#250



NAVAL RESEARCH LOGISTICS QUARTERLY VOL. 11 NO. 1 MARCH 1964

# ANALYSIS OF A TIME-SHARED PROCESSOR*

Leonard Kleinrock†

*Department of Engineering*
*University of California*
*Los Angeles, California*

## ABSTRACT

This paper analyzes a queueing structure for a time-shared service facility (or processor) and compares these results with a straightforward first-come first-served discipline. The assumption is that the processing time for each job is chosen from a geometric distribution. This time-shared discipline shares the desirable features of a first-come first-served principle, as well as that of a discipline which services short jobs first. It is shown that those jobs with shorter than average processing requirements, spend less time in the queue than they would in a strict first-come first-served system, and conversely for longer than average jobs.

## INTRODUCTION

Recently a great deal of interest has been expressed in various forms of time-shared computing systems. The motivation for such interest is directed primarily toward encouraging the interaction between the user (programmer) and the computer itself. Specifically the hope is that time-shared systems will economize (in some sense) the user's time and will encourage the formulation of new problems for computer solution; it should be noted that no claims can be made that a system of time-sharing will improve the efficiency in use of the central computer itself (except insofar as more efficient programs may emerge from such a system).

This article considers a simple "round-robin" time-shared service facility and compares its behavior with that of a strict first-come first-served system. The model of the round robin system is set up so as to include the advantages of a first-come first-served system, as well as those of a discipline which services short computational jobs first.

## THE MODEL

Let time be quantized into segments each Q seconds in length. At the end of each time interval, a new unit (or job) arrives in the system with probability $\lambda Q$ (result of a Bernoulli trial); thus, the average number of arrivals per second is $\lambda$. The service time (i.e., the required processing time) of a newly arriving unit is chosen independently from a geometric distribution such that for $\sigma < 1$

$$s_n = (1-\sigma)\,\sigma^{n-1} \qquad n = 1, 2, 3, \ldots, \tag{1}$$

*This material is to appear in a forthcoming book in the Lincoln Laboratory Publication Series entitled, Message Delay in Communication Nets with Storage, by Leonard Kleinrock (McGraw-Hill Book Co. Inc., New York).

†This work was done while the author was employed at Lincoln Laboratory (Operated with support from the U.S. Army, Navy, and Air Force), Massachusetts Institute of Technology, Lexington, Massachusetts.

where $s_n$ is the probability that a unit's service time is exactly n time intervals long (i.e., that its service time is nQ seconds).

The procedure for servicing is as follows: a newly arriving unit joins the end of the queue, and waits in line in a first-come first-served fashion until it finally arrives at the service facility. The server picks the next unit in the queue, and performs one unit of service upon it (i.e., it services this job for exactly Q seconds). At the end of this time interval, the unit leaves the system if its service (processing) is finished; if not, it joins the end of the queue with its service partially completed, as shown in Figure 1. Obviously, a unit whose processing requirement is n time units long will be forced to join the queue n times in all before its service is completed.

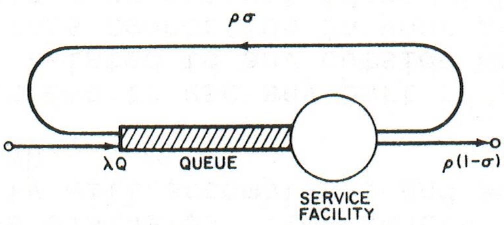

Figure 1 - The round-robin time-shared service system

Another assumption must now be made regarding the order in which events take place at the end of a time interval. We consider two types of systems: The first system allows the unit in service to be ejected from the service facility (and then allows it to join the end of the queue, if more service is required for this unit), and instantaneously after that a new unit arrives (with probability $\lambda Q$). We call this a late arrival system. The second system reverses the order in which these events are allowed to occur, giving rise to the early arrival system. In both systems, a new unit is taken into service at the beginning of a time interval.

