Queues Served in Cyclic Order: Waiting Times

By R. B. COOPER

(Manuscript received December 2, 1969)

This paper extends the results of a previous paper in which two models of a system of queues served in cyclic order were studied. One model is an exhaustive service model, in which the server waits on all customers in a queue before proceeding to the next queue in cyclic order. The other is a gating model, in which a gate closes behind the waiting units when the server arrives, and the server waits on only those customers in front of the gate, deferring service of later arrivals until the next cycle.

In the present paper, the Laplace–Stieltjes transforms of the order-of-arrival waiting time distribution functions and, for the exhaustive service model, the mean waiting time for a unit arriving at a queue, are obtained.

I. INTRODUCTION

In a recent paper[1] we studied two models of a system of queues served in cyclic order:

In each model, the ith queue is characterized by general service time distribution function $H_i(\cdot)$ and Poisson input with parameter λ_i . In the exhaustive service model, the server continues to serve a particular queue until for the first time there are no units in service or waiting in that queue; at this time the server advances to and immediately starts service on the next nonempty queue in the cyclic order. The gating model differs from the exhaustive service model in that when the server advances to a nonempty queue, a gate closes behind the waiting units. Only those units waiting in front of the gate are served during this cycle, with the service of subsequent arrivals deferred to the next cycle.

In Ref. 1 we found, for the exhaustive service model, expressions for the mean number of units in a queue at the instant it starts service, the mean cycle time, and the Laplace–Stieltjes transform of the cycle time distribution function.

In the present paper, we extend the analysis to obtain, for each model, the Laplace–Stieltjes transform of the order-of-arrival waiting time distribution function and, for the exhaustive service model, the mean waiting time for a unit arriving at the ith queue.

In Ref. 1 we defined a switch point as a time epoch at which the server finishes serving a queue; and we defined $P_i(n_1, \cdots, n_N)$ as the joint probability that at a switch point the server has just completed a visit at queue i ($i = 0, 1, \cdots, N$) and n_1 units are waiting in queue $i + 1$, n_2 units in queue $i + 2$, $\cdots$, and n_N units in queue $i + N$.

The central results of Ref. 1 were an iterative algorithm for the calculation of the probability generating functions

$$g_i(x_1, \cdots, x_N) = \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} P_i(n_1, \cdots, n_N) x_1^{n_1} \cdots x_N^{n_N} \qquad (i = 0, 1, \cdots, N) \qquad (1)$$

and, for the exhaustive service model, an expression for the mean number $\bar{n}_i$ of units waiting in queue $i + 1$ when the server completes a visit at queue i. In particular, it was shown for the exhaustive service model that these generating functions satisfy the functional equations

$$g_i(x_1, \cdots, x_N) = g_{i-1}\left(\beta_i\left(\sum_{m=1}^{N} \lambda_{i+m}(1 - x_m)\right), x_1, \cdots, x_{N-1}\right) + \frac{\lambda_i}{\lambda} \beta_i\left(\sum_{m=1}^{N} \lambda_{i+m}(1 - x_m)\right) P(\mathbf{0}) - P_{i-1}(0, \cdots, 0) \qquad (i = 0, 1, \cdots, N) \qquad (2)$$

and that $\bar{n}_i$ is given by $\bar{n}_i = \bar{m}_i / g_i(1, \cdots, 1)$, where

$$\bar{m}_i = \frac{\lambda_{i+1}}{\lambda} P(\mathbf{0}) \frac{\rho - \rho_{i+1}}{1 - \rho} \qquad (i = 0, 1, \cdots, N) \qquad (3)$$

and where λ_i is the rate of arrivals of units at queue i, ρ_i is the traffic intensity at queue i,

$$\lambda = \sum_{i=0}^{N} \lambda_i,$$

$$\rho = \sum_{i=0}^{N} \rho_i,$$

$$P(\mathbf{0}) = \sum_{i=0}^{N} P_i(0, \cdots, 0),$$

and $\beta_i(\cdot)$ is the Laplace–Stieltjes transform of the distribution function of the length of the busy period at queue i. Equations (1), (2), and (3) appear in Ref. 1 as equations (3), (5), and (34), respectively.

