1975

Analysis & Applications of the Delay Cycle for the M/M/c Queueing System

K. Omahen

V. Marathe

Report Number:
75-165

Analysis & Applications of the Delay Cycle for the M/M/c Queueing System

K. Omahen
Department of Computer Science
Purdue University
W. Lafayette, IN 47907

and

V. Marathe
Department of Business Administration
University of Illinois
Urbana, IL 61801

CSD-TR 165

Keywords: Delay Cycle Analysis, Busy Period Analysis, M/M/c Queueing System, Waiting Time Distributions, Multiprocessor Models.

1. Introduction

This paper addresses the problem of solving multiple-server queueing models through the use of busy-period analyses. In the past, this technique of busy-period analysis has been extensively used for the M/G/1 queueing system and found to be a powerful tool for dealing with a great many complex scheduling rules, particularly those involving preemptive and nonpreemptive priorities. The results presented in this paper demonstrate that the method of busy-period analysis can be extended to deal with certain instances of the M/M/c queueing system as well. In references [1,2], the authors independently made use of this general technique to solve multiple-server models for which jobs have simultaneous resource requirements; an overview of this work is contained in reference [3].

The method of analysis involving decomposition of busy periods was first used by Cobham [4] and later by Avi-Itzhak, Maxwell, and Miller [5] and others; the text by Conway, Maxwell, and Miller [6] contains many of the results obtained for the M/G/1 queueing system through the use of this technique. In order to motivate the analysis to be presented, the typical sequence of steps in such a busy-period analysis will be briefly covered.

When the term "busy-period" is used, we will be referring to an interval during which one or more jobs are in service. There will generally be a number of different busy-periods for a system, and system states will be defined so as to include the idle state and a number of mutually exclusive and exhaustive busy states corresponding to these different types of busy-periods. The system states are defined so that the state transition process is Markovian, and the limiting probability of the system being in a particular state may be determined in a straightforward manner (e.g. using results from the theory of Semi-Markov processes [7]). The analysis typically begins by first finding the Laplace-Stieltjes Transform (LST) for the distribution of busy period length for each case and then obtaining the LST for the flow time or waiting time conditioned upon the arrival finding a specified type of busy-period in progress. If jobs are generated by means of a Poisson source, the distribution of system states at arrival epochs will be identical to the steady-state distribution of system states (see Strauch [8]). Using this result, the unconditional LST for the distribution of flow time or waiting time may be directly obtained. In principle, the LST completely describes a distribution, but the inversion of the transform is usually so difficult as to require that a numerical transform inversion software package be employed. In most cases the first two moments of a random variable may be found from the LST without excessive effort.

The system to be analyzed in this paper is the multiple-server Poisson-Exponential queue, commonly denoted as the M/M/c queueing system. The model may be described as follows:

> There are c identical servers, each capable of performing one job at a time. The incoming jobs have independent and exponentially distributed processing requirements with mean $1/\mu$. The inter-arrival times are independent and exponentially distributed with mean $1/\lambda$. It is assumed that the servers process the jobs continuously as long as there are jobs in the system.

For each of the busy period types to be defined in later sections, the LST and first two moments will be obtained. In addition, the LST and first moment will be derived for the waiting time of a job conditioned upon the arrival finding a specified busy-period in progress.

2. Busy Periods for the M/M/c Queueing System

In a single-server system, a "normal" busy period is usually defined to be an interval which begins with the arrival of a job to an empty system and which terminates when the system again becomes idle; this busy period represents the length of time that the (single) server is busy. A similar type of busy period for a multiple-server system may be defined which is an interval during which one or more processors are busy; however, there are additional busy periods which are of interest. Define the following type of busy period for the M/M/c queueing system:

$$T_k = \min \left[t \; : \; \begin{matrix} (k) \text{ jobs in system at time } 0^+ \\ (k-1) \text{ " " " " time } t \end{matrix} \right]$$

where integer $k \geq 1$.

The random variable (r.v.) may be considered as an interval during which the number in system is greater than or equal to k; alternatively, this variable represents the length of time necessary to achieve an overall reduction of jobs in system by one. Due to the memoryless property of the exponential processing times, the distribution of T_k will not depend on whether or not any of the k jobs have received prior processing. For each random variable T_k, we also define:

$H_k(t) = \Pr[T_k \leq t]$ = cumulative distribution function (cdf) for T_k,

$\eta_k(s)$ = LST for the distribution of T_k.

The properties of Poisson processes summarized in Appendix 1 will often be cited in the remainder of the paper, and the following notation will be employed:

λ = (Poisson) arrival rate for jobs,

μ = (Exponential) processing rate for jobs,

$\sim$ = denotes 'distributed as' e.g. $X \sim Y$ is interpreted to mean that random variable X has the same distribution as r.v. Y.

Before proceeding into the analysis, a r.v. will be introduced which simplifies the presentation; for $1 \leq k \leq c$, define:

P_k = r.v. denoting an interevent time associated with an aggregate of k simultaneous Poisson processes, each with rate μ.

$G_k(p)$ = cdf for P_k,

$\gamma_k(s)$ = LST for the distribution of P_k.

Referring to Appendix 1, we easily find the following:

$\gamma_k(s) = k\mu/(s+k\mu)$,

$E(P_k) = 1/(k\mu)$,

$E(P_k^2) = 2/(k\mu)^2$.

Given the above results, we may next analyze the properties of busy period T_k.

LEMMA 1. For the M/M/c queueing system, the distribution of busy period T_k has the following characteristics:

(i) For $k > c$,

$$T_k \sim T_c.$$

(ii) For $k = c$,

$$\eta_c(s) = \gamma_c(s + \lambda - \lambda\eta_c(s)),$$

$$E(T_c) = E(P_c)/\left[1 - \lambda E(P_c)\right],$$

$$E(T_c^2) = E(P_c^2)/\left[1 - \lambda E(P_c)\right]^3.$$

(iii) For $1 \leq k < c$,

$$\eta_k(s) = \gamma_k(s + \lambda - \lambda\eta_{k+1}(s)),$$

$$E(T_k) = E(P_k) + \lambda E(P_k)E(T_{k+1}),$$

$$E(T_k^2) = E(P_k^2) \left[1+E(T_{k+1})\right]^2 + \lambda E(P_k)E(T_{k+1}^2).$$

Proof. We divide the proof into three sections corresponding to the statement of the lemma.

(i) It is trivial to verify that, for $k > c$, r.v. T_k has the same distribution as r.v. T_c. There will be exactly c processors active throughout the duration of these busy periods, and these busy periods also have identical arrival processes for jobs.

(ii) Note that busy period T_c consists of an integer number of intervals which have the distribution of P_c. If no jobs arrive during the initial interval, the busy period T_c completes. If N arrivals occur during the initial interval (distributed as P_c), the busy period also includes N subintervals as shown in Figure 1.

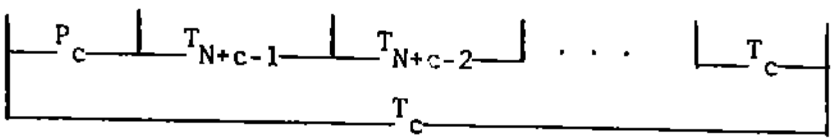

Figure 1. Subintervals with a busy period T_c.

At the end of the initial interval there will be (N+c-1) jobs in system, and the subintervals T_{N+c-1} through T_c will be needed to reduce the number in system to (c-1). These subintervals each have the same distribution as T_c(see part (i)), and from the convolution property of the LST we have

$$\eta_c(s \mid P_c=p, N=n) = \exp(-sp)\left[\eta_c(s)\right]^n.$$

Removing the conditioning on N and P_c, the unconditional LST becomes

$$\eta_c(s) = \int_{p=0}^{\infty} \left\{ \sum_{n=0}^{\infty} \left[(\lambda p)^n/n!\right] \exp(-\lambda p)\,\eta_c(s \mid P_c=p, N=n) \right\} dG_c(p),$$

$$= \gamma_c(s + \lambda - \lambda\eta_c(s)).$$

The properties of the LST allow the first two moments for T_c to be found from $-\eta_c'(0)$ and $\eta_c''(0)$, respectively. Solving the resulting expressions for $E(T_c)$ and $E(T_c^2)$ gives the desired results.
(iii) For $1 \leq k < c$, each busy period T_k may be decomposed into the intervals shown in Figure 2. The first of these intervals is an interevent time for an aggregate of k Poisson processes, each with rate μ, and one Poisson arrival process with rate λ.

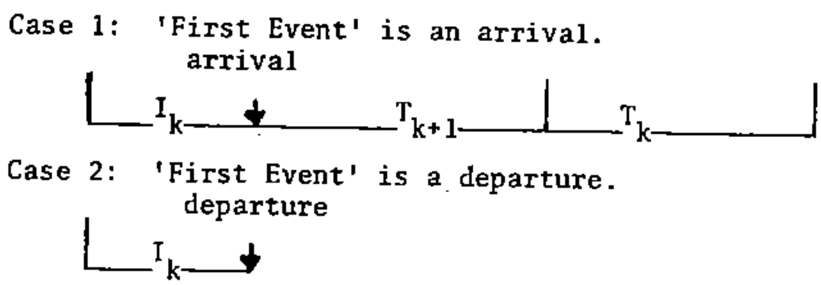

Figure 2. Intervals within busy period T_k ($1 \leq k < c$).

Random variable I_k is therefore the time until the first arrival or departure event in the busy period. If the 'first event' is an arrival, the remainder of the busy period is distributed as the sum of busy periods T_{k+1} plus T_k. Otherwise, the busy period ends if the 'first event' is a departure. The probability of the 'first event' being an arrival is $\lambda/(\lambda+k\mu)$ and that of a service completion event is $k\mu/(\lambda+k\mu)$. The LST for the distribution of I_k is found from Appendix 1 to be $(\lambda+k\mu)/(s+\lambda+k\mu)$, and the convolution property of the LST gives:

$$\eta_k(s) = \left[(\lambda+k\mu)/(s+\lambda+k\mu)\right] \left[(\lambda/(\lambda+k\mu))\eta_{k+1}(s)\eta_k(s)+(k\mu/(\lambda+k\mu)\right].$$

Solving for $\eta_k(s)$, we obtain

$$\eta_k(s) = k\mu/(k\mu+s+\lambda-\lambda\eta_{k+1}(s)).$$

Noting that $\gamma_k(s) = k\mu/(k\mu+s)$, the above result may be written as

$$\eta_k(s) = \gamma_k(s + \lambda - \lambda\eta_{k+1}(s)).$$

The first two moments for T_k are directly found by evaluating $-\eta_k'(0)$ and $\eta_k''(0)$, respectively. Q.E.D.

