i) - r_2 z_{i+1}}{V(y_i, z_{i+1})} \tag{II-20b}$$

We start with either $z_0=0$ or $y_0=0$ and proceed through increasing values of i as indicated.

It is shown in Appendix II that

$$\lim_{i \to \infty} z_i = 1 \tag{II-21}$$

regardless of the starting point (either $z_o=0$ or $y_0=0$), as long as the condition

$$1 - \rho_1 - \rho_2 > 0 \tag{II-22}$$

is satisfied. This means that $\lim_{i \to \infty} \sigma^1(z_i) = \sigma^1(1)$, again regardless of the starting point. The important point to note is this-- if the wrong value of b is used, $\lim_{i \to \infty} \sigma^1(z_i)$ will <u>not</u> be independent of the starting point. In fact Appendix III shows that the difference between the quantity

$$\lim_{i \to \infty} \sigma^1(z_i), \text{ starting at } z_o = 0.$$

and the quantity

$$\lim_{i \to \infty} \sigma^1(z_i), \text{ starting at } y_o = 0$$

is linearly related to b. Taking advantage of this fact, it is easy to determine the value of b which causes $\lim_{i \to \infty} \sigma^1(z_i)$ to be the same, independent of the starting point. A more detailed description of this iterative procedure for finding b is given in Appendix III.

We have thus discovered a method whereby the value of the constant b can be found. Since now b is known, and the functions $\sigma^1(z)$ and $\sigma^2(y)$ are completely specified, our problem is in principle solved.

The trouble is that in general it is impossible to solve the functional equations (Eqs. II-18) explicitly. However, despite this, much useful information can be obtained from them. We have already seen how they can be used to obtain b as well as the values of $\sigma^1(z)$ and $\sigma^2(y)$ for specific values of y and z. In particular, $\sigma^1(1)$ was obtained as a by-product of the iterative procedure described above. In the next section we shall see that still more use may be made of these equations.

Before we do this, let us first examine one special case where it is possible to get explicit solutions to Eqs. (II-18). We let $\mu_1 = \mu_2 = \infty$. This corresponds to a situation where the service times are extremely short. For this case

$$\sigma^1(z) = r_1 + bS_{21}(\lambda_2 - \lambda_2 z)\,(\sigma^2(1) - r_2) \qquad \text{(II-23a)}$$

$$b\sigma^2(y) = br_2 + S_{12}(\lambda_1 - \lambda_1 y)(\sigma^1(1) - r_1) \qquad \text{(II-23b)}$$

so that for y=z=0

$$1 = r_1 + bS_{21}(\lambda_2)\,(\sigma^2(1) - r_2) \qquad \text{(II-24a)}$$

$$b = br_2 + S_{12}\;(\lambda_1)(\sigma^1(1) - r_1) \qquad \text{(II-24b)}$$

This is consistent with Eq. (II-28) (derived in the next section) only if

$$b = \frac{r_2}{r_1}\;\frac{S_{12}(\lambda_1)}{S_{21}(\lambda_2)} \qquad \text{(II-25)}$$

And it follows that

$$\sigma^1(z) = r_1 + r_2\;\frac{S_{21}(\lambda_2 - \lambda_2 z)}{S_{21}(\lambda_2)} \qquad \text{(II-26a)}$$

$$\sigma^2(y) = r_2 + r_1\;\frac{S_{12}(\lambda_1 - \lambda_1 y)}{S_{12}(\lambda_1)} \qquad \text{(II-26b)}$$

which is the solution we were seeking.

II-d Derivatives of $\sigma^1(z)$ and $\sigma^2(y)$ at $y = z = 1$

In this section we find that it is possible to obtain sufficiently much information about $\sigma^1(z)$ and $\sigma^2(y)$ from Eqs. (II-18) so that all moments of the waiting times may be determined.

We note, first of all, that the moments of the waiting times can be expressed in terms of derivatives of $\pi^1(y, z)$ and $\pi^2(y, z)$ evaluated at the point $y = z = 1$. $\pi^1(y, z)$ and $\pi^2(y, z)$ are given in terms of $\sigma^1(z)$ and $\sigma^2(y)$, and therefore we can obtain the moments of the waiting times if we can find the following quantities

$$\begin{array}{ll} \sigma^1(1) & \sigma^2(1) \\ \sigma^1(1)' & \sigma^2(1)' \\ \sigma^1(1)'' & \sigma^2(1)'' \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \end{array} \qquad \text{(II-27)}$$

where $\sigma^1(1)' \equiv \left.\dfrac{d\sigma^1(z)}{dz}\right|_{z=1}$ etc. It turns out, that since b and $\tau^1(1)$ are already known (from the methods of the previous section), all the other quantities listed above may be gotten by taking appropriate derivatives of Eqs. (II-18). The remainder of this section is devoted to showing this in detail.

If y = z = 1 is substituted into Eqs. (II-18), the result of both equations is

$$\sigma^1(1) - r_1 = b(\sigma^2(1) - r_2) \qquad \text{(II-28)}$$

which gives $\sigma^2(1)$ in terms of known quantities. Incidentally, an interesting interpretation of this equation may be obtained by deriving it in another way. The probability that a given service is followed by a changeover from 1 to 2 is

$$\sum_{n=1}^{\infty} \pi^1_{on} + \pi^1_{oo}\, r_2 = \sum_{n=o}^{\infty} \pi^1_{on} - \pi^1_{oo}\, r_1$$

$$= \pi^1_{oo}(\sigma^1(1) - r_1) \qquad \text{(II-29a)}$$

This is true because a switch from 1 to 2 is made only if a service completion leaves the system in one of the states ${}_{o}^{1}{}_{n}$ $n \geq 1$, or if it leaves the system in state ${}_{o}^{1}{}_{o}$ and the next arrival is of type 2. Similarly the probability for a 2 to 1 changeover is

$$\sum_{m=1}^{\infty} \pi^2_{mo} + \pi^2_{oo}\, r_1 = \sum_{m=o}^{\infty} \pi^2_{mo} - \pi^2_{oo}\, r_2$$

$$= \pi^2_{oo}(\sigma^2(1) - r_2) \qquad \text{(II-29b)}$$

The number of changeovers in either direction must be equal. Equating the probabilities of Eqs. (II-29) and taking note of the definition of b in Eq. (II-9), produces Eq. (II-28).

Continuing, we take the first derivative of Eqs. (II-18) and evaluate the result at y = z = 1. The outcome is

$$\sigma^1(1)' = \frac{\lambda_2}{\mu_1 - \lambda_1} \left[(1+b)r_1 + b\frac{\mu_1}{\mu_{21}} (\sigma^2(1) - r_2) + b\sigma^2(1)' \right] \qquad \text{(II-30a)}$$

$$\sigma^2(1)' = \frac{\lambda_1}{\mu_2 - \lambda_2} \left[(1 + \frac{1}{b})r_2 + \frac{1}{b}\frac{\mu_2}{\mu_{12}} (\sigma^1(1) - r_1) + \frac{1}{b}\sigma^1(1)' \right] \qquad \text{(II-30b)}$$

One notes that this pair of simultaneous equations may be solved for $\sigma^1(1)'$ and $\sigma^2(1)'$ in terms of known quantities.

In a like manner the second derivatives of Eqs. (II-18) may be taken. The result is

$$\begin{aligned}
\sigma^1(1)'' = {} & r_1\lambda_2^2 E(S_1^2) \left(\frac{\mu_1}{\mu_1 - \lambda_1}\right)^3 (1+b) + 2b\, r_1 \frac{\mu_1}{\mu_{21}} \left(\frac{\lambda_2}{\mu_1 - \lambda_1}\right)^2 \\
& + b \left[\left(\frac{\mu_1\lambda_2}{\mu_1 - \lambda_1}\right)^2 E(S_{21}^2) + \frac{\lambda_1\lambda_2^2}{\mu_{21}} \left(\frac{\mu_1}{\mu_1 - \lambda_1}\right)^3 E(S_1^2) \right] (\sigma^2(1) - r_2) \\
& + b \left[\frac{\lambda_2^2\mu_1}{\mu_{21}(\mu_1 - \lambda_1)^2} + \lambda_2^2 \left(\frac{\mu_1}{\mu_1 - \lambda_1}\right)^3 E(S_1^2) \right] \sigma^2(1)' \\
& + b \left(\frac{\lambda_2}{\mu_1 - \lambda_1}\right)^2 \sigma^2(1)''
\end{aligned} \qquad \text{(II-31a)}$$

with an obvious counterpart for $\sigma^2(1)''$ which we shall call Eq. (II-31b). The only unknown quantities in this pair of simultaneous equations are $\sigma^1(1)''$ and $\sigma^2(1)''$. Solving the equations gives the desired derivatives.

One notes that the n-th derivative of Eqs. (II-18) expresses $\sigma^1(1)^{(n)}$ in terms of $\sigma^2(1)^{(n)}, \ldots, \sigma^2(1)', \sigma^2(1)$ and $\sigma^2(1)^{(n)}$ in terms of $\sigma^1(1)^{(n)}, \ldots, \sigma^1(1)', \sigma'(1)$. If the derivatives through the (n-1)st have been determined, the nth derivatives can thus be found by solving a pair of linear simultaneous equations. By induction therefore, any number of derivatives can be obtained. Of course, the manipulations involved in calculating the higher derivatives become increasingly tedious.

Our knowledge of the behavior of the state probability generating functions $\pi^1(y, z)$ and $\pi^2(y, z)$ is now quite complete. In particular, we now have at our disposal any derivatives of these functions evaluated at $y = z = 1$, and can thus determine moments of waiting times and number-in-system. Before proceeding with this, however, there is the one final matter of normalization which needs to be taken care of.

II-e Non-Saturation Condition

The quantity π^1_{oo} remains to be determined. In terms of generating functions, Eq. (II-5a) is

$$r_1 = \lim_{\substack{y \to 1 \\ z \to 1}} \pi^1(y, z) \qquad \text{(II-32)}$$

Using the expression of Eq. (II-11a) in the above, one finds that both the numerator and denominator go to zero so that L' Hopital' s rule must be employed. One then obtains

$$\pi^1_{oo} = \frac{(1-\rho_1)r_1}{(1+b)\, r_1 + b\, \frac{\lambda_1}{\mu_{21}}\, (\sigma^2(1)-r_1)+b\sigma^2(1)'} \qquad \text{(II-33)}$$

Substituting $\sigma^2(1)'$ which comes from the solution to Eqs. (II-30) and also using Eq. (II-28), there results

$$\pi^1_{oo} = \frac{1-\rho_1 - \rho_2}{1 + b + (\lambda_1+\lambda_2)(\frac{1}{\mu_{12}} + \frac{1}{\mu_{21}})\, (\sigma^1(1)-r_1)} \qquad \text{(II-34)}$$

where the values of both b and $\sigma^1(1)$ are produced by the iterative procedure discussed in Section II-c.

It was shown in Chapter 1 that the general-time probability that the system is empty is $\pi^1_{oo} + \pi^2_{oo} = (1+b)\pi^1_{oo}$. Thus the system will remain unsaturated as long as π^1_{oo} is positive. The non-saturation condition, then is

$$1 - \rho_1 - \rho_2 > 0 \qquad \text{(II-35)}$$

independent of the changeover time. This can be explained by noticing that when traffic is extremely heavy, the time spent in switching becomes negligibly small relative to the time spent servicing customers.

II-f Mean Waiting Times and Other Quantities

In this section we use the results of the previous sections to obtain expressions for the mean waiting times and several other quantities of interest.

The Laplace-Steiltjes transform of a 1-customer was related to the generating function of the state probabilities in Eq. (I-16a). Making the appropriate substitutions here one finds

$$W_1(s) = \frac{\pi^1_{oo}}{r_1} \frac{r_1\left(1-\frac{s}{\lambda_1}\right) - bS_{21}(s)\left(r_2+r_1\frac{s}{\lambda_1}\right) - \sigma^1(1) + bS_{21}(s)\sigma^2\left(1-\frac{s}{\lambda_1}\right)}{1-\frac{s}{\lambda_1} - S_1(s)} \qquad \text{(II-36)}$$

where π^1_{oo} is given in Eq. (II-34). The mean waiting time for a 1-customer is

$$\bar{W}_1 = \lim_{s \to o} -\frac{dW_1(s)}{ds} \qquad \text{(II-37)}$$

The limit in the above requires the double application of L'Hopital's rule. The outcome is an expression containing $\sigma^2(1)'$ and $\sigma^2(1)''$, which are found by solving the simultaneous equations indicated in Section II-d. The final result is

$$\bar{W}_1 = \frac{1}{(1-\rho_1-\rho_2)(1-\rho_1-\rho_2+2\rho_1\rho_2)(C_1+C_2)} \Big[\frac{(1-\rho_1)(1-\rho_2)^2+\rho_1\rho_2^2}{\mu_{21}} (\rho_2 C_1+(1-\rho_1)C_2)$$

$$+ \frac{\rho_2(1-\rho_1)}{\mu_{12}} ((1-\rho_2)C_1 + \rho_1 C_2)$$

$$+\frac{1}{2} (C_1+C_2)(((1-\rho_1)(1-\rho_2)^2+\rho_1\rho_2^2 + \rho_2^2) \lambda_1 E(S_1^2) +(1-\rho_1)\lambda_2 E(S_2^2))$$

$$+\frac{1}{2} C(1-\rho_1-\rho_2)(((1-\rho_1)(1-\rho_2)^2+\rho_1\rho_2^2 + \rho_2^2) E(S_{21}^2) +(1-\rho_1)E(S_{12}^2)\Big]$$

(II-38)

where we have used

$$C \equiv (\lambda_1 + \lambda_2)(\sigma^1(1) - r_1) \qquad \text{(II-39a)}$$

$$C_1 \equiv 1 + \frac{C}{\mu_{21}} \qquad \text{(II-39b)}$$

$$C_2 \equiv b + \frac{C}{\mu_{12}} \qquad \text{(II-39c)}$$

The same equation with subscripts switched holds for $\bar{W}_2$. The mean waiting time for an arbitrary customer is $\bar{W} = r_1\bar{W}_1 + r_2\bar{W}_2$. Using the above definitions π^1_{oo} may be rewritten

$$\pi^1_{oo} = \frac{1-\rho_1-\rho_2}{C_1+C_2} \qquad \text{(II-40)}$$

and of course

$$\pi^2_{oo} = b\ \pi^1_{00} \tag{II-41}$$

LIMITING CASES

1) Zero Changeover Times

In the special case $\mu_{12}=\mu_{21}=\infty$, the expression for $\bar{W}$ becomes

$$\bar{W}_1 = \frac{[(1-\rho_1)(1-\rho_2)^2 + \rho_2^2(1+\rho_1)]\ \lambda_1 E(S_1^2) + (1-\rho_1)\lambda_2 E(S_2^2\ )}{2(1-\rho_1-\rho_2)(1-\rho_1-\rho_2+2\ \rho_1\rho_2)} \tag{II-42}$$

This agrees with the results of Avi-Itzhak, Maxwell, and Miller.[2] No iteration need be performed in this case unless the value of b, defined in Eq. (II-9) is desired. Also in this case

$$\pi^1_{oo}(1+b) = 1-\rho_1-\rho_2 \tag{II-43}$$

as expected, since the probability that the system is empty cannot depend upon the queue discipline, as long as the server is never idle when customers are waiting.

2) Zero Service Times

In the case $\mu_1=\mu_2=\infty$, results can again be obtained without iteration as shown in Section II-c. There follows

$$\bar{W}_1 = \frac{\lambda_1\lambda_2}{\lambda_2 S_{12}(\lambda_1) + \lambda_1 S_{21}(\lambda_2) + \lambda_1\lambda_2 \frac{1}{\mu_{12}} + \frac{1}{\mu_{21}}}$$

$$\times \left[\frac{S_{12}(\lambda_1)}{\lambda_1 \mu_{21}} + \frac{1}{\mu_{12}\mu_{21}} + \frac{E(S_{12}^2) + E(S_{21}^2)}{2} \right] \qquad \text{(II-44)}$$

and

$$\pi_{oo}^1 = \frac{\lambda_1 S_{21}(\lambda_2)}{\lambda_1 S_{21}(\lambda_2) + \lambda_2 S_{12}(\lambda_1) + \lambda_1\lambda_2 \left(\frac{1}{\mu_{12}} + \frac{1}{\mu_{21}} \right)} \qquad \text{(II-45)}$$

3) Symmetry

If the system is symmetric, considerable simplification is achieved. The functions $\sigma^1(.)$ and $\sigma^2(.)$ are identical. Thus the simultaneous equations discussed in Section II-d reduce to a single equation in a single unknown. Also we know from symmetry that b = 1. The results for this case are:

$$\bar{W} = \frac{1}{1-2\rho} \left[\frac{1}{2\mu_T} + \lambda E(S^2) + C\pi_{oo}^1 E(S_T^2) \right] \qquad \text{(II-46)}$$

and

$$\pi_{oo}^1 = \pi_{oo}^2 = \frac{\frac{1-2\rho}{2C}}{1 + \mu_T} \qquad \text{(II-47)}$$

where

$$\rho \equiv \rho_1 = \rho_2 \qquad E(S^2) \equiv E(S_1^2) = E(S_2^2)$$

$$\mu_T \equiv \mu_{12} = \mu_{21} \qquad E(S_T^2) \equiv E(S_{21}^2) = E(S_{12}^2) \qquad \text{(II-48)}$$

$$\lambda \equiv \lambda_1 = \lambda_2$$

END OF LIMITING CASES

We can readily obtain several other quantities of interest. There are $\lambda_1 + \lambda_2$ service completions per unit time on the average. With probability $\pi^1_{oo}\,(\sigma^1(1) - r_1)$ a given service is followed by a switch from 1 to 2. (See Eq. (II-29a). Thus the fraction of the time that the server spends crossing from 1 to 2 is

$$(\lambda_1+\lambda_2)\,\pi^1_{oo}(\sigma^1(1)-r_1)\;/\;\mu_{12} = \frac{C\pi^1_{oo}}{\mu_{12}} \tag{II-49}$$

and the fraction spent crossing from 2 to 1 is $C\pi^1_{oo}/\mu_{21}$.

