THE UNIVERSITY OF MICHIGAN

INDUSTRY PROGRAM OF THE COLLEGE OF ENGINEERING

ANALYSIS OF SYSTEMS OF QUEUES IN PARALLEL

Erhan Çınlar

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in the
University of Michigan
Department of Industrial Engineering
1965

November, 1965

IP-723

ensm

UMR1202

Doctoral Committee:

Associate Professor Ralph L. Disney, Chairman
Associate Professor A. Bruce Clarke
Associate Professor John N. Darroch
Professor Herbert P. Galliher
Professor Robert M. Thrall
Professor Richard C. Wilson

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to Professor R. L. Disney who has been a most inspiring counselor. Thanks are also due to Professor A. B. Clarke and Professor J. N. Darroch who read a preliminary version of the manuscript and offered several helpful suggestions. The Industry Program of the College of Engineering at the University of Michigan has provided help in the preparation and printing of this manuscript, and I would like to acknowledge their assistance.

TABLE OF CONTENTS

		Page
ACKNOWLEDGEMENTS		iii
LIST OF ILLUSTRATIONS		v
INTRODUCTION		1
	PART I. DECOMPOSITION OF A SEMI-MARKOVIAN STREAM INTO R STREAMS	
Chapter I.	PRELIMINARIES	22
II.	DECOMPOSITION OF A SEMI-MARKOVIAN STREAM UNDER A MARKOVIAN DECISION RULE	35
III.	DECOMPOSITION OF A SEMI-MARKOVIAN STREAM UNDER A STATE DEPENDENT DECISION RULE	59
	PART II. BALKING IN THE QUEUEING SYSTEM SM/M/1	
IV.	QUEUEING SYSTEM SM/M/1 WITH BALKING	78
V.	STREAM BALKING FROM THE QUEUEING SYSTEM SM/M/1	98
VI.	GENERALIZATIONS AND SPECIALIZATIONS	122
	PART III. APPLICATIONS	
VII.	A SIMPLE SYSTEM: AN EXAMPLE	141
BIBLIOGRAPHY		158

LIST OF ILLUSTRATIONS

Figure		Page
1.	The Overflow Problem.....	4
2.	A System of Parallel Queues.....	6
3.	Form of Some of the Matrices.....	85
4.	Form of the Matrix P.....	87
5.	Matrices in the Overflow Problem.....	135
6.	A Simple System.....	141

INTRODUCTION

1. HISTORICAL BACKGROUND

Considerable attention has been given to the theory of simple queues. The easy problems have been solved and the methods that have been developed are sufficiently powerful to analyze simple models consisting of a group of servers with one waiting line in front of them under some special types of inputs and service mechanisms. We refer to treatises by L. Takàcs [31] and T. L. Saaty [26] and their impressive lists of references.

Although the subject of queues seems to be approaching a state of maturity, the general problem of a network of queues remains in its infancy. A network of queues is an arrangement of queues in parallel and in series. So far most of the interest has been on queues in series. The solution of such a system requires the knowledge of the output process of a queue, and furthermore since this output becomes the input to the next queue, it is required (for reasons of analytical ease) that the output process have independent increments. This greatly limits the kind of inputs and service times that can be assumed. Invariably, it is assumed that the input to the system is Poisson, all the servers are negative exponential, and all the queues are of infinite length, so that as P. Burke [4] shows, the outputs are also Poisson processes. Since very little is known about queues with general inputs, and since the output is a dependent process in general, no substantial breakthrough has been made in this field of queues in series. We refer to [10,11,12,13,14,21,24,27] for discussion of series systems.

The field of parallel queues is equally barren. Aside from studies of Palm and Disney on the overflow problem (which we will mention below) the only work is of Kingman [17]. Kingman has studied the queueing properties of a system of two queues in parallel subject to a recurrent input where the arriving customers join the queue of shortest length (in case of a tie, decisions are based on the outcome of a coin tossing experiment.) Each of the two queues has one negative exponential server and a waiting room of infinite size. No jockeying (changing the queues after joining one) is allowed. Kingman shows that a steady state exists and gives the joint distribution of the queue sizes in the steady state by means of a double generating function.

A different approach to queueing networks is taken by R. Syski and V. E. Beneš. They put forward the idea that a network should be treated as a whole, and not as a combination of separate queues. They argue that this type of approach has an advantage in that it would eventually formulate general laws governing the behavior of the whole complex system of queues. For example, Syski [30], introduces the "generalized congestion process" $\{X(t),\ t \geq 0\}$ whose random variable takes values in the abstract space $\mathfrak{X}$ of "patterns". A pattern describes the state of the network taking into consideration actual situations prevailing at each queue of the network (number of customers, their waiting times, number of rejected customers, input law, queue disciplines, structure of interconnections, etc.). Partial order is then introduced into the space $\mathfrak{X}$ of all patterns, and the operations

of addition and scalar multiplications of patterns are defined. It is shown that $\mathcal{X}$ forms a σ-complete vector lattice under order topology, and the expectation of $X(t)$ is defined by means of the McShane integral.

In the same vein but with different interests, Beneš [1,2], considers the mathematical problems related to the theory of connecting systems. He pays special attention to partial order, and goes on to study the algebraic and topological properties of connecting networks. He introduces three topologies (induced by partial ordering by inclusion, its dual, and by a distance function on the set of states) and uses them to characterize the properties of non-blocking and rearrangeability.

Although this approach to queueing networks may in time prove to be valuable, it will not offer any analytical tools for studying the characteristics of individual queues. Ideally, what is needed is a body of analytical methods that somehow enables one to study each queue in the network individually and yet is also able to bring about the system characteristics such as correlation among the queues. In this paper we contribute several such methods that can be used successfully in the analysis of parallel queues.

The approach we have used may be called the method of decomposition. Instead of considering a system of R queues in parallel as a whole, we decompose the system so that each queue may be studied in isolation. This idea of decomposition seems to be the implicit motive in the studies on the decomposition of a Poisson process into R

streams by independent trials or by every other principle, etc. On the same line, Palm [20] has studied a system with no storage of R single server negative exponential queues in parallel subject to a recurrent input where the overflows from the i-th queue become the arrivals to the $i + 1$ st queue. He did this by finding the distribution of the time between overflows from the i-th queue as a function of the distribution of the time between arrivals to that queue. Since the overflows from the i-th queue are the arrivals to the $i + 1$ st queue he was able to repeat this for $i = 1, 2, \ldots, R$ starting with the arrivals into the first queue which is the same as the arrivals into the system.

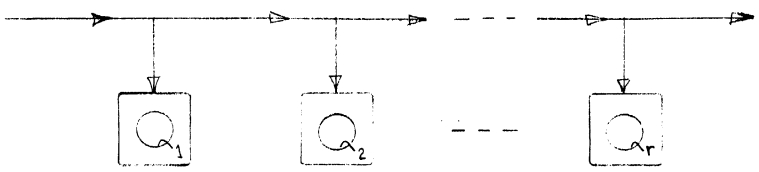

Figure 1. The Overflow Problem.

Disney [6] has generalized this problem by allowing storage in the system. He emphasized the usefulness of a decomposition approach to study such systems of parallel queues. He has shown that the overflow stream from a finite queue with negative exponential servers subject to a recurrent input is itself a recurrent process by a qualitative proof. He found the distribution of the time between overflows in the case of Poisson input, negative exponential servers, and a waiting room of size N. However, the problem of determining the distribution of the time between overflows in the case of a recurrent input has been left untreated. Since the overflows from a negative exponential server form a recurrent process

which is not Poisson even if the input is Poisson, one cannot use Disney's methods for the second queue or beyond because of the unavailability of the distribution function for the overflows from the second queue. Disney's problem (and therefore Palm's problem as its special case) happens to be a special case of one of the problems investigated in this paper, so that we supply the distribution function needed, thus completing Disney's treatment.

2. PROBLEM IN GENERAL

Analysis of a system of parallel queues by the method of decomposition is essentially a study of the decomposition of the arrival process of that system. In general this decomposition of the arrival stream occurs as a result of a series of decisions made either by the customers themselves, or by the system, or by both. Thus the general decomposition problem is actually a series of decompositions affected one after another.

To illustrate these ideas we refer to the figure below. A circle represents a point of decision making, either by the customer or by the management. Squares indicate queues, a queue consisting of several servers with a waiting room in front of them. Assume that at decision point D_1 customers are sent in one of the three possible directions by some rule such as "every other" rule. At decision points D_2, D_3, D_4 customers decide on whether or not to join the queue they are before on the basis of queue size there. For example, given that he has arrived at D_3, a customer decides to join the queue number three,

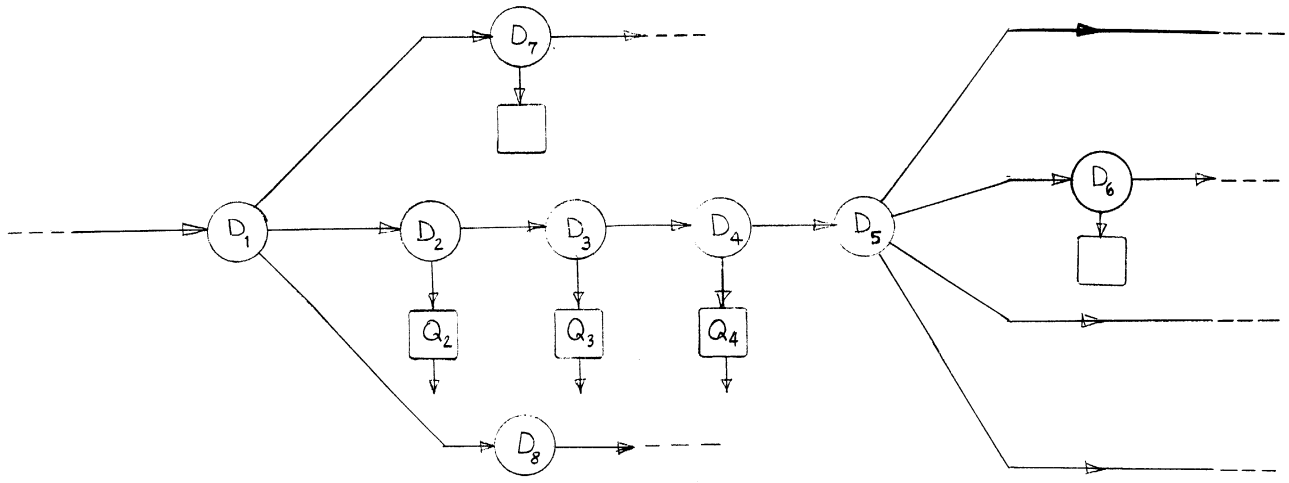

Figure 2. A System of Parallel Queues.

Q_3 , with a probability that depends on the number of people already there in Q_3 ; if he decides not to join Q_3 , then he proceeds to D_4 and the same things are repeated there.

The customers arriving at D_5 will include, in general, some customers who needed to be served in one of the queues Q_2, Q_3, Q_4 but got rejected for some reason such as the unavailability of space in waiting room. Assume the system directs such customers towards D_6 for serving them in that branch.

Notice that the customers arriving at D_1 go through a series of decision makings until they finally join a queue. These decisions are of a sequential character: only those customers who have been through points D_1, D_2, D_3, D_4 are allowed to make a decision at point D_5 . Hence, the general decomposition problem can be thought as a series of decompositions affected one after another. Then, the

problem reduces to the investigation of the decompositions affected at each one of these sequence of decision points. For example, assuming that the service times in each queue are independent of each other and no recycling is permitted, we would proceed to analyze the system of Figure 1 as follows. First, given the process of arrivals at D_1, we would determine each one of the three streams emerging at D_1. Then, the process of arrivals at D_2, D_7, D_8 are all known. Consider only the stream arriving at D_2. Some of the customers will join the queue number 2, and some will move to D_3. Here we need to analyze the properties of the queueing processes going on at Q_2, as well as characterizing the stream of customers that did not join Q_2. Once this balking process is characterized, then the process of arrivals at D_3 is known, and everything that is done at point D_2 and Q_2 can be repeated there. The same can be done for D_4 and Q_4 to analyze the processes of interest in Q_4, and to characterize the stream of customers who balk at Q_4 to become arrivals at D_5. Next, given this arrival stream and the decomposition rule used at D_5, we can obtain the streams emerging from D_5; and continue the analysis further.

This method of analysis requires that one determine each of the streams emerging from a decomposition of a known stream by a known rule. Clearly, decompositions by different rules require different tools of analysis. In general, we distinguish two types of decomposition rules.

Decision points D_1, D_5 are examples of what may be called "junctions". Decisions at such points are about "what direction to take". Although this decision may be dependent on the kind of service(s) offered in subsequent queues at each direction, we assume that it is made without the actual knowledge of the states of the queueing processes going on. This assumption is quite realistic in cases where the direction to be taken is dictated by the needs of customers for a special kind of service which may be offered only in the queues at certain directions, (which is usually the case in industrial applications.) Or, sometimes it may be economically infeasible, or even physically impossible, to obtain precise information about the sizes of the queues, etc. in the subsystems in each direction.

The second type of decomposition rules are exemplified by D_2, D_3, D_4 . Such decisions are about "whether to join the queue". It is assumed that upon arrival, a customer makes his decision as to whether to join the queue or balk on the basis of number of people already present in the queue. Although it is convenient to think of the decision as customer's decision, in many actual cases, it is the system which makes the decision. If we assume that the queue discipline everywhere is first come, first served, then balking on the basis of queue length is equivalent to balking on the basis of expected waiting times, since the expected waiting time is a constant times the number of people already there.

In the present paper we concentrate on the problem of characterization of the stream emerging from decompositions of these two types, as well as the queueing properties of queues with such inputs.

Before going on to describe the problems attacked in this paper we should point out some of the difficulties to be encountered.

Notice that if we are to be general as to the order of decision points of different types, our analysis will have to assume that arrivals to a decision point are of the most general type of all streams encountered within such systems. In the system of Figure 2, for example, even if the arrivals to D_2 form a Poisson process, the arrivals to D_5 will form a general stream with dependent interarrival times. Therefore we should be prepared to deal with dependent arrival streams at points D_5, D_6, ...

Another requirement to be met with respect to arrivals is the need to identify the customers with different needs and experiences. Returning to the system of Figure 2 for motivation, let us note that the customers arriving at D_5 will include, in general, some customers who needed to be served in one of the queues Q_2, Q_3, Q_4 but got rejected for some reason such as the unavailability of space in waiting room. Assume the system directs such customers towards D_6 for serving them in that branch. Then we need to somehow identify such customers at point D_5. One way to do this is to attach to each customer a tag on which the states of the queues from which he balked (or was rejected) are recorded. In this way we are able to keep a record of the needs of a customer as well as his experiences within the system.

Such a system of identifying customers with their needs and experiences is invaluable for three reasons. First, it is of value whenever the system wants to act (or customers themselves make their

decisions) on the basis of conditions existing in the queues previous to the decision point. Second, it makes it possible to study the correlation between the queues of a system. For example, in the system of Figure 2, the conditions in the queue Q_6 are closely related to those existing in queues Q_2, Q_3, Q_4, since all of the customers that needed to be served at one of Q_2, Q_3, Q_4 are directed to Q_6 if they cannot be served in the queues of their first choice. Clearly, the load on Q_6 is directly correlated with the load on Q_2, Q_3, Q_4, and it may be of interest to study the nature of that correlation. Lastly, this identification of customers with their experiences brings an order to otherwise helplessly complicated streams within the system. We have already mentioned that arrivals to D_5 in Figure 2 will in general form a dependent stream even if the arrivals to D_2 form a Poisson stream. But if the service times in Q_2, Q_3, Q_4 are all negative exponentially distributed, then the two dimensional stochastic process composed of the times between arrivals to D_5 and the queue sizes in Q_2, Q_3, Q_4 at instants of these arrivals to D_5 turns out to be semi-Markovian. All ordered combinations of possible experiences make up the state space of this semi-Markovian process, and the times between arrivals can be called the times between transitions of the process.

If we assume that the arrivals into the system form a semi-Markov process with a finite space, and if only finite lengths of queues are allowed, then all the semi-Markov processes within the system will have finite state spaces. So that the states of a particular SMP can be numbered 1, 2, 3, ..., M where M is finite. Then we

can talk about "customers of type j" in referring to those customers whose needs and experiences are identically the same as the combination numbered j .

Some decision rules will treat customers of different types differently. An example to this was at D_5 where certain types of customers were directed in one direction with a probability equaling unity. Another example would be the balking type of decomposition where the probabilities of balking would depend not only on the queue size encountered but also on the type of customers.

Dependence or non-dependence on the arrival process is a second criteria in classifying decision rules. Put together with the first classification according to dependence on the queueing processes, we obtain the following four different types of decision rules in general.

1) Independent of arrival process and the queueing processes. Examples are: "every other" rule, assignment by independent trials, assignment by Markovian trials as in D_2 , etc.

2) Dependent on arrival process but independent of the queueing processes. Examples are: assignment by customer type as in D_5 , etc.

3) Independent of arrival process but dependent on some of the queueing processes. For example, in Palm and Disney's overflow problem customers join the queue if they can find space; i.e., whether they join the queue or overflow is dependent on queue size. Faced with R queues, customers may decide to join the queue with a minimal length. Faced with a queue as at D_3 of Figure 2, customer may decide to join the queue or move on to the next queue dependent on the size of the queue there.

4) Dependent both on the arrival process and on the queueing processes. Example are: balking from a queue with a probability that depends both on the type of the customer himself and on the queue size at the instant of arrival. Another example would be balking from a queue on the basis of the virtual waiting time at the instant of arrival where different types of customers react differently to the same waiting time.

In the present paper we are concerned with decompositions under these four different types of decision rules as well as the queueing processes in a system with these types of decision rules. Specifically we consider systems of parallel queues with following characteristics:

a) Arrivals into the system form a semi-Markov process with a finite number of states. A semi-Markov process with only one state is equivalent to an ordinary recurrent process; so that renewal type inputs are special cases of the general arrival process we assume.

b) Queue number j has r_j servers whose service times are negative exponentially distributed with mean μ_j ; the waiting room is of size N_j , N_j finite. (We also give a generalization for general independent service times.)

c) Decision points are of the following four types:

1. Sequence of decisions forms a Markov chain which is independent of both the arrival process to that point and the queueing processes in the subsystems.

2. Dependent on arrival process but independent of the queueing processes.

3. Independent of arrival process but dependent on the size of the queue. Only one queue at a time is considered for possible entry.

4. Same as in 3, but the decision as to balk from the queue is dependent on the customer type also.

d) Travel time between two sequential decision points is constant for all customers.

e) No two streams are ever superimposed to make one stream.

This then is the general problem of the paper. As we have mentioned earlier our method of attack is to study the decomposition problem under each different rule separately. Precise descriptions of these problems and a summary of major results will be given in the next section.

3. ORGANIZATION OF THE PAPER AND SUMMARY OF RESULTS

The processes of interest in this paper have finite state spaces, so that we are able to use matrix theory and notation profitably. In the first half of Chapter I, we define and review concepts such as convergence of matrix sequences and series, derivatives and integrals of matrices over function spaces, etc. In the second half of the chapter, rigorous definitions and some properties of Markov Renewal processes and semi-Markov processes are given and some terminology regarding such processes are introduced. We make no claims of originality in this chapter; our object is to introduce some terminology, and to put together some material to make later references to them easier.

In the remaining chapters streams of arrivals to decision points are taken to form semi-Markovian processes. We assume they have M states, (i.e., there are M types of customers,) the first M_1 of which are ergodic and the remaining ones are transient states. Now consider a decision point, and starting with an arrival, number the consecutive arrivals as $0, 1, 2, 3, \ldots$. Let X_n denote the time between the n-th and the $n-1^{st}$ arrivals, and let Z_n be the type of the n-th customer arriving at that point. We take the arrival process as well defined, i.e., we assume

a) the initial distribution

$$p_o(j) = \Pr\{Z_o=j\} ; \quad p_o(j) \geq 0 , \quad (j = 1, 2, \ldots, M);$$
$$\Sigma_j\, p_o(j) = 1$$

and

b) the mass functions

$$A_{ij}(x) = \Pr\{Z_n=j,\ X_n \leq x \mid Z_{n-1}=i\} ; \quad A_{ij}(x) \geq 0 ,$$
$$\Sigma_{j=1}^{M} A_{ij}(\infty) = 1 ; \quad (i,j = 1, 2, \ldots, M;\ x \geq 0)$$

are given.

In Chapter II, the problem of decomposing such an arrival stream into R streams according to a Markovian decision rule is investigated. Let Y_n be the random variable associated with the n-th arrival so that $Y_n = j$ if and only if the n-th customer is sent in the j-th direction, $(j = 1, 2, \ldots, R.)$ Then, we assume $\{Y_n\}$ is a Markov chain each one of whose R states are persistent non-null. We

further assume that the decisions neither depend nor have any effects on the arrival process; i.e., we assume

$$\Pr\{Y_n=j \mid Y_0,Y_1, \ldots, Y_{n-1} ; Z_0, \ldots, Z_n ; X_1, \ldots, X_n\} = \Pr\{Y_n=j \mid Y_{n-1}\} .$$

Then, we show that each one of the R resulting streams is generalized Markov renewal processes with the same state space as the arrival process. Initial distributions and transition matrices are derived for each stream (thus making them well defined), and a complete classification of their states is given, showing that the resulting streams also have M_1 ergodic states and $M-M_1$ transient states. Some steady state results and the mean occupancy times are also given. The decomposition rule is a fairly general one. As its special cases it includes the often used rules such as "every other" discipline, ($Y_n=j$ if and only if $n = j(\text{mod } R)$,) and the independent assignments rule. With such special rules and Poisson arrivals our results agree with the well known results.

Chapter III treats a decomposition problem of the second type, i.e., the decision rule depends on the arrival process. Arrivals to the decision point are again taken to form a SMP with M_1 ergodic and $M-M_1$ transient states. The decision process, $\{Y_n ; n = 0,1,2,\ldots\}$ is such that

$$\Pr\{Y_n=j \mid Y_0, \ldots, Y_{n-1} ; Z_0, \ldots, Z_n ; X_1, \ldots, X_n\} = \Pr\{Y_n=j \mid Z_n\} \quad (j = 1, 2, \ldots, R).$$

We then show that each one of the R resulting streams is again SMP with the same state space as the arrival process. Initial distributions and transition matrices are derived for each stream, a complete classification of their states is given. Steady state results and the mean occupancy times are shown in their relation to the original arrival stream.

The decomposition rule is quite general in its dependence on the type of customer. Some special cases of this rule, for example when $\Pr\{Y_n=j \mid Z_n\} = \Pr\{Y_n=j\}$, overlap with some special cases of the problem of Chapter II, and we note and compare these results.

In Chapter IV we examine the queueing properties of a finite queue (of maximum length N) with a single, negative exponential server subject to a SMP input. As usual the SMP is taken to have M (M is finite) states some of which may be transient. Customers are permitted to balk, i.e., upon their arrival at the decision point they may or may not join the queue. This balking process is taken to be of the most general type, namely a customer of type j balks with a probability $b(j,k)$ if the number of people in the queue at the instant of his arrival is k. Of course, $b(j,k)$ equals unity if $k = N$, i.e., if the queue is full.

We embed the queue size process at epochs of arrivals and write the transition matrix for the resulting semi-Markov process $\{Z_n, S_n, X_n\}$ where S_n is the number of people just before the arrival of the n-th customer. Then, $\{Z_n, S_n\}$ is a Markov chain, and we

investigate the properties of that chain extensively; we classify its states, show the existence of a limiting distribution. In this way we have the joint distribution of the queue size with the customer type. Marginal distribution of the queue size is then easy to obtain.

Aside from the queue size process, we examine and give the distribution of waiting times under first in first out servicing discipline. We also investigate the law of the busy period. We give the distribution of the length of a busy period that was started by a customer of type j, $(j = 1, 2, \ldots, M)$, and find the absolute distribution of the length of a busy period from that.

A special case where the arrival process is a recurrent input (a SMP with $M=1$) has been partly investigated by Finch [9] previously. Finch has shown the existence of the steady state distribution of the queue size and gave that distribution in terms of generating functions.

A special case where balking probabilities depend only on the queue size can be investigated by setting $b(j,k) = b(k)$ everywhere. A further special case where $b(j,k) = 0$ whenever $k < N$ gives the properties of a finite queue with no balking.

Chapter V concentrates on the stream of customers who have balked from the queue of Chapter IV. If the k-th balking customer is the n-th arriving one, we set $\zeta_k = Z_n$, $\psi_k = S_n$; i.e., ζ_k is the type of the k-th balking customer and ψ_k is the state of the

queue size process when he arrived. Further letting Θ_k be the time between the k-th and the $k - 1^{st}$ balkings we prove that

$$\Pr\{\zeta_k=i,\ \psi_k=j,\ \Theta_k \le x \mid \zeta_0, \ldots, \zeta_{k-1};\ \Theta_1, \ldots, \Theta_{k-1};\ \psi_0, \ldots, \psi_{k-1}\}$$
$$= \Pr\{\zeta_k=i,\ \psi_k = j,\ \Theta_k \le x \mid \zeta_{k-1},\ \psi_{k-1}\};$$

i.e., the balking stream is a Markov renewal process whose state space is the Cartesian product of the state spaces of the arrival process and the queue size process. The one-to-one mapping $(\zeta_k,\psi_k) \to \psi_k M + \zeta_k$ maps this space into the set of positive integers, and we can call $\psi_k M + \zeta_k = \overline{Z}_k$ as the type of the k-th balking customer.

We completely define the balking process, i.e., give the initial distribution and the transition matrix. Then, we give a complete characterization of its states, give some steady state results, investigate the absolute distribution and the mean of interbalking times, and derive the distribution of number of customers joining the queue within a balking interval.

This, as far as we know, is the first time the stream of customers balking from a queue has ever been investigated. We do this under quite general assumptions. The question, "when is the balking stream of renewal type?" is answered by giving necessary and sufficient conditions.

In Chapter VI the results of Chapters IV and V are extended in some respects, and some special cases are reviewed. The results of IV and V were for a queue with a single negative exponential server. First, we generalize the results to the case where there are r servers

($r \leq N$) in the queueing system. Then we discuss the problems introduced by letting the single server of queue IV have general independent service times. We show that the processes of interest can still be studied by introducing a dummy variable called unexpended service time at epochs of arrivals.

Next, some special cases of the problem of Chapter V are discussed. Palm's overflow problem [20] is reviewed and his formulas are recovered. Disney's extension [6] of Palm's problem is discussed, and the distribution of the time between overflows is obtained first in the notation of this paper, and then by a formula using no special notation.

Chapter VII is reserved for illustrating the use of the method of this paper in the analysis of a system of parallel queues. A simple system that can nevertheless illustrate the methods of each of the Chapters II, III, IV, V is chosen; and the queueing processes, etc. are analyzed.

4. CONCLUSION

Notice that in all the decompositions studied inputs are taken to be SMP's and it turns out that each decomposed stream is also a SMP. So that, no matter what the order of decision points within the system is, we are able to decompose the network stream by stream and queue by queue. Furthermore, in doing this we lose no information that may be used in studying the correlation between the queue sizes of various queues. In this we are indebted to balking customers for carrying the information about the size of the queue from which they balked in addition to retaining the information they had when they arrived.

Note that we did not need to keep a record of directions taken at decompositions of types examined in Chapters II and III, since for any point within a system of parallel queues there exists only one combination of such directions that lead to that point. This of course is a result of not allowing any two streams to merge.

Although this paper attacks quite a few of the most common problems encountered in parallel queueing systems, there remain several problems to be solved. One is the general problem of R queues in parallel where customers join the queue of shortest length . Another is the investigation of the queueing properties of a single server queue subject to a MRP input of impatient customers, and the characterization of the stream of customers that renege from that queue.

The superposition problem for SMP remains to be solved since in many systems several streams are merged to make one stream. To date, this problem is not even defined.

PART I

DECOMPOSITION OF A SEMI-MARKOVIAN STREAM INTO R STREAMS

CHAPTER I

PRELIMINARIES

1. INTRODUCTION

In this paper we make considerable use of the theory of semi-Markov processes and Markov renewal processes. We assume semi-Markovian arrivals into the systems we study. We show that the balking stream, or decomposed stream etc., is also semi-Markovian. Therefore, we give a rigorous definition and a listing of the more important properties of these processes in Section 3 of this chapter.

The processes of interest in this paper have finite state spaces. Therefore, we are able to use matrix theory and matrix notation profitably. A brief review of those parts of matrix theory which we use will be given next in Section 2.

We make no claims of originality in this chapter. The section on matrix theory is largely from Mirsky [19], whereas the section on Markov renewal and semi-Markov processes is essentially a condensation of papers by Pyke [22,23] and Smith [28,29] on these topics.

2. SOME RESULTS IN MATRIX ANALYSIS

Our objective here is to define and develop concepts such as convergence of matrix sequences, matrix series, etc., as much as we make use of in our work. Proofs and further work on these topics are found in Mirsky [19].

Throughout the following we speak of square matrices of degree M over the field of complex numbers.

a. On the location of eigenvalues

Let A be any matrix. The polynomial $f(x) = \det(xI-A)$ is called the characteristic function of the matrix A. Its zeros are called the eigenvalues of A. A vector y that satisfies $yA = \lambda y$ is said to be a left hand eigenvector of A (corresponding to the eigenvalue λ.)

Let A and B be two matrices of the same degree M over the field of real numbers. Then we will write $A \leq B$ if the inequality sign holds for all corresponding elements of A and B, i.e., if $A = [A_{ij}]$ and $B = [B_{ij}]$, then $A \leq B$ if $A_{ij} \leq B_{ij}$ for all i,j. A matrix A is non-negative if every element of A is a non-negative real number. Let $A = [A_{ij}]$ be a complex matrix, and let $B = [B_{ij}]$ be a non-negative matrix. Then B is said to dominate A if $|A_{ij}| \leq B_{ij}$, $(i,j = 1,2,\ldots,M)$. (Notation: $A \ll B$).

Theorem I.2.1. Let $A = [A_{ij}]$ be a non-negative matrix with $\Sigma_j A_{ij} < 1$ for all i. Then all eigenvalues of A are less than unity in absolute value.

Theorem I.2.2. Let A be a complex matrix and let B be a non-negative matrix such that A is dominated by B, $(A \ll B.)$ Let λ and μ be eigenvalues with maximum absolute value of A and B respectively. Then, $|\lambda| \leq |\mu|$.

b. Convergence of a sequence of matrices

Definition. The sequence $\{A_n\}$ $(A = [A_{ij}^{(n)}])$ of matrices converges to $A = [A_{ij}]$ (notation: $A_n \to A$ as $n \to \infty$, or $\lim_{n\to\infty} A_n = A$) if $\lim_{n\to\infty} A_{ij}^{(n)} = A_{ij}$ $(i,j = 1,2,\ldots,M)$. A sequence that does not converge is said to diverge.

Theorem I.2.3. Let A be a matrix such that $A \ll B$. Then, $A^m \to 0$ if $B^m \to 0$.

Theorem I.2.4. Let A be any matrix. Then, $A^m \to 0$ if and only if all the eigenvalues of A are less than unity in absolute value.

c. Convergence of a series of matrices

Definition. The series of matrices $A_o+A_1+A_2+\ldots$ is said to converge to, or to have the sum, S if the sequence of partial sums $\{S_n\}$, $S_n = A_o+A_1+\ldots+A_n$, converges to S as $n \to \infty$. Clearly, the convergence of $\Sigma_n A_n$ is the same as convergence of series $\Sigma_n A_{ij}^{(n)}$ for all i,j.

Almost all the series we will encounter will be power series. An easy but very useful result is the following.

Theorem I.2.5. Let A be any matrix such that $A^m \to 0$ as $m \to \infty$. Then $I-A$ is non-singular, and the power series $I+A+A^2+A^3+\ldots$ converges to $(I-A)^{-1}$.

More generally, there exists a very close relation between the matrix power series $\Sigma_{m=o}^{\infty} f_m A^m$ and the corresponding scalar power series $\Sigma_{m=o}^{\infty} f_m z^m$.

Theorem I.2.6. If all the eigenvalues of A lie within the circle of convergence of the power series $\phi(z) = \Sigma_{m=0}^{\infty} f_m z^m$, then the matrix power series $\Sigma_{m=0}^{\infty} f_m A^m$ converges absolutely. If at least one eigenvalue of A lies outside the circle of convergence of $\Sigma_{m=0}^{\infty} f_m z^m$, then $\Sigma_{m=0}^{\infty} f_m A^m$ diverges.

This theorem is due to Weyr (1887). A more complete result is the following theorem found by Hensel (1926).

Theorem I.2.7. Let A be a complex matrix. Then the matrix power series $\Sigma_{m=0}^{\infty} f_m A^m$ converges if and only if all eigenvalues of A lie within or on the circle of convergence of the power series $\phi(z) = \Sigma_{m=0}^{\infty} f_m z^m$ and satisfy the further condition that, for every k-fold eigenvalue λ on the circle of convergence, the power series $\phi^{(k-1)}(\lambda)$ is convergent, where $\phi^{(n)}(z)$ is the n-th derivative of $\phi(z)$.

Definition. Let A be any matrix such that $\Sigma_m f_m A^m$ converges, and let $\phi(z) = \Sigma_m f_m z^m$. Then, $\phi(A)$ is defined as the sum of the series $\Sigma_m f_m A^m$.

d. Differentiation and integration of matrices

Let $A(x) = [A_{ij}(x)]$ be a matrix valued function. Another way of thinking of $A(x)$ is as a matrix whose (i,j) entry is a single valued function $A_{ij}(x)$.

Definition. The matrix $A(x)$ is said to be differentiable if all its elements $A_{ij}(x)$ are differentiable. Its derivative is then defined by the formula

$$\frac{d}{dx} A(x) = \left[\frac{d}{dx} A_{ij}(x)\right] .$$

The simple formula $\frac{dz^n}{dz} = nz^{n-1}$ fails to hold for every $A(x)$ in general. Instead we have

$$\frac{d}{dx}\,(A(x)\,B(x)) \;=\; \frac{dA(x)}{dx}\,B(x) + A(x)\frac{dB(x)}{dx} \quad .$$

The following formula can be obtained by successive use of this formula; we have, for $n \geq 1$,

$$\frac{d}{dx}\,A(x)^n \;=\; \sum_{k=0}^{n-1} A(x)^{n-k-1}\,\frac{dA(x)}{dx}\,A(x)^k \quad . \tag{I.2.1}$$

Let $A(x)$ be a matrix such that $A(x)^n \to 0$ as $n \to \infty$ uniformly for all x . Then by Theorem I.2.5 we can write

$$(I-A(x))^{-1} \;=\; I + A(x) + A(x)^2 + A(x)^3 + \ldots \quad ;$$

then, by using the formula (I.2.1) we obtain

$$\frac{d}{dx}\,(I-A(x))^{-1} \;=\; (I-A(x))^{-1}\,\frac{dA(x)}{dx}\,(I-A(x))^{-1}. \tag{I.2.2}$$

Integration of matrices is defined in a way similar to that of differentiation. We denote by $A(x)\,dx$ the matrix whose (i,j) entry is $A_{ij}(x)\,dx$. The matrix $A(x)$ is called integrable, continuous, or bounded respectively if all its elements have the property in question. One special integral is very helpful. Let $A(x)$ be any matrix such that Laplace-Stieltjes transforms of its elements exist.

