THE UNIVERSITY OF CALGARY

THE BUSY PERIOD DISTRIBUTION IN THE GENERAL SINGLE-SERVER QUEUE

by

IBRAHIM OSMAN

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF

MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

APRIL, 1989

ISBN 0-315-50369-6

THE UNIVERSITY OF CALGARY
FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "*The Busy Period Distribution in the General Single-Server Queue*", submitted by Ibrahim Osman in partial fulfillment of the requirements for the degree of Master of Science.

Supervisor, Dr. E.G. Enns,
Department of Mathematics and Statistics

Dr. P. Ehlers, Department of Mathematics and Statistics

M. D. Burke

Dr. M.D. Burke, Department of Mathematics and Statistics

Dr. E.S. Edgington,
Department of Psychology

25 April 1988

ABSTRACT

Individual customers arrive at a single–channel servicing facility demanding service. Explicit derivations of the distribution of the busy period of this facility are attained, as well as the distribution for the number of customers served during such a period. Different mathematical approaches are considered but the main emphasis is on utilizing the idea of generating functions and Laplace transforms. The appropriate moments of these two distributions are derived particularly when the arbitrary general independent distribution is assumed for the arrival or interarrival pattern of this single–server queue.

ACKNOWLEDGEMENT

I am grateful to my supervisor Dr. E.G. Enns for his continuous encouragement and guidance. I am also indebted to the Graduate School and the Department of Mathematics and Statistics, University of Calgary for their support of this effort.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

In this chapter I will discuss first the common characteristics, description, and application of queueing processes; in particular the busy period distribution in the multi server queue.

The conventional notation that was developed by D.G. Kendall [8] simplified the description of every queueing system. This notation is written as

input distribution/service time distribution/number of servers.

Some of the conventional notation that has been used ever since are, M for the Poisson distribution (arrivals) or the negative exponential distribution (for interarrival or service times); G for an arbitrary general distribution; D for a deterministic service or interarrival time and E_k for the Erlang distribution. As an example if the arrivals are Poisson, service times have a negative exponential distribution and there are c servers in the system, then this queue is denoted by M/M/c.

As illustrated by this simplified notation, the common characteristics of a queueing system are respectively, the input process, the service mechanism, and the queue discipline. The input process constitutes the arrivals to the system. These arrivals are always random and controlled by factors outside the scope of the system. On the other hand, the service mechanism includes the number of customers getting served at a particular time epoch, as well as the queue discipline or order in which customers are served, and the duration of each customer's service. In certain cases, the queue discipline might be revised in order to simplify some mathematical results.

Most queueing systems, in fact, could be viewed as a renewal process over time, where the main interest is focused upon counting "occurrences" that take place as a function of time. At a renewal point the past history of the queue is no longer relevant for predicting its future evolution. In the M/M/c queue any point in time is a renewal point. In the M/G/1 queue, the end of a busy period and end of the busy cycle are renewal points. For the G/G/1 queue, the end of the busy cycle or beginning of a busy period is the only renewal point.

According to what is mentioned so far, I could define queueing systems to be a mathematically descriptive theory that have a wide range of applications in the real world. This was best described by Kendall [8], when he wrote, "The theory of queues has a special appeal to the mathematician interested in stochastic processes ....". He also added that by studying queueing systems, one can benefit by gaining more insight into other stochastic processes as well.

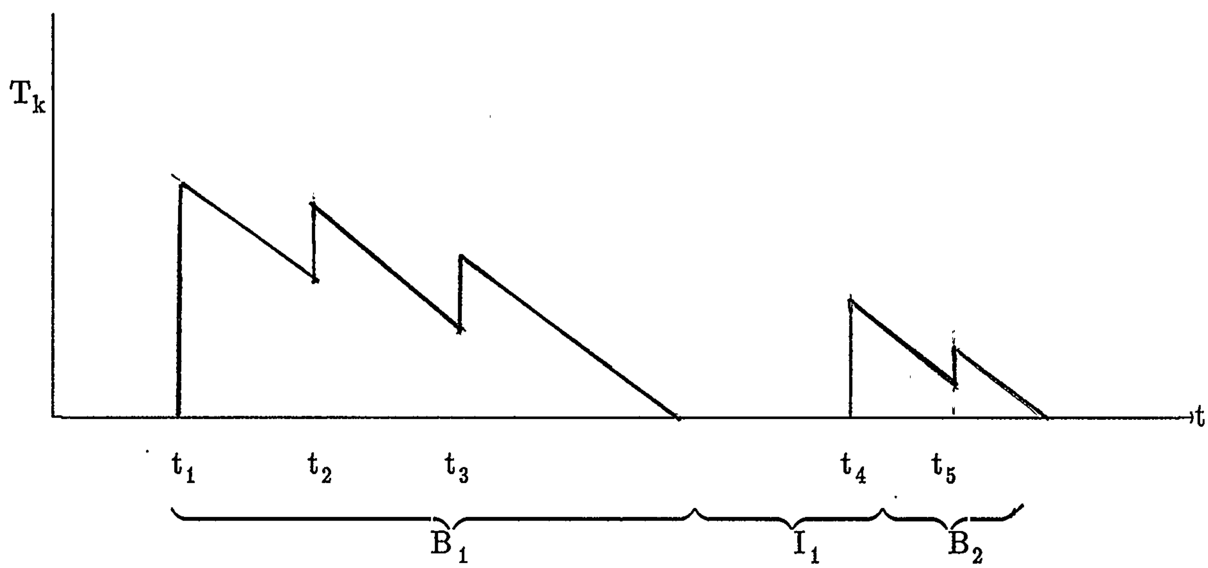
Queues occur in any system, when at a given time, the number of "arrivals" demanding a certain service exceeds the capacity of the service facility. What is required for a stable queue is that the average capacity of the service facility should be sufficient to deal with the average rate of customer arrivals. The variations in the time intervals between arrivals and the variable duration of service times result in waiting lines occurring from time to time. The busy period which is defined as the time that a single–

server channel remains continuously busy, is the main theme of my thesis. I would focus my attention on a set of two random variables, one of them is the discrete random variable, K, which represents the number of customers served during the busy period. The other random variable is B, which is the time during which the single–server is continuously busy. A good example describing such a situation is where we have aircraft landing in a small airport with a single runway, and we are interested to know how long does the landing operation take for all aircraft to land. Another example is where we have ships arriving at a port with a single berth and the service being performed by the cranes at the berth occupied by the ship. The port is busy as long as boats are waiting for the berth.

The two examples above trigger the question of how long such a single channel (runway or berth) is being occupied, or what is the distribution of the periods during which this single channel is being utilized, and most important is how to determine such a distribution according to the different situations assumed. And out of this distribution we will be able to calculate the necessary parameters e.g. mean, variance, and covariance of K and B. The busy period commences once an arrival or a customer moves into service and it is terminated whenever the last customer or unit in the queue has their service completed. The durations of the different busy periods are denoted by the random variables B_1, B_2, ..., B_r which are independently and identically distributed. And since we are interested in the time duration until all customers are served, it does not matter if we reconsider the order in which customers are being served. However, unless otherwise mentioned, I would assume that the order in which customers are served is on the first–in–first–out (FIFO) basis. Though the busy period is in fact independent of the order in which arrivals are being served. In most of the situations FIFO is the natural order for service, but also there are some cases where random selection or priorities exist.

The idea of the busy period, or more accurately the length of this period could be well explained by the stochastic process $\{T_k(t),\ t > 0\}$. This stochastic process symbolizes the duration or interval of time required to empty the system of all customers. In other words, it is the amount of the unfinished work at time t. T_k might also be viewed as the remaining time before the single server becomes idle, and k stands for the number of customers served during that particular time interval.

If we let x_k be the service time needed by the k^{th} customer entering the system, and t_k be the arrival time of that customer. Then graph (1.1) shows clearly how the length of the busy period accumulates over time. The first customer enters the system at time, t_1, and the amount of time required to serve him is x_1. But before the single–server finishes the work load of the first customer, a second arrival may enter the system at time, t_2, with a work



Graph (1.1) : the stochastic process T_k

load that requires x_2 of time to be done. Now the first arrival will definitely push T_k upwards by x_1 from the zero axis, after which T_k starts declining towards the zero line again. This decline is the result of the fact that the single server starts working on the first customer's work load. However, upon the arrival of the second customer, T_k will be pushed up to a new height by the presence of both customers in the system. And, as long as the single server is handling both customers' work load, T_k decreases again with a slope of negative one. Likewise, and as time passes, and before the server finishes the available work load, a third customer might step in the system at time t_3 with an extra work load of x_3. Ultimately, T_k will jump to a new vertical height as shown in graph (1.1). And, if and only if, the server finishes all the piled work load of the three customers before the next arrival takes place, then this first phase of elapsed time, represented by B_1 is considered to be the first busy period. That means the single server has successfully emptied the system of the first group of arrivals, after which he is apt to enjoy his first idle period denoted by I_1.

Upon an arrival of a new customer to the system (fourth in the example), the first idle period will be terminated and the second busy period, B_2 commences. The new period will consist of a random number of customers (two in the example, the fourth and fifth in the system) as being illustrated by the graph. However, to conclude the above example, I would say that it took the single server a time B_1 to clear his desk in the first phase, and he remained idle a period of time of length I_1, which represents the first idle period. In the second phase, it took him B_2 of time to finish all the compiled work load. In other words, the whole time interval under investigation, that is, (0,t), goes through alternating cycles of a busy period followed by an idle period, which is in turn followed by another busy period, and so on and so forth. It is a series of alternating busy and idle periods.

Taking into consideration the above illustration of the stochastic process, T_k, which is regarded as a continuous state Markov process that contains some discontinuities at certain epochs of the whole time period. And, due to the fact that we are only concerned with the busy periods, it is worth noting that any sub–busy period, B_i, is in fact identical with the main one, B, for the M/M/c queue which is covered in the third chapter of this thesis.

The second chapter of the thesis deals with the busy period distribution for the M/M/1 queue which is tackled through different mathematical approaches. It is derived first by the absorbing Markov chain where state 0, is considered to be the absorbing state. Bailey [1] was the first to introduce the idea of generating functions and Laplace transforms which facilitate the derivation of the p.d.f. of the busy period for almost every queue. In fact I will use his methods with some simplifications in inverting the Laplace transform. I will also consider the bivariate distribution of B, the length of the busy period and K, the number of customers served during this period, after setting up the distribution of K by the reflection principle, Feller [7]. This principle takes into account the number of transitions in the system, where the up–transitions are the result of arrivals and the down–transitions are due to services accomplished.

The third chapter deals with the busy period distribution for the M/M/2 and the M/M/3 queues. It shows how the busy period gets more and more complicated as the number of servers increases not only in derivation of the distribution but also in the definition of the busy period as well. The argument that was applied to the M/M/2 will be extended partly to cover the general case of M/M/c. The reason behind this is the vital importance of the M/M/c queue in practical applications in telephone, computer, and the air–traffic industries.

In the fourth chapter, I will discuss the derivation of the busy period distribution for two types of queues, namely the G/M/1 and the M/G/1. However, sometimes it is difficult to invert the Laplace transform of this distribution, but, it is possible to derive the important moments from the transform. In the case of M/G/1, the bivariate distribution of B and K is derived, while for the G/M/1, the distribution of B is derived by considering the system to contain r arrivals demanding service.