As pointed out in Chapter I (Eq. I-23a) the general-time probability that the system is empty is $\pi^1_{oo} + \pi^2_{oo} = (1+b)\,\pi^1_{oo}$. State $o\,{}^{2}\,o$ is entered on the average b times as often as state $o\,{}^{1}\,o$. Since the average time spent in each state is the same (each state remains occupied until the next arrival, of either type, occurs), the probability that the system is empty and the server is at line 1 must be

$$\frac{1}{1+b}\,(1+b)\pi^1_{oo} = \pi^1_{oo} \tag{II-50}$$

and the probability that the system is empty and the server is at line 2 must be

$$\frac{b}{1+b}\,(1+b)\pi^1_{oo} = b\pi^1_{oo} \tag{II-51}$$

Thus the probabilities of the states $o\,{}^{1}\,o$ and $o\,{}^{2}\,o$ are the same for the imbedded process and for general-time.

The fraction of the server's time spent at line 1 is $\pi^1_{oo} + \rho_1$. A fraction π^1_{oo} is spent idle and a fraction ρ_1 is spent busy. Similarly, the fraction of the server's time spent at line 2 is $b\pi^1_{oo} + \rho_2$.

The probability that the service completion of a 1-customer is followed by a switch is

$$\frac{\Pr(\text{customer is of type 1 } \underline{\text{and}} \text{ customer is followed by switch from 1 to 2})}{\Pr(\text{customer is of type 1})}$$

$$= \frac{\pi^1_{oo}(\sigma^1(1) - r_1)}{r_1}$$

Similarly, the probability that the service completion of a 2-customer is followed by a switch is

$$\frac{\pi^1_{oo} \quad (\sigma^1(1)-r_1)}{r_2}$$

The probability that an arbitrary customer is followed by a switch is

$$r_1 \Pr(\text{1-customer is followed by a switch}) + r_2 \Pr(\text{2-customer is followed by a switch})$$

$$= 2\pi^1_{oo}(\sigma^1(1) - r_1).$$

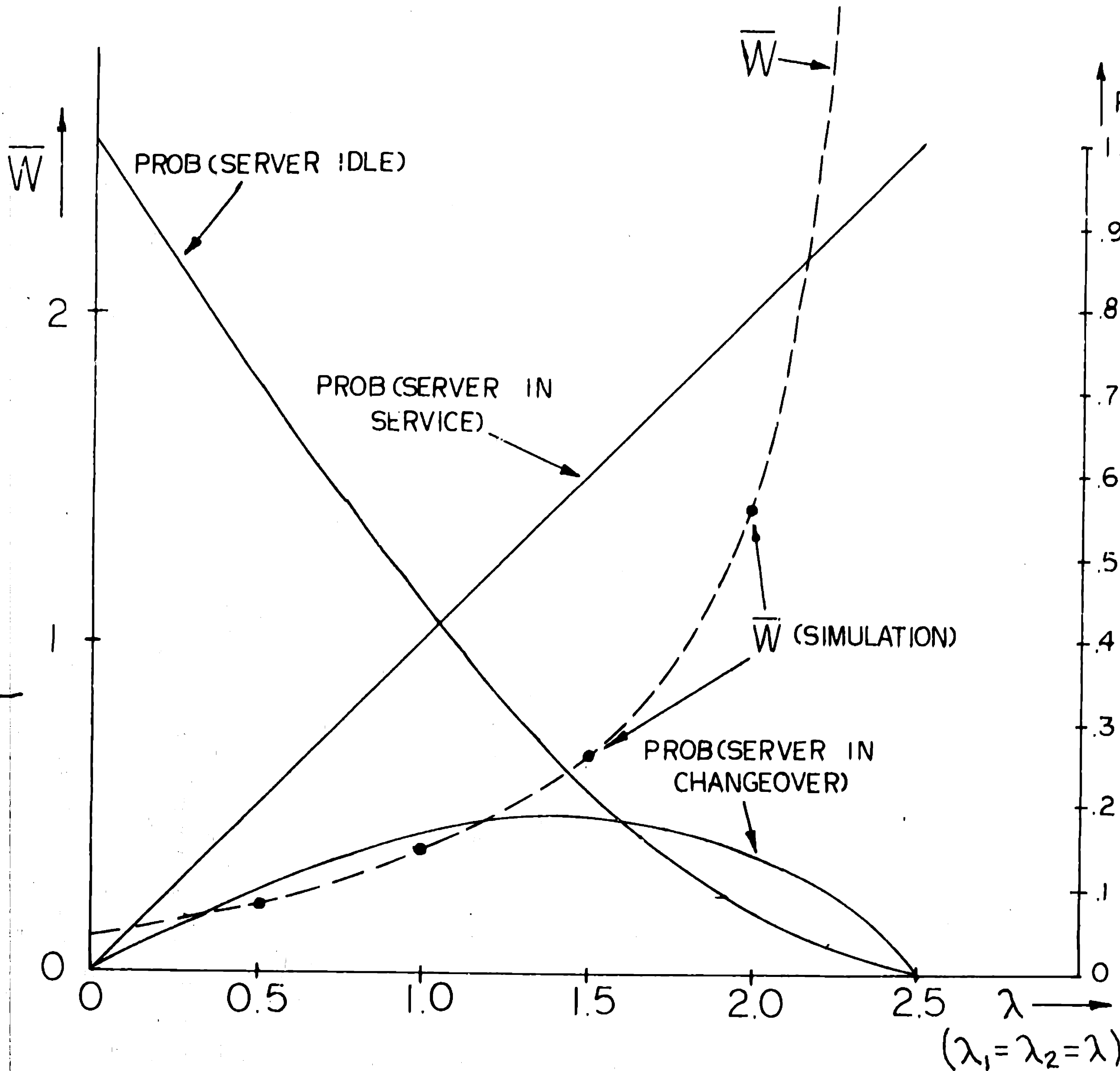
These latter quantities are of interest if a cost as well as a time is associated with changeovers.

A program has been written in the FORTRAN IV language which carries out all of the computation described in this chapter. Run in double-precision mode on an IBM 360/65 computer, each calculation for all of the quantities mentioned, for a given set of input parameters requires somewhat less than one second, and is accurate to more than 8 significant digits.

Fig. 2 is a plot of the waiting time versus arrival rate for a symmetrical system. Also plotted is the fraction of time that the server spends idle, serving customers, and switching. It is of interest to note that, as expected, at near saturation levels, the server spends only a very small portion of his time crossing between lines. Results of simulation runs are superimposed on the plots to verify the analytical predictions.

II-g Summary

An application of generating functions to the imbedded Markov process has enabled us to find expressions for the mean waiting time for each type of customer, as well as other interesting quantities. The complexity of the problem necessitated the utilization of some novel mathematical techniques, but did not otherwise affect the over-all approach.



$$\frac{1}{\mu_1} = \frac{1}{\mu_2} = \frac{1}{\mu_{12}} = \frac{1}{\mu_{21}} = .2$$

EXPONENTIALLY DISTRIBUTED SERVICE AND CHANGEOVER

PLOT OF MEAN WAITING TIME AND SERVER PROBABILITIES vs. ARRIVAL RATE
— ALTERNATING PRIORITY —

FIG. 2

CHAPTER III

STRICT PRIORITY

The strict priority rule, as opposed to the alternating priority rule, always gives customers of one type higher priority than the other. We shall choose customers of type 1 to be "high priority" and those of type 2 to be "low priority." Thus, whenever both types are present, the server will attend to a customer of type 1.

The motivations for the investigation of strict priority are that it is simple and that situations exist when it is optimum or near-optimum. It is known for example, that in the absence of changeover times, a strict priority discipline, assigning high priority to those customers with the shorter mean service time, is optimum (if performance is judged on the basis of the mean waiting time of an arbitrary customer.)[8] We would expect, therefore, if changeover times were not too large relative to service times and if the mean service time of one type of customer differed greatly from the mean service time of the other type, the use of the strict priority discipline would be desirable.

The introduction of changeover times necessitates some further specification of the server's behavior. There is the question of whether, after crossing from 1 to 2, the server may cross back to 1 without serving any 2-customers. We shall assume that this is not allowed and that at least one 2-customer must be served. Also we must specify the behavior of the server when the system becomes empty. We shall assume that when service is terminated on the last customer in the system, the server will remain stationary at that line. It should be possible, using the techniques we employ, to easily handle modifications of these two assumptions.

The steps taken in this chapter parallel very closely those of Chapter II. We again concentrate on the imbedded Markov process formed at service-completion instants. From the transition probabilities, expressions for the generating functions of the state probabilities are obtained and as in Chapter II contain unknown boundary conditions. These boundary conditions are observed to satisfy two simultaneous equations which are solved, producing a complete solution to the problem. It is then easy to obtain expressions for the mean waiting times and certain other measures of the system's effectiveness.

III-a The Imbedded Process

The transition probabilities for the strict priority discipline are

1 → 1 TRANSITIONS

<u>m≥1, all n</u>

$$P(\overset{1}{mn} \rightarrow \overset{1}{m'n'}) = \begin{cases} 0 & , m'<m-1 \\ 0 & , n'<n \\ p_{m'-m+1,n'-n} , & \text{otherwise} \end{cases}$$

<u>m=o, n≥1</u>

$$P(\overset{1}{on} \rightarrow \overset{1}{m'n'}) = 0$$

<u>m=o, n=o</u>

$$P(\overset{1}{00} \rightarrow \overset{1}{m'n'}) = r_1 \, p_{m'n'}$$

1 → 2 TRANSITIONS

<u>m≥1, all n</u>

$$P(\overset{1}{mn} \rightarrow \overset{2}{m'n'}) = 0$$

<u>m=o, n≥1</u>

$$P(\overset{1}{on} \rightarrow \overset{2}{m'n'}) = \begin{cases} 0 & , n'<n-1 \\ \sum_{i=o}^{m'} \sum_{j=o}^{n'-n+1} v_{ij} q_{m'-i,n'-n+1-j}, & \text{otherwise} \end{cases}$$

<u>m=o, n=o</u>

$$P(\overset{1}{oo} \rightarrow \overset{2}{m'n'}) = r_2 \sum_{i=o}^{m'} \sum_{j=o}^{n'} v_{ij} q_{m'-i,n'-j}$$

2 → 2 TRANSITIONS

m≥1, all n

$$P(mn \overset{2}{\rightarrow} \overset{2}{m'n'}) = 0$$

m=o, n≥1

$$P(on \overset{2}{\rightarrow} \overset{2}{m'n'}) = \begin{cases} 0 & , n'<n-1 \\ q_{m',n'-n+1} & , \text{otherwise} \end{cases}$$

m=o, n=o

$$P(oo \overset{2}{\rightarrow} \overset{2}{m'n'}) = r_2 \, q_{m'n'}$$

2 → 1 TRANSITIONS

m≥1, all n

$$P(mn \overset{2}{\rightarrow} \overset{1}{m'n'}) = \begin{cases} 0 & , m'<m-1 \\ 0 & , n'< n \\ \sum_{i=o}^{m'-m+1} \sum_{j=o}^{n'} u_{ij} p_{m'-m+1-i, n'-j}, & \text{otherwise} \end{cases}$$

m=o, n≥1

$$P(on \overset{2}{\rightarrow} \overset{1}{m'n'}) = 0$$

m=o, n=o

$$P(oo \overset{2}{\rightarrow} \overset{1}{m'n'}) = r_1 \sum_{i=o}^{m'} \sum_{j=o}^{n'} u_{ij} p_{m'-i, n'-j}$$

where p_{ij}, q_{ij}, u_{ij} and v_{ij} were defined at the beginning of Chapter II. The difference between these probabilities and those for alternating priority appear in transitions from line 2. For example $P(\overset{2}{mn} \rightarrow \overset{2}{m'n'}) = 0$ when $m \geq 1$ since the server will always go to line 1 if there are any 1-customers waiting. When the server is at line 1, he treats 1-customers as "high priority" and 2-customers as "low priority" under either strict or alternating priority. Thus the transition probabilities from line 1 are the same for both disciplines.

The equilibrium equations of Chapter II (Eqs. (II-3)) again hold, as well as the normalization conditions (Eq. (II-4) and Eqs. (II-5)). When the above transition probabilities are introduced into the equilibrium equations, and generating functions taken, there results

$$\pi^1(y,z) = P(y,z) \Big\{ [\pi^1_{oo} r_1 + \frac{1}{y}(\pi^1(y,z) - \pi^1(o,z))] + U(y,z)[\pi^2_{oo} r_1 + \frac{1}{y}(\pi^2(y,z) - \pi^2(o,z))] \Big\} \qquad \text{(III-1a)}$$

$$\pi^2(y,z) = Q(y,z) \Big\{ V(y,z)[\pi^1_{oo} r_2 + \frac{1}{z}(\pi^1(o,z) - \pi^1_{oo})] + [\pi^2_{oo} r_2 + \frac{1}{z}(\pi^2_{oo} r_2 + \frac{1}{z}(\pi^2(o,z) - \pi^2_{oo})] \Big\} \qquad \text{(III-1b)}$$

where the functions $P(y,z)$, $Q(y,z)$, $U(y,z)$ and $V(y,z)$ are given by Eqs. (II-2).

One notes the appearance of the boundary conditions $\pi^1(o, z)$ and $\pi^2(o, z)$.

In order to isolate the problem of normalization from the other more fundamental issues, we treat π^1_{oo} for the time being as arbitrary and deal with the ratios of the state probabilities. To facilitate this, the following definitions are made

$$b = \frac{\pi^2_{oo}}{\pi^1_{oo}} \tag{III-2}$$

$$\sigma^1(z) = \frac{\pi^1(o, z)}{\pi^1_{oo}} \tag{III-3a}$$

$$b\sigma^2(z) = \frac{\pi^2(o, z)}{\pi^1_{oo}} \tag{III-3b}$$

The quantity b, the ratio of the idle time spent by the server at line 2 to the idle time spent at line 1, has the same interpretation as in Chapter II. Also the definition of $\sigma^1(z)$ is the same. It should be noticed that $b\sigma^2(z)$ is the generating function of π^2_{on}/π^1_{oo}, whereas in Chapter II we used a similar notation for the generating function of π^2_{mo}/π^1_{oo} . One observes that $\sigma^1(o) = \sigma^2(o) = 1$, since $\pi^1(o, o) = \pi^1_{oo}$ and $\pi^2(o, o) = b\pi^1_{oo}$. Making use of these definitions Eqs. (III-1) become

$$\pi^1(y,z) = \frac{\pi^1_{oo}\, P(y,z)}{y-P(y,z)} \left[r_1 y - \sigma^1(z) + U(y,z)\left(br_1 y - b\sigma^2(z) + \frac{\pi^2(y,z)}{\pi^1_{oo}} \right) \right] \qquad \text{(III-4a)}$$

$$\pi^2(y,z) = \pi^1_{oo}\, Q(y,z) \left[V(y,z) \left(\frac{\sigma(z) + zr_2 - 1}{z} \right) + b \left(\frac{\sigma^2(z) + zr_2 - 1}{z} \right) \right] \qquad \text{(III-4b)}$$

Eqs. (III-4) are the generating functions for the state probabilities, expressed in terms of the (as yet unknown) quantities $\sigma^1(z)$, $\sigma^2(z)$ and b.

III-b Determination of Boundary Conditions

In this section we derive and solve equations which specify the functions $\sigma^1(z)$ and $\sigma^2(z)$. These equations also serve to determine the constant b, since it happens that only one value of b leads to an acceptable solution.

The denominator of Eq. (III-4a) is zero when

$$y = P(y,z) \qquad \text{(III-5)}$$

so that a necessary condition is that the numerator be zero when Eq. (III-5) is satisfied. In Section II-c of Chapter II this equation was solved, the result being

$$y = B_1(\lambda_2 - \lambda_2 z) \qquad \text{(III-6)}$$

where we recall that $B_1(s)$ is the Laplace-Steiltjes transform of a type-1 busy period in isolation. Making the substitution of Eq. (III-6) in the numerator of Eq. (III-4a) and setting the result to zero yields

$$\sigma^1(z)\,(z - S_{21}(\alpha)S_{12}(\alpha)S_2(\alpha)\,) = zr_1S_1(\alpha)$$

$$+S_{21}(\alpha)\,[bzr_1S_1(\alpha) + S_2(\alpha)\,(zr_2-1)(S_{12}(\alpha)+b)+b\sigma^2(z)(S_2(\alpha)-z)\,]$$

(III-7a)

where for the sake of simplicity we have used

$$\alpha = \alpha(z) \equiv \lambda_1 - \lambda_1 B_1\,(\lambda_2 - \lambda_2 z) + \lambda_2 - \lambda_2 z \qquad \text{(III-8)}$$

and also

$$B_1(\lambda_2 - \lambda_2 z) = S_1(\alpha) \qquad \text{(III-9)}$$

the latter equation coming from Eq. (II-15). Another equation satisfied by $\sigma^1(z)$ and $\sigma^2(z)$ can be obtained by requiring the expression for $\pi^2(y, z)$ in Eq. (III-4b) to reduce to $\pi^2(o, z) = b\pi^1_{oo}\sigma^2(z)$ when y is set to zero. From this condition one finds

$$b\sigma^2(z)\,(z-S_2(\beta)\;) = (zr_2-1)\,S_2(\beta)(S_{12}(\beta)+b) + S_{12}(\beta)\,S_2(\beta)\,\sigma^1(z) \qquad \text{(III-7b)}$$

where we have used

$$\beta = \beta(z) \equiv \lambda_1 + \lambda_2 - \lambda_2 z \qquad \text{(III-10)}$$

One notes that Eqs. (III-7) are linear and may be solved. Considerable simplification is achieved if the equations are solved for the quantities

$$\frac{\sigma^1(z) + zr_2-1}{z} \quad \text{and } b\,\frac{\sigma^2(z)+ zr_2-1}{z}$$

rather than $\sigma^1(z)$ and $\sigma^2(z)$. The solution is

$$\frac{\sigma^1(z) + zr_2-1}{z} = \frac{A_1(z)}{D(z)} \qquad \text{(III-11a)}$$

$$b\,\frac{\sigma^2(z) + zr_2-1}{z} = \frac{A_2(z)}{D(z)} \qquad \text{(III-11b)}$$

where

$$A_1(z) = [z-S_2(\beta)(1+b\,S_{21}(\alpha))](r_1S_1(\alpha) +zr_2-1)+ bS_{21}(\alpha)[zr_1S_1(\alpha)+(zr_2-1)S_2(\alpha)]$$

(III-12a)

$$D(z) = (z-S_2(\beta))(z-S_{21}(\alpha)S_{12}(\alpha)S_2(\alpha)) + (z-S_2(\alpha))S_{21}(\alpha)S_{12}(\beta)S_2(\beta)$$

(III-12b)

$$A_2(z) = S_{12}(\beta)S_2(\beta)(1+bS_{21}(\alpha))(r_1S_1(\alpha) + zr_2-1) + b(zr_2-1)(z-S_{21}(\alpha)S_{12}(\alpha)S_2(\alpha))$$

(III-12c)

We now know $\sigma^1(z)$ and $\sigma^2(z)$ and only b remains to be determined. In Appendix III it is proved that D(z) is zero for some value of z between o and 1. It follows that $A_1(z)$ and $A_2(z)$ must also be zero at that point. This condition is sufficient to determine b.