Then

$$A^*(s) = \int_0^\infty e^{-sx}\, dA(x)$$

is called the Laplace-Stieltjes transform of $A(x)$.

e. Kronecker product

Let $A = [A_{ij}]$ and $B = [B_{ij}]$ be any two matrices. Then the Kronecker product of A and B , denoted $A \oplus B$, is defined as

$$A \oplus B = \begin{bmatrix} A_{11}B & \cdots & A_{1M}B \\ \vdots & \cdots & \vdots \\ A_{M1}B & \cdots & A_{MM}B \end{bmatrix}$$

It can be shown, (cf. Thrall and Tornheim [32],) that if AC and BD are defined, then

$$(A \oplus B)(C \oplus D) = (AC) \oplus (BD) \quad . \tag{I.2.3}$$

f. Notation

Throughout this paper Roman capital letters A, B, ... will denote non-negative matrices; the notation $A(x)$, $B(x)$, ... will be used for matrix valued functions defined over non-negative real numbers. The Laplace-Stieltjes transform of a matrix valued function will be denoted by the same letter with an asterisk above it, e.g., the Laplace-Stieltjes transform of $A(x)$ will be $A^*(s)$. Occasionally we will use $L\{A(x)\}$ for the same thing.

The identity matrix will be denoted by I irrespective of its degree. E will denote a column vector of ones. Where its size becomes important we will write it as a subscript; e.g., E_n is an $n \times 1$ column vector of ones.

3. SEMI-MARKOV PROCESSES AND MARKOV RENEWAL PROCESSES WITH FINITE STATE SPACES

Consider a stochastic process $\{Z(t),\ t \geq 0\}$ which moves from one to another of a finite number of states with the successive states forming a Markov chain, and the process staying in a given state a random length of time, the distribution function of which depends on the initial state as well as on the one to be visited next. Such a process is called a semi-Markov process.

Associated with the process described above is a Markov renewal process which records at each time t, the number of times the process $\{Z(t)\}$ has visited each of the possible states up to time t. We will formalize these descriptions below.

We define a $\{Z,X\}$ process as any two dimensional stochastic process $\{Z_n,\ X_n;\ n \geq 0\}$ defined on a complete probability space $(\Omega, \mathcal{B}, P)$ that satisfies $X_0 = 0$ a.s. and,

$$\Pr\{Z_0=j\} = p_0(j), \quad j = 1,2, \ldots, M;$$

$$p_0(j) \geq 0, \quad \Sigma_{j=1}^{M} p_0(j) = 1;$$

and,

$$\Pr\{Z_n=j,\ X_n \leq x \mid Z_0,Z_1,\ldots,Z_{n-1};\ X_1,X_2,\ldots,X_{n-1}\} = \Pr\{Z_n=j,\ X_n \leq x \mid Z_{n-1}\}.$$

The row vector $p_o = [p_o(1)\ p_o(2)\ \ldots\ p_o(M)]$ is called the vector of initial probabilities. Next let

$$A_{ij}(x) = \Pr\{Z_n=j\ ,\ X_n \le x \mid Z_{n-1} = i\}\ ,\quad n \ge 1\ ;$$

and let $A(x)$ be the matrix whose (i,j) entry is $A_{ij}(x)$, $i,j = 1,2,\ldots,M$; $x \ge 0$. Then the $A_{ij}(x)$ are mass functions satisfying

$$\Sigma_{j=1}^{M} A_{ij}(\infty) = 1\ ,\quad i = 1,2,\ldots,M\ .$$

Further define

$$H_i(x) = \Sigma_{j=1}^{M} A_{ij}(x)\ ,$$

$$\mu_i = \int_o^{\infty} x\, dH_i(x)$$

$$A_{ij} = A_{ij}(\infty)$$

$$F_{ij}(x) = A_{ij}^{-1} A_{ij}(x)\ .$$

Then clearly,

$$H_i(x) = \Pr\{X_n \le x \mid Z_{n-1} = i\}$$

$$F_{ij}(x) = \Pr\{X_n \le x \mid Z_{n-1} = i\ ,\ Z_n = j\}$$

$$A_{ij} = \Pr\{Z_n = j \mid Z_{n-1} = i\}.$$

Now let $T_n = X_o + X_1 + X_2 + \ldots + X_n$, $n \ge 0$. The instants T_n are called regeneration points for the process $\{Z_n,\ X_n\}$.

Next, define an integer valued stochastic process $\{N(t),\ t \geq 0\}$ by

$$N(t) = \sup \{n : T_n \leq t\} ,$$

i.e., $N(t)$ is the number of regenerations that have occurred within the interval $(0,T_n)$. We further define $N_j(t)$ as the number of times $Z_n = j$ for $0 < n \leq N(t)$. Let

$$\mathcal{N}(t) = [N_1(t)\ N_2(t) \ldots N_M(t)] .$$

The stochastic process $\{\mathcal{N}(t),\ t \geq 0\}$ is called the Markov renewal process defined by $(p_o,\ A(x))$.

Related to this Markov renewal process is the stochastic process which records the state of the process at each instant. Let $Z(t) = Z_{N(t)}$, $t \geq 0$. Then, the process $\{Z(t),\ t \geq 0\}$ is called the semi-Markov process defined by $(p_o,\ A(x))$. Clearly there is a one-to-one correspondence between any pair of the processes $\{Z_n,\ X_n\}$, $\{\mathcal{N}(t)\}$, and $\{Z(t)\}$, so that given one the other two can be obtained easily. The process $\{Z_n,\ X_n\}$ is the easiest to formulate and therefore it is used more often. We will say the process $\{Z_n,\ X_n\}$ is equivalent to a semi-Markov process if there exists a semi-Markov process $\{Z(t)\}$ defined by $(p_o,\ A(x))$ where p_o is the vector of initial probabilities and $A(x)$ is the transition matrix for the process $\{Z_n,\ X_n\}$.

Some of the properties of semi-Markov processes are given below as theorems. Proofs of these and further results are found in Pyke [22,23] and Smith [28,29].

Theorem I.3.1. The process $\{Z_n, T_n;\ n \geq 0\}$ is a Markov process, and the process $\{Z_n;\ n \geq 0\}$ is a Markov chain. In particular for $1 \leq j \leq M$, $n \geq 0$

$$\Pr\{Z_{n+1} = j,\ T_{n+1} \leq t \mid Z_0,\ldots,Z_n;\ T_1,\ldots,T_n\} = A_{Z_n\, j}(t-T_n)$$

and,

$$\Pr\{Z_{n+1} = j \mid Z_0,Z_1,\ldots,Z_n\} = A_{Z_n\, j}(\infty) = A_{Z_n\, j} \ .$$

The process $\{Z_n\}$ is said to be the corresponding Markov chain for the process $\{Z(t)\}$. The following properties of the process $\{Z_n, X_n\}$ can easily be obtained from the definitions and the theorem above.

$$\Pr\{X_{n+1} \leq x \mid Z_0,Z_1,\ldots Z_n\} = \Pr\{X_{n+1} \leq x \mid Z_n\} = H_{Z_n}(x) \tag{I.3.1}$$

$$\Pr\{X_{n+1} \leq x \mid Z_0,Z_1,\ldots,Z_n,Z_{n+1}\} =$$
$$= \Pr\{X_{n+1} \leq x \mid Z_n,Z_{n+1}\} = F_{Z_nZ_{n+1}}(x) \ . \tag{I.3.2}$$

$$\Pr\{X_{n_1} \leq x_1,\ X_{n_2} \leq x_2,\ \ldots,\ X_{n_k} \leq x_k \mid Z_n,\ n \geq 0\} =$$
$$= \Pr\{X_{n_1} \leq x_1,\ \ldots,\ X_{n_k} \leq x_k \mid Z_0,Z_1,\ldots Z_{n_k}\}$$
$$= \Pr\{X_{n_1} \leq x_1 \mid Z_{n_1-1},Z_{n_1}\} \cdots \Pr\{X_{n_k} \leq x_k \mid Z_{n_k-1},Z_{n_k}\}$$
$$= \prod_{i=1}^{k} F_{Z_{n_i-1}\, Z_{n_i}}(x_i) \ . \tag{I.3.3}$$

Property given by (I.3.3) is of great importance; it states that $X_{n_1},X_{n_2},\ldots,X_{n_k}$ are "mutually conditionally independent" given $Z_{n_1-1},Z_{n_1};\ Z_{n_2-1},Z_{n_2};\ \cdots;\ Z_{n_k-1},Z_{n_k}$.

a. Classification of states

We now will give a classification of the states of a Markov renewal process or semi-Markov process in much the same way as is done for Markov chains. We try to retain the terminology for Markov chains as introduced by Feller [8].

For $t \geq 0$, let

$$B_{ij}(t) = \Pr\{Z(t) = j \mid Z_0 = i\},$$

$$G_{ij}(t) = \Pr\{N_j(t) > 0 \mid Z_0 = i\},$$

and

$$\lambda_{ij} = \int_0^\infty t \, dG_{ij}(t).$$

According to these definitions, $B_{ij}(t)$ is the probability that a semi-Markov process initially in state i, is in state j at time t. On the other hand, $G_{ij}(t)$ is the probability distribution of the first passage time from state i to state j. Then, λ_{ij} is the expected value of the first passage time from state i to j; in particular, λ_{ii} will be called the mean recurrence time of state i.

Defintions:

a) State j is said to be reachable from state i if $G_{ij}(\infty)$ is positive.

b) States i and j are said to communicate if they are reachable from each other.

c) Communication is an equivalence relation, and the equivalence classes are called classes.

d) A class is said to be closed if no state outside that class is reachable from any state inside.

e) State i is persistent if $G_{ii}(\infty) = 1$. A state i is transient if it is not persistent.

f) State i is said to be ergodic if it is persistent and λ_{ii} is finite.

As can be seen easily, the properties defined here for a semi-Markov process or a Markov renewal process are very closely related to those of the corresponding Markov chains, namely the processes $\{Z_n\}$. In fact we have

Theorem I.3.2. In a semi-Markov process, (i) a state j is persistent (transient) if and only if the state j is persistent (transient) in the corresponding Markov chain; (ii) a class is closed if and only if it is closed in the corresponding Markov chain; (iii) a state j is ergodic if and only if state j is ergodic in the corresponding Markov chain, and μ_k is finite for all states k in the same class as j.

b. Special cases

A semi-Markov process with $X_1 = X_2 = X_3 = \ldots = 1$ is equivalent to a Markov chain. On the other hand, if X_n are independent and identically distributed as $1 - e^{-ax}$, then the process $\{Z(t)\}$ becomes a Markov process. If the number of states M is equal to one, then the Markov renewal process $\{N(t)\}$ becomes equivalent to an ordinary renewal process.

c. Generalized semi-Markov processes

In some cases the process $\{Z_n, X_n\}$ is such that the first step transition probabilities

$$\tilde{A}_{ij}(x) = \Pr\{Z_1 = j , X_1 \leq x \mid Z_o = i\}$$

are different than the probabilities

$$A_{ij}(x) = \Pr\{Z_n = j , X_n \leq x \mid Z_{n-1} = i\} , \quad n \geq 2 .$$

In such a case, the process $\{Z(t)\}$ is called a generalized semi-Markov process; the process $\{Z_n, X_n\}$ is said to be equivalent to a generalized semi-Markov process; and the process $\{\mathcal{N}(t)\}$ is called a generalized Markov renewal process defined by $(p_o, \tilde{A}(x), A(x))$.

4. CONCLUSION

In this chapter we have given some definitions and results that will become useful in our work in succeeding chapters. We have limited ourselves to those subjects that are recurring often enough in the paper, leaving theorems of a special nature to later chapters where they are needed.

CHAPTER II

DECOMPOSITION OF A SEMI-MARKOVIAN STREAM UNDER A MARKOVIAN DECISION RULE

1. INTRODUCTION

Consider a semi-Markov process $\{Z(t),\ t \geq 0\}$. Let the times the process $\{Z(t)\}$ makes its transitions be $T_0, T_1, T_2, \ldots$. Assume that at each of the points $\{T_n\}$ one of R possible events, $E_1, E_2, \ldots, E_R$, can happen where the occurrences of successive events form a Markov chain that neither depends nor has any effect on $\{Z(t)\}$. For a fixed r , let the times the event E_r happens be $\tau_1, \tau_2, \tau_3, \ldots$, and let $\tau_0 = 0$.

In this chapter we are interested in the process $\{\zeta(t),\ t \geq 0\}$, where $\zeta(t) = Z(\tau_k)$ for $\tau_k \leq t < \tau_{k+1}$ $(k \geq 1)$, and $\zeta(t) = \zeta(0)$ for $0 = \tau_0 \leq t < \tau_1$. We will show that $\{\zeta(t)\}$ is a semi-Markov process with the same state space as $\{Z(t)\}$; and we will derive the necessary functions that define the process $\{\zeta(t)\}$ from the ones defining $\{Z(t)\}$.

The application we have in mind is the following. Assume the input to a system of R parallel queues is semi-Markovian. Upon arrival a customer is assigned to one of the R queues so that the probability of the n-th customer being assigned to the j-th queue depends only upon where the n-1st customer was assigned. Then, the times

$T_o, T_1, T_2, \ldots$ are the times of arrivals to the system, event E_r corresponds to an assignment of an arrival to the r-th queue, and $\tau_1, \tau_2, \ldots$ are the times of arrivals to the r-th queue. Then $\{\zeta(t)\}$ is the arrival process to the r-th queue.

This chapter, then, discusses the problem of decomposing a given semi-Markovian stream into R streams by a Markovian decision rule.

2. DEFINITIONS AND NOTATION

Let $\{Z(t),\ t \geq 0\}$ be a semi-Markov process. Let the times $\{Z(t)\}$ makes its transitions be $T_o, T_1, T_2, \ldots$ where $0 = T_o < T_1 < T_2 < T_3 < \ldots$. Let $X_n = T_n - T_{n-1}$, $n \geq 1$, and set $X_o = 0$. Let $Z_n = Z(T_n+0)$, $n \geq 0$. Assume there are M states of $\{Z(t)\}$, states $1,2,\ldots,M_1$ forming an ergodic class, and the remaining ones, M_1+1, M_1+2, ..., M , being transient states. Let $a_o = [a_o(1), a_o(2) \ldots a_o(M)]$ be the vector of initial probabilities, i.e.,

$$a_o(j) = \Pr\{Z_o = j\} \quad (j = 1,2,\ldots,M)\ ; \quad a_o \geq 0\ ; \quad a_o E = 1\ .$$

Define the mass functions

$$A_{ij}(x) = \Pr\{Z_n = j\ ,\ X_n \leq x \mid Z_{n-1} = i\} \quad (i,j = 1,2,\ldots,M)\ ;$$

and let

$$A(x) = [A_{ij}(x)]\ ,$$

and

$$A^*(s) = \int_o^\infty e^{-sx}\, dA(x)\ .$$

The process $\{Z(t)\}$ is well defined once a_o and either of $A(x)$ or $A^*(s)$ are known. Further, let

$$H_i(x) = \Sigma_{j=1}^{M} A_{ij}(x) , \quad \text{and}$$

$$\mu_i = \int_o^\infty x \, dH_i(x) \qquad i = 1,2,\ldots,M .$$

$A_{ij}(x)$ is the joint probability that the process $\{Z(t)\}$, given that it is in state i at instant T_{n-1}, stays there for a time not exceeding x and then enters the state j. $H_i(x)$ is the distribution of the occupancy time of state i; then μ_i becomes the expected value of that occupancy time.

Now assume that at each point T_n one of R possible events, $E_1, E_2, \ldots, E_R$, can happen. Let Y_n be an integer valued random variable which assumes the value j if the event that happens at time T_n is E_j. We assume $\{Y_n; n=0,1,2,\ldots\}$ is a Markov chain with stationary transition probabilities

$$Q_{ij} = \Pr\{Y_n = j \mid Y_{n-1} = i\} \qquad i,j = 1,2,\ldots,R ,$$

and the initial distribution

$$q_o(j) = \Pr\{Y_o = j\} \qquad j = 1,2,\ldots,R .$$

We let Q be the square matrix of degree R whose (i,j) entry is Q_{ij} and let $q_o = [q_o(1) \ldots q_o(R)]$. We assume that every state of $\{Y_n\}$ is persistent non-null; and that the process $\{Y_n\}$ neither depends nor has any effect whatsoever on the process $\{Z(t)\}$.

Choose an integer r, $1 \leq r \leq R$, and fix it once and for all. Let the times the event E_r happens be $\tau_1, \tau_2, \tau_3, \cdots$, $0 \leq \tau_1 < \tau_2 < \cdots$, and set $\tau_o = 0$. Let $\theta_k = \tau_k - \tau_{k-1}$, $k \geq 1$, and set $\theta_o = 0$. We define a new stochastic process $\{\zeta(t),\ t \geq 0\}$ such that

$$\zeta(t) = \begin{cases} \zeta_o & \text{for} \quad 0 \leq t < \tau_1 \quad \text{and} \\ Z(\tau_k) & \text{for} \quad \tau_k \leq t < \tau_{k+1} \quad (k \geq 1), \end{cases}$$

where,

$$\Pr\{\zeta_o = i\} = b_o(i) \qquad i = 1,2,\ldots,M \quad \text{are given.}$$

The object of this chapter is to show that the process $\{\zeta(t),\ t \geq 0\}$ is a generalized semi-Markov process, to determine it and to analyze its properties with respect to those of $\{Z(t)\}$ and $\{Y_n\}$. Accordingly, let $\zeta_k = \zeta(\tau_k + 0)$, $k \geq 1$ and define

$$\tilde{B}_{ij}(x) = \Pr\{\zeta_1 = j,\ \theta_1 \leq x \mid \zeta_o = i\}$$

$$B_{ij}(x) = \Pr\{\zeta_{k+1} = j,\ \theta_{k+1} \leq x \mid \zeta_k = i\} \qquad k \geq 1 ;$$

$$\tilde{B}(x) = [\tilde{B}_{ij}(x)], \qquad B(x) = [B_{ij}(x)],$$

$$\tilde{B}^*(s) = \int_o^\infty e^{-sx}\, d\,\tilde{B}(x), \qquad B^*(s) = \int_o^\infty e^{-sx}\, d\,B(x),$$

$$\tilde{B} = \tilde{B}(\infty) = \tilde{B}^*(0), \qquad B = B(\infty) = B^*(0).$$

3. THE PROCESS $\{\zeta(t)\}$

In this section we will characterize the process $\{\zeta(t)\}$ and will determine it completely. First we give the following main result.

Theorem II.3.1. The process $\{\zeta(t),\ t \geq 0\}$ is a generalized semi-Markov process defined by the triplex $(b_o,\ \tilde{B}(x),\ B(x))$.

Proof. By the definition of a generalized semi-Markov process we only need to prove that

(i) $\Pr\{\zeta(0) = j\} = b_o(j)\ , \qquad j = 1,2,\ldots,M\ ;$

(ii) $\Pr\{\zeta_1 = j\ ,\ \theta_1 \leq x \mid \zeta_o\} = \tilde{B}_{\zeta_o j}(x) \qquad ;$ and

(iii) $\Pr\{\zeta_{k+1} = j\ ,\ \theta_{k+1} \leq x \mid \zeta_o,\zeta_1,\ldots,\zeta_k\ ;\ \theta_1,\theta_2,\ldots,\theta_k\} =$

$= \Pr\{\zeta_{k+1} = j\ ,\ \theta_{k+1} \leq x \mid \zeta_k\}$

$= B_{\zeta_k j}(x) \quad$ for $\quad k \geq 1\ .$

Now, (i) and (ii) follow from definitions. To prove the statement (iii) assume $k \geq 1$. Then,

$$\Pr\{\zeta_{k+1} = j\ ,\ \theta_{k+1} \leq x \mid \zeta_o,\zeta_1,\ldots,\zeta_k\ ;\ \theta_1,\theta_2,\ldots,\theta_k\} =$$

$$= \sum_{m=1}^{\infty} \Pr\{Z_{n+m} = j\ ;\ X_{n+1} + \ldots + X_{n+m} \leq x\ ;\ Y_{n+1} \neq r\ ,\ Y_{n+2} \neq r, \ldots,$$
$$Y_{n+m-1} \neq r\ ,\ Y_{n+m} = r \mid Z_n = \zeta_k,\ Y_n = r;\ \zeta_o,\ldots,\zeta_k;\ \theta_1,\ldots,\theta_k\}$$

$$= \sum_{m=1}^{\infty} \Pr\{Z_{n+m} = j\ ;\ X_{n+1} + \ldots + X_{n+m} \leq x\ ;\ Y_{n+1} \neq r, \ldots, Y_{n+m-1} \neq r,$$
$$Y_{n+m} = r \mid Z_n = \zeta_k,\ Y_n = r\} \quad \text{since } \{Z(t)\} \text{ is a S.-M.P.}$$

$$= \Pr\{\zeta_{k+1} = j\ ,\ \theta_{k+1} \leq x \mid \zeta_k\}$$

$$= B_{\zeta_k j}(x) \qquad \text{, and the proof is complete.}$$

Corollary II.3.1.A. The process $\{\zeta_k,\ k=0,1,2,\ldots\}$ is a Markov chain.

The proof follows from Theorem I.3.1:

$$\Pr\{\zeta_{k+1} = j \mid \zeta_o, \zeta_1, \ldots, \zeta_k\} = \lim_{x\to\infty} \Pr\{\zeta_{k+1} = j,\ \theta_k \le x \mid \zeta_o, \zeta_1, \ldots, \zeta_k\}$$

$$= \lim_{x\to\infty} \Pr\{\zeta_{k+1} = j,\ \theta_k \le x \mid \zeta_k\}$$

$$= \Pr\{\zeta_{k+1} = j \mid \zeta_k\}$$

$$= \begin{cases} \tilde{B}_{\zeta_k j}(\infty) & \text{if } k = 0 \\ B_{\zeta_k j}(\infty) & \text{if } k \ge 1 \ . \end{cases}$$

Corollary II.3.1.B. The two dimensional process $\{\zeta_k, \tau_k\ ;\ k=0,1,2,\ldots\}$ is a Markov process.

The proof again follows from Theorem I.3.1:

$$\Pr\{\zeta_{k+1} = j\ ,\ \tau_{k+1} \le x \mid \zeta_o, \zeta_1, \ldots, \zeta_k\ ;\ \tau_o, \tau_1, \ldots, \tau_k\} =$$

$$= \Pr\{\zeta_{k+1} = j\ ,\ \theta_{k+1} \le x - \tau_k \mid \zeta_o, \ldots, \zeta_k\ ;\ \tau_o, \ldots, \tau_k\}$$

$$= \Pr\{\zeta_{k+1} = j\ ,\ \theta_{k+1} \le x - \tau_k \mid \zeta_k\}$$

$$= \begin{cases} \tilde{B}_{\zeta_k j}(x) & \text{for } k = 0 \\ B_{\zeta_k j}(x-\tau_k) & \text{for } k \ge 1 \ . \end{cases}$$

This theorem and its corollaries characterize the process $\{\zeta(t)\}$. Notice that a study of the process $\{\zeta_n, \theta_n\}$ is equivalent to studying $\{\zeta(t)\}$ since $\{\zeta_n, \theta_n\}$ uniquely defines $\{\zeta(t)\}$. Next, we will derive the probability mass functions $\tilde{B}_{ij}(x)$ and $B_{ij}(x)$ thus completely determining the processes $\{\zeta_n, \theta_n\}$ and $\{\zeta(t)\}$ since the initial distribution vector is already given as b_o. Before, however, we will define several useful quantities and derive them.

a. m-step probabilities in $\{Z_n, X_n\}$

Let the m-step transition probabilities for the process $\{Z_n, X_n\}$ be defined as

$$A_{ij}^{(m)}(x) = \begin{cases} \delta_{ij}\, U(x) & \text{for } m = 0, \text{ and} \\ \Pr\{Z_{n+m} = j,\ X_{n+1} + \ldots + X_{n+m} \le x \mid Z_n = i\} & \text{for } m \ge 1 \end{cases}$$

where

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

is Dirac's delta function, and where

$$U(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \ge 0 \end{cases}$$

is Heaviside's unit step function. Further, let

$$A^{(m)}(x) = \left[A_{ij}^{(m)}(x)\right], \quad L\{A^{(m)}(x)\} = \int_0^\infty e^{-sx}\, dA^{(m)}(x).$$

Now, for $m = 1$, $A_{ij}^{(1)}(x) = A_{ij}(x)$ trivially. For $m \ge 2$ we have,

$$\begin{aligned} A_{ij}^{(m)}(x) &= \Pr\{Z_{n+m} = j,\ X_{n+1} + \ldots + X_{n+m} \le x \mid Z_n = i\} \\ &= \sum_{h=1}^{M} \int_{(o,x)} \Pr\{Z_{n+m} = j\ ;\ X_{n+2} + \ldots + X_{n+m} \le x-y\ ; \\ &\qquad Z_{n+1} = h;\ y-dy < X_{n+1} \le y \mid Z_n = i\} \\ &= \sum_{h=1}^{M} \int_{(o,x)} \Pr\{Z_{n+m} = j;\ X_{n+2} + \ldots + X_{n+m} \le x-y \mid Z_{n+1} = h\} \\ &\qquad \Pr\{Z_{n+1} = h;\ y-dy < X_{n+1} \le y \mid Z_n = i\} \\ &= \sum_{h=1}^{M} \int_0^x A_{hj}^{(m-1)}(x-y)\, dA_{ih}(y). \end{aligned}$$

Using matrix notation and taking Laplace-Stieltjes transforms we obtain,

$$L\{A^{(m)}(x)\} = \begin{cases} I & \text{for } m = 0 \\ A^*(s) & \text{for } m = 1 \\ L\{A^{(m-1)}(x)\} \cdot A^*(s) & \text{for } m \geq 2 . \end{cases}$$

From here, by a simple induction on m, we obtain

$$L\{A^{(m)}(x)\} = A^*(s)^m \qquad m \geq 0$$

where as usual $A^*(s)^0 = I$.

b. First passage times in $\{Y_n\}$

Next, related with the Markov chain $\{Y_n\}$, we define the distribution of the first passage time from state E_i to state E_j:

$$F_{ij}^{(m)} = \Pr\{Y_{n+1} \neq j, Y_{n+2} \neq j, \ldots, Y_{n+m-1} \neq j; Y_{n+m} = j \mid Y_n = i\} \quad m = 1,2,3,\ldots .$$

Then, defining

$$Q_{ij}^{(m)} = \begin{cases} \delta_{ij} & \text{for } m = 0 \\ \Pr\{Y_{n+m} = j \mid Y_n = i\} & \text{for } m \geq 1 \end{cases}$$

we easily have, (cf. Feller [8],)

$$F_{ij}^{(m)} = \begin{cases} Q_{ij} & \text{for } m = 1 \\ Q_{ij}^{(m)} - \sum_{n=1}^{m-1} F_{ij}^{(n)} Q_{jj}^{(m-n)} & \text{for } m \geq 2 . \end{cases}$$

From the Chapman-Kolmogorov equations

$$Q_{ij}^{(m)} = \sum_{h=1}^{R} Q_{ih}^{(m-1)} Q_{hj} \qquad (m \geq 1)$$

$Q_{ij}^{(m)}$ can be calculated, and then $F_{ij}^{(m)}$ can be obtained by the above formula. Let, however,

$$q_{ij}(z) = \sum_{m=o}^{\infty} Q_{ij}^{(m)} z^m \qquad \text{for } |z| < 1 ,$$

and

$$f_{ij}(z) = \sum_{m=1}^{\infty} F_{ij}^{(m)} z^m \qquad \text{for } |z| \leq 1 ,$$

and further let

$$q(z) = [q_{ij}(z)] .$$

Then, since $Q_{ij}^{(m)}$ is the (i,j) entry of Q^m, we have

$$q(z) = \sum_{m=o}^{\infty} Q^m z^m .$$

Lemma II.3.2. For $|z| < 1$, $I-zQ$ is non-singular, and the matrix power series $q(z) = \sum_{m=o}^{\infty} (zQ)^m$ converges to $(I-zQ)^{-1}$. Then $q_{ij}(z)$ is the (i,j) entry of $(I-zQ)^{-1}$.

Proof. For $|z| < 1$, the matrix zQ is dominated by $|z|Q$. Then, since $QE = E$, $|z|QE = |z|E < E$. Thus, by Theorem I.2.1, all eigenvalues of $|z|Q$ are less than unity in absolute value; and hence, (by Theorem I.2.4,) $(|z|Q)^m \to 0$ as $m \to \infty$. Then, (by Theorem I.2.3,) $(zQ)^m \to 0$ as $m \to 0$ since $zQ \ll |z|Q$; and, hence by Theorem I.2.5, $I-zQ$ is non-singular, and the matrix power series $q(z) = \sum_{m=o}^{\infty} (zQ)^m$ converges to $(I-zQ)^{-1}$ for all z with $|z| < 1$. From the definition of $q(z)$ its (i,j) entry is $q_{ij}(z)$.

Next, from the formulas for $F_{ij}^{(m)}$, for the generating function $f_{ij}(z)$ we have

$$f_{ij}(z) = Q_{ij}\, z + \sum_{m=2}^{\infty} \left(Q_{ij}^{(m)} - \sum_{n=1}^{m-1} F_{ij}^{(n)}\, Q_{jj}^{(m-n)}\right) z^m$$

$$= \sum_{m=1}^{\infty} Q_{ij}^{(m)}\, z^m - \sum_{n=1}^{\infty} \sum_{m=n+1}^{\infty} F_{ij}^{(n)}\, z^n\, Q_{jj}^{(m-n)}\, z^{m-n}$$

$$= q_{ij}(z) - Q_{ij}^{(o)} - f_{ij}(z)\,(q_{jj}(z) - Q_{jj}^{(o)})$$

$$= q_{ij}(z) - \delta_{ij} - f_{ij}(z)\,(q_{jj}(z) - 1)$$

so that

$$f_{ij}(z)\, q_{jj}(z) = q_{ij}(z) - \delta_{ij}\,,$$

and hence

$$f_{ij}(z) = \begin{cases} q_{ij}(z)/q_{jj}(z) & \text{for } i \neq j \\ 1 - 1/q_{jj}(z) & \text{for } i = j\,. \end{cases}$$

These equations are well known (cf. Kemperman [15].) Notice that, since $\{Y_n\}$ was assumed to have only persistent non-null states, $f_{ij}(z)$ converge for $|z| = 1$. These results we put below as

Theorem II.3.3. The generating function $f_{ij}(z) = \sum_{m=1}^{\infty} F_{ij}^{(m)}\, z^m$ is given by

$$f_{ij}(z) = \begin{cases} q_{ij}(z)/q_{jj}(z) & \text{if } i \neq j \\ 1 - 1/q_{jj}(z) & \text{if } i = j \end{cases}$$

where $q_{ij}(z)$ is the (i,j) entry of $(I - zQ)^{-1}$, $(|z| < 1)$.

c. Derivation of $B_{ij}(x)$ and $\tilde{B}_{ij}(x)$

We now return to the derivation of $B_{ij}(x)$. From the proof of Theorem II.3.1 we already have

$$B_{ij}(x) = \Pr\{\zeta_k = j\ ,\ \Theta_k \le x \mid \zeta_{k-1} = i\}$$
$$= \sum_{m=1}^{\infty} \Pr\{Z_{n+m} = j\ ;\ X_{n+1} + \ldots + X_{n+m} \le x\ ;\ Y_{n+1} \neq r\ ,\ \ldots,$$
$$Y_{n+m-1} \neq r\ ,\ Y_{n+m} = r \mid Z_n = i\ ,\ Y_n = r\}\ .$$

Since the processes $\{Z_n, X_n\}$ and $\{Y_n\}$ are completely independent of each other we can write

$$B_{ij}(x) = \sum_{m=1}^{\infty} \Pr\{Z_{n+m} = j,\ X_{n+1} + \ldots + X_{n+m} \le x \mid Z_n = i\}$$
$$\Pr\{Y_{n+1} \neq r,\ Y_{n+2} \neq r,\ \ldots,\ Y_{n+m-1} \neq r,\ Y_{n+m} = r \mid Y_n = r\}$$
$$= \sum_{m=1}^{\infty} A_{ij}^{(m)}(x)\ F_{rr}^{(m)}\ ,$$

or in terms of Laplace-Stieltjes transforms in matrix notation

$$B^*(s) = \sum_{m=1}^{\infty} F_{rr}^{(m)}\ A^*(s)^m\ .$$

Of course, for this formula to be meaningful, the matrix power series $\Sigma_{m=1}^{\infty} F_{rr}^{(m)}\ A^*(s)^m$ must converge. This we will prove below. First, however, we give the following

Lemma II.3.4. For the matrix power series $\Sigma_{m=1}^{\infty} F_{rr}^{(m)}\ A^*(s)^m$ $(\mathrm{Re}\{s\} \ge 0)$ to converge, it is sufficient that $\Sigma_{m=1}^{\infty} F_{rr}^{(m)}\ A^m$ converge.

Proof. Consider the (i,j) entries of $A^*(s)^m$ and A^m, i.e., $L\{A_{ij}^{(m)}(x)\}$ and $A_{ij}^{(m)} = A_{ij}^{(m)}(\infty)$ respectively.

$$|L\{A_{ij}^{(m)}(x)\}| = \left|\int_0^{\infty} e^{-sx}\ d\ A_{ij}^{(m)}(x)\right| \le \int_0^{\infty} |e^{-sx}|\ dA_{ij}^{(m)}(x) \le$$
$$\le \int_0^{\infty} dA_{ij}^{(m)}(x) = A_{ij}^{(m)}(\infty) = A_{ij}^{(m)} \qquad \text{for all}$$

s with non-negative real parts. Hence, by Weierstrass' dominated convergence theorem, $\Sigma_{m=1}^{\infty} F_{rr}^{(m)}\ L\{A_{ij}^{(m)}(x)\}$ converges absolutely if $\Sigma_{m=1}^{\infty} F_{rr}^{(m)}\ A_{ij}^{(m)}$ converges. Restating in matrix language we obtain the Lemma.

Theorem II.3.5. The matrix power series $B^*(s) = \Sigma_{m=1}^{\infty} F_{rr}^{(m)} A^*(s)^m$ converges to $f_{rr}(A^*(s))$ for all s with $\mathrm{Re}\{s\} \geq 0$, where $f_{rr}(z) = \Sigma_{m=1}^{\infty} F_{rr}^{(m)} z^m$.