The last chapter covers the G/G/1 queue, where there is an independent arbitrary general distribution in the arrival pattern as well as the inter–arrival times. The Wiener–Hopf decomposition is used, where the complex plane is separated into two halves.

CHAPTER II

THE BUSY PERIOD DISTRIBUTION FOR THE M/M/1 QUEUE

Out of all topics in the queueing theory, the probability density function of the busy period was subjected to a wide range of theoretical approaches. However, explicit derivation for this density function was given by Palm [10], Kendall [8], Bailey [1], Prabhu [13], and Champernowne [2]. In fact, such a distribution could be derived directly from the difference – differential equations utilizing the idea of generating functions and Laplace transforms, which were first introduced by Bailey [1].

In the M/M/1 queue, customers arrive randomly one at a time according to a Poisson distribution with a parameter λ as the rate of arrivals. All arrivals receive a certain service at a single counter. The service times are independent and identically distributed with a negative exponential distribution with a rate of service, μ. Let us assume for the time being that customers are served in the order of their arrival, that is FIFO, (first in, first out) unless otherwise stated. It is also assumed that service is not interrupted as long as there are some customers waiting in the queue. Then, let N_t be the number of customers in the system at time t and $p_n(t) = P\{N_t=n\}$. Moreover, if i is the initial number of customers in the system at the opening time, t = 0, then,

$$p_i(0) = \delta_{k,i} \text{ where } \delta_{k,i} = 1 \quad \text{if} \quad k = i$$
$$= 0 \quad \text{if} \quad k \neq i.$$

In order to develop the difference equations for this queue using $p_n(t)$, we have to consider the probability postulates of the different situations of arrivals and departures from the system at time, t. For example, the probability that one new customer arrives during the time interval $(t, t + \Delta t)$ is $\lambda\Delta t$, hence, the probability that no customer arrives during the same period is $1 - \lambda\Delta t$, and the probability that more than one customer arrive is assumed to be small and negligible as $0(\Delta t)$. Such that $\frac{0(\Delta t)}{\Delta t} \longrightarrow \emptyset$,whenever $\Delta t \longrightarrow \emptyset$. Likewise, the probability that there is only one service being completed during the same period of time is $\mu\Delta t$, and $1 - \mu\Delta t$ is, therefore, the probability that there is no service done during that period. And, as we mentioned earlier, service completion is independent of the time at which it started, (Chapter 1).

Now, we could derive what became to be known as the Chapman–Kolmogorov or the difference equations for the M/M/1 queue, as follows,

$$p_0(t+\Delta t) = p_0(t)(1-\lambda\Delta t) + p_1(t)\ \mu\Delta t, \tag{2.1}$$

and,

$$p_n(t+\Delta t) = p_n(t)(1-\lambda\Delta t)(1-\mu\Delta t) + p_{n+1}(t)\ \mu\Delta t(1-\lambda\Delta t)$$
$$+ p_{n-1}(t)\ \lambda\Delta t(1-\mu\Delta t),\ n \geq 1. \tag{2.2}$$

In the limit as $\Delta t \rightarrow 0$ one obtains:

$$\frac{dp_0(t)}{dt} = -\lambda\ p_0(t) + \mu\ p_1(t) \tag{2.3}$$

and,

$$\frac{dp_n(t)}{dt} = -(\lambda+\mu)\ p_n(t) + \mu\ p_{n+1}(t) + \lambda\ p_{n-1}(t). \tag{2.4}$$

<u>THE ABSORBING STATE METHOD</u>:

In order to derive the distribution of the length of the busy period, or time that the server is continuously busy, define the state when $N_t = 0$ as an absorbing state. Hence the differential difference equations become:

$$p_n(t) = P\{N_t = n | N_s > 0,\ 0 \leq s < t\} \tag{2.5}$$

with the following boundary conditions,

$$p_1(0) = 1 \quad \text{and} \quad p_n(0) = 0 \quad \text{for} \quad n \geq 2. \tag{2.6}$$

$$p_0(t+\Delta t) = p_1(t)\ \mu\Delta t \tag{2.7}$$

$$p_1(t+\Delta t) = p_1(t)(1-(\lambda+\mu)\Delta t)+p_2(t)\mu\Delta t \tag{2.8}$$

$$p_n(t+\Delta t) = p_n(t)(1-(\lambda+\mu)\Delta t)+p_{n+1}(t)\mu\Delta t+p_{n-1}(t)\lambda\Delta t \quad \text{for } n \geq 2. \tag{2.9}$$

Again as $\Delta t \to 0$, we obtain

$$\frac{dp_0(t)}{dt} = \mu p_1(t) \tag{2.10}$$

$$\frac{dp_1(t)}{dt} = -(\lambda+\mu)\ p_1(t) + \mu p_2(t) \tag{2.11}$$

$$\frac{dp_n(t)}{dt} = -(\lambda+\mu)\ p_n(t) + \lambda p_{n-1}(t) + \mu p_{n+1}(t), \quad n \geq 2. \tag{2.12}$$

Now the busy period p.d.f. for this absorbing system is:

$$f_B(t) = \mu p_1(t) \tag{2.13}$$

or from (2.10):

$$f_B(t) = \frac{dp_0(t)}{dt} . \tag{2.14}$$

Moreover, define the generating function of $p_n(t)$ as

$$G(z,t) = \sum_{n=1}^{\infty} z^n \, p_n(t) \tag{2.15}$$

and taking the derivative of G(z,t) w.r.t. time t, will give,

$$\frac{d}{dt} G(z,t) = \sum_{n=1}^{\infty} z^n \frac{dp_n(t)}{dt} . \tag{2.16}$$

Now, all we need to do is to substitute all the different possible values of $\frac{dp_n(t)}{dt}$ from (2.11) – (2.12) into the definition of the generating function of (2.16) with some rearrangements of similar terms, we get,

$$\frac{dG(z,t)}{dt} = -(\lambda+\mu)\, G(z,t)+\lambda\, z\, G(z,t)+\frac{\mu}{z}\, [G(z,t)-z\, p_1(t)] . \tag{2.17}$$

At this point I define and denote the Laplace transform of the p.d.f., f(t), as

$$\mathscr{L}[f(t)] = \int_0^{\infty} e^{-st} f(t)\, dt = \hat{f}(s). \tag{2.18}$$

Hence, taking the Laplace transform of (2.17), will give the following,

$$\mathscr{L}\left[\frac{d}{dt} G(z,t)\right] = -(\lambda+\mu)\, \hat{G}(z,s)+\lambda\, z\, \hat{G}(z,s)+\frac{\mu}{z}\, [\hat{G}(z,s)-z\, \hat{p}_1(s)]. \tag{2.19}$$

But, we know that the Laplace transform of a derivative could, in fact, be written as (i.e. the L.H.S. of (2.19)),

$$\mathscr{L}\left[\frac{d}{dt} G(z,t)\right] = - G(z,o) + s \hat{G}(z,s). \tag{2.20}$$

Suppose that the initial number in the system is one customer, then we could apply the initial condition,

$$G(z,o) = z. \tag{2.21}$$

From (2.19), (2.20) and (2.21) we can arrive at the result

$$- z + s \hat{G}(z,s) = - (\lambda+\mu) \hat{G}(z,s)+\lambda \hat{G}(z,s)+\frac{\mu}{z} [\hat{G}(z,s)-z \hat{p}_1(s)]. \tag{2.22}$$

When it is rearranged again, it will give

$$\hat{G}(z,s) = \frac{z(\mu \hat{p}_1(s)-z)}{\lambda z^2-(s+\lambda+\mu)z+\mu} . \tag{2.23}$$

The objective here is to find a solution for the p.d.f. of the busy period as stated in (2.13).

Rouche's Theorem, Saaty [14], says that if f(s) and g(s) are two analytic functions of s inside and on a closed contour C, and also if $g(s) < f(s)$ on C, then f(s) and $f(s) + g(s)$ have the same number of zeros inside the contour C. Applying this theorem to (2.23) indicates that if the denominator of $\hat{G}(z,s)$ vanishes when $\hat{G}(z,s)$ is finite, then the numerator must also vanish at the same point where the former vanished. Therefore, all we need to do is to equate the denominator of $\hat{G}(z,s)$ to zero and solve it for the right root, at which $\hat{G}(z,s)$ converges inside and on the unit circle $|Z| = 1$. This approach results in the following two roots for the quadratic equation formed by equating the denominator of (2.23) to zero,

$$\xi_{1,2}(s) = \frac{(s+\lambda+\mu) \pm \sqrt{(s+\lambda+\mu)^2 - 4\lambda\mu}}{2\lambda} \tag{2.24}$$

at s = 0, both roots are

$$\xi_{1,2}(0) = \frac{(\lambda+\mu) \pm \sqrt{(\lambda-\mu)^2}}{2\lambda} = \frac{\mu}{\lambda} \text{ or } 1 \tag{2.25}$$

but $\rho = \frac{\lambda}{\mu} < 1$, that is the traffic intensity is always less than one, otherwise the queue will explode with a heavy backlog. Therefore $\frac{\mu}{\lambda} > 1$, and we will consider only the root $\xi_2(s)$ with the negative sign which gives the value of one when $s = 0$. Hence, applying Rouche's theorem means that the numerator must vanish at the point where $z = \xi_2(s)$. This results in

$$\mu\, \hat{p}_1(s) = \xi_2(s). \tag{2.26}$$

Then by considering both results in (2.13) and (2.26), we conclude,

$$\hat{f}_B(s) = \xi_2(s) = \frac{(s+\lambda+\mu) - \sqrt{(s+\lambda+\mu)^2 - 4\lambda\mu}}{2\lambda} \tag{2.27}$$

where we have $\hat{f}_B(0) = 1$, and all that is required at this point is to invert the above Laplace transform in order to get an expression for the p.d.f., $f_B(t)$. That is,

$$f_B(t) = \mathscr{L}^{-1}[\xi_2(s)]. \tag{2.28}$$

However, in order to invert the Laplace transform of $\xi_2(s)$, first we need to expand its expression after rewriting $\hat{f}_B(s)$ as,

$$\hat{f}_B(s) = \frac{s+\lambda+\mu}{2\lambda}\left[1 - \sqrt{1 - \frac{4\lambda\mu}{(s+\lambda+\mu)^2}}\right]. \tag{2.29}$$

The above (2.29) could be expanded using the binomial series to give the following,

$$\hat{f}_B(s) = \sum_{k=1}^{\infty} \frac{\lambda^{k-1}\mu^k (2k-2)!}{(s+\lambda+\mu)^{2k-1}\, k!(k-1)!} \tag{2.30}$$

then by using the following Laplace transform,

$$\mathscr{L}\,[t^{2k-2}] = \int_0^{\infty} e^{-(s+\lambda+\mu)t}\, t^{2k-2}\, dt = \frac{(2k-2)!}{(s+\lambda+\mu)^{2k-1}}$$

we can invert (2.30) to get,

$$f_B(t) = \sum_{k=1}^{\infty} \frac{\lambda^{k-1}\mu^k t^{2k-2}}{k!\,(k-1)!}\, e^{-(\lambda+\mu)t}\,. \tag{2.31}$$

And, if we let $j = k-1$, we can rewrite (2.31) as,

$$f_B(t) = \frac{1}{t}\sum_{j=0}^{\infty} \frac{\lambda^j \mu^{j+1} t^{2j+1}}{j!\,(j+1)!}\, e^{-(\lambda+\mu)t}$$

or

$$f_B(t) = \frac{1}{t}\, e^{-(\lambda+\mu)t}\sqrt{\frac{\mu}{\lambda}}\; I_1\,(2t\sqrt{\lambda\mu}) \tag{2.32}$$

where,

$$I_1\,(2t\sqrt{\lambda\mu}) = \sum_{j=0}^{\infty} \frac{(t\sqrt{\lambda\mu})^{2j+1}}{j!\,(j+1)!}$$

which is known as the modified Bessel function of the first order, and (2.32) is the required p.d.f. of the busy period for the M/M/1 queue.