RESULTS

First we consider the late arrival system, which is similar to a system considered by Jackson [2] for a different class of priority systems. He arrives at the solution for the steady state probability, $r_k$, that there are k units in the system just before the time when an arrival is allowed to occur (i.e., just after the time when a unit is ejected from service if there was a unit in service); Jackson's result is

(2) $$r_k = (1 - a)\, a^k ,$$

where

$$a = \frac{\rho\sigma}{1 - \lambda Q}$$

and

$$\rho = \frac{\lambda Q}{1 - \sigma}.$$

This definition of $\rho$ is the product of the average arrival rate $\lambda$ and the mean service time, $Q/(1 - \sigma)$; in queueing theory, $\rho$ is referred to as the utilization factor and, as we shall see below, plays a crucial role in determining queue lengths, waiting times, and so on. The notation of Jackson's result has been altered to correspond to that used in this article. From this we quickly obtain the expected value, E, of the number k as*

(3) $$E = \frac{\rho\sigma}{1-\rho}.$$

These results also apply to the time-shared service facility under consideration. For the time-shared system, we now state the theorem:

THEOREM 1:* The expected value, $T_n$, of the total time† spent in the late arrival system for a job whose service time is nQ seconds, is

(4) $$T_n = \frac{nQ}{1-\rho} - \frac{\lambda Q^2}{1-\rho}\left[1 + \frac{(1-\sigma\alpha)(1-\alpha^{n-1})}{(1-\sigma)^2(1-\rho)}\right],$$

where

$$\alpha = \sigma + \lambda Q.$$

In the appendix we show that $\alpha < 1$. An upper bound for $T_n$ is easily obtained (by lower bounding the bracket above, by unity) as

(5) $$T_n \le \frac{Q}{1-\rho}(n - \lambda Q).$$

We now consider the early arrival system. Let $r_k$ be the steady-state probability that there are k units in the system just after the time when an arrival is allowed to occur (i.e., just before the time when a unit is ejected from service if there is a unit in service). In the appendix we show that

(6) $$r_k = \begin{cases} 1-\rho & k = 0 \\ \dfrac{1-\rho}{\sigma}a^k & k = 1, 2, \ldots, \end{cases}$$

where a and $\rho$ are defined just as in the late arrival system. From this we obtain E', the expected value of the number k, as

(7) $$E' = \frac{\rho}{1-\rho}(1-\lambda Q).$$

*See the appendix for proof of Eq. (3) and of Theorem 1.
†$T_n$ is the sum of the time spent in the queue and the time spent in the service facility.

THEOREM 2:* The expected value, $T_n$, of the total time spent in the early arrival system for a unit whose service time is nQ seconds is

(8) $$T_n = \frac{nQ}{1-\rho} - \rho Q - \frac{\lambda Q^2 \rho}{1-\rho}\left[1 + \frac{(1-\sigma\alpha)(1-\alpha^{n-1})}{(1-\sigma)^2(1-\rho)}\right],$$

where $\alpha$ is defined as before. An upper bound for $T_n$ is easily obtained (by lower bounding the bracket above the unity) as

(9) $$T_n \leq \frac{Q}{1-\rho}(n - \lambda Q \rho) - \rho Q .$$

We now consider the case in which all units wait for service in order of arrival, and once in service, each unit remains until it is completely serviced. It is then easy to show that with $T_n$ defined as before, we get Theorem 3.

THEOREM 3:* The expected value, $T_n$, of the total time spent in the strict first-come first-serve system for a unit whose service time is nQ seconds is

(10) $$T_n = \frac{QE}{1-\rho} + nQ ,$$

where, once again,

$$E = \frac{\rho\sigma}{1-\rho} .$$

Note that the distinction between the early and late arrival systems has disappeared, as, of course, it must. Note also that the expression defining E is the same as that in Eq. (3) which is the average number of units in the late arrival system.