Another type of busy period found to be useful in the analysis of the M/G/1 system under non-preemeptive scheduling rules is the delay cycle (cf. reference [6], Chapter 8). A delay cycle is a generalized busy period which arises in situations where jobs arrive while the processor is unavailable due to some reason; for example, the processor(s) may be temporarily broken down or busy servicing higher priority jobs. Figure 3 gives an example of a delay cycle for the M/M/c queueing system.

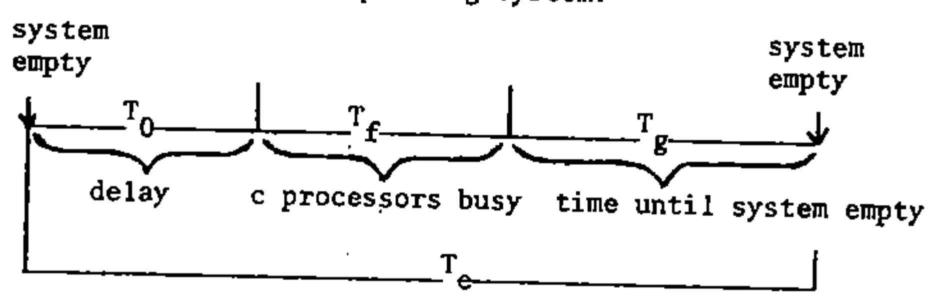

Figure 3. Delay Cycle T_e for the M/M/c queueing system.

The delay T_0 represents an amount of time during which all c processors are unavnilable. At the conclusion of this delay, the system will begin processing any jobs which arrived during the initial interval. The remainder of the delay cycle is the sum of two subintervals T_f and T_g during which jobs are serviced. Subinterval T_f begins immediately after the delay and lasts until there are fewer than c jobs in system (this subinterval has length zero if less than c jobs arrive during the delay). Subinterval T_g begins immediately after T_f and represents the time necessary to clear the system of jobs. The following notation is introduced to deal with the delay cycle described above:

T_0 = Delay during which the c processors are unavailable,

$H_0(t)$ = cdf for T_0,

$\eta_0(s)$ = LST for the distribution of T_0.

It is assumed that the LST and moments for r.v. T_0 are given.

$$T_e = \min \left[t: \begin{array}{l} \text{system empty at } 0^- \\ \text{delay interval } T_0 \text{ commences at } 0^+ \\ \text{system again empty at time } t \geq T_0 \end{array} \right]$$

$H_e(t)$ = cdf for T_e,

$\eta_e(s)$ = LST for the distribution of T_e;

T_f = length of interval commencing immediately after delay T_0 and terminating when there are less than c jobs in system,

$H_f(t)$ = cdf for T_f,

$\eta_f(s)$ = LST for the distribution of T_f;

T_g = interval which begins at the conclusion of T_f and which lasts until the system is empty,

$H_g(t)$ = cdf for T_g,

$\eta_g(s)$ = LST for the distribution of T_g.

We next present a lemma describing the distributions for the random variables given above.

LEMMA 2. The LST and first moment for random variables T_e, T_f, and T_g are given by:

$$\begin{aligned}
\eta_e(s) = {} & \left[\eta_0(s+\lambda-\lambda\eta_c(s)) \prod_{j=1}^{c-1} \eta_j(s)\right] / \left[\eta_c(s)\right]^{c-1} \\
& + \eta_0(s+\lambda)\left[1-\left\{\left[\prod_{j=1}^{c-1} \eta_j(s)\right] / \left[\eta_c(s)\right]^{c-1}\right\}\right] \\
& + \sum_{n=1}^{c-1} \left[(-1)^n\lambda^n/n!\right] \eta_0(s+\lambda)\left\{\prod_{j=1}^{n} \eta_j(s) - \left[\prod_{j=1}^{c-1} \eta_j(s)\right] / \left[\eta_c(s)\right]^{c-1-n}\right\}, \\
\eta_f(s) = {} & \sum_{n=0}^{c-1} \left[(-1)^n\lambda^n/n!\right] \eta_0^{(n)}(\lambda) \\
& + \left\{\eta_0(\lambda-\lambda\eta_c(s)) - \sum_{n=0}^{c-1} \left[(-1)^n\lambda^n/n!\right]\left[\eta_c(s)\right]^n \eta_0^{(n)}(\lambda)\right\} / \left[\eta_c(s)\right]^{c-1},
\end{aligned}$$

$$\eta_g(s) = \eta_0(\lambda) + \sum_{n=1}^{c-2} \left[(-1)^n\lambda^n/n!\right] \eta_0^{(n)}(\lambda) \prod_{j=1}^{n} \eta_j(s)$$

$$+ \prod_{j=1}^{c-1} \eta_j(s) \left[1 - \sum_{n=0}^{c-2} \left[(-1)^n\lambda^n/n!\right] \eta_0^{(n)}(\lambda)\right],$$

$$E(T_e) = E(T_0) + E(T_f) + E(T_g),$$

$$E(T_f) = E(T_c)\left\{\left[\lambda E(T_0) - \sum_{n=0}^{c-1} \left[n(-1)^n\lambda^n/n!\right] \eta_0^{(n)}(\lambda)\right]\right.$$

$$\left. - (c-1)\left[1 - \sum_{n=0}^{c-1} \left[(-1)^n\lambda^n/n!\right] \eta_0^{(n)}(\lambda)\right]\right\},$$

$$E(T_g) = \sum_{k=1}^{c-1} E(T_k)\left\{1 - \sum_{n=0}^{k-1} \left[(-1)^n\lambda^n/n!\right] \eta_0^{(n)}(\lambda)\right\}.$$

Proof. The above LSTs are derived using similar techniques; given the length of the delay and number of arrivals during the delay, the conditional LST is found for each type of interval, and the conditioning is then removed to find the final result.

Delay cycle T_e consists of delay T_0 plus the time (subintervals T_f and T_g) needed to clear the system of all jobs. Recall that a busy period T_k ($k \geq 1$) represents the time necessary to reduce the number of jobs in system from k jobs to (k-1) jobs. If N jobs arrive during the delay T_0, the remainder of T_e is the sum of busy periods T_N through T_1. It follows that the conditional LST for the distribution of T_e equals

$$\eta_e(s|T_0=t,N=n) = \begin{cases} \exp(-st) \prod_{j=1}^{n} \eta_j(s) & \text{for } n > 0, \\ \exp(-st) & \text{for } n = 0. \end{cases}$$

Since $\eta_j(s) = \eta_c(s)$ for all j greater than c, we have for $n \geq c$

$$\eta_e(s|T_0=t,N=n) = \exp(-st) \left[\eta_c(s)\right]^{n-c+1} \prod_{j=1}^{c-1} \eta_j(s) .$$

Removing the conditioning on the number of Poisson arrivals during the delay and on the delay length, the unconditional LST is found to be

$$\eta_e(s) = \int_{t=0}^{\infty} \left\{ \sum_{n=0}^{\infty} \left[(\lambda t)^n/n!\right] \exp(-\lambda t)\eta_e(s|T_0=t,N=n)\right\} dH_0(t),$$

$$= \eta_0(s+\lambda) + \sum_{n=1}^{c-1} \left\{\left[(-1)^n\lambda^n/n!\right]\eta_0^{(n)}(s+\lambda) \prod_{j=1}^{n} \eta_j(s)\right\}$$
$$+\left\{\left[\prod_{j=1}^{c-1} \eta_j(s)\right]/\left[\eta_c(s)\right]^n\right\}\left\{\eta_0(s+\lambda-\lambda\eta_c(s))-\eta_0(s+\lambda)\right.$$
$$\left. - \sum_{n=1}^{c-1} \left[(-1)^n\lambda^n/n!\right]\left[\eta_c(s)\right]^n\eta_0{}^{(n)}(s+\lambda)\right\}.$$

In obtaining this result, it was necessary to make use of the following property of the Laplace-Stieltjes Transofrm (see reference [9] , p. 57):

$$\int_{t=0}^{\infty} \exp(-st)t^n dH_0(t) = (-1)^n \eta_0^{(n)}(s).$$

The expression for $\eta_e(s)$ is, with minor rearrangement of terms, the desired LST for the distribution of T_e.

We next derive the LSTs for T_f and T_g, again conditioned upon the number of arrivals during the delay and on the delay length. Subinterval T_f has length zero if there are fewer than c arrivals during delay T_0; if there are N arrivals, where N is greater than or equal to c, subinterval T_f consists of the sequence of busy periods $T_N, T_{N-1},\ldots,T_c$. Since each of these busy periods has the same distribution as T_c, the conditional LST for the distribution of T_f becomes

$$\eta_f(s|T_0=t,N=n) = \begin{cases} 1 & \text{for } 0 \leq n < c, \\ \left[\eta_c(s)\right]^{n-c+1} & \text{for } c \leq n. \end{cases}$$

Subinterval T_g also has length zero if no jobs arrive during the delay. If N jobs arrive during the delay, where N is some positive integer less than c, subinterval T_g is the sum of busy periods $T_n, T_{N-1},\ldots, T_1$. Otherwise, T_g begins immediately after a subinterval T_f having length greater than zero. Since subinterval T_f ends with (c-1) jobs in system, T_g is the sum of busy periods T_{c-1} through T_1. The conditional LST for the distribution of T_g is therefore given by

$$\eta_g(s|T_0=t,N=n) = \begin{cases} 1 & \text{for } n = 0, \\ \prod_{j=1}^{n} \eta_j(s) & \text{for } 0 < n \leq c-1, \\ \prod_{j=1}^{c-1} \eta_j(s) & \text{for } c \leq n. \end{cases}$$

The conditioning on number of arrivals and on delay length for the above can be removed in exactly the same manner as used for delay cycle T_e, and the desired LSTs are obtained.

The expected values for T_e, T_f, and T_g follow directly from $-\eta_e'(0)$, $-\eta_f'(0)$, and $-\eta_g'(0)$, respectively. Q.E.D.

At first glance, the terms appearing in the expressions for the expected lengths of intervals T_f and T_g may seem somewhat puzzling. The terms are better understood if one notes that

$$\Pr\left[N=n \text{ arrivals during delay } T_0\right] = \left[(-1)^n\lambda^n/n!\right]\eta_0^{(n)}(\lambda).$$

Since the arrival process is Poisson, the expected number of arrivals during the delay is given by $\lambda E(T_0)$; using this information we can rewrite the expressions for the expected lengths of T_f and T_g as shown below, where N = number of arrivals during T_0:

$$E(T_f) = E(T_c)\left\{\Pr\left[N \geq c\right]E(N|N \geq c) - (c-1)\Pr\left[N \geq c\right]\right\},$$

and

$$E(T_g) = \sum_{n=1}^{c-1} E(T_n)\Pr\left[N \geq n\right].$$

The above interpretation is consistent with the type of result that one intuitively expects.