THE BENCH METHOD:

The Bench method is used when there are initially i customers in the system where each can be considered as generating his own busy period. Hence B and K could be partitioned as follows

$$B = B_1 + B_2 + \dots + B_i$$

$$K = K_1 + K_2 + \dots + K_i.$$

All of the i customers are considered to be awaiting on the bench and the busy period of the i^{th} customer includes his service time duration and those customers who join the queue until the queue is empty. And in this case, the bivariate distribution of B and K

$$F_i(k,t) = P\{K=k,\ B \leq t \mid N_0 = i\}$$

could be derived directly by partitioning it in terms of the first event which could be either an arrival or a departure. For example, at the beginning when there is only one customer (the first) in the system, we can move one step further to consider the next situation of having two customers in the system, one is being served and the other is awaiting service at the bench. Or, we can move one step backward to the situation where there is no service done but there is one arrival to the system. Hence,

$$f_1(k,t) = \mu\, e^{-(\lambda+\mu)t}\, \delta_{k,1} + \lambda\, e^{-(\lambda+\mu)t} * f_2(k,t). \qquad (2.33)$$

where the joint distribution and density are:

$$f_i(k,t) = \frac{dF_i(k,t)}{dt} .$$

The first term on the L.H.S. represents the situation of one service being completed during which there was no arrival, where by $\delta_{k,1}$, the Kronecker's symbol

$$\delta_{k,1} = \begin{cases} 1, & \text{if} \quad k = 1 \\ 0, & \text{if} \quad k \neq 1 \end{cases} . \tag{2.34}$$

The second term of (2.33) is the case of an arrival but no service being completed. The star (*) stands for the convolution where,

$$a(t) * b(t) = \int_0^t a(u)b(t-u)\, du.$$

The convolution in (2.33) takes care of future services of those customers who joined the queue during each service time.

Now taking Laplace transforms of both sides of (2.33), will give

$$\hat{f}_1(k,s) = \frac{\mu}{s+\lambda+\mu}\, \delta_{k,1} + \frac{\lambda}{s+\lambda+\mu}\, \hat{f}_2(k,s). \tag{2.35}$$

And to define the generating function,

$$\phi_i(z,s) = \sum_{k=i}^{\infty} z^k\, \hat{f}_i(k,s) \tag{2.36}$$

then applying (2.36) to (2.35) when there is initially one customer in the system, that is, $i = 1$, gives

$$\phi_1(z,s) = \frac{\mu}{s+\lambda+\mu} z + \frac{\lambda}{s+\lambda+\mu} \phi_2(z,s). \tag{2.37}$$

However, it should be understood that the general generating function for any number i is

$$\phi_i(z,s) = E[z^K e^{-sB}/I=i] \tag{2.38}$$

where B and K are as partitioned above. Then B_i is the time to serve the i^{th} customer plus those who join the queue until the queue is empty and all of these customers are denoted by K_i. So, B_r and K_r, $r = 1, \ldots, i$ are i.i.d. so that (2.38) could be factored as

$$\begin{aligned}\phi_i(z,s) &= E[z^K e^{-sB}] \\ &= \prod_{r=1}^{i} E[z^{K_r} e^{-sB_r}] \\ &= \left[E[z^{K_1} e^{-sB_1}]\right]^i \\ &= \left[\phi_1(z,s)\right]^i. \end{aligned} \tag{2.39}$$

Moreover, if we let

$$\phi_1(z,s) = \phi$$

then by considering (2.37) and (2.39) together, we have

$$\phi(s+\lambda+\mu) = \mu z + \lambda\phi^2 \tag{2.40}$$

which is a quadratic equation in ϕ that could be written as

$$\lambda\phi^2 - (s+\lambda+\mu)\,\phi + \mu z = 0 \tag{2.41}$$

for which the appropriate root is,

$$\phi(z,s) = \frac{(s+\lambda+\mu) - \sqrt{(s+\lambda+\mu)^2-4\lambda\mu z}}{2\lambda}\,. \tag{2.42}$$

And, by applying the binomial expansion, as we did in (2.29), we arrive at

$$\phi(z,s) = \sum_{k=1}^{\infty} \frac{(2k-2)!}{k!(k-1)!}\,\frac{\lambda^{k-1}\mu^{k}z^{k}}{(s+\lambda+\mu)^{2k-1}}\,. \tag{2.43}$$

Hence

$$\hat{f}_1(k,s) = \frac{(2k-2)!}{k!(k-1)!}\,\frac{\lambda^{k-1}\mu^{k}}{(s+\lambda+\mu)^{2k-1}}. \tag{2.44}$$

Taking the inverse Laplace Transform, one obtains:

$$f_1(k,t) = \frac{\lambda^{k-1}\mu^{k}}{k!(k-1)!}\; t^{2k-2}\, e^{-(\lambda+\mu)t}. \tag{2.45}$$

And in order to derive the marginal p.d.f. of B all we have to do is to sum over all the possible values of K, that is,

$$f_B(t) = \sum_{k=1}^{\infty} \frac{\lambda^{k-1}\mu^{k}}{k!(k-1)!}\, t^{2k-2}\, e^{-(\lambda+\mu)t} \cdot \tag{2.46}$$

which is exactly the same result we had already derived in (2.31). Likewise, to get the marginal distribution of K, the number of customers served during a busy period, we integrate over all t, to get the following,

$$P\{K=k\} = \frac{\lambda^{k-1}\mu^{k}}{k!(k-1)!}\,\frac{(2k-2)!}{(\lambda+\mu)^{2k-1}} \tag{2.47}$$

for $\rho = \frac{\lambda}{\mu}$, the traffic intensity, the above can be written as,

$$P\{K=k\} = \frac{(2k-2)!}{k!(k-1)!}\,\frac{\rho^{k-1}}{(1+\rho)^{2k-1}}\,. \tag{2.48}$$

<u>**THE REFLECTION METHOD**</u>:

The above result could also be derived by the use of the Reflection principle where we assume that there are i customers initially in the system. Here, we look to the system in terms of the number of transitions that had occurred. A transition is defined as the occurrence of an arrival or a departure in the system. Obviously the busy period terminates the moment that we have had sufficient departures to empty the system. Figure (2.1) represents the number in the system based upon a hypothetical number of transitions. Now, whenever we have a departure from the system, this is represented by a "down" transition. In order to serve k ($k \geq i$) customers during a busy period, there must be k "down" transitions and hence k–i arrivals or "up" transitions. As represented by figure (2.1), there must be a total of 2k–i transitions when we reach the zero axis in the system for the first time. Let us assume that the system contains a total number of n transitions, where r is the number of up transitions that occurred after time, t = 0. That makes the total number

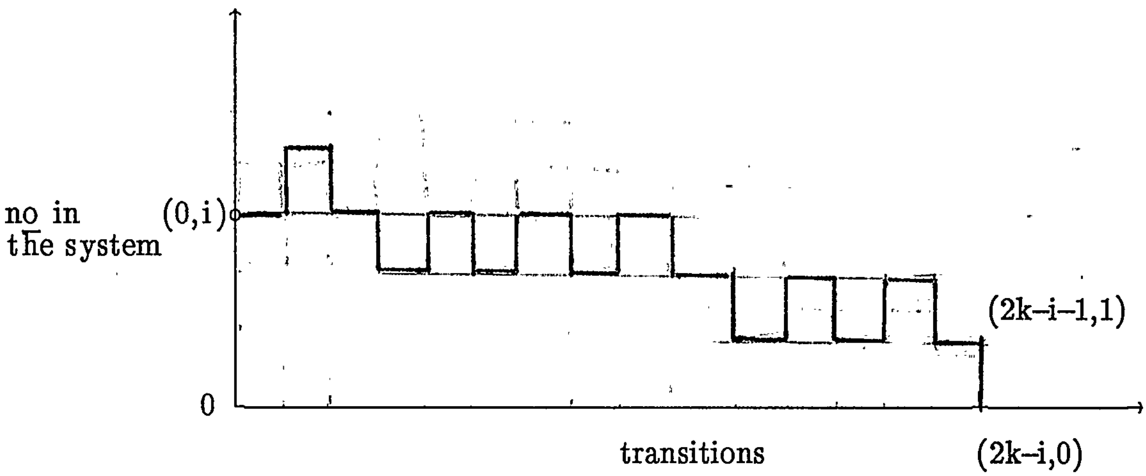


Figure (2.1): The number of paths in the Reflection method

of up transitions to be i+r, while on the other hand the number of down transitions is s. Therefore the number of customers served during the busy period should be equal to the number of services performed, which is the total number of down transitions.

$$k = i + r = s. \tag{2.49}$$

Let $f_+(t)$ be the joint probability density function and the probability that an up transition occurs at time t since the last transition. Similarly, define $f_-(t)$ as the joint probability density and the probability that a down transition occurs (but no arrival) at time t. Then,

$$f_+(t) = \lambda\ e^{-(\lambda+\mu)t}$$

and (2.50)

$$f_-(t) = \mu\ e^{-(\lambda+\mu)t}$$

with the following Laplace transforms, respectively,

$$\hat{f}_+(s) = \frac{\lambda}{(s+\lambda+\mu)}$$

and, (2.51)

$$\hat{f}_-(s) = \frac{\mu}{(s+\lambda+\mu)}.$$

Hence the probability of an up transition, that is, the probability of one arrival and no service is

$$p\{up\} = \int_0^\infty f_+(t)\ dt = \frac{\rho}{1+\rho},\ \text{if } \rho = \frac{\lambda}{\mu}. \tag{2.52}$$

Similarly,

$$p\{down\} = \int_0^\infty f_-(t)\ dt = \frac{1}{1+\rho}. \tag{2.53}$$

Let N = the number of configurations, see example in figure (2.1), that lead from an initial point (0,i) to the point (2k–i, 0) where the axis is touched for the first time at (2k–i, 0). This is equivalently and better stated as the number of paths from (0,i) to (2k–i–1, 1) which never touch the axis. Hence, the distribution of K, the number served during the busy period could be written as

$$P(K{=}k) = N\left[\frac{\rho}{1+\rho}\right]^r \left[\frac{1}{1+\rho}\right]^{n-r}$$

or,

$$P(K{=}k/N_0{=}i) = N\left[\frac{\rho}{1+\rho}\right]^{k-i} \left[\frac{1}{1+\rho}\right]^{n-k+i}. \tag{2.54}$$