Let us now compare some of these results for time-shared systems. First, we compare the value of E and E'. Let $\Delta$ be the difference between the expected number of units in the early and late arrival systems. Then

$$\Delta = \frac{\rho}{1-\rho}(1-\lambda Q) - \frac{\rho}{1-\rho}\sigma ,$$

and so

(11) $$\Delta = \rho(1-\sigma) = \lambda Q .$$

This result is quite reasonable, since for $\sigma$ equal to zero (which says that each service time equals one time interval exactly) the difference $\Delta$ should be the probability of finding a unit in

*See the appendix for proof of theorems.

the early arrival system (which is merely $\rho$);* and for $\sigma$ approaching unity, the difference approaches zero since, with probability $1 - \sigma$ a unit will leave the system before (after) the next arrival. Note that $\Delta$ is always less than unity.

Now, if one wishes an approximate solution to the round-robin system, one might argue as follows: A unit whose service time is nQ seconds must enter the end of the queue exactly n times. Roughly speaking, each time the unit enters the queue it finds E (or E') units ahead of it (this approximation is evaluated presently). The time spent waiting for service each time around is then approximately QE. The time actually spent in service is exactly nQ. Thus, the approximation to $T_n$, which we label as $T'_n$ is

$$T'_n = nQE + nQ . \tag{12}$$

Upon comparing this to Eq. (10) for the strict first-come first-served system in which

$$T_n = \frac{1}{1 - \sigma} QE + nQ ,$$

we see that for units which require a number of service intervals less (greater) than $1/(1 - \sigma)$, the round-robin waiting time (for the late arrival system) is less (greater) than the strict first-come first-served system. One notes, however, that the average number of service intervals is exactly $1/(1 - \sigma)$. Thus, for this approximate solution, the crossover point for waiting time is at the mean number of service intervals. An evaluation of this approximation may be obtained by comparing the quantity $T_n/Q$ as given in Eq. (4) and $T'_n/Q$ as given in Eq. (12). That is, the approximation is only as good as the agreement between these two (for the late arrival system†):

$$\frac{n}{1 - \rho} - \frac{\lambda Q x}{1 - \rho} \longleftrightarrow \frac{n}{1 - \rho} - \frac{\lambda Q n}{1 - \rho} , \tag{13}$$

where

$$x = 1 + \frac{(1 - \sigma\alpha)(1 - \alpha^{n-1})}{(1 - \sigma)^2 (1 - \rho)} .$$

In Figures 2-4, curves of $(1 - \sigma)/(\sigma Q)\, W_n(\rho)$ are plotted‡ to show the general behavior of the round-robin structure for the late arrival system. On each graph, points corresponding to the first-come first-served case have also been included. The normalization $(1 - \sigma)/(\sigma Q)$ is used for convenience so that for the first-come first-served case, we obtain the curve

*That is, a well-known result in queueing theory (see for example, Rice [5], p. 272) is that the probability of finding a non-empty system is equal to $\rho$.

†For the early arrival system, we compare

$$\frac{n}{1 - \rho} - \frac{\lambda Q \rho x}{1 - \rho} - \rho \longleftrightarrow \frac{n}{1 - \rho} - \frac{\lambda Q n}{1 - \rho} .$$

‡In all the results of this article, the expected value $W_n$ of the time spent in the queue for a unit which requires n intervals of service is obtained from $T_n$ by $W_n = T_n - nQ$.

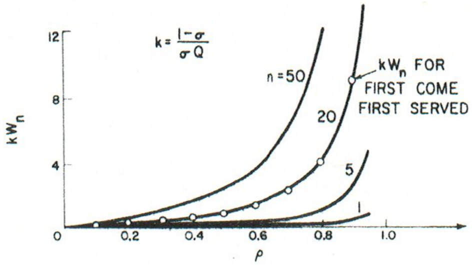


Figure 2 - $[(1 - \sigma)/(\sigma Q)\ W_n\ (\rho)]$ for the late-arrival time-shared service system. ($\sigma$ = 19/20)