The distribution generated by $g_i(x_1, \cdots, x_N)$ is defined with respect to a Markov chain imbedded at the switch points. An analysis by Takács[2] of the exhaustive service model for the special case of two queues is based on a Markov chain imbedded at the set of service completion points. Clearly, the set of switch points is a proper subset of the set of service completion points. Our use of switch points instead of service completion points enabled us in Ref. 1 to analyze the multiqueue model with about the same degree of mathematical complexity as Takács required for the analysis of the 2-queue model. On the other hand, one would expect that our use of a chain imbedded in a "smaller" set of points would result in a corresponding loss of useful information.

Takács' analysis yielded waiting time results for the 2-queue model. At the time Ref. 1 was written, it was not apparent to us that our method of analysis provided enough information to enable us to obtain corresponding waiting time results for the multiqueue model. Accordingly, we concentrated on the cycle time, a quantity that seemingly gives the same kind of information as the waiting time. Unfortunately, we did not have complete freedom in choosing a precise definition of the quantity we would call cycle time. The mathematical formulation of the model dictated that the cycle time for queue i be defined, roughly speaking, as the length of time between two successive instants at which the server completes service at queue i, without regard to whether or not the server is continuously busy throughout this time interval. This definition introduces, among others, the following difficulty in the interpretation of realized values of the cycle time: A long cycle time could have resulted either from heavy congestion or from no congestion.

No such ambiguities exist with respect to interpretation of the waiting time, which is simply the elapsed time from the arrival instant of a unit to the instant at which service on this unit begins. Therefore, we would like to obtain waiting time results. Furthermore, we would like to obtain these results, if possible, without directly extending the previous analysis to include the entire set of service completion points.

In the present paper, we obtain the desired waiting time results without recourse to a complicated reformulation of the original analysis based on the complete set of service completion points. Rather, to obtain the waiting times at queue i, we use the generating function $g_{i-1}(x_1, \cdots, x_N)$, calculated in Ref. 1, to append to the original set of switch points only those service completion points that correspond to departures from queue i; and this is sufficient for our waiting time calculations.

The essence of the method is to calculate the probability generating function of the number of units left in queue i by an arbitrary departure from queue i, using only the (known) probability generating function of the number of units waiting in queue i when the server arrives. The Laplace–Stieltjes transform of the order-of-arrival waiting time distribution function for units at queue i is then easily obtained by a standard argument.

The preceding discussion refers mainly to the exhaustive service model, which was discussed in detail in Ref. 1. The gating model was shown to be characterized by equations that are essentially the same as those of the exhaustive service model, and was therefore not developed in detail. In the present paper, the waiting times for the gating model will also be discussed.

II. THE M/G/1 QUEUE WITH SERVER VACATION TIMES

In preparation for calculation of the waiting times in the exhaustive service model, we first consider the following generalization of the $M/G/1$ queue:

As usual, the server serves the queue continuously as long as there is at least one unit in the system (waiting or in service). When the server finishes serving a unit and finds the system empty, however, it goes away for a length of time called a vacation. At the end of the vacation the server returns to the queue, and begins to serve those units, if any, that have arrived during the vacation. If the server finds the system empty at the end of a vacation, it immediately takes another vacation, and continues in this manner until it finds at least one waiting unit upon return from a vacation.