3. Waiting Time Under the FCFS Discipline

Having obtained results pertaining to the distribution of the two busy period types introduced in the previous section, we consider the (conditional) waiting time for a job which arrives to find one of these busy periods in progress. It is assumed that the First-Come-First-Served (FCFS) discipline is employed and that the system is not saturated. Define:

W = Waiting time for a job (i.e. time between the arrival of a job and the instant that it first goes into service),

A(w) = cdf for random variable W,

$\alpha(s)$ = LST for the distribution of W.

In order to avoid repetitious derivations, a new type of delay cycle T_p will be introduced; this delay cycle is of interest because both a busy period T_c and a subinterval T_f (within a delay cycle T_e) are special cases of this type of interval. Define interval T_p as given below:

$$T_p = \min \left[t: \begin{array}{l} \text{(c-1) jobs in system when delay } T_{p,0} \text{ commences at time } 0^+ \\ \text{(c-1) jobs in system at time t, where } t \geq T_{p,0} \end{array} \right]$$

That is, there are (c-1) "initial" jobs in system when T_p begins; interval T_p then consists of a delay $T_{p,0}$ which is assumed to have a general distribution plus a "delay busy period" during which jobs are serviced and for which c or more jobs are in system.

THEOREM 1. The conditional LST for the distribution of waiting time W in the M/M/c system, given that a job arrives during a delay cycle T_p, is equal to

$$\alpha(s|\text{arrival during } T_p) = \left[1-\eta_{p,0}(s)\right] / \left\{E(T_p)\left[\lambda\gamma_c(s)-\lambda+s\right]\right\}$$

where $\eta_{p,0}(s)$ = LST for the distribution of delay $T_{p,0}$, and the expected value for the conditional waiting time is

$$E(W|\text{arrival during } T_p) = \lambda E(P_c^2)/\left\{2\left[1-\lambda E(P_c)\right]\right\} + E(T_{p,0}^2)/\left[2E(T_{p,0})\right].$$

Proof. Interval T_p will be viewed as the sum of subintervals $T_{p,j}$, where $j \geq 0$; Figure 4 illustrates the situation.

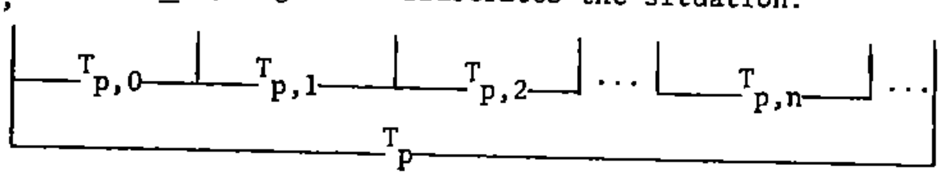

Figure 4. Subintervals with delay cycle T_p.

The initial subinterval is the delay $T_{p,0}$; at the conclusion of this subinterval there will be the (c-1) initial jobs plus those jobs which arrived during the delay. The number of jobs arriving during any subinterval $T_{p,j}$ will be denoted by N_j, and each subinterval $T_{p,j}$ (where $j > 0$) is defined as the sum of N_{j-1} inter-event times P_c (i.e. the time for N_{j-1} departures). For $j \geq 0$, define:

$T_{p,j}$ = length of subinterval-j of delay cycle T_p,

$H_{p,j}(t)$ = cdf for $T_{p,j}$,

$\eta_{p,j}(s)$ = LST for the distribution of $T_{p,j}$.

The LST for the distribution of subinterval $T_{p,j}$ will now be derived; using the convolution property of the Laplace-Stieltjes Transform, we have

$$\eta_{p,j}(s|T_{p,j-1}=t,N_{j-1}=n) = \left[\gamma_c(s)\right]^n \qquad \text{for } j \geq 1.$$

The conditioning is easily removed to obtain

$$\eta_{p,j}(s) = \int_{t=0}^{\infty} \left\{ \sum_{n=0}^{\infty} \left[(\lambda t)^n/n!\right] \exp(-\lambda t) \left[\gamma_c(s)\right]^n \right\} dG_{p,j-1}(t) \quad ,$$

$$= \eta_{p,j-1}(\lambda-\lambda\gamma_c(s)).$$

If the system is operating under nonsaturated conditions, there will be a finite j for which $T_{p,j}$ is zero, and it will also be true that

$$\underset{j\to\infty}{\text{limit}}\ \eta_{p,j}(s) = 1.$$

A job which arrives during subinterval $T_{p,j}$ of busy cycle T_p will encounter the situation illustrated in Figure 5.

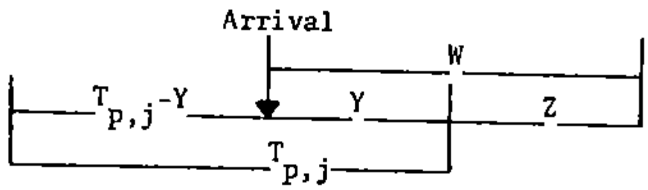

Figure 5. Waiting time for a job arriving during subinterval-j of delay cycle T_p.

The job arriving during subinterval $T_{p,j}$ will be required to wait in queue for an amount of time Y until the end of the subinterval plus an additional amount Z which represents the time for N departures to occur, where N is the number of jobs which arrived previously in time $T_{p,j}$-Y. Quantity Y is a random modification (see Appendix 1), and r.v. Z is the sum of N interevent intervals P_c. The convolution property of the Laplace-Stieltjes Transform gives the following conditional LST for the waiting time distribution:

$$\alpha(s|T_{p,j}=t,Y=y,N=n) = \exp(-sy)\left[\gamma_c(s)\right]^n .$$

The conditioning on N can be easily eliminated because the arrival process is Poisson; therefore, we obtain

$$\alpha(s|T_{p,j}=t,Y=y) = \sum_{n=0}^{\infty} \left[\left[\lambda(t-y)\right]^n/n!\right] \exp(-\lambda(t-y))\alpha(s|T_{p,j}=t,Y=y,N=n),$$
$$= \exp(-sy)\exp(-\lambda(t-y))\exp(\lambda(t-y).\gamma_c(s)).$$

Making use of the fact that Y is a random modification, we have

$$\Pr\left[y \le Y \le y+dy,\ t \le T_{p,j} \le t+dt\right] = dH_{p,j}(t)dy/E(T_{p,j}).$$

The conditional LST for the waiting time of a job, given that the arrival occurs during the jth subinterval of delay cycle T_p, equals

$$\alpha(s|\text{arrival during } T_{p,j}) = \int_{t=0}^{\infty}\left\{\int_{y=0}^{t} \alpha(s|T_{p,j}=t,Y=y)dy\right\}dH_{p,j}(t)/E(T_{p,j}),$$
$$= \left[\eta_{p,j}(\lambda-\lambda\gamma_c(s))-\eta_{p,j}(s)\right]/\left\{E(T_{p,j})\left[\lambda\gamma_c(s)-\lambda+s\right]\right\},$$
$$= \left[\eta_{p,j+1}(s)-\eta_{p,j}(s)\right]/\left\{E(T_{p,j})\left[\lambda\gamma_c(s)-\lambda+s\right]\right\}.$$

The probability π_j that a job arrives during subinterval $T_{p,j}$, given that the arrival takes place during delay cycle T_p, is equal to the steady-state probability that interval $T_{p,j}$ is in progress (see Strauch [8]).

$$\pi_j = \Pr\left[\text{arrival during } T_{p,j}|T_p \text{ in progress}\right] = E(T_{p,j})/E(T_p).$$

The conditional LST for the waiting time of a job arriving during T_p is therefore equal to

$$\alpha(s|\text{arrival during } T_p) = \sum_{j=0}^{\infty} \pi_j\alpha(s|\text{arrival during } T_{p,j}),$$
$$= \sum_{j=0}^{\infty} \left[\eta_{p,j+1}(s)-\eta_{p,j}(s)\right] / \left\{ E(T_p) \left[\lambda\gamma_c(s)-\lambda+s\right]\right\},$$
$$= \left[1-\eta_{p,0}(s)\right] / \left\{ E(T_p) \left[\lambda\gamma_c(s)-\lambda+s\right]\right\}.$$

The first moment for the conditional waiting time is found by making use of l 'Hospital's Rule to evaluate $-\alpha'(0|\text{arrival during } T_p)$.

Q.E.D.

The above Theorem will now be used to obtain the conditional LST and first moment for the distribution of waiting time for a job which arrives during a busy period T_c.

<u>THEOREM 2.</u> The conditional LST for the distribution of the waiting time W of a job in the M/M/c system, given that the job arrives during busy period T_c is given by

$$\alpha(s|\text{arrival during } T_c) = \left[1-\gamma_c(s)\right] / \left\{E(T_c)\left[\lambda\gamma_c(s)-\lambda+s\right]\right\},$$

and the expected value for the conditional waiting time is

$$E(W|\text{arrival during } T_c) = \lambda E(P_c^2)/\left[2\left[1-\lambda E(P_c)\right]\right] + E(P_c^2)/\left[2E(P_c)\right].$$

<u>Proof.</u> Observe that busy period T_c is a special case of delay cycle T_p in which the delay $T_{p,0}$ is an interevent time P_c. Replacing $\eta_{p,0}(s)$, $E(T_{p,0})$ and $E(T_{p,0}^2)$ in THEOREM 1 by the corresponding terms $\gamma_c(s)$, $E(P_c)$, and $E(P_c^2)$, we obtain the desired results.

Q.E.D.

Given the LST and expectation associated with the waiting time of a job conditioned upon its arrival during a busy period T_c, sufficient information is available to determine corresponding results for busy period T_k, where $1 \leq k < c$.

<u>THEOREM 3.</u> The conditional LST for the distribution of waiting time W for a job in the M/M/c system which arrives during a busy period T_k, where $1 \leq k \leq c-1$, is given by

$$\alpha(s|\text{arrival during } T_k) = \pi_1(k) + \pi_2(k)\alpha(s|\text{arrival during } T_{k+1}),$$

where $\pi_1(k) = E(P_k)/E(T_k)$ and $\pi_2(k) = E(P_k)E(T_{k+1})/E(T_k)$; therefore, the expected value for the conditional waiting time is

$$E(W|\text{arrival during } T_k) = \pi_2(k)E(W|\text{arrival during } T_{k+1}).$$

<u>Proof.</u> During busy period T_k, the system will be in one of two possible states:

State-1: exactly k jobs in system,

State-2: (k+1) or more jobs in system.