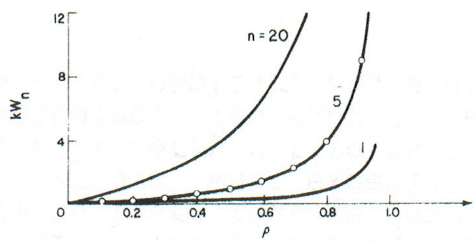


Figure 3 - $[(1 - \sigma)/(\sigma Q)\ W_n\ (\rho)]$ for the late-arrival time-shared service system. ($\sigma$ = 4/5)

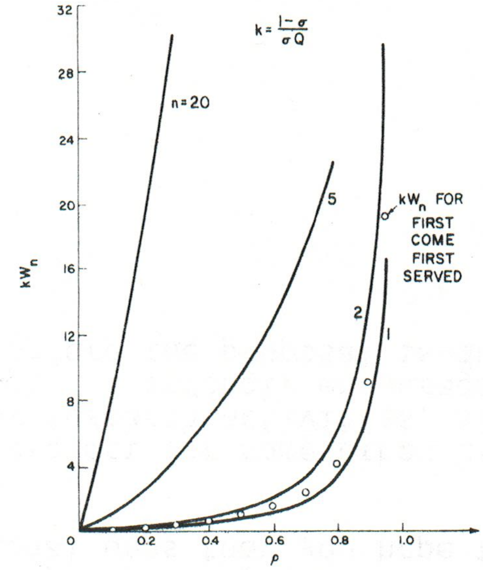


Figure 4 - $[(1 - \sigma)/(\sigma Q)\ W_n\ (\rho)]$ for the late-arrival time-shared service system. ($\sigma$ = 1/5)

$\rho/(1 - \rho)$, which is a function only of $\rho$. Note that the only parameter change among Figures 2-4 is the value of $\sigma$.

Figures 2-4 indicate the accuracy of the approximation discussed above in which the crossover point for waiting times is at the mean number of service intervals, $1/(1 - \sigma)$. In Figures 2 and 3 there is no noticeable difference (on the scale used) between the first-come first-served points, and the curve for $n = 1/(1 - \sigma)$; moreover, in Figure 4 the points fall between the curves for $n = 1$ and $n = 2$, since $1/(1 - \sigma) = 1.25$.

It is interesting to note that both round-robin disciplines, along with the first-come first-served discipline offer an example of the validity of the conservation law (see Kleinrock [3]). That is, if we define

$T_n$(FCFS) as given by Eq. (10),

$T_n$(LAS) as given by Eq. (4),

$T_n$(EAS) as given by Eq. (8),

and also

$$W_n(\cdot) = T_n(\cdot) - nQ,$$

then it is a simple algebraic exercise to show that

$$\sum_{n=1}^{\infty} \rho_n W_n(\cdot) = \text{constant} = \frac{Q\rho^2 \sigma}{(1-\rho)(1-\sigma)}, \tag{14}$$

where

$$\cdot = \{\text{FCFS, LAS, EAS}\}$$

and

$$\rho_n = \rho s_n = \rho(1-\sigma)\,\sigma^{n-1}.$$

But $\lambda_n = \lambda s_n$ = average arrival rate of units which require n service intervals. Thus, Eq. (14) may be rewritten as

$$W \equiv \sum_{n=1}^{\infty} (\lambda_n/\lambda)\, W_n(\cdot) = \text{constant} = \frac{Q\rho\sigma}{(1-\rho)(1-\sigma)}. \tag{15}$$

This equation states that the mean waiting time,* W, is a constant which is independent of the three disciplines discussed. Thus, as is to be expected, one does not improve the mean wait, W; however, by introducing the round-robin system analyzed in this article, one manipulates the relative waiting time for different jobs (while maintaining a constant W), and thus imposes a time-sharing system which gives preferential treatment to short jobs.