Let the value of z for which D(z) is zero be z_o, i. e. $D(z_o) = 0$. Using either $A_1(z_o) = 0$ or $A_2(z_o) = 0$, one finds

$$b = \frac{(z_o-S_2(\beta_o))(1-r_1S_1(\alpha_o) - r_2z_o)}{S_{21}(\alpha_o)[\,r_1S_1(\alpha_o)(z_o-S_2(\beta_o)) + (r_2z_o-1)(S_2(\alpha_o)-S_2(\beta_o))\,]} \qquad \text{(III-13)}$$

where

$$\alpha_o \equiv \alpha(z_o)$$

$$\beta_o \equiv \beta(z_o)$$

The boundary conditions have now been determined completely. The root of D(z) is found by standard methods, and b, $\sigma^1(z)$, and $\sigma^2(z)$ are then given by Eq.(III-13), Eq. (III-11a), and Eq. (III-11b) respectively. The only unknown in the expressions for $\pi^1(y,z)$ and $\pi^2(y,z)$ (Eqs. (III-4)) is π^1_{oo}. We now turn our attention to the determination of this quantity.

III-c Non-Saturation Condition

Proper normalization requires π^1_{oo} to be chosen so that

$$\pi^2(1,1) = r_2 \qquad \text{(III-14)}$$

This follows from Eq. (II-5b).

In light of Eq. (III-4b) this means

$$\pi^1_{oo} = \frac{r_2}{\sigma^1(1) - r_1 + b\,(\sigma^2(1) - r_1\,)} \qquad \text{(III-15)}$$

We may derive this equation in another way. The probability that a given service is followed by a changeover from 1 to 2 is

$$\sum_{n=1}^{\infty} \pi^1_{on} + \pi^1_{oo} r_2 = \sum_{n=o}^{\infty} \pi^1_{on} - \pi^1_{oo} r_1$$

$$= \pi^1_{oo}(\,\sigma^1(1) - r_1) \qquad \text{(III-16)}$$

This follows from the fact that a service will be followed by a switch from 1 to 2 only if it leaves the system in one of the states $\overset{1}{on}$, $n \geq 1$, or if it leaves the system in the state $\overset{1}{o\,o}$

and the next customer to arrive is of type 2. The probability that a given service is followed by a changeover from 2 to 1 is

$$\sum_{m=1}^{\infty}\sum_{n=o}^{\infty} \pi^2_{mn} + \pi^1_{oo} r_1 = r_2 \sum_{n=o}^{\infty} \pi^2_{on} + \pi^2_{oo}\, r_2$$

$$= r_2 + br_2 \pi^1_{oo} - b\pi^1_{oo}\sigma^2(1) \qquad \text{(III-17)}$$

since a changeover from 2 to 1 is made only if the system has been left in one of the states $\overset{2}{mn}$, $m \geq 1$, or if it is left in the state $\overset{2}{o\,o}$ and the next arrival is of type 1. The number of changeovers in either direction must be the same. Equating the probabilities of Eqs. (III-16) and (III-17) produces Eq. (III-15) and sheds some light on its significance.

One observes that

$$\sigma^1(1) - r_1 = \lim_{z\to 1} \frac{\sigma^1(z) + zr_2 - 1}{z} = \lim_{z\to 1} \frac{A_1(z)}{D(z)} \qquad \text{(III-18a)}$$

$$b(\sigma^2(1) - r_1) = \lim_{z\to 1} b\,\frac{\sigma^2(z) + zr_2 - 1}{z} = \lim_{z\to 1} \frac{A_2(z)}{D(z)} \qquad \text{(III-18b)}$$

Since $A_1(1) = A_2(1) = D(1) = 0$, L' Hopital's rule must be applied.

Thus

$$\sigma^1(1) - r_1 = \frac{A_1'(1)}{D'(1)} \qquad \text{(III-19a)}$$

$$b(\sigma^2(1)-r_1)=\frac{A'_2(1)}{D'(1)} \qquad \text{(III-19b)}$$

One finds that

$$A_1'(1) = \frac{1}{1-\rho_1}\left[\, r_2(1+b)(1-S_2(\lambda_1)\,)\,+r_1 b(1-\rho_1-\rho_2)\,\right] \qquad \text{(III-20a)}$$

$$D'(1)= \frac{1}{1-\rho_1}\left[\,(1-S_2(\lambda_1)+S_2(\lambda_1)S_{12}(\lambda_1))(1-\rho_1-\rho_2)-(1-S_2(\lambda_1))\,\lambda_2\left(\frac{1}{\mu_{12}}+\frac{1}{\mu_{21}}\right)\,\right]$$

(III-20b)

$$A_2'(1) = \frac{1}{1-\rho_1}\left[\, S_{12}(\lambda_1)S_2(\lambda_1)(1+b)r_2-br_1(1-\rho_1-\rho_2)+br_1\lambda_2\left(\frac{1}{\mu_{12}}+\frac{1}{\mu_{21}}\right)\right]$$

(III-20c)

so that there results from Eqs. (20), Eqs. (19), and Eq. (15)

$$\pi_{oo}^1= \frac{r_2D'(1)}{A_1'(1)+A_2'(1)} = \frac{(1-S_2(\lambda_1)+S_2(\lambda_1)S_{12}(\lambda_1))(1-\rho_1-\rho_2)-(1-S_2(\lambda_1))\lambda_2\left(\frac{1}{\mu_{12}}+\frac{1}{\mu_{21}}\right)}{(1-S_2(\lambda_1)+S_2(\lambda_1)S_{12}(\lambda_1))(1+b)+ b\lambda_1\left(\frac{1}{\mu_{12}}+\frac{1}{\mu_{21}}\right)}$$

(III-21)

The system will operate without saturation as long as $\pi^1_{oo} + \pi^2_{oo} = (1+b)\pi^1_{oo} > o$. Since $A'_1(1) + A'_2(1)$ is always positive (see Eq. III-21)) this condition becomes $\Gamma'(1) > 0$.

$$1-\rho_1-\rho_2 - \lambda_2 \frac{1-S_2(\lambda_1)}{1-S_2(\lambda_1)+S_2(\lambda_1)S_{12}(\lambda_1)} \left(\frac{1}{\mu_{12}} + \frac{1}{\mu_{21}} \right) > o$$

(III-22)

This is the non-saturation condition.

The significance of the last term in the above inequality may be better understood by obtaining it in another fashion. Let us consider the period of time that elapses between the instant at which a changeover from 1 to 2 begins and that at which a changeover from 2 to 1 terminates. In this interval the server crosses to line 2, serves some number of 2-customers (at least one), and returns to line 1. We shall be interested in the number of 2-customers served in this "trip." For this purpose we note

$$S_2(\lambda_1) = \int_o^\infty e^{-\lambda_1 t} \, dF_{S_2}(t) = \text{Prob(exactly zero 1-customers arrive during service of a 2-customer)}$$

$$\equiv q_o \qquad \text{(III-23a)}$$

$$S_{12}(\lambda_1) = \int_o^\infty e^{-\lambda_1 t} \, dF_{S_{12}}(t) = \text{Prob(exactly zero 1-customers arrive during a changeover from 1 to 2)}$$

$$\equiv v_o \qquad \text{(III-23b)}$$

Since we are concerned with near-saturation conditions, we shall neglect the possibility that the server spends idle time at line 2 during his trip. When the server departs from line 1, no 1-customers are present. If a 1-customer arrives either during the changeover from 1 to 2 or during service of the first 2-customer, the server will return to line 1 as soon as the service is completed. In this case only one 2-customer is served. Thus

Prob (exactly one 2-customer served per trip) = $1-q_o v_o$

Exactly two 2-customers will be served if no 1-customers arrive during the changeover time from 1 to 2, and none arrive during the service time of the first 2-customer, but that at least one 1-customer arrives during the service of the second 2-customer. Thus

Prob(serve exactly two 2-customers before returning to line 1)

$$= v_o q_o (1-q_o)$$

We can likewise find the probability that exactly n-customers are served in succession

Prob(exactly n 2-customers served per trip)

$$= v_o q_o^{\,n-1} (1-q_o) \qquad n \geq 2$$

The average number of 2-customers served per trip then is

$$1-q_o v_o + \sum_{n=2}^{\infty} n\, v_o q_o^{\,n-1}(1-q_o) = \frac{1-q_o+q_o v_o}{1-q_o} \qquad \text{(III-24)}$$

The reciprocal of this is the average number of trips made per 2-customer. Since 2-customers arrive at a rate λ_2 per unit time and each trip requires an average crossing time of $\frac{1}{\mu_{12}} + \frac{1}{\mu_{21}}$,

the server spends a fraction of his time

$$\lambda_2 \frac{1-q_o}{1-q_o+q_o v_o}\left(\frac{1}{\mu_{12}} + \frac{1}{\mu_{21}}\right) = \frac{1-S_2(\lambda_1)}{1-S_2(\lambda_1)+S_2(\lambda_1)S_{12}(\lambda_1)}\left(\frac{1}{\mu_{12}} + \frac{1}{\mu_{21}}\right)$$

(III-25)

crossing between lines. To have non-saturation, the total portion of the server's time servicing 1-customers, serving 2-customers, and crossing must be less than 100 per cent. This explains Eq. (III-22).

III-d Waiting Times and Other Quantities

The Laplace-Steiltjes transform for the high priority waiting time is obtained by substitution of Eq. (III-4a) into Eq. (I-16a).

$$W_1(s) = \frac{\pi^1_{oo}}{r_1} \; \frac{(\sigma^1(1)-r_1)(S_{21}(s)S_{12}(s)S_2(s)-1)+b(\sigma^2(1)-r_1)S_{21}(s)(S_2(s)-1) + \frac{r_1 s}{\lambda_1}(1+bS_{21}(s))}{1-\frac{s}{\lambda_1}-S_1(s)}$$

(III-26)

The mean wait for 1-customers may be obtained by two applications of L'Hopital's rule to the derivative of the above expression. The result is

$$\overline{W}_1 = \frac{\lambda_1 E(S_1^2)+\lambda_2 E(S_2^2)}{2(1-\rho_1)} + \frac{b\pi^1_{oo}+\rho_2}{(1-\rho_1)\mu_{21}} + \frac{(\lambda_1+\lambda_2)\pi^1_{oo}(\sigma^1(1)-r_1)}{(1-\rho_1)}$$

(III-27)

$$\times \left[\frac{E(S_{12}^2)+E(S_{21}^2)}{2} + \frac{1}{\mu_2\mu_{12}} + \frac{1}{\mu_{12}\mu_{21}} \right]$$

The quantity $\sigma^1(1)-r_1$ is given in Eq. (III-19a), b is given in Eq. (III-13), and π^1_{oo} is given in Eq. (III-21).

Using a technique introduced by Cobham,[2] and refined by Holley,[5] this equation may be rederived: An arriving 1-customer (high priority) has a wait which can be broken into two components:

that contributed by the delay, if any, between the time of arrival and the time at which the server is prepared to attend to the next 1-customer ; and that contributed by the delay involved in servicing all those 1-customers already waiting in line. We call the first of these W_0, and deal with it below. The second component can be obtained by observing that the average number of 1-customers in line 1, not including the one which may be in service, is $\lambda_1 \bar{W}_1$ (7). Since each customer requires an average service time of $1/\mu_1$ the total average time required for their service is $\frac{\lambda_1 \bar{W}_1}{\mu_1}$

From this it follows that

$$\bar{W}_1 = W_o + \rho_1 \bar{W}_1 \qquad \text{(III-28)}$$

and thus

$$\bar{W}_1 = \frac{W_o}{1-\rho_1} \qquad \text{(III-29)}$$

Now we consider the quantity W_o. It is well known that in a renewal process, the mean time between a random instant and the occurrence of the next event is $\frac{\mu E(S^2)}{2}$, where $1/\mu$ is the mean and $E(S^2)$ is the second moment of the inter-event times. (2) An arriving 1-customer enters the system at a random instant and may find the server in any one of six situations:

1) The server may be idle at line 1. The probability of this is π^1_{oo} and in this case service can begin immediately.

2) The server may be idle at line 2, and this occurs with probability $\pi^2_{oo} = b\pi^1_{oo}$. An average delay of $\frac{1}{\mu_{21}}$ is required before the server is prepared to service the 1-arrival.

3) The server may be busy at line 1, which has a probability ρ_1. An average time $\frac{\mu_1 E(S_1^2)}{2}$ is required before the service of the next 1-customer can begin. 4) The server may be busy at line 2. The probability of this is ρ_2 and an average delay of $\frac{\mu_2 E(S_2^2)}{2} + \frac{1}{\mu_{21}}$ is required. 5) The server may be crossing from 2 to 1. By the same reasoning leading to Eq. (II-49), the probability of this is $\frac{(\lambda_1+\lambda_2)\pi^1_{oo}(\sigma^1(1)-r_1)}{\mu_{21}}$. The average delay involved in this case is $\frac{\mu_{21}E(S_{21}^2)}{2}$. 6) The server may be crossing from 1 to 2, the probability being $\frac{(\lambda_1+\lambda_2)\pi^1_{oo}(\sigma^1(1)-r_1)}{\mu_{12}}$. The average delay is $\frac{\mu_{12}E(S_{12}^2)}{2} + \frac{1}{\mu_2} + \frac{1}{\mu_{21}}$. Thus weighing the delays in each case by their probabilities,

$$W_o = \frac{\lambda_1 E(S_1^2) + \lambda_2 E(S_2^2)}{2} + \frac{b\pi^1_{oo} + \rho_2}{\mu_{21}} + (\lambda_1+\lambda_2)\pi^1_{oo}(\sigma^1(1)-r_1) \times \left(\frac{E(S_{12}^2)+E(S_{21}^2)}{2} + \frac{1}{\mu_{21}\mu_2} + \frac{1}{\mu_{12}\mu_{21}} \right) \tag{III-30}$$

This, in conjunction with Eq. (II-29) gives Eq. (III-27).

For the low priority customers, the waiting time transform is

$$W_2(s) = \frac{\pi^1_{oo}}{r_2(1-\frac{s}{\lambda_2})} \left[S_{12}(s)(\sigma^1(1-\frac{s}{\lambda_2})- \frac{r_2 s}{\lambda_2} -r_1)+b(\sigma^2(1- \frac{s}{\lambda_2})-\frac{r_2 s}{\lambda_2}-r_1)\right]$$

(III-31)

The mean wait is

$$\overline{W}_2 = \frac{\pi^1_{oo}}{r_2} \left[\frac{\sigma^1(1)'+r_2-(\sigma^1(1)-r_1)}{\lambda_2} +b \frac{\sigma^2(1)+r_2-(\sigma^2(1)-r_1)}{\lambda_2} + \frac{\sigma^1(1)-r_1}{\mu_{12}} \right]$$

(III-32)

One notes that

$$\sigma^1(1)' +r_2-(\sigma^1(1)-r_1) = \frac{d}{dz} \left. \frac{\sigma^1(z)+r_2 z-1}{z} \right|_{z=1} = \frac{d}{dz} \left. \frac{A_1(z)}{D(z)} \right|_{z=1}$$

(III-33a)

$$b\,[\sigma^2(1)'+r_2-(\sigma^2(1)-r_1)]= \frac{d}{dz} \left. b\,\frac{\sigma^2(z)+r_2 z-1}{z} \right|_{z=1} = \frac{d}{dz} \left. \frac{A_2(z)}{D(z)} \right|_{z=1}$$

(III-33b)

The expressions become quite complicated so they are omitted here. For completeness the expression for $\bar{W}_2$ is included in Appendix V. No simple interpretation of the result for $\bar{W}_2$ is evident to the author.