Let $\pi_1(k)$ and $\pi_2(k)$ denote the steady-state probability that the system is in State-1 and State-2, respectively, given that busy period T_k is in progress. Using LEMMA 1, we have

$$E(T_k) = E(P_k) + \lambda E(P_k)E(T_{k+1}).$$

Observe that the first term on the right-hand side of the equation represents the expected time that the system is in State-1 during interval T_k, and the second term is the expected time in State-2.

Using the method described in Appendix 2, we obtain the representation for $\pi_1(k)$ and $\pi_2(k)$ given above.

Because a job arriving when the system is in State-1 can immediately go into service (i.e. $k < c$) on an available processor, the waiting time of the job is zero, and therefore the conditional LST is

$$\alpha(s|\text{State-1}) = 1.$$

The conditional LST for jobs which arrive during State-2 is $\alpha(s|\text{arrival during } T_{k+1})$, and the random property of Poisson arrivals gives:

$$\alpha(s|\text{arrival during } T_k) = \pi_1(k) + \pi_2(k)\alpha(s|\text{arrival during } T_{k+1})$$

and the expected value for the waiting time conditioned upon arrival during T_k is found to be

$$E(W|T_k) = \pi_2(k)E(W|T_{K+1}). \qquad \text{Q.E.D.}$$

Using THEOREM 2 and THEOREM 3, the expected values for the conditional waiting time in the M/M/c system under the FCFS discipline may be found for arrival during busy periods $T_c, T_{c-1}, \ldots, T_1$ in a straightforward fashion. This completes the waiting time analysis for the first type of busy period introduced in the previous section, and we next examine the delay cycle T_e.

THEOREM 4. Given a M/M/c system, the conditional LST for the distribution of waiting time for a job which arrives during delay T_0 (within delay cycle T_e) is given by

$$\alpha(s|\text{arrival during } T_0) =$$
$$= \Big\{ \big[\eta_0(\lambda-\lambda\gamma_c(s)\big] / \Big\{ \big[\gamma_c(s)\big]^{c-1}\big[s-\lambda+\lambda\gamma_c(s)\big]\Big\} - \eta_0(s)\theta(0,s) + \sum_{n=0}^{c-2} \big[(-\lambda)^n/n!\big]\, \eta_0^{(n)}(\lambda)\theta(n,s) \Big\} / E(T_0),$$

where $\theta(n,s) = \Big\{1-\big[-\lambda/(s-\lambda)\big]^{c-1-n}\Big\}/s$
$$-\Big\{1-\big[-\lambda\gamma_c(s)/(s-\lambda)\big]^{c-1-n}\Big\}/\Big\{\big[\gamma_c(s)\big]^{c-1-n}\big[s-\lambda+\lambda\gamma_c(s)\big]\Big\}.$$

It follows that

$$E(W|\text{arrival during } T_0) = E(T_0^2)/\big[2E(T_0)\big] + E(P_c)\Big\{\lambda E(T_0^2)/\big[2E(T_0)\big] - \phi_1 - (c-1)(1-\phi_2)\Big\},$$

where

$$\phi_1 = \left\{ (c-2)(c-1)/2 - \sum_{n=0}^{c-2} \left[(-\lambda)^n/n!\right] \eta_0^{(n)}(\lambda) \left[(c-2)(c-1)/2-(n-1)\ n/2\right]\right\} / \left[\lambda E(T_0)\right]$$

and

$$\phi_2 = \left\{ (c-1) - \sum_{n=0}^{c-2} \left[(-\lambda)^n/n!\right] \eta_0^{(n)}(\lambda) \left[c-1-n\right] / \left[\lambda E(T_0)\right] \right. .$$

<u>Proof.</u> Recall that delay cycle T_e starts with delay T_0 and that the system is initially empty of jobs. A job arriving during the delay portion of the delay cycle encounters the situation shown in Figure 6.

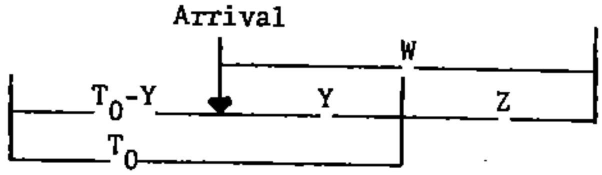

Figure 6. Waiting time for a job arriving during delay T_0.

Interval Y in the above diagram represents the time between the arrival of the job and the end of delay T_0; because the arrival process is Poisson, interval Y has the distribution of a random modification (cf. Appendix 1). If N jobs arrived to the system during T_0-Y, interval Z represents the amount of time that the job will be delayed due to jobs which arrived earlier during the delay. Interval Z equals zero if N is less than the necessary for N-(c-1) departures to occur, and this time is the sum of N-(c-1) interevent times P_c. From the convolution property of the Laplace-Stieltjes Transofrm, we have

$$\alpha(s|T_0=t,N=n) = \begin{cases} \exp(-sy) & \text{for } 0 \leq n \leq c-2, \\ \exp(-sy)\left[\gamma_c(s)\right]^{n-(c-1)} & \text{for } c-1 \leq n\ . \end{cases}$$

The conditioning on the number of arrivals N can be removed by taking into account the probability of any specified value of N for the interval T_0-Y.

$$\alpha(s|T_0=t,Y=y) = \sum_{n=0}^{\infty} \left[(\lambda(t-y))^n/n!\right] \exp(-\lambda(t-y))\alpha(s|T_0=t,Y=y,N=n) ,$$

$$= \exp(-\lambda t) \sum_{n=0}^{c-2} \left[(\lambda(t-y))^n/n!\right] \exp(-(s-\lambda)y)$$
$$+ \left[\gamma_c(s)\right]^{-(c-1)} \exp(-(\lambda-\lambda\gamma_c(s))t)\exp(-(s-\lambda+\lambda\gamma_c(s))y)$$
$$- \left[\gamma_c(s)\right]^{-(c-1)} \exp(-\lambda t) \sum_{n=0}^{c-2} \left[(\lambda(t-y)\gamma_c(s))^n/n!\right] \exp(-(s-\lambda)y) .$$

Because interval Y is a random modification, the conditioning on Y and T_0 may be removed to give

$$\alpha(s|\text{arrival during } T_0) = \int_{t=0}^{\infty} \int_{y=0}^{t} \alpha(s|T_0=t, Y=y)dy \, dH_0(t) \,/E(T_0).$$

We first consider some terms which pose a problem in the evaluation of the above integral. Define $Q_1(t,n)$, where $0 \leq n \leq c-2$, as shwon below:

$$Q_1(t,n) = \int_0^t \left[(\lambda(t-y))^n/n!\right] \exp(-(s-\lambda)y) \; dy.$$

For n=0, we have

$$Q_1(t,0) = \left[1-\exp(-(s-\lambda)t)\right] /(s-\lambda).$$

For $1 \leq n \leq c-2$, we obtain the following result by using integration by parts:

$$Q_1(t,n) = \left[(\lambda t)^n/n!\right] / (s-\lambda) - (\lambda/(s-\lambda))Q_1(t,n-1).$$

From this result it follows that

$$\sum_{n=0}^{c-2} Q_1(t,n) = Q_1(t,0) \sum_{k=0}^{c-2} (-\lambda/(s-\lambda))^k$$
$$+\left\{\sum_{n=1}^{c-2} \left[(\lambda t)^n/n!\right] \sum_{k=0}^{c-2-n} (-\lambda/(s-\lambda))^k\right\} /(s-\lambda).$$

Noting that several partial sums of geometric series appear in the above expression and substituting for $Q_1(t,0)$, we obtain:

$$\sum_{n=0}^{c-2} Q_1(t,n) = Q_1(t,0) \left[1-(-\lambda/(s-\lambda))^{c-1}\right]\left[1-(-\lambda/(s-\lambda))\right]$$
$$+ (1/(s-\lambda)) \sum_{n=1}^{c-2} \left[(\lambda t)^n/n!\right] \left[1-(-\lambda/(s-\lambda))^{c-1-n}\right]/\left[1-(-\lambda/(s-\lambda))\right] ,$$
$$= - \exp(-(s-\lambda)t)\left[1 - (-\lambda/(s-\lambda))^{c-1}\right]/s$$
$$+ \sum_{n=0}^{c-2} \left[(\lambda t)^n/n!\right] \left[1-(-\lambda/(s-\lambda))^{c-1-n}\right] /s .$$

Define another function $Q_2(t,n)$ to be

$$Q_2(t,n) = \int_0^t \left[(\lambda\gamma_c(s)(t-y))^n/n!\right] \exp(-(s-\lambda)y)dy \ .$$

Using the same procedure employed for the function $Q_1(t,n)$, we obtain

$$\sum_{n=0}^{c-2} Q_2(t,n) = -\exp(-(s-\lambda)t)\left\{1-\left[-\lambda\gamma_c(s)/(s-\lambda)\right]^{c-1}\right\}/\left\{s-\lambda+\lambda\gamma_c(s)\right\}$$
$$+\sum_{n=0}^{c-2}\left[(\lambda\gamma_c(s)t)^n/n!\right]\left\{1-\left[-\lambda\gamma_c(s)/(s-\lambda)\right]^{c-1-n}\right\}/\left\{s-\lambda+\lambda\gamma_c(s)\right\}.$$

Returning to the evaluation of $\alpha(s|\text{arrival during } T_0)$, we have

$$\alpha(s|\text{arrival during } T_0)$$
$$= \int_{y=0}^{\infty}\left\{\exp(-\lambda y)\sum_{n=0}^{c-2} Q_1(t,n) + \left[\gamma_c(s)\right]^{-(c-1)}\exp(-(\lambda-\lambda\gamma_c(s))t)\int_0^t \exp(-(s-\lambda+\lambda\gamma_c(s))y)dy - \left[\gamma_c(s)\right]^{-(c-1)}\exp(-\lambda t)\sum_{n=0}^{c-2} Q_2(t,n)\right\} dH_0(t)/E(T_0) \ .$$

Substituting the results for the sums of functions $Q_1(t,n)$ and $Q_2(t,n)$ into the above equation, the evaluation of the integral may be completed. Rearranging terms in the result gives the desired expression for $\alpha(s|\text{arrival during } T_0)$.

The expected value for the conditional waiting time is found by making use of l'Hospital's Rule to evaluate $-\alpha'(0|\text{arrival during } T_0)$. Q.E.D.