## APPENDIX

Let us first prove Eq. (3), which is the expected value of the distribution $r_k$, where

$$r_k = (1-a)\,a^k:$$

clearly,

$$E = \sum_{k=0}^{\infty} k r_k$$

$$E = \frac{a}{1-a}$$

But

$$a = \frac{\sigma\rho}{1-\lambda Q}$$

and so

*This mean waiting time is an appropriate average of $W_n$ since it weights the waiting time $W_n$ by the fractional number of jobs $(\lambda_n/\lambda)$ which must suffer that waiting time.

$$E = \frac{\rho\sigma}{1 - \rho}$$

where we have used the definition of $\rho = \lambda Q/(1 - \sigma)$ as before. This establishes Eq. (3).

PROOF OF THEOREM 1: The following arguments are based exclusively on expected values. Consider the arrival of a unit (the tagged unit) whose service time is nQ seconds. Let $D_i$ be the expected value of the delay (or time spent) between the completion of its $(i-1)^{st}$ service interval, and the completion of its $i^{th}$ service interval. We complete this definition by assuming that the completion of its $0^{th}$ service interval occurs at its time of arrival. Clearly then, $T_n$, the expected value of the total time spent in the system for such a unit, will be

$$T_n = \sum_{i=1}^{n} D_i .$$

Let us further define $N_i$ as the expected number of units which are serviced between the completion of the $(i - 1)^{st}$ and $i^{th}$ service interval of the tagged unit, i.e.,

$$N_i = \frac{D_i}{Q} ,$$

and so

$$T_n = Q \sum_{i=1}^{n} N_i . \tag{16}$$

We now derive a general form for $N_i$. Upon its arrival to the system, the tagged unit finds a certain number of units in the queue, the expected value of which is E by definition. Note that the service facility is empty whenever a new unit enters the system. Thus

$$N_1 = E + 1 .$$

The addition of unity is due to the service interval used up in serving the tagged unit's first time interval. Now, each of these E units will remain in the system with probability $\sigma$, and so $\sigma(N_1 - 1)$ of them will contribute to $N_2$. In addition, during the time $Q(N_1 - 1)$, devoted to servicing these E units, we expect $\lambda$ new units to arrive per second, and so we must also add $\lambda Q(N_1 - 1)$ more units to $N_2$. Besides all this, for $n > 1$, we must add one more (the tagged unit itself) to $N_2$, giving

$$\begin{aligned} N_2 &= \sigma(N_1 - 1) + \lambda Q(N_1 - 1) + 1 \\ &= (\sigma + \lambda Q)\, E + 1 . \end{aligned}$$

In calculating $N_3$, we see that a fraction $\sigma$ of the units which were served before the second time interval of the tagged unit will remain in the system, i.e., $\sigma(N_2 - 1)$. In addition during the time $Q(N_2 - 1)$ devoted to servicing these units, $\lambda Q(N_2 - 1)$ new units will arrive. Also, for $n > 2$, we must add one more (the tagged unit again) to $N_3$. However, we, now notice a new effect entering, namely, the presence of a unit which arrived (with probability $\lambda Q$) at the conclusion of the first service interval of the tagged unit. This additional unit was placed in back of the tagged unit when it arrived, and therefore did not appear in $N_2$. From now on, however, it will appear as an additional $\lambda Q$ added to each $N_i$ for $i \geq 3$. Thus

$$N_3 = \sigma(N_2 - 1) + \lambda Q(N_2 - 1) + 1 + \lambda Q$$

$$= (\sigma + \lambda Q)^2 E + \lambda Q + 1 .$$

For $N_i$, we merely repeat the arguments used in finding $N_3$, with the substitution $N_i$ for $N_3$ and $N_{i-1}$ for $N_2$. This gives us, for $i = 3, 4, \ldots, n$,