Let X_k $(k = 1, 2, \cdots)$ be the number of units left behind by the kth departing unit. Then

$$P\{X_{k+1} = n\} = \sum_{\nu=0}^{n+1} P\{X_k = \nu\}P\{X_{k+1} = n \mid X_k = \nu\}$$
$$(k = 1, 2, \cdots ; n = 0, 1, \cdots). \qquad (4)$$

Let $P(j)$ be the probability that at the end of a vacation the server finds $j \geqq 0$ units waiting for service. If the arrival rate and the service time distribution function are denoted by λ and $H(\cdot)$, respectively, then

$$P\{X_{k+1} = n \mid X_k = \nu > 0\} = \int_0^\infty \frac{(\lambda\xi)^{n+1-\nu}}{(n+1-\nu)!} \exp(-\lambda\xi)\, dH(\xi)$$
$$(n \geqq \nu - 1) \qquad (5)$$

and

$$P\{X_{k+1} = n \mid X_k = 0\} = \sum_{j=1}^{n+1} \frac{P(j)}{1 - P(0)} \int_0^\infty \frac{(\lambda\xi)^{n+1-j}}{(n+1-j)!} \exp(-\lambda\xi)\, dH(\xi). \tag{6}$$

The expression $P(j)/[1 - P(0)]$ in equation (6) is the conditional probability that when the server starts serving the queue there are j units waiting, given that at least one unit is waiting.

When the traffic intensity ρ is less than unity ($\rho = \lambda h$, where h is the mean service time), there exists a unique distribution

$$\pi_n = \lim_{k\to\infty} P\{X_k = n\} \qquad (n = 0, 1, \cdots) \tag{7}$$

that satisfies both the normalization equation

$$\sum_{n=0}^{\infty} \pi_n = 1 \tag{8}$$

and the limiting set of equations obtained from equation (4),

$$\pi_n = \pi_0 \sum_{j=1}^{n+1} \frac{P(j)}{1 - P(0)} \int_0^\infty \frac{(\lambda\xi)^{n+1-j}}{(n+1-j)!} \exp(-\lambda\xi)\, dH(\xi)$$
$$+ \sum_{\nu=1}^{n+1} \pi_\nu \int_0^\infty \frac{(\lambda\xi)^{n+1-\nu}}{(n+1-\nu)!} \exp(-\lambda\xi)\, dH(\xi) \qquad (n = 0, 1, \cdots). \tag{9}$$

Define the probability generating functions

$$f(x) = \sum_{n=0}^{\infty} \pi_n x^n \tag{10}$$

and

$$\psi(x) = \sum_{j=0}^{\infty} P(j) x^j. \tag{11}$$

Substitution of equation (9) into equation (10) yields, after some manipulation,

$$f(x) = \frac{\left\{\dfrac{\psi(x) - P(0)}{1 - P(0)} - 1\right\}\eta(\lambda - \lambda x)}{x - \eta(\lambda - \lambda x)}\, \pi_0 \tag{12}$$

where $\eta(\cdot)$ is the Laplace–Stieltjes transform of the service time distribution function $H(\cdot)$. Observe that the expression $[\psi(x) - P(0)]/[1 - P(0)]$ is the probability generating function of the number of units waiting when service commences.

The unknown probability π_0 is determined from equation (12) by the normalization condition $f(1) = 1$. Application of l'Hospital's rule to equation (12) yields

$$\pi_0 = \frac{1 - P(0)}{\psi'(1)} (1 - \rho). \tag{13}$$

Thus, the probability generating function $f(\cdot)$ of the number of units left behind by an arbitrary departure, and the probability generating function $\psi(\cdot)$ of the number of units waiting at the end of a vacation are related as follows:

$$f(x) = \frac{[\psi(x) - 1]\eta(\lambda - \lambda x)}{x - \eta(\lambda - \lambda x)} \frac{(1 - \rho)}{\psi'(1)}. \tag{14}$$

We now apply a standard argument to obtain the Laplace–Stieltjes transform $\omega(\cdot)$ of the order-of-arrival waiting time distribution function from the generating function (14).