Observing the result for $E(W|\text{arrival during } T_0)$, one sees certain terms which are easily explained; e.g. we might anticipate that the waiting time would consist of the expected length of the random modification, $E(T_0^2)/\left[2E(T_0)\right]$, which is the time between the arrival of the job and the end of delay T_0. The remaining terms in the expression are explained by making the following observation: The time interval between the start of the delay and the arrival of the job whose progress we are following has the same distribution as the random modification. Referring to Figure 6, this is equivalent to stating that the distribution of interval T_0-Y is identical to that for interval Y. Define random variable X as follows:

X = interval T_0-Y in Figure 6,

$\tau(s)$ = LST for the distribution of X.

Because X has the distribution of a random modification, we have the following results from Appendix 1:

$$E(X) = E(T_0^2) / \left[2E(T_0)\right] ,$$

and

$$\tau(s) = \left[1-\eta_0(s)\right] / \left[sE(T_0)\right] .$$

We define a variable N to be

N = Number of previous arrivals to the system during interval X.

A patient person can verify that the following equations are true:

$$\sum_{n=0}^{c-2} n.\Pr\left[N=n \text{ arrivals during } X\right] = \sum_{n=0}^{c-2} \left[(-\lambda)^n/n!\right] \tau^{(n)}(\lambda).n,$$

$$= \phi_1,$$

and

$$\sum_{n=0}^{c-2} \Pr\left[N=n \text{ arrivals during } X\right] = \sum_{n=0}^{c-2} \left[(-\lambda)^n/n!\right] \tau^{(n)}(\lambda) ,$$

$$= \phi_2 .$$

If we interpret terms ϕ_1 and ϕ_2 in the above manner, the expected length of the conditional waiting time becomes (cf. Figure 6)

$$E(W | \text{arrival during } T_0) = E(Y) + E(Z) ,$$

where the expected length of interval Z is

$$E(Z) = E(P_c) \left\{\Pr\left[N \geq c-1\right] \left[E(N | N \geq c-1)\right] - (c-1)\Pr\left[N \geq c-1\right]\right\}.$$

Thus, we have a satisfying explanation for all terms in the expression for the expected value of the conditional waiting time.

<u>THEOREM 5.</u> Given a M/M/c system, the conditional LST for the distribution of waiting time for a job which arrives during interval T_f (within delay cycle T_e) is equal to

$$\alpha(s | \text{arrival during } T_e) = \left[1-\eta_{f,0}(s)\right] / \left\{E(T_f)\left[\lambda\gamma_c(s)-\lambda+s\right]\right\},$$

where $\eta_{f,0}(s) = \sum_{n=0}^{c-2} \left[(-\lambda)^n/n!\right] \eta_0^{(n)}(\lambda)$

$$+ \left\{\eta_0(\lambda-\lambda\gamma_c(s)) - \sum_{n=0}^{c-2} \left[(-\lambda\gamma_c(s))^n/n!\right] \eta_0^{(n)}(\lambda)\right\} / \left[\gamma_c(s)\right]^{c-1} .$$

The expected length of the conditional waiting time is given by

$$E(W|\text{arrival during } T_f) = E(P_c^2)/\left[2(1-\lambda E(P_c))\right] + E(T_{f,0}^2)/\left[2E(T_{f,0})\right],$$

where

$$E(T_{f,0}) = E(P_c)\left\{\lambda E(T_0) - \sum_{n=0}^{c-2}\left[(-\lambda)^n/n!\right]\eta_0^{(n)}(\lambda)\cdot n - (c-1)\left[1-\sum_{n=0}^{c-2}\left[(-\lambda)^n/n!\right]\eta_0^{(n)}(\lambda)\right]\right\},$$

and

$$\begin{aligned} E(T_{f,0}^2) = \left[E(P_c)\right]^2 \Bigg\{ & -2(c-1)\left[\lambda E(T_0) - \sum_{n=0}^{c-2}\left[(-\lambda)^n/n!\right]\eta_0^{(n)}(\lambda)\cdot n\right. \\ & + \lambda^2 E(T_0^2) - \sum_{n=0}^{c-2}\left[(-\lambda)^n/n!\right]\eta_0^{(n)}(\lambda)\cdot n\cdot(n-1) \\ & + c(c-1)\left[1-\sum_{n=0}^{c-2}\left[(-\lambda)^n/n!\right]\eta_0^{(n)}(\lambda)\right]\Bigg\} \\ + E(P_c^2)\Bigg\{ & \lambda E(T_0) - \sum_{n=0}^{c-2}\left[(-\lambda)^n/n!\right]\eta_0^{(n)}(\lambda)\cdot n \\ & - (c-1)\left[1-\sum_{n=0}^{c-2}\left[(-\lambda)^n/n!\right]\eta_0^{(n)}(\lambda)\right]\Bigg\}. \end{aligned}$$

<u>Proof.</u> Interval T_f within delay cycle T_e begins immediately following delay T_0 and lasts until there are (c-1) or fewer jobs in system. This means that the interval has length zero if there are fewer than c arrivals during the delay. Interval T_f can be represented as the sum of an infinite number of subintervals $T_{f,k}$, where $k \geq 0$, as shown in Figure 7.

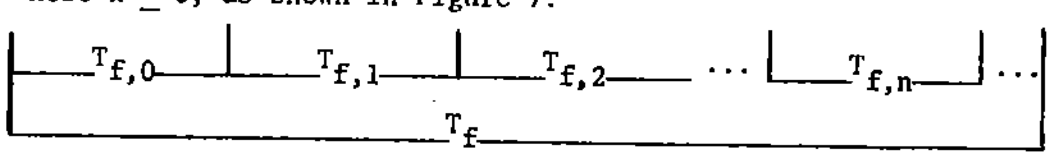

Figure 7. Subintervals of interval T_f.

Define a variable N as follows:

N = Number of jobs which arrive during delay T_0.

Subinterval $T_{f,0}$ is the time necessary for N-(c-1) departures to occur, where $N \geq c-1$. At the conclusion of $T_{f,0}$ there will be (c-1) jobs in system (i.e. the jobs which arrived during the delay) plus any jobs which arrived during $T_{f,0}$.

Interval T_f may now be seen to be a special case of a delay cycle T_p, and by replacing $\eta_{p,0}(s)$, $E(T_{p,0})$, and $E(T^2_{p,0})$ in THEOREM 1 with corresponding terms $\eta_{f,0}(s)$, $E(T_{f,0})$ and $E(T^2_{f,0})$, respectively, we immediately find the conditional LST the waiting time of a job which finds interval T_f in progress. However, we have yet to derive the LST and moments associated with subinterval $T_{f,0}$. The conditional LST for the distribution of $T_{f,0}$, given the number of arrivals N which occur during delay T_0, is given by

$$\eta_{f,0}(s|T_0=t,N=n) = \begin{cases} 1 & \text{for } 0 \leq n \leq c-2 , \\ \left[\gamma_c(s)\right]^{n-(c-1)} & \text{for } c-1 \leq n. \end{cases}$$

The conditioning on the number of arrivals and length of delay T_0 may be removed to give

$$\eta_{f,0}(s) = \int_{t=0}^{\infty} \sum_{n=0}^{\infty} \left[(\lambda t)^n/n!\right] \eta_{f,0}(s|T_0=t,N=n)\, dH_0(t) ,$$

$$= \int_{t=0}^{\infty} \left\{ \sum_{n=0}^{c-2} \left[(\lambda t)^n/n!\right] \exp(-\lambda t) + \left[\gamma_c(s)\right]^{-(c-1)} \left\{ \exp(-(\lambda-\lambda\gamma_c(s))t) - \sum_{n=0}^{c-2} \left[(\lambda\gamma_c(s)t)^n/n!\right] \exp(-\lambda t) \right\} \right\} dH_0(t),$$

$$= \sum_{n=0}^{c-2} \left[(-\lambda)^n/n!\right] \eta_0^{(n)}(\lambda)$$

$$+ \left\{ \eta_0(\lambda-\lambda\gamma_c(s)) - \sum_{n=0}^{c-2} \left[(-\lambda\gamma_c(s))^n/n!\right] \eta_0^{(n)}(\lambda) \right\} / \left[\gamma_c(s)\right]^{c-1} .$$

The expressions for $E(T_{f,0})$ and $E(T^2_{f,0})$ are found directly from $-\eta'_{f,0}(0)$ and $\eta''_{f,0}(0)$.

The last theorem in this section deal with the waiting time for a job arriving during delay cycle T_e. This result has a number of interpretations which will be discussed in the next section; at this point we note that this result can arise in the analysis of certain nonpreemptive priority disciplines in which there is a job class whose resource requests are so large that no concurrent processing is possible when that class of job is in service.

THEOREM 6. Given a M/M/c system, the conditional LST for the distribution of waiting time W of a job which arrives during delay cycle T_e is given by

$$\alpha(s|\text{arrival during } T_e) = \pi_0\alpha(s|\text{arrival during } T_0) + \pi_f\alpha(s|\text{arrival during } T_f) + \sum_{k=1}^{c-1} \pi_k\alpha(s|\text{arrival during } T_k),$$

where $\pi_0 = E(T_0)/E(T_e)$,

$\pi_f = E(T_f)/E(T_e)$,

and, for $1 \leq k \leq c-1$,

$$\pi_k = E(T_k)\left\{1- \sum_{n=0}^{k-1} \left[(-\lambda)^n/n!\right] \eta_0^{(n)}(\lambda)\right\}/E(T_e) .$$

The expected waiting time for a job which arrives during delay cycle T_e is therefore equal to

$$E(W|\text{arrival during } T_e) = \pi_0 E(W|\text{arrival during } T_0) + \pi_f E(W|\text{arrival during } T_f) + \sum_{k=1}^{c-1} \pi_k E(W|\text{arrival during } T_k).$$

Proof. Given that a Poisson arrival occurs during interval T_e, the probability that the arriving job finds the system in a particular state equals the steady-state probability of that state within the delay cycle.

Delay cycle T_e is the sum of delay T_0 and two intervals T_f and T_g which constitute a delay busy period and which have been previously defined. Recall that interval T_f starts immediately after the delay and lasts until there are less than c jobs in system. Interval T_g begins at the conclusion of T_f and terminates when the system is empty. Interval T_g is the sum of (c-1) subintervals $T_{g,j}$, where $1 \leq j \leq c-1$; these subintervals are illustrated in Figure 8.

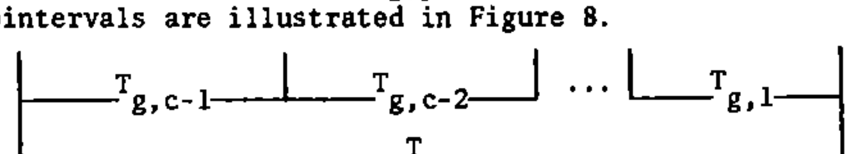

Figure 8. Subintervals within interval T_g.