(17)
$$N_i = \sigma(N_{i-1} - 1) + \lambda Q(N_{i-1} - 1) + \lambda Q + 1$$

$$= (\sigma + \lambda Q)(N_{i-1} - 1) + \lambda Q + 1 .$$

Now, letting $\alpha = \sigma + \lambda Q$, we assert that

(18)
$$N_i = \alpha^{i-1} E + \lambda Q \sum_{j=0}^{i-3} \alpha^j + 1$$

is the solution of Eq. (17) for $i = 3, 4, \ldots, n$. Let us prove this by induction. Clearly, it holds for $i = 3$. Assuming its validity for $N_{i-1}$, we will show its validity for $N_i$ as follows:

$$N_i = \alpha(N_{i-1} - 1) + \lambda Q + 1$$

$$= \alpha \left[ \alpha^{i-2} E + \lambda Q \sum_{j=0}^{i-4} \alpha^j \right] + \lambda Q + 1$$

$$= \alpha^{i-1} E + \lambda Q \sum_{j=0}^{i-4} \alpha^{j+1} + \lambda Q + 1$$

$$= \alpha^{i-1} E + \lambda Q \sum_{j=0}^{i-3} \alpha^j + 1 ,$$

which proves the assertion. Now if we take the usual definition of

$$\sum_{i=a}^{b} x_i = 0 \qquad \text{for} \qquad b < a,$$

we see that Eq. (18) also holds for $i = 1, 2$. Thus, we find that $N_i (i = 1, 2, \ldots, n)$ as given by Eq. (18) is the general form we were looking for. Recognizing that, for $0 \leq b$ and $|x| < 1$,

$$\sum_{k=a}^{b+a} x^k = x^a \frac{1 - x^{b+1}}{1 - x} \tag{19}$$

we can readily evaluate $N_i$. First let us display that $x < 1$ (i.e., that $\alpha < 1$). This is easily done by recalling that we are dealing with systems in equilibrium (steady-state). This implies* that $\rho < 1$. Substituting for $\rho$ we get

$$\rho = \frac{\lambda Q}{1 - \sigma} < 1$$

or

$$\lambda Q + \sigma < 1$$

This shows, of course, that $\alpha < 1$.

We now use Eq. (19) in Eq. (18) and obtain

$$N_i = \begin{cases} E + 1 & i = 1 \\ \alpha^{i-1} E + \lambda Q \dfrac{1 - \alpha^{i-2}}{1 - \alpha} + 1 & i = 2, 3, \ldots, n. \end{cases} \tag{20}$$

Substituting for E, and collecting terms, we get

$$N_i = \begin{cases} \dfrac{1 - \lambda Q}{1 - \rho} & i = 1 \\ \dfrac{1}{1 - \rho} - \dfrac{\rho(1 - \sigma\alpha)}{1 - \rho} \alpha^{i-2} & i = 2, 3, \ldots, n. \end{cases} \tag{21}$$

*As discussed on p. 60, $\rho$ is the product of the average arrival rate and the average service time. Thus, it represents the average number of seconds of service that enter the system each second. Clearly, if $\rho > 1$, the input demand is in excess of the processor's capacity and an infinite queue will form. It may be shown that a similar overload occurs at $\rho = 1$.

We are now in a position to evaluate $T_n$, from Eq. (16) by substituting in Eq. (21). By performing the required operations, and recognizing that $1 - \alpha = (1-\sigma)(1-\rho)$ we are led to

$$T_n = \frac{nQ}{1-\rho} - \frac{\lambda Q^2}{1-\rho}\left[1 + \frac{(1-\sigma\alpha)(1-\alpha^{n-1})}{(1-\sigma)^2(1-\rho)}\right],$$

which completes the proof of Theorem 1.