Let $F(\cdot)$ be the distribution function of an arbitrary unit's sojourn time, defined as the elapsed time between the unit's arrival and departure epochs, and denote by $\phi(\cdot)$ the Laplace–Stieltjes transform of $F(\cdot)$. Since the sojourn time is the sum of the waiting time and the service time, and since these latter times are independent, therefore

$$\phi(s) = \omega(s)\eta(s). \tag{15}$$

When units are served in their arrival order, each departing unit leaves behind it precisely those units that arrived during the departure's sojourn time. Further, these remaining units arrived according to a Poisson process, independent of the sojourn time. Therefore, the probability π_n that a departing unit leaves behind n other units is

$$\pi_n = \int_0^\infty \frac{(\lambda\xi)^n}{n!} \exp(-\lambda\xi)\, dF(\xi). \tag{16}$$

Substitution of equation (16) into (10) gives the well known and fundamental relation

$$f(x) = \phi(\lambda - \lambda x). \tag{17}$$

Equations (14), (15), and (17) together give the Laplace–Stieltjes transform $\omega(\cdot)$ of the waiting time distribution function in terms of the probability generating function $\psi(\cdot)$ of the number of units waiting at the end of a vacation:

$$\omega(s) = \frac{\lambda}{\psi'(1)} \left[1 - \psi\left(\frac{\lambda - s}{\lambda}\right)\right] \frac{1 - \rho}{s - \lambda + \lambda\eta(s)}. \tag{18}$$

Note that for the ordinary $M/G/1$ queue, in which the vacation ends immediately whenever a unit arrives and finds the server idle, $\psi(x) = x$ and equation (18) reduces to the well known Pollaczek–Khinchin formula,

$$\omega(s) = \frac{s(1-\rho)}{s - \lambda + \lambda\eta(s)}. \tag{19}$$

Finally, the mean wait for service $\bar{W} = -\omega'(0)$, obtained from equation (18), is given by

$$\bar{W} = \frac{\lambda\eta''(0)}{2(1-\rho)} + \frac{\psi''(1)}{2\lambda\psi'(1)}. \tag{20}$$

The first term on the right side of equation (20) is identical with the mean waiting time in the ordinary $M/G/1$ queue, as would be obtained directly from the Pollaczek–Khinchin formula (19). The second term in equation (20) represents the component of the mean wait that arises because of the variability in the number of units waiting when service begins. Although service in arrival order was assumed in its derivation, equation (20) is valid for any order of service that is independent of the service times.

Note that the result (14) is true regardless of any relationship between the vacation lengths and the arrival process, whereas equations (18) and (20) are valid only when the vacation lengths are determined without regard to the arrival process. For example, if one were to consider a mechanism such that service begins as soon as a fixed number j ($j \geqq 2$) units are waiting, then equation (14) with $\psi(x) = x^j$ would correctly give the probability generating function of the number of units in the system just after a service completion epoch. On the other hand, equations (18) and (20) would not apply, because (16), and therefore (17), would no longer be true. (For if $\psi(x) = x^j$, then the first departing unit would always leave behind at least those $j - 1$ other units that were present when service commenced. Thus, for the first departing unit, $\pi_n = 0$ for $n < j - 1$, and this contradicts the assumption (16) if $j \geqq 2$.)

III. LAPLACE–STIELTJES TRANSFORM OF WAITING TIME DISTRIBUTION FUNCTION FOR EXHAUSTIVE SERVICE MODEL

We now proceed to apply the results of Section II to the analysis of waiting times in the exhaustive service model. In essence, the "vacation time" of Section II is the length of time that the server spends idle or working on other queues before registering a switch point at queue $i - 1$ and beginning service on queue i.

For want of a better word, let us define as a supercycle the elapsed time between the arrival epoch of a unit at any queue when the system is completely empty, and the first instant at which the whole system again becomes empty. Then units that arrive at queue i can be classified into two exclusive and exhaustive categories:

(1) arrivals at queue i that either initiate a supercycle or occur during the 1-busy period generated by an arrival at queue i that initiated a supercycle; or
(2) all other arrivals at queue i.

Equivalently, units in category (1) are those arrivals at queue i whose service begins prior to the occurrence of the first switch point of a supercycle, whereas units in category (2) are arrivals at queue i whose service times begin after the first switch point of a supercycle.