Proof. If $\Sigma_m F_{rr}^{(m)} A^*(s)^m$ converges, its sum is $f_{rr}(A^*(s))$ by the definition of $f_{rr}(A^*(s))$. To prove that $\Sigma_m F_{rr}^{(m)} A^*(s)^m$ converges, in view of Lemma II.3.4, it is sufficient to prove that $\Sigma_m F_{rr}^{(m)} A^m$ converges. Now, by the assumptions about $\{Z(t)\}$, the states $1,2,\ldots,M_1$ are ergodic and form one class. By Theorem I.3.2, these states are ergodic and form one class in the corresponding Markov chain $\{Z_n\}$. Then, since A is the transition matrix for this chain, $\lim_{n\to\infty} A^n$ exists. Hence, by Perron's theorem, the eigenvalue of A with maximum absolute value is 1 itself, which is a simple root of the characteristic function. Since state E_r is persistent, $f_{rr}(1) = 1$. Hence, by the theorem of Hensel (Theorem I.2.7), $\Sigma_{m=1}^{\infty} F_{rr}^{(m)} A^m$ converges. Proof is complete.

This completes the derivation of transition probabilities $B_{ij}(x)$, actually we have obtained their Laplace-Stieltjes transforms; in matrix notation,

$$B^*(s) = f_{rr}(A^*(s)) \tag{II.3.1}$$

where

$$f_{rr}(z) = \Sigma_{m=1}^{\infty} F_{rr}^{(m)} z^m .$$

To complete the identification of the process $\{\zeta(t)\}$ we next derive the first-step transition probabilities $\tilde{B}_{ij}(x)$. Now,

$$\tilde{B}_{ij}(x) = \Pr\{\zeta_1 = j,\ \theta_1 \le x \mid \zeta_0 = i\}$$

$$= \Pr\{Z_0 = j\}\ \Pr\{Y_0 = r\} \qquad \text{if } x = 0$$

$$= \sum_{m=1}^{\infty} \sum_{\substack{h=1 \\ h\neq r}}^{R} \sum_{k=1}^{M} \Pr\{Y_0 = h\}\ \Pr\{Y_1 \neq r,\ Y_2 \neq r, \ldots, Y_{m-1} \neq r,\ Y_m = r \mid Y_0 = h\} \cdot$$

$$\Pr\{Z_0 = k\}\ \Pr\{Z_m = j,\ X_1 + \ldots + X_m \le x \mid Z_0 = k\} +$$

$$\Pr\{Z_0 = j\} \cdot \Pr\{Y_0 = r\} \qquad \text{if } x > 0 .$$

Hence,

$$\tilde{B}_{ij}(x) = a_0(j)\ q_0(r)\ U(x) + \sum_{m=1}^{\infty} \sum_{\substack{h=1 \\ h\neq r}}^{R} \sum_{k=1}^{M} q_0(h)\ F_{hr}^{(m)}\ a_0(k)\ A_{kj}^{(m)}(x) \quad .$$

Note that $\tilde{B}_{ij}(x)$ is independent of i, thus $\tilde{B}(x)$ will have all of its rows identically equal to

$$a_0 \left(q_0(r) I + \sum_{m=1}^{\infty} \sum_{\substack{h=1 \\ h\neq r}}^{R} q_0(h)\ F_{hr}^{(m)}\ A^{(m)}(x) \right) \quad .$$

Taking Laplace-Stieltjes transforms we obtain as a row of $\tilde{B}^*(s)$,

$$a_0 \left(q_0(r) I + \sum_{\substack{h=1 \\ h\neq r}}^{R} q_0(h) \left(\sum_{m=1}^{\infty} F_{hr}^{(m)}\ A^*(s)^m \right) \right).$$

By a theorem very similar to Theorem II.3.5, $\Sigma_m\ F_{hr}^{(m)}\ A^*(s)^m$ can be shown to converge to $f_{hr}(A^*(s))$. Substituting this in the formula, we obtain

$$\tilde{B}^*(s) = E_M \oplus \left(a_0 \left(q_0(r)\ I + \sum_{\substack{h=1 \\ h\neq r}}^{R} q_0(h)\ f_{hr}\ (A^*(s)) \right) \right) \quad \text{(II.3.2)}$$

where "$\oplus$" stands for the Kronecker product (cf. page 27).

The formulas II.3.1 and II.3.2 completely identify the process $\{\zeta(t)\}$ since b_0, the initial distribution vector, is given.

4. THE PROCESS $\{\zeta_n\}$

The Markov chain $\{\zeta_n\}$, (cf. Corollary II.3.1A,) is closely related to the process $\{\zeta(t)\}$. Especially, the classification of the states of the process $\{\zeta(t)\}$ is very much the same as the classification of the states of the process $\{\zeta_n\}$. Furthermore, limiting properties of $\{\zeta(t)\}$ depend on $\{\zeta_n\}$ very strongly. In this section, we will examine the process $\{\zeta_n\}$, classify its states, and find its limiting distribution, leaving the examination of $\{\zeta(t)\}$ to the next section.

First of all, notice that the state space for $\{\zeta_n\}$ is the same as the one for the chain $\{Z_n\}$. $\{\zeta_n\}$ is defined by $(b_o, \tilde{B}, B)$ where $\tilde{B} = \tilde{B}^*(0)$ and $B = B^*(0)$. Further, in studying $\{\zeta_n\}$ we will only concentrate on $\{\zeta_n; n = 1,2,\ldots\}$ since the first step transition probabilities do not depend on the initial distribution as given by b_o anyway.

By the assumptions about the states of the process $\{Z(t)\}$, the states $1,2,\ldots,M_1$ are ergodic, and the remaining states $M_1+1,\ldots,M$ are transient. Then, by Theorem I.3.2, the states $1,2,\ldots,M_1$ are ergodic, and $M_1+1,\ldots,M$ are transient in the chain $\{Z_n\}$. Thus the transition matrix A of the chain $\{Z_n\}$ can be partitioned accordingly to obtain

$$A = \begin{bmatrix} A_1 & 0 \\ A_3 & A_2 \end{bmatrix}$$

where A_1 is obtained by keeping only the first M_1 columns and rows of the matrix A, and the matrices A_2, A_3 are defined accordingly.

Now, since

$$A^n = \begin{bmatrix} A_1^n & 0 \\ A_3^{(n)} & A_2^n \end{bmatrix}$$

we have

$$B = \sum_{n=1}^{\infty} F_{rr}^{(n)} A^n = \sum_{n=1}^{\infty} F_{rr}^{(n)} \begin{bmatrix} A_1^n & 0 \\ A_3^{(n)} & A_2^{(n)} \end{bmatrix} = \begin{bmatrix} \sum_{n=1}^{\infty} F_{rr}^{(n)} A_1^n & 0 \\ \sum_{n=1}^{\infty} F_{rr}^{(n)} A_3^{(n)} & \sum_{n=1}^{\infty} F_{rr}^{(n)} A_2^n \end{bmatrix}.$$

Thus, setting $B_1 = f_{rr}(A_1) = \sum_{n=1}^{\infty} F_{rr}^{(n)} A_1^n$, $B_2 = f_{rr}(A_2) = \sum_{n=1}^{\infty} F_{rr}^{(n)} A_2^n$, and $B_3 = \sum_{n=1}^{\infty} F_{rr}^{(n)} A_3^{(n)}$, we can write

$$B = \begin{bmatrix} B_1 & 0 \\ B_3 & B_2 \end{bmatrix}.$$

This partitioning suggests the theorem below.

Theorem II.4.1. State j of the process $\{\zeta_n\}$ is ergodic (transient) if it is ergodic (transient) in the process $\{Z_n\}$. Furthermore, all ergodic states in $\{\zeta_n\}$ are in the same class.

Proof. We will prove that the states $1,2,\ldots,M_1$ in $\{\zeta_n\}$ are ergodic and form one class, and that the states $M_1+1, M_1+2,\ldots,M$ in $\{\zeta_n\}$ are transient.

Since the states $1,2,\ldots,M_1$ in $\{Z_n\}$ are in the same class of ergodic states the matrix A_1 is regular, i.e., there exists a positive integer k such that A_1^k has no zero elements. Now let h be the smallest integer for which $F_{rr}^{(h)}$ is positive. Clearly, $h \geq 1$. Then,

$$B_1 = \Sigma_m F_{rr}^{(m)} A_1^m \geq F_{rr}^{(h)} A_1^h .$$

Hence, since $A_1^k > 0$,

$$B_1^k \geq (F_{rr}^{(h)} A_1^h)^k = (F_{rr}^{(h)})^k (A_1^k)^h > 0 .$$

Thus $B_1^n > 0$ for some n , (for example k;) hence all of the states $1,2,\ldots,M_1$ form one class, and they are all ergodic since not all of the states of a finite Markov chain, (i.e., the chain defined by B_1 in part,) can be transient.

To show that the states $M_1+1,\ldots,M$ of $\{\zeta_n\}$ are transient we need only to show that $B_2^n \to 0$ as $n \to \infty$. Now, let TA_2T^{-1} be in the classical canonical form. Then, $T A_2^n T^{-1} = (TA_2T^{-1})^n$ is upper triangular, and hence so is

$$TB_2T^{-1} = T(\Sigma_n F_{rr}^{(n)} A_2^n)T^{-1} = \Sigma_n F_{rr}^{(n)} T A_2^n T^{-1} .$$

Now let the eigenvalues of B_2 be $y_{M_1+1},\ldots,y_M$, and those of A_2 be $x_{M_1+1},\ldots,x_M$. Clearly, these are the diagonal elements of TB_2T^{-1} and TA_2T^{-1} respectively. Hence,

$$y_i = \Sigma_n F_{rr}^{(n)} x_i^n \qquad i = M_1+1,\ldots,M .$$

Now, since the states $M_1+1,\ldots,M$ are transient in $\{Z_n\}$, $A_2^n \to 0$ as $n \to \infty$. Thus, (by Theorem I.2.4,) $|x_i| < 1$; hence,

$$|y_i| = |\Sigma_n F_{rr}^{(n)} x_i^n| \leq \Sigma_n F_{rr}^{(n)} |x_i|^n < \Sigma_n F_{rr}^{(n)} = 1 .$$

But, by Theorem I.2.4 again, $|y_i| < 1$ implies that $B_2^n \to 0$ as $n \to \infty$. Hence, the states $M_1+1,\ldots,M$ are transient in $\{\zeta_n\}$. Proof is complete.

From the above theorem we are assured of the existence of a limiting distribution for $\{\zeta_n\}$. Let a be the limiting distribution vector for $\{Z_n\}$, i.e., let $a = [a(1)\ a(2)\ \ldots\ a(M)]$ where $a(j) = \lim_{n\to\infty} \Pr\{Z_n=j\}$.

Theorem II.4.2. $\lim_{n\to\infty} B^n$ exist, and $\lim_{n\to\infty} B^n = \lim_{n\to\infty} A^n$, or equivalently, $\lim_{n\to\infty} \Pr\{\zeta_n=j\} = \lim_{n\to\infty} \Pr\{Z_n=j\}$ for $j = 1,2,\ldots,M$.

Proof. The existence of $\lim B^n$ is obvious in view of the preceding theorem. Now let $\overset{\infty}{A} = \lim_{n\to\infty} A^n$. Then, clearly, every row of $\overset{\infty}{A}$ is equal to the limiting distribution vector a, and $aA = a$ and $aE = 1$. Now from $B = \Sigma_n F_{rr}^{(n)} A^n$ we have, since $\Sigma_n F_{rr}^{(n)} = 1$,

$$aB = a \sum_n F_{rr}^{(n)} A^n = \sum_n F_{rr}^{(n)} aA^n = \sum_n F_{rr}^{(n)} a = a .$$

Since the limiting distribution for $\{\zeta_n\}$ is unique, and $aB = a$ and $aE = 1$, a is the limiting distribution vector for $\{\zeta_n\}$ also. Thus, every row of $\overset{\infty}{B} = \lim_{n\to\infty} B^n$ is equal to a, and hence, $\overset{\infty}{B} = \overset{\infty}{A}$. Proof is complete.

The similarity of the chains $\{\zeta_n\}$ and $\{Z_n\}$ is obvious since the process $\{Y_n\}$ is completely independent of the process $\{Z_n\}$. The difference in the processes $\{\zeta(t)\}$ and $\{Z(t)\}$ lie in their respective interval processes. We examine this in the next section.

5. CLASSIFICATION OF THE STATES OF $\{\zeta(t)\}$

We have already given the classification of the states of the corresponding Markov chain, $\{\zeta_n\}$. Then, by Theorem I.3.2, the states M_1+1, M_1+2, ..., M are all transient in the process $\{\zeta(t)\}$ since they are transient in the corresponding Markov chain $\{\zeta_n\}$. On the other hand, to say anything about the ergodicity of the states $1,2,\ldots,M_1$ we need to examine the mean occupancy times in those states.

Let

$$G_i(x) = \Pr\{\theta_{n+1} \le x \mid \zeta_n = i\} \qquad n \ge 1 ,$$

$$\tilde{G}_i(x) = \Pr\{\theta_1 \le x \mid \zeta_o = i\} ,$$

and

$$\eta_i = \int_o^\infty x \, d\, G_i(x) ,$$

$$\tilde{\eta}_i = \int_o^\infty x \, d\, \tilde{G}_i(x) .$$

$G_i(x)$ is the distribution function of the occupancy times of state i, and η_i is the mean value of it, both defined for steps later than the first one. $\tilde{G}_i(x)$ and $\tilde{\eta}_i$ are the same things as $G_i(x)$ and η_i except that they are defined for the first step. We already have defined the analogs of $G_i(x)$ and η_i for the process $\{Z(t)\}$ as $H_i(x)$ and μ_i in page 37. Clearly, $H_i(x) = \Sigma_j A_{ij}(x)$, $G_i(x) = \Sigma_j B_{ij}(x)$, $\tilde{G}_i(x) = \Sigma_j \tilde{B}_{ij}(x)$. Now letting μ, η, $\tilde{\eta}$ as the column vectors with μ_i, η_i, $\tilde{\eta}_i$ as elements respectively, we have

$$\mu = - \frac{d}{ds} A^*(s)\Big|_{s=o} E , \quad \eta = - \frac{d}{ds} B^*(s)\Big|_{s=o} E , \quad \tilde{\eta} = - \frac{d}{ds} \tilde{B}^*(s)\Big|_{s=o} E .$$

Now;

$$\eta = - \frac{d}{ds} B^*(s)\Big|_{s=o} E = - \frac{d}{ds} \left(\Sigma_{n=1}^\infty F_{rr}^{(n)} A^*(s)^n\right)\Big|_{s=o} E$$

$$= \Sigma_{n=1}^\infty F_{rr}^{(n)} \left(- \frac{d}{ds} A^*(s)^n\right)_{s=o} E , \quad \text{now using the formula I.2.1,}$$

$$= \Sigma_{n=1}^\infty F_{rr}^{(n)} \left(\Sigma_{k=o}^{n-1} A^*(s)^{n-k-1} \left(- \frac{dA^*(s)}{ds}\right) A^*(s)^k\right)_{s=o} E$$

$$= \Sigma_{n=1}^\infty F_{rr}^{(n)} \Sigma_{k=o}^{n-1} A^{n-k-1} \left(- \frac{dA^*(s)}{ds}\right)_{s=o} A^k E$$

$$= \Sigma_{n=1}^\infty F_{rr}^{(n)} \Sigma_{k=o}^{n-1} A^{n-k-1} \left(- \frac{dA^*(s)}{ds}\right)_{s=o} E \qquad \text{since}$$

$A^k E = E$ for all k. Futher noticing that $\Sigma_{k=o}^{n-1} A^{n-k-1} = \Sigma_{k=o}^{n-1} A^k$, and that $\left(- \frac{dA^*(s)}{ds}\right)_{s=o} E = \mu$ we obtain

$$\eta = \Sigma_{n=1}^{\infty} \Sigma_{k=o}^{n-1} F_{rr}^{(n)} A^k \mu \qquad \text{(II.5.1)}$$

Now multiply both sides of (II.5.1) by the limiting distribution vector a, and we get, since $aA^k = a$,

$$\begin{aligned} a\eta &= \Sigma_{n=1}^{\infty} \Sigma_{k=o}^{n-1} F_{rr}^{(n)} aA^k\mu \\ &= \Sigma_{n=1}^{\infty} \Sigma_{k=o}^{n-1} F_{rr}^{(n)} a\mu \\ &= (\Sigma_{n=1}^{\infty} n F_{rr}^{(n)}) a\mu \\ a\eta &= \lambda_r a\mu \end{aligned} \qquad \text{(II.5.2)}$$

where $\lambda_r = \Sigma_{n=1}^{\infty} n F_{rr}^{(n)}$ is the mean recurrence time for state E_r in the Markov chain $\{Y_n\}$.

Next multiplying both sides of (II.5.1) by $I-A$, and noting that $(I-A)(I+A+A^2+\ldots+A^{n-1}) = I-A^n$ we get

$$\begin{aligned} (I-A)\eta &= \Sigma_{n=1}^{\infty} F_{rr}^{(n)} (I-A)(I+A+A^2+\ldots+A^{n-1}) \mu \\ &= \Sigma_{n=1}^{\infty} F_{rr}^{(n)} (I-A^n) \mu \\ &= (I - \Sigma_{n=1}^{\infty} F_{rr}^{(n)} A^n) \mu \\ (I-A)\eta &= (I - B) \mu \end{aligned} \qquad \text{(II.5.3)}$$

Notice that both $I-A$ and $I-B$ are of the same rank, namely of rank $M-1$. Hence, the non-homogeneous system of linear equations (II.5.3) in M unknowns, $\eta_1, \eta_2, \ldots, \eta_M$ does have a solution. Since a is independent of the rows of $I-A$ or $I-B$, the equation system

$$\begin{aligned} (I-A)\eta &= (I-B)\mu \\ a\eta &= \lambda_r a \mu \end{aligned} \qquad \text{(II.5.4)}$$

has a unique solution for η. (Clearly, one of the first M_1 equations provided by $(I-A)\eta = (I-B)\mu$ is superfluous and can safely be discarded.)

We next need the mean occupancy times in the first step. From our derivation of $\tilde{B}_{ij}(x)$ we found that $\tilde{B}_{ij}(x)$ were independent of i. Thus, the expected time of Θ_1 is the same for all values of ζ_o, i.e., $\tilde{\eta}_1 = \tilde{\eta}_2 = \ldots = \tilde{\eta}_M$. Then,

$$\begin{aligned}
\tilde{\eta}_1 &= -\frac{d}{ds}\left\{a_o\,(I\, q_o(r) + \sum_{\substack{h=1\\h\neq r}}^{R}\sum_{m=1}^{\infty} q_o(h)\, F_{hr}^{(m)}\, A^*(s)^m)\right\}\Big|_{s=o}\mathbf{E} \\
&= a_o \sum_{\substack{h=1\\h\neq r}}^{R}\sum_{m=1}^{\infty} q_o(h)\, F_{hr}^{(m)} \sum_{k=o}^{m-1} A^{m-k-1}\left(-\frac{d}{ds}A^*(s)\right)\Big|_{s=o} A^k\mathbf{E} \\
&= \sum_{h\neq r}\sum_{m} q_o(h)\, F_{hr}^{(m)} \sum_{k=o}^{m-1} a_o A^k\,\mu \\
&= \sum_{h\neq r}\sum_{m} q_o(h)\, F_{hr}^{(m)} \sum_{k=o}^{m-1} a_k\,\mu \qquad \text{(II.5.5)}
\end{aligned}$$

where $a_k = a_o A^k$ is the row vector of probabilities $\Pr\{Z_k = i\}$.

Lemma II.5.1. The expected value of Θ_1 is finite.

Proof. The expected value of Θ_1 is given by (II.5.5). To show that it is finite let $\mu_{max} = \sup\{\mu_i\ ;\ i=1,2,\ldots,M\}$, and $\lambda_{max} = \sup\{\lambda_{hr}\ ;\ h=1,2,\ldots,R\}$ where $\lambda_{hr} = \sum_{m=1}^{\infty} m\, F_{hr}^{(m)}$ is the mean first passage time from state E_h to E_r in the chain $\{Y_n\}$. Then clearly $\mu \leq \mu_{max}\mathbf{E}$, and hence,

$$\begin{aligned}
\tilde{\eta}_1 &= \sum_{h\neq r}\sum_{m=1}^{\infty}\sum_{k=o}^{m-1} q_o(h)\, F_{hr}^{(m)}\, a_k\mu \leq \sum_{h\neq r}\sum_{m=1}^{\infty}\sum_{k=o}^{m-1} q_o(h)\, F_{hr}^{(m)}\, a_k\mathbf{E}\,\mu_{max} = \\
&= \sum_{h\neq r}\sum_{m=1}^{\infty}\sum_{k=o}^{m-1} q_o(h)\, F_{hr}^{(m)}\,\mu_{max} = \sum_{h\neq r}\sum_{m=1}^{\infty} q_o(h)\, m\, F_{hr}^{(m)}\,\mu_{max} = \\
&= \sum_{h\neq r} q_o(h)\,\lambda_{hr}\,\mu_{max} \leq \sum_{h\neq r} q_o(h)\,\lambda_{max}\,\mu_{max} \leq \lambda_{max}\,\mu_{max}\,.
\end{aligned}$$

Since state $M_1+1,\ldots,M$ are transient in $\{Z(t)\}$, $\mu_{M_1+1},\ldots,\mu_M$ are finite; and since states $1,2,\ldots,M_1$ are ergodic in $\{Z(t)\}$,

$\mu_1, \ldots, \mu_M$ are finite by Theorem I.3.2. Hence, μ_i is finite for all i, and hence μ_{max} is finite. On the other hand, since all of the states of $\{Y_n\}$ are persistent non-null, λ_{hr} is finite for all h, and hence λ_{max} is finite. Thus, $\tilde{\eta}_1 \leq \lambda_{max}\ \mu_{max} < \infty$ also.

Lemma II.5.2. The mean occupancy times η_i are finite.

Proof. From the proof of the preceding lemma μ_i is finite for all i. Also, since E_r is persistent non-null in $\{Y_n\}$, λ_r is finite. Then, the solution η of the system of linear equations (II.5.4) is finite, i.e., η_i is finite for all i.

Theorem II.5.3. State j of the process $\{\zeta(t)\}$ is ergodic transient) if it is ergodic (transient) in the process $\{Z(t)\}$. Furthermore, all ergodic states in $\{\zeta(t)\}$ are in the same class.

Proof follows from Theorems II.4.1 and I.3.2, and Lemmas II.5.1 and II.5.2.

6. SPECIAL CASES AND APPLICATIONS

Consider R queues in parallel. Assume the arrivals into such a system form a semi-Markov process. Each arriving customer is assigned to one of the queues on the basis of the assignment of the last customer. Then, we have shown that the arrivals to the r-th queue form a semi-Markov process.

Example II.6.1. One of the most common rules of assignment is "every other" rule, i.e., the first customer goes to the first queue, second customer to second queue, and so on. In this case, the transition probabilities for the chain $\{Y_n\}$ are given as

$$Q_{ij} = \begin{cases} 1 & \text{if } j \equiv i + 1 \pmod R \\ 0 & \text{otherwise.} \end{cases}$$

Then, clearly,

$$F_{rr}^{(n)} = \begin{cases} 1 & \text{for } n = R \\ 0 & \text{otherwise,} \end{cases}$$

and $$F_{1r}^{(n)} = \begin{cases} 1 & \text{for } n = r-1 \\ 0 & \text{otherwise,} \end{cases}$$

and

$$f_{1r}(z) = z^{r-1} , \quad f_{rr}(z) = z^R , \quad (r > 1) .$$

Since $\Pr\{Y_o = 1\} = 1$, for the r-th queue we obtain

$$\tilde{B}^*(s) = E_M \Theta (a_o A^*(s)^{r-1}) , \quad \text{(cf. formula II.3.2,)}$$

and

$$B^*(s) = A^*(s)^R , \quad \text{(cf. formula II.3.1.)}$$

This last formula is very well known in at least one special case: when arrivals form a Poisson process. Then, there exists only one state $\{Z(t)\}$, and the X_n are distributed as $1 - e^{-\alpha x}$ where α is the mean arrival rate. It is well known that the arrivals to any of the queues form an Erlang-R process; i.e., $A^*(s) = \alpha/(\alpha+s)$ and thus $B^*(s) = \alpha^R/(\alpha+s)^R$.

Example II.6.2. Another special assignment rule is to assign customers to queues on the basis of independent trials, i.e., $\{Y_n\}$ is an independent trials process. Then letting

$$\Pr\{Y_n = j \mid Y_{n-1} = i\} = \Pr\{Y_n = j\} = q_j ,$$

and writing $p_j = 1-q_j$ we get

$$F_{rr}^{(n)} = q_r p_r^{n-1} ,$$

so that,

$$f_{rr}(z) = q_r z (1-zp_r)^{-1} .$$

Thus, by formula II.3.1,

$$B^*(s) = q_r A^*(s) (1 - p_r A^*(s))^{-1} .$$

If the arrival process is a recurrent process, i.e., $\{Z(t)\}$ has only one state, then the semi-Markov process $\{\zeta(t)\}$ is equivalent to a recurrent process. Then, the Laplace-Stieltjes transform of the distribution of the time between arrivals to the r-th queue is $B^*(s)$ for which an explicit formula is given above. This result, even if known, has never been published.

A further special case of this, however, is very well known; and that is where the arrivals form a Poisson process. In that case $A^*(s) = \alpha/(\alpha+s)$, so that

$$B^*(s) = q_r \frac{\alpha}{\alpha+s} \left(1 - p_r \frac{\alpha}{\alpha+s}\right)^{-1} = \frac{q_r\alpha}{\alpha+s} \cdot \frac{\alpha+s}{\alpha+s-p_r\alpha} = \frac{q_r\alpha}{q_r\alpha+s} ,$$

i.e., the arrivals to the r-th queue is Poisson with parameter $q_r\alpha$.

7. CONCLUSION

In this chapter we have solved the problem of identifying the process that resulted from choosing, according to a Markovian rule, some of the points of a known semi-Markov process. Furthermore, we have discussed some of the more important properties of the resulting process as well as deriving formulas for the mean occupancy times, etc.

Semi-Markov processes arise naturally as the balking streams from negative exponential servers subject to recurrent inputs as well as semi-Markovian inputs as will be shown in Chapter IV. If one is interested in decomposing such a stream into R sub-streams, this

chapter provides him with an analytical tool. One can, by means of the tools provided here, examine each one of the R sub-streams, so that the sub-systems with these streams as inputs may be studied in isolation from each other.

In this chapter the process $\{Y_n\}$, (the decision process, or the assignment rule,) was taken to be completely independent of $\{Z(t)\}$, the arrival process. In the next chapter we will discuss the same problem with the process $\{Y_n\}$ being dependent on the arrival process.

CHAPTER III

DECOMPOSITION OF A SEMI-MARKOVIAN STREAM UNDER A STATE DEPENDENT DECISION RULE

1. INTRODUCTION

Consider a semi-Markov process $\{Z(t),\ t \geq 0\}$. Let the times the process makes its transitions be $T_0, T_1, T_2, \ldots$. Assume that at each one of the points $\{T_n\}$, one of R possible events, $E_1, E_2, \ldots, E_R$, can happen. The probability of E_j happening at the instant T_n is conditional on the state of the process $\{Z(t)\}$ at that instant, but what event occurs at the instant T_n has no effects on the process $\{Z(t)\}$.

For a fixed r, let the times the event E_r happens be $\tau_1, \tau_2, \tau_3, \ldots$, and set $\tau_0 = 0$. In this chapter we are interested in the process $\{\zeta(t)\}$, where $\zeta(t) = Z(\tau_k)$ for $\tau_k \leq t < \tau_{k+1}$, $k \geq 1$. We will show that $\{\zeta(t)\}$ is a semi-Markov process, will identify it, and examine its properties.

The application we have in mind is as follows. Assume the input to a system of R queues is a semi-Markov process. Upon arrival, a customer is assigned to one of these R queues depending, in part, on the customer's type. Then $\{Z(t)\}$ is the arrival process, states of $\{Z(t)\}$ are the customer types, event E_r is the event that a customer is assigned to the r-th queue; $T_0, T_1, T_2, \ldots$ are the times of arrivals into the system, and $\tau_1, \tau_2, \ldots$ are the times of arrivals to

the r-th queue. By means of this chapter, we will be able to find the arrival processes to each one of the R queues, so that, each queue can be studied in isolation.

2. DEFINITIONS AND NOTATION

Let $\{Z(t),\ t \geq 0\}$ be a semi-Markov process with M states, the states $1,2,\ldots,M_1$ forming one ergodic class, and the remaining ones, $M_1+1,\ldots,M$, being transient states. Let the times $\{Z(t)\}$ makes its transitions be $T_0,T_1,T_2,\ldots$ where $0 = T_0 < T_1 < T_2 < \ldots$. Let $X_n = T_n - T_{n-1}$, $n \geq 1$, and set $X_0 = 0$. Let $Z_n = Z(T_n+0)$.

Let $a_0 = [a_0(1)\ a_0(2)\ \ldots\ a_0(M)]$ be the vector of initial probabilities, and let $A(x)$ be the transition matrix whose (i,j) entry is

$$A_{ij}(x) = \Pr\{Z_n = j,\ X_n \leq x \mid Z_{n-1} = i\}.$$

Further, define $A^*(s)$ as the Laplace-Stieltjes transform of $A(x)$, i.e., $A^*(s) = L\{A(x)\} = \int_0^\infty e^{-sx}\, dA(x)$; let $H_i(x) = \sum_{j=1}^{M} A_{ij}(x)$, and $\mu_i = \int_0^\infty x\, dH_i(x)$.

Now assume that at each point T_n one of R possible events, $E_1,E_2,\ldots,E_R$, can happen. Let Y_n be an integer valued random variable which assumes the value j if the event that happens at time T_n is E_j. We assume Y_n depends only on Z_n, i.e.,

$$\Pr\{Y_n=j \mid Y_0,Y_1,\ldots,Y_{n-1};\ Z_0,Z_1,\ldots,Z_n;\ X_1,X_2,\ldots,X_n\} = \Pr\{Y_n=j \mid Z_n\}.$$

We assume that what value Y_n takes has no effect on the future of the process $\{Z(t)\}$ or $\{Y_n\}$. Now choose an integer r, $1 \leq r \leq R$, and

fix it once and for all. Let $q_j = \Pr\{Y_n = r \mid Z_n = j\}$; and define q to be the column vector whose j-th entry is q_j . Clearly, q_j $(j=1,2,\ldots,M)$ satisfy $0 \le q_j \le 1$; further we assume that at least one of $q_1, q_2, \ldots, q_{M_1}$ is positive. This assumption is made in order to make the problem meaningful, (cf. p. 69.)

Let the times the event E_r happens be $\tau_1, \tau_2, \ldots$, and set $\tau_o = 0$. We define a new stochastic process $\{\zeta(t),\ t \ge 0\}$, such that,

$$\zeta(t) = \begin{cases} \zeta_o & \text{if } 0 \le t < \tau_1 \\ Z(\tau_k) & \text{if } \tau_k \le t < \tau_{k+1}\ , \quad k \ge 1\ , \end{cases}$$

where

$$\Pr\{\zeta_o = i\} = b_o(i)\ , \qquad i=1,2,\ldots,M\ , \quad \text{are given}\ .$$

We let $\zeta_n = \zeta(\tau_n+0)$, $n \ge 0$, and $\theta_n = \tau_n - \tau_{n-1}$, $n \ge 1$, and set $\theta_o = 0$.

The object of this chapter is to show that the process $\{\zeta(t)\}$ is a generalized semi-Markov process, to identify it, and examine some of its properties. Accordingly we let

$$\begin{aligned}
\tilde{B}_{ij}(x) &= \Pr\{\zeta_1 = j\ ,\ \theta_1 \le x \mid \zeta_o = i\}\ , \\
B_{ij}(x) &= \Pr\{\zeta_{n+1} = j\ ,\ \theta_{n+1} \le x \mid \zeta_n = i\}\ , \quad n \ge 1\ ; \\
\tilde{B}(x) &= [\tilde{B}_{ij}(x)]\ , \qquad B(x) = [B_{ij}(x)]\ ; \\
\tilde{B}^*(s) &= \int_o^\infty e^{-sx}\, d\tilde{B}(x), \quad B^*(s) = \int_o^\infty e^{-sx}\, dB(x)\ ; \\
\tilde{B} = \tilde{B}(\infty) &= \tilde{B}^*(0)\ , \qquad B = B(\infty) = B^*(0)\ .
\end{aligned}$$

Further, let $\tilde{G}_i(x) = \Sigma_{j=1}^M \tilde{B}_{ij}(x)$, $G_i(x) = \Sigma_{j=1}^M B_{ij}(x)$; $\tilde{\eta}_i = \int_o^\infty x\, d\tilde{G}_i(x)$, and $\eta_i = \int_o^\infty x\, dG_i(x)$.

3. THE PROCESS $\{\zeta(t)\}$

In this section we will characterize and define the process $\{\zeta(t)\}$.

Theorem III.3.1. The process $\{\zeta(t)\}$ is a generalized semi-Markov process defined by the triplex $(b_o, \tilde{B}(x), B(x))$.

Proof. $\Pr\{\zeta_o=i\} = b_o(i)$ follows by the definition of $b_o(i)$. $\Pr\{\zeta_1=j,\ \theta_1 \le x \mid \zeta_o = i\} = B_{ij}(x)$ again by definition. Next, we need to prove that, for $k \ge 1$,

$$\Pr\{\zeta_{k+1}=j,\ \theta_{k+1} \le x \mid \zeta_o,\zeta_1,\ldots,\zeta_k\ ;\ \theta_1,\theta_2,\ldots,\theta_k\} = \Pr\{\zeta_{k+1}=j\ ,\ \theta_{k+1} \le x \mid \zeta_k\} \quad .$$

Now,

$$\Pr\{\zeta_{k+1}=j,\ \theta_{k+1} \le x \mid \zeta_o,\ldots,\zeta_k\ ;\ \theta_1,\theta_2,\ldots,\theta_k\} =$$

$$= \sum_{m=1}^{\infty} \Pr\{Z_{n+m}=j\ ;\ X_{n+1}+\ldots+X_{n+m} \le x\ ;\ Y_{n+1}\neq r\ ,\ Y_{n+2}\neq r,\ldots,Y_{n+m-1}\neq r\ ,$$
$$Y_{n+m}=r \mid \zeta_o,\ldots,\zeta_k\ ;\ \theta_1,\theta_2,\ldots,\theta_k\ ;\ Z_n = \zeta_k\ ;\ Y_n = r\}$$

$$= \sum_{m=1}^{\infty} \Pr\{Z_{n+m}=j\ ;\ X_{n+1}+\ldots+X_{n+m} \le x\ ;\ Y_{n+1}\neq r,\ldots,Y_{n+m-1} \neq r\ ,$$
$$Y_{n+m}=r \mid Z_n = \zeta_k\ ,\ Y_n = r\} \quad \text{(since } \{Z(t)\} \text{ is a S.-M.P.)}$$

$$= \Pr\{\zeta_{k+1} = j\ ,\ \theta_{k+1} \le x \mid \zeta_k\}$$

$$= B_{\zeta_k j}(x) \ . \qquad \text{The proof is complete.}$$

Corollary III.3.1.A. The process $\{\zeta_k\ ;\ k = 0,1,2,\ldots\}$ is a Markov chain defined by $(b_o, \tilde{B}, B)$.