Subinterval $T_{g,c-1}$ begins immediately after the conclusion of T_f, and this subinterval ends when there are less than c-1 jobs in system. For $1 \leq j \leq c-2$, interval $T_{g,j}$ begins immediately following subinterval $T_{g,j+1}$ and terminates when fewer than j jobs remain in system. Subinterval $T_{g,j}$ has length zero if less than j jobs arrive during the delay; otherwise, the length of $T_{g,j}$ has the distribution of a busy period T_j (cf. LEMMA 1). For $1 \leq j \leq c-1$, define:

$T_{g,j}$ = length of subinterval-j of T_g shown in Figure 8,

$H_{g,j}(t)$ = cdf for $T_{g,j}$,

$\eta_{g,j}(s)$ = LST for the distribution of $T_{g,j}$.

The conditional LST for the distribution of $T_{g,j}$ given that N jobs arrived during delay T_0 is equal to

$$\eta_{g,j}(s|T_0=t,N=n) = \begin{cases} \eta_j(s) & \text{for } j \leq n, \\ 1 & \text{for } n < j. \end{cases}$$

Removing the conditioning on the number of arrivals and length of delay T_0 gives the following result:

$$\begin{aligned} \eta_{g,j}(s) &= \int_{t=0}^{\infty} \sum_{n=0}^{j-1} \left[(\lambda t)^n/n!\right] \exp(-\lambda t)+\eta_j(s) \sum_{n=j}^{\infty} \left[(\lambda t)^n/n!\right] \exp(-\lambda t)dH_0(t) \\ &= \sum_{n=0}^{j-1} \left[(-\lambda)^n/n!\right] \eta_0^{(n)}(\lambda) \left[1-\eta_j(s)\right] + \eta_j(s). \end{aligned}$$

By LEMMA 2, we have

$$\begin{aligned} E(T_e) &= E(T_0) + E(T_f) + E(T_g) , \\ &= E(T_0) + E(T_f) + \sum_{j=1}^{c-1} E(T_{g,j}) . \end{aligned}$$

The steady-state probabilities that the system is in any given state during delay cycle T_e will be defined and calculated using the method of Appendix 2:

$$\begin{aligned} \pi_0 &= \Pr\left[\text{delay } T_0 | \text{delay cycle } T_e \text{ in progress}\right] , \\ &= E(T_0)/E(T_e); \end{aligned}$$

$$\begin{aligned} \pi_f &= \Pr\left[\text{interval } T_f | \text{delay cycle } T_e \text{ in progress}\right] , \\ &= E(T_f)/E(T_e); \end{aligned}$$

and, for $1 \leq j \leq c-1$,

$$\pi_j = \Pr\left[\text{subinterval } T_{g,j} \middle| \text{delay cycle } T_e \text{ in progress}\right],$$
$$= E(T_{g,j})/E(T_e).$$

Evaluating $-\eta'_{g,j}(0)$, the expected length of subinterval $T_{g,j}$ is found to be

$$E(T_{g,j}) = E(T_j)\left\{1 - \sum_{n=0}^{j-1}\left[(-\lambda)^n/n!\right]\ \eta_0^{(n)}(\lambda)\right\}.$$

Substituting the above into the corresponding expression for π_j, we obtain the steady-state probabilities as given in the statement of the theorem. By the random property of Poisson arrivals, these probabilities will also be the probabilities that an arrival finds the system in that particular state, given that delay cycle T_e is in progress. We therefore have

$$\alpha(s|\text{arrival during } T_e) = \pi_0\alpha(s|\text{arrival during } T_0)$$
$$+ \pi_f\alpha(s|\text{arrival during } T_f)$$
$$+ \sum_{j=1}^{c-1} \pi_j\alpha(s|\text{arrival during } T_j).$$

The expected waiting time for a job arriving during delay cycle T_e follows directly from $-\alpha'(0|\text{arrival during } T_e)$. Q.E.D.

4. Applications and Extensions of the Method

In this section we consider a representative sample of applications for the results derived in previous portions of this paper. Furthermore, a discussion is given of ways in which these results may be easily extended to deal with other M/M/c queueing models as well.

4.1 Comparison With the M/G/1 Queueing System

We begin by pointing out similarities in certain results for the M/G/1 and M/M/c queueing systems which have been noted by a number of authors. Consider a busy period T for the M/G/1 system which is initiated by the arrival of a job to an empty system, and define the following:

$$T = \min \left[t: \begin{array}{l} 0 \text{ jobs in the M/G/1 system at time } 0^- \\ 1 \text{ job in the M/G/1 system at time } 0^+ \\ 0 \text{ jobs in the M/G/1 system at time } t \end{array} \right]$$

Λ = Poisson input rate for jobs to the M/G/1 system;

P = General processing time for a job in the M/G/1 system.

If we examine results for busy period T given in reference [6] pp. 149-155, it may be seen from LEMMA 1 that busy period T_c in the M/M/c system (with arrival rate λ and interevent time P_c) appears to have the same distribution as a busy period T in the M/G/1 system when Λ is equal to λ and where P has the same distribution as P_c. Let us use symbol '$\sim$' to denote that two random variables have the same distribution and symbol '$\equiv$' to denote the equivalence of parameters. The above-mentioned situation may then be described as given below:

$$T_c \sim T \quad \text{when } P \sim P_c \text{ and } \Lambda \equiv \lambda \quad .$$

If we examine the expected waiting time for a job in the M/M/c system which arrives to find busy period T_c in progress, it is again found that the waiting time distribution is identical to that for an arrival to the M/G/1 system under the FCFS discipline which finds busy period T taking place if again $\Lambda \equiv \lambda$ and $P \sim P_c$. If we compare the derivations of these results for the M/G/1 and M/M/c systems, it becomes obvious that other M/G/1 results may be easily extended to the M/M/c system as illustrated in the example given below.

<u>Example. Waiting Time for the M/M/c System Under the LCFS Rule.</u>

Consider a M/M/c system which employs the Last-Come-First-Served (LCFS) discipline at the queue so that, at a scheduling epoch, the most recent arrival is chosen for servicing; we assume here that once a job goes into service it is processed to completion. Under both the LCFS and FCFS rules a job arriving to find fewer than c jobs in system immediately goes into service and therefore encounters a waiting time of zero. Jobs are required to wait only when arrival occurs during an interval T_c, and the distribution of busy period T_c is identical under the FCFS and LCFS rules.

Results for the M/G/1 system under the LCFS rule given in reference [6], pp. 155-158 allow us to state the following for the

M/M/c system (using the same notation as used for the M/M/c system under the FCFS rule but with subscripts denoting the scheduling rule):

$$\alpha_{LCFS}(s|\text{arrival during } T_c) = \left[1-\eta_c(s)\right] \Big/ \left\{E(P_c)\left[s+\lambda-\lambda\gamma_c(s)\right]\right\};$$

$$E_{LCFS}(W|\text{arrival during } T_c) = E(P_c^2)\Big/\left\{2[1-\lambda E(P_c)]\,E(P_c)\right\};$$

$$E_{LCFS}(W|\text{arrival during } T_c) = E(P_c^3)\Big/\left\{3[1-\lambda E(P_c)]^2 E(P_c)\right\} + \lambda[E(P_c^2)]^2\Big/\left\{2[1-\lambda E(P_c)]^3 E(P_c)\right\}.$$

The results of THEOREM 3 may then be used to obtain the conditional waiting time distribution for arrival during busy period T_k, where $1 \leq k < c$. It also follows that for the M/M/c queueing system we have similar results to those for the M/G/1 system:

$$E_{LCFS}(W) = E_{FCFS}(W)$$

and

$$E_{LCFS}(W^2) = E_{FCFS}(W^2)/[1-\lambda E(P_c)].$$

This type of analysis can also be applied to analyze waiting time distribution for the M/M/c system under other rules such as the random rule as well.

4.2 Variations Using Delay Cycles

Many modifications of a simple M/M/c queueing system seem to include the concept of delay cycles in one way or another. The following are examples of such modifications.

Example. Multiprocessor Facility With 'Down' Pauses at the End of Busy Periods.

Consider a multiprocessor service facility which can simultaneously process up to c jobs at one time. Whenever the system is empty, all c processors are assigned to some other obligation which takes a time T_0 having a general distribution. The distribution of waiting time in such a system is that for arrival during a delay cycle T_e derived earlier (cf. THEOREMs 4,5,6). We assume here that at the conclusion of the interval T_0 the queue is examined and if empty another interval T_0 is initiated. This will guarantee that every arrival to the system finds a delay cycle T_e in progress.

Example. Multichannel Facility With 'Warm-Up' Time. Consider a multichannel system in which the processor, once idle, needs some warm-up time or a setup time T_0^*, of a general distribution, after the arrival of the first job. After this time T_0^*, it starts processing the jobs and continues processing until the system is empty. The analysis of this system is done essentially in the same manner as for the delay cycle T_e. The difference lies in the small changes required to account for the first job which was already in the system at the beginning of the delay T_0^*. This job is the one responsible for initiating the delay cycle in which the warm-up time acts as the delay interval. Figure 9 illustrates the important random variables involved in the analysis.

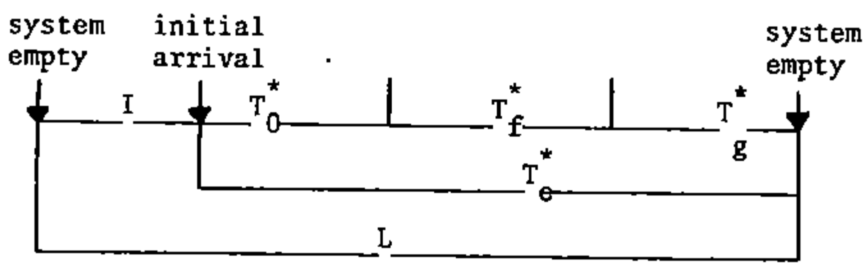

Figure 9. Busy-Idle Cycle and Subintervals.