PROOF OF THEOREM 2: Let us first establish the distribution of $r_k$ as given by Eq. (6). Using methods similar to those of Morse [4], we derive the following equilibrium relationships among the $r_k$:

$$\lambda Q r_0 = (1-\sigma)(1-\lambda Q)\, r_1$$

$$[(1-\sigma)(1-\lambda Q) + \lambda Q\sigma]\, r_1 = \lambda Q r_0 + (1-\sigma)(1-\lambda Q)\, r_2$$

$$[(1-\sigma)(1-\lambda Q) + \lambda Q\sigma]\, r_k = \lambda Q\sigma r_{k-1} + (1-\sigma)(1-\lambda Q)\, r_{k+1} \qquad k \geq 2.$$

As before, let

$$a = \frac{\rho\sigma}{1-\lambda Q}$$

and

$$\rho = \frac{\lambda Q}{1-\sigma}.$$

It is then a simple matter to show that the solution to the above equation is

$$r_k = \begin{cases} 1-\rho & k = 0 \\ \\ \dfrac{1-\rho}{\sigma}\, a^k & k = 1, 2, \ldots, \end{cases}$$

which proves Eq. (6).

The expected value of the above distribution is

$$E' = \sum_{k=0}^{\infty} k\, r_k$$

$$= \frac{1-\rho}{\sigma}\, a \sum_{k=0}^{\infty} k\, a^{k-1}$$

$$E' = \frac{(1-\rho)a}{(1-a)^2\sigma} .$$

But

$$1 - a = 1 - \frac{\rho\sigma}{1 - \lambda Q}$$

$$= \frac{(1-\lambda Q-\sigma) + \sigma(1-\rho)}{1 - \lambda Q}$$

$$= \frac{1-\rho}{1-\lambda Q} .$$

Thus

$$E' = \frac{\rho}{1-\rho}(1-\lambda Q) ,$$

which proves Eq. (7).

Now, for the theorem. The arguments needed here are quite similar to those used in Theorem 1, and therefore will be shortened considerably. In particular, define $T_n$, $D_i$, and $N_i$ as previously, thereby establishing Eq. (16) again. Let us now derive a general form for $N_i$. Upon entering the system, the tagged unit finds $E'$ units in the system. Now, if there is a unit in the service facility (which occurs with probability $1 - r_0 = \rho$) only $E'$ minus the expected value of the number in the service facility will contribute to $N_1$ (since any unit in service must be on the verge of being ejected from service). Well, this expected value is just

$$0 \cdot (r_0) + 1 \cdot (1 - r_0) = \rho ,$$

and so

$$N_1 = E' - \rho + 1 ,$$

where the $+1$ term is due to the tagged unit itself. Following the same reasoning as in Theorem 1, we find that

$$N_2 = \lambda Q(N_1) + \sigma(N_1 - 1) + \sigma\rho$$

where the $\sigma\rho$ term is due to the unit (if there is one) found in service at the time of arrival of the tagged unit. Using the same type of argument, we find for $i > 2$,

$$N_i = \lambda Q(N_{i-1}) + \sigma(N_{i-1} - 1) + 1$$

where, for $i > 2$, we omit the $\sigma\rho$ term since it is fully accounted for in $N_2$. We assert that the solution to this set of equations is

$$
(22) \qquad N_i = \begin{cases} E' + 1 - \rho & i = 1 \\ \alpha^{i-1} E' + \alpha^{i-2} \lambda Q(1-\rho) + \lambda Q \sum_{j=0}^{i-3} \alpha^j + 1 & i > 1 . \end{cases}
$$

That this is indeed the solution is easily shown by induction on $i$ as follows. Clearly, it is true for $i = 1, 2$. Now, assuming its validity for $N_{i-1}$, we will show its validity for $N_i$ as follows:

$$
N_i = \alpha N_{i-1} + 1 - \sigma
$$

$$
= \alpha \left[ \alpha^{i-2} E' + \alpha^{i-3} \lambda Q(1-\rho) + \lambda Q \sum_{j=0}^{i-4} \alpha^j + 1 \right] + 1 - \sigma
$$

$$
= \alpha^{i-1} E' + \alpha^{i-2} \lambda Q(1-\rho) + \lambda Q \sum_{j=1}^{i-3} \alpha^j + \alpha + 1 - \sigma
$$