Furthermore, if we let N_1 = the total number of paths from (0,i) to (2k–i–1, 1). These may touch or go below the axis one or more times. And, N_2 = the total number of paths from (0, –i) to (2k–i–1, 1). Then

$$N = N_1 - N_2$$

where

$$N_1 = \begin{bmatrix} 2k-i-1 \\ k-1 \end{bmatrix} \text{ and } N_2 = \begin{bmatrix} 2k-i-1 \\ k \end{bmatrix}$$

hence,

$$N = \frac{i(2k-i-1)!}{k!\,(k-i)!}\,.$$

As a result (2.54) becomes

$$P(K=k|I=i) = \frac{i(2k-i-1)!}{k!\,(k-i)!}\left[\frac{\rho}{1+\rho}\right]^{k-i}\left[\frac{1}{1+\rho}\right]^{k} \tag{2.55}$$

when there are i customers initially in the system. However, when i = 1,

$$P(K=k) = \frac{(2k-2)!}{k!(k-1)!}\,\frac{\rho^{k-1}}{(1+\rho)^{2k-1}} \tag{2.56}$$

which is the same result derived earlier. Or, the same result could be derived from the joint distribution of B and K. And to derive this distribution, we begin by assuming $f_k(t)$ be the joint probability that a particular configuration, as represented by figure (2.1) occurs and it is the probability density that it touches the axis for the first time. Then we have the convolution of (k–i) "ups" and k "downs" of those densities in (2.50). Hence the Laplace transform of $f_k(t)$, considering (2.51), is

$$\hat{f}_k(s) = (\hat{f}_+(s))^{k-i} (\hat{f}_-(s))^k$$

$$= \frac{\lambda^{k-i} \mu^k}{(s+\lambda+\mu)^{2k-i}} . \tag{2.57}$$

In order to find $f_i(k,t)$ we need to enumerate all possible configurations that allow exactly k customers to be served, and this is where we use the reflection principle as described in Feller [7]. Hence the Laplace transform of the joint density is,

$$f_i(k,s) = N \hat{f}_k(s)$$

which by inversion gives

$$f_i(k,t) = N f_k(t)$$

$$= \frac{i \lambda^{k-i} \mu^k t^{2k-i-1} e^{-(\lambda+\mu)t}}{k! (k-i)!} \tag{2.58}$$

for an initial, i, number of customers in the system at time t = 0. And by integrating over t, or summing over k, we can derive the marginal densities of K and B, respectively, as given previously by (2.47) and (2.46).

THE MODELLING METHOD:

I have suggested this name because the pattern of arrivals and departures from the single–server queue, could, in fact, be modeled in such a way that departures from the system can be regarded as arrivals with a rate of μ. Champernowne [2] suggested a direct approach to find the p.d.f. of the busy period for the M/M/1 queue. He set, N_t, the number of customers in the system at time t, to be equal to the number of arrivals to the queue minus the number of departures from the system. That is,

$$N_t = A_t - D_t \tag{2.59}$$

where arrivals follow a Poisson distribution with an arrival rate λ. Hence,

$$P(A_t=k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \tag{2.60}$$

and,

$$P(D_t=k) = \frac{(\mu t)^k e^{-\mu t}}{k!}\,. \tag{2.61}$$

Though Champernowne used the above results with a combination of some modified Bessel functions to prove his argument, I would rather simplify the analysis by considering the 'peak' time or the 'rush' time to be uniformly distributed at any point of time as u(0,t). Consequently, if there are i–1 customers in the queue at time t, where i = 1,2,...; then there must be i departures at the other end of the system. In other words, the busy period service time must accomodate the service time of the last customer at the service desk.

Now, the distribution of the busy period could be modeled as

$$f_B(t) = u(0,t)\ P\{N_t = A_t - D_t\}.$$

And due to the set up of the pattern of arrivals and departures in the system both occur independently of each other, then

$$f_B(t) = u(0,t) \sum_{i=1}^{\infty} P\{A_t = i-1\}\ P\{D_t=i\} \qquad i = 1,2,\ldots.$$

$$= \frac{1}{t} \sum_{i=1}^{\infty} \frac{(\lambda t)^{i-1} e^{-\lambda t}}{(i-1)!} \frac{(\mu t)^{i} e^{-\mu t}}{i!}$$

$$= \sqrt{\frac{\mu}{\lambda}}\ \frac{1}{t}\ e^{-(\lambda+\mu)t}\ I_1(2t\sqrt{\lambda\mu}) \tag{2.62}$$

which is also the same result as of (2.32) obtained earlier.

To get the average length of the busy period we simply find the negative derivative of the Laplace transform of the bivariate density of B and K w.r.t. to s,

$$E(B) = \frac{-\partial\ \hat{\phi}_1(1,0)}{\partial s} = \frac{1}{\mu(1-\rho)}. \tag{2.63}$$

Consequently the variance is

$$\mathrm{Var}(B) = E(B^2) - [E(B)]^2$$

$$= \frac{1+\rho}{\mu^2(1-\rho)^3}\ . \tag{2.64}$$

On the other hand, the average number of the customers served during the busy period could be derived from the Laplace transform of the same density with the derivative w.r.t. z,

$$E(K) = \frac{\partial \hat{\Phi}_1(1,0)}{\partial z} = \frac{1}{(1-\rho)} \tag{2.65}$$

and,

$$Var(K) = \frac{3\rho-1}{(1-\rho)^3} \; . \tag{2.66}$$

Another important parameter is the covariance of B and K, which is derived as follows,

$$E(BK) = \frac{-\partial \hat{\Phi}_1(1,0)}{\partial z \partial s} = \frac{1+\rho}{\mu(1-\rho)^3}$$

then,

$$Cov(B,K) = E(BK) - E(B)E(K) = \frac{2\rho}{\mu(1-\rho)^3} \; . \tag{2.67}$$

And since ρ is always positive, that is, $0 < \rho < 1$, then the covariance of B and K is always positive as well, due to the fact that as the busy period duration increases so does the number of customers served during this period. That is, both of them must generally move in the same direction.

CHAPTER III

THE M/M/C QUEUE

If one or more servers are added to the single–server queue of the previous chapter, then we need to define carefully what a busy period really is. In some instances, statisticians define the busy period for the M/M/2 queue as the time period during which both servers are busy. They ignore the time period when one of the two servers is busy and the other idle. However, to me, the real busy period starts when one of the two servers becomes busy and terminates the instant both of them become idle. The latter is the natural extension of the busy period of the single server queue discussed earlier. I will consider the busy period analysis for the M/M/2, M/M/3 queues and then generalize the argument to cover the M/M/c queue.

THE M/M/2 QUEUE:

In this queue, we have two identical servers each with a service rate of μ. Both servers will only be busy when two or more customers are in the system. But the busy period terminates the moment the last customer leaves and both servers are idle. The system will therefore alternate between none, one, or both servers being busy. But since we are interested only in the busy period, we consider only the last two cases, that is, when one or both servers are busy. Let,

$$p_{10}(t) = \text{p.d.f. of the length of a service and that no new customer arrives}$$

hence,

$$p_{10}(t) = \mu\, e^{-(\lambda+\mu)t} . \tag{3.1}$$

Similarly,

$p_{12}(t)$ = p.d.f. of time until next arrival and that the customer in service has not reached completion,

therefore,

$$p_{12}(t) = \lambda\ e^{-(\lambda+\mu)t} \tag{3.2}$$

and,

$$p_{10}(t) + p_{12}(t) = (\lambda+\mu)\ e^{-(\lambda+\mu)t}$$

= p.d.f. of time spent in state 1, where state 1 represents the situation of having one customer in the system.

Moreover, let

$$p_{21}(t) = \phi_2(t) \tag{3.3}$$

where,

$\phi_i(t)$ = p.d.f. of the period of time of entry to state 1, if we begin initially with i customers in the queue.

Then we can partition $\phi_2(t)$ in terms of one service and k arrivals as follows,

$$\phi_2(t) = \sum_{k=0}^{\infty} \left[\frac{(\lambda t)^k e^{-\lambda t}}{k!} 2\mu e^{-2\mu t} \right] * \phi_{k+1}(t) \tag{3.4}$$

where,

$$\phi_k(t) = \begin{cases} 1, & \text{if} \quad k \geq 1 \\ 0, & \text{if} \quad \text{o.w.} \end{cases} .$$

And it should be noted that the above service is in terms of both servers being busy. Now applying Laplace transforms to (3.1) – (3.4) respectively, we get,

$$\hat{p}_{10}(s) = \frac{\mu}{s+\lambda+\mu} \tag{3.5}$$

$$\hat{p}_{12}(s) = \frac{\lambda}{s+\lambda+\mu} \tag{3.6}$$

$$\hat{p}_{21}(s) = \hat{\phi}_2(s) \tag{3.7}$$

$$\hat{\phi}_2(s) = \mathscr{L}\left[\sum_{k=0}^{\infty} \frac{(\lambda t)^k e^{-\lambda t}}{k!} 2\mu e^{-2\mu t}\right] \hat{\phi}_{k+1}(s). \tag{3.8}$$

Let

$$\hat{\phi}_2(s) = \hat{\phi}(s)$$

then

$$\hat{\phi}_{k+1}(s) = (\hat{\phi}(s))^k. \tag{3.9}$$

Now (3.8) becomes

$$\hat{\phi}(s) = \mathscr{L}\left[\sum_{k=0}^{\infty} \frac{(\lambda t)^k e^{-\lambda t}}{k!} 2\mu e^{-2\mu t}\right] (\hat{\phi}(s))^k$$

$$= \mathscr{L}\left[2\mu\, e^{-(\lambda(1-\hat{\phi}(s))+2\mu)t}\right]$$

$$= \frac{2\mu}{s+\lambda(1-\hat{\phi}(s))+2\mu} \tag{3.10}$$

With some basic algebra (3.10) gives the following quadratic equation,

$$\lambda\, \hat{\phi}^2 - (s+\lambda+2\mu)\, \hat{\phi} + 2\mu = 0$$

with the following appropriate root,

$$\hat{p}_{21}(s) = \hat{\phi}(s) = \frac{(s+\lambda+2\mu) - \sqrt{(s+\lambda+2\mu)^2-8\lambda\mu}}{2\lambda} \tag{3.11}$$

which is the Laplace transform of the p.d.f. of the busy period if it were considered to be the total occupation of the two servers. However, with the proper definition of the busy period, its p.d.f. could be partioned in the following way,

$$f_B(t) = p_{10}(t) + p_{12}(t) * p_{21}(t) * p_{10}(t) + \ldots$$

for which the Laplace transform is,

$$\hat{f}_B(s) = \hat{p}_{10}(s)\ [1 + \hat{p}_{12}(s)\ \hat{p}_{21}(s) + (\hat{p}_{12}(s)\ \hat{p}_{21}(s))^2 + \ldots]$$

$$= \frac{\hat{p}_{10}(s)}{1-\hat{p}_{12}(s)\hat{p}_{21}(s)} \tag{3.12}$$

in which, we substitute the different values of $\hat{p}_{10}(s)$, $\hat{p}_{12}(s)$, and $\hat{p}_{21}(s)$, respectively, from (3.5), (3.6) and (3.11), we get

$$\hat{f}_B(s) = \frac{2\mu}{s+\lambda+\sqrt{(s+\lambda+2\mu)^2-8\lambda\mu}} \tag{3.13}$$

the Laplace transform of the real busy period as defined previously. However, it is difficult to invert such a Laplace transform, but it is worth noting the necessary condition for a density function, that is,

$$\hat{f}_B(0) = 1.$$

THE M/M/2 BIVARIATE DISTRIBUTION OF B AND K:

The bivariate density of the busy period, B, and the number served during this period, K, may be written in the form of (2.35) of the Bench

Method. And its derivation is identical to that of (2.40). However, in this case the Bench Method does not apply. Now, applying the generating function approach, as we did before,

$$\hat{\phi}_2(z,s) = E[z^K e^{-sB}/I=2] \tag{3.14}$$

where $K = K_1 + K_2$ and $B = B_1 + B_2$. B_1 and K_1 are defined as the length of time and the number served until the moment we have one customer in the queue for the first time. While B_2 and K_2 are the length of the busy period and the number of customers served when we begin with one customer. Accordingly, B_1 and K_1 can be interpreted as the length of the busy period and the number served in a M/M/1 queue with a service rate of 2μ. Hence (3.14) becomes

$$\hat{\phi}_2(z,s) = \hat{\phi}^{(2)}(z,s)\ E(z^{K_2} e^{-sB_2}) \tag{3.15}$$

where $\hat{\phi}^{(2)}(z,s)$ is defined by (2.42) with μ being replaced by 2μ. And, by definition of B_2 and K_2,

$$E[z^{K_2} e^{-sB_2}] = \hat{\phi}_1(z,s). \tag{3.16}$$

Now recall the result (2.37), which is written as

$$\hat{\phi}_1(z,s) = \frac{\mu z}{s+\lambda+\mu} + \frac{\lambda\hat{\phi}_2(z,s)}{s+\lambda+\mu}\ . \tag{3.17}$$

Then, by substituting (3.15), and (3.16) into (3.17), we get,

$$\hat{\phi}_1(z,s) = \frac{2\mu\ z}{s+\lambda+\sqrt{(s+\lambda+2\mu)^2-8\lambda\mu z}} \tag{3.18}$$

which is the Laplace transform of the bivariate density of B and K for the M/M/2 queue. As a result we would be able to derive the important moments of B and K for this queue as

$$E(B) = -\frac{\partial\hat{\phi}_1(1,0)}{\partial s} = \frac{1}{\mu(1-\rho_2)} \tag{3.19}$$

where

$$\rho_2 = \frac{\lambda}{2\mu} < 1 \text{ for stationarity}$$

is the traffic intensity for the M/M/2 queue. Similarly.

$$E(K) = -\frac{\partial\hat{\phi}_1(1,0)}{\partial z} = \frac{1+\rho_2}{1-\rho_2}\ . \tag{3.20}$$

And the variance of B is

$$\mathrm{Var}(B) = \frac{1}{\mu^2(1-\rho_2)^3}\ . \tag{3.21}$$

The covariance of B and K is

$$\mathrm{Cov}(B,K) = \frac{\rho_2(3-\rho_2)}{\mu(1-\rho_2)^3}\ .$$

<u>THE M/M/3 BIVARIATE DISTRIBUTION OF B AND K</u>:

We will provide the derivation of the bivariate distribution in this case as it illustrates the generalization. Here we will begin with the arrival of a single customer that makes one of the three servers busy. We vary between one, two, and three servers being busy until the busy period terminates when the last customer leaves the system. Let $f_i(k,t)$ be the bivariate density of K and B when there are i customers initially in the system. Again we can partition densities in terms of the first event as,

$$f_1(k,t) = \mu\, e^{-(\lambda+\mu)t}\, \delta_{k,1} + \lambda\, e^{-(\lambda+\mu)t} * f_2(k,t) \qquad (3.22)$$

and

$$f_2(k,t) = 2\mu\, e^{-(\lambda+2\mu)t} * f_1(k-1,t) + \lambda\, e^{-(\lambda+2\mu)t} * f_3(k,t). \qquad (3.23)$$

Let the Laplace transform of the generating function,

$$\hat{\phi}_i(z,s) = E[z^K e^{-sB}/I=i] = \mathscr{L}[f_i(k,t)]$$

which when applied to (3.22), (3.23) respectively,

$$(s+\lambda+\mu)\, \hat{\phi}_1(z,s) = \mu z + \lambda\, \hat{\phi}_2(z,s) \qquad (3.24)$$

and,

$$(s+\lambda+2\mu)\, \hat{\phi}_2(z,s) = 2\mu z\, \hat{\phi}_1(z,s) + \lambda\, \hat{\phi}_3(z,s). \qquad (3.25)$$

Now, we define,

$$\hat{\phi}_3(z,s) = E\left[Z^{K_1+K_2}\, e^{-s(B_1+B_2)}/I=3\right] \qquad (3.26)$$

where K_1 and B_1 are the number served and the length of time until we have two customers in the queue for the first time. This is equivalent to the busy period in the M/M/1 queue with μ being replaced by 3μ. Also, we could write

$$E\left[e^{K_2}\, e^{-sB_2}\right] = \phi_2(z,s). \qquad (3.27)$$

Therefore, using (3.27) and (3.26), we have

$$\phi_3(z,s) = \phi_2(z,s) \frac{(s+\lambda+3\mu) - \sqrt{(s+\lambda+3\mu)^2-12\lambda\mu z}}{2\lambda} . \tag{3.28}$$

Substitution of (3.28) into (3.25) and (3.24) and eliminating $\phi_2(z,s)$ in both equations, we will obtain the desired result,

$$\phi_1(z,s) = \frac{\mu z\left[s+\lambda+\mu + \sqrt{(s+\lambda+3\mu)^2-12\lambda\mu z}\right]}{(s+\lambda+\mu)\left[(s+\lambda+\mu) + \sqrt{(s+\lambda+3\mu) -12\lambda\mu z}\right]-4z\lambda\mu} \tag{3.29}$$

which is the Laplace transform of the bivariate density of B and K for the M/M/3 queue.

This procedure may be continued for queues of this type with additional servers. But as the number of servers gets large, it will be difficult to follow the required substitution. Therefore, I will consider another procedure that will cover all cases; but only for the r.v. B.

THE M/M/C QUEUE:

It is not difficult to extend the concept of the real busy period to the multi servers case. In a similar fashion, the busy period begins with an arrival to the system that makes at least one of the c servers busy. I would use the method of the absorbing Markov chain for the M/M/2 queue and then generalize it for the M/M/c case. The difference–differential equations for the M/M/2 assuming state 0 is absorbing

$$\frac{dp_0(t)}{dt} = \mu\, p_1(t), \qquad n = 0$$

$$\frac{dp_1(t)}{dt} = -(\lambda+\mu)\, p_1(t) + 2\mu\, p_2(t), \qquad n = 1$$

$$\frac{dp_n(t)}{dt} = -(\lambda+2\mu)\ p_n(t) + \lambda\ p_{n-1}(t) + 2\mu\ p_{n+1}(t),\ n \geq 2$$

and again the terms that contain $p_0(t)$ were dropped from the first two equations, because zero is the absorbing state.

In order to solve the above system of equations, we need the following generating function,

$$G(z,t) = \sum_{n=1}^{\infty} z^n\ p_n(t) \tag{3.30}$$

for which the derivative w.r.t. t is,

$$\frac{d}{dt}\ G(z,t) = \sum_{n=1}^{\infty} z^n\ \frac{dp_n(t)}{dt}\ . \tag{3.31}$$

Considering $\frac{dp_n(t)}{dt}$ from the above system of equations after multiplying the second equation by z and the third by z^n and summing over all the possible values of n. After some algebraic manipulations, this will give us the following result,

$$\frac{dG(z,t)}{dt} = -\lambda\ G(z,t) - 2\mu\ G(z,t) + \mu\ z\ p_1(t) +$$

$$\frac{2\mu}{z}\ [G(z,t)-z\ p_1(t)]\ + \lambda\ z\ G(z,t). \tag{3.32}$$

Moreover let us assume that the initial number of customers at time t = 0 is one, that is,

$$G(z,0) = z,. \tag{3.33}$$

Using the initial condition (3.33) and taking the Laplace transform of (3.32) yields:

$$\hat{G}(z,s) = \frac{- z^2 - \mu z^2 \ \hat{p}_1(s) + 2\mu \ z \ \hat{p}_1(s)}{\lambda z^2 - (s+\lambda+2\mu) \ z + 2\mu} . \tag{3.34}$$

Applying Rouche's theorem to (3.34) gives the appropriate root for the denominator as

$$\hat{\phi}(s) = \frac{(s+\lambda+2\mu) - \sqrt{(s+\lambda+2\mu)^2-8\lambda\mu}}{2\lambda} .$$

And by equating the numerator to zero in (3.34), when $z = \phi$, the appropriate root gives,

$$\hat{f}_B(s) = \mu \ \hat{p}_1(s) = \frac{\hat{\phi}(s)}{2 - \hat{\phi}(s)} \tag{3.35}$$

which is the Laplace transform of the p.d.f. of the busy period for the M/M/2 queue, in (3.13).

Similarly, the difference–differential equations for the M/M/c queue are

$$\frac{dp_0(t)}{dt} = \mu \ p_1(t)$$

$$\frac{dp_n(t)}{dt} = - (\lambda+n\mu) \ p_n(t) + \lambda \ p_{n-1}(t) + (n+1) \ \mu \ p_{n+1}(t), \qquad 1 \leq n < c$$

$$\frac{dp_n(t)}{dt} = - (\lambda+c\mu) \ p_n(t) + \lambda \ p_{n-1}(t) + c\mu \ p_{n+1}(t), \qquad \text{for } n \geq c$$

where $\lambda \ p_{n-1}(t)$ is to be excluded whenever $n = 1$ in the second equation. Again, and as we did before in the M/M/2, the first equation constitutes the p.d.f. of the busy period. Equation two on the other hand could be rewritten as

$$\frac{dp_n(t)}{dt} = -(\lambda + c\mu)\ p_n(t) + (c-n)\ \mu\ p_n(t) +$$

$$\lambda\ p_{n-1}(t) + c\mu\ p_{n+1}(t) - (c-n-1)\ \mu\ p_{n+1}(t),$$

$$1 \leq n < c \qquad (3.36)$$

which in this form is quite similar to equation three. Now the generating function equivalent of this becomes

$$\frac{d}{dt}\ G(z,t) = -(\lambda + c\mu)\ G(z,t) + \lambda\ z\ G(z,t) +$$

$$\frac{c\mu}{z}\ [G(z,t) - z\ p_1(t)] + \mu\,(c-1)\ z\ p_1(t)$$

$$-\mu \left[\frac{1}{z} - 1\right] \sum_{n=2}^{c-1} (c-n)\ z^n\ p_n(t). \qquad (3.37)$$

Taking the Laplace transform of both sides of (3.37), gives

$$-G(z,0) + s\ \hat{G}(z,s) = -(\lambda + c\mu)\ \hat{G}(z,s) + \lambda z\ \hat{G}(z,s)$$

$$+ \frac{c\mu}{z}\ [\hat{G}(z,s) - z\ \hat{p}_1(s)] + \mu\,(c-1)\ z\ \hat{p}_1(s)$$

$$-\mu \left[\frac{1}{z} - 1\right] \sum_{n=2}^{c-1} (c-n)\ z^n\ \hat{p}_n(s). \qquad (3.38)$$

And at the beginning of the busy period if there are i customers in the system at time $t = 0$, then

$$G(z,0) = z^i. \qquad (3.39)$$

While just before the busy period terminates, there should be only one customer remaining in the system and that is why

$$f_B(t)\Delta t = P\{t \leq B \leq t + \Delta t\} = \mu\Delta t \; p_1(t)$$

and

$$\hat{p}_n(s) = \varepsilon \cdot \delta_{n,1} \tag{3.40}$$

where $\varepsilon > 0$.