Proof. $\Pr\{\zeta_0=j\} = b_0(j)$ by definition.

$\Pr\{\zeta_1=j \mid \zeta_0\} = \Pr\{\zeta_1=j,\ \theta_1 \le \infty \mid \zeta_0\} = \tilde{B}_{\zeta_0 j}(\infty) = \tilde{B}_{\zeta_0 j}$. Next, for $k \ge 1$,

$\Pr\{\zeta_{k+1}=j \mid \zeta_0,\ldots,\zeta_k\} = \Pr\{\zeta_{k+1}=j\ ,\ \theta_{k+1} \le \infty \mid \zeta_k\} = B_{\zeta_k j}$.

Corollary III.3.1.B. The process $\{\zeta_k,\theta_k\ ;\ k \ge 0\}$ is a Markov process.

Proof follows from the general theory, (see for example the proof of Corollary II.3.1.B.)

This theorem and its corollaries completely characterize the process $\{\zeta(t)\}$. Next we will derive the matrices $\tilde{B}(x)$ and $B(x)$ from already given things such as $A(x)$, q , etc.

First, we define

$$Q_{ij}(x) = \Pr\{Z_{n+1}=j\ ,\ X_{n+1} \le x\ ,\ Y_n \neq r \mid Z_n = i\} .$$

Then, since Y_n depends only on Z_n ,

$$Q_{ij}(x) = \Pr\{Z_{n+1}=j\ ,\ X_{n+1} \le x \mid Z_n = i\}\ \Pr\{Y_n \neq r \mid Z_n = i\}$$

$$= A_{ij}(x)\ (1-q_i) .$$

Let D be the diagonal matrix of degree M whose i-th diagonal entry is q_i ; and let $Q(x) = [Q_{ij}(x)]$, $Q^*(s) = L\{Q(x)\} = \int_0^\infty e^{-sx}\, dQ(x)$, $\mathrm{Re}\{s\} \ge 0$. Then, we have

$$Q(x) = (I-D)\ A(x)$$

$$Q^*(s) = (I-D)\ A^*(s) .$$

Next define m-step probabilities

$$Q_{ij}^{(m)}(x) = \begin{cases} \delta_{ij}\, U(x) & \text{for } m = 0 \\ \Pr\{Z_{n+m}=j,\ X_{n+1}+\ldots+X_{n+m} \le x\ ;\ Y_n \neq r\ ,\ldots,Y_{n+m-1} \neq r \mid Z_n=i\} & \text{for } m \ge 1\ , \end{cases}$$

where δ_{ij} is the Dirac's delta function and $U(x)$ is the Heaviside's unit step function. Clearly, for $m = 1$, $Q_{ij}^{(1)}(x) = Q_{ij}(x)$; and for $m \ge 2$ we get

$$\begin{aligned} Q_{ij}^{(m)}(x) &= \sum_{h=1}^{M} \int_{(o,x)} \Pr\{Z_{n+m}=j,\ X_{n+2}+\ldots+X_{n+m} \le x-y,\ y-dy < X_{n+1} \le y\ ; \\ &\qquad Y_n \neq r,\ Y_{n+1} \neq r,\ldots,Y_{n+m-1} \neq r,\ Z_{n+1}=h \mid Z_n=i\} \\ &= \sum_{h=1}^{M} \int_{(o,x)} \Pr\{Z_{n+m}=j,\ X_{n+2}+\ldots+X_{n+m} \le x-y;\ Y_{n+1} \neq r,\ Y_{n+2} \neq r,\ldots, \\ &\qquad Y_{n+m-1} \neq r \mid Z_{n+1}=h\}\ d_y \Pr\{Z_{n+1}=h,\ X_{n+1} \le y,\ Y_n \neq r \mid Z_n=i\} \\ &= \sum_{h=1}^{M} \int_{o}^{x} Q_{hj}^{(m-1)}(x-y)\ dQ_{ih}(y)\ . \end{aligned}$$

In matrix notation using Laplace-Stieltjes transforms we have, with $Q^{(m)}(x) = [Q_{ij}^{(m)}(x)]$,

$$L\{Q^{(m)}(x)\} = \begin{cases} I & \text{for } m=o \\ Q^*(s) & \text{for } m=1 \\ L\{Q^{(m-1)}(x)\}\ Q^*(s) & \text{for } m \ge 2\ . \end{cases}$$

Then, by a simple induction on m , we obtain

$$L\{Q^{(m)}(x)\} = Q^*(s)^m \quad , \qquad m \ge 0\ .$$

Further, we let $Q = Q(\infty) = Q^*(0) = (I-D)A^*(0) = (I-D)A$.

One expects that $Q_{ij}^{(m)}(x)$ should go to zero uniformly for all x as m goes to infinity. This is true and we have

Theorem III.3.2. The m-step probabilities $Q_{ij}^{(m)}(x)$ $(i,j=1,2,\ldots,M)$ converge to zero uniformly for all $x \geq 0$ as m approaches infinity.

Proof. Since $0 \leq Q_{ij}^{(m)}(x) \leq Q_{ij}^{(m)}(\infty) = Q_{ij}^{(m)}$, by Weierstrass' uniform convergence theorem, it is sufficient to show that $Q_{ij}^{(m)} \to 0$ as $m \to \infty$.

By our assumptions, at least one of the quantities $q_1, q_2, \ldots, q_{M_1}$, say q_N, is positive. The states $1,2,\ldots,M_1$ form an ergodic set in $\{Z(t)\}$, and hence in $\{Z_n\}$, (by Theorem I.3.2.) Thus, $\lim_{m \to \infty} \Pr\{Z_{n+m} = N \mid Z_n = j\}$ is positive. Therefore,

$$\lim_{m \to \infty} \Pr\{Y_{n+m}=r \mid Z_n=i\} = \lim_{m \to \infty} \Sigma_{j=1}^{M} \Pr\{Y_{n+m}=r \mid Z_{n+m}=j\} \Pr\{Z_{n+m}=j \mid Z_n=i\} \geq \lim_{m \to \infty} \Pr\{Y_{n+m}=r \mid Z_{n+m}=N\} \Pr\{Z_{n+m}=N \mid Z_n=i\} =$$

$$q_N \lim_{m \to \infty} \Pr\{Z_{n+m}=N \mid Z_n = i\} > 0 .$$

Consequently, the probability that $Y_{n+h}=r$ for at least one $h \geq 0$, given that $Z_n = i$, is equal to one. Hence, the complementary event that $Y_{n+h} \neq r$ for all $h \geq 0$, given $Z_n = i$, has probability zero, i.e.,

$$\Pr\{Y_{n+h} \neq r, \ h \geq 0 \mid Z_n = i\} = 0 .$$

Now,

$$0 \leq Q_{ij}^{(m)} \leq \Sigma_j Q_{ij}^{(m)} = \Sigma_j \Pr\{Z_{n+m}=j ; Y_{n+h} \neq r, \ h = 0,1,\ldots,m-1 \mid Z_n=i\} = \Pr\{Y_{n+h} \neq r, \ h = 0,1,\ldots,m-1 \mid Z_n=i\} .$$

Then, taking limits as $m \to \infty$,

$$0 \leq \lim_{m \to \infty} Q_{ij}^{(m)} \leq \Pr\{Y_{n+h} \neq r \ , \ h = 0,1,\ldots \mid Z_n = i\} = 0 \ .$$

Hence, $\lim_{m \to \infty} Q_{ij}^{(m)} = 0$ for all i and j . Proof is complete.

We restate this theorem in matrix language as

Theorem III.3.2'. The sequence of matrix valued functions $Q^{(m)}(x)$ converges to zero uniformly for all $x \geq 0$. Or, equivalently, the sequence $\{Q^*(s)^m\}$ converges to zero uniformly for all s with $\mathrm{Re}\{s\} \geq 0$.

We now return to the derivation of $B_{ij}(x)$ and $\tilde{B}_{ij}(x)$. From the proof of Theorem III.3.1 we already have

$$B_{ij}(x) = \sum_{m=1}^{\infty} \Pr\{Z_{n+m}=j \ ; \ X_{n+1}+\ldots+X_{n+m} \leq x \ ; \ Y_{n+1} \neq r, \ldots, Y_{n+m-1} \neq r \ , \ Y_{n+m}=r \mid Z_n = i \ , \ Y_n = r\} \ .$$

Then,

$$B_{ij}(x) = \sum_{m=1}^{\infty} \sum_{k=1}^{M} \int_0^x \Pr\{Z_m=j; \ X_2+\ldots+X_m \leq x-y; \ Y_1 \neq r, \ldots, Y_{m-1} \neq r \mid Z_1=k\} \Pr\{Y_m=r \mid Z_m=k\} \, d_y \Pr\{Z_1=k, \ X_1 \leq y \mid Z_o = i\}$$

$$= \sum_{m=1}^{\infty} \sum_{k=1}^{M} \int_0^x q_j \, Q_{kj}^{(m-1)}(x-y) \, dA_{ik}(y) \ .$$

Or in matrix notation, using Laplace-Stieltjes transforms,

$$B^*(s) = \sum_{m=1}^{\infty} A^*(s) \, Q^*(s)^{m-1} \, D \ .$$

Lemma III.3.3. The matrix $I-Q^*(s)$ in non-singular, and the matrix power series $\Sigma_{m=0}^{\infty} Q^*(s)^m$ converges to $(I-Q^*(s))^{-1}$ for all s with $\mathrm{Re}\{s\} \geq 0$.

Proof. By Theorem III.3.2', $Q^*(s)^m \to 0$ as $m \to \infty$ for all s with non-negative real parts. Then, the lemma follows from Theorem I.2.5.

Using this lemma, we obtain as the Laplace-Stieltjes transform of the transition matrix $B(x)$,

$$B^*(s) = A^*(s)(I-Q^*(s))^{-1}D \ . \qquad \text{(III.3.1)}$$

We next derive $\tilde{B}_{ij}(x)$. We have ,

$$\begin{aligned}
\tilde{B}_{ij}(x) &= \Pr\{\zeta_1 = j \ , \ \Theta_1 \le x \mid \zeta_o = i\} \\
&= \Pr\{Z_o = j \ , \ Y_o = r\} + \sum_{m=1}^{\infty} \sum_{h=1}^{M} \Pr\{Z_m = j \ ; \ X_1 + \ldots + X_m \le x \ ; \\
&\qquad Y_o \neq r \ , \ Y_1 \neq r, \ \ldots, \ Y_{m-1} \neq r \ , \ Y_m = r \ , \ Z_o = h\} \\
&= \Pr\{Y_o = r \mid Z_o = j\} \Pr\{Z_o = j\} + \sum_{m=1}^{\infty} \sum_{h=1}^{M} \Pr\{Y_m = r \mid Z_m = j\} \cdot \\
&\qquad \Pr\{Z_m = j \ ; \ X_1 + \ldots + X_m \le x \ ; \ Y_o \neq r \ , \ldots, \ Y_{m-1} \neq r \mid Z_o = h\} \Pr\{Z_o = h\} \\
&= a_o(j)\, q_j + \sum_{m=1}^{\infty} \sum_{h=1}^{M} a_o(h)\, Q_{hj}^{(m)}(x)\, q_j \ .
\end{aligned}$$

In matrix notation, taking Laplace-Stieltjes transforms, we have $\tilde{B}^*(s) = E \oplus (a_o D + \Sigma_{m=1}^{\infty} a_o Q^*(s)^m D)$ where "$\oplus$" stands for the Kronecker product. Noting that $Q^*(s)^o = I$, we get $\tilde{B}^*(s) = E \oplus (a_o(\sum_{m=o}^{\infty} Q^*(s)^m)D)$; and using the Lemma III.3.3, we finally obtain

$$\tilde{B}^*(s) = E \oplus (a_o(I-Q^*(s))^{-1}D) \ . \qquad \text{(III.3.2)}$$

This completes the problem of identifying the process $\{\zeta(t)\}$ since of the defining triplex $(b_o, \tilde{B}(x), B(x))$, we have derived $\tilde{B}(x)$ and $B(x)$, and the initial distribution vector b_o was already given.

4. THE PROCESS $\{\zeta_n\}$

By the Corollary III.3.1.A the process $\{\zeta_n ; n = 0,1,2,\ldots\}$ is a Markov chain. In this section we will classify its states and give some limiting results.

First, consider the result of setting $q_j = 0$ for some j. Then, the j-th column of D will be completely zero. This would cause the j-th columns of $\tilde{B}$ and B to be completely zero, (see the formulas III.3.1 and III.3.2.) Hence, a state j for which $q_j = 0$ cannot be reached from any state of the process $\{\zeta_n\}$. We shall call all such states empty states, since $\{\zeta_n\}$ can never visit any such state. An empty state is transient in a trivial manner, namely, $\Pr\{\zeta_n = j\} = 0$ for all $n \geq 1$ if j is an empty state.

Lemma III.4.1. A state j of the process $\{\zeta_n\}$ is transient if it is transient in $\{Z(t)\}$, or if $q_j = 0$.

Proof. If the state j of $\{Z(t)\}$ is transient, then that state is transient in $\{Z_n\}$ also. For the process $\{\zeta_n\}$ to visit a state i, first Z_n must be i, and then Y_n should be r. As n approaches infinity $\Pr\{Z_n = i\}$ approaches zero; hence by the preceding argument, $\Pr\{\zeta_n = i\}$ approaches zero also. Thus, any state j which is transient in $\{Z(t)\}$ is transient in $\{\zeta_n\}$. On the other hand, if $q_j = 0$, by the paragraph preceding this lemma, the state j is transient in a trivial manner.

Actually, these are the only transient states in $\{\zeta_n\}$.

Theorem III.4.2. State j of the process $\{\zeta_n\}$ is ergodic if it is ergodic in $\{Z(t)\}$ and $q_j > 0$. State j is transient if

either it is transient in $\{Z(t)\}$ or $q_j = 0$. Furthermore, there exists only one ergodic class in $\{\zeta_n\}$.

Proof. The second statement is the preceding lemma restated. Now, assume a state k can be reached from a state h in the process $\{Z(t)\}$, (and therefore in $\{Z_n\}$.) Then, there exists an N such that $A_{hk}^{(N)} = \Pr\{Z_N = k \mid Z_o = h\} > 0$. Now, assume $\zeta_m = h$. Then, there exists an n such that $Z_n = \zeta_m = h$. Then, there exists an $\ell \leq N$ such that $\Pr\{\zeta_{m+\ell}=k \mid \zeta_m=h\} = \Pr\{Y_N=r \mid Z_N=k\} \Pr\{Z_N=k \mid Z_o=h\} = q_k A_{hk}^{(N)}$ which is positive if q_k is positive. Hence, a state k can be reached from a state h in the process $\{\zeta_n\}$ if $q_k > 0$ and state k can be reached from h in $\{Z_n\}$. Now consider the set of all states j such that j is ergodic in $\{Z_n\}$ and $q_j > 0$. Then, all these states can be reached from one another. Hence they are all in the same class. These states must all be ergodic since the complementing set of states are all transient by Lemma III.4.1, and not all of the states of a finite Markov chain can be transient. Proof is complete.

We can now give the rational behind an assumption made earlier. We have assumed that at least one of $q_1, q_2, \ldots, q_{M_1}$ is positive. If these quantities were allowed to vanish altogether, then by the preceding theorem there would be only transient states in $\{\zeta_n\}$, i.e., $\{\zeta_n\}$ cannot be a Markov chain.

In view of the preceding theorem, the limiting probabilities $\lim_{n\to\infty} \Pr\{\zeta_n=j\}$ exist. Let b be the row vector whose j-th element is $b(j) = \lim_{n\to\infty} \Pr\{\zeta_n=j\}$. Similarly, let a be the row vector whose j-th entry is $a(j) = \lim_{n\to\infty} \Pr\{Z_n=j\}$.

Theorem III.4.3. The limiting probabilities $b(j)$, $j = 1,2,\ldots,M$, exist and are indepedent of the initial distribution given by b_o. Furthermore, $b = \frac{1}{aq}\, aD$ where a is the limiting distribution vector for the chain $\{Z_n\}$.

Proof. Existence of the limiting distribution is guaranteed by the preceding theorem. Furthermore, since there is only one ergodic set in $\{\zeta_n\}$, these limiting probabilities do not depend on the initial conditions. Hence, the unique eigenvector b corresponding to the simple eigenvalue one, ($bB=b$,) with $bE = 1$ is the vector of steady state distribution.

Now, consider the steady state distribution vector a of the process $\{Z_n\}$; $aA=a$, $aE=1$. Note that,

$$a(I-Q) = a(I-(I-D)A) = a-a(A-DA) = a-aA+aDA = aDA .$$

Then,

$$(aD)B = (aD)(A(I-Q)^{-1}D) = (aDA)(I-Q)^{-1}D = a(I-Q)(I-Q)^{-1}D = aD .$$

Hence, aD is an eigenvector of B corresponding to the simple eigenvalue one. Now dividing each term of aD by the sum of the terms, $aDE = aq$, we obtain $\frac{1}{aq}\, aD$ which is a probability vector besides being an eigenvector for the eigenvalue one; (note that, $aq = \Sigma_{j=1}^{M} a(j)\, q_j = \Sigma_{j=1}^{M_1} a(j)q_j > 0$ since $a(j) > 0$ for $j = 1,2,\ldots,M_1$, and at least one of $q_1, q_2, \ldots, q_{M_1}$ is positive.) From the uniqueness of b, then, $b = \frac{1}{aq}\, aD$. Proof is complete.

We restate this theorem below in matrix form by first letting $A^{\infty} = \lim_{n\to\infty} A^n$, and $B^{\infty} = \lim_{n\to\infty} B^n$.

Theorem III.4.3'. $\overset{\infty}{B} = \lim_{n\to\infty} B^n$ exists, and $\overset{\infty}{B} = \frac{1}{aq} \overset{\infty}{A} D$.

5. THE CLASSIFICATION OF THE STATES OF $\{\zeta(t)\}$

In this section we will give a complete classification for the states of the process $\{\zeta(t)\}$. First, however, we need to examine the occupancy times of each state.

The mean occupancy time of state i by the process $\{\zeta(t)\}$ was defined to be $\tilde{\eta}_i$ for the first step and η_i for the succeeding steps, (cf. page 61 .) The occupancy time of state i by $\{Z(t)\}$ was μ_i . Now let μ , η , $\tilde{\eta}$ be the column vectors whose i-th entries are μ_i , η_i , $\tilde{\eta}_i$ respectively. Then,

$$\mu = - \frac{dA^*(s)}{ds}\Big|_{s=o} E , \quad \eta = - \frac{dB^*(s)}{ds}\Big|_{s=o} E , \quad \text{and} \quad \tilde{\eta} = - \frac{d\tilde{B}^*(s)}{ds}\Big|_{s=o} E .$$

Now,

$$\begin{aligned}\frac{d}{ds} B^*(s) &= \frac{d}{ds} (A^*(s) (I - Q^*(s))^{-1} D) \\ &= (\frac{d}{ds} A^*(s))(I - Q^*(s))^{-1} D + A^*(s)(\frac{d}{ds}(I-Q^*(s))^{-1})D \\ &= \frac{dA^*(s)}{ds} (I-Q^*(s))^{-1} D + A^*(s)(I-Q^*(s))^{-1} \frac{dQ^*(s)}{ds} (I-Q^*(s))^{-1} D\end{aligned}$$

by formula (I.2.2). Then,

$$- \frac{d}{ds} B^*(s)\Big|_{s=o} = \left(- \frac{dA^*(s)}{ds}\right)_{s=o} (I-Q)^{-1} D + A(I-Q)^{-1} \left(- \frac{dQ^*(s)}{ds}\right)_{s=o} (I-Q)^{-1} D.$$

Notice that,

$$\frac{dQ^*(s)}{ds} = \frac{d}{ds} (I-D) A^*(s) = (I-D) \frac{dA^*(s)}{ds} ,$$

so that,

$$-\frac{d}{ds}B^*(s)\Big|_{s=o}E = (I+A(I-Q)^{-1}(I-D))\left(-\frac{dA^*(s)}{ds}\right)_{s=o}(I-Q)^{-1}DE .$$

Noting that,

$$DE = DE+E-E = E-(I-D)E = E-(I-D)AE = E-QE = (I-Q)E ,$$

and putting this in the preceding equation, we obtain

$$\eta = -\frac{d}{ds}B^*(s)\Big|_{s=o}E = (I+A(I-Q)^{-1}(I-D))\left(-\frac{d}{ds}A^*(s)\right)_{s=o}E$$

$$= (I+A(I-Q)^{-1} - A(I-Q)^{-1}D)\mu$$

$$\eta = (I-B+A(I-Q)^{-1})\mu . \qquad \text{(III.5.1)}$$

Given μ , η can be computed from (III.5.1) uniquely as a function of μ . It should perhaps be noted that the mean occupancy time of a state i may be positive even though the state i is an empty state. This is possible since η_i is the mean time spent in state i given that the process is in state i already whether it is possible to be in state i at all or not.

Next we derive the mean occupancy times in the first step. Since $\tilde{B}_{ij}(x)$, $(i,j = 1,2,\ldots,M,)$ do not depend on i, $\tilde{G}_1(x) = \tilde{G}_2(x) = \ldots = \tilde{G}_M(x)$, so that $\tilde{\eta}_1 = \tilde{\eta}_2 = \ldots = \tilde{\eta}_M$; and hence,

$$\tilde{\eta}_1 = -\frac{d}{ds}(a\,(I-Q^*(s))^{-1}D)\Big|_{s=o}E \qquad \text{(see formula III.3.2)}$$

$$= a_o\left((I-Q^*(s))^{-1}\,(-\frac{dQ^*(s)}{ds})\,(1-Q^*(s))^{-1}\right)_{s=o}DE$$

$$= a_o\,(I-Q)^{-1}\left(-\frac{dQ^*(s)}{ds}\right)_{s=o}(I-Q)^{-1}\,DE .$$

Now, noting that $DE = (I-Q)E$, and $\frac{dQ^*(s)}{ds} = (I-D)\frac{dA^*(s)}{ds}$

$$\tilde{\eta}_1 = a_o (I-Q)^{-1} (I-D) \left(-\frac{dA^*(s)}{ds}\right)_{s=o} E$$

$$\tilde{\eta}_1 = a_o (I-Q)^{-1} (I-D) \mu \quad . \tag{III.5.2}$$

Given μ , $\tilde{\eta}_1 = \tilde{\eta}_2 = \ldots = \tilde{\eta}_M$ can all be computed from this formula uniquely.

Lemma III.5.1. The mean occupancy time of state i is finite for all $i = 1,2,\ldots,M$ in the process $\{\zeta(t)\}$.

Proof. Since $\{Z(t)\}$ had its first M_1 states ergodic, and the remaining $M-M_1$ states are transient, μ_i is finite $(i = 1,2,\ldots,M_1)$ by Theorem I.3.2. Then, clearly, $\tilde{\eta}_1 = \ldots = \tilde{\eta}_M$ as computed from (III.5.2) are finite. Also, η as obtained from (III.5.1) is finite since μ is finite. Hence, the occupancy time of state i has finite expectation both in the first step and in the succeeding ones.

Theorem III.5.2. A state j of the process $\{\zeta(t)\}$ is transient if either j is a transient state in $\{Z(t)\}$ or $q_j = 0$. A state j of $\{\zeta(t)\}$ is ergodic if it is ergodic in $\{Z(t)\}$ and $q_j > 0$. Furthermore, all ergodic states are in the same class.

Proof. The first statement follows from Theorems II.3.2 and III.4.2. By Theorem II.3.2, a state is ergodic in $\{\zeta(t)\}$ if it is ergodic in $\{\zeta_n\}$ and η_i , $\tilde{\eta}_i$ are finite for all i in the same class with that state. Hence, the second statement follows from Theorems II.3.2, III.4.2, and Lemma III.5.1. The fact that there exists only one ergodic class follows from Theorem III.4.2 in view of Theorem II.3.2.

Before ending this section we should point out that in the processes $\{\zeta(t)\}$ and $\{\zeta_n\}$ the empty states can be completely discarded. If in the matrices $B^*(s)$, $B(s)$, and B all columns which are completely zero and all corresponding rows are thrown out, the remaining matrices will still represent the processes fully. This is clearly desirable whenever these processes are the inputs in other problems.

6. SPECIAL CASES AND APPLICATIONS

In multi-server queueing systems, it may happen that certain servers specialize in the service of certain types of customers. This is advantageous if specialization enables the service times to be reduced appreciably. In such cases, there are separate waiting lines forming before different types of servers. The treatment in this chapter enables us to obtain the process of arrivals to each one of the queues. Once the arrival process for a queue is obtained, that queue can be studied in isolation from the rest of the queues.

Semi-Markov processes arise naturally as the streams of balking customers from negative exponential servers. There, the states of the semi-Markov process reflect the number of people in the queues from which customers balked. If such a balking stream is the input to a system of queues in parallel, it may be desirable to assign a customer to a certain queue on the basis of the number of people in the previous queues from which he balked, (or was not accepted.) This chapter, by letting the probability of assigning a customer to a certain

queue be dependent on the type of the customer, provides a valuable model for such a problem.

In the case the number of states, M, is one, the process $\{X_n\}$ becomes a renewal process; and the dependence of $\{Y_n\}$ on $\{Z_n\}$ becomes meaningless. Then, the results of this chapter coincides with the results of Chapter II for the case $M=1$ and the decision process is an independent trials process.

A special case of interest is where $0 < q_1 = q_2 = \ldots = q_M \leq 1$, i.e., Y_n does not depend on Z_n. Then,

$$q = q_1 E , \quad D = q_1 I ; \quad Q^*(s) = (I-D)A^*(s) = (1-q_1)A^*(s) .$$

Hence, letting $1-q_1 = p_1$

$$\begin{aligned} B^*(s) &= A^*(s)\,(I-Q^*(s))^{-1} D \\ &= A^*(s)\,(I-p_1 A^*(s))^{-1}\, q_1 I \\ &= q_1 A^*(s)(I-p_1 A^*(s))^{-1} , \end{aligned}$$

which is the same as the formula obtained in Example II.6.2.

7. CONCLUSION

We have solved the problem of identifying and deriving some of the more important properties of a process which is constructed from a given semi-Markov process by sampling its points by a rule that depended on the nature of points.

Both the present chapter and the preceding one treated the problem of decomposing a stream of arrivals into R streams, each one

of which may become the arrival process for some subsystem. So far, the rules of decomposition, as reflected in the process $\{Y_n\}$, have been chosen to be independent of the states of the queues involved.

In the next chapters, we will take up the case where the process $\{Y_n\}$ depends on the state of the queueing process.

PART II

BALKING IN THE QUEUEING SYSTEM SM/M/1

CHAPTER IV

QUEUEING SYSTEM SM/M/1 WITH BALKING

1. INTRODUCTION

Consider a single server queue with negative exponentially distributed service times with mean $1/\beta$. Let the capacity of the waiting room be $N-1$. Let the arrivals into the system form a semi-Markov process with M states, the first M_1 of which are ergodic and form one class, and the remaining states, $M_1+1,\ldots,M$, are transient.

Upon his arrival, a customer may or may not join the queue depending on the number of people already there and his own type. The object of this chapter is to analyze the queueing properties of such a system. We will examine the queue size process, waiting times, and the busy period.

2. DEFINITIONS AND NOTATION

Let the arrival process $\{Z(t)\}$ be a semi-Markov process with M states. Let the times $\{Z(t)\}$ makes its transitions, (i.e., the instants of arrivals into the system,) be $T_0, T_1, T_2, \ldots$ where $0 = T_0 < T_1 < T_2 < \ldots$. Let $X_n = T_n - T_{n-1}$ $(n \geq 1,)$ and set $X_0 = 0$. Let $Z_n = Z(T_n)$; Z_n can be referred to as the type of the n-th arrival. We assume that the states $1,2,\ldots,M_1$ form an ergodic class, and that

the remaining states, $M_1+1,\ldots,M$, are transient. Let $a_o=[a_o(1)\ldots a_o(M)]$ be the vector of initial probabilities, $a_o \geq 0$, $a_oE = 1$; and let $A(x)$ be the matrix whose (i,j) entry is

$$A_{ij}(x) = \Pr\{Z_{n+1}=j,\ X_{n+1} \leq x \mid Z_n=i\}, \quad n \geq 0 .$$

As usual, we set $A^*(s) = \int_o^\infty e^{-sx}dA(x)$, and $A = A(\infty) = A^*(0)$.

Let the state of the system at time t be denoted by $S(t)$; we write $S(t) = j$ if the number of people in the system, (number waiting plus the one being served,) at time t is j, $j = 0,1,2,\ldots,N$. We especially denote by S_n the state of the system just before the arrival of the n-th customer; i.e., $S_n = S(T_n-0)$, $n = 0,1,2,\ldots$. We take $s_o(j) = \Pr\{S_o=j\}$, $s_o(j) \geq 0$, $\Sigma_{j=o}^N s_o(j) = 1$, as given.

An arriving customer is permitted to balk, i.e., upon his arrival a customer may or may not join the system depending on his wish and the number of people already there. Associated with the n-th customer we define a random variable Y_n which takes on the values of one or zero according as the n-th customer did or did not join the queue. We assume that Y_n depends only on Z_n and S_n, i.e.,

$$\Pr\{Y_n=0 \mid Z_o,\ldots,Z_n;\ S_o,\ldots,S_n;\ Y_o,\ldots,Y_{n-1}\} = \Pr\{Y_n=0 \mid Z_n,\ S_n\} .$$

We let b be the column vector of balking probabilities, so that, the $jM+i^{th}$ element of b is $b(i,j)$ where

$$b(i,j) = \Pr\{Y_n=0 \mid Z_n=i,\ S_n=j\} .$$

In particular, $b(i,N) = 1$ for all i since everyone must balk when the system is full.

We let,

$$P(i,j;h,k,x) = \Pr\{Z_n=h,\ S_n=k,\ X_n \le x \mid Z_{n-1}=i,\ S_{n-1}=j\},$$

$$P(i,j;\ h,k) = \Pr\{Z_n=h,\ S_n=k \mid Z_{n-1}=i,\ S_{n-1}=j\},$$

$$p_n(i,j) = \Pr\{Z_n=i,\ S_n=j\}.$$

We define $P(x) = [P_{jk}(x)]$ where the submatrix $P_{jk}(x)$ has as its (i,h) entry $P(i,j;h,k,x)$. We define $P = [P_{jk}]$ similarly. We further let p_n be the row vector whose $(jM+i)$ entry is $p_n(i,j)$; and let D be the diagonal matrix whose $jM+i^{th}$ diagonal entry is $b(i,j)$.

3. THE PROCESS $\{Z_n, S_n\}$

The major process of interest in this chapter is $\{Z_n, S_n;\ n = 0,1,2,\ldots\}$. First, however, we will look into another process, $\{Z_n, S_n, X_n\}$.

Theorem IV.3.1. The process $\{Z_n, S_n, X_n\}$ is equivalent to a semi-Markov process defined by $(p_o,\ P(x))$.

Proof. We need to show that $\Pr\{Z_n=h,\ S_n=k,\ X_n \le x \mid Z_o,\ldots,Z_{n-1};\ S_o,\ldots,S_{n-1};\ X_1,\ldots,X_{n-1}\} = \Pr\{Z_n=h,\ S_n=k,\ X_n \le x \mid Z_{n-1}, S_{n-1}\}$. Now,

$$\Pr\{Z_n=h,\ S_n=k,\ X_n \le x \mid Z_o,\ldots,Z_{n-1};\ S_o,\ldots,S_{n-1};\ X_1,\ldots,X_{n-1}\} =$$

$$= \int_o^x \Pr\{S_n=k \mid Z_o,\ldots,Z_n;\ S_o,\ldots,S_{n-1};\ X_o,\ldots,X_{n-1};\ X_n=y\} \cdot$$

$$d_y \Pr\{Z_n=h,\ X_n \le y \mid Z_o,\ldots,Z_{n-1};\ S_o,\ldots,S_{n-1};\ X_o,\ldots,X_{n-1}\}.$$

S_n depends only on X_n, S_{n-1}, and Y_{n-1}. But Y_{n-1} depends on Z_{n-1} and S_{n-1} only. Hence, given Z_{n-1}, S_{n-1}, X_n, all other information lose their predictive value as far as predicting S_n is concerned, i.e., $\Pr\{S_n=k \mid Z_o,\ldots,Z_n;\ S_o,\ldots,S_{n-1};\ X_o,\ldots,X_n\} = \Pr\{S_n=k \mid Z_{n-1},\ S_{n-1},\ X_n\}$.

On the other hand, since $\{Z_n, X_n\}$ is equivalent to a semi-Markov process, and the arrival process is independent of the queueing process we get $\Pr\{Z_n=h,\ X_n \le y \mid Z_o,\ldots,Z_{n-1};\ S_o,\ldots,S_{n-1};\ X_o,\ldots,X_{n-1}\} = \Pr\{Z_n=h,\ X_n \le y \mid Z_{n-1}\}$.

Putting these into the above we obtain,

$$\Pr\{Z_n=h,\ S_n=k,\ X_n \le x \mid Z_o,\ldots,Z_{n-1};\ S_o,\ldots,S_{n-1};\ X_1,\ldots,X_{n-1}\} =$$

$$= \int_o^x \Pr\{S_n=k \mid Z_{n-1},\ S_{n-1},\ X_n=y\}\, d_y \Pr\{Z_n=h,\ X_n \le y \mid Z_{n-1}\}$$

$$= \Pr\{Z_n=h,\ S_n=k,\ X_n \le x \mid Z_{n-1},\ S_{n-1}\}$$

as was to be demonstrated.

A very important corollary of this theorem follows.

Corollary IV.3.1.A. The process $\{Z_n, S_n;\ n = 0,1,2,\ldots\}$ is a Markov chain defined by $(p_o,\ P)$.

Proof. $\Pr\{Z_n=h,\ S_n=k \mid Z_m,\ S_m;\ m < n\} =$

$$= \lim_{x\to\infty} \Pr\{Z_n=h,\ S_n=k,\ X_n \le x \mid Z_m,\ S_m;\ m < n\}$$

$$= \lim_{x\to\infty} \Pr\{Z_n=h,\ S_n=k,\ X_n \le x \mid Z_{n-1},\ S_{n-1}\}$$

$$= \Pr\{Z_n=h,\ S_n=k \mid Z_{n-1},\ S_{n-1}\}$$

$$= P\ (Z_{n-1},\ S_{n-1};\ h,\ k) \qquad \text{as needed.}$$

Corollary IV.3.1.B. The process $\{S_n\}$ is a Markov chain if and only if $M = 1$, i.e., the arrival process is a recurrent process.