Consider now the waiting times of units that arrive at queue i. Those units in category (1) are served during a busy period originated by one unit. Therefore, the Laplace–Stieltjes transform $\omega_i^{(1)}(\cdot)$ of the order-of-arrival waiting time distribution function for units at queue i that belong to category (1) is given by the Pollaczek–Khinchin formula (19):

$$\omega_i^{(1)}(s) = \frac{s(1 - \rho_i)}{s - \lambda_i + \lambda_i \eta_i(s)} \tag{21}$$

where ρ_i, λ_i, and $\eta_i(\cdot)$ are the corresponding quantities in equation (19) defined now with respect to queue i.

Units in category (2) are served during a busy period originated by those units waiting in queue i when the server leaves queue $i - 1$. Let $\psi_i(\cdot)$ be the probability generating function of the number of units waiting in queue i when the server leaves queue $i - 1$; $\psi_i(\cdot)$ is the probability generating function of the number of units waiting for service in queue i when the server finishes a vacation, and is given by

$$\psi_i(x) = \frac{g_{i-1}(x, 1, \cdots, 1)}{g_{i-1}(1, 1, \cdots, 1)}. \tag{22}$$

[Note that $\psi_i'(1) = \bar{n}_{i-1} = \bar{m}_{i-1}/g_{i-1}(1, 1, \cdots, 1)$.] Thus, the Laplace–Stieltjes transform $\omega_i^{(2)}(\cdot)$ of the waiting time distribution function for units in category (2) is given by equation (18):

$$\omega_i^{(2)}(s) = \frac{\lambda_i}{\psi_i'(1)} \left[1 - \psi_i\left(\frac{\lambda_i - s}{\lambda_i}\right)\right] \frac{1 - \rho_i}{s - \lambda_i + \lambda_i \eta_i(s)}. \tag{23}$$

Let $p_i^{(1)}$ be the proportion of all arrivals at queue i that are in category (1). The mean number of units that arrive at queue i during an interval of length t is $\lambda_i t$. The probability that an arbitrary arrival at queue i finds the whole system empty is $1 - \rho$ ($\rho = \rho_0 + \cdots + \rho_N$), so that $\lambda_i t(1 - \rho)$ is the mean number of arrivals at queue i that initiate a supercycle during any elapsed time t. The mean number of units served at queue i during the 1-busy period generated by each such arrival is $(1 - \rho_i)^{-1}$, and hence the mean number of units in category (1) served at queue i during an elapsed time t is $\lambda_i t(1 - \rho)/(1 - \rho_i)$. Therefore, the probability is $[\lambda_i t(1 - \rho)/(1 - \rho_i)]/\lambda_i t$ that an arbitrary arrival at queue i is in category (1); that is,

$$p_i^{(1)} = \frac{1 - \rho}{1 - \rho_i}, \tag{24}$$

and the probability $p_i^{(2)} = 1 - p_i^{(1)}$ that an arbitrary arrival at queue i is in category (2) is

$$p_i^{(2)} = \frac{\rho - \rho_i}{1 - \rho_i}. \tag{25}$$

The Laplace–Stieltjes transform $\omega_i(\cdot)$ of the waiting time distribution function for an arbitrary unit at queue i is the weighted sum of the transforms for each category:

$$\omega_i(s) = p_i^{(1)}\omega_i^{(1)}(s) + p_i^{(2)}\omega_i^{(2)}(s). \tag{26}$$

Finally, equation (26) becomes, with the help of equations (21) through (25) and equation (3),

$$\omega_i(s) = \frac{1 - \rho}{s - \lambda_i + \lambda_i \eta_i(s)} \cdot \left\{ \frac{\lambda}{P(\mathbf{0})} \left[g_{i-1}(1, 1, \cdots, 1) - g_{i-1}\left(\frac{\lambda_i - s}{\lambda_i}, 1, \cdots, 1 \right) \right] + s \right\} \qquad (i = 0, 1, \cdots, N). \tag{27}$$