Limiting Cases

1) In the case $\mu_{12}=\mu_{21}=\infty$, the expressions for the mean waiting time become

$$\bar{W}_1 = \frac{\lambda_1 E(S_1^2)+\lambda_2 E(S_2^2)}{(1-\rho_1)} \qquad \text{(III-34a)}$$

$$\bar{W}_2 = \frac{\lambda_1 E(S_1^2)+\lambda_2 E(S_2^2)}{(1-\rho_1)(1-\rho_1-\rho_2)} \qquad \text{(III-34b)}$$

The probability that the server is idle is

$$\pi_{oo}^1 + \pi_{oo}^2 = 1-\rho_1-\rho_2 \qquad \text{(III-35)}$$

These are the rather well-known results for a simple priority queue and were first derived by Cobham [(2)]. See also Miller [(10)]. Eqs. (III-7) take on a particularly simple form in this case, and the value of z_o, defined in Section III-b, is zero. The constant b is then

$$b= \frac{1}{\frac{S_2(\lambda_1-\lambda_1 B_1(\lambda_2)+\lambda_2)}{S_2(\lambda_1+\lambda_2)(1-r_1 B_1(\lambda_2))}-1} \qquad \text{(III-36)}$$

2) The case $\mu_1=\mu_2=\infty$ is not of interest because the strict priority discipline forces the server to wait on not more than one 2-customer if a 1-customer is waiting. The waiting times could certainly be reduced by allowing service to take place on all 2-customers present, since this takes no more time than the service of a single 2-customer. Thus strict priority is not an acceptable policy in this situation.

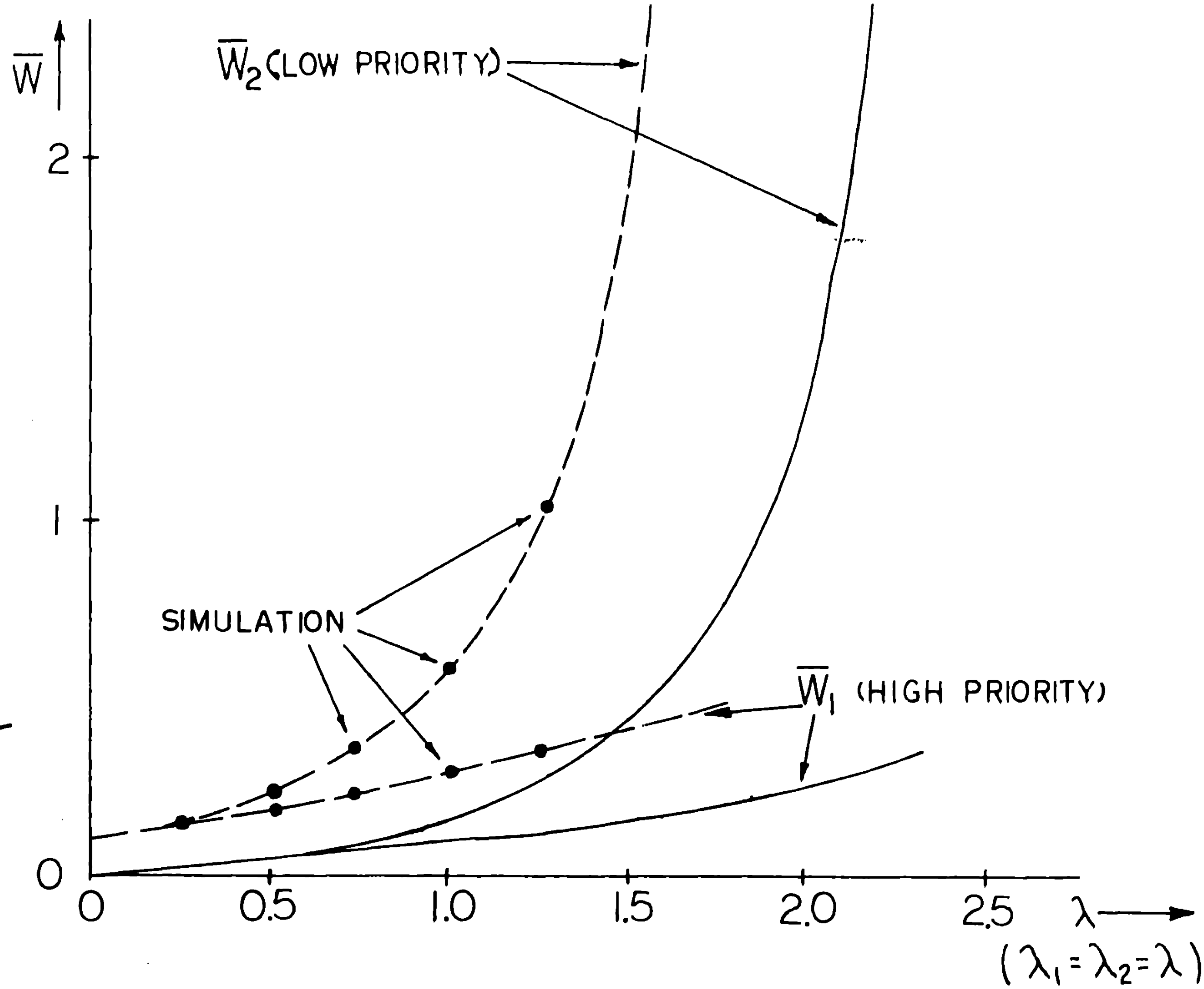
End of Limiting Cases

If a cost is associated with changeovers we would be interested in the following quantities: Probability that a 1-customer is followed by a changeover $= \frac{\pi^1_{oo}(\sigma^1(1)-r_1)}{r_1}$. Probability that a 2-customer is followed by a changeover $= \frac{\pi^1_{oo}(\sigma^1(1)-r_1)}{r_2}$. Probability that an arbitrary customer is followed by a changeover $= 2\pi^1_{oo}(\sigma^1(1)-r_1)$.

Fig. 3 plots the waiting time of the high priority and low priority customers versus arrival rates for the cases of no changeover time and a fairly large changeover time. The low priority customer suffers most of the added delay when the changeover time is increased. Superimposed on the plot are the results of some simulation runs.

III-e Summary

The study of the imbedded process has resulted in the complete solution for the transform of the waiting-time distributions. This was possible due to the simple form of the equations which specified the boundary conditions.



———— $\frac{1}{\mu_{12}} = \frac{1}{\mu_{21}} = 0$ $\frac{1}{\mu_1} = \frac{1}{\mu_2} = .2$

– – – – $\frac{1}{\mu_{12}} = \frac{1}{\mu_{21}} = .2$ EXPONENTIALLY DISTRIBUTED SERVICE AND CHANGEOVER

PLOT OF MEAN WAITING TIMES vs. ARRIVAL RATE FOR ZERO AND FINITE CHANGEOVER TIMES — STRICT PRIORITY —

FIG. 3

CHAPTER IV

ZERO SERVICE TIMES

The assumption is made throughout this Chapter that $\mu_1=\mu_2=\infty$ i.e. , the service times are zero. This is done so that we may study the effect of more sophisticated queue disciplines than we have studied previously without at the same time adding tremendous complexity to the analysis. The validity of the assumption in actual practice depends on the extent to which the service times are small with respect to both the changeover times and the inter-arrival times.

We consider the 2-line model of Fig. 1 with $F_{S_1}(t) = F_{S_2}(t)=\mu_{-1}(t)$, the unit step function. We restrict attention to those queue disciplines which can be described as follows: 1) All customers in a given line are served immediately upon arrival of the server to that line.
2) Unless the number of waiting customers in the other line is greater than some critical value, the server will remain stationary ; the moment this critical value is exceeded, a changeover is begun. To be specific, the server will remain at line 2 until at least M customers are present at line 1. Likewise, he will remain at line 1 until at least N customers are present at line 2. One notes that there are never any waiting customers at the line which is receiving service, since the assumption of zero service times allows these customers to be served immediately.

The optimum policy-- in the sense of minimum mean waiting time of an arbitrary customer-- falls within this set of disciplines. It is certainly true that the optimum discipline always clears all customers in a given line since this takes zero time and the waiting times could only be increased by refusing to service some of those who are present. Also, the fact that arrivals are Poisson means that any information, other than the number of waiting customers in the other line, is irrelevant to the decision of when to cross. One notes that setting $M=N=1$ here gives for the case of zero service time, the alternating priority discipline discussed in Chapter II.

We again employ the technique of working with the process imbedded at the instants of service completion. We find the boundary conditions to have a more complicated form than previously, but additional considerations lead to their complete determination. Waiting times and other properties of the system are then obtained.

IV-a The Imbedded Process

The state of the imbedded process is again either $\overset{1}{mn}$ or $\overset{2}{mn}$ and the transition probabilities for above discipline are listed below:

1 → 1 TRANSITIONS

$m\leq$, all n

$$P(\overset{1}{mn} \to \overset{1}{m'n'}) = \delta_{m',m-1}\,\delta_{n'n}$$

$m=o,\ n\leq n$

$$P(o\,n \to m'n') = 0$$

m=o, n<N

$$P(\overset{1}{o}\;n \rightarrow \overset{1}{m'}n') = \delta_{om'}[r_1\delta_{nn'} + r_2 r_1\,\delta_{n+1,n'} + r_2^2 r_1\,\delta_{n+2,n'} + \dots\dots + r_2^{N-n-1} r_1 \delta_{N-1,n'}]$$

1→2 TRANSITIONS

m≤1, all n

$$P(\overset{1}{m}\;n \rightarrow \overset{2}{m'}n') = 0$$

m=o, n≤N

$$P(\overset{1}{o}\;n \rightarrow \overset{2}{m'}n') = \begin{cases} 0\ , & n' < n-1 \\ v_{m',n'-n+1}, & \text{otherwise} \end{cases}$$

m=o, n<N

$$P(\overset{1}{o}\;n \rightarrow \overset{2}{m'}n') = \begin{cases} 0\ , & n' < N-1 \\ r_2^{N-n}\ v_{m',\,n'-N+1}, & \text{otherwise} \end{cases}$$

2 → 2 TRANSITIONS

n>o, all m

$$P(\overset{2}{m}\;n \rightarrow \overset{2}{m'}n') = \delta_{m'm}\ \delta_{n',n-1}$$

n=o, m≤M

$$P(\overset{2}{m}\;o \rightarrow \overset{2}{m'}n') = 0$$

n=o, m<M

$$P(\overset{2}{m}\,\overset{2}{o} \to m'n') = \delta_{on'}\,[r_2\delta_{mm'} + r_1r_2\,\delta_{m+1,m'} + r_1^2\,r_2\delta_{m+2,m'} + \ldots\ldots + r_1^{M-m-1}\,r_2\,\delta_{M-1,m}]$$

2 → 1 TRANSITIONS

n>o, all m

$$P(\overset{2}{m}\,n \to \overset{1}{m'}n') = 0$$

n=o, m≤M

$$P(\overset{2}{m}\,o \to \overset{1}{m'}n') = \begin{cases} 0 & m<m-1 \\ u_{m'-m+1,n'} & \end{cases}$$

n=o, m<M

$$P(\overset{2}{m}\,o \to \overset{1}{m'}n') = \begin{cases} 0 & m'< M-1 \\ r_1^{M-m} & u_{m'-M+1,n'} \end{cases}$$

The quantity $\delta_{ij} = \begin{cases} 1 & i=j \\ o & i\neq j \end{cases}$ is the kroneker delta function of its arguments i and j.

Let us examine some of the above expressions. If the present state of the system is ${}^{1}_{m\,n}$, $m \leq 1$ then when the next customer is served, the state of the system must be ${}^{1}_{m-1\,n}$. This is so because service times are zero and no arrivals could have occurred during the service. This explains the expression for $P({}^{1}_{m\,n} \rightarrow {}^{1}_{m'\,n'})$. If the present state is ${}^{1}_{o\,n}$, $n < N$, the next state (at a service-completion) depends upon the order of future arrivals. If the next arrival is of type 1, he will be served immediately and the resulting state will again be ${}^{1}_{o\,n}$. If a 2-arrival occurs, followed by a 1-arrival, the state becomes ${}^{1}_{o\,n+1}$. As many as N-n-1 2-arrivals might occur, followed by a 1-arrival. In this case the next state is ${}^{1}_{o\,N-1}$. If more than N-n-1 2-arrivals occur in succession, the server will move to line 2. These possibilities are reflected in the expressions for $P({}^{1}_{o\,n} \rightarrow {}^{1}_{m'n'})$, $n<N$ and $P({}^{1}_{o\,n} \rightarrow {}^{2}_{m'n'})$, $n<N$. (It should be remembered that the probability an arrival is of type 1 or 2 is r_1 or r_2 respectively.)

Substitution of the above transition probabilities into the equilibrium equations (Eqs. (II-3)) allows us to find the generating functions of the steady-state probabilities. The result is

$$\pi^1(y,z) = \frac{\pi^1_{oo}}{y-1}\Big[\frac{yr_1}{1-r_2 z}\left(\sigma^1_N(z) - z^N r_2^N \sigma^1_N\left(\frac{1}{r_2}\right)\right)$$

$$+ bU(y,z)\left(\sigma^2(y)-\sigma^2_M(y)+y^M r_1^M \sigma^2_M\left(\frac{1}{r_1}\right)\right)-\sigma^1(z)\Big] \qquad \text{(IV-1a)}$$

$$\pi^2(y,z) = \frac{\pi^1_{oo}}{z-1}\Big[\frac{bzr_2}{1-r_1y}\Big(\sigma^2_M(y) - y^M r_1{}^M \sigma^2_M\Big(\frac{1}{r_1}\Big)$$

$$+ V(y,z)\Big(\sigma^1(z) - \sigma^1_N(z) + z^N r^N_2\, \sigma^1_N\Big(\frac{1}{r_2}\Big)\Big) - b\sigma^2(y)\Big]$$

(IV-1b)

where we have used the following

$$b \equiv \frac{\pi^2_{oo}}{\pi^1_{oo}} \qquad \text{(IV-2)}$$

$$\sigma^1(z) \equiv \frac{\pi^1(o,z)}{\pi^1_{oo}} = \frac{1}{\pi^1_{oo}} \sum_{n=o}^{\infty} \pi^1_{on}\, z^n \qquad \text{(IV-3a)}$$

$$b\sigma^2(y) \equiv \frac{\pi^2(y,o)}{\pi^1_{oo}} = \frac{1}{\pi^1_{oo}} \sum_{m=o}^{\infty} \pi^2_{mo} y^m \qquad \text{(IV-3b)}$$

$$\sigma^1_N(z) = \frac{1}{\pi^1_{oo}} \sum_{n=o}^{N-1} \pi^1_{on} z^n \qquad \text{(IV-4a)}$$

$$b\sigma^2_M(y) = \frac{1}{\pi^1_{oo}} \sum_{m=o}^{M-1} \pi^2_{mo} y^m \qquad \text{(IV-4b)}$$

$$\sigma^1(o) = \sigma^2(o) = \sigma^1_N(o) = \sigma^2_M(o) = 1 \qquad \text{(IV-5)}$$

The quantities U(y, z) and V(y, z) are the generating functions for the probabilities u_{ij} and v_{ij} defined in Eqs. (II-1) and (II-2). The quantities $\sigma^1(z)$, $\sigma^2(y)$, and b have the same interpretation that they had in Chapter II. The functions $\sigma^1_N(z)$ and $\sigma^2_M(y)$ are the power series in z and y for $\sigma^1(z)$ and $\sigma^2(y)$ truncated after the Nth and Mth terms respectively.

We now give a preview of the steps that will be taken in the next few sections. $\pi^1(y, z)$ and $\pi^2(y, z)$, the generating functions which we seek, are expressed in Eqs. (IV-1) in terms of $\sigma^1(z)$, $\sigma^2(y)$, $\sigma^1_N(z)$, $\sigma^2_M(y)$, b and π^1_{oo} in addition to the known quantities. In section IV-b we derive equations which express $\sigma^1(z)$ and $\sigma^2(y)$ in terms of $\sigma^1_N(z)$ and $\sigma^2_M(y)$. In Section IV-c we derive equations which determine $\sigma^1_N(z)$ and $\sigma^2_M(y)$ in terms of b. These equations are solved and upon examination of the results, it is seen that b is uniquely specified. The outcome of this section is a complete knowledge of the quantities $\sigma^1_N(z)$, $\sigma^2_M(y)$, and b. In Section IV-d the last remaining unknown, π^1_{oo}, is determined from the condition of normalization.

IV-b Determination of $\sigma^1(z)$ and $\sigma^2(y)$

The purpose of this section is the determination of the functions $\sigma^1(z)$ and $\sigma^2(y)$. Equations for these quantities may be obtained by noticing that the numerators of Eqs. (IV-1) must be zero, when y = 1 and z = 1 since the denominators are zero at this point. Thus

$$\sigma^1(z) = \frac{r_1}{1-r_2 z}\left(\sigma^1_N(z) - z^N r_2^N \sigma^1_N\left(\frac{1}{r_2}\right)\right) + bU(z)\left(\sigma^2(1) - \sigma^2_M(1) + r_1^M \sigma^2_M\left(\frac{1}{r_1}\right)\right) \quad \text{(IV-6a)}$$

$$b\sigma^2(y) = \frac{br_2}{1-r_1 y}\left(\sigma^2_M(y) - y^M r_1^M \sigma^2_M\left(\frac{1}{r_1}\right)\right) + V(y)\left(\sigma^1(1) - \sigma^1_N(1) + r_2^N \sigma^1_N\left(\frac{1}{r_2}\right)\right) \quad \text{(IV 6b)}$$

where we have used the notation

$$U(z) \equiv U(1,z) = \sum_{i=o}^{\infty}\sum_{j=o}^{\infty} u_{ij} z^j \quad \text{(IV-7a)}$$

$$V(y) \equiv V(y,1) = \sum_{i=o}^{\infty}\sum_{j=o}^{\infty} v_{ij} y^i \quad \text{(IV-7b)}$$

It is convenient to define the quantities

$$u_j \equiv \sum_{i=o}^{\infty} u_{ij} \quad = \text{Prob (j 2-customers arrive during a changeover from 2 to 1)} \quad \text{(IV-8a)}$$

$$v_i \equiv \sum_{j=o}^{\infty} v_{ij} \quad = \text{Prob (i 1-customers arrive during a changeover from 1 to 2)} \qquad \text{(IV-8b)}$$

U(z) and V(y) then are the generating functions for the probabilities defined above.