The modified delay cycle T_e^* is comprised of subintervals T_0^* (warm-up time), T_f^* (during which c or more jobs are in system), and T_g^* defined in a manner similar to the subintervals of delay cycle T_e. Idle period I is an exponentially distributed interarrival time with mean $1/\lambda$, and busy-idle cycle L is the sum of an idle period I and a modified delay cycle T_e^*. It follows that

$$E(I) = 1/\lambda \ ;$$

$$E(L) = E(I) + E(T_e^*) \ ;$$

$$E(T_e^*) = E(T_0^*) + E(T_f^*) + E(T_g^*) \ ;$$

and the probabilities of an arrival finding an idle period I or modified delay cycle T_e^*, given that busy-idle cycle L is in progress, are

$$\pi_I = \Pr[\text{idle period I in progress}],$$
$$= E(I)/E(L);$$

$$\pi_e = \Pr[\text{delay cycle } T_e^* \text{ in progress}],$$
$$= E(T_e^*)/E(L).$$

Since every job arrives during a busy-idle cycle L, the above are also the unconditional steady-state probabilities. Assume that the FCFS discipline is employed; a job arriving during idle period I has waiting time equal to the warm-up time T_0^*, and the unconditional expected waiting time is

$$E(W) = \pi_I E(T_0^*) + \pi_e E(W|\text{arrival during } T_e^*).$$

The conditional expected waiting time of a job arriving during modified delay cycle T_e^* can be obtained in essentially the same manner as for delay cycle T_e. Refer to the derivation and discussion for LEMMA 2; if N jobs arrive during delay T_0^*, the remainder of T_e^* is the sum of busy periods $T_{N+1}, T_N, \ldots, T_1$ (rather than T_N through T_1 as for delay cycle T_e). The presence of the initial job at the start of interval T_e^* requires the changes shown below for the expected length of subintervals T_f^* and T_g^*:

$$E(T_f^*) = E(T_c)\left\{\Pr[N\geq c-1]\ E(N|N\geq c-1)-(c-2)\Pr[N\geq c-1]\right\},$$

$$= E(T_c)\left\{\left[\lambda E(T_0^*) - \sum_{n=0}^{c-2}\left[(-\lambda)^n/n!\right]\eta_0^{(n)}(\lambda)n\right] - (c-2)\left[1-\sum_{n=0}^{c-2}\left[(-\lambda)^n/n!\right]\eta_0^{(n)}(\lambda)\right]\right\},$$

$$E(T_g^*) = \sum_{k=1}^{c-1} E(T_k)\Pr[N\geq k-1],$$

$$= E(T_1) + \sum_{k=2}^{c-1} E(T_k)\left\{1 - \sum_{n=0}^{k-2}\left[(-\lambda)^n/n!\right]\eta_0^{(n)}(\lambda)\right\}.$$

The expected conditional waiting time for an arrival during T_0^* requires similar changes as compared to the results given in THEOREM 4:

$$E(W|\text{arrival during } T_0^*) = E((T_0^*)^2)/[2E(T_0^*)]$$
$$+ E(P_c)\left\{[\lambda E((T_0^*)^2)/[2E(T_0^*)] - \sum_{n=0}^{c-3}[(-\lambda)^n/n!]\tau^{(n)}(\lambda)n]\right.$$
$$\left. - (c-2)\left[1 - \sum_{n=0}^{c-3}[(-\lambda)^n/n!]\tau^{(n)}(\lambda)\right]\right\},$$

where $\tau(s) = [1-\eta_0(s)]/[sE(T_0^*)]$ and $\eta_0(s)$ = LST for T_0^*.

The expected conditional waiting times for arrival during T_f^* have the same representation as given in THEOREM 5 except that we must include changes to the distribution of subinterval-0 of T_f^*. That is, we view random variable T_f^* as being composed of subintervals $T_{f,j}^*$ (cf. Figure 7) for $j \geq 0$. If N jobs arrive during the warm-up time T_0^* (in addition to the first arrival which initiated T_0^*), subinterval $T_{f,0}^*$ is the time needed for (N+1)-(c-1) departures to occur, assuming that N is greater than or equal to (c-1). If we denote the number of arrivals during subinterval-j of T_f^* by N_j ($j \geq 0$), subinterval $T_{f,j}^*$ is the time needed for N_{j-1} departures to occur. Comparing these definitions with those given for the subintervals of T_f of THEOREM 5, we see that only the definition for subinterval-0 has been changed. The waiting time analysis for T_f applies also to T_f^* if we substitute results for subinterval-0 of T_f^*. Denote the LST for the distribution of $T_{f,0}^*$ (i.e. subinterval-0 of T_f^*) by $\eta_{f,0}^*(s)$; the conditional LST given that N arrivals occur during the warm-up time T_0^* is given as

$$\eta_{f,0}^*(s|T_0^*=t, N=n) = \begin{cases} 1 & \text{if } 0 \leq n < c-2, \\ [\gamma_c(s)]^{n-(c-2)} & \text{if } n > c-2 . \end{cases}$$

Removing the conditioning we obtain

$$\eta_{f,0}^*(s) = \sum_{n=0}^{c-3} [(-\lambda)^n/n!]\eta_0^{(n)}(\lambda) + \left\{ \eta_0(\lambda-\lambda\gamma_c(s)) - \sum_{n=0}^{c-3} [(-\lambda\gamma_c(s))^n/n!]\eta_0^{(n)}(\lambda) \right\} / [\gamma_c(s)]^{c-2}.$$

The first two moments of $T_{f,0}^*$ are directly found to be

$$E(T_{f,0}^*) = E(P_c) \left\{ \lambda E(T_0^*) - \sum_{n=0}^{c-3} [(-\lambda)^n/n!]\eta_0^{(n)}(\lambda)\cdot n - (c-2) \left[1 - \sum_{n=0}^{c-3} [(-\lambda)^n/n!]\eta_0^{(n)}(\lambda)\right] \right\};$$

$$\begin{aligned} E((T_{f,0}^*)^2) = [E(P_c)]^2 \Bigg\{ & -2(c-2) \left[\lambda E(T_0^*) - \sum_{n=0}^{c-3} [(-\lambda)^n/n!]\eta_0^{(n)}(\lambda)\cdot n\right] \\ & + \lambda^2 E((T_0^*)^2) - \sum_{n=0}^{c-3} [(-\lambda)^n/n!]\eta_0^{(n)}(\lambda) n\cdot(n-1) \\ & + (c-2)(c-1) \left[1 - \sum_{n=0}^{c-3} [(-\lambda)^n/n!]\eta_0^{(n)}(\lambda)\right] \Bigg\} \\ + E(P_c^2) \Bigg\{ & \lambda E(T_0^*) - \sum_{n=0}^{c-3} [(-\lambda)^n/n!]\eta_0^{(n)}(\lambda)\cdot n \\ & - (c-2) \left[1 - \sum_{n=0}^{c-3} [(-\lambda)^n/n!]\eta_0^{(n)}(\lambda)\right] \Bigg\}, \end{aligned}$$

and THEOREM 5 gives $E(W|\text{arrival during } T_f^*)$ by replacing $E(T_{f,0})$ and $E(T_{f,0}^2)$ with $E(T_{f,0}^*)$ and $E((T_{f,0}^*)^2)$, respectively.

Random variable T_g^* may be viewed as being a sequence of subintervals $T_{g,c-1}^*, T_{g,c-2}^*, \ldots, T_{g,1_*}^*$ (defined as the subintervals of T_{g_*} given in THEOREM 6). Given that $T_{g,k}^*$ has length greater than zero, $T_{g,k}^*$ has the distribution of busy period T_k. In order to apply THEOREM 6, we need only modify the probability π_k, for $1 \le k \le c-1$, that an arrival during T_e^* finds subinterval $T_{g,k}^*$ in progress; the expected length of $T_{g,k}^*$ is given by

$$E(T_{g,k}^*) = E(T_k)*\Pr[N \ge k-1],$$

where N is the number of arrivals during T_0^*.

It follows from THEOREM 6 that

$$\begin{aligned} E(W|\text{arrival during } T_e^*) = {} & \pi_0 E(W|\text{arrival during } T_0^*) \\ & + \pi_f E(W|\text{arrival during } T_f^*) \\ & + \sum_{k=1}^{c-1} \pi_k E(W|\text{arrival during } T_k) \ , \end{aligned}$$

where $\pi_0 = E(T_0^*)/E(T_e^*)$, $\pi_f = E(T_f^*)/E(T_e^*)$, $\pi_1 = E(T_1)/E(T_e^*)$,

and for $2 \le k \le c-1$, $\pi_k = E(T_k)\left\{1 - \sum_{n=0}^{k-2} [(-\lambda)^n/n!]\eta_0^{(n)}(\lambda)\right\}/E(T_e^*)$.

This completes the waiting time analysis for the example problem.

4.3 Variations Using Busy Periods

We next consider an example in which results for the busy period T_k of LEMMA 1 find application. This will allow an analysis of the waiting time results of THEOREMS 2 and 3.

Example. Multiprocessor Facility With Start-Up Based on Number in System.

Consider a multichannel system in which due to the high cost of starting and servicing, the processor waits until its use is warranted by a certain number, M, of jobs which have arrived to the system. Once the processor starts processing, it continues in operation until the system is empty. Let us assume that $M \le c$ in the analysis which follows; Figure 10 illustrates various random variables of interest for the given example problem.

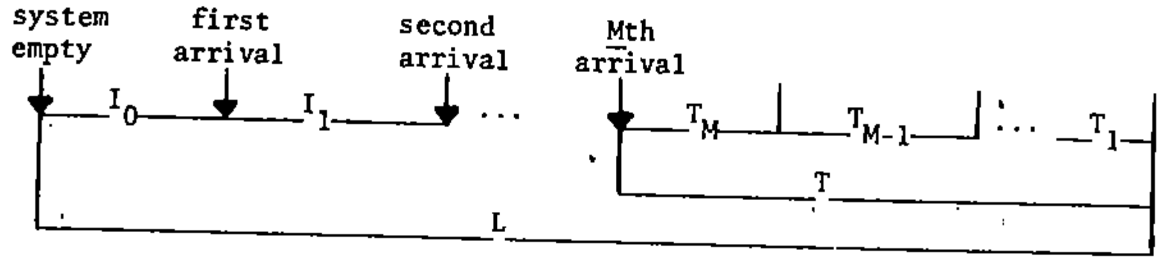

Figure 10. Busy-idle Cycle and Subintervals.

A busy period T in this case consists of the sum of busy periods T_M, $T_{M-1}, \ldots, T_1$ (here it is appropriate to interpret T_k as the time needed to achieve an overall reduction, by one, of the jobs in system). We identify M idle periods, defined as given below for $0 \leq j \leq M-1$:

I_j = idle period during which j jobs are in system; this interval represents an interarrival time for a Poisson process with rate λ and therefore $E(I_j) = 1/\lambda$.

Busy-idle cycle L is the sum of busy period T and the M idle periods, from which it follows that

$$E(L) = \sum_{j=0}^{M-1} E(I_j) + \sum_{k=1}^{M} E(T_k)$$, where $E(T_k)$ is given by LEMMA 1.