$$
= \alpha^{i-1} E' + \alpha^{i-2} \lambda Q(1-\rho) + \lambda Q \sum_{j=0}^{i-3} \alpha^j + 1 ,
$$

which proves the assertion. Substituting for $E'$, performing the indicated summation, and collecting terms gives us

$$
(23) \qquad N_i = \begin{cases} \dfrac{\rho}{1-\rho} (1 - \lambda Q) + 1 - \rho & i = 1 \\[2ex] \dfrac{1}{1-\rho} - \dfrac{\rho^2 (1 - \sigma\alpha)}{1-\rho} \alpha^{i-2} & i = 2, 3, \ldots, n . \end{cases}
$$

We are now in a position to evaluate $T_n$, from Eq. (16) by substituting in Eq. (23). Performing the required operations leads us to

$$
T_n = \frac{nQ}{1-\rho} - \rho Q - \frac{\lambda Q^2 \rho}{1-\rho} \left[ 1 + \frac{(1-\sigma\alpha)(1-\alpha^{n-1})}{(1-\sigma)^2 (1-\rho)} \right]
$$

which proves Theorem 2.

PROOF OF THEOREM 3: Let us first consider the late arrival system. Arguing on an expected value basis, we recognize that, upon entry, the tagged unit finds $E = (\rho\sigma/1-\rho)$ units in the system. Each unit in the queue has an expected service time of $Q/(1-\sigma)$. Now, as far as the unit in service is concerned, we appeal to the Markovian property of the geometric distribution (e.g., see Ref. [1]). That is, we assert that the expected additional service time for the unit in service is $Q/(1-\sigma)$ (given that more service is required). Thus, each of the E units in the system (queue plus service) will delay the tagged unit by $Q/(1-\sigma)$ seconds, and this unit will spend nQ seconds in service itself. Hence, for the late arrival system,

$$T_n = Q \frac{\rho\sigma}{(1-\sigma)(1-\rho)} + nQ$$

which proves Theorem 3 for the late arrival system.

For the early arrival system, we recognize that, upon entry, the tagged unit finds $E' = [\rho/(1-\rho)](1-\lambda Q)$ in the system. Now, as before, given that the unit in service requires additional service, the expected value of this additional service is $Q/(1-\sigma)$ seconds. But now, we cannot be sure (as we were in the late arrival system) that the unit in service will require more service; that is, with probability $\sigma$, the unit in service will remain for more service. Also, the expected number in service is merely $\rho$ (that is, the probability of finding one unit in service) and so, the delay suffered by the tagged unit due to the unit in service is $\rho\sigma\, Q/(1-\sigma)$. Each of the units in the queue (the expected number of which is $E' - \rho$) will, on the average, delay the tagged unit by $Q(1-\sigma)$ seconds. In addition, the tagged unit will spend nQ seconds in service. Hence, for the early arrival system,

$$T_n = (E' - \rho)\frac{Q}{1-\sigma} + \rho\sigma \frac{Q}{1-\sigma} + nQ$$

$$= \frac{Q}{1-\sigma}\left[E' - \rho(1-\sigma)\right] + nQ$$

Note that

$$\Delta = \rho(1-\sigma) = \frac{\rho}{1-\rho}(1-\lambda Q) - \frac{\rho\sigma}{1-\rho}$$

and so, we see that

$$T_n = Q \frac{\rho\sigma}{(1-\sigma)(1-\rho)} + nQ$$

which proves Theorem 3 for the early arrival system.

## REFERENCES

[1] Feller, W., An Introduction to Probability Theory and Its Applications (John Wiley and Sons, Inc., New York, 1957), 2d ed., Vol. 1, Chap. XIII, 9.

[2] Jackson, J. R., "Some Problems in Queueing with Dynamic Priorities," Naval Research Logistics Quarterly, 7, 235 (1960).

[3] Kleinrock, L., "A Conservation Law for a Wide Class of Queue Disciplines" (to be published).