One observes that Eqs. (IV-6) express $\sigma^1(z)$ and $\sigma^2(y)$ in terms of $\sigma^1(1)$ and $\sigma^2(1)$ as well as $\sigma^1_N(z)$, $\sigma^2_M(y)$, and b. The quantities $\sigma^1(1)$ and $\sigma^2(1)$ may be eliminated by requiring the expressions for $\sigma^1(z)$ and $\sigma^2(y)$ to be equal to 1 when y = z = o. It follows that

$$\sigma^1(z) = \frac{r_1}{1-r_2 z} \left(\sigma^1_N(z) - z^N r_2^N \sigma^1_N \left(\frac{1}{r_2}\right)\right) + U(z) \frac{r_2}{u_o} \qquad \text{(IV-9a)}$$

$$\sigma^2(y) = \frac{r_2}{1-r_1 y} \left(\sigma^2_M(y) - y^M r_1^M \sigma^2_M\left(\frac{1}{r_1}\right)\right) + V(y) \frac{r_1}{v_o} \qquad \text{(IV-9b)}$$

These equations express $\sigma^1(z)$ and $\sigma^2(y)$ completely in terms of $\sigma^1_N(z)$ and $\sigma^2_M(y)$. Thus the former functions can easily be determined once the latter have been found.

IV-c Determination of $\overset{1}{\sigma}_N(z)$, $\overset{2}{\sigma}_M(y)$ and b

The purpose of this section is the determination of the functions $\overset{1}{\sigma}_N(z)$ and $\overset{2}{\sigma}_M(y)$ as well as the constant b. Progress toward this end could no doubt be made by taking advantage of the fact in Eqs. (IV-9) that $\overset{1}{\sigma}(z)$ and $\overset{1}{\sigma}_N(z)$ agree up to the (N-1)st power of z, and that $\overset{2}{\sigma}(y)$ and $\overset{2}{\sigma}_M(y)$ agree to the (M-1)st power of y. However, it seems easier in this case to use the following method.

We make use of a more deeply imbedded process. This new process is formed by taking note of the system <u>only</u> at that subset of service-completion instants which cause a line to become empty but do not cause a changeover to begin. That is, we ignore all service-completion instants except those that leave the system in one of the states $\overset{1}{o\,n}$ with $n < N$ or $\overset{2}{m\,o}$ with $m < M$ (a total of M+N states). We use a prime (') on the transition probabilities to remind us that only those states mentioned above are being considered. The transition probabilities are listed below.

1 → 1 TRANSITIONS

$$P'(\overset{1}{o\,n} \rightarrow \overset{1}{o\,n'}) = r_1\delta_{nn'} + r_2r_1\delta_{n+1,n'} + r_2^2r_1\delta_{n+2,n'} + \cdots + r_2^{N-n-1}r_1\delta_{N-1,n'}$$

$$+ r_2^{N-n}\,\bar{V}_M u_{n'}\left(1 + \bar{U}_N\bar{V}_M + (\bar{U}_N\bar{V}_M)^2 + (\bar{U}_N\bar{V}_M)^3 + \cdots\right)$$

1 → 2 TRANSITIONS

$$P'(\overset{1}{o}\, n \rightarrow m'\overset{2}{o}) = r_2^{N-n} v_{m'}(1+(\bar{V}_M \bar{U}_N) + (\bar{V}_M \bar{U}_N)^2 + \ldots)$$

2 → 2 TRANSITIONS

$$P'(m\,\overset{2}{o} \rightarrow m'\overset{2}{o}) = r_2 \delta_{mm'} + r_1 r_2\, \delta_{m+1,m'} + r_1^2\, r_2 \delta_{m+2,m'} + \ldots + r_1^{M-m-1} r_2 \delta_{M-1,m'}$$

$$+ r_1^{M-m} \bar{U}_N V_{m'}\, (1+ \bar{V}_M \bar{U}_N + (\bar{V}_M \bar{U}_N)^2 + \ldots)$$

2 → 1 TRANSITIONS

$$P'(m\,\overset{2}{o} \rightarrow \overset{1}{o}n') = r_1^{M-m} u_{n'}(1+ \bar{U}_N \bar{V}_M + (\bar{U}_N \bar{V}_M)^2 + \ldots)$$

where

$\bar{U}_N$ = Prob(N or more 2-customers arrive during a changeover from 2 to 1)

$$= \sum_{j=N}^{\infty} u_j \qquad \text{(IV-10a)}$$

$\bar{V}_M$ = Prob(M or more 1-customers arrive during a changeover from 1 to 2

$$= \sum_{i=M}^{\infty} v_i$$

$$\delta_{ij} = \begin{cases} 1 & i=j \\ o & i \neq j \end{cases} \quad \text{kroneker delta}$$

The special form of these expressions is interesting. Suppose the present state is $\overset{1}{o\ n}$. If exactly j 2-customers arrive before the next 1-customer, where $j \leq N-n-1$, the next state will be $\overset{1}{o\ n+j}$. This explains the first part of the expression for $P'(\overset{1}{o\ n} \rightarrow \overset{1}{o\ n'})$. The rest of this expression may be understood by considering what happens in the event that N-n 2-customers arrive in succession.

If the state is $\overset{1}{o\ n}$, and if N-n successive 2-arrivals occur, the server will cross to line 2. Depending upon the number of 1-customers that arrive during the changeover from 1 to 2, the server may or may not stop at line 2. If more than M 1-arrivals occur, the server will leave line 2 immediately. Likewise the server will leave line 1 immediately if N or more 2-customers arrive during the changeover from 2 to 1. The server may travel back and forth in this fashion many times. We are interested in the state of the system when the server finally stops.

The probability that line 2 is left immediately after a 1 to 2 changeover is $\bar{V}_M$. The probability that line 1 is left immediately after a 2 to 1 changeover is $\bar{U}_N$. The last part of the expression for $P'(\overset{1}{o\ n} \rightarrow \overset{1}{o\ n'})$ reflects the possibility that the server crosses to line 2 and back any number of times until finally a 2 to 1 crossing is made during which exactly n' 2-customers arrive. The expressions for $P'(\overset{1}{o}\ n \rightarrow \overset{2}{m'}o)$, $P'(m\overset{2}{\ }o \rightarrow \overset{1}{o}n')$, and $P'(m\overset{2}{\ }o \rightarrow \overset{2}{m'}o)$ were obtained through identical reasoning.

The steady-state probabilities must satisfy the following equilibrium equations

$$\pi^1_{on'} = \sum_{n=o}^{N-1} \pi^1_{on} P'(\overset{1}{o\,n} \rightarrow \overset{1}{o\,n'}) + \sum_{m=o}^{M-1} \pi^2_{mo} P'(\overset{2}{m\,o} \rightarrow \overset{1}{o\,n'}) \qquad n'<N$$

(IV-11a)

$$\pi^2_{m'o} = \sum_{n=o}^{N-1} \pi^1_{on} P'(\overset{1}{o\,n} \rightarrow \overset{2}{m'o}) + \sum_{m=o}^{M-1} \pi^2_{mo} P'(m^2 o \rightarrow m'^2 o) \qquad m< M$$

(IV-11b)

which serve to determine the ratios of the indicated probabilities. This is true since the <u>ratios</u> of the probabilities do not depend upon whether or not all service-completions are taken into consideration of just the subset we have chosen. Dividing Eqs. (IV-11) by π^1_{oo}, substituting the expressions for the transition probabilities, and taking generating functions produces the result

$$\sigma^1_N(z) = \frac{1}{r_2(1-z)} \left[\frac{(1-r_2 z)U_N(z)}{1-\bar{U}_N \bar{V}_M} \left(\bar{V}_M r_2^{\,N} \sigma^1_N\left(\frac{1}{r_2}\right) + b r_1^{\,M} \sigma^2_M\left(\frac{1}{r_1}\right) - z^N r_1 r_2^{\,N} \sigma^1_N\left(\frac{1}{r_2}\right)\right)\right]$$

(IV-12a)

$$b\sigma^2_M = \frac{1}{r_1(1-y)} \left[\frac{(1-r_1 y)V_M(y)}{1-\bar{U}_N \bar{V}_M} \left(b\bar{U}_N r_1^{\,M} \sigma^2_M\left(\frac{1}{r_1}\right) + r_2^{\,N} \sigma^1_N\left(\frac{1}{r_2}\right) - b y^M r_2 r_1^{\,M} \sigma^2_M\left(\frac{1}{r_1}\right)\right)\right]$$

(IV-12b)

where we have used

$$U_N(z) \equiv \sum_{j=o}^{N-1} u_j z^j \qquad \text{(IV-13a)}$$

$$V_M(y) \equiv \sum_{i=o}^{M-1} v_i y^i \qquad \text{(IV-13b)}$$

Note that $U_N(z)$ and $V_M(y)$ are merely truncations of $U(z)$ and $V(y)$ defined in Eqs. (IV-7). Note also that

$$U_N \equiv U_N(1) = 1-\bar{U}_N = \sum_{j=o}^{N-1} u_j \qquad \text{(IV-14a)}$$

= Prob(less than N 2-customers arrive during a changeover from 2 to 1)

$$V_M \equiv V_M(1) = 1-\bar{V}_M = \sum_{i=o}^{M-1} v_i \qquad \text{(IV-14b)}$$

=Prob (less than M 1-customers arrive during a changeover from 1 to 2)

Eqs. (IV-12) express $\sigma^1_N(z)$ and $\sigma^2_M(y)$ in terms of $\sigma^1_N(\frac{1}{r_2})$ and $\sigma^2_M(\frac{1}{r_1})$ as well as b. The quantities $\sigma^1_N(\frac{1}{r_2})$ and $\sigma^2_M(\frac{1}{r_1})$ may be eliminated by requiring the expressions for $\sigma^1_N(z)$ and $\sigma^2_M(y)$ to be equal to 1 when y = z = o. The outcome is

$$\sigma^1_N(z) = \frac{1}{1-z} \left[\frac{(1-r_2 z)\, U_N(z) + z^N r_1 \bar{U}_N}{u_o} - \frac{z^N r_1{}^2 b}{r_2 v_o} \right] \qquad \text{(IV-15a)}$$

$$\sigma^2_M(y) = \frac{1}{1-y} \left[\frac{(1-r_1 y)\, V_M(y) + y^M r_2 \bar{V}_M}{v_o} - \frac{y^M r_2{}^2}{b r_1 u_o} \right] \qquad \text{(IV-15b)}$$

which expresses $\sigma^1_N(z)$ and $\sigma^2_M(y)$ only in terms of b and known quantities.

It is now possible to determine b exactly by observing that the numerators of Eqs. (IV-15) must be equal to zero when y = z = 1. Using either Eq. (IV-15a) or Eq. (IV-15b) we find

$$b = \frac{r_2 v_o}{r_1 u_o} \qquad \text{(IV-16)}$$

One notes that

u_o = Prob (no 2-customers arrive during a changeover from 2 to 1)

$$= \int_o^\infty e^{-\lambda_2 t}\, dF_{S_{21}}(t) = S_{21}(\lambda_2) \qquad \text{(IV-17a)}$$

v_o = Prob (no 1-customers arrive during a changeover from 1 to 2)

$$= \int_o^\infty e^{-\lambda_1 t}\, dF_{S_{12}}(t) = S_{12}(\lambda_1) \qquad \text{(IV-17b)}$$

so that

$$b = \frac{r_2 S_{12}(\lambda_1)}{r_1 S_{21}(\lambda_2)} \qquad \text{(IV-18)}$$

We originally found this same value of b when the alternating priority discipline was used and $\mu_1 = \mu_2 = \infty$. See Eq. (II-25). b is the ratio of the time the system spends empty with the server at line 2, to the time the system spends empty with the server at line 1. It is interesting to find that b remains unchanged for a much wider class of disciplines.

Using the expression of Eq. (III-16) in place of b in Eqs. (IV-15) there results

$$\sigma^1_N(z) = \frac{1}{1-z}\left[\frac{(1-r_2 z)U_N(z) - z^N r_1 U_N}{u_o}\right] \qquad \text{(IV-19a)}$$

$$\sigma^1_M(y) = \frac{1}{1-y}\left[\frac{(1-r_1 y)\, V_M(y) - y^M r_2 V_M}{v_o}\right] \qquad \text{(IV-19b)}$$

which gives $\sigma^1_N(z)$ and $\sigma^2_M(y)$ completely in terms of known quantities.

The important results of this section are Eqs. (IV-19) and Eq. (IV-16). We have obtained these equations through the analysis of a more deeply imbedded process. These results may now be used in conjunction with Eqs. (IV-9), to specify the behavior of $\pi^1(y, z)$ and $\pi^2(y, z)$ as given in Eqs. (IV-1). Only the quantity π^1_{oo} remains to be determined, and this is done by using the proper normalization condition.

IV-d Non-Saturation Condition

Proper normalization requires

$$\pi^{1}(1,1) = r_{1} \tag{IV-20}$$

This follows from Eq. (5a) of Chapter II.

One finds that L'Hopital's rule must be used in applying the above condition to Eq. (IV-20). One obtains

$$\pi^{1}_{oo} = \frac{u_o/\lambda_2}{\frac{1}{\lambda_1}\sum_{i=o}^{M-1}(M-i)v_i + \frac{1}{\lambda_2}\sum_{j=o}^{N-1}(N-j)u_j + \frac{1}{\mu_{12}} + \frac{1}{\mu_{21}}} \tag{IV-21}$$

Saturation can never take place since it is always true that π^{1}_{oo} is greater than zero. It is obvious that this must be the case since all customers in a given line can be cleared from the system at once, and a given customer need never wait longer than the time required for two changeovers.

IV-e Waiting Times and Other Quantities

The generating functions of Eqs. (IV-1) can be used in the expression of Eq. (I-16) to determine the Laplace-Steiltjes transform of the waiting times. Taking the appropriate derivatives, the mean waiting time may be found. The results are

$$\bar{W}_1 = \frac{\frac{1}{\lambda_1\mu_{21}} \sum_{i=o}^{M-1} (M-i)v_i + \frac{1}{2\lambda_1^2} \sum_{i=o}^{M-1} [M(M-1)-i(i-1)]v_i + \frac{E(S_{12}^2)+E(S_{21}^2)}{2} + \frac{1}{\mu_{12}\mu_{21}}}{\frac{1}{\lambda_1} \sum_{i=o}^{M-1} (M-i)\, v_i + \frac{1}{\lambda_2} \sum_{j=o}^{N-1} (N-j)u_j + \frac{1}{\mu_{12}} + \frac{1}{\mu_{21}}} \tag{IV-22a}$$

$$\bar{W}_2 = \frac{\frac{1}{\lambda_2\mu_{12}} \sum_{j=o}^{N-1} (N-j)u_j + \frac{1}{2\lambda_2^2} \sum_{j=o}^{N-1} [N(N-1)-j(j-1)]\, u_j + \frac{E(S_{12}^2)+E(S_{21}^2)}{2} + \frac{1}{\mu_{12}\mu_{21}}}{\frac{1}{\lambda_1} \sum_{i=o}^{M=1} (M-i)v_i + \frac{1}{\lambda_2} \sum_{j=o}^{N-1} (N-j)\, u_j + \frac{1}{\mu_{12}} + \frac{1}{\mu_{21}}} \tag{IV-22b}$$

$$\bar{W} = r_1\bar{W}_1 + r_2\bar{W}_2 \tag{IV-23}$$

where, to repeat

u_j = Prob. (j 2-customers arrive during a changeover from 1 to 2)

$$= \int_o^\infty \frac{(\lambda_2 t)^j}{j!} e^{-\lambda_2 t} \, dF_{S_{12}}(t) \tag{IV-24a}$$

v_i = Prob (i 1-customers arrive during a changeover from 2 to 1)

$$= \int_o^\infty \frac{(\lambda_1 t)^i}{i!} e^{-\lambda_1 t} \, dF_{S_{21}}(t) \tag{IV-24b}$$

One can determine the optimum policy by treating Eq. (IV-23) as a function of M and N and determining by a direct search which values of M and N produce the minimum average waiting time.

If we set M = N = 1 in the above, we find

$$\overline{W}_1 = \frac{\frac{1}{\lambda_1\mu_{21}} v_o + \frac{1}{\mu_{12}\mu_{21}} + \frac{E(S_{12}^2) + E(S_{21}^2)}{2}}{\frac{1}{\lambda_1} v_o + \frac{1}{\lambda_2} u_o + \frac{1}{\mu_{12}} + \frac{1}{\mu_{21}}} \qquad \text{(IV-25)}$$

which in light of Eq. (IV-17) becomes

$$\overline{W}_1 = \frac{\lambda_1\lambda_2}{\lambda_2 S_{12}(\lambda_1) + \lambda_1 S_{21}(\lambda_2) + \lambda_1\lambda_2\left(\frac{1}{\mu_{12}} + \frac{1}{\mu_{21}}\right)} \left[\frac{S_{12}(\lambda_1)}{\lambda_1\mu_{21}} + \frac{1}{\mu_{12}\mu_{21}} + \frac{E(S_{12}^2) + E(S_{21}^2)}{2}\right] \qquad \text{(IV-26)}$$

This checks with Eq. (44) of Chapter II, as it should, since the discipline in this case is nothing more than alternating priority.

Although we have tacitly assumed M≥1 and N≥1 throughout this analysis, it turns out that Eqs. (IV-22) are valid for M=0 and N=0 as well, if the terms with summations are taken to be zero. Physically M=0 means that the server will always depart from line 2 immediately upon arrival, regardless of the number of customers waiting at line 1. The equivalent holds if N =0.