Next we will give formulas for obtaining the transition matrix for the Markov chain $\{Z_n, S_n\}$. First let,

$$\gamma_k(i,j,x) = \Pr\{S_{n+1}=j \mid Y_n=k, S_n=i, X_{n+1}=x\}, \quad k = 0,1.$$

Then,

$$\gamma_k(i,j,x) = \begin{cases} \alpha_{i-j+k}(x) & \text{for } 0 < j \le i+k \le N+k \\ \sum_{m=i+k}^{\infty} \alpha_m(x) & \text{for } j = 0,\ i \le N \\ 0 & \text{otherwise,} \end{cases}$$

where

$$\alpha_n(x) = \frac{e^{-\beta x}(\beta x)^n}{n!}, \qquad n = 0,1,2\ldots .$$

Now let $\Gamma_0(x)$ and $\Gamma_1(x)$ be the square matrices of degree $N+1$ whose (i,j) entries are $\gamma_0(i,j,x)$ and $\gamma_1(i,j,x)$ respectively.

Returning to the derivation of $P(i,j;h,k)$, from the proof of Theorem IV.3.1, we have

$$P(i,j;h,k,x) = \Pr\{Z_n=h, S_n=k, X_n \le x \mid Z_{n-1}=i, S_{n-1}=j\} =$$

$$= \int_0^x \Pr\{S_n=k \mid Z_{n-1}=i, S_{n-1}=j, X_n=y\}\, d_y \Pr\{Z_n=h, X_n \le y \mid Z_{n-1}=i\}.$$

Now,

$$\Pr\{S_n=k \mid Z_{n-1}=i, S_{n-1}=j, X_n=y\} =$$

$$= \sum_{\delta=0,1} \Pr\{S_n=k, Y_{n-1}=\delta \mid Z_{n-1}=i, S_{n-1}=j, X_n = y\}$$

$$= \sum_{\delta=0,1} \Pr\{Y_{n-1}=\delta \mid Z_{n-1}=i, S_{n-1}=j\} \Pr\{S_n=k \mid Y_{n-1}=j, X_n=y\}$$

$$= b(i,j)\,\gamma_0(j,k,y) + (1-b(i,j))\,\gamma_1(j,k,y).$$

Hence we have,

$$P(i,j;h,k,x) = \int_0^x \{b(i,j)\ \gamma_0(j,k,y) + (1-b(i,j))\ \gamma_1(j,k,y)\}\ dA_{ih}(y).$$

Using matrix notation we obtain

$$P(x) = D(\int_0^x \Gamma_0(y) \oplus dA(y)) + (I-D)(\int_0^x \Gamma_1(y) \oplus dA(y)), \qquad (IV.3.1)$$

where $\oplus$ stands for the Kronecker product.

In many cases, it is easier to work with Laplace-Stieltjes transforms. Note that the elements of $\Gamma_0(y)$ and $\Gamma_1(y)$ are in terms of $\alpha_n(y)$, $n = 0,1,2,\ldots$ only. Thus, in giving the Laplace-Stieltjes transform of $P(x)$ we first need

$$\begin{aligned} L\left\{\int_0^x \alpha_n(y)\ dA(y)\right\} &= \int_0^\infty e^{-sx}\ \alpha_n(x)\ dA(x) \\ &= \int_0^\infty e^{-sx}\ e^{-\beta x}\ \frac{\beta^n}{n!}\ x^n\ dA(x) \\ &= (-1)^n\ \frac{\beta^n}{n!}\ \frac{d^n}{ds^n}\ A^*(s+\beta)\ . \end{aligned}$$

Let $\overline{D}$ be the differential operator $\frac{d}{ds}$; and let $\overline{\alpha}_n = \frac{(-\beta\overline{D})^n}{n!}$. Then,

$$L\left\{\int_0^x \alpha_n(y)\ dA(y)\right\} = \overline{\alpha}_n\ A^*(s+\beta)\ .$$

Now, define (for $k = 0,1$),

$$\overline{\gamma}_k(i,j) = \begin{cases} \overline{\alpha}_{i-j+k} & \text{for } 0 < j \le i+k \le N+k \\ \sum_{m=i+k}^{\infty} \overline{\alpha}_m & \text{for } j = 0,\ i \le N \\ 0 & \text{otherwise,} \end{cases}$$

and let $\overline{\Gamma}_k$ be the matrix whose (i,j) entry is the operator $\overline{\gamma}_k(i,j)$. With this notation, from (IV.3.1),

$$P^*(s) = D(\overline{\Gamma}_o \oplus A^*(s+\beta)) + (I-D)(\overline{\Gamma}_1 \oplus A^*(s+\beta)) . \qquad \text{(IV.3.2)}$$

Then, the transition matrix can be obtained by $P = P(\infty) = P^*(0)$. To check to see if P is a stochastic matrix, i.e., if $PE = E$, we do the following calculations. (We use the rule I.2.3 repeatedly; also remember that $\Sigma_{n=o}^{\infty} \overline{\alpha}_n = \Sigma_{n=o}^{\infty} \frac{(-\beta\overline{D})^n}{n!} = e^{-\beta\overline{D}}$, where $e^{-\beta\overline{D}}[f(x)] = f(x-\beta)$ for any function f .)

$$P^*(s)E = P^*(s)\, E_{M(N+1)} = P^*(s)(E_{N+1} \oplus E_M)$$

$$= D\,(\overline{\Gamma}_o \oplus A^*(s+\beta))(E_{N+1} \oplus E_M) + (I-D)(\overline{\Gamma}_1 \oplus A^*(s+\beta))(E_{N+1} \oplus E_M)$$

$$= D\,((\overline{\Gamma}_o\, E_{N+1}) \oplus (A^*(s+\beta)E_M)) + (I-D)((\overline{\Gamma}_1\, E_{N+1}) \oplus (A^*(s+\beta)\, E_M))$$

$$= D\,(e^{-\beta\overline{D}}\, E_{N+1} \oplus (A^*(s+\beta)E_M)) + (I-D)(e^{-\beta\overline{D}}\, E_{N+1} \oplus (A^*(s+\beta)\, E_M))$$

$$= (D + I - D)\,(E_{N+1} \oplus (A^*(s)\, E_M)) \cdot$$

$$= E_{N+1} \oplus (A^*(s)E_M) .$$

Thus,

$$PE = E_{N+1} \oplus (AE_M) = E_{N+1} \oplus E_M = E_{M(N+1)}$$

as was needed for P to be a stochastic matrix. This partially completes the problem of identifying the Markov chain $\{Z_n, S_n\}$. To completely define the process $\{Z_n, S_n\}$ we next give the initial distribution $p_o(i,j)$; $i=1,2,\ldots,M$, $j=0,1,2,\ldots,N$. Since the initial distributions of $\{Z_n\}$ and $\{S_n\}$ are assumed to be independent, we have

$$p_o(i,j) = \Pr\{Z_o=i\}\ \Pr\{S_o=j\} = a_o(i)s_o(j) .$$

$$\bar{\Gamma}_o = \begin{bmatrix} \Sigma_o^\infty & \bar{\alpha}_k & 0 & 0 & \cdots & 0 \\ \Sigma_1^\infty & \bar{\alpha}_k & \bar{\alpha}_o & 0 & \cdots & 0 \\ \Sigma_2^\infty & \bar{\alpha}_k & \bar{\alpha}_1 & \bar{\alpha}_o & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ \Sigma_N^\infty & \bar{\alpha}_k & \bar{\alpha}_{N-1} & \bar{\alpha}_{N-2} & \cdots & \bar{\alpha}_o \end{bmatrix}$$

$$\bar{\Gamma}_1 = \begin{bmatrix} \Sigma_1^\infty & \bar{\alpha}_k & \bar{\alpha}_o & 0 & \cdots & 0 & 0 \\ \Sigma_2^\infty & \bar{\alpha}_k & \bar{\alpha}_1 & \bar{\alpha}_o & \cdots & 0 & 0 \\ \Sigma_3^\infty & \bar{\alpha}_k & \bar{\alpha}_2 & \bar{\alpha}_1 & \cdots & 0 & 0 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots \\ \Sigma_N^\infty & \bar{\alpha}_k & \bar{\alpha}_{N-1} & \bar{\alpha}_{N-2} & \cdots & \bar{\alpha}_1 & \bar{\alpha}_o \\ \Sigma_{N+1}^\infty & \bar{\alpha}_k & \bar{\alpha}_N & \bar{\alpha}_{N-1} & \cdots & \bar{\alpha}_2 & \bar{\alpha}_1 \end{bmatrix}$$

$$b = \begin{bmatrix} b(0,1) \\ \vdots \\ b(0,M) \\ b(1,1) \\ \vdots \\ b(1,M) \\ \vdots \\ b(N,1) \\ \vdots \\ b(N,M) \end{bmatrix} \qquad D = \begin{bmatrix} b(0,1) & & & & & & 0 \\ & \ddots & & & & & \\ & & b(0,M) & & & & \\ & & & b(1,1) & & & \\ & & & & \ddots & & \\ & & & & & b(1,M) & \\ & & & & & & \ddots \\ & & & & & & & b(N,1) & \\ & & & & & & & & \ddots \\ & & & & & & & & & b(N,M) \end{bmatrix}$$

Figure 3. Some of the Matrices Defined.

4. DISTRIBUTION OF QUEUE SIZE

Let $s_n(j) = \Pr\{S_n=j\}$. Then, $s_n(j) = \Sigma_{i=1}^{M} p_n(i,j)$. Hence, the queue size distribution can be obtained easily once $p_n(i,j)$ are known. But the calculation of $p_n(i,j)$, or the vector p_n , is easy. By the general theory of Markov chains, $p_n = p_o P^n$ $(n \geq 0)$. The behavior of p_n as n approaches infinity is of much interest, and we will give such results in a later section. In some cases, the correlation between Z_n and S_n may be of interest. We are able to obtain any quantities of interest in measuring this correlation once p_n are found.

5. CLASSIFICATION OF THE STATES OF $\{Z_n, S_n\}$

In the following sections we will say the state of $\{Z_n, S_n\}$ is (i,j) if $Z_n = i$ and $S_n = j$.

Before classifying the states of $\{Z_n, S_n\}$, a study of the transition matrix is useful. P_{jk} was defined to be the submatrix of P whose (i,h) entry is $P(i,j;h,k)$. Now, partition the matrix D accordingly, so that the j-th diagonal submatrix of D is D_j whose i-th diagonal entry is $b(i,j)$. Then, from the formulas of page 83, we can write

$$P_{jk} = D_j \int_o^\infty \gamma_o(j,k,y)\, dA(x) + (I-D_j) \int_o^\infty \gamma_1(j,k,y)\, dA(x) . \qquad \text{(IV.5.1)}$$

From the formulas for $\gamma_o(j,k,y)$ and $\gamma_1(j,k,y)$ we see that $\gamma_o(j,j+1,y) = \gamma_o(j,j+2,y) = \ldots = \gamma_o(j,N,y) = 0$, as well as $\gamma_1(j,j+2,y) = \gamma_1(j,j+3,y) = \ldots = \gamma_1(j,N,y) = 0$. Hence,

$P_{j,j+2} = P_{j,j+3} = \cdots = P_{j,N} = 0$ for all $j \leq N-2$. Furthermore, we see that no one of the submatrices P_{jk} with $j \geq k$ can be completely zero no matter what $b(i,j)$ are. But, if $b(i,j) = 1$ for all i for some fixed j, i.e., if $D_j = I$, then $P_{j,j+1} = 0$. There exists at least one j with $D_j = 0$, namely $j = N$. Let v be the smallest such integer, i.e., let $v = \inf \{j : b(i,j) = 1, \ i=1,2,\ldots,M\}$.

$$P = \begin{bmatrix} P_{oo} & P_{o1} & 0 & \cdots & 0 \\ P_{1o} & P_{11} & P_{12} & \cdots & 0 \\ P_{2o} & P_{21} & P_{22} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ P_{No} & P_{N1} & P_{N2} & \cdots & P_{NN} \end{bmatrix}$$

Figure 4. Form of the Matrix P.

Lemma IV.5.1. If $v < N$, then all states (i,j) with $j > v$ are transient in the chain $\{Z_n, S_n\}$.

Proof. Let

$$P_1 = \begin{bmatrix} P_{oo} & P_{o1} & & & \\ \cdot & \cdot & \cdot & & \\ \cdot & & \cdot & & P_{v-1,v} \\ \cdot & & & \cdot & \\ P_{vo} & \cdot & \cdot & \cdot & P_{v,v} \end{bmatrix},$$

and define P_2 and P_3 in the obvious appropriate manner so that

$$P = \begin{bmatrix} P_1 & 0 \\ P_2 & P_3 \end{bmatrix} .$$

The zero matrix in the upper right hand corner of P is there since, by the discussion preceding this lemma, $P_{j,j+2} = P_{j,j+3} = \ldots = P_{jN} = 0$ for all $j \leq N-2$ and $P_{v,v+1} = 0$. Hence no state (i,j) with $j > v$ can be reached from any of the states (h,k) with $k \leq v$. But, since P_2 is not zero, it is possible to reach the set of states (h,k) with $k \leq v$ from the states (i,j) with $j > v$. Hence, all states (i,j) with $j > v$ are transient states in $\{Z_n, S_n\}$. Proof is complete.

Next, let $A_1(x)$ be the $M_1 \times M_1$ submatrix of $A(x)$ corresponding to the ergodic class of states in the process $\{Z(t)\}$. Then, by a suitable selection of $A_2(x)$ and $A_3(x)$ we can write

$$A(x) = \begin{bmatrix} A_1(x) & 0 \\ A_2(x) & A_3(x) \end{bmatrix} .$$

Putting this in the formula (IV.5.1), we obtain

$$P_{jk} = D_j \begin{bmatrix} \int_0^\infty \gamma_0(j,k,y)\, dA_1(y) & 0 \\ \int_0^\infty \gamma_0(j,k,y)\, dA_2(y) & \int_0^\infty \gamma_0(j,k,y)\, dA_3(y) \end{bmatrix} +$$

$$+ (I-D_j) \begin{bmatrix} \int_0^\infty \gamma_1(j,k,y)\, dA_1(y) & 0 \\ \int_0^\infty \gamma_1(j,k,y)\, dA_2(y) & \int_0^\infty \gamma_1(j,k,y)\, dA_3(y) \end{bmatrix}$$

$$\equiv \begin{bmatrix} P_{jk}^{(1)} & 0 \\ P_{jk}^{(2)} & P_{jk}^{(3)} \end{bmatrix} .$$

From this partitioning of P_{jk} $(j,k = 0,1,2,\ldots,N)$ the following lemma is obvious.

Lemma IV.5.2. A state (i,j) of the process $\{Z_n, S_n\}$ is transient if $i > M_1$.

The partitioning of the paragraph preceding the Lemma IV.5.2 suggests the next lemma which has Lemma IV.5.1 as a corollary. Let N_1 be the smallest integer such that $b(i,N_1) = 1$ for $i = 1,2,\ldots,M_1$, i.e., let $N_1 = \inf \{j : b(i,j) = 1 , \quad i = 1,2,\ldots,M_1\}$. Clearly, such a number exists and $N_1 \leq v$.

Lemma IV.5.3. If $N_1 < N$, then all states (i,j) with $j > N_1$ are transient in the chain $\{Z_n, S_n\}$.

Proof. Consider only $P_{jk}^{(1)}$. Clearly, $P_{jk}^{(1)}$ is zero whenever P_{jk} is zero, i.e., whenever $k \geq j + 2$. Furthermore, $P_{N_1,N_1+1}^{(1)}$ is zero since $b(i,N_1) = 1$ for all $i = 1,2,\ldots,M_1$. Hence, $P(h,k;i,j) = 0$ whenever $h \leq M_1$, $k \leq N_1$ and $j > N_1$. Thus none of the states (i,j) with $j > N_1$ can be reached from any of the states (h,k) with $h \leq M_1$, $k \leq N_1$. But, it is easy to see that it is possible to reach a state (h,k) with $h \leq M_1$ and $k \leq N_1$ from any of the states (i,j) with $j > N_1$. Hence, all states (i,j) with $j > N_1$ are transient in the chain $\{Z_n, S_n\}$.

We put the Lemmas IV.5.1, IV.5.2, and IV.5.3 together in the form of a theorem.

Theorem IV.5.4. A state (i,j) of the process $\{Z_n, S_n\}$ is transient if either $i > M_1$ or $j > N_1$.

Next we will show that these are the only transient states by proving that all states (i,j) with $i \leq M_1$ and $j \leq N_1$ are ergodic. In the proof of the following theorem we will write $(i,j) \to (h,k)$ if the state (h,k) can be reached from the state (i,j).

If h is an ergodic state of $\{Z_n\}$, i.e., if $h \leq M_1$, then $(i,j) \to (h,j)$ for all $i = 1,2,\ldots,M$ and $j = 0,1,\ldots,N$ since state h can be reached from any state i in the process $\{Z_n\}$. On the other hand, if $b(i,j) < 1$, i.e., if balking is not certain given the state (i,j), then $(i,j) \to (i,j+1)$ since $P(i,j;i,j+1) > 0$ in that case. With these in mind we will prove the following theorem, thus completing the classification of the states of the process $\{Z_n, S_n\}$.

Theorem IV.5.5. All states (i,j) with $i \leq M_1$ and $j \leq N_1$ are ergodic states in the chain $\{Z_n, S_n\}$; furthermore, they all belong to the same class.

Proof. For $j < N_1$, by the definition of N_1, there exists at least one k $(1 \leq k \leq M_1)$, say k_j, such that $b(k_j, j) < 1$. Now let (h,i) and (k,j) be any two states with $h,k \leq M_1$ and $i,j \leq N_1$. If $i = j$, then $(h,i) \to (k,j)$ and $(k,j) \to (h,i)$ clearly, since both h and k are ergodic states belonging to the same class in $\{Z_n\}$. Assume next that $i < j$. Then, since P_{ji} is not zero and h,k are ergodic in $\{Z_n\}$, $(k,j) \to (h,i)$. On the other hand, by the comments of the paragraph preceding this theorem, we have

$$(h,i) \to (k_i,i) \to (k_i,i+1) \to (k_{i+1},i+1) \to \ldots \to (k_{j-1},j-1) \to (k_{j-1},j) \to (k,j).$$

Hence, $(h,i) \to (k,j)$ also. Thus, $(h,i) \to (k,j)$ and $(k,j) \to (h,i)$ for all (h,i) and (k,j) with $i,j \leq N_1$, $k,h \leq M_1$. Therefore, all

such states belong to the same class and hence are all of the same type. Thus, all these states are ergodic, since otherwise there would be only transient states in view of Theorem IV.5.4. Proof is complete.

Corollary IV.5.5A. The process $\{Z_n, S_n\}$ is an irreducible Markov chain if and only if $M = M_1$ and $N = N_1$.

Proof. If $M = M_1$ and $N = N_1$, all states communicate with each other, (Theorem IV.5.5;) hence, $\{Z_n, S_n\}$ is an irreducible chain. On the other hand, if $M_1 < M$, the state (M_1+1, j) is transient; or if $N_1 < N$, the state (i, N_1+1) is transient (Theorem IV.5.4.) Then, it is impossible that $\{Z_n, S_n\}$ be irreducible when $M_1 < M$ or $N_1 < N$. Thus, the condition of the corollary is necessary.

In the special case where $b(i,j) = 0$ except for $b(i,N) = 1$, i.e., no balking is allowed unless the system is full, clearly, $N_1 = N$. If further $M = 1$, then the chain $\{S_n\}$ is an irreducible Markov chain by the preceding corollary. This is well known (cf. Takàcs [31].)

6. STEADY STATE BEHAVIOR OF $\{Z_n, S_n\}$

From the theorems of the preceding section the following is obvious from the general theory of Markov chains.

Theorem IV.6.1. The limiting probabilities $\lim_{n\to\infty} \Pr\{Z_n=i, S_n=j\}$, $(i=1,2,\ldots,M\ ;\ j=0,1,2,\ldots,N,)$ exist and are independent of the initial distribution $p_0(i,j)$.

Let $p(i,j) = \lim_{n\to\infty} p_n(i,j) = \lim_{n\to\infty} \Pr\{Z_n=i,\ S_n=j\}$. Clearly, $p(i,j) = 0$ whenever $i > M_1$ or $j > N_1$. Letting p to be the row vector whose $(jM+i)$ entry is $p(i,j)$ we have

$$p = \lim_{n\to\infty} p_n = \lim_{n\to\infty} p_o P^n = p_o \lim_{n\to\infty} P^n$$

$$pP = p \ , \quad pE = 1 \ .$$

Once P is known, p can be calculated as the unique solution of the system of linear equations

$$pP = p$$

$$pE = 1 \ .$$

Then, $\overset{\infty}{P} = \lim_{n\to\infty} P^n$ can be constructed easily as a square matrix of degree $M(N+1)$ whose every row is p .

7. WAITING TIMES

If we suppose, in particular, that the customers are served in the order of their arrival, then the waiting time behavior can be immediately deduced from the queue size.

Let $W_n(x)$ be the distribution of the waiting time of the n-th customer, given that he has joined the system; and let $W_n^*(s) = \int_o^\infty e^{-sx}\, dW_n(x)$. Then, $W_n^*(s)$ is $\beta^j/(\beta+s)^j$ given that the n-th arrival found j customers in the system. Hence,

$$W_n^*(s) = \sum_{i=1}^{M} \sum_{j=o}^{N} \left(\frac{\beta}{\beta+s}\right)^j p_n(i,j) \ .$$

Let $W(x) = \lim_{n\to\infty} W_n(x)$, and let $W^*(s) = \int_o e^{-sx}\, dW(x)$. This limiting distribution exists and is given below without proof.

Theorem IV.7.1. Waiting time of a customer in the steady state has the distribution $W(x)$ whose Laplace-Stieltjes transform is

$$W^*(s) = \sum_{i=1}^{M} \sum_{j=0}^{N} p(i,j) \left(\frac{\beta}{\beta+s}\right)^j \ .$$

8. BUSY PERIODS

Let Φ denote the length of a busy period in the steady state, and define $G(x)$ as its distribution function. Let $G_n(x)$ be the joint probability that a busy period is n services long, and that these n service times, $V_1, V_2, \ldots, V_n$, add to at most x units of time. Further, let Φ_m be the remaining busy period at time T_m, and introduce the distribution function

$$F(i,j,x) = \Pr\{\Phi_m = V_1 + \ldots + V_{j+n} \leq x \mid Z_m = i,\ S_m = j\}$$

for $n \geq 1$, and for $n = 0$,

$$F_o(i,j,x) = \begin{cases} 0 & \text{for } j = 0,\ i = 1,2,\ldots,M\ ; \\ \Pr\{\Phi_m = V_1+\ldots+V_j \leq x \mid Z_m = i,\ S_m = j\} & \text{otherwise.} \end{cases}$$

In other words, $F_n(i,j,x)$ is the joint probability that the server will be idle for the first time after $j+n$ services, and these services will total at most x units of time, given that the state of the $\{Z_n, S_n\}$ now is $Z_m = i$ and $S_m = j$.

Further let $H(i,x)$ be the distribution function of a busy period started by a customer of type i, and let $H_n(i,x)$ be the joint probability that a busy period started by a customer of type i is n services long and these n services take at most x units of time. Reminding that $a(i) = \lim_{m\to\infty} \Pr\{Z_m=i\}$, we write the interrelationships of these functions below.

$$H_n(i,x) = F_n(i,0,x) \tag{IV.8.1}$$

$$H(i,x) = \Sigma_{n=1}^{\infty} H_n(i,x) = \Sigma_{n=o}^{\infty} F_n(i,0,x) \tag{IV.8.2}$$

$$G_n(x) = \Sigma_{i=1}^{M} a(i)\, H_n(i,x) \tag{IV.8.3}$$

$$G(x) = \Sigma_{n=1}^{\infty} G_n(x) = \Sigma_{i=1}^{M} a(i)\, H(i,x) \tag{IV.8.4}$$

We will first derive $F_n(i,j,x)$; then the above relations can be used to calculate other functions of interest. Now, first let

$$R_1(i,j;h,k,x) = \Pr\{Z_{m+1}=h,\ S_{m+1}=k,\ X_{m+1} \le x,\ Y_m=0 \mid Z_m=i,\ S_m=j\}$$

$$Q_1(i,j;h,k,x) = \Pr\{Z_{m+1}=h,\ S_{m+1}=k,\ X_{m+1} \le x,\ Y_m=1 \mid Z_m=i,\ S_m=j\} .$$

Then, from the developments of page 82, we have

$$R_1(i,j;h,k,x) = \int_0^x b(i,j)\ \gamma_0(j,k,y)\ dA_{ih}(y)$$

$$Q_1(i,j;h,k,x) = \int_0^x (1-b(i,j))\ \gamma_1(j,k,y)\ dA_{ih}(y) .$$

Further, let $R_1(x)$ and $Q_1(x)$ be the square matrices of degree $M(N+1)$ whose $(jM+i,\ kM+h)$ entries are $R_1(i,j;h,k,x)$ and $Q_1(i,j;h,k,x)$ respectively for all i,j,h,k meaningful except for $k = 0$ for which case we define the $(jM+i,h)$ entries to be zero. Also, let $R_1^*(s)$ and Q_1^* be the Laplace-Stieltjes transforms of $R_1(x)$ and $Q_1(x)$ respectively. These can be obtained readily through the developments of pages 82,83.

Returning to the derivation of $F_n(i,j,x)$ we have, for $n = 0$, $j > 0$:

$$F_0(i,j,x) = \Pr\{\Phi_m = V_1 + \ldots + V_j \le x \mid Z_m = i,\ S_m = j\} =$$

$$= \int_0^x \Pr\{X_{m+1}>y \mid Z_m=i\}\ d_y\Pr\{V_1+\ldots+V_j\le y\}\ \Pr\{Y_m=0 \mid Z_m=i,\ S_m=j\}$$

$$+ \sum_{h=1}^{M}\sum_{k=1}^{N} \int_0^x \Pr\{\Phi_{m+1}=V_1+\ldots+V_k \le y \mid Z_{m+1}=h,\ S_{m+1}=k\}$$

$$d_y\ \Pr\{Z_{m+1}=h,\ S_{m+1}=k,\ X_{m+1}\le y,\ Y_m=0 \mid Z_m=i,\ S_m=j\} =$$

$$= \int_0^x b(i,j)\,(1 - \sum_{\ell=1}^{M} A_{i\ell}(y))\,\frac{e^{-\beta y}(\beta y)^j}{j!}\,\beta\,dy + \sum_{h=1}^{M}\sum_{k=1}^{N} \int_0^x F_0(h,k,x-y)\,dR_1(i,j;h,k,y)\ ;$$

and further for $n \geq 1$,

$$F_n(i,j,x) = \sum_{h=1}^{M}\sum_{k=1}^{N} \int_0^x \{F_n(h,k,x-y)\,dR_1(i,j;h,k,y) + F_{n-1}(h,k,x-y)\,dQ_1(i,j;h,k,y)\}\ .$$

Now let $F_n(x)$ be the column vector whose $(jM+i)$ entry is $F_n(i,j,x)$, and let $K(x)$ be the column vector whose $(jM+i)$ entry is

$$\int_0^x b(i,j)\,(1 - \sum_{\ell=1}^{M} A_{i\ell}(y))\,\frac{e^{-\beta y}(\beta y)^j}{j!}\,\beta\,dy\ .$$

Further, let $F_n^*(s)$ and $K^*(s)$ be the Laplace-Stieltjes transforms of $F_n(x)$ and $K(x)$ respectively. $K^*(s)$ can be computed by the help of the developments in page 83. Then the equations for $F_n(i,j,x)$ can be put in matrix notation as

$$F_0(x) = \int_0^x (dR_1(y))\,F_0(x-y) + K(x)$$

$$F_n(x) = \int_0^x (dR_1(y))\,F_n(x-y) + \int_0^x (dQ_1(y))\,F_{n-1}(x-y), \quad n \geq 1\ .$$

Taking Laplace-Stieltjes transforms everywhere, we obtain

$$F_0^*(s) = R_1^*(s)\,F_0^*(s) + K^*(s)$$

$$F_n^*(s) = R_1^*(s)\,F_n^*(s) + Q_1^*(s)\,F_{n-1}^*(s)\ , \quad n \geq 1\ . \qquad \text{(IV.8.5)}$$

Define the generating function $L(s,z) = \Sigma_{n=0}^{\infty} z^n F_n^*(s)$, $|z| \leq 1$; then, from (IV.8.5) ,

$$L(s,z) = \sum_{n=0}^{\infty} z^n R_1^*(s) F_n^*(s) + \sum_{n=1}^{\infty} z^n Q_1^*(s) F_{n-1}^*(s) + K^*(s)$$

$$= R_1^*(s) L(s,z) + zQ_1^*(s) L(s,z) + K^*(s) ,$$

so that,

$$(I-R_1^*(s) - z Q_1^*(s)) L(s,z) = K^*(s) . \qquad \text{(IV.8.6)}$$

Lemma IV.8.1. $I-R_1^*(s) - zQ_1^*(s)$ is non-singular for $|z| \leq 1$ and $\mathrm{Re}\{s\} \geq 0$.

Proof. Let $R_1 = R_1^*(0)$ and $Q_1 = Q_1^*(0)$. Clearly, then $R_1^*(s)$ is dominated by R_1 and $Q_1^*(s)$ is dominated by Q_1 , (i.e., $R_1^*(s) \ll R_1$, $Q_1^*(s) \ll Q_1$). Then, since $|z| \leq 1$, $zQ_1^*(s) \ll |z| Q_1 \ll Q_1$; and hence $R_1^*(s) + zQ_1^*(s) \ll R_1 + Q_1$ for all s with $\mathrm{Re}\{s\} \geq 0$ and $|z| \leq 1$.

Notice that R_1+Q_1 is the transition matrix P discussed earlier with the first M columns replaced by zeros. Since no row was completely zero in these first M columns, we have $(R_1+Q_1)E<PE = E$. Hence all eigenvalues of $R_1 + Q_1$ are less than one in absolute value (Theorem I.2.1). Then, from Theorem I.2.2, since $R_1^*(s) + z Q_1^*(s) \ll R_1 + Q_1$, all eigenvalues of $R_1^*(s) + z Q_1^*(s)$ are less than unity in absolute value. Therefore, $I - (R_1^*(s) + z Q_1^*(s))$ is non-singular for $|z| \leq 1$ and $\mathrm{Re}\{s\} \geq 0$.

Using this lemma we can write (IV.8.6) as:

$$L(s,z) = (I-R_1^*(s) - z Q_1^*(s))^{-1} K^*(s) . \qquad \text{(IV.8.7)}$$

Once $L(s,z)$ is obtained all the other functions of interest can be computed as follows. Let $L_0(s,z)$ be the $M \times 1$ column vector

obtained from $L(s,z)$ be keeping only the first M entries. Then, clearly from the definitions,

$$L_o(s,z) = \sum_{n=o}^{\infty} \int_o^{\infty} z^n e^{-sx} d H_n(x)$$

where $H_n(x)$ is the column vector whose i-th entry is $H_n(i,x)$. Thus $\int_o^{\infty} e^{-sx} dH_n(x)$, and hence $\int_o^{\infty} e^{-sx} dH_n(i,x)$ as its i-th entry, can be obtained from a Taylor's series expansion of $L_o(s,z)$ around a suitable point z within the unit circle. Calculation of $H(i,x)$ is easier: $\int_o^{\infty} e^{-sx} dH(i,x)$ can be obtained as the i-th entry of the vector $L_o(s,1)$. The distribution function $G(x)$ of the busy period can be obtained from $\int_o^{\infty} e^{-sx} dG(x) = a\, L_o(s,1)$ where a is the steady state distribution vector of $\{Z_n\}$.

9. CONCLUSION

We have investigated the queueing processes (queue size, waiting times, busy period) of a finite queue with a negative exponential server subject to a semi-Markov process input and a decision rule about joining the queue which depended on both the arrival process and the queue size.

Some generalizations with respect to number of servers and the service time distributions will be given in Chapter VI. Also in that chapter, some special cases, representing specializations with respect to balking mechanism and the arrival process, will be discussed.

In the next chapter we aim at an investigation of the properties of the stream of customers that decided to balk from the queue of this chapter.

CHAPTER V

STREAM BALKING FROM THE QUEUEING SYSTEM SM/M/1

1. INTRODUCTION

In the preceding chapter we have discussed the properties of a queueing system with a negative exponential server subject to an input which formed a semi-Markov process where customers are permitted to balk.

In this chapter we will investigate the stream formed by the customers who have chosen not to join the queue. We will show that this stream is equivalent to a semi-Markov process whose state space is the Cartesian product of the state spaces of the arrival process and the queue size process.

Some generalizations with respect to the number of servers and the service times, and some special cases such as the overflow problem will be treated in the next chapter.

2. NOTATION AND DEFINITIONS

We will use the notation and definitions already introduced in the preceding chapter. We assume the arrivals form a semi-Markov process identified by a_o as the vector of initial distribution, and $A(x)$ as the transition matrix. We assume that the arrival process $\{Z_n, X_n\}$ have M states, the first M_1 of which are ergodic, and the remaining ones are transient.

The state of the system at time t, denoted by $S(t)$, is the number of people in the queueing system at time t. It is assumed that the service times are negative exponentially distributed with mean $1/\beta$.

Customers are permitted to balk, not to join the queue. Let the instants of arrivals be $T_o, T_1, T_2, T_3, \ldots$, and let the instants of balkings be $\tau_o, \tau_1, \tau_2, \ldots$. We take, for convenience, $T_o = \tau_o = 0$. Obviously, for every k there exists an n such that $\tau_k = T_n$. We let $\zeta_k = Z_n$ and $\psi_k = S_n$ if $\tau_k = T_n$. Further, we let $X_n = T_n - T_{n-1}$, and $\theta_n = \tau_{n-1}$ $(n \geq 1)$, and $\theta_o = X_o = 0$. We take all other terms introduced in Chapter IV as they are without re-introducing them here.