A job arriving during idle period I_j becomes the (j+1)-st job in system and so must wait until (M-(j+1) additional jobs arrive before going into service. It follows that for $0 \leq j \leq M-1$,

$$E(W|\text{arrival during } I_j) = (1/\lambda)(M-j-1).$$

The unconditional waiting time is easily obtained by utilizing the method of Appendix-2 to find the steady-state probability that an arrival finds the system in any particular state.

$$E(W) = \sum_{j=0}^{M-1} [E(I_j)/E(L)]*E(W|\text{arrival during } I_j) + \sum_{k=1}^{M} [E(T_k)/E(L)]*E(W|\text{arrival during } T_k) ,$$

where $E(W|\text{arrival during } T_k)$ is found using THEOREMs 2 and 3.

4.4 Unequal Channel-Service-Rates

In some practical situations it is possible to have unequal channel-service-rates in multichannel systems. For example, a job with highest internal priority in a multiprogrammed computer system receives preferential treatment and may in effect get a higher service rate.

Let the service-rate for channel-j be μ_j, where $1 \leq j \leq c$, and let the channels be ordered such that

$$\mu_1 \geq \mu_2 \geq \mu_3 \geq \ldots > \mu_c .$$

Assume that it is possible to have instantaneous switching of channels and that, when there are j jobs in service, the first j channels will be servicing jobs. A new job, when started, receives the highest available service rate. The service rate μ_i of a job, being processed, is instantaneously changed to the higher rate μ_{i-1} as soon as this rate is available. Any rate μ_i is said to be available when the job with that rate is either finished or when that job's service rate is changed to a higher rate. For example, consider a 3-channel system with all three servers busy. Suppose that the job with rate μ_1 is finished; the job with rate μ_2 will then have rate μ_1, and the job with rate μ_3 will then have rate μ_2. If the queue is not empty, the first job in the queue will have its processing started and will receive service at rate μ_3.

The analysis of busy period T_k of LEMMA 1 and corresponding conditional waiting time results given in THEOREMs 2 and 3 may be easily modified to deal with this case. In LEMMA 1, THEOREM 2, and THEOREM 3, it is only necessary to change the distribution of random variable P_k (the aggregate of k Poisson processes associated with the jobs in service) as shown below for $1 \leq k \leq c$:

$$\gamma_k(s) = \left\{ \sum_{i=i}^{k} \mu_i \right\} / \left\{ s + \sum_{i=i}^{k} \mu_i \right\} ,$$

$$E(P_k) = 1/\left\{ \sum_{i=i}^{k} \mu_i \right\} ,$$

$$E(P_k^2) = 2/\left\{ \sum_{i=i}^{k} \mu_i \right\}^2 .$$

If it is not possible to have switching of service rates, the analysis quickly becomes cumbersome for increasing values of c, although small values such as c =2 can be readily handled.

5. Summary.

This paper has demonstrated that the method of busy period analysis previously used for treating the M/G/1 queueing system can be extended to deal with the M/M/c queueing system as well. Closed-form results have been presented for the distribution of two major types of busy periods arising in the M/M/c system and for the distribution of waiting time (under the FCFS discipline) for an arriving job which finds a particular type of busy period in progress.

A number of examples have been presented which show that these results may be usefully applied and extended to deal with a number of different models for multiprocessor systems. These examples included the following:

--- M/M/c System Under the LCFS Rule.

--- Multiprocessor Facility With 'Down' Pauses at the End of Busy Periods.

--- Multichannel Facility With 'Warm-Up' Time.

--- Multiprocessor Facility With Start-Up Based on Number of Jobs in System.

--- Multichannel System With Unequal Channel-Service-Rates.

It is the hope of the authors that the results in this paper will serve not only to illustrate the usefulness of the method of busy period analysis for M/M/c systems but also to give the reader insights in the characteristics of this class of multiprocessor queueing systems.

References

1. Omahen, K. J. Analytic models of multiple resource systems. Ph.D. Thesis, Committee on Information Sciences, University of Chicago, June, 1973.

2. Marathe, V. P. Priority queueing systems with simultaneous server requirements. Ph.D. Thesis, Operations Research, Cornell University, May, 1972.

3. Omahen, K. and Marathe, V. A queueing model for a multiprocessor system with partitioned memory. Technical Report CSD-TR 132, Dept. of Computer Science, Purdue University, January, 1975.

4. Cobham, A. Priority assignment in waiting line problems. Operations Research 2, 1 (1954).

5. Avi-Itzhak, B., Maxwell, W. L., and Miller, L. W. Queueing with alternating priorities. Operations Research 13 (1965), 306-318.

6. Conway, R. W., Maxwell, W. L., and Miller, L. W. Theory of Scheduling. Addison-Wesley, Reading, Massachusetts, 1967.

7. Cox, D. R. Renewel Theory. London: Methuen, 1962.

8. Strauch, R. E. When a queue looks the same to an arriving customer as to an observer. Management Science 17 (1970), 140-141.

9. Widder, D. V. The Laplace Transform. Princeton University Press, Princeton, New Jersey, 1946.

10. Little, J. D. C. A proof of the queueing formula $L = \lambda W$. Operations Research 9 (1961), 383-387.

APPENDIX-1

PROPERTIES OF POISSON PROCESSES

Consider a Poisson process with rate λ; such a process is characterized by a sequence of interevent times which are independent and exponentially distributed. Define:

T = time between successive events for the Poisson process.

$G(t) = Pr[T \leq t] = 1 - \exp(-\lambda t)$, $t \geq 0$.

$\gamma(s)$ = LST for the distribution of $T = \lambda/(s+\lambda)$.

The first and second moments for the distribution of interevent times are

$E(T) = 1/\lambda$, and $E(T^2) = 2/\lambda^2$.

Define another random variable as follows:

$N(t)$ = Number of events which take place during interval t for a Poisson process with rate λ.

The distribution for $N(t)$ has the following characteristics:

$Pr[N(t) = n] = [(\lambda t)^n/n!] \exp(-\lambda t)$ for integer $n \geq 0$,

$E(N(t)) = \lambda t$ for $t \geq 0$.

The Poisson process has many properties which are utilized in the body of this paper; these properties are summarized below. The proofs for these properties are available in a number of textbooks (e.g. see Reference [6]).

A1.1 The Memoryless Property

Suppose that we are interested in the distribution for interevent time T, given that a certain amount of time y has already passed without an event taking place. Poisson processes have the unique property that

$$Pr[T < y+t | T > y] = 1 - \exp(-\lambda t).$$

The process is memoryless in the sense that the distribution for the remaining time until the next event does not depend upon the amount of time which has passed without an event taking place.

A1.2 Aggregation and Branching of Poisson Processes

A1.2.1 Assume that there are n Poisson processes simultaneously in progress and that events associated with the kth process are taking place at rate λ_k, where $k = 1,2,\ldots,n$. The aggregate of these processes will be defined such that events associated with each of the n Poisson processes will be considered to be events for the aggregate. The aggregate of these Poisson processes is also Poisson with rate $\lambda = \sum_{i=i}^{n} \lambda_i$.

A1.2.2 Consider the situation in which events associated with a Poisson process with rate λ are subjected to a decision process whereby each event is (instantly) mapped into one of n classes. Every time an event occurs, the decision process maps the event into class-k independently and with probability λ_k, where $k = 1,2,\ldots,n$ and where the sum of these probabilities is equal to one. If we define n processes, where the kth process consists of those events mapped into class-k, each process constitutes a Poisson process with rate λ_k. Therefore, we have the situation in which a Poisson process branches into independent Poisson processes.

A1.2.3 As a consequence of A1.2.1 and A1.2.2, we obtain the following result. Given that we have an aggregate of n Poisson processes and an event occurs, the probability that the event is associated with the kth Poisson process ($1 \le k \le n$) is equal to λ_k/λ.

A1.3 The Random Property

Poisson events are frequently referred to as "random events" because of the property described below. Given that an event occurs during an interval of length t, the instant at which the event takes place is uniformly distributed over the length of the interval, i.e.

$$\Pr[y \le \text{instant of event occurrence} \le y+dy \mid \text{event in } t]$$
$$= dy/t \text{ for } 0 \le y \le t.$$

A1.4 The Random Modification

Assume that an event associated with a Poisson process takes place during interval X. Consider the time Y between the occurrence of the Poisson event and the end of interval X; this interval Y will be called the random modification. Define:

X = length of some interval having an arbitrary distribution,

$G(x)$ = cdf for r.v. X,

$\gamma(s)$ = LST for the distribution of X;

Y = length of random modification of variable X,

$H(y)$ = cdf for r.v. Y,

$\tau(s)$ = LST for the distribution of Y.

We have the following results for the random modification:

(a) $\Pr[y \le Y \le y+dy,\ x \le X \le x+dx] = dG(x)dy/E(X),\ 0\le y\le x,\ 0\le x\le\infty.$

(b) $dH(y) = [1-G(y)]dy/E(X)$ for $0 \le y \le \infty$.

(c) $\tau(s) = [1-\gamma(s)]/[sE(X)]$, and $E(Y^k) = E(X^{k+1})/\{(k+1)E(X)\}, k \ge 1$.

APPENDIX-2

A METHOD FOR DETERMINING STEADY-STATE PROBABILITIES

We will often be interested in determining the steady-state probability that the system is in some specified state, given that an interval T is in progress. The possible states of the system during interval T will be denoted by $S_1, S_2, \ldots, S_n$. Associated with each state S_k is a random variable X_k which represents the amount of time that the system remains in state S_k upon a transition to that state, where $k = 1,2,\ldots,n$. For $1 \leq k \leq n$, define:

$$N_k = \begin{cases} 1 & \text{if } X_k \text{ is in progress within interval } T, \\ 0 & \text{otherwise.} \end{cases}$$

The steady-state probability that the system is in state S_k given that interval T is in progress is therefore

$$\Pr[S_k|T] = E(N_k) \quad \text{for } k = 1,2,\ldots,n.$$

If we examine the system operation only during those times that intervals of type T are in progress, the intervals of type T appear to be initiated at rate λ given by

$$\lambda = 1/E(T),$$

and intervals of type X_k appear to be initiated at rate λ_k given as

$$\lambda_k = r_k\lambda,$$

where r_k is the relative rate at which intervals X_k are initiated given that an interval T is in progress.

Using Little's Equation [10], we find

$$E(N_k) = \lambda_k E(X_k),$$
$$= r_k E(X_k)/E(T);$$

therefore, we have the following result:

$$\Pr[S_k|T] = r_k E(X_k)/E(T).$$

If there are n mutually exclusive and exhaustive system states during T, it will obviously be the case that

$$\sum_{k=1}^{n} r_k E(X_k)/E(T) = 1 \qquad \text{or} \qquad E(T) = \sum_{k=1}^{n} r_k E(X_k) \;.$$