We now obtain some other interesting measures of the system's behavior. The probability that a particular customer is followed by a changeover from 1 to 2 is

$$\pi^1_{oo} r_2^{\,N} + \pi^1_{o1} r_2^{\,N-1} + \pi^1_{o2} r_2^{\,N-2} + \ldots\ldots + \pi^1_{on-1} r_2 + \sum_{n=N}^{\infty} \pi^1_{on} \qquad \text{(IV-27)}.$$

This is explained as follows: A customer will be followed by a 1 to 2 changeover if its service completion leaves the system in one of the states $\,^1_{o\,n}$ with $n \geq N$. Also a changeover will follow a customer who leaves the system in state $\,^1_{o\,o}$ with N 2-customers then arriving before a 1-customer, or who leaves the system in state $\,^1_{o\,1}$ with N-1 2-customers then arriving, etc. Manipulation of the above expression yields

$$\pi^1_{oo}[\sigma^1(1) - \sigma^1 N(1) + r_2^{\,N}\, \sigma^1_N(\tfrac{1}{r_2})] \qquad \text{(IV-28)}$$

which from Eq. (IV-9a) is found to be

$$\frac{\pi^1_{oo} r_2}{u_o} \qquad \text{(IV-29)}$$

The probability that a given customer is followed by a 2 to 1 changeover is the same as the above since the number of changeovers in either direction must be the same.

The probability that a 1-customer is followed by a changeover is then

$$\frac{1}{r_1} \quad \frac{\pi^1_{oo} r_2}{u_o} = \frac{\pi^1_{oo} r_2}{r_1 u_o} \qquad \text{(IV-30)}$$

and the probability that a 2-customer is followed by a changeover is

$$\frac{1}{r_2} \quad \frac{\pi^1_{oo} r_2}{u_o} = \frac{\pi^1_{oo}}{u_o} \qquad \text{(IV-31)}$$

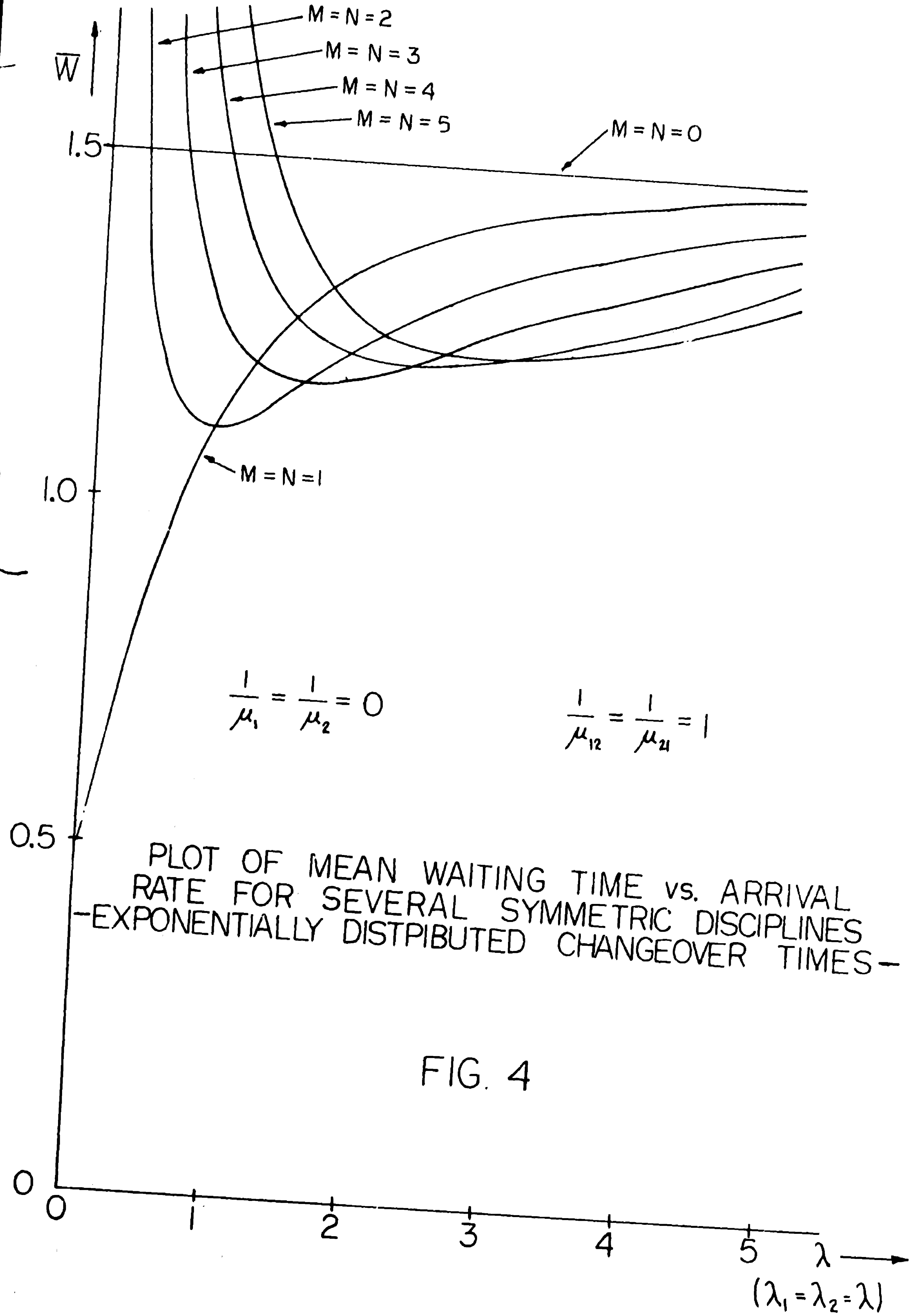
The probability that an arbitrary customer is followed by a changeover is

$$r_1 \frac{\pi^1_{oo} r_2}{r_1 u_o} + r_2 \frac{\pi^1_{oo}}{u_o} = \frac{2\pi^1_{oo} r_2}{u_o} \qquad \text{(IV-32)}$$

In Figs. 4 and 5 we plot the average waiting time for an arbitrary customer versus arrival rate for a number of different policies. (We recall that the policy is specified by fixing M and N). The system considered is symmetrical, i. e., $\lambda_1 = \lambda_2 = \lambda$ and $\mu_{12} = \mu_{21} = \mu_T$, so that we need only deal with symmetrical disciplines (M=N). In Fig. 4 the changeover times are exponentially distributed and in Fig. 5 they are discrete. One observes that for low arrival rates, the alternating priority discipline is best (M=N=1). As λ increases, it begins to pay to wait until more customers are present at the other line before a changeover is begun. For very high arrival rates, the optimum policy has a large value for M and N, but in this case there is very little improvement over the alternating priority discipline. The reason for this is that when $\lambda >> \mu_T$ a very large number of customers will probably arrive during a changeover, so that the server will almost always leave a line as soon as he arrives. Under these circumstances, the behavior of the server is virtually indistinguishable from what it would be under alternating priority.

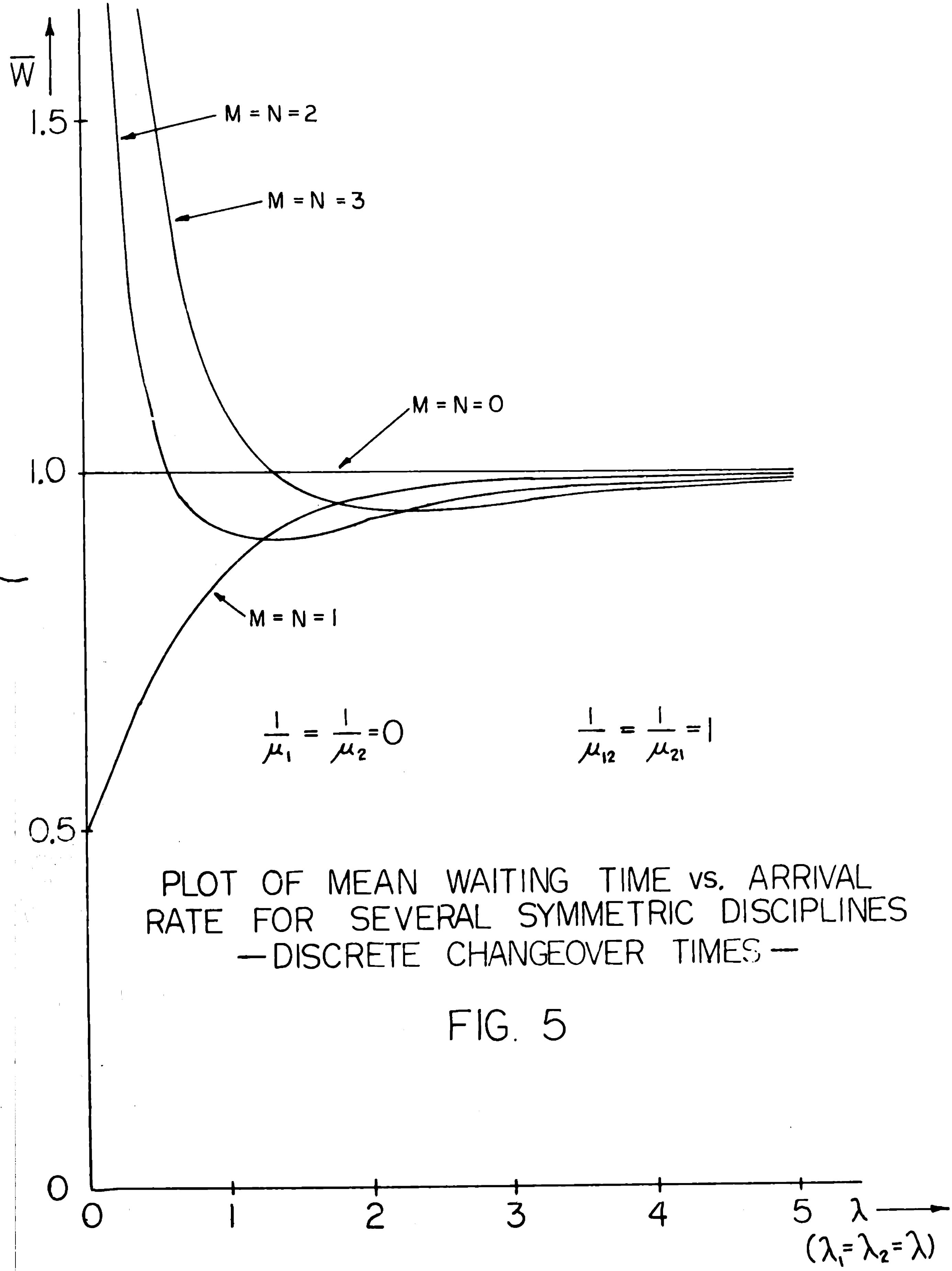
IV-f Summary

A wide class of disciplines was studied. For each of these disciplines, the analysis of the imbedded process resulted in an expression for the generating function of the state probabilities in terms of boundary conditions having a fairly complicated form. A number of techniques, including the study of a more deeply imbedded process, were used to determine the boundary conditions, yielding a complete solution.



PLOT OF MEAN WAITING TIME vs. ARRIVAL RATE FOR SEVERAL SYMMETRIC DISCIPLINES
-EXPONENTIALLY DISTPIBUTED CHANGEOVER TIMES-

FIG. 4



PLOT OF MEAN WAITING TIME vs. ARRIVAL RATE FOR SEVERAL SYMMETRIC DISCIPLINES
—DISCRETE CHANGEOVER TIMES—

FIG. 5

CHAPTER V

SELECTION OF A DISCIPLINE

In this chapter we look more closely at the question of choosing the queue discipline. Although we are not able to find the optimum discipline we do discover certain characteristics that it must possess. We show that the policies studied in Chapters II, III, and IV are special cases of a general classification. To demonstrate these ideas we give a specific example and compare the performance of several disciplines.

V-a The Optimum Discipline

In the absence of changeover times the optimum discipline is strict priority (SP), assigning high priority to the customers with the shorter mean service time. The fact is to some extent, intuitively obvious. (It must be undesirable to keep customers waiting that could be served very quickly.) When the changeover times become large, however, the primary concern becomes the minimization of the number of changeovers. In this case SP is unacceptable due to the high amount of switching associated with it. Under these circumstances more desirable discipline is alternating priority (AP).

We see from this that there are two competing factors that must must be taken into account in choosing the queue discipline.

One is the desirability of always serving customers with the shorter mean service time first. The other is the desirability of minimizing the number of changeovers. The SP and AP disciplines represent "limiting" cases in the sense that they each give much more weight to one of these two factors than to the other factor.

To make the above reasoning more precise, let us consider a situation where the mean service time of 1-customers is less than the mean service time of 2-customers. Were it not for the presence of changeover times, the server would always choose to serve a 1-customer if any are available. If, however, the server is at line 2, and there are only a few 1-customers present, the server may decide to serve a 2-customer rather than spend the extra time required to cross to line 1. Even if line 2 is empty, the server may refuse to cross to line 1 if a very few 1-customers are waiting. If the server is at line 1, he will, of course, not consider leaving until he has cleared all 1-customers. When line 1 becomes empty, the server may not cross to line 2 unless some minimum number of 2-customers are waiting. From this it follows that the optimum policy of the server will take the following form:

<u>At Line 1</u>

> Cross to line 2 only if line 1 is empty and line 2 contains N or more customers.

At Line 2

Cross to line 1 if it contains M_n or more customers where M_n may depend upon n, the number of 2-customers in the system.

If the service times for customers of type 1 and type 2 are equal, then it never pays to depart from a line which is occupied, and the optimum policy takes the form:

At Line 1

Cross to line 2 only if line 1 is empty and line 2 contains N or more customers.

At Line 2

Cross to line 1 only if line 2 is empty and line 1 contains M or more customers.

Note that the latter form is really a special case of the former with

$$N = N$$

$$M_0 = M$$

$$M_n = \infty \qquad n = 1, 2, 3, \ldots \qquad \text{(V-1)}$$

This is also recognized as the discipline we studied in Chapter IV without the assumption of zero service times.

Using the above notation we can identify SP as

$$N = 1$$

$$M_o = 1$$

$$M_n = 1 \quad , \qquad n = 1, 2, 3\ldots \qquad \text{(V-2)}$$

and AP as

$$N = 1$$

$$M_o = 1$$

$$M_n = \infty \qquad , \qquad n = 1, 2, 3 \ldots \qquad \text{(V-3)}$$

FCFS cannot be described in this manner since it depends on the order in which customers arrive. Hence, it cannot possibly be optimum in the sense we have defined.

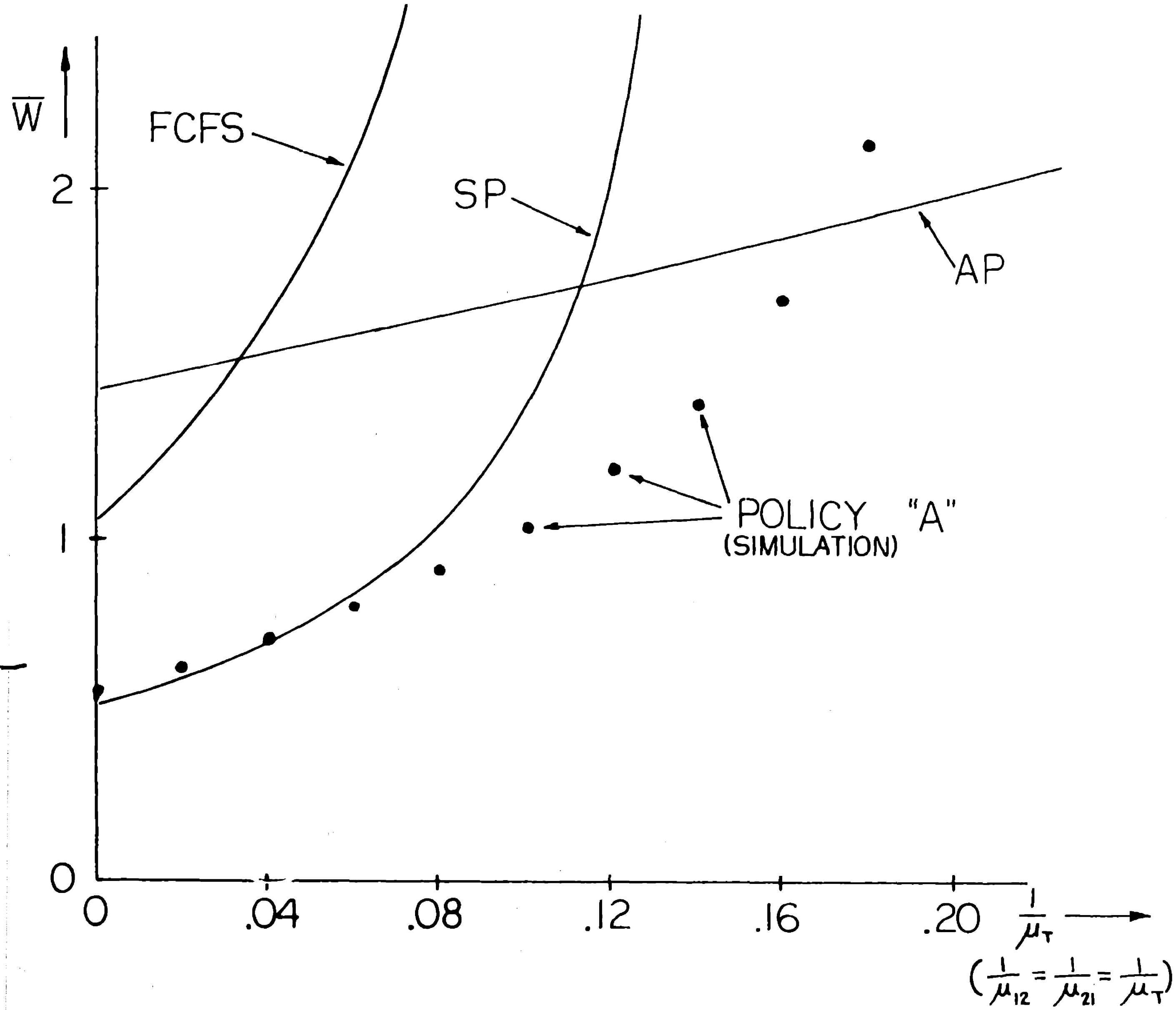
When the changeover times are zero, the SP discipline is optimum. As the changeover times are increased it becomes desirable to increase the quantities N, M_n so that less switching takes place. (Note the difference between Eq. (2) and Eq. (3).) When the changeover times are very large, we would want to use either AP or a discipline that causes even less switching, such as Eq. (1) with $M, N > 1$. It should be observed that policies having $M = M_o > 1$ or $N > 1$ have the property that at times the server is idle while the system is not empty. That this could be a desirable situation was demonstrated in Chapter IV.