Now let us center our attention on the underlying process $\{Z_n, S_n, X_n;\ n = 0,1,2,\ldots\}$. Assume that at time T_n there occurred a balking, i.e., $T_n = \tau_m$ for some m; then

$$R(i,j;h,k,x) = \Pr\{Z_{n+1}=h,\ S_{n+1}=k,\ X_{n+1} \leq x \mid Z_n=i,\ S_n=j,\ Y_n=0\},$$

is the joint probability that the arrival process will move from state i to h, and the queue size will change from j to k during the next interarrival time which is at most x long, $(i,h = 1,2,\ldots,M;\ j,k = 0,1,\ldots,N;\ x \geq 0)$. On the other hand, the joint probability that no balking occurs at time T_n and the same kind of changes occur in the process $\{Z_n, S_n, X_n\}$ is given by

$$Q(i,j;h,k,x) = \Pr\{Z_{n+1}=h,\ S_{n+1}=k,\ X_{n+1} \leq x,\ Y_n=1 \mid Z_n=i,\ S_n=j\},$$

$(i,h=1,2,\ldots,M;\ j,k=0,1,\ldots,N;\ x \geq 0)$. These probabilities can be

calculated by the following equations (see page 82)

$$R(i,j;h,k,x) = \int_0^\infty \gamma_0(j,k,y)\, d\, A_{ih}(y) ,$$

$$Q(i,j;h,k,x) = (1-b(i,j)) \int_0^\infty \gamma_1(j,k,y)\, d\, A_{ih}(y) .$$

Let $R(x)$ and $Q(x)$ be the square matrices of degree $M(N+1)$ whose $(jM+i, kM+h)$ entries are $R(i,j;h,k,x)$ and $Q(i,j;h,k,x)$ respectively. Then

$$R(x) = \int_0^x \Gamma_0(y) \oplus dA(y),$$

$$Q(x) = (I-D) \left(\int_0^x \Gamma_1(y) \oplus dA(y)\right) .$$

Thus, the Laplace-Stieltjes transforms of $R(x)$ and $Q(x)$ can be obtained as

$$R^*(s) = \overline{\Gamma}_0 \oplus A^*(s+\beta) \qquad (V.2.1)$$

$$Q^*(s) = (I-D)\, (\overline{\Gamma}_1 \oplus A^*(s+\mu)). \qquad (V.2.2)$$

The probabilities $Q(i,j;h,k,x)$ defined were "one-step" probabilities. Now let $Q^{(m)}(i,j;h,k,x)$ be the m-step probabilities defined by

$$Q^{(m)}(i,j;h,k,x) = \begin{cases} \delta_{ih}\, \delta_{jk}\, U(x) & \text{for } m = 0 \\ \Pr\{Z_{n+m}=h,\ S_{n+m}=k,\ X_{n+1}+\ldots+X_{n+m} \le x,\ Y_n=Y_{n+1}=\ldots=Y_{n+m-1}=1 \mid Z_n=i,\ S_n=j\} & \text{for } m \ge 1 . \end{cases}$$

$Q^{(m)}(i,j;h,k,x)$ is the joint probability that the arrival process will move from state i to h, and the queue size will change from j to

k during the next m interarrival times whose total is at most x units, and none of the customers balk. Further, let $Q^{(m)}(x)$ be the matrix whose $(jM+i,\ kM+h)$ entry is $Q^{(m)}(i,j;h,k,x)$.

Our primary aim in this chapter is to show that the process $\{\zeta_n, \psi_n, \theta_n\}$ is equivalent to a semi-Markov process. Accordingly we define

$$B(i,j;h,k,x) = \Pr\{\zeta_{n+1}=h,\ \psi_{n+1}=k,\ \theta_{n+1} \le x \mid \zeta_n=i,\ \psi_n=j\} ,$$

and define $B(x)$ as the matrix whose $(jM+i,\ kM+h)$ entry is $B(i,j;h,k,x)$, and further define $B^*(s)$ as the Laplace-Stieltjes transform of $B(x)$.

3. DERIVATION AND THE LIMIT OF $Q^{(m)}(x)$

We have already derived expressions for $R^*(s)$ and $Q^*(s)$, and hinted at the derivation of $Q^{(m)}(x)$. First, note in passing that $R = R^*(0)$ is a stochastic matrix:

$$\begin{aligned} R^*(s)E_{M(N+1)} &= (\bar{\Gamma}_o \oplus A^*(s+\beta))(E_{N+1} \oplus E_M) \\ &= (\bar{\Gamma}_o\ E_{N+1}) \oplus (A^*(s+\beta)\ E_M) \\ &= E_{N+1} \oplus (e^{-\bar{D}}\ A^*(s+\beta)\ E_M) \\ &= E_{N+1} \oplus (A^*(s)E_M) , \end{aligned}$$

thus,

$$RE = E_{N+1} \oplus A^*(0)E_M = E_{N+1} \oplus E_M = E \quad .$$

But this does not hold for $Q = Q^*(0)$. In fact,

$$Q^*(s)E_{M(N+1)} = (I-D)(\overline{\Gamma}_1 \oplus A^*(s+\beta))(E_{n+1} \oplus E_M)$$
$$= (I-D)((\overline{\Gamma}_1 E_{N+1}) \oplus (A^*(s+\beta)E_M))$$
$$= (I-D)(E_{N+1} \oplus (A^*(s)E_M)) \quad ,$$

so that

$$QE = (I-D)(E_{N+1} \oplus A E_M) = (I-D)E = E-b \ .$$

We now show that the probabilities $Q^{(m)}(i,j;h,k,x)$ obey the Chapman-Kolmogorov equations. From the definition, $Q^{(1)}(i,j;h,k,x) = Q(i,j;h,k,x)$. Now assume $m \geq 2$; then,

$$Q^{(m)}(i,j;h,k,x) = \Pr\{Z_m=h,\ S_m=k,\ X_1+X_2+\ldots+X_m \leq x\ ,$$
$$Y_o=Y_1=\ldots=Y_{m-1}=1 \mid Z_o=i,\ S_o=j\} =$$
$$= \sum_{g=1}^{M} \sum_{\ell=o}^{N} \int_o^x \Pr\{Z_{m-1}=g,\ S_{m-1}=\ell,\ X_1+\ldots+X_{m-1} \leq x-y,$$
$$Y_o=Y_1=\ldots=Y_{m-2}=1 \mid Z_o=i,\ S_o=j\}$$
$$d_y \Pr\{Z_m=h,\ S_m=k,\ X_m \leq y,\ Y_{m-1}=1 \mid Z_{m-1}=g, S_{m-1}=\ell\}$$
$$= \sum_{g=1}^{M} \sum_{\ell=o}^{N} \int_o^x Q^{(m-1)}(i,j;g,\ell,x-y)\ d_y Q(g,\ell;h,k,y) \ .$$

In matrix notation then,

$$Q^{(m)}(x) = \begin{cases} I\ U(x) & \text{for } m = 0 \\ Q(x) & \text{for } m = 1 \\ \int_o^x Q^{(m-1)}(x-y)\ dQ(y) & \text{for } m \geq 2 \ . \end{cases}$$

Now taking Laplace-Stieltjes transforms,

$$L\{Q^{(m)}(x)\} = \begin{matrix} I \\ \{ \\ L\{Q^{(m-1)}(x)\} \end{matrix} \quad Q^*(s) \quad ,$$

so that by a simple induction on m we obtain

$$L\{Q^{(m)}(x)\} = Q^*(s)^m , \qquad m \geq 0 . \tag{V.3.1}$$

The m-step probabilities $Q^{(m)}(i,j;h,k,x)$ were defined so that they are joint in the event that each one of the m customers joins the queue. Intuitively, it can be seen that this cannot go on forever, i.e., somebody would balk at some instant. We formalize this below.

Theorem V.3.1. Sequence of mass functions $Q^{(m)}(i,j;h,k,x)$ converge to zero uniformly for all $x \geq 0$ as m approaches infinity.

Proof. By Weierstrass' uniform convergence theorem, it is sufficient to prove that $Q^{(m)}(i,j;h,k)$ converges to zero as m approaches infinity since $0 \leq Q^{(m)}(i,j;h,k,x) \leq Q^{(m)}(i,j;h,k,\infty) = Q^{(m)}(i,j;h,k)$. Again by the same theorem, since $0 \leq Q^{(m)}(i,j;h,k) \leq \Sigma_h \Sigma_k Q^{(m)}(i,j;h,k)$, it is sufficient to show that $\Sigma_h \Sigma_k Q^{(m)}(i,j;h,k) \to 0$ as $m \to \infty$. But now,

$$\Sigma_h \Sigma_k Q^{(m)}(i,j;h,k) = \Sigma_h \Sigma_k \Pr\{Z_m=h;\ S_m=k;\ Y_o=\ldots=Y_{m-1}=1 \mid Z_o=i,\ S_o=j\}$$
$$= \Pr\{Y_o=Y_1=\ \ldots\ =Y_{m-1}=1 \mid Z_o=i,\ S_o=j\} .$$

On the other hand let N_1 and M_1 be as defined in Chapter IV; then, from Theorem IV.5.5, probability of finding $Z_n = M_1$ and $S_n = N_1$ for at least one n is equal to 1 no matter what Z_o, S_o are. Since the probability of balking when $Z_n=M_1$, $S_n = N_1$ is $b(M_1,N_1) = 1$, the probability that $Y_n=0$ for at least one $n \geq 0$ is 1 no matter what Z_o, S_o are. Hence, the complementary event $Y_o = Y_1 = \ldots = 1$ has probability zero, i.e.,

$$\Pr\{Y_n=1\ ,\quad n \geq 0 \mid Z_o=i,\ S_o=j\} = 0 \qquad \text{for all } i,j\ .$$

Hence,

$$\lim_{m\to\infty} \Sigma_h \Sigma_k\, Q^{(m)}(i,j;h,k) = \lim_{m\to\infty} \Pr\{Y_o=\ldots=Y_{m-1}=1 \mid Z_o,\ S_o=j\} = 0,$$

and the proof is complete.

We now restate this theorem in matrix language below without proof.

Theorem V.3.1'. The matrix sequence $\{Q^{(m)}(x)\}$ converges to zero uniformly for all $x \geq 0$ as m approaches infinity. Equivalently, the matrix sequence $\{Q^*(s)^m\}$ converges to zero uniformly for all s with $\mathrm{Re}\{s\} \geq 0$ as m approaches infinity.

This completes our examination of the sequence $\{Q^{(m)}(x)\}$ and enables us to start our investigation of the main process of interest in this chapter.

4. BALKING STREAM

Assume at time T_n a balking has occurred, i.e., $T_n = \tau_k$ for some k, and the arrival process and the queue size were in states $Z_n = \zeta_k = i$ and $S_n = \psi_k = j$ respectively. With this knowledge, one can predict the future of $\{Z_n, X_n\}$ and $\{Z_n, S_n, X_n\}$ in all future arrival epochs T_{n+1}, T_{n+2}, etc. In particular, one can compute the probabilities related to ζ_{k+1}, ψ_{k+1}, θ_{k+1} with only a knowledge of ζ_k, ψ_k. Subject of this section is to prove these assertions and identify the balking process $\{\zeta_k, \psi_k, \theta_k\}$.

Theorem V.4.1. The process $\{\zeta_k, \psi_k, \Theta_k;\ k = 0,1,\ldots\}$ is equivalent to a semi-Markov process defined by the vector of initial probabilities p_o and the transition matrix $B(x)$.

Proof. Since $T_o=\tau_o=0$, $\zeta_o=Z_o$ and $\psi_o=S_o$. Thus $\Pr\{\zeta_o=i,\ \psi_o=j\} = \Pr\{Z_o=i,\ S_o=j\} = p_o(i,j)$.

On the other hand,

$$\Pr\{\zeta_{k+1}=i,\ \psi_{k+1}=j,\ \Theta_{k+1} \le x \mid \zeta_o,\ldots,\zeta_k;\ \psi_o,\ldots,\psi_k;\ \Theta_1,\ldots,\Theta_k\} =$$

$$= \sum_{m=1}^{\infty} \Pr\{Z_{n+m}=i,\ S_{n+m}=j,\ X_{n+1} + \ldots + X_{n+m} \le x,\ Y_{n+1} = \ldots = Y_{n+m-1}=1,\ Y_{n+m}=0 \mid \zeta_o,\ldots,\zeta_{k-1},\ \zeta_k=Z_n;\ \psi_o,\ldots,\psi_{k-1},\ \psi_k=S_n;\ \Theta_1,\ldots,\Theta_k;\ Y_n=0\}$$

$$= \sum_{m=1}^{\infty} \Pr\{Z_{n+m}=i,\ S_{n+m}=j,\ X_{n+1} + \ldots + X_{n+m} \le x,\ Y_{n+1} = \ldots = Y_{n+m-1}=1,\ Y_{n+m}=0 \mid Z_n=\zeta_k,\ S_n=\psi_k,\ Y_n=0\}$$

$$= \Pr\{\zeta_{k+1}=i,\ \psi_{k+1}=j,\ \Theta_{k+1} \le x \mid \zeta_k,\ \psi_k\}$$

$$= B(\zeta_k,\ \psi_k\ ;\ i,j,x) \qquad \text{as claimed.}$$

Proofs of the following corollaries follow immediately from the general theory of semi-Markov processes.

Corollary V.4.1.A. The process $\{\zeta_k,\ \psi_k\}$ is a Markov chain defined by (p_o, B).

Corollary V.4.1.B. The process $\{\zeta_k,\ \psi_k,\ \tau_k\}$ is a Markov process.

Corollary V.4.1.C. The interbalking times $\Theta_1, \Theta_2, \ldots$ are mutually conditionally independent given $\zeta_o, \zeta_1, \ldots$ and $\psi_o, \psi_1, \ldots$.

During the proof of the above theorem we already have derived a relation that can be used in computing the transition probabilities. We have

$$B(i,j;h,k,x) = \Pr\{\zeta_{n+1}=h,\ \psi_{n+1}=k,\ \theta_{n+1} \le x \mid \zeta_n=i,\ \psi_n=j\} =$$

$$= \sum_{m=1}^{\infty} \Pr\{Z_m=h;\ S_m=k;\ X_1+\ldots+X_m \le x;\ Y_1=Y_2=\ldots=Y_{m-1}=1,$$
$$Y_m=0 \mid Z_o=i,\ S_o=j,\ Y_o=0\}$$

$$= \sum_{m=1}^{\infty} \sum_{g=1}^{M} \sum_{\ell=o}^{N} \int_o^x \Pr\{Z_1=g,\ S_1=\ell,\ X_1 \le x-y \mid Z_o=i,\ S_o=j,\ Y_o=0\} \cdot$$
$$d_y \Pr\{Z_m=h,\ S_m=k,\ X_2+\ldots+X_m \le y;\ Y_1=\ldots=Y_{m-1}=1;$$
$$Y_m=0 \mid Z_1=g,\ S_1=\ell\}$$

$$= \sum_{m=1}^{\infty} \sum_{g=1}^{M} \sum_{\ell=o}^{N} \int_o^x \Pr\{Z_1=g,\ S_1=\ell,\ X_1 \le x-y \mid Z_o=i,\ S_o=j,\ Y_o=0\}$$
$$d_y \Pr\{Z_m=h,\ S_m=k,\ X_2+\ldots+X_m \le y;\ Y_1=\ldots=Y_{m-1}=1 \mid$$
$$Z_1=g,\ S_1=\ell\}\ \Pr\{Y_m=0 \mid Z_m=h,\ S_m=k\}$$

$$= \sum_{m=1}^{\infty} \sum_{g=1}^{M} \sum_{\ell=o}^{N} \int_o^x R(i,j;\ g,\ell,x-y)\ d_y Q^{(m-1)}(g,\ell;h,k,y)\ b(h,k) .$$

In matrix notation this can be written concisely as

$$B(x) = \sum_{m=1}^{\infty} \int_o^x R(x-y)\ dQ^{(m-1)}(y) D .$$

Now taking Laplace-Stieltjes transforms, remembering that $L\{Q^{(m)}(y)\} = Q^*(s)^m$ by (V.3.1), we obtain

$$B^*(s) = \sum_{m=1}^{\infty} R^*(s)\ Q^*(s)^{m-1}\ D . \qquad (V.4.1)$$

Of course we need to show that $\sum_{m=o}^{\infty} Q^*(s)^m$ converges. This we do next.

Lemma V.4.2. $I-Q^*(s)$ is non-singular, and the power series $\Sigma_{m=0}^{\infty} Q^*(s)^m$ converges to $(I-Q^*(s))^{-1}$ for all s with $Re\{s\} \geq 0$.

Proof. From Theorem V.3.1' , $Q^*(s)^m \to 0$ as $m \to \infty$ for all s with $Re\{s\} \geq 0$. Then, the lemma follows from Theorem I.2.5.

Using this lemma in (V.4.1) we obtain

$$B^*(s) = R^*(s)(I-Q^*(s))^{-1} D . \qquad (V.4.2)$$

Clearly, $B = B^*(0)$ should be a stochastic matrix. To check: $B = R(I-Q)^{-1}D$; hence $BE = R(I-Q)^{-1}DE$. Now from page 102, $QE = E - DE$; so that, $DE = E - QE = (I-Q)E$. Hence, $BE = R(I-Q)^{-1}(I-Q)E = RE = E$ as needed.

Our identification of the process $\{\zeta_k, \psi_k, \Theta_k\}$ is now complete, since the initial distribution vector p_o was already computed in Chapter IV.

5. THE PROCESS $\{\zeta_n, \psi_n\}$

The Markov chain $\{\zeta_n, \psi_n\}$ is well defined by p_o and B, (see Corollary V.4.1.A.) In this section we will classify its states and give some steady state results.

For every $n \geq 0$, there exists a k such that $\zeta_n = Z_k$ and $\psi_n = S_k$. Hence the state space for $\{\zeta_n, \psi_n\}$ is a subset of the state space of the process $\{Z_n, S_n\}$. Note that, for any n , $\zeta_n=i$ and $\psi_n=j$ only if there exists a k such that $T_k=\tau_n$, $Z_k=i$, $S_k=j$, and $Y_k=0$. Therefore, if a state (i,j) is transient in the process $\{Z_n, S_n\}$, then that state is transient in $\{\zeta_n, \psi_n\}$ also. Thus, the lemma below follows from Theorem IV.5.4.

Lemma V.5.1. A state (i,j) of the process $\{\zeta_n,\psi_n\}$ is transient if either $i > M_1$ or $j > N_1$.

However, these are not the only transient states. Assume for some i and j $b(i,j)=0$. Then, no customer can balk if $Z_n=i$ and $S_n=j$. Hence, it is impossible that such a state ever be visited by the process $\{\zeta_n,\psi_n\}$. This can be seen easily from an examination of the transition matrix $B = R(I-Q)^{-1}D$. Note that if $b(i,j)=0$, then the $jM+i^{th}$ diagonal entry of D is zero, and hence the $jM+i^{th}$ column of B is completely filled with zeros, i.e., such a state can never be reached from any state. In Chapter III, we called such a state an "empty state", a state which the process never visits. Clearly, an empty state is a trivial transient state. In fact, such states can safely be discarded from the state space without any loss. Adding the arguments of this paragraph to Lemma V.5.1 we obtain:

Theorem V.5.2. A state (h,k) of the process $\{\zeta_n,\psi_n\}$ is transient if

(i) $h > M_1$, or

(ii) $k > N_1$, or

(iii) $b(h,k) = 0$.

Actually these are the only transient states. We will show this by proving that all other states are ergodic.

Theorem V.5.3. A state (h,k) of the process $\{\zeta_n,\psi_n\}$ is ergodic if and only if

(i) $h \leq M_1$ and

(ii) $k \leq N_1$ and

(iii) $b(h,k) > 0$.

Furthermore, all the ergodic states are in the same class.

Proof. If anyone of the three conditions is not fulfilled, then by Theorem V.5.2, (h,k) is a transient state. Hence the conditions are necessary.

Proof of sufficiency: Consider the set C of all states (h,k) with $h \leq M_1$, $k \leq N_1$, and $b(h,k) > 0$. Let (i,j) and (h,k) belong to C . Then, by Theorem IV.5.5, (h,k) can be reached from (i,j) and vice versa in the process $\{Z_n, S_n\}$; i.e.,

$$\Sigma_{n=1}^{\infty} \Pr \{Z_n=h,\ S_n=k \mid Z_o=i,\ S_o=j\} = \infty .$$

But now, assume $\zeta_o = i$, $\psi_o = j$; then,

$$\Pr \{\zeta_m=h,\ \psi_m=k \mid \zeta_o=i,\ \psi_o=j\} =$$
$$= \Sigma_{n=1}^{\infty} \Pr \{\tau_m=T_n \mid Z_n=h,\ S_n=k\} \Pr\{Z_n=h,\ S_n=k \mid Z_o=i,\ S_o=j\} .$$

Hence,

$$\Sigma_{m=1}^{\infty} \Pr \{\zeta_m=h,\ \psi_m=k \mid \zeta_o=i,\ \psi_o=j\} =$$
$$= \Sigma_{n=1}^{\infty} b(h,k) \Pr\{Z_n=h,\ S_n=k \mid Z_o=i,\ S_o=j\} = \infty$$

since $b(h,k) > 0$. Thus, (h,k) can be reached from (i,j) . We have shown that any two states of C can be reached from each other. Hence, C has only one type of states, either all ergodic or all transient. But they cannot be only transient since then, in view of

Theorem V.5.2, all the states of $\{\zeta_n, \psi_n\}$ would be transient: an impossibility since in a finite Markov chain not all states are transient. Therefore, all states belonging to C are ergodic. Proof is complete.

In view of this last theorem, limiting probabilities exist and are independent of the initial distribution as given by p_o. Now let $\pi(i,j) = \lim_{n\to\infty} \Pr\{\zeta_n=i, \psi_n=j\}$, and let Π be the row vector whose $(jM+i)$ entry is $\pi(i,j)$. Clearly Π is the unique solution of the system of linear equations

$$\Pi B = \Pi$$

$$\Pi E = 1 .$$

There exists a very close relationship between Π and p where p was defined as the limiting distribution vector for the process $\{Z_n, S_n\}$, i.e., $(jM+i)$ entry of p was $p(i,j) = \lim_{n\to\infty} \Pr\{Z_n=i, S_n=j\}$.

Theorem V.5.4. Limiting probabilities $\pi(i,j)$, $(i=1,2,\ldots,M;$ $j=0,1,\ldots,N,)$ exist and are independent of the initial distribution $p_o(i,j)$. Furthermore,

$$\pi(i,j) = \frac{p(i,j)\, b(i,j)}{\sum_{i=1}^{M} \sum_{j=0}^{N} p(i,j)\, b(i,j)} .$$

Proof. Existence and the independence from the initial distribution are obvious. Now, by the definition of p, $pP=p$. Note that $P = DR+Q$, so that $pP = pDR + pQ = p$, therefore $pDR = p - pQ = p(I-Q)$. Then,

$$(pD)B = pD(R(I-Q)^{-1}D) = (pDR)(I-Q)^{-1}D = p(I-Q)(I-Q)^{-1}D = pD .$$

Hence, pD is an eigenvector of B corresponding to the eigenvalue 1. Then, dividing each element of pD by the sum of its elements, $pDE = pb$, we obtain $\frac{1}{pb} pD$; (notice that $pb \geq p(M_1,N_1)\, b(M_1,N_1) = p(M_1,N_1) > 0$ since the state (M_1,N_1) is an ergodic state in $\{Z_n,S_n\}$.) Then, clearly,

$$(\frac{1}{pb} pD)B = \frac{1}{pb} pD ,$$

and,

$$(\frac{1}{pb} pD)E = 1 .$$

From the uniqueness of Π, then, $\Pi = \frac{1}{pb} pD$. Proof is complete.

Next, let $\overset{\infty}{B} = \lim_{n\to\infty} B^n$. Then, $\overset{\infty}{B}$ has all its rows identically equal to Π. From the last theorem, we also obtain:

$$\overset{\infty}{B} = E \oplus \Pi = \frac{1}{pb} E \oplus pD = \frac{1}{pb} \overset{\infty}{P} D .$$

6. THE PROCESS $\{\zeta_n,\psi_n,\theta_n\}$

Theorem V.4.1 show that $\{\zeta_n,\psi_n,\theta_n\}$ is equivalent to a semi-Markov process and is defined by the vector of initial probabilities p_o and the transition matrix $B(x)$ whose Laplace-Stieltjes transform is given by $B^*(s) = R^*(s)\,(I-Q^*(s))^{-1}D$. In this section we will classify the states of this process and give some steady state results.

First let,

$$H_i(x) = \sum_{j=1}^{M} A_{ij}(x) , \qquad G(i,j,x) = \sum_{h=1}^{M} \sum_{k=0}^{N} B(i,j;h,k,x) ,$$

$$\mu_i = \int_0^\infty x\, dH_i(x), \qquad \eta(i,j) = \int_0^\infty x\, dG(i,j,x) .$$

Then, $H_i(x)$ is the distribution of the occupancy time of state i in the process $\{Z(t)\}$, and μ_i is the mean of that occupancy time; $G(i,j,x)$ is the distribution of the occupancy time of the state (i,j) by the process $\{\zeta_n, \psi_n, \theta_n\}$; and $\eta(i,j)$ is the expected value of that occupancy time. Further, let μ be the M by 1 column vector of μ_i's, and let η be the column vector whose $jM+i^{th}$ entry is $\eta(i,j)$. Clearly, then,

$$\mu = -\left.\frac{dA^*(s)}{ds}\right|_{s=o} E, \quad \text{and} \quad \eta = -\left.\frac{dB^*(s)}{ds}\right|_{s=o} E .$$

Now,

$$\begin{aligned}\frac{d}{ds} B^*(s) &= \frac{d}{ds}\left(R^*(s)(I-Q^*(s))^{-1}D\right) \\ &= \frac{dR^*(s)}{ds}(I-Q^*(s))^{-1}D + R^*(s)(I-Q^*(s))^{-1}\frac{dQ^*(s)}{ds}(I-Q^*(s))^{-1}D \\ &= \left(\frac{dR^*(s)}{ds} + R^*(s)(I-Q^*(s))^{-1}\frac{dQ^*(s)}{ds}\right)(I-Q^*(s))^{-1}D ,\end{aligned}$$

so that,

$$\left.\frac{d}{ds}B^*(s)\right|_{s=o} = \left(\left.\frac{dR^*(s)}{ds}\right|_{s=o} + R(I-Q)^{-1}\left.\frac{dQ^*(s)}{ds}\right|_{s=o}\right)(I-Q)^{-1}D .$$

Remembering that $(I-Q)E = DE$, we obtain

$$\left.\frac{d}{ds}B^*(s)\right|_{s=o} E = \left(\left.\frac{dR^*(s)}{ds}\right|_{s=o} + R(I-Q)^{-1}\left.\frac{dQ^*(s)}{ds}\right|_{s=o}\right) E .$$

Now, from page 101,

$$\begin{aligned}R^*(s)E &= E_{n+1} \oplus A^*(s)E_M \\ Q^*(s)E &= (I-D)\,R^*(s)E ;\end{aligned}$$

hence,

$$- \frac{dR^*(s)}{ds}\Big|_{s=o} E = E_{n+1} \oplus \frac{-dA^*(s)}{ds}\Big|_{s=o} E_M = E_{N+1} \oplus \mu ,$$

and

$$- \frac{dQ^*(s)}{ds}\Big|_{s=o} E = (I-D)\,(E_{N+1} \oplus \mu).$$

Putting these in the above, we have

$$\eta = - \frac{dB^*(s)}{ds}\Big|_{s=o} E = (E_{N+1} \oplus \mu) + R(I-Q)^{-1}(I-D)(E_{N+L} \oplus \mu)$$

$$\eta = (I-B+R(I-Q)^{-1})(E_{n+1} \oplus \mu) . \qquad (V.6.1)$$

Next is a theorem on the classification of the states of the process $\{\zeta_n, \psi_n, \Theta_n\}$.

Theorem V.6.1. A state (i,j) of the process $\{\zeta_n, \psi_n, \Theta_n\}$ is ergodic if and only if $i \leq M_1$ and $j \leq N_1$ and $b(i,j) > 0$. Furthermore, all ergodic states are in the same class. A state is transient if it is not ergodic.

Proof. By Theorem I.3.2, a state of a semi-Markov process is transient if it is transient in the corresponding Markov chain. Then, whenever $i > M_1$, or $j > N_1$, or $b(i,j) = 0$, the state (i,j) is transient since it is transient in $\{\zeta_n, \psi_n\}$ by Theorem V.5.2. Hence the conditions of the theorem are necessary.

To prove the sufficiency, consider the state (i,j) with $i \leq M_1$, $j \leq N_1$, and $b(i,j) > 0$. Then, from Theorem V.5.3, (i,j) is ergodic in $\{\zeta_n, \psi_n\}$. Further, the solution of η given by (V.6.1) shows that η is finite since μ is finite, i.e., $\eta(h,k)$ is finite for all h and k . Thus, by Theorem I.3.2, the state (i,j) is ergodic.

The fact that all ergodic states are in the same class follows from Theorem V.5.3.

7. LENGTHS OF INTERBALKING TIMES

In this section we will put together all the results relating to $\{\theta_n\}$, the sequence of interbalking times. We have already shown that points of balking are regeneration points for the balking process. Time between balkings are dependent on the states of the arrival process and the queue size. Although balking intervals are dependent on each other, this dependence is through the states of the other processes $\{\zeta_n\}$ and $\{\psi_n\}$, so that, given $\{\zeta_n, \psi_n\}$ and $(\zeta_{n+m}, \psi_{n+m})$, the intervals θ_{n+1} and θ_{n+m+1} are stochastically independent.

We have derived the expected value of θ_{n+1} conditional on $\{\zeta_n, \psi_n\}$ in the preceding section, (formula V.6.1). To find the expected value of θ_{n+1} with no conditions we need to find $\Pi_n \eta$ where Π_n is a probability vector whose $jM+i^{th}$ element is $\Pr\{\zeta_n=i, \psi_n=j\}$. Thus, in general θ_n $(n=1,2,\ldots)$ have different means. But, in the steady state, these means become equal to $\Pi\eta$. In fact, the marginal distributions of $\{\theta_n\}$ become identical in the steady state: from,

$$\Pr\{\theta_n \leq x\} = \Sigma_i \Sigma_j G(i,j,x)\, \pi_{n-1}(i,j)$$

we get

$$\lim_{n\to\infty} \Pr\{\theta_n \leq x\} = \Sigma_i \Sigma_j G(i,j,x)\, \pi(i,j) . \qquad (V.7.1)$$

Let $\overline{\eta}$, $\overline{\mu}$ denote the expected values of θ_n and X_n respectively in the steady state. Then, clearly,

$\overline{\eta} = \Pi\eta$ and $\overline{\mu} = p\mu$. But now, from (V.6.1),

$$\begin{aligned}\overline{\eta} = \Pi\eta &= \Pi\,(I-B+R(I-Q)^{-1})(E_{N+1} \oplus \mu) \\ &= (\Pi - \Pi B + \frac{1}{pb}\, pDR(I-Q)^{-1})(E_{N+1} \oplus \mu) \\ &= (\Pi - \Pi + \frac{1}{pb}\, p)\,(E_{n+1} \oplus \mu) \\ &= \frac{1}{pb}\, p(E_{N+1} \oplus \mu)\end{aligned}$$

Now,

$$\begin{aligned}p(E_{n+1} \oplus \mu) &= \sum_{i=1}^{M} \sum_{j=0}^{N} p(i,j)\, \mu_i \\ &= \sum_{i=1}^{M} a(i)\, \mu_i \\ &= \overline{\mu} \ .\end{aligned}$$

Thus, we obtain $\overline{\eta} = \frac{1}{pb}\, \overline{\mu}$. Noticing that pb is the probability that a customer balks in the steady state, we can state this above result in words as: "in the steady state, the expected value of a balking interval times the ratio of balking customers is equal to the mean interarrival time."

Next we give a characterization of balking processes with independent interbalking times in the following three theorems.

Theorem V.7.1. The interbalking times $\theta_1, \theta_2, \theta_3, \ldots$ are independent and identically distributed if and only if

(i) $M_1 = 1$

(ii) $b(1,j) = 0$ for $j < N_1$; and

(iii) $Z_0 = 1$, $S_0 = N_1$.

Theorem V.7.2. The interbalking times $\theta_1, \theta_2, \theta_3, \ldots$ form a modified renewal process, (i.e., $\theta_1, \theta_2, \ldots$ are independent and $\theta_2, \theta_3, \ldots$ are indentically distributed,) if and only if

(i) $M_1 = 1$

(ii) $b(1,j) = 0$ for $j < N_1$, and

either (iii) $Z_o = 1$, $S_o \leq N_1$

or (iii') $Z_o = 1$, $S_o > N_1$, and $b(1,S_o) = 1$, $b(1,j) = 0$ for $j = N_1+1, N_1+2, \ldots, S_o-1$.

Theorem V.7.3. Assume the process has started at time $-\infty$. Then, the interbalking times $\theta_1, \theta_2, \ldots$ are independent and identically distributed if and only if

(i) $M_1 = 1$

(ii) $b(1,j) = 0$ for $j < N_1$.

Proof of Theorem V.7.1.

Proof of sufficiency. Assume $M_1 = 1$, $b(1,j) = 0$ for $j < N_1$, and $Z_o=1$, $S_o=N_1$. Then, from Theorem V.6.1, the only ergodic state in the process $\{\zeta_n, \psi_n, \theta_n\}$ is $(1, N_1)$. Since the process is initially in that state, it can never get to any other state; hence $\zeta_n=1$ and $\psi_n=N_1$ for all $n \geq 0$. There is only one state of the process $\{\zeta_n, \psi_n, \theta_n\}$, and hence the balking process $\{\zeta_n, \psi_n, \theta_n\}$ is equivalent to a simple renewal process; i.e., $\theta_1, \theta_2, \theta_3, \ldots$ are independent and identically distributed as $G(1, N_1, x)$.

Proof of sufficiency. Assume $\{\zeta_n, \psi_n, \theta_n\}$ is a simple renwal process. Then, there exists only one state. By Theorem V.6.1, the state (M_1, N_1) is an ergodic state; hence this must be the only

state. Then by Theorem V.6.1 again, we must have $b(i,j) = 0$ for $i < M_1$ and $j < N_1$ in order for (M_1,N_1) to be the only state. But for $M_1 > 1$, by the definition of N_1, $b(M_1-1,N_1) = 1$. Thus, it is necessary that $M_1=1$, and, we only need $b(1,j) = 0$ for $j < N_1$. Furthermore, since $(1,N)$ is the only state permissible, it is necessary that $\psi_0 = S_0 = N_0$ and $\zeta_0 = Z_0 = M_1$. Proof of Theorem V.7.1 is complete.

Proof of Theorem V.7.2.

Proof of sufficiency. Assume (i), (ii), (iii) or (iii'). Then,

$$B(i,j;h,k) = \begin{cases} 1 & \text{for } h = M_1 = 1, \ k = N_1 \\ 0 & \text{otherwise .} \end{cases}$$

Thus $(\zeta_n,\psi_n) = (M_1,N_1) = (1,N_1)$ for all $n \geq 1$; and hence $\{\zeta_n,\psi_n,\theta_n;\ n \geq 1\}$ is equivalent to a renewal process, i.e., $\theta_2,\theta_3,\theta_4,\ldots$ are independent and indentically distributed as $G(1,N_1,x)$. Since (ζ_n,ψ_n) is known for all $n \geq 1$, θ_1 is independent of all the others.

Proof of necessity. Assume $\theta_2,\theta_3,\ldots$ are independent and identically distributed. Then, by Theorem V.7.1, it is necessary that $M_1 = 1$, $b(1,j) = 0$ for $j < N_1$, and $\zeta_1 = 1$, $\psi_1 = N_1$. Now then, we need to have

$$B(i,j;h,k) = \begin{cases} 1 & \text{if } h=1, \ k=N_1 \\ 0 & \text{otherwise.} \end{cases}$$

This is possible only if $i=1$ and $j \leq N_1$, or, j could be greater than N_1 but then we need condition (iii'). Proof is complete.

Proof of Theorem V.7.3.

Sufficiency. Assume (i) and (ii). Then the only ergodic state is $(1,N_1)$, hence if the process has started at $-\infty$, then at time zero it is already absorbed into state $(1,N_1)$, i.e., $\zeta_o = 1$, $\psi_o = N_1$. Then the sufficiency follows from Theorem V.7.1.