Using (3.39) and (3.40) into (3.38) and with some rearrangement, we get,

$$\hat{G}(z,s) = \frac{-z^{i+1} + c\mu \, z\hat{p}_1(s) - \mu(c-1) \, z^2\hat{p}_1(s)}{\lambda z^2 - (s+\lambda+c\mu)z + c\mu} \tag{3.41}$$

and applying Rouche's theorem will get the appropriate root for the denominator as

$$\hat{\phi}(s) = \frac{(s+\lambda+c\mu) - \sqrt{(s+\lambda+c\mu)^2 - 4c\lambda\mu}}{2\lambda} \tag{3.42}$$

which is the same Laplace transform of the p.d.f. of the M/M/1 queue busy period with μ being replaced $c\mu$. It is the Laplace transform of the busy period if it is defined as the total occupation of the c servers available.

However, equating the numerator of (3.41) to zero, at the point where $z = \phi$, we will get

$$\phi^i - c\mu \; \hat{p}_1(s) + \mu(c-1) \; \phi \; \hat{p}_1(s) = 0 \tag{3.43}$$

which could be written as

$$\hat{f}_B(s) = \mu\ \hat{p}_1(s) = \frac{\phi^i}{c-(c-1)\phi} \tag{3.44}$$

which represents the Laplace transform of the p.d.f. of the busy period for the M/M/c queue. It should be noted that when i = 1 and c = 1 then (3.44) becomes

$$\hat{f}_B(s) = \mu\ \hat{p}_1(s) = \phi$$

where ϕ as defined in (3.42) with c = 1, is the Laplace transform of the p.d.f. for the M/M/1 queue as obtained previously in (2.42). Similarly for i = 1 and c = 2, (3.44) gives

$$\hat{f}_B(s) = \mu\ \hat{p}_1(s) = \frac{\phi}{2-\phi}$$

which exactly the Laplace transform of the p.d.f. of the busy period of the M/M/2 queue in (3.35).

An important derivation from (3.44) is the average length of the busy period, it is now possible to find a general expression for E(B) as follows,

$$E(B) = \left.\frac{-\partial \hat{f}_B(s)}{\partial s}\right|_{s=0} = \frac{i+c-1}{c\mu-\lambda} \tag{3.45}$$

Hence for the M/M/c queue, if the initial number of customers is one, i = 1 then

$$E(B) = \frac{c}{c\mu-\lambda} = \frac{1}{\mu(1-\rho_c)},\ \rho_c = \frac{\lambda}{c\mu} \tag{3.46}$$

and,

$$\mathrm{Var}(B) = c\ \phi''(0) + 2c(c-1)[\phi'(0)]^2 - [c\phi'(0)]^2. \tag{3.47}$$

CHAPTER IV

THE M/G/1 AND THE G/M/1 QUEUES

THE M/G/1 QUEUE:

The features of this queue are that customers arrive randomly according to a Poisson process and obtain service from a single server, who follows an independent general service time distribution. The pioneer work in this area was done by Erlang [6], Pollaczek [11], and Khintchine [9]. Though Takacs [17] gave an explicit form of the busy period density function, his opening remarks considered a complicated combinatorial analysis technique. However, for this type of queue and the G/M/1, the Chapman–Kolomogorov difference – differential equations, as illustrated previously, are not possible due to the fact of the relaxation of the Poisson and exponential distributions within the arrival and service times, respectively. However, for the G/M/1, a modification for the equations is done by Cox [5] by using what is known as the supplementary variables technique, which is out of the scope of my thesis.

The direct approach is to model the M/G/1 queue as a Markov chain, and needs a complete description of the state of the system at any particular moment of time. Therefore, we require information regarding the number of customers in the system as well as the remaining or elapsed service time of the customer currently being served. To understand the general behavioral pattern of this system, one can study it at some discrete points in order to simplify the analysis. For example, if the time interval (0,t) is split at the points of departure of customers from the system, that is, t_1, t_2, ...

At these points, the remaining service time of the customer is zero, and therefore the queue length could be studied independently. Accordingly, let us consider the system exactly at the moment when one particular customer's service is completely done (say, the k^{th} customer), and the next service (of the $(k+1)^{th}$ customer) is about to commence. For sake of convenience, allow me to adopt and define the following notations,

N_k = number of customers in the queue at the completion of the k^{th} service.

L_{k+1} = number of arrivals during the $(k+1)^{th}$ service.

$h(t)$ = the probability density function of the i.i.d. service times.

We can now write

$$N_{k+1} = \begin{cases} N_k + L_{k+1} - 1, & \text{if} \quad N_k > 0 \\ L_{k+1} \quad , & \text{if} \quad N_k = 0 \end{cases} .$$

So that $\{N_k\}$ has the properties of a Markov chain, and it is called the imbedded Markov chain, Kendall [8].

In order to attain the M/G/1 busy period distribution, we will partition $f_1(k,t)$ in terms of the outcome of the first service. Therefore one can write $f_1(k,t)$ as

$$f_1(k,t) = \sum_{r=0}^{\infty} \left[\frac{(\lambda t)^r e^{-\lambda t}}{r!} h(t)\right] * f_r(k-1,t) \tag{4.1}$$

where $f_0(k-1,t) = \delta_{k,1}\, \delta(t)$. Then the Laplace transform of (4.1) is,

$$\hat{f}_1(k,s) = \sum_{r=0}^{\infty} \mathscr{L}\left[\frac{(\lambda t)^r e^{-\lambda t}}{r!} h(t)\right] \hat{f}_r(k-1,s). \tag{4.2}$$

Using the generating function of (2.36), then (4.2) becomes

$$\phi_1(z,s) = z \sum_{r=0}^{\infty} \mathscr{L}\left[\frac{(\lambda t)^r e^{-\lambda t}}{r!} h(t)\right] \phi_r(z,s). \tag{4.3}$$

From (2.39) applying the Bench method argument we have that

$$\phi_r(z,s) = (\phi_1(z,s))^r. \tag{4.4}$$

And, substituting (4.4) into (4.3), and performing the sum, one yields,

$$\phi_1(z,s) = z \mathscr{L}\left[e^{-\lambda(1-\phi_1(z,s))t} h(t)\right] \tag{4.5}$$

or

$$\phi_1(z,s) = z\, \hat{h}(s + \lambda(1-\phi_1(z,s))) \tag{4.6}$$

which is the Laplace transform of the bivariate density of the busy period, B, and K, the number served during such period; which was obtained by identifying sub–busy periods within the main busy period, and all of them having the same distribution as the main one.

In order to derive the moments of B and K, first, let us assume that the moments of the arbitrary general service time distribution exist as

$$E(X^k) = (-1)^k \hat{h}^{(k)}(0)$$

where the mean is denoted by

$$\alpha = E(X) = -\hat{h}'(0)$$

and the traffic intensity is therefore $\rho = \lambda\alpha$. Hence, from (4.6),

$$E(B) = \left. \frac{-\partial \phi_1(z,s)}{\partial s} \right|_{\substack{z=1 \\ s=0}} = \alpha(1 + \lambda\ E(B)) \text{ or } E(B) = \frac{\alpha}{1-\rho}. \quad (4.7)$$

It should be noted that for the M/M/1 queue, $\alpha = \frac{1}{\mu}$ for the exponential distribution, therefore

$$E(B) = \frac{1}{\mu(1-\rho)}$$

as obtained earlier in (2.63). In the same way, the average number of customers served during the busy period is

$$E(K) = \left. \frac{\partial \phi_1(z,s)}{\partial z} \right|_{\substack{z=1 \\ s=0}} = 1 + \alpha\lambda\ E(K) = \frac{1}{1-\rho}. \quad (4.8)$$

the same result for the M/M/1 queue as (2.65) with $\alpha = \frac{1}{\mu}$. Moreover, the variance of B is

$$\text{Var}(B) = \frac{\sigma_x^2 + \rho\ \alpha^2}{(1-\rho)^3} \quad (4.9)$$

where σ_x^2 represents the variance of the service time distribution. Similarly the variance of K is

$$\text{Var}(K) = \frac{\lambda^2\ E(X^2) + (1-\rho)(2\rho-1)}{(1-\rho)^3} \quad (4.10)$$

and

$$E(BK) = \frac{\alpha(1-\rho) + \lambda\ E(X^2)}{(1-\rho)^3}$$

which makes the covariance of B and K,

$$\text{Cov}(B,K) = \frac{\lambda\ E(X^2)}{(1-\rho)^3} > 0 \text{ as expected.} \quad (4.11)$$

Going back to invert the transformed generating function, we can use the Lagrange expansion on (4.6). This implies that we have a function $\mathscr{S}$ such that

$$\mathscr{S} = a + t\ \psi(\mathscr{S}) \tag{4.12}$$

then,

$$\mathscr{S} = a + \sum_{n=1}^{\infty} \frac{t^n}{n!} \frac{d^{n-1}}{d\ a^{n-1}} (\psi(a))^n . \tag{4.13}$$

Now, let $\psi(s) = \hat{h}(s)$, $a = s+\lambda$, $t = -\lambda z$, and $\mathscr{S} = s + \lambda(1-\phi_1(z,s))$. Then (4.6) is in the form of (4.12), so that (4.13) will be

$$\mathscr{S} = s+\lambda - \lambda\phi_1(z,s) = s+\lambda + \sum_{n=1}^{\infty} \frac{(-1)^n(\lambda z)^n}{n!} \frac{d^{n-1}(\hat{h}(s+\lambda))^n}{d(s+\lambda)^{n-1}}. \tag{4.14}$$

Now,

$$(\hat{h}(s))^n = h_n(s)$$

where,

$$\hat{h}_n(s) = \mathscr{L}[h_n(t)]$$

where $h_n(t)$ is the n–fold convolution of $h(t)$ with itself. So, (4.14) will be

$$\phi_1(z,s) = \sum_{n=1}^{\infty} \frac{(-\lambda)^{n-1} z^n}{n!} \int_0^{\infty} \frac{d^{n-1} e^{-(s+\lambda)t}}{d(s+\lambda)^{n-1}} h_n(t)dt$$

$$= \sum_{n=1}^{\infty} \frac{\lambda^{n-1} z^n}{n!} \int_0^{\infty} e^{-(s+\lambda)t}\ t^{n-1}\ h_n(t)dt . \tag{4.15}$$

From (4.15), one can readily find $f_1(k,t)$, namely,

$$f_1(k,t) = \frac{(\lambda t)^{k-1} e^{-\lambda t}}{k!} h_k(t) . \tag{4.16}$$

This agrees with the M/M/1 result of (2.45) with

$$h(t) = \mu e^{-\mu t}.$$

Then, $h_k(t)$ is the k–fold convolution of $h(t)$,

$$h_k(t) = \frac{\mu(\mu t)^{k-1} e^{-\mu t}}{(k-1)!} . \tag{4.17}$$

The marginal distributions of (4.16) are obtained in the usual way.