The author has been unable to determine, other than in this general way, a method for finding the optimum discipline. In general, it will depend on not only the arrival rates and the average service and changeover times, but also on the distributions of these times. (One example of this can be seen in Fig. 4 and 5. When $\lambda = \mu_T = 1$, the optimum policy for exponentially distributed changeovers is M = N = 2, while for discrete changeovers it is M = N = 1.)

V-b Comparison of Disciplines

To illustrate the ideas of the previous section we now consider a specific example. We study a system with arrival rates $\lambda_1 = 5.0$ and $\lambda_2 = 1.0$ and exponentially distributed service times with means $\frac{1}{\mu_1} = 0.05$ and $\frac{1}{\mu_2} = 0.5$. The changeover times are also exponentially distributed with means $\frac{1}{\mu_{12}} = \frac{1}{\mu_{21}}$ (the same in either direction). Fig. 6 plots the average waiting time of an arbitrary customer in this system versus the mean changeover time for each of the disciplines FCFS, AP, and SP. For the case of SP, type 1-customers have "high priority."

One observes that SP has the lowest waiting time when the changeover times are zero. This is to be expected since the service time for 1-customers is less than that for 2-customers. It is interesting to note that FCFS, while not better than SP, is to be preferred over AP when the changeover times are small. This is true because AP has the undesirable characteristic that it often spends time serving 2-customers when 1-customers are



$\lambda_1 = 5.0 \qquad \lambda_2 = 1.0$

EXPONENTIALLY DISTRIBUTED
SERVICE AND CHANGEOVER

$\frac{1}{\mu_1} = .05 \qquad \frac{1}{\mu_2} = .5$

PLOT OF MEAN WAITING TIME vs. CHANGEOVER TIME FOR SEVERAL DISCIPLINES

FIG. 6

waiting. Although FCFS will occasionally do the same thing, it is more likely than AP to switch back to a 1-customer next.

As the changeover times increase, the advantage of SP and FCFS relative to AP decreases. In fact, with each of the former disciplines the queue saturates for a certain value of the changeover time. Using AP the queue remains unsaturated no matter how large the changeover times are made.

Also shown in Fig. 6 are some simulation results with the policy

$$\text{Policy "A"} \qquad \begin{aligned} N &= 1 \\ M_o &= 1 \\ M_n &= 3 \quad , \quad n=1,2,3.\ldots \end{aligned} \tag{V-4}$$

Policy "A" is the same as SP, except that if the server is at line 2 and line 2 is occupied, at least 3 1-customers must be waiting for a 2 to 1 changeover to take place. This policy is "between" SP and AP in the sense that it involves less switching than SP, but places more emphasis than AP does on the importance of serving 1-customers first. It can be seen that at certain points this policy is better than either SP or AP.

Another measure of the system behavior is the general-time probability that the system is empty $\pi^1_{oo}+\pi^2_{oo}$ (which we do not plot). When the changeover times are zero in the example of Fig.6, $\pi^1_{oo}+\pi^2_{oo}$ is the same for all three disciplines. This, of course, is expected since the fraction of time spent servicing customers is $\rho_1+\rho_2$ regardless of the queue discipline. For non-zero changeover times it is interesting to observe that $\pi_{oo}^1+\pi_{oo}^2$ is greater for AP than for either of the other two disciplines, even though the mean waiting time for AP is also greater. This result is somewhat unexpected, since queues with a good deal of server idleness are usually associated with short waiting times, and vice versa. The reason is that AP does not "waste time" crossing back and forth between lines. It clears the system in the fastest way possible. The trouble is that many 1-customers are kept waiting till last when they could have been served very quickly.

V-c Summary

In this chapter we have attempted to improve our understanding of the relationship between the queue discipline and the behavior of the system. We have catagorized the disciplines and demonstrated some of their more important characteristics. However, due to the large number of parameters that can be varied, we have only been able to investigate a small part of the problem. Further study could reveal the importance of such things as service distributions, changeover distributions, and arrival rates.

CHAPTER VI

CONCLUSIONS

A model was defined which had application to many queuing situations where server changeover times exist. In order to determine the effect of the server policy we analyzed some specific disciplines for the two-line case. For the general model we studied alternating priority and strict priority. With the assumption of zero service times, we studied a class of disciplines, specified by a pair of integers, M and N.

The method used was to concentrate on the "imbedded" process formed at service-completion instants. For each discipline considered, the transition probabilities were calculated and the generating functions for the state probabilities were obtained. In each case the result was expressed in terms of unknown boundary conditions. Through a variety of mathematical techniques, enough information about these boundary conditions could be found to determine the saturation condition, the mean waiting times, and certain other quantities.

The disciplines were compared and judged on the basis of the mean wait of an arbitrary customer. One interesting outcome was the observation that a discipline with a greater idle time is not necessarily a better discipline. We concluded that two factors must be taken into account in choosing the server's policy: one is the desirability to serve first those customers with the smaller mean service time, the other is the desirability to minimize the number of changeovers. From this we were able to classify potentially optimum disciplines in a general way and to show how each of the disciplines we examined fell into this classification.

In an appendix, the mean waiting time of an arbitrary customer was obtained for a specialized system with K lines. Although the particular results are not of great value, the method used to obtain them is instructive.

The problem of queuing systems with changeover times has been studied by Gaver[(4)], Miller[(9)], and Skinner[(11)]. Gaver considered the disciplines first-come, first-served and preemptive strict priority. Miller examined alternating priority. Skinner considered strict priority. The author believes that the results of this thesis represent a significant additional contribution. One of the most important outcomes of this research, in the opinion of the author, is the demonstration of the power of the method used. Much of the previous work on this problem has been highly specialized and of little generality. In some cases unrealistic assumptions were necessary due to the technique of solution. The method of the imbedded process, on the other hand, has been shown to be useful for a number of different situations and different disciplines. It is felt that the method could also be profitably employed even if the model were considerably expanded. One addition that appears easy to handle is the existence of "setup times"-- extra time required to serve a customer if he arrives at an empty queue, even if the server is at the same line.

Further research topics include the investigation of more general classes of policies for the two-line case, and perhaps some simple policies for the case of three or more lines. Also the question of what constitutes the optimum policy is extremely interesting and deserves additional study.

APPENDIX I

LIST OF NOTATION

λ_1	=	average arrival rate of 1-customers
λ_2	=	average arrival rate of 2-customers
r_1	=	$\frac{\lambda_1}{\lambda_1+\lambda_2}$
r_2	=	$\frac{\lambda_2}{\lambda_1+\lambda_2}$
$\overset{1}{m\ n}$	=	state of system "server at line 1, m customers at line 1, n customers at line 2"
$\overset{2}{m\ n}$	=	State of system "server at line 2, m customers at line 1, n customers at line 2"
π^1_{mn}	=	imbedded probability of state $\overset{1}{m\ n}$
π^2_{mn}	=	imbedded probability of state $\overset{2}{m\ n}$
$\pi^1(y,z)$	=	$\sum_{m=o}^{\infty}\sum_{n=o}^{\infty} \pi^1_{mn}\ y^m z^n$
$\pi^2(y,z)$	=	$\sum_{m=o}^{\infty}\sum_{n=o}^{\infty} \pi^2_{mn}\ y^m z^n$
b	=	π^2_{oo}/π^1_{oo}
$\sigma^1(z)$	=	$\frac{1}{\pi^1_{oo}} \sum_{n=o}^{\infty} \pi^1_{on}\ z^n$
$b\sigma^2(y)$	=	$\frac{1}{\pi^1_{oo}} \sum_{m=o}^{\infty} \pi^2_{mo} y^m$ (used in Chapters II and IV)
$b\sigma^2(z)$	=	$\frac{1}{\pi^1_{oo}} \sum_{n=o}^{\infty} \pi^2_{on}\ z^n$ (used in Chapter III)

$$\sigma^1_N(z) = \frac{1}{\pi^1_{oo}} \sum_{n=o}^{N-1} \pi^1_{on} z^n$$

$$b\sigma^2_M(y) = \frac{1}{\pi^1_{oo}} \sum_{m=o}^{M-1} \pi^2_{mo} y^m$$

$F_{S_1}(t)$ = Prob(service time of 1-customer $\leq$ t)

$F_{S_2}(t)$ = Prob(service time of 2-customer $\leq$ t)

$F_{S_{12}}(t)$ = Prob(changeover time from 1 to 2 $\leq$ t)

$F_{S_{21}}(t)$ = Prob(changeover time from 2 to 1 $\leq$ t)

$F_{W_1}(t)$ = Prob (waiting time of 1-customer $\leq$ t)

$F_{W_2}(t)$ = Prob (waiting time of 2-customer $\leq$ t)

$F_{B_1}(t)$ = Prob (length of type 1 busy period in isolation $\leq$ t)

$F_{B_2}(t)$ = Prob(length of type 2 busy period in isolation $\leq$ t)

$$S_1(s) = \int_o^\infty e^{-st} dF_{S_1}(t)$$

$$S_2(s) = \int_o^\infty e^{-st} dF_{S_2}(t)$$

$$S_{12}(s) = \int_o^\infty e^{-st} dF_{S_{12}}(t)$$

$$S_{21}(s) = \int_o^\infty e^{-st} dF_{S_{21}}(t)$$

$$W_1(s) = \int_o^\infty e^{-st} dF_{W_1}(t)$$

$$W_2(s) = \int_o^\infty e^{-st} dF_{W_2}(t)$$

$$B_1(s) = \int_0^\infty e^{-st} dF_{B_1}(t)$$

$$B_2(s) = \int_0^\infty e^{-st} dF_{B_2}(t)$$

$$\frac{1}{\mu_1} = \int_0^\infty t\, dF_{S_1}(t)$$

$$\frac{1}{\mu_2} = \int_0^\infty t\, dFS_2(t)$$

$$\frac{1}{\mu_{12}} = \int_0^\infty t\, dF_{S_{12}}(t)$$

$$\frac{1}{\mu_{21}} = \int_0^\infty t dF_{S_{21}}(t)$$

$$E(S_1^{\ 2}) = \int_0^\infty t^2 dF_{S_1}(t)$$

$$E(S_2^{\ 2}) = \int_0^\infty t^2 dF_{S_2}(t)$$

$$E(S_{12}^2) = \int_0^\infty t^2 dF_{S_{12}}(t)$$

$$E(S_{21}^2) = \int_0^\infty t^2 dF_{S_{21}}(t)$$

$$\rho_1 = \lambda_1/\mu_1$$

$$\rho_2 = \lambda_2/\mu_2$$

$\bar{m}$ = average number of 1-customers in system

$\bar{n}$ = average number of 2-customers in system

$\bar{W}_1$ = average waiting time of 1-customer (not including service)

$\bar{W}_2$ = average waiting time of 2-customer (not including service)

$\bar{W}$ = $r_1\bar{W}_1 + r_2\bar{W}_2$

$$p_{ij} = \int_o^\infty \frac{(\lambda_1 t)^i}{i!} \frac{(\lambda_2 t)^j}{j!} e^{-(\lambda_1+\lambda_2)t} \, dF_{S_1}(t)$$

$$q_{ij} = \int_o^\infty \frac{(\lambda_1 t)^i}{i!} \frac{(\lambda_2 t)^j}{j!} e^{-(\lambda_1+\lambda_2)t} \, dF_{S_2}(t)$$

$$u_{ij} = \int_o^\infty \frac{(\lambda_1 t)^i}{i!} \frac{(\lambda_2 t)^j}{j!} e^{-(\lambda_1+\lambda_2)t} \, dF_{S_{21}}(t)$$

$$v_{ij} = \int_o^\infty \frac{(\lambda_1 t)^i}{i!} \frac{(\lambda_2 t)^j}{j!} e^{-(\lambda_1+\lambda_2)t} \, dF_{S_{12}}(t)$$

$$u_j = \sum_{i=o}^\infty u_{ij}$$

$$v_i = \sum_{j=o}^\infty v_{ij}$$

$$P(y,z) = \sum_{m=o}^\infty \sum_{n=o}^\infty p_{mn} y^m z^n$$

$$Q(y,z) = \sum_{m=o}^\infty \sum_{n=o}^\infty q_{mn} y^m z^n$$

$$U(y,z) = \sum_{m=o}^\infty \sum_{n=o}^\infty u_{mn} y^m z^n$$

$$V(y,z) = \sum_{m=o}^{\infty} \sum_{n=o}^{\infty} v_{mn} y^m z^n$$

$$U(z) = \sum_{n=o}^{\infty} u_n z^n$$

$$V(y) = \sum_{m=o}^{\infty} v_m y^m$$

$$U_N(z) = \sum_{n=o}^{-1} u_n z^n$$

$$V_M(y) = \sum_{m=o}^{M-1} v_m y^m$$

$$U_N = U_N(1)$$

$$V_M = V_M(1)$$

$$\bar{U}_N = 1-U_N$$

$$\bar{V}_M = 1-V_M$$

APPENDIX II

PROOF OF EQ. (II-21)

We have

$$y_i = B_1(\lambda_2 - \lambda_2 z_i) \qquad \text{(A-II-1a)}$$

$$z_{i+1} = B_2(\lambda_1 - \lambda_1 y_i) \qquad \text{(A-II-1b)}$$

and we desire to find under what conditions

$$\lim_{i \to \infty} z_i = 1 \qquad \text{(A-II-2)}$$

We define the quantities

$$\delta_i = 1 - y_i \qquad \text{(A-II-3a)}$$

$$\epsilon_i = 1 - z_i \qquad \text{(A-II-3b)}$$

and determine the conditions for which

$$\lim_{i \to \infty} \epsilon_i \to 0 \qquad \text{(A-II-4)}$$

from which Eq. (A-II-2) immediately follows.

From Eqs. (A-II-1) and (A-II-3)

$$\delta_i = 1 - B_1(\lambda_2 \epsilon_i) \qquad \text{(A-II-5a)}$$

$$\epsilon_{i+1} = 1 - B_2(\lambda_1 \delta_i) \qquad \text{(A-II-5b)}$$

Using the finite Taylor series expansion (with error term) about the point zero

$$\delta_i = 1-B_1(0)-B_1'(0)\,\lambda_2\epsilon_i - \frac{B_1''(\xi_i)}{2}\,(\lambda_2\epsilon_i)^2 \tag{A-II-6a}$$

$$\epsilon_{i+1} = 1-B_2(0)-B_2'(0)\lambda_1\delta_i - \frac{B_2''(\eta_i)}{2}\,(\lambda_1\delta_i)^2 \tag{A-II-6b}$$

where

$$0 < \xi_i < \lambda_2\epsilon_i$$

$$0 < \eta_i < \lambda_1\delta_i$$

We must have

$$B_1(0) = B_2(0) = 1 \tag{A-II-7}$$

(since B(s) is the transform of a probability distribution) and from Eq. (II-13) one finds

$$B_1'(0) = -\frac{1}{\mu_1-\lambda_1} \tag{A-II-8a}$$

$$B_2'(0) = -\frac{1}{\mu_2-\lambda_2} \tag{A-II-8b}$$

Eqs. (A-II-6) thus reduce to

$$\delta_i = \frac{\lambda_2}{\mu_1-\lambda_1}\,\epsilon_i - \frac{B_1''(\xi_i)}{2}\,(\lambda_2\epsilon_i)^2 \tag{A-II-9a}$$

$$\epsilon_{i+1} = \frac{\lambda_1}{\mu_2-\lambda_2}\,\delta_i - \frac{B_2''(\eta_i)}{2}\,(\lambda_1\delta_i)^2 \tag{A-II-9b}$$

Making the substitution of Eq. (A-II-9a) into Eq. (A-II-9b)

$$\epsilon_{i+1} = \frac{\lambda_1\lambda_2}{(\mu_1-\lambda_1)(\mu_2-\lambda_2)} \epsilon_i - \frac{\lambda_1}{\mu_2-\lambda_2} \frac{B_1''(\xi_i)}{2} (\lambda_2\epsilon_i)^2 - \frac{B''_2(\eta_i)}{2} (\lambda_1\delta_i)^2$$

(A-II-10)

From the convexity of Laplace transforms

$$B_1''(\xi_i) \geq 0 \qquad \text{(A-II-11a)}$$

$$B''_2(\eta_i) \geq 0 \qquad \text{(A-II-11b)}$$

so that

$$\epsilon_{i+1} \leq \frac{\lambda_1\lambda_2}{(\mu_1-\lambda_1)(\mu_2-\lambda_2)} \epsilon_i \qquad \text{(A-II-12)}$$

It follows that

$$\epsilon_{i+1} \leq \left[\frac{\lambda_1\lambda_2}{(\mu_1-\lambda_1)(\mu_2-\lambda_2)}\right]^i \epsilon_o \qquad \text{(A-II-13)}$$

Since $\epsilon_o = 1$ or $\delta_o = 1$, (depending on the starting point)

Eqs. (A-II-5) show that also

$$\epsilon_i \geq 0 \qquad \text{(A-II-14)}$$

Thus from Eqs. (A-II-13) and (A-II-14)

$$\lim_{i\to\infty} \epsilon_i = 0 \qquad \text{(A-II-15)}$$

for either starting point as long as

$$\frac{\lambda_1\lambda_2}{(\mu_1-\lambda_1)(\mu_2-\lambda_2)} = \frac{\rho_1\rho_2}{(1-\rho_1)(1-\rho_2)} < 1 \qquad \text{(A-II-16)}$$

The above inequality can be rewritten

$$1-\rho_1-\rho_2 < 1 \qquad \text{(A-II-17)}$$

which is the condition for which Eq. (A-II-2) is valid. (Incidentally, the magnitude of the quantity in Eq. (A-II-16) gives an indication of the speed at which the iteration will converge.)