Necessity. There needs to be only one ergodic state. Since (M_1,N_1) is one, it must be the only one. This is possible only if $M_1 = 1$ and $b(1,j) = 0$ for $j < N_1$ by Theorem V.6.1. Proof is complete.

8. NUMBER OF CUSTOMERS JOINING THE QUEUE DURING AN INTERBALKING TIME

Let ξ_{ij} be the number of customers that join the queue during an interbalking time starting with $\zeta_n=i$, $\psi_n=j$; and let ξ be a column vector of random variables whose $jM+i^{th}$ element is ξ_{ij}. Then,

$$\Pr\{\xi_{ij}=m\} = \sum_h \sum_k \Pr\{Z_{m+1}=h,\ S_{m+1}=k;\ Y_1 = \cdots = Y_m = 1,\ Y_{m+1}=0 \mid Z_o=i,\ S_o=j,\ Y_o=0\}$$

$$= \sum_h \sum_k \sum_g \sum_\ell R(i,j;g,\ell)\, Q^{(m)}(g,\ell;h,k)\, b(h,k) \quad ;$$

or in matrix notation,

$$\Pr\{\xi=mE\} = RQ^m b \ . \tag{V.8.1}$$

Now let $\phi_{ij}(z)$ be the generating function of ξ_{ij}, i.e., let $\phi_{ij}(z) = \Sigma_{m=o}\, z^m \Pr\{\xi_{ij}=m\}$; and further let $\phi(z)$ be the column vector of these generating functions. Then,

$$\phi(z) = \sum_{m=0}^{\infty} z^m \Pr\{\xi = mE\}$$

$$= \sum_{m=0}^{\infty} z^m R Q^m b \quad .$$

From Theorem V.3.1', $Q^m \to 0$ as $m \to \infty$. Then, since $zQ \ll |z| Q \ll Q$, $(zQ)^m \to 0$ as $m \to \infty$ for $|z| \leq 1$. Thus, by Theorem I.2.5, the matrix power series $\Sigma_{m=0}^{\infty} (zQ)^m$ converges to $(I-zQ)^{-1}$. Putting this above, we obtain

$$\phi(z) = R(I-zQ)^{-1} b , \qquad |z| \leq 1 . \tag{V.8.2}$$

Note that $\phi(1) = R(I-Q)^{-1}b = RE = E$ as needed.

Mean number of customers joining the queue during a balking interval starting in state (i,j) can now be computed as $\nu_{ij} = \frac{d}{dz} \phi_{ij}(z) \Big|_{z=1}$. Let ν be the column vector with these means as entries in the usual arrangement. Then,

$$\nu = \frac{d}{dz} \phi(z)\Big|_{z=1} = RQ (I-zQ)^{-2} b \Big|_{z=1}$$

$$= RQ (I-Q)^{-2} b$$

$$= RQ (I-Q)^{-1} E .$$

Since Q and $(I-Q)^{-1}$ are commutative,

$$\nu = R(I-Q)^{-1}QE = R(I-Q)^{-1}(I-(I-Q))E = R(I-Q)^{-1}E - E .$$

The unconditional distribution of the number of customers joining the queue can also be obtained easily, especially in the steady state. Now let $\bar{\xi}$ be the number of customers joining the queue during an interbalking time. Then,

$$\Pr\{\bar{\xi}=m\} = \sum_i \sum_j \pi(i,j) \Pr\{\xi_{ij}=m\} = \Pi \Pr\{\xi=m\} = \Pi R Q^m b .$$

The generating function for $\bar{\xi}$ can be calculated easily as

$$\bar{\phi}(z) = \sum_{m=0}^{\infty} z^m \Pr\{\bar{\xi}=m\}$$

$$= \Pi R (I-zQ)^{-1} b .$$

Then the expected value $\bar{\nu}$ of $\bar{\xi}$ is

$$\bar{\nu} = \Pi R(I-Q)^{-1}E - \Pi E$$

$$= \frac{1}{pb} pDR (I-Q)^{-1}E - 1$$

$$= \frac{1}{pb} - 1 .$$

Noting that 100 pb is the percent of customers balking, $\frac{1}{pb}$ is the expected number of customers arriving per balking period. Then, subtracting one for the one who balks we have the expected number of customers that do join the queue in a balking period.

9. CONCLUSION

In this chapter we have discussed the properties of the balking stream. We have shown that the balking stream forms a semi-Markov process. We have indentified the process, gave a classification of its states and some steady state results. Further we have provided a characterization of all balking streams with independent and identically distributed interbalking times.

All this was done under most general balking rules. Some more special balking processes will be investigated in the next chapter.

On the other hand we have assumed a single server queue with negative exponential service times. We will give some generalizations with respect to the number of servers and the service times in the next chapter.

CHAPTER VI

GENERALIZATIONS AND SPECIALIZATIONS

1. INTRODUCTION

In the preceding chapters properties of a single server queue and a stream of customers balking from it were considered. Arrival process was taken to be a semi-Markov process and the balking rule was of the most general type: it depended not only on the queue size but also on the type of customer.

In this chapter we will remark on some generalizations with respect to number of servers and service times, and some special cases with respect to arrival process, queue size allowed, and balking rules.

2. MANY SERVER QUEUEING PROCESSES

Restriction of the number of servers to one in the Chapters IV and V was not necessary at all but it made the presentation of the problem easier because of its simplicity as well as helping the intuition.

Main theorems of Chapters IV and V, Theorems IV.3.1 and V.4.1 remain unchanged under the assumption of many servers. Hence the results of both chapters can be easily generalized for many server queues.

Let the number of servers in the system be r and let the waiting room be of size $N-r \geq 0$, so that, the maximum number of customers allowed into the system is N, $N \geq 1$. Assume all the servers

are identical, they all have negative exponential holding times with the same mean $1/\beta$.

a. Generalizations of the processes of Chapter IV

Then, in Section IV.3, the formulas given for $\gamma_\delta(j, k, x) = \Pr\{S_{n+1}=k \mid S_n=j, Y_n=\delta, X_n=x\}$ do not hold. Instead we have

$$\gamma_0(j,k,x) = \begin{cases} 1 & \text{for } j=0 \\ \gamma_1(j-1,k,x) & \text{otherwise;} \end{cases}$$

$$\gamma_1(j,k,x) = \begin{cases} \binom{j+1}{k} e^{-k\beta x} (1-e^{-\beta x})^{j+1-k} & \text{for } j < r \\ \binom{r}{k} e^{-k\beta x} \left\{ \int_0^x \frac{(r\beta y)^{j-r}}{(j-r)!} (e^{-\beta y}-e^{-\beta x})^{r-k} \, r\mu dy \right\} & \text{for } j \geq r \text{ , and } k < r \\ e^{-r\beta x} \frac{(r\beta x)^{j+1-k}}{(j+1-k)!} & \text{for } j \geq r \text{ , and } k \geq r \text{ .} \end{cases}$$

(See Takàcs [31] for these formulas.)

The matrices $\Gamma_0(x)$ and $\Gamma_1(x)$ now have these above defined quantities as elements. All the other formulas of Section IV.3 hold with no change excepting the ones for $\overline{\Gamma}_0$ and $\overline{\Gamma}_1$ which need to be changed in accordance with the definitions here.

Sections IV.4, IV.5, IV.6 remain unchanged, all the theorems hold without changes. Section IV.7 on waiting times needs to be revised. With r servers we see that a customer's wait is zero if there are less than r customers in the system upon his arrival. Assuming that customers are served on a first come, first served basis, the waiting time of the n-th customer is zero if $S_n < r$, and is equal to the service time of S_n-r+1 customers if $S_n \geq r$. Hence, the distribution of the waiting time of the n-th customer is:

$$W_n(x) = \sum_{j=o}^{r-1} \Pr\{S_n = j\} + \sum_{j=r}^{N-1} \Pr\{S_n=j\} \int_o^x e^{-r\beta y} \frac{(r\beta y)^{j-r}}{(j-r)!} r\beta dy.$$

A limiting distribution exists: as $n \to \infty$, $W_n(x) \to W(x)$ and is given by

$$W(x) = \sum_{i=1}^{M} \{ \sum_{j=o}^{r-1} p(i,j) + \sum_{j=r}^{N-1} p(i,j) \int_o^x \frac{r\beta e^{-r\beta y}(r\beta y)^{j-r}}{(j-r!)} dy \}.$$

Section IV.8 on the length of a busy period no longer holds. In fact, a new meaning for busy period is necessary. If we define the busy period as the period where at least one server is busy at any time, then we are able to modify the section for the general case of r servers. We let all the functions introduced in IV.8 remain as they are defined. We only need to supply new formulas for the functions $F_n(i,j,x)$. We have now,

$$F_o(i,j,x) \begin{cases} = 0 \quad \text{for} \quad j=0 \\ = \int_o^x b(i,j)\, \gamma_o(j,0,y)(1-\Sigma_{\ell=1}^{M} A_{i\ell}(y))\, dy + \\ \quad \int_o^x F_o(h,k;\, x-y)\, dR_1(i,j;\, h,k,y) \qquad \text{for} \quad j > 0; \end{cases}$$

and the formula for $n \geq 1$ is the same as in page 95. We define $R_1(x)$ and $Q_1(x)$ in the same way as in Chapter IV, (note that $\gamma_o(j,k,y)$ and $\gamma_1(j,k,y)$ are different now.) Further, define $F_n(x)$ as the column vector with $F_n(i,j,x)$ as its $jM+1^{th}$ entry as before. Define $K(x)$ so that its $jM+i^{th}$ entry is

$$\int_o^x b(i,j)\, \gamma_o(j,0,y)(1-\Sigma_{\ell=1}^{M} A_{i\ell}(y))\, dy .$$

Then, just as before, we obtain in terms of Laplace-Stieltjes transforms

$$F_0^*(s) = R_1^*(s)\, F_0^*(s) + K^*(s)$$
$$F_n^*(s) = R_1^*(s)\, F_n^*(s) + Q_1^*(s)\, F_{n-1}^*(s) \ , \ n \geq 1$$

which are the same as (IV.8.5). All the rest of the section remain the same with the new meanings attached to the quantities $K^*(s)$, $R_1^*(s)$, $Q_1^*(s)$, etc.

This then is the generalization of Chapter IV to many server case.

Chapter V holds without change except that in the derivations of $R(x)$ and $Q(x)$ the matrices $\Gamma_0(x)$ and $\Gamma_1(x)$ are now different and should be calculated through formulas given in this section. With this change made everything looks exactly as in Chapter V.

A special case of the many server case is interesting. Let $N = r$, i.e., no storage is allowed. Then the resulting problem can be solved by the general method outlined in this chapter. If, however, the probability of balking when the system is not full is zero, then this problem can be looked as r queues in parallel, each one being a single-server queue with no storage, and the overflow of from one queue becoming the arrival stream for the next one.

3. SINGLE SERVER QUEUE WITH GENERAL SERVICE DISTRIBUTION

Let us return to the single-server queue of Chapters IV and V. We have assumed, in those chapters, that the service times are negative exponentially distributed. Next, retain the assumptions

regarding independence of service times, but assume that service times are identically distributed positive random variables with $V(x)$ as their distribution function.

Then, $\{S_n, Z_n, X_n\}$ is no longer equivalent to a semi-Markov process; neither, therefore, is the process $\{\zeta_n, \psi_n, \theta_n\}$ equivalent to a semi-Markov process. Therefore, methods of Chapters IV and V fail to be useful.

We can, however, remedy for this by the addition of a dummy variable U_n denoting the remaining service time on the customer being served at the instant of n-th arrival, T_n. U_n is zero if and only if $S_n = 0$. Further let Υ_k denote the remaining service time on the customer being served at the instant of k-th balking τ_k, i.e., $\Upsilon_k = U_n$ if $\tau_k = T_n$. For convenience we take $U_o = \Upsilon_o = 0$.

Theorem VI.3.1. The process $\{Z_n, S_n, U_n, X_n\}$ is equivalent to a semi-Markov process whose state space is the Cartesian product of the set of non-negative real numbers with the set of integers $\{1,2,\ldots, M(N+1)\}$.

Proof. We shall show that

$$\Pr\{Z_n=j, S_n=k, U_n \le x, X_n \le y \mid Z_k, S_k, U_k, X_k;\ k \le n-1\} =$$
$$\Pr\{Z_n=j, S_n=k, U_n \le x, X_n \le y \mid Z_{n-1}, S_{n-1}, U_{n-1}\}.$$

First note that S_n, U_n depend only on S_{n-1}, U_{n-1}, Y_{n-1}. But Y_{n-1} depends on Z_{n-1} and S_{n-1}. Hence,

$$\Pr\{S_n=k, U_n \le x \mid Z_k, S_k, U_k, X_{k+1};\ k \le n-1\} =$$
$$\Pr\{S_n=k, U_n \le x \mid Z_{k-1}, S_{k-1}, U_{k-1}, X_k\}.$$

Then,

$$\Pr\{Z_n=j,\ S_n=k,\ U_n \le x,\ X_n \le y \mid Z_k,\ S_k,\ U_k,\ X_k;\ k \le n-1\} =$$

$$= \int_0^y \Pr\{S_n=k,\ U_n \le x \mid Z_n=j,\ X_n=w;\ Z_k,\ S_k,\ U_k,\ X_k,\ k \le n-1\}$$
$$d_w \Pr\{Z_n=j,\ X_n \le w \mid Z_k,\ S_k,\ U_k,\ X_k,\ k \le n-1\}$$

$$= \int_0^y \Pr\{S_n=k,\ U_n \le x \mid Z_{n-1},\ S_{n-1},\ U_{n-1},\ X_n=w\}\ d_w \Pr\{Z_n=j,$$
$$X_n \le w \mid Z_{n-1}\}$$

$$= \Pr\{Z_n=j,\ S_n=k,\ U_n \le x, X_n \le y \mid Z_{n-1},\ S_{n-1}, U_{n-1}\}$$

as was to be demonstrated.

We next give a few corollaries without proofs.

Corollary VI.3.1.A. The process $\{Z_n,\ S_n,\ U_n\}$ is a Markov process.

Corollary VI.3.1.B. The process $\{Z_n,\ S_n,\ U_n,\ T_n\}$ is a Markov process.

If one defines the balking process as $\{\phi_n,\ \Theta_n\}$ where $\Theta_n=\tau_n-\tau_{n-1}$ as before, and where $\phi_n = (\zeta_n,\ \psi_n,\ T_n)$, then with a little alteration of proof of Theorem V.4.1, we obtain the next theorem. We omit the proof.

Theorem VI.3.2. The balking process $\{\phi_n,\ \Theta_n\}$ is equivalent to a semi-Markov process, i.e.,

$$\Pr\{\phi_n \in v,\ \Theta_n \in (0,x) \mid \phi_k,\ \Theta_k,\ k \le n-1\} = \Pr\{\phi_n \in v,\ \Theta_n \in (0,x) \mid \phi_{n-1}\}$$

In view of these two theorems, both the queueing properties of the system and the balking process can be investigated by almost the same methods as outlined in Chapters IV and V. However, because of the ugly nature of the state space of the processes of interest, the calculations involved become much too involved. We do not carry out investigation of this general case any further.

4. SINGLE NEGATIVE EXPONENTIAL SERVER WITH SOME SPECIALIZED BALKING RULES

In this section we take up the problem of Chapters IV and V. In those chapters balking rules were taken to be most general, namely, the probability of a customer balking from the queue depended on not only the state of the queue but also on the state of the arrival process. This can be specialized in at least two directions.

a) Probability of balking depends only on the state of the queue; i.e., $\Pr\{Y_n=0 \mid Z_n=i, S_n=j\} = \Pr\{Y_n=0 \mid S_n=j\}$. This case will be encountered in applications where the customers themselves are homogeneous. For this case, proofs of theorems given in Chapters IV and V become much easier. However, we have no new theorems.

b) A further specialization on this would be to set $\Pr\{Y_n=0 \mid Z_n=i, S_n=j\}$ zero if $j < N$ and unity if $j=N$, i.e., customers join as long as there is room for them. In this case, Chapter IV gives the queueing properties of a finite queue with a single negative exponential server, no balking allowed. Of course this is of interest in its own right; and Chapter IV retains its originality even in this case.

Under this type of a balking rule, Chapter V gives the properties of the overflow stream from a single negative exponential server subject to a semi-Markov process input. In this case, the steady state solution becomes quite interesting. From Theorem V.6.1 the only ergodic states are (i,N) $i=1,2,\ldots,M_1$. So that, the overflow stream will have many of the characteristics of the arrival stream, but will differ from it in its interval process. Our research remains original in this special case also. A further specialization of this problem by letting the interarrival process be a simple renewal process will be discussed later.

c) Another specialization of the general problem may be made at the other extreme by letting balking probabilities be dependent on the state of the arrival process only; i.e.,

$$\Pr\{Y_n=0 \mid Z_n=i,\ S_n=j\} = \begin{cases} \Pr\{Y_n=0 \mid Z_n=i\} & \text{if } j < N \\ 1 & \text{if } j = N \ . \end{cases}$$

This case will be quite useful if the customers differ from each other greatly in their needs so that it is advantageous to send only certain types of customers to a given queue.

This special problem can also be solved in a different way by the methods of Chapter III and the special case b above of Chapters IV and V. First, by the method of Chapter III, the arrivals can be partitioned into two streams, one of which would become the arrival stream to the queue of interest. Then by the use of special case b above, things of interest can be obtained.

These then are some of the special cases obtained by restricting the balking rules to certain forms.

5. SINGLE NEGATIVE EXPONENTIAL SERVER SUBJECT TO A RENEWAL INPUT

In this section we take up the interesting simple case where the interarrival times are independent, identically distributed random variables. This is equivalent to a semi-Markov process with only one state, i.e., M=1. Note that in this case dependence of balking rule on the arrival process becomes of a trivial nature.

In this special case, Chapter IV gives the queueing properties of a GI/M/1 queueing system with balking. By Corollary B to Theorem IV.3.1, the process $\{S_n\}$ $(n \geq 0)$ is a Markov chain. This was also noted by Finch [9] who investigated the properties of queue size process for this case. Finch, further has proven that a steady state distribution exists and has given this limiting distribution through generating functions. We in Chapter IV give the solutions for these by a different treatment through the use of matrix theory. Because of different tools used, it is difficult to show that our solutions are the same as given by Finch. In Chapter IV, in this special case, we are not completely original. However our treatment of busy period still remains original.

In Chapter V, through further specializations we obtain some already investigated problems. One such further specialization is obtained by letting the balking probabilities be zero as long as the system is not full. Then we obtain the overflow problem investigated by Disney [6] for a GI/M/1 system with storage, and by Palm [20] for the same system with no storage.

a. Palm's overflow problem

For this case N is set to be 1. Then,

$$\bar{\Gamma}_0 = \begin{bmatrix} \sum_0^\infty \bar{\alpha}_k & 0 \\ \sum_1^\infty \bar{\alpha}_k & \bar{\alpha}_0 \end{bmatrix} = \begin{bmatrix} e^{-\beta\bar{D}} & 0 \\ e^{-\beta\bar{D}} - 1 & 1 \end{bmatrix};$$

$$\bar{\Gamma}_1 = \begin{bmatrix} \sum_1^\infty \bar{\alpha}_k & \bar{\alpha}_0 \\ \sum_2^\infty \bar{\alpha}_k & \bar{\alpha}_1 \end{bmatrix} = \begin{bmatrix} e^{-\beta\bar{D}} - 1 & 1 \\ e^{-\beta\bar{D}} - 1 + \beta\bar{D} & -\beta\bar{D} \end{bmatrix};$$

$$b = \begin{bmatrix} 0 \\ 1 \end{bmatrix}; \qquad D = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix};$$

so that,

$$R^*(s) = \bar{\Gamma}_0 \oplus A^*(s+\beta) = \begin{bmatrix} A^*(s) & 0 \\ A^*(s) - A^*(s+\beta) & A^*(s+\beta) \end{bmatrix}$$

$$Q^*(s) = (I-D)(\bar{\Gamma}_1 \oplus A^*(s+\beta)) = \begin{bmatrix} A^*(s) - A^*(s+\beta) & A^*(s+\beta) \\ 0 & 0 \end{bmatrix}.$$

In this case, by Theorem V.7.1, the times between overflows are independent and identically distributed. This was shown by Palm. Now to get $L(x)$, the distribution function of the time between overflows, we first obtain $B^*(s)$, then $L^*(s) = \pi B^*(s) E$ by the formula (V.7.1).

$$I - Q^*(s) = \begin{bmatrix} 1 - A^*(s) + A^*(s+\beta) & -A^*(s+\beta) \\ 0 & 1 \end{bmatrix},$$

$$(I-Q^*(s))^{-1} = (1-A^*(s)+A^*(s+\beta))^{-1} \begin{bmatrix} 1 & A^*(s+\beta) \\ 0 & 1-A^*(s)+A^*(s+\beta) \end{bmatrix} .$$

Then, from (V.4.2),

$$B^*(s) = R^*(s)(I-Q^*(s))^{-1} D$$

$$= (1-A^*(s)+A^*(s+\beta))^{-1} \begin{bmatrix} 0 & A^*(s)A^*(s+\beta) \\ 0 & A^*(s+\beta) \end{bmatrix}$$

$$B^*(s) = \frac{A^*(s+\beta)}{1-A^*(s)+A^*(s+\beta)} \begin{bmatrix} 0 & A^*(s) \\ 0 & 1 \end{bmatrix} .$$

Since $B = B^*(0) = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$, and $\Pi_0 = p_0 = \begin{bmatrix} 0 & 1 \end{bmatrix}$, then $\Pi_n = \Pi_0 B^n = \begin{bmatrix} 0 & 1 \end{bmatrix}$. Also, clearly, $\Pi = \begin{bmatrix} 0 & 1 \end{bmatrix}$. Thus,

$$L^*(s) = \Pi B^*(s) E = \frac{A^*(s+\beta)}{1-A^*(s)+A^*(s+\beta)} \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & A^*(s) \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$L^*(s) = \frac{A^*(s+\beta)}{1-A^*(s)+A^*(s+\beta)} .$$

In Palm's problem, the overflow stream from the i-th server constitutes the arrival stream for the $i+1^{st}$ server. Then, denoting the Laplace-Stieltjes transform of the interarrival distribution to the i-th server by $A_i^*(s)$, we obtain

$$A_{i+1}^*(s) = \frac{A_i^*(s+\beta)}{1- A_i^*(s)+ A_i^*(s+\beta)} \qquad (i=1,2,\ldots,r) .$$

These are the well known difference equations of Palm. For a more complete treatment of Palm's overflow problem we refer to [6, 16, 20].

b. Disney's overflow problem

Disney [6] has generalized the results of Palm by letting storage in the system. Since $M=M_1=1$, $b(1,j)=0$ for $j < N$, and $\psi_o=S_o=N$, we have, from Theorem V.7.1, that the times between overflows, θ_1, θ_2,..., are independent and identically distributed. This result was proved by Disney qualitatively. But he did not give the distribution of the time between overflows. In the following we are going to derive this distribution.

Since $\Pi_o=p_o=[0...0\ 1]$, $b=[0...0\ 1]^T$, clearly, $\Pi=[0...0\ 1]$. Thus, $L^*(s)$, the Laplace-Stieltjes transform of the distribution function of the time between overflows can be calculated from

$$L^*(s) = \Pi B^*(s)E = \Pi R^*(s)(I-Q^*(s))^{-1}DE = \Pi R^*(s)(I-Q^*(s))^{-1}b$$

which reduces, by the special nature of Π and b , to the inner product of the last row of $R^*(s)$ and the last column of $(I-Q^*(s))^{-1}$. (See also the figure on page 135)

Though this description of the solution should suffice, we next go on to give a formula for $G^*(s)$ in terms of $A^*(s)$. First, we introduce some special notation: let,

$$g_k(s) = \bar{\alpha}_k\, A^*(s+\beta) = \int_0^\infty e^{-sx}\, \frac{e^{-\beta x}(\beta x)^k}{k!}\, dA(x)\ ,\ \mathrm{Re}\{s\} \geq 0\ ;$$

$$\phi(s,z) = \sum_{k=0}^{\infty} z^k g_k(s) , \quad |z| < 1 ;$$

and

$$\Omega(s,z) = \sum_{n=0}^{\infty} z^n \Delta_n(s) , \quad |z| < 1 , \text{ where}$$

$$\Delta_n(s) = \begin{cases} 1 & \text{for } n = 0 \\ \det\ (I - Q_n^*(s)) & \text{for } n \geq 1 , \end{cases} \quad \text{Re}\{s\} \geq 0, \text{ where}$$

$$Q_n^*(s) = \begin{bmatrix} \Sigma_1^{\infty} g_k(s) & g_0(s) & & & & & \\ \Sigma_2^{\infty} g_k(s) & g_1(s) & g_0(s) & & & & \\ \Sigma_3^{\infty} g_k(s) & g_2(s) & g_1(s) & g_0(s) & & & \\ \vdots & \vdots & \vdots & \ddots & \ddots & & \\ \Sigma_{n-2}^{\infty} g_k(s) & g_{n-3}(s) & g_{n-4}(s) & \cdots & g_1(s) & g_0(s) & \\ \Sigma_{n-1}^{\infty} g_k(s) & g_{n-2}(s) & g_{n-3}(s) & \cdots & g_2(s) & g_1(s) & g_0(s) \\ \Sigma_n^{\infty} g_k(s) & g_{n-1}(s) & g_{n-2}(s) & \cdots & g_3(s) & g_2(s) & g_1(s) \end{bmatrix} .$$

First,

$$\phi(s,z) = \sum_{k=0}^{\infty} z^k \int_0^{\infty} e^{-sx} e^{-\beta x} \frac{(\beta x)^k}{k!} dA(x)$$

$$= \int_0^{\infty} e^{-sx} e^{-\beta x} e^{z\beta x} dA(x)$$

$$\phi(s,z) = A^*(s-\beta(1-z)) , \quad |z| < 1 , \quad \text{Re}\{s\} \geq 0 \qquad \text{(VI.5.1)}$$

Next, expanding $\Delta_n(s)$ about the last row we obtain, (we suppress "s" for convenience,)

$$\Delta_n = (-1)^{n+1} (-\Sigma_n^{\infty} g_k)(-g_0)^{n-1} + (-1)^{n+2}(-g_{n-1})(-g_0)^{n-2} \Delta_1 +$$

$$(-1)^{n+3} (-g_{n-2}) (-g_0)^{n-3}\Delta_2 + \cdots + (-1)^{n+n-2} (-g_3)(-g_0)^2 \Delta_{n-3}$$

$$+ (-1)^{n+n-1} (-g_2)(-g_0) \Delta_{n-2} + (-1)^{n+n} (1-g_1)(-g_0)^0 \Delta_{n-1}$$

$$= \Delta_{n-1} - g_0^{-1} \sum_{j=1}^{n} g_j g_0^j \Delta_{n-j} - \sum_{k=n+1}^{\infty} g_k g_0^{k-1}, \quad n \geq 1 .$$

$$R^* = \begin{bmatrix} \Sigma_0^\infty g_k & & & & & & & \\ \Sigma_1^\infty g_k & g_0 & & & & & & \\ \Sigma_2^\infty g_k & g_1 & g_0 & & & & & \\ \Sigma_3^\infty g_k & g_2 & g_1 & g_0 & & & & \\ \vdots & \vdots & \vdots & \vdots & \ddots & & & \\ \Sigma_N^\infty g_k & g_{N-1} & g_{N-2} & g_{N-3} & \cdots & \cdots & \cdots & g_0 \end{bmatrix}$$

$$Q^* = \begin{bmatrix} \Sigma_1^\infty g_k & g_0 & & & & & & \\ \Sigma_2^\infty g_k & g_1 & g_0 & & & & & \\ \Sigma_3^\infty g_k & g_2 & g_1 & g_0 & & & & \\ \vdots & \vdots & \vdots & \ddots & \ddots & & & \\ \Sigma_N^\infty g_k & g_{n-1} & g_{N-2} & \cdots & \cdots & g_1 & & g_0 \\ 0 & 0 & 0 & \cdots & \cdots & 0 & & 0 \end{bmatrix}$$

Figure 5 . Matrices in the Overflow Problem
(The argument "s" is suppressed everywhere.)

Thus,

$$\Omega(s,z) = 1 + \sum_{n=1}^{\infty} \Delta_n z^n$$

$$= 1 + z \sum_{n=1}^{\infty} \Delta_{n-1} z^{n-1} - g_o^{-1} \sum_{n=k}^{\infty} \sum_{j=1}^{n} g_j g_o^j \Delta_{n-j} z^n$$

$$- \sum_{n=1}^{\infty} \sum_{k=n+1}^{\infty} g_k g_o^{k-1} z^n \quad . \qquad \text{(VI.5.2)}$$

Now,

$$\sum_{n=1}^{\infty} \Delta_{n-1} z^{n-1} = \Omega(s,z) \; ; \qquad \text{(VI.5.3)}$$

$$\sum_{n=1}^{\infty} \sum_{j=1}^{n} g_j\, g_o^j\, \Delta_{n-j}\, z^n = \sum_{j=1}^{\infty} \sum_{n=j}^{\infty} g_j (z g_o)^j\, \Delta_{n-j}\, z^{n-j}$$

$$= \sum_{j=1}^{\infty} g_j (z g_o)^j \sum_{k=o}^{\infty} \Delta_k\, z^k$$

$$= \Omega(s,z)(\phi(s, z g_o(s)) - g_o(s)); \qquad \text{(VI.5.4)}$$

and finally,

$$\sum_{n=1}^{\infty} \sum_{j=n+1}^{\infty} g_o^{n-1}\, z^n\, g_j = \sum_{n=1}^{\infty} g_o^{n-1}\, z^n\, (A^* - \sum_{k=o}^{n} g_k)$$

$$= \frac{zA^*}{1-zg_o} - g_o^{-1} \sum_{n=o}^{\infty} g_o^n\, z^n \sum_{k=o}^{n} g_k + 1$$

$$= 1 + \frac{zA^*}{1-zg_o} - g_o^{-1} \sum_{k=o}^{\infty} \sum_{n=k}^{\infty} g_k\, (zg_o)^k\, (zg_o)^{n-k}$$

$$= 1 + \frac{zA^*(s)}{1-zg_o(s)} - \frac{\phi(s, zg_o(s))}{g_o(s)(1-zg_o(s))} \qquad \text{(VI.5.5)}$$

Putting (VI.5.3), (VI.5.4), and (VI.5.5) in (VI.5.2), we obtain

$$\Omega(s,z) = 1 + z\Omega(s,z) - \frac{\phi(s, zg_o(s)) - g_o(s)}{g_o(s)}\, \Omega(s,z) - 1 - \frac{zg_o(s)A^*(s) - \phi(s, zg_o(s))}{g_o(s)(1-zg_o(s))}$$

Introducing $\omega = \omega(s) = zg_o(s)$;

$$\left(1 + \frac{\phi(s,\omega) - g_o - \omega}{g_o}\right)\Omega(s,z) = \frac{\phi(s,\omega) - \omega A^*(s)}{g_o(1-\omega)},$$

Hence,

$$\Omega(s,z) = \frac{\phi(s,\omega) - \omega A^*(s)}{(1-\omega)(\phi(s,\omega) - \omega)}$$

where,

$$\omega = zg_o(s) = zA^*(s+\beta), \qquad \text{(VI.5.6)}$$

and

$$\phi(s,\omega) = A^*(s+\beta(1-\omega))$$
$$= A^*(s+\beta(1-z\ A^*(s+\beta))) \ .$$

$\Omega(s,z)$ has a removable singularity at $z=A^*(s+\beta)^{-1}$, (this corresponds to $\omega=1$.) Further, $\phi(s,\omega)-\omega = A^*(s+\beta(1-\omega))-\omega$ has, by Rouché's theorem, only one root within the unit circle; denote that root by $\omega_o(s)$. Then, $\Omega(s,z)$ has at most one pole within the unit circle, since from $\omega_o(s) = z_o(s)\ A^*(s+\beta)$ we have $z_o(s) = \left|\frac{\omega_o(s)}{A^*(s+\beta)}\right| \geq |\omega_o(s)|$. Therefore, $\Omega(s,z)$ is analytic for all z with $0 < |z| \leq |z_o(s)|$, $\text{Re}\{s\} \geq 0$. Hence, $\Delta_n(s)$ can be obtained as the coefficient of z^n in a series expansion of $\Omega(s,z)$ around the origin with a suitably chosen z .

Next we obtain the last column of $(I-Q^*(s))^{-1}$, (by using the equation $M^{-1} = (\det M)^{-1}\ \text{adj}\ M$ for a non-singular matrix M .) We then have as the last column

$$\frac{1}{\Delta_N}\begin{bmatrix} (-1)^N & (-g_o)^N & \\ (-1)^{N+1} & (-g_o)^{N-1} & \Delta_1 \\ (-1)^{N+2} & (-g_o)^{N+2} & \Delta_2 \\ & \vdots & \\ (-1)^{N+N-1} & (-g_o) & \Delta_{N-1} \\ (-1)^{N+N} & (-g_o)^o & \Delta_N \end{bmatrix} = \frac{1}{\Delta_N}\begin{bmatrix} g_o^N & \Delta_o \\ g_o^{N-1} & \Delta_1 \\ g_o^{N-2} & \Delta_2 \\ \vdots & \\ g_o & \Delta_{N-1} \\ g_o & \Delta_N \end{bmatrix} .$$

The last row of $R^*(s)$ on the other hand is

$$[\Sigma_{k=N}^{\infty} g_k \; g_{N-1} \; g_{N-2} \cdots g_2 \; g_1 \; g_o] \; .$$

Then, the Laplace-Stieltjes transform of the distribution function of the time between overflows, $L^*(s)$, can be obtained by taking the product of the last row of $R^*(s)$ with the last column of $(I-Q^*(s))^{-1}$. Doing this we obtain

$$L^*(s) = \frac{1}{\Delta_N} (\Sigma_N^{\infty} g_k \; g_o^N + g_{N-1} \; g_o^{N-1} \; \Delta_1 + g_{N-2} \; g_o^{N-2} \; \Delta_2 + \cdots +$$

$$g_2 \; g_o^2 \; \Delta_{N-2} + g_1 \; g_o \; \Delta_{N-1} + g_o \; \Delta_N) \; .$$

$$= \frac{g_o}{\Delta_N} (\Delta_{N-1} - \Delta_N + \Delta_N)$$

Hence, since $g_o(s) = A^*(s+\beta)$

$$L^*(s) = \frac{\Delta_{N-1}(s)}{\Delta_N(s)} A^*(s+\beta) \qquad \text{(VI.5.7)}$$

This formula can be used by calculating the determinants $\Delta_{N-1}(s)$ and $\Delta_N(s)$ first. This can be done easily by direct computation if N is small. However, for larger values of N, it would be easier to use the generating function $\Omega(s,z)$ given by (VI.5.6) to calculate Δ_N and Δ_{N-1}.