<u>THE G/M/1 QUEUE</u>:

In this single server queuing system the service times have a negative exponential distribution but the i.i.d. arrival times follow an arbitrary distribution which has a c.d.f.,

$$A(t), \qquad 0 \leq t < \infty$$

with a p.d.f. $a(t)$. Again, we could identify an imbedded Markov chain considering the arrival points as the discrete set of points. Then,

N_k = number of customers in the system ahead of a newly arrived customer.

L_{k+1} = number of customers served between the arrival of the k^{th} customer and the $(k+1)^{th}$.

As the case in M/G/1, then

$$N_{k+1} = N_k + 1 - L_{k+1} \quad \text{for all } N_k.$$

This makes $\{N_k\}$ as a Markov chain with renewal points taken at the instances of customers arrival. That is, the G/M/1 is modeled by considering the system immediately after an arrival occurs. The busy period distribution had been covered extensively through different approaches; an interested reader may consult Takacs [18], Kendall [8], Prabhu [13], and Conolly [4]. The p.d.f. of the busy period is considered in terms of one arrival and (r–1) services being completed or, no arrivals occurred and there are r services being done. Therefore,

$$f_r(t) = \sum_{k=0}^{r-1}\left[\frac{(\mu t)^k e^{-\mu t}}{k!} a(t)\right] * f_{r-k+1}(t) + \frac{\mu(\mu t)^{r-1} e^{-\mu t}}{(r-1)!} \bar{A}(t) \qquad (4.18)$$

where $\bar{A}(t) = P(T > t)$ is the probability of no arrivals during time t, or arrivals have occurred after time t.

Now, taking the Laplace transform of both sides of (4.18),

$$\hat{f}_r(s) = \mathscr{L}\left[\sum_{k=0}^{r-1}\frac{(\mu t)^k e^{-\mu t}}{k!} a(t)\right]\hat{f}_{r-k+1}(s) + \mathscr{L}\left[\frac{\mu(\mu t)^{r-1} e^{-\mu t}}{(r-1)!} \bar{A}(t)\right]. \quad (4.19)$$

The Laplace transform of the necessary generating function is

$$\hat{\phi}(z,s) = \sum_{r=1}^{\infty} z^r \hat{f}_r(s) \tag{4.20}$$

and,

$$\mathscr{L}[\bar{A}(t)] = \frac{1 - \hat{a}(s)}{s}. \tag{4.21}$$

Hence applying (4.20) and (4.21) to the second term on the R.H.S. of equation (4.19) yields

$$\mathscr{L}\left[\mu z \sum_{r=1}^{\infty} \frac{(\mu zt)^{r-1}}{(r-1)!} e^{-\mu t} \bar{A}(t)\right] = \mathscr{L}\left[\mu z \, e^{-\mu t(1-z)} \bar{A}(t)\right]$$

$$= \frac{\mu z(1 - \hat{a}(s+\mu(1-z)))}{s+\mu(1-z)}. \tag{4.22}$$

The first term of the R.H.S. of (4.19) can be written as follows,

$$\sum_{r=1}^{\infty} \sum_{k=0}^{r-1} z^r \mathscr{L}\left[\frac{(\mu t)^k e^{-\mu t}}{k!} a(t)\right] \hat{f}_{r-k+1}(s) =$$

$$= \sum_{k=0}^{\infty} \sum_{r=k+1}^{\infty} z^r \mathscr{L}\left[\frac{(\mu t)^k e^{-\mu t}}{k!} a(t)\right] \hat{f}_{r-k+1}(s) \tag{4.23}$$

that is after changing the order of the two sums as

$$\sum_{r=1}^{\infty}\sum_{k=0}^{r-1} a_{rk} = \sum_{k=0}^{\infty}\sum_{r=k+1}^{\infty} a_{rk}$$

then (4.23) can be written as

$$\sum_{k=0}^{\infty}\sum_{m=1}^{\infty} z^{m+k}\, \mathscr{L}\left[\frac{(\mu t)^k e^{-\mu t}}{k!}\, a(t)\right]\, \hat{f}_{m+1}(s)$$

$$= \frac{1}{z}\, \mathscr{L}\left[\sum_{k=0}^{\infty} \frac{(\mu z t)^k e^{-\mu t}}{k!}\, a(t)\right] \sum_{m=1}^{\infty} z^{m+1}\, \hat{f}_{m+1}(s)$$

$$= \frac{1}{z}\, [\hat{\phi}(z,s) - z\, \hat{f}_1(s)]\, \hat{a}(s+\mu(1-z)). \tag{4.24}$$

So, if we combine the whole expression from (4.19), (4.22), and (4.24), we get

$$\hat{\phi}(z,s) = \frac{\hat{a}(s+\mu(1-z))}{z}\, [\hat{\phi}(z,s) - z\, \hat{f}_1(s)] + \frac{\mu z(1-\hat{a}(s+\mu(1-z)))}{s+\mu(1-z)}. \tag{4.25}$$

And, suppose we let $s' = s + \mu(1-z)$ and rearranging (4.25),

$$\hat{\phi}(z,s) = \frac{\mu z^2(1-\hat{a}(s')) - \hat{a}(s')\, \hat{f}_1(s)\, zs'}{(z-\hat{a}(s'))\, s'}. \tag{4.26}$$

Again by applying Rouche's theorem, denominator of (4.26) gives

$$z_0 = \hat{a}(s') = \hat{a}(s + \mu(1-z_0)) \tag{4.27}$$

on the other hand, the numerator yields,

$$\hat{f}_1(s) = \frac{\mu(1-z_0)}{s+\mu(1-z_0)} \tag{4.28}$$

which is the Laplace transform of the p.d.f. of the busy period for the G/M/1 queue.

The average length of the busy period is

$$E(B) = \left.\frac{-d\hat{f}_1(s)}{ds}\right|_{s=0} = \frac{1}{\mu[1 - \hat{a}(\mu(1-z_0))]} \,. \tag{4.29}$$

CHAPTER V

THE G/G/1 QUEUE

In this queueing system, the interarrival times are independent and identically distributed with an arbitrary distribution, A(t). The service times are independent and have an identical arbitrary distribution denoted by H(t). There is a single server offering service on a FIFO basis. In order to model the G/G/1 queue, we need to split the time interval (0,t) at the points of customers arrival, t_n, namely, t_0, t_1, t_2, ... Also if y_n represents the n^{th} customer service completion time, and, x_n, is his service duration time. Then

$$A(t) = P(t_n - t_{n-1} < t) \tag{5.1}$$

and,

$$H(t) = P(y_n - y_{n-1} < t). \tag{5.2}$$

THE WIENER–HOPF DECOMPOSITION:

The probability generating function of K, the number served in a busy period is derived via the Wiener–Hopf decomposition, where the complex plane is divided into two halves H_γ^+ and H_γ^-, such that if $s = \sigma + i\tau$ is a line in the complex plane, then

$$H_\gamma^+ = \{s: \ \sigma \leq \gamma\}$$

and,

$$H_\gamma^- = \{s: \sigma \geq \gamma\}.$$

At this point let me introduce a common distribution function, $F(x)$, for which the n–fold convolution is $F_n(x)$. Moreover, let the following transform to be defined as follows.

$$\phi(s) = \int_{-\infty}^{\infty} e^{sx}\, dF(x) \tag{5.3}$$

where $s = \sigma + i\tau$, σ, τ are real. The transform of the n–fold convolution, therefore, is

$$\phi^n(s) = \int_{-\infty}^{\infty} e^{sx}\, dF_n(x). \tag{5.4}$$

If t denotes a fixed complex number, then γ which separates the half planes H_γ^- and H_γ^+ is chosen so that

$$\phi(\gamma)\ |t| < 1, \qquad \phi(\gamma) < \infty.$$

If two functions are defined and analytic in H_γ^- and H_γ^+ respectively, then they are analytic extensions of each other whenever they have the same value at each point s with $\mathrm{Re}(s) = \gamma$. Define

$$f_n^-(s) = \int_{-\infty}^{0-} e^{sx}\, dF_n(x) \tag{5.5}$$

which is analytic in H_γ^-, and

$$f_n^+(s) = \int_{0-}^{\infty} e^{sx}\, dF_n(x) \tag{5.6}$$

which is analytic in H_γ^+. Another two functions corresponding to each one above could be defined as

$$L^-(t,s) = \sum_{n=1}^{\infty} \frac{t^n}{n} f_n^-(s) \tag{5.7}$$

and,

$$L^+(t,s) = \sum_{n=1}^{\infty} \frac{t^n}{n} f_n^+(s). \tag{5.8}$$

Similarly for each above, let

$$M^-(t,s) = e^{-L^-(t,s)} \tag{5.9}$$

and,

$$M^+(t,s) = e^{L^+(t,s)}. \tag{5.10}$$

Now, at this point we need to state the following theorem.

<u>Theorem 5.1</u>:

On the boundary γ, $\mathrm{Re}(s) = \gamma$, we have

(a) $(1 - t\ \ \phi(s))\, M^+(t,s) = M^-(t,s)$

(b) $\lim_{\sigma \to \infty} M^-(t,s) = 1$

(c) $M^-(t,s)$ and $M^+(t,s)$ are uniquely determined by (a) and (b).

Proof:

By definition of the function in (5.4) we have

$$\phi^n(s) = \int_{-\infty}^{\infty} e^{sx}\, d_nF(x) = f_n^-(s) + f_n^+(s). \tag{5.11}$$

Similarly by the definitions in (5.6) – (5.8), we can write

$$\sum_{n=1}^{\infty} \frac{t^n}{n} \phi^n(s) = L^-(t,s) + L^+(t,s) = -\ell n(1 - t\,\phi(s)). \tag{5.12}$$

Then,

$$-L^-(t,s) = L^+(t,s) + \ell n(1 - t\,\phi(s)). \tag{5.13}$$

Taking the exponential function of both sides of (5.13), we get,

$$M^-(t,s) = M^+(t,s)\,(1 - t\,\phi(s)). \tag{5.14}$$

which is the proof of the first part of the theorem. For the second part, it is clear from (5.5), (5.7), and (5.9) that

$$\lim_{\sigma\to\infty} M^-(t,s) = 1. \tag{5.15}$$

For part (c), we need to state the following theorem:

Theorem 5.2:

Liouvilles' theorem states that if $\alpha(s)$ is any entire function (differentiable everywhere) and $|\alpha(s)|$ is bounded for every s in the complex plane then $\alpha(s)$ is a constant. (For proof consult any standard text on complex variables).