APPENDIX III

ITERATIVE METHOD FOR FINDING b

If b were known, it would be possible to determine the functions $\sigma^1(z)$ and $\sigma^2(y)$ for many different values of their arguments by the iteration defined by Eqs. (II-19) and (II-20). In particular, since (as shown in Appendix II)

$$\lim_{i \to \infty} z_i = 1 \qquad \text{(A-III-1)}$$

we could obtain $\sigma^1(1)$. The same value for this quantity must result regardless of the starting point of the iteration (either $y_o = 0$ or $z_o = 0$. If, however, an incorrect value of b is used in the iteration, the resulting value of $\sigma^1(1)$ does depend upon the starting point. The requirement that the convergent values of $\sigma^1(1)$ be independent of the starting point is sufficient to determine b uniquely. We now show this in detail.

Substitution of Eq. (II-20a) into Eq. (II-20b) yields

$$\sigma^1(z_{i+1}) = 1 - r_2 z_{i+1} - \frac{r_1 y_i}{U(y_i, z_i)V(y_i, z_{i+1})} + \frac{\sigma^1(z_i)}{U(y_i, z_i)V(y_i, z_{i+1})} + b\,\frac{1-r_1 y_i}{V(y_i, z_{i+1})}$$

(A-III-2)

which expresses $\sigma^1(z_{i+1})$ in terms of $\sigma^1(z_i)$. If the iteration starts with $z_0=0$, we have from Eq. (II-2)

$$\sigma^1(z_1) = 1 - r_2 z_1 - \frac{r_1 y_o}{U(y_o, 0)V(y_o, z_1)} + \frac{1}{U(y_o, o)V(y_o, z_1)} + b\,\frac{1-r_1 y_o}{V(y_o, z_1)}$$

$(z_o = 0)$

(A-III-3)

whereas if the iteration begins with $y_0=0$, we have from Eq. (II-20b)

$$\sigma^1(z_1) = 1-r_2 y_1 + b \left[\frac{1-r_2 z_1}{V(0, z_1)} \right] \qquad (y_o=0) \qquad \text{(A-III-4)}$$

It should be observed, even though it is not indicated by the notation, that the value of z_i when $z_o = 0$ is used to begin the iteration is <u>not</u> the same as z_i when $y_0 = 0$ is used to begin the iteration. Thus, for example, there is no need for $\sigma^1(z_1)$ of Eq. (A-III-3) to be equal to $\sigma^1(z_1)$ of Eq. (A-III-4) since in the former case

$$z_1 = B_2(\lambda_1-\lambda_1 B_1(\lambda_2)) \qquad (z_0=0) \qquad \text{(A-III-5)}$$

and for the latter

$$z_1 = B_2(\lambda_1) \qquad (y_0=0) \qquad \text{(A-III-6)}$$

Note that both Eq. (A-III-3) and Eq. (A-III-4) are of the form

$$\sigma^1(z_1) = A_o + B_o b \qquad \text{(A-III-7)}$$

where z_1, A_0, and B_0, depend on the starting point ($z_0 = 0$ or $y_0=0$) but do not depend on b. Using Eq. (A-III-7) one finds by induction that the form of Eq. (A-III-2) is

$$\sigma^1(z_{i+1}) = A_i + B_i\, b \qquad \text{(A-III-8)}$$

Again z_{i+1}, A_i, and B_i, depend upon the starting point (as well as i) but do not depend upon b .

From this it follows that the difference between the quantities

$\sigma^1(z_{i+1})$, starting at $z_0=0$

and

$\sigma^1(z_{i+1})$, starting at $y_0=0$

is of the form

$$C_i+D_i\, b \tag{A-III-9}$$

One notes that Eq. (A-III-1) is true independent of the starting point so that

$\sigma^1(1)$, starting at $z_0=0$ - $\sigma^1(1)$, starting at $y_0=0$

$$= \lim_{i\to\infty}[C_i+D_i b] = C+Db \equiv F(b) \tag{A-III-10}$$

The difference between the two convergent values of $\sigma^1(1)$ is linearly dependent upon b. For an acceptable solution, the two convergent values must be the same, forcing F(b) in Eq. (A-III-10) to be equal to zero.

The method to find b then is as follows. First we assume an arbitrary value for b, say b_1. Then we compute $F(b_1)$ (by performing the two iterations and taking the difference of the results). If $F(b_1)$ is zero then we are done. If $F(b_1)$ is not zero, we try a new value of b, say b_2, and compute $F(b_2)$.

The correct value of b is then

$$b = \frac{b_1 F(b_2) - b_2 F(b_1)}{F(b_2) - F(b_1)} \qquad \text{(A-III-11)}$$

which follows directly from the linearity of F(b).

APPENDIX IV

PROOF THAT $D(z_0) = 0$

We desire to show that there exists a quantity z_0

$$0 \leq z_0 < 1 \tag{A-IV-1}$$

such that

$$D(z_0) = 0 \tag{A-IV-2}$$

where D(z) is given by Eqs. (III-12b), (III-8), and (III-10).

One has for $z = 0$

$$D(0) = S_2(\beta(0))\, S_{21}(\alpha(0))\, S_2(\alpha(0))\, [\, S_{12}(\alpha(0)) - S_{12}(\beta(0))\,] \tag{A-IV-3}$$

One notes that

$$\beta(0) = \lambda_1+\lambda_2 \qquad > \qquad \alpha(0) = \lambda_1 + \lambda_2 - \lambda_1 B_1(\lambda_2) \tag{A-IV-4}$$

and since $S_{12}(s)$ is a non-increasing function of s

$$D(0) \geq 0 \tag{A-IV-5}$$

One also notes that

$$D(1) = 0 \tag{A-IV-6}$$

We have, for the case of non-saturation

$$D'(1) > 0 \tag{A-IV-7}$$

See Eq. (III-21). Eqs. (A-IV-5), (A-IV-6), and (A-IV-7) can only be satisfied by a function having the property mentioned at the beginning of this appendix.

APPENDIX V

EXPRESSION FOR $\bar{W}_2$

Since $A_1(1) = A_2(1) = D(1) = 0$ in Eqs. (III-33), the use of L'Hopital's rule is required. Thus

$$\sigma^1(1)'+r_2-(\sigma^1(1)-r_1) = \frac{A''_1(1)D'(1) - A'_1(1)D''(1)}{2(D'(1))^2} \qquad \text{(A-V-1a)}$$

$$b[\sigma^2(1)'+r_2- (\sigma^2(1)-r_1)] = \frac{A''_2(1)D'(1) - A'_2(1)D''(1)}{2(D'(1))^2} \qquad \text{(A-V-1b)}$$

The quantities of Eqs. (A-V-1) are

$$A''_1(1)=r_1y_1(1+b)(1-S_2(\lambda_1))+2(r_1\lambda_1+r_2)[(1+b)(1-\lambda_2S'_2(\lambda_1))+b\ X_{21}(1-S_2(\lambda_1))]$$

$$+ 2b(1-X_2)(r_1X_{21}-r_2) - r_1by_2 \qquad \text{(A-V-2a)}$$

$$A''_2(1) = r_1y_1(1+b)S_2(\lambda_1)S_{12}(\lambda_1)+2(r_1X_1+r_2)[(1+b)\lambda_2S'_{12}(\lambda_1)S_2(\lambda_1)$$

$$+ \lambda_2S'_2(\lambda_1)S_{12}(\lambda_1)+bX_{21}S_2(\lambda_1)S_{12}(\lambda_1)] \qquad \text{(A-V-2a)}$$

$$+2b(1-X_2)r_2+r_1by_2-2br_2(X_{12}+X_{21}) \qquad \text{(A-V-2b)}$$

$$+br_1(y_{12}+y_{21}+ 2X_{21}X_2+ 2X_{21}X_{12}+ 2X_2X_{12})$$

$$D''(1) = 2(1-X_2)(1-\lambda_2 S_2'(\lambda_1)-S_2'(\lambda_1)\lambda_2 S_{12}(\lambda_1)+S_2(\lambda_1)\lambda_2 S_{12}'(\lambda_2)+X_{21}S_2(\lambda_1)S_{12}(\lambda_1))$$

$$- 2(1-\lambda_2 S_2'(\lambda_1))\,(X_2+X_{21}) - (1-S_2(\lambda_1)+S_2(\lambda_1)S_{12}(\lambda_1))y_2 \qquad \text{(A-V-2c)}$$

$$-(1-S_2(\lambda_1))\,(y_{12}+y_{21}+2X_{21}X_2+2\,X_{21}X_{12}+2X_2X_{12})$$

where

$$X_i \equiv \frac{\lambda_2/\mu_i}{1-\rho_1} \qquad i = 1,\ 2,\ 12,\ 21$$

$$y_i \equiv \left(\frac{\lambda_2}{1-\rho_1}\right)^2 \left[\, E(S_i^2) + \frac{\lambda_1/\mu_i}{1-\rho_1} E(S_1^2)\right] \quad i = 1,\ 2,\ 12,\ 21$$

Combining Eq. (A-V-1) and (A-V-2) and Eqs. (III-20) with Eqs. (III-18) and (III-32), results in an expression for $\overline{W}_2$.

APPENDIX VI

SOME RESULTS FOR A QUEUING SYSTEM WITH K LINES

In Chapters II, III, and IV, our attention was focused on the "imbedded" process. In this appendix we concentrate instead on the actual continuous-time process for a queuing system with K lines. By making some highly specialized assumptions, we are able to obtain an expression for the mean waiting time of an arbitrary customer.

Consider a system having the following characteristics:

1) There are K lines.

2) The arrivals at each line are Poisson with the average arrival rate λ.

3) Service times at each line are zero.

4) The changeover time from any line to any other line is exponentially distributed with mean $1/\mu$.

5) The queue discipline is

    a) All customers at a line are served immediately upon arrival of the server to that line.
    b) The server will remain stationary if, and only if, the system is empty.
    c) The line that the server chooses to go to next is the one containing the earliest arriving customer.

The waiting times could obviously be reduced by changing 5c) to read "the line containing the largest number of customers." The method of solution makes the assumption of 5c) necessary. It is reassuring to some extent at least, that the line with the earliest arriving customer is also the one which has a priori the greatest probability of being the longest.

We first show that the average waiting time for an arbitrary customer is completely specified by k, the number of occupied lines. We then find the general-time probability that k lines are occupied. By multiplying the expected wait in each case by the probability of its occurrence and summing, the mean wait is determined.

If a customer arrives when the system is empty, he will have to wait an average time

$$W/o = \frac{K-1}{K} \cdot \frac{1}{\mu} + \frac{1}{K} \cdot 0 \qquad \text{(A-VI-1)}$$

where W/0 denotes "average wait given the customer arrives when 0 lines are occupied." Eq. (A-IV-1) is obtained by noticing that with probability $\frac{K-1}{K}$ the server is at a different line than the customer and thus an average wait $\frac{1}{\mu}$ is required. If the server is at the same line, the customer is served immediately.

If exactly one line is occupied when a customer arrives, then his average wait is

$$W/1 = \frac{K-1}{K} \cdot \frac{2}{\mu} + \frac{1}{K} \cdot \frac{1}{\mu} \qquad \text{(A-VI-2)}$$

where W/1 denotes "average wait given the customer arrives when 1 line is occupied." This expression is explained as follows: When the customer arrives, the server must be traveling toward the occupied line. If the customer enters this line he will experience an average delay of $1/\mu$. If he enters any other line, he must first wait for the occupied line to be cleared, and then must wait for the server to cross to the line which the customer entered. This takes a total average time of $2/\mu$.

If exactly $k = 1, 2. \dots, K-1$, lines are occupied when a customer arrives, his average wait is

$$W/k = \frac{1}{K} \cdot \frac{1}{\mu} + \frac{1}{K} \cdot \frac{2}{\mu} + \dots + \frac{1}{K} \cdot \frac{k}{\mu} + \frac{K-k}{K} \cdot \frac{k+1}{\mu} \qquad \text{(A-VI-3)}$$

$$k = 1, 2. \dots, K-1$$

This is obtained by noticing that the arrival could have entered the line which is about to receive service, or the one which will be served after that, or the next, etc., each with probability $1/K$. Or the customer might have entered an unoccupied line with probability $\frac{K-k}{K}$.

Finally, if all lines are occupied at the instant of arrival, the mean wait is

$$W/K = \frac{1}{K} \cdot \frac{1}{\mu} + \frac{1}{K} \cdot \frac{2}{\mu} + \dots + \frac{1}{K} \cdot \frac{K}{\mu} \qquad \text{(A-VI-4)}$$

which is derived using the same reasoning as above.

We now find the general-time probability that k lines are occupied. Let $p_k(t)$ be the probability that k lines are occupied at time t. We can write

$$p_o(t+dt) = p_o(t)\,(1-(K-1)\lambda\, dt) + p_1(t)\,\mu dt \qquad \text{(A-VI-5)}$$

This is obtained as follows: The system will be empty at time t +dt if it was empty at time t and between t and t +dt no customers arrive at any of the lines other than the one occupied by the server. (Note that arrivals to the line receiving service are cleared immediately, leaving the system empty.) The system could also be empty at time t + dt if at time exactly one line was occupied, and this line was served between t and t + dt. We can also write

$$p_1(t+dt) = p_o(t)(K-1)\lambda dt + p_1(t)(1-(K-1)\lambda^{dt}-\mu dt)+ p_2(t)\mu dt \qquad \text{(A-VI-6)}$$

$$p_k(t+dt)=p_{k-1}(t)\,(K-k+1)\lambda dt +p_k(t)(1-(K-k)\lambda dt-\mu\, dt)+p_{k+1}(t)\,\mu dt \qquad \text{(A-VI-7)}$$

$$k= 2, 3. \dots , K-1$$

$$p_K(t+dt) = p_{K-1}(t)\,\lambda\, dt + p_K(t)\,(1-\mu dt) \qquad \text{(A-VI -8)}$$

The derivations are completely similar to the above and an explanation is therefore omitted.

By rearranging terms, dividing by dt, and letting dt → 0 we obtain expressions for $\frac{dp_k(t)}{dt}$. In the steady state, $\frac{d}{dt}= 0$ and $p_k(t) \rightarrow p_k$. It is easy to show that

$$p_k = \frac{(K-1)\,(K-1)\,!}{(K-k)!} \quad \left(\frac{\lambda}{\mu}\right)^k p_o \qquad k=1, 2, \ldots . K \qquad \text{(A-VI-9)}$$

p_o is obtained by the requirement that

$$\sum_{k=0}^{K} p_k = 1 \qquad \text{(A-VI-10)}$$

The result is

$$p_o = \quad 1+(K-1)(K-1)\,! \sum_{k=1}^{K} \left(\frac{\lambda/\mu}{K-k}\right)^k \qquad \text{(A-VI-11)}$$

The mean waiting time for an arbitrary customer is

$$\overline{W} = \sum_{k=1}^{K} p_k \quad W/k \qquad \text{(A-VI-12)}$$

Substitution of Eqs. (A-VI-1), (A-VI-3), (A-VI-4), (A-VI-11), and (A-VI-9) into Eq. (A-VI-12) yields the desired result. The assumption of zero service times insures that the queue will never saturate.

The results of this appendix are of little value in themselves due to the highly restricted nature of the problem. The reason for including this analysis here is the possibility that the technique might have applicability to some more general problems.

BIBLIOGRAPHY

1. B. Avi Itzhak, W. L. Maxwell, and L. W. Miller, "Queuing with Alternating Priorities", Opns. Res. 13, 306-318 (1965).

2. A. Cobham, "Priority Assignment in Waiting Lines", Opns. Res. 2, 70-76 (1954).

3. D. R. Cox and H. D. Miller, The Theory of Stochastic Processes, Wiley, New York, (1965).

4. D. P. Gaver, "A Comparison of Queue Disciplines When Service Orientation Times Occur," Naval Log. Quart. 10, 219-235(1963).

5. J. L. Holley, "Waiting Line Subject to Priorities", Opns. Res. 2, 341-343 (1954).

6. D. G. Kendall, "Some Problems in the Theory of Queues", J. Roy. Stat. Soc. B., 13, 151-173 (1951).

7. J. D. C. Little, "A Proof for the Queuing Formula $L = \lambda W$", Opns. Res. 9, 383-387 (1961).

8. J. D. C. Little, "A Markov Decision Process in a Single Channel Queue with Poisson Arrivals", (An Unpublished Paper) (1961).

9. L. W. Miller, "Alternating Priorities in Multi-Class Queues," Ph. D. Thesis, Cornell University, September 1964.

10. R. G. Miller Jr., "Priority Queues", Ann. Math. Statist. 31, 86-103 (1960).

11. C. E. Skinner "A Priority Queuing System with Server-Walking Time", Opns. Res. 15, 278-285 (1967).

12. L. Takács, "Investigation of Waiting Time Problems by Reduction to Markov Processes", Acta. Math. Acad. Sci. Hung. 6, 101-129 (1955).

BIOGRAPHICAL NOTE

Martin Eisenberg was born in Neptune, New Jersey, on October 27, 1941. He attended public schools in Marlboro, New Jersey and was graduated from Freehold Regional High School in June 1959. He entered M. I. T. in the fall of 1959 in the co-operative course in electrical engineering with IBM. He received the degrees S. B. and S. M. in Electrical Engineering in September 1964, and Ph. D. in September 1967.

He served as a teaching assistant for three years and was promoted to an instructor in June 1966. He is a member of Eta Kappa Nu, Tau Beta Pi, and Sigma Xi. In 1966 he received a Supervised Investors Services award for accomplishment in teaching.

Mr. Eisenberg was married to Esther Joan Shupe in June 1963. The Eisenbergs have one child, Diane Elizabeth, born in March 1966.