As an illustration of the use of these we return to Palm's problem, where $N=1$. Then, $L^*(s) = \frac{\Delta_o(s)}{\Delta_1(s)} A^*(s+\beta)$. Now,

$$\Omega(s,z) = \frac{1}{1-zA^*(s+\beta)} \left[1 + \frac{zA^*(s+\beta)(1-A^*(s))}{A^*(s+\beta-\beta zA^*(s+\beta))-zA^*(s+\beta)} \right] .$$

$\Delta_0(s) = \Omega(s,0) = 1$, and $\Delta_1(s) = \left.\frac{d\Omega(s,z)}{dz}\right|_{z=0}$.

$$\frac{d\Omega(s,z)}{dz} = \frac{A^*(s+\beta)}{(1-zA^*(s+\beta))^2}\left\{1+\frac{zA^*(s+\beta)(1-A^*(s))}{A^*(s+\beta-\beta zA^*(s+\beta))-zA^*(s+\beta)}\right\} +$$

$$\frac{1}{(1-zA^*(s+\beta))}\left\{\frac{A^*(s+\beta)(1-A(s))}{A^*(s+\beta-\beta zA^*(s+\beta))-zA^*(s+\beta)} - \right.$$

$$\left.\frac{zA^*(s+\beta)(1-A^*(s))\left[\frac{d}{dz}A^*(s+\beta-\beta zA^*(s+\beta))-A^*(s+\beta)\right]}{[A^*(s+\beta-\beta zA^*(s+\beta))-zA^*(s+\beta)]^2}\right\};$$

so that,

$$\Delta_1(s) = \left.\frac{d\Omega(s,z)}{dz}\right|_{z=0} = A^*(s+\beta) + \frac{A^*(s+\beta)(1-A^*(s))}{A^*(s+\beta)}$$

$$= 1 - A^*(s) + A^*(s+\beta) .$$

Hence we have,

$$L^*(s) = \frac{\Delta_0(s)A^*(s+\beta)}{\Delta_1(s)} = \frac{A^*(s+\beta)}{1-A^*(s) + A^*(s+\beta)}$$

once again.

6. CONCLUSION

In this chapter we have given some generalizations of Chapter IV and V as well as some more important special cases. This completes our treatment of the decomposition of streams under queue dependent decision rules.

PART III

APPLICATIONS

CHAPTER VII

A SIMPLE SYSTEM: AN EXAMPLE

1. INTRODUCTION

In this last chapter we will illustrate the use of the analytical methods developed in the thesis by an example. We choose a simple system as shown below in Figure 6. Customers arriving into the system at D_1 form a recurrent process. The decision rule at D_1 is of the type discussed in Chapter II, namely, the sequence of outcomes of decisions forms a Markov chain with two states. One of the resulting two streams becomes the arrival stream to a single negative exponential server at Q_1. Customers arriving at D_2 either join the queue or balk with probabilities depending upon the number of people already there. Those who do not join Q_1 form the arrival stream to D_3. Decisions at D_3 are made on the basis of customers' experiences, as in Chapter III.

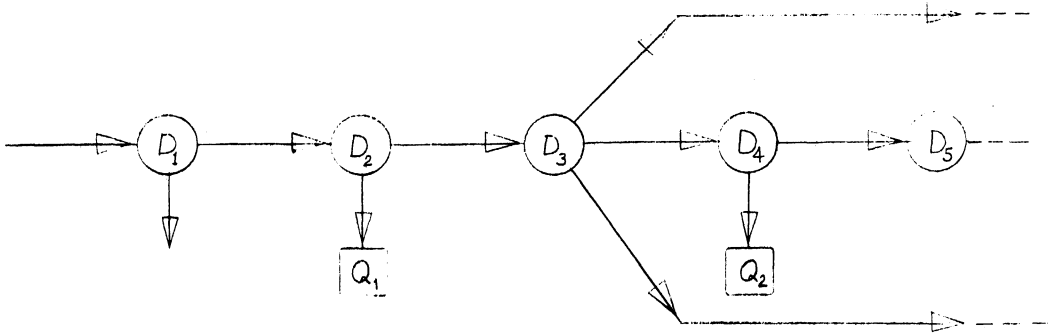

Figure 6. A Simple System.

We will only be interested in the stream which constitutes the arrival stream at Q_2 ; we will not consider the queueing properties of Q_2 but will give the overflow stream from Q_2 .

2. DEFINITIONS AND NOTATION

We will denote the distribution function, (single or matrix valued,) of the interarrival times at the k-th decision point by $A_k(x)$, $k = 1,2,3,4,5$. We take the arrivals into the system to be a recurrent process with the time between arrivals distributed as $A_1(x)$.

We assume the decision process at D_1 , the process $\{Y_n^1\}$ is a Markov-chain with two states with the initial distribution vector $[1 \quad 0]$ and the transition matrix $Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix}$, $Q_{ij} > 0$.

We take the server of Q_1 to have negative exponentially distributed holding times with mean $1/\beta$. There is no storage allowed in Q_1 so that the maximum number allowed in Q_1 is $N = 1$. Customers arriving at D_2 balk Q_1 with probability c (< 1) if the server is idle and with probability 1 if the server is busy. Hence the balking vector is $b = \begin{bmatrix} c \\ 1 \end{bmatrix}$, and $D = \begin{bmatrix} c & o \\ o & 1 \end{bmatrix}$.

The customers arriving at D_3 will be of two types: those who have balked Q_2 when it was empty, and those who balked when it was full. Assume only the customers of second type are allowed in the direction of D_4 , and that with probability d (d is positive.) Hence the vector q is given as $\begin{bmatrix} o \\ d \end{bmatrix}$.

Since only one type of customer is permitted to arrive at D_4 , the arrival stream to D_4 will be a recurrent process. We assume

the maximum number allowed in Q_2 is 2, and only those customers who find the system full are turned away, so that, the arrivals to D_5 are the overflows from Q_2.

3. DECOMPOSITION AT D_1 : ARRIVALS INTO Q_1

The decision process at D_1 is a Markov chain defined by $[1\ 0]$ and Q. The event E_1 corresponds to a customer being sent to D_2, thus we only care about the occurrences of E_1.

The arrivals at D_1 form a recurrent process, i.e., a semi-Markov process with only one state. Since the decomposed streams have the same state space as the original stream, (cf. Theorem II.3.1,) the stream arriving at D_2 will have only one state also, i.e., it will be a recurrent process. Hence the distribution of the time between arrivals to D_2, $A_2(x)$, will be given, by (II.3.1), by its Laplace-Stieltjes transform as

$$A_2^*(s) = f_{11}(A_1^*(s))$$

where $f_{11}(z)$ is the generating function of distribution of the first passage time from state E_1 to E_1 in the chain $\{Y_n^1\}$.

From Theorem II.3.3, the generating function $f_{11}(z)=1-1/q_{11}(z)$ where $q_{11}(z)$ is the $(1,1)$ entry of $(I-zQ)^{-1}$, $|z| < 1$. Now,

$$I-zQ = \begin{bmatrix} 1-zQ_{11} & -zQ_{12} \\ -zQ_{21} & 1-zQ_{22} \end{bmatrix},$$

thus,

$$(I-zQ)^{-1} = \frac{1}{(1-zQ_{11})(1-zQ_{22}) - Q_{12}\,Q_{21}\,z^2} \begin{bmatrix} 1-zQ_{22} & zQ_{12} \\ zQ_{21} & 1-zQ_{11} \end{bmatrix} ,$$

so that,

$$q_{11}(z) = \frac{1-zQ_{22}}{1 - (Q_{11}+Q_{22})z + (Q_{11}Q_{22}-Q_{12}Q_{21})z^2}$$

$$= \frac{1-zQ_{22}}{1 - (Q_{11}+Q_{22})z - (1-Q_{11}-Q_{22})z^2} .$$

Then, we obtain

$$f_{11}(z) = 1 - \frac{1}{q_{11}(z)} = \frac{Q_{11}z + (1-Q_{11}-Q_{22})z^2}{1-zQ_{22}} ,$$

and hence,

$$A_2^*(s) = \frac{Q_{11}A_1^*(s) + (1-Q_{11}-Q_{22})\,A_1^*(s)^2}{1 - Q_{22}A_1^*(s)} . \qquad \text{(VII.3.1)}$$

Let the expected interarrival time at D_1 and D_2 be μ_1 and μ_2 respectively. Then, by (II.5.4), we have $\mu_2 = \lambda_1 \mu_1$. Now,

$$\lambda_1 = \frac{d}{dz} f_{11}(z)\Big|_{z=1}$$

$$= \frac{(Q_{11}+2(1-Q_{11}-Q_{22}))(1-Q_{22}) + Q_{22}(Q_{11}+1-Q_{11}-Q_{22})}{(1-Q_{22})^2}$$

$$= \frac{2 - Q_{11} - Q_{22}}{1 - Q_{22}}$$

$$= 1 + \frac{Q_{12}}{Q_{21}} .$$

Thus,

$$\mu_2 = (1 + \frac{Q_{12}}{Q_{21}})\,\mu_1 . \qquad \text{(VII.3.2)}$$

4. QUEUEING SYSTEM Q_2

Now that we have the arrival process for Q_2 completely indentified, we turn to an investigation of its properties. Since $N=1$, the only states of the system are 0 and 1. Now, from equations in page 83 we have

$$\overline{\Gamma}_o = \begin{bmatrix} e^{-\beta\overline{D}} & 0 \\ e^{-\beta\overline{D}}-1 & 1 \end{bmatrix}, \quad \overline{\Gamma}_1 = \begin{bmatrix} e^{-\beta\overline{D}}-1 & 1 \\ e^{-\beta\overline{D}}-1+\beta\overline{D} & \beta\overline{D} \end{bmatrix};$$

and we are given $b = \begin{bmatrix} c \\ 1 \end{bmatrix}$, and hence $D = \begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix}$.

Then,

$$P^*(s) = D(\overline{\Gamma}_o \oplus A_2^*(s+\beta)) + (I-D)(\overline{\Gamma}_1 \oplus A_2^*(s+\beta))$$

$$= (D\overline{\Gamma}_o + (I-D)\overline{\Gamma}_1)\, A_2^*(s+\beta)$$

$$= \begin{bmatrix} c\, e^{-\beta\overline{D}} + (1-c)(e^{-\beta\overline{D}}-1) & 1-c \\ e^{-\beta\overline{D}} - 1 & 1 \end{bmatrix} A_2^*(s+\beta)$$

$$= \begin{bmatrix} e^{-\beta\overline{D}} - 1 + c & 1-c \\ e^{-\beta\overline{D}} - 1 & 1 \end{bmatrix} A_2^*(s+\beta)$$

$$= \begin{bmatrix} A_2^*(s) - (1-c)\, A_2^*(s+\beta) & (1-c)A_2^*(s+\beta) \\ A_2^*(s) - A_2^*(s+\beta) & A_2^*(s+\beta) \end{bmatrix}.$$

Clearly, the process $\{S_n\}$, (where S_n is the queue size at the n-th arrival epoch,) is a Markov chain (cf. Corollary IV.3.1.B.) The transition matrix P for $\{S_n\}$, then, is

$$P = P^*(0) = \begin{bmatrix} 1 - (1-c)\, A_2^*(\beta) & (1-c)A_2^*(\beta) \\ 1 - A_2^*(\beta) & A_2^*(\beta) \end{bmatrix}.$$

Let $c' = 1-c$, and $v = A_2^*(\beta)$. Then,

$$p = \begin{bmatrix} 1 - c'v & c'v \\ 1-v & v \end{bmatrix}. \qquad \text{(VII.4.1)}$$

a. Distribution of queue size

Assuming that Q_1 was empty initially, i.e., $p_o = [1 \;\; 0]$, we find the distribution of the queue size at the n-th arrival epoch from $p_n = p_o P^n$ as follows. Let $p(z) = \Sigma_{n=o}^{\infty} p_n z^n$ for $|z| < 1$. Then,

$$p(z) = \Sigma_{n=o}^{\infty} p_o P^n z^n = p_o \Sigma_{n=o}^{\infty} (zP)^n .$$

The matrix zP is dominated by $|z|\, P$ whose rows all sum to $|z| < 1$; thus the series $\Sigma_{n=o}^{\infty} (zP)^n$ converges to $(I-zP)^{-1}$. We, thus, have

$$p(z) = p_o (I - zP)^{-1} .$$

Now,

$$(I-zP)^{-1} = \begin{bmatrix} 1 - z(1-c'v) & -c'vz \\ -(1-v)z & 1-vz \end{bmatrix}^{-1}$$

$$= \frac{1}{(1-z)(1-cvz)} \begin{bmatrix} 1-vz & c'vz \\ (1-v)z & 1-z(1-c'v) \end{bmatrix}.$$

Hence,

$$p(z) = [1 \;\; 0]\, (I-zP)^{-1}$$

$$p(z) = \left[\frac{1 - vz}{(1-z)(1-cvz)} \qquad \frac{c'vz}{(1-z)(1-cvz)} \right] .$$

From a power series expansions of the entries of $p(z)$ the probabilities $p_n(0)$ and $p_n(1)$ can be obtained as the coefficients of z^n. Now, expanding $\frac{1-vz}{(1-z)(1-cvz)}$ we get

$$\frac{1 - vz}{(1-z)(1-cvz)} = \frac{1-v}{1-cv} \cdot \frac{1}{1-z} + \frac{c'v}{1-cv} \cdot \frac{1}{1-cvz}$$

for $|z| < 1$, $|cvz| < 1$ also since $c < 1$ and $v < 1$; thus,

$$\frac{1-v}{1-cv}\,\frac{1}{1-z} = \frac{1-v}{1-cv}\,(1 + z + z^2 + z^3 + \ldots)$$

$$\frac{c'v}{1-cv}\,\frac{1}{1-cvz} = \frac{c'v}{1-cv}\,(1+cvz+(cv)^2z^2 + (cv)^3z^3+ \ldots) .$$

Hence,

$$\frac{1 - vz}{(1-z)(1-cvz)} = \sum_{n=o}^{\infty} \frac{1-v+c'v(cv)^n}{1 - cv}\, z^n .$$

Thus we have

$$p_n(0) = \frac{1-v+c'v(cv)^n}{1 - cv}$$

$$p_n(1) = 1 - p_n(0) = \frac{c'v(1-(cv)^n)}{1 -cv} . \qquad \text{(VII.4.2)}$$

The matrix P is positive, hence both of the states are ergodic and form one class as they should be in view of Theorem IV.5.5 since $M_1 = M = 1$ and $N_1 = N = 1$. The distribution of the queue size in the steady state can be obtained from $pP = p$ and $pE = 1$ uniquely. Doing this we obtain

$$p = [\frac{1-v}{1-cv} \qquad \frac{c'v}{1- cv}] . \qquad \text{(VII.4.3)}$$

The same could have been obtained by taking limits as n approaches infinity of the Equations (VII.4.2) above: since $c < 1$, $v < 1$ we have $cv < 1$, and thus $(cv)^n \to 0$ as $n \to \infty$; hence, $p_n(0) \to \frac{1-v}{1-cv}$ and $p_n(1) \to \frac{c'v}{1-cv}$.

b. Length of the busy period

Since the arrival process is a recurrent process, the busy period is the busy period started by a customer of type 1. Hence the Laplace-Stieltjes transform of the distribution of the busy period is the first entry of the 2×1 column vector $F^*(s)$ where

$$F^*(s) = (I - R_1^*(s) - Q_1^*(s))^{-1} K^*(s)$$

with R_1^*, Q_1^*, K^* as defined in Section IV.8. Now;

$$R_1^*(s) = \begin{bmatrix} 0 & c \cdot 0 \\ 0 & 1 \cdot \bar{\alpha}_o \end{bmatrix} A_2^*(s+\beta) = \begin{bmatrix} 0 & 0 \\ 0 & A_2^*(s+\beta) \end{bmatrix}$$

$$Q_1^*(s) = \begin{bmatrix} 0 & (1-c)\bar{\alpha}_o \\ 0 & 0 \cdot \alpha_1 \end{bmatrix} A_2^*(s+\beta) = \begin{bmatrix} 0 & c' A_2^*(s+\beta) \\ 0 & 0 \end{bmatrix}$$

And from,

$$K(x) = \begin{bmatrix} \int_0^x c\beta e^{-\beta y} (1 - A_2(y)) dy \\ \int_0^x \beta^2 e^{-\beta y} y (1 - A_2(y)) dy \end{bmatrix},$$

we get,

$$K^*(s) = \begin{bmatrix} \beta c \int_0^\infty e^{-sx}\, e^{-\beta x}\, (1-A_2(x))\, dx \\ \beta^2 \int_0^\infty e^{-sx}\, e^{-\beta x}\, x\, (1-A_2(x))\, dx \end{bmatrix}$$

$$= \begin{bmatrix} \beta c \int_0^\infty e^{-ux}\, (1-A(x))dx \Big|_{u=s+\beta} \\ -\beta^2 \frac{d}{du} \int_0^\infty e^{-ux}\, (1-A_2(x))dx \Big|_{u=s+\beta} \end{bmatrix}$$

$$K^*(s) = \begin{bmatrix} \beta c\, (1-A_2^*(s+\beta))/(s+\beta) \\ \beta^2\, (1-A_2^*(s+\beta) + (s+\beta) \frac{d}{ds} A_2^*(s+\beta))/(s+\beta)^2 \end{bmatrix} .$$

Now,

$$(I-R_1^*(s)-Q_1^*(s))^{-1} = \begin{bmatrix} 1 & -c'A_2^*(s+\beta) \\ 0 & 1 - A_2^*(s+\beta) \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 1 & c'A_2^*(s+\beta)/(1-A_2^*(s+\beta)) \\ 0 & 1/(1-A_2^*(s+\beta)) \end{bmatrix} ;$$

so that, the first entry of $F^*(s) = (I-R_1^*(s)-Q_1^*(s))^{-1} K^*(s)$ becomes, (by letting $s+\beta = u$ and $A_2^*(s+\beta) = w$,)

$$\frac{\beta c(1-w)}{u} \quad \frac{c'w}{1-w} \quad \frac{\beta^2(1-w+u\frac{dw}{du})}{u^2} = \frac{1}{u}\,(\beta c(1-w) - c'w \frac{d}{du} \log \frac{w-1}{u}) .$$

Hence we have as the Laplace-Stieltjes transform of the distribution of a busy period:

$$H^*(s) = \frac{1}{s+\beta} \left(\beta c\, (1-A_2^*(s+\beta)) - c'A_2^*(s+\beta) \frac{d}{ds} \log \frac{A_2^*(s+\beta)-1}{s+\beta} \right) .$$

This completes our treatment of the properties of Q_2 ; if the properties of this queue is desired in terms of the original arrival stream, that can be accomplished by replacing $A_s^*(s)$ by its equivalent in terms of $A_1^*(s)$ as given by (VII.3.1).

5. BALKING PROCESS FROM Q_2 : ARRIVALS AT D_3

The customers arriving at D_3 are those who have balked from the first queue. Thus by the use of the methods of Chapter V, we can obtain the arrival process at D_3 as the balking process from Q_1 . We now have

$$R^*(s) = \overline{\Gamma}_o \oplus A_2^*(s+\beta) = \begin{bmatrix} A_2^*(s) & 0 \\ A_2^*(s)-A_2^*(s+\beta) & A_2^*(s+\beta) \end{bmatrix},$$

and,

$$Q^*(s) = (I-D)\overline{\Gamma}_1 \oplus A^*(s+\beta) = \begin{bmatrix} c'(A_2^*(s)-A_2^*(s+\beta)) & c'A_2^*(s+\beta) \\ 0 & 0 \end{bmatrix}.$$

Let $u = A_2^*(s)$, and $w = A_2^*(s+\beta)$; then

$$I - Q^*(s) = \begin{bmatrix} 1-c'(u-w) & -c'w \\ 0 & 1 \end{bmatrix},$$

so that

$$I-Q^*(s))^{-1} = \begin{bmatrix} (1-c'(u-w))^{-1} & c'w(1-c'(u-w))^{-1} \\ 0 & 1 \end{bmatrix}.$$

Hence, from (V.4.2), we have as the matrix of balking process:

$$A_3^*(s) = R^*(s)(I-Q^*(s))^{-1}D$$

$$= \begin{bmatrix} u & 0 \\ u-w & w \end{bmatrix} \frac{1}{1-c'(u-w)} \begin{bmatrix} 1 & c'w \\ 0 & 1-c'(u-w) \end{bmatrix} \begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix}$$

$$= \frac{1}{1-c'(u-w)} \begin{bmatrix} u & 0 \\ u-w & w \end{bmatrix} \begin{bmatrix} c & c'w \\ 0 & 1-c'(u-w) \end{bmatrix}$$

$$= \frac{1}{1-c'(u-w)} \begin{bmatrix} cu & c'uw \\ c(u-w) & w \end{bmatrix} ,$$

i.e.,

$$A_3^*(s) = \frac{1}{1-c'A_2^*(s) + c'A_2^*(s+\beta)} \begin{bmatrix} cA_2^*(s) & c'A_2^*(s)\, A_2^*(s+\beta) \\ cA_2^*(s)-cA_2^*(s+\beta) & A_2^*(s+\beta) \end{bmatrix} . \qquad \text{(VII.5.1)}$$

Thus, the arrivals to D_3 form a semi-Markov process with two states, both of which are ergodic and in the same class.

To find the distribution of the type of an arrival in the steady state we solve $\Pi A_3 = \Pi$, $\Pi E = 1$ for Π. Letting $v = A_2^*(\beta)$,

$$A_3 = \frac{1}{c + c'v} \begin{bmatrix} c & c'v \\ c(1-v) & v \end{bmatrix} ; \qquad \text{(VII.5.2)}$$

hence $\Pi A_3 = \Pi$ gives:

$$\frac{v}{c + c'v} (c'\pi(0) + \pi(1)) = 1$$

and, since $\pi(0) + \pi(1) = \Pi E = 1$,

$$\frac{v}{c + c'v} (c' + c\pi(1)) = \pi(1) .$$

Hence;

$$\pi(0) = \frac{c(1-v)}{c+(1-2c)v} \quad , \qquad \pi(1) = \frac{vc'}{c+(1-2c)v} \quad .$$

Note that $\Pi = \frac{1}{pb}\, pD = \frac{1 - cv}{(1-v)c + c'v} \begin{bmatrix} \frac{(1-v)c}{1-cv} & \frac{vc'}{1-cv} \end{bmatrix}$

$$= \begin{bmatrix} \frac{c(1-v)}{c + (1-2c)v} & \frac{vc'}{c + (1-2c)v} \end{bmatrix}$$

as needed in order to satisfy the Theorem V.5.4.

Let μ_3 denote the 2×1 vector of expected interarrival times at D_3 started by a customer who balked Q_1 when it was empty, and by a customer who balked when it was busy. Then, from Equation (V.6.1) we have

$$\mu_3 = (I - A_3 + R(I-Q^*(0))^{-1})(E_2 \oplus \mu_2)$$

where μ_2, to remind, is the mean interarrival time at D_2. Now,

$$R(I-Q)^{-1} = \frac{1}{c + c'v} \begin{bmatrix} 1 & 0 \\ 1-v & v \end{bmatrix} \begin{bmatrix} 1 & c'v \\ 0 & c+c'v \end{bmatrix}$$

$$= \frac{1}{c + c'v} \begin{bmatrix} 1 & c'v \\ 1-v & v \end{bmatrix} \quad ;$$

thus,

$$(I-A_3+R(I-Q)^{-1}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \frac{1}{c + c'v} \begin{bmatrix} c' & 0 \\ c'(1-v) & 0 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1+c'v}{c+c'v} & 0 \\ \frac{c'(1-v)}{c+c'v} & 1 \end{bmatrix} \quad ;$$

so that,

$$\mu_3 = \frac{1}{c+c'v} \begin{bmatrix} 1+c'v & 0 \\ c'(1-v) & c+c'v \end{bmatrix} \begin{bmatrix} \mu_2 \\ \mu_2 \end{bmatrix}$$

$$\mu_3 = \frac{\mu_2}{c+c'v} \begin{bmatrix} 1+c'v \\ 1 \end{bmatrix}. \qquad \text{(VII.5.3)}$$

6. DECOMPOSITION AT D_3 : ARRIVALS INTO Q_2

Arrivals into Q_2 constitute one of the streams resulting from the decomposition effected at D_3. Assume that we start with an instant of arrival at D_4 as the origin, i.e., $0 = T_o = \tau_1$. Arrival process at D_3 has two states both of which are ergodic, i.e., $M_1 = M = 2$. And the vector of conditional probabilities of an assignment is $q = [0 \quad d]^T$. From Theorem III.5.2 we immediately have that the first state is empty, (since $q_1 = 0$,) and the second state is ergodic. Hence the arrivals into Q_2 will form a semi-Markov process with only one state, i.e., a recurrent process. Now, to find the distribution function of the interarrival times we use the formula (III.3.1). We have

$$A_3^*(s) = \frac{1}{1-c'(v-w)} \begin{bmatrix} cu & c'uw \\ c(u-w) & w \end{bmatrix}$$

where $u = A_2^*(s)$, and $w = A_2^*(s+\beta)$, and $c' = 1-c$. Then, further letting $d' = d$;

$$Q^*(s) = (I-D)A_3^*(s) = \frac{1}{1-c'(v-w)} \begin{bmatrix} cu & c'uw \\ d'c(u-w) & d'w \end{bmatrix},$$

so that,

$$(I-Q^*(s))^{-1} =$$

$$= \frac{1-c'u+c'w}{(1-u+c'w)(1-c'u+c'w-d'w)-cc'd'uw(u-w)} \begin{bmatrix} 1-c'u+c'w-d'w & c'uw \\ cd'(u-w) & 1-u+c'w \end{bmatrix}$$

hence, letting the denominator of the terms in $(I-Q^*(s))$ be denoted by Δ , we get

$$B^*(s) = A_3^*(s)(I-Q^*(s))^{-1}D$$

$$= \frac{1}{\Delta} \begin{bmatrix} cu & c'uw \\ c(u-w) & w \end{bmatrix} \begin{bmatrix} 1-c'u+c'w-d'w & c'uw \\ cd'(u-w) & 1-u+c'w \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & d \end{bmatrix}$$

$$= \frac{1}{\Delta} \begin{bmatrix} 0 & c'uvwd(cu+1-u+c'w \\ 0 & wd(cc'u(u-w)+1-u+c'w) \end{bmatrix} .$$

Hence we have

$$A_4^*(s) = \frac{dw(1-u+c'w+cc'u(u-w))}{(1-u+c'w)(1-c'u+c'w-d'w) - cc'dvw(v-w)} \qquad \text{(VII.6.1)}$$

where

$$u = A_2^*(s), \quad w = A_2^*(s+\beta), \quad c' = 1-c, \quad \text{and} \quad d' = 1-d .$$

7. OVERFLOW STREAM FROM Q_2 : ARRIVALS AT D_5

Above we have given the distribution function, (or rather its Laplace-Stieltjes transform,) of the time between arrivals into Q_2 . This second queue is a finite queue with a negative exponential server with mean service time $1/\sigma$, subject to a recurrent input. The queueing

properties of such a system have been examined extensively in the already existing literature; we refer to Saaty [26] and Takàcs [31], and their lists of references.

We shall not examine the queueing processes in Q_2 ; we are only interested in the overflow stream from that queue. We have, from Theorem V.7.1, that the overflow stream is a recurrent one, i.e., the times between overflows form a renewal process. We shall, here, derive the distribution function $A_5(x)$ of the time between overflows by using the formulas developed in Chapter VI for this special case.

From Formula (VI.5.7), we have for the Laplacee-Stieltjes transform of $A_5(x)$

$$A_5^*(s) = \frac{\Delta_1}{\Delta_2} A_4^*(s+\sigma)$$

where,

$$\Delta_1 = 1 - \sum_{k=1}^{\infty} g_k(s) \quad , \quad \Delta_2 = \det \begin{bmatrix} 1 - \Sigma_1^{\infty} g_k(s) & -g_o(s) \\ - \Sigma_2^{\infty} g_k(s) & 1-g_1(s) \end{bmatrix} ,$$

where

$$g_k(s) = \frac{(-\sigma \overline{D})^k}{k!} A^*(s+\sigma) .$$

Now,

$$g_o(s) = A_4^*(s+\sigma)$$

$$g_1(s) = - \sigma \frac{d}{ds} A_4^*(s+\sigma)$$

$$\Sigma_o^{\infty} g_k(s) = e^{-\sigma \overline{D}} A_4^*(s+\sigma) = A_4^*(s) .$$

Thus,

$$\Delta_1 = 1 - (\Sigma_o^\infty g_k(s) - g_o(s)) = 1 - A_4^*(s) + A_4^*(s+\sigma) ,$$

and

$$\Delta_2 = \det \begin{bmatrix} 1-A_4^*(s) + A_4^*(s+\sigma) & - A_4^*(s+\sigma) \\ -(A_4^*(s) - A_4^*(s+\sigma) + \sigma \frac{d}{ds} A_4^*(s[\sigma)) & 1 + \sigma \frac{d}{ds} A_4^*(s+\sigma) \end{bmatrix}$$

$$= (1-A_4^*(s)+A_4^*(s+\sigma))(1+\sigma \frac{d}{ds} A_4^*(s+\sigma))-A_4^*(s+\sigma)(A_4^*(s)-A_4^*(s+\sigma)+\sigma \frac{d}{ds} A_4^*(s+\sigma))$$

$$\Delta_2 = (1-A_4^*(s)+A_4^*(s+\sigma))A_4^*(s+\sigma) + (1-A_4^*(s))(1+\sigma \frac{d}{ds} A_4^*(s+\sigma)) .$$

Hence,

$$A_5^*(s) = \frac{(1-A_4^*(s)+A_4^*(s+\sigma))A_4^*(s+\sigma)}{(1-A_4^*(s)+A_4^*(s+\sigma))A_4^*(s+\sigma) + (1-A_4^*(s))(1+\sigma \frac{d}{ds} A_4^*(s+\sigma))}$$

$$A_5^*(s) = [1 + (1+\sigma \frac{d}{ds} A_4^*(s+\sigma))(A_4^*(s+\sigma)^{-1} - (1-A_4^*(s)+A_4^*(s+\sigma))^{-1}]^{-1} \quad \text{(VII.7.1)}$$

If one wants to express (VII.7.1) in terms of the original interarrival distribution, it can be done by first expressing $A_4^*(s)$ in terms of $A_2^*(s)$ as given in (VII.6.1) and then replacing $A_2^*(s)$ by its equivalent in terms of $A_1^*(s)$ as given in (VII.3.1).

8. CONCLUSION

We have taken a very simple system and illustrated the step by step analysis of the system by the methods and tools developed in this thesis. Although the computations involved are tedious (if done by

hand of course,) they require basically simple operations. Happily we live in the age of computers, and there exist programming languages that can handle the calculations involved quite efficiently. The fact that we only need the results of the last step to go on to the next step is helpful in reducing the size of the problem. Further, if one is interested only in steady state solutions, all transient states created in the course of decompositions can be discarded, as is done in Section 6 of this chapter, thus reducing the size of the matrices involved.

BIBLIOGRAPHY

1. Beneš, V. E. "Heuristic Remarks and Mathematical Problems Regarding the Theory of Connecting Systems," Bell System Tech. J. vol. 41 (1963) 1701-1748.

2. _____. "Algebraic Properties of Connecting Networks," Bell System Tech. J. vol. 42 (1963) 567-607.

3. Berge, C. Théorie des Graphes et ses Applications. 2nd ed. Paris: Dunod, 1963.

4. Burke, P. J. "The Output of a Queueing System," Opr. Res. vol.4 (1956) 699-704.

5. Chung, K. L. Markov Chains with Stationary Transition Probabilities. Berlin: Springer-Verlag, 1960.

6. Disney, R. L. "Some Problems in the Theory of Conveyors and Their Analysis by the Method of Decomposition of Queueing Networks." Doctoral dissertation, Johns Hopkins University, 1964.

7. Doob, J. L. Stochastic Processes. New York: Wiley, 1953.

8. Feller, W. An Introduction to Probability Theory and Its Applications. 2nd ed. vol. 1. New York: Wiley, 1957.

9. Finch, P. D. "Balking in the Queueing System GI/M/1," Acta Math. Acad. Sci. Hung. vol. 10 (1959) 241-247.

10. Ghosal, A. "Queues in Series," J. Roy. Stat. Soc. Ser. B, vol. 24 (1962) 359-363.

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

12. Jackson, J. R. "Networks of Waiting Lines," Opr. Res. vol.5 (1957) 518-521.

13. _____. "Jobshop-Like Queueing Systems," Mgt. Sci. vol. 10 (1963) 131-142.

14. Jackson, R. R. P. "Queueing Processes with Phase Type Service," J. Roy. Stat. Soc. Ser. B, vol. 18 (1956) 129-132.

15. Kemperman, J. H. B. The Passage Problem for a Stationary Markov Chain. Chicago: University of Chicago Press, 1961.

16. Khinchine, A. Y. Mathematical Methods in the Theory of Queueing. London: Griffin, 1960.

17. Kingman, J. F. C. "Two Similar Queues in Parallel," Ann. Math. Stat. vol. 32 (1961) 1314-1323.

18. Lévy, P. "Processus Semi-Markoviens," Proc. Int. Cong. Math. (Amsterdam, 1954), vol. 3, 416-426.

19. Mirsky, L. An Introduction to Linear Algebra. Oxford: Oxford University Press, 1963.

20. Palm, C. "Intensitätschwankungen im Fernsprechverkehr," Ericsson Technics, vol. 44 (1943) 1-189.

21. Patterson, R. L. "Some Analytical Methods in the Study of n-Stage Stochastic Service Systems with Applications to the Optimization Problems." Doctoral dissertation, University of Michigan, 1963.

22. Pyke, R. "Markov Renewal Processes: Definitions and Preliminary Properties," Ann. Math. Stat. vol. 32 (1961) 1231-1242.

23. _____. "Markov Renewal Processes with Finitely Many States," Ann. Math. Stat. vol. 32 (1961) 1243-1259.

24. Reich, E. "Waiting Times When Queues are in Tandem," Ann. Math. Stat. vol. 28 (1957) 768-773.

25. Rosenblatt, M. Random Processes. New York: Oxford University Press, 1962.

26. Saaty, T. L. Elements of Queueing Theory with Applications. New York: McGraw-Hill, 1961.

27. _____. "Stochastic Network Flows." Paper read at the Symposium on Congestion Theory, Chapel Hill, August 1964. (Mimeographed.)

28. Smith, W. L. "Regenerative Stochastic Processes," Proc. Roy. Soc. (London), Ser. A, vol. 232 (1955) 6-31.

29. _____. "Renewal Theory and Its Ramifications," J. Roy. Stat. Soc. Ser. B, vol. 20 (1958) 243-302.

30. Syski, R. "Congestion in Telephone Exchanges," in Information Theory, ed. E. C. Charry (New York: Butterworth, 1961), 85-98.

31. Takàcs, L. Introduction to the Theory of Queues. New York: Oxford University Press, 1962.

32. Thrall, R. M., and Tornheim, L. Vector Spaces and Matrices. New York: Wiley, 1957.