One may have to show that $M^-(t,s)$ and $M^+(t,s)$ are uniquely defined by (5.14). Namely assume M^- and M^+ relations also hold for A^- and A^+. Then:

$$\frac{A^+}{M^+} = \frac{A^-}{M^-} \text{ when } \sigma = \gamma .$$

Hence define
$$g(s) = \frac{A^+(s)}{M^+(s)} \text{ on } s \in H_\gamma^+$$
$$= \frac{A^-(s)}{M^-(s)} \text{ on } s \in H_\gamma^-.$$

Then g(s) is a bounded entire function. Now $g(s) \longrightarrow 1$ as $\sigma \longrightarrow \infty$ by (5.15). Therefore $g(s) \equiv 1$ by Liouvilles' theorem. Hence M^- and M^+ are unique.

QUEUEING APPLICATION OF THE DECOMPOSITION:

We will use the Wiener–Hopf decomposition to derive the probability generating function of K (the number served) in a G/G/1 busy period. If we denote the beginning of a busy period as the event ε, then ε is a persistent recurrent event for a stationary process.

The first customer arrives at time $t_0 = 0$ and has his service completed at $y_1 > 0$. Similarly the n^{th} customer arrives at t_{n-1} and has his service completion at y_n where:

$$T_n = t_n - t_{n-1}, \qquad t_0 = 0$$

$$Y_n = y_n - y_{n-1}, \qquad y_0 = 0.$$

Let,

$$X_n = Y_n - T_n$$

and,

$$Z_n = \sum_{i=1}^{n} X_i = y_n - t_n.$$

Therefore if $K = \inf\{n: z_n < 0\}$, then K denotes the number of customers served in the G/G/1 busy period. And z_n is a process of independent increments where $X_n = z_n - z_{n-1}$ and $z_0 = 0$. Define

$$p_n(s) = \int_{-\infty}^{0-} p\{K=n,\ z_n < 0\}\ e^{sz_n}\ dz_n \text{ on } H_\gamma^-$$

and,

$$q_n(s) = \int_{0-}^{\infty} p\{K \geq n,\ z_n \geq 0\}\ e^{sz_n}\ dz_n \text{ on } H_\gamma^+$$

where

$$K = \begin{cases} \text{number served in the busy period} \\ \infty, \text{ if the busy period never ends} \end{cases}.$$

Then

$$p_n(s) + q_n(s) = q_{n-1}(s)\ \phi(s) \quad \text{if} \quad \sigma = \gamma \quad \text{and} \quad n \geq 1. \qquad (5.16)$$

Also define:

$$p(t,s) = \sum_{n=0}^{\infty} p_n(s)\ t^n$$

and

$$q(t,s) = \sum_{n=0}^{\infty} q_n(s)\ t^n$$

where

$$p_0(s) = 0 \quad \text{and} \quad q_0(s) = 1.$$

Then multiplying (5.16) by t^n and summing, one finds:

$$p(t,s) + q(t,s) - 1 = t\ \phi(s)\ q(t,s)$$

or

$$1 - p(t,s) = (1 - t\ \phi(s))\ q(t,s) \quad \text{on} \quad \sigma = \gamma.$$

Therefore from (5.14),

$$1 - p(t,s) = \frac{M^-(t,s)}{M^+(t,s)}\ q(t,s). \tag{5.17}$$

Let us define the entire bounded function:

$$g(s) = \frac{1 - p(t,s)}{M^-(t,s)} \quad \text{on} \quad H_\gamma^-$$

and

$$g(s) = \frac{q(t,s)}{M^+(t,s)} \quad \text{on} \quad H_\gamma^+.$$

Then by Theorem 5.1,

$$\lim_{\sigma \to \infty} g(s) = 1, \text{ hence } g(s) \equiv 1$$

for all s. Hence

$$p(t,s) = 1 - M^{-}(t,s) \quad \text{on} \quad H_{\gamma}^{-}.$$

The probability generating function of K is:

$$G(t) = \sum_{k=1}^{\infty} t^k \, P(K=k) = p(t,0).$$

Hence

$$G(t) = 1 - M^{-}(t,0). \tag{5.18}$$

Now

$$\phi(s) = E[e^{sX_n}] = E[e^{sY_n - sT_n}]$$
$$= \hat{h}(s) \, \hat{a}(-s). \tag{5.19}$$

The sequence of T_n and Y_n are i.i.d. random variables with distributions A(t) and H(t) respectively and their corresponding transforms are

$$\hat{h}(s) = \int_{-\infty}^{\infty} e^{sx} \, d \, H(x)$$

and,

$$\hat{a}(s) = \int_{-\infty}^{\infty} e^{sx} \, d \, A(x).$$

Now A(t) and H(t) with the corresponding densities a(t) and h(t) are zero if $t < 0$. Hence,

$$f_n(t) = \int_{u=\max(-t,0)}^{\infty} h_n(t+u)\ a_n(u)\ du$$

and

$$\hat{f}_n^-(s) = \int_{-\infty}^{0} f_n(t)\ e^{st}\ dt$$

$$= \int_{-\infty}^{0} e^{st} \int_{u=-t}^{\infty} h_n(t+u)\ a_n(u)\ du.$$

Rearranging the integral yields:

$$\hat{f}_n^-(s) = \int_0^{\infty} e^{-st}\ dt \int_0^{\infty} h_n(v)\ a_n(v+t)\ dv$$

or

$$\hat{f}_n^-(0) = \int_0^{\infty} \bar{A}_n(v)\ h_n(v)\ dv.$$

One can therefore write (5.18) as

$$G(t) = 1 - \exp\left[- \sum_{n=1}^{\infty} \frac{t^n}{n} \int_0^{\infty} \bar{A}_n(v)\ h_n(v)\ dv \right]$$

which represents the generating function of K, the number served during the busy period for the G/G/1 queue.

TABLE OF IMPORTANT RESULTS

Queue \ Results	densities	means	Variances	Covariances
M/M/1	$f_B(t) = \frac{1}{t} e^{-(\lambda+\mu)t} \sqrt{\frac{\mu}{\lambda}}\, I_1(2t\sqrt{\lambda\mu})$ $P(K=k) = \frac{(2k-2)!}{k!(k-1)!} \frac{\lambda^k \mu^k}{(\lambda+\mu)^{2k-1}}$	$E(B) = \frac{1}{\mu(1-\rho)}$ $E(K) = \frac{1}{1-\rho}$	$Var(B) = \frac{1+\rho}{\mu^2(1-\rho)^3}$ $Var(K) = \frac{3\rho-1}{(1-\rho)^3}$	$Cov(B,K) = \frac{2\rho}{\mu(1-\rho)^3}$
M/M/2	$\hat{f}_B(s) = \frac{\phi}{2-\phi}$	$E(B) = \frac{1}{\mu(1-\rho_2)}$ $E(K) = \frac{1+\rho_2}{1-\rho_2}$	$Var(B) = \frac{1}{\mu^2(1-\rho_2)^3}$	$Cov(B,K) = \frac{\rho_2(3-\rho_2)}{\mu(1-\rho_2)^3}$
M/M/C	$\hat{f}_B(s) = \frac{\phi^i}{c-(c-1)\phi}$	$E(B) = \frac{i+c-1}{c\mu-\lambda}$	$Var(B) = c\,\phi''(0)+2c(c-1)$ $[\phi'(0)]^2-[c\phi'(0)]^2$	
M/G/1	$f_1(k,t) = \frac{(\lambda t)^{k-1} e^{-\lambda t}}{k!} h_k(t)$ is the bivariate density of B and K.	$E(B) = \frac{\alpha}{1-\rho}$ $E(K) = \frac{1}{1-\rho}$	$Var(B) = \frac{\sigma_X^2+\rho\alpha^2}{(1-\rho)^3}$ $Var(K) = \frac{\lambda^2 E(X^2)+(1-\rho)(2\rho-1)}{(1-\rho)^3}$	$Cov(B,K) = \frac{\lambda E(x^2)}{(1-\rho)^3}$
G/M/1	$\hat{f}_1(s) = \frac{\mu(1-z_0)}{s+\mu(1-z_0)}$	$E(B) = \frac{1}{\mu(1-\hat{a}(\mu(1-z_0)))}$		
G/G/1	$G(t) = 1 - \exp\left[-\sum_{n=1}^{\infty} \frac{t^n}{n} \int_0^{\infty} h_n(v)\, \bar{A}_n(v)\, dv\right]$			

BIBLIOGRAPHY

[1] Bailey, N.T.J., "A continuous time treatment of a simple queue using generating functions". J. Roy. Statist. Society, Ser. B, vol. 16 (1954), pp.288–291.

[2] Champernowne, D.G., "An elementary method of the solution of the queueing problem with a single server and a constant parameter". J. Roy. Statist. Society, Ser. B, vol. 18 (1956), pp.125–128.

[3] Chaudry, M.L. and Templeton, J.G.C., "A note on the distribution of a busy period for the M/M/c queue system". Math. Operationsforsch Statist. 4, (1973), pp. 75–79.

[4] Conolly, B.W., "The busy period distribution in relation to the queuing process GI/M/1". Biometrika vol. 46 (1959), pp. 246–251.

[5] Cox, D.R., "The analysis of non–Markovian stochastic processes by the inclusion of supplementary variables". Proc. Cambridge Philos. Soc. 51 (1955), pp.433–441.

[6] Erlang, A.K., "Probability and Telephone Calls". Nyt. Tidsskr., Mat., Ser. B, vol 20 (1909), pp.33–39.

[7] Feller, W., "An Introduction to Probability Theory and its Applications". Vol. 1, 3rd ed. (1968), John Wiley and Sons.

[8] Kendall, D.G., "Some problems in the theory of queues". J. Roy. Statist. Soc., B 13 (1951), pp. 151–185.

[9] Khintchine, A.Y., "Mathematical Methods in the Theory of Queueing". Griffin, London 1960).

[10] Palm, C., "Some investigations in waiting times in telephone plants". Tek. Medd. Fran. Kugnl. Teleg. Vol. 7–9 (1937) Swedish.

[11] Pollaczek, F., in french, Memoriale de sci. Math. (1957), Paris, no. 136.

[12] Prabhu, N.U., "Some results for the queue with poisson arrivals". J. Roy. Statist. Soc., Ser. B, Vol. 22 (1960), pp. 104–107.

[13] Prabhu, N.U., "Queues and Inventories". Wiley (1965).

[14] Saaty, T.L., "Elements of Queueing Theory with Applications". (1961), Mcgraw–Hill Co.

[15] Saaty, T.L., "Time–dependent solution of the many server Poisson queue". Oper. Res. 8, (1960), pp. 755–771.

[16] Takacs, L., "The probability law of the busy period for two types of queueing processes". Oper. Res. 9, (1961), pp. 402–407.

[17] Takacs, L., "A single server queue with poisson input". Oper. Res. 10, (1962), pp. 388–394.

[18] Takacs, L., "A single server queue with recurrent input and exponentially distributed service times". Oper. Res. 10, (1960), pp.395–399.