Inherent in equation (27) is the assumption that units in queue i are served in their arrival order, but no assumption is made regarding the order of service of units in other queues. If at each queue units are served in their arrival order, then the waiting time distribution function for an arbitrary unit, without regard to the identity of the queue in which it is served, has Laplace–Stieltjes transform $\omega(\cdot)$ given by

$$\omega(s) = \lambda^{-1} \sum_{i=0}^{N} \lambda_i \omega_i(s). \tag{28}$$

IV. MEAN WAITING TIMES FOR EXHAUSTIVE SERVICE MODEL

Denote by $\bar{W}_i$ the mean wait for service suffered by units arriving at queue i. The mean wait for service for units in category (1) is $[\lambda_i \eta_i''(0)/2(1 - \rho_i)]$; the mean wait for service for units in category (2) is, in analogy with equation (20), $[\lambda_i \eta_i''(0)/2(1 - \rho_i)] + [\psi_i''(1)/2\lambda_i \psi_i'(1)]$. Weighting these values according to equations (24) and (25), respectively, we have

$$\bar{W}_i = \frac{\lambda_i \eta_i''(0)}{2(1 - \rho_i)} + \frac{\psi_i''(1)}{\psi_i'(1)} \frac{\rho - \rho_i}{2\lambda_i(1 - \rho_i)} \qquad (i = 0, 1, \cdots, N). \tag{29}$$

In equation (26) of Ref. 1 we defined

$$\bar{m}_i(k) = \frac{\partial}{\partial x_k} g_i(x_1, \cdots, x_N) \Big|_{x_1 = \cdots = x_N = 1} \qquad (i = 0, 1, \cdots, N; \quad k = 1, \cdots, N) \tag{30}$$

and $\bar{m}_i = \bar{m}_i(1)$. Let us also define

$$\bar{m}_i(j, k) = \frac{\lambda(1 - \rho)}{P(\mathbf{0})} \frac{\partial^2}{\partial x_j \, \partial x_k} g_i(x_1, \cdots, x_N) \Big|_{x_1 = \cdots = x_N = 1}$$
$$(i = 0, 1, \cdots, N; \quad j = 1, \cdots, N; \quad k = 1, \cdots, N). \tag{31}$$

Then it follows from equation (22) and these definitions that

$$\frac{\psi_i''(1)}{\psi_i'(1)} = \frac{P(\mathbf{0})}{\lambda(1 - \rho)} \frac{\bar{m}_{i-1}(1, 1)}{\bar{m}_{i-1}(1)}. \tag{32}$$

Using equations (3) and (32), we can rewrite (29):

$$\bar{W}_i = \frac{\lambda_i \eta_i''(0)}{2(1 - \rho_i)} + \frac{\bar{m}_{i-1}(1, 1)}{2\lambda_i^2(1 - \rho_i)} \qquad (i = 0, 1, \cdots, N). \tag{33}$$

It remains to calculate the quantity $\bar{m}_{i-1}$ (1, 1) in (33). To this end, we define

$$\bar{\beta}_i(k) = \frac{\partial}{\partial x_k} \beta_i\left(\sum_{m=1}^{N} \lambda_{i+m}(1 - x_m)\right) \Big|_{x_1 = \cdots = x_N = 1} \qquad (i = 0, 1, \cdots, N; \quad k = 1, \cdots, N) \tag{34}$$

and

$$\bar{\beta}_i(j, k) = \frac{\partial^2}{\partial x_j \, \partial x_k} \beta_i\left(\sum_{m=1}^{N} \lambda_{i+m}(1 - x_m)\right) \Big|_{x_1 = \cdots = x_N = 1}$$
$$(i = 0, 1, \cdots, N; \quad j = 1, \cdots, N; \quad k = 1, \cdots, N). \tag{35}$$

Note that in terms of the given parameters,

$$\bar{\beta}_i(k) = \lambda_{i+k} \frac{h_i}{1 - \rho_i} \tag{36}$$

and

$$\bar{\beta}_i(j, k) = \lambda_{i+j}\lambda_{i+k} \frac{\eta_i''(0)}{(1 - \rho_i)^3} \tag{37}$$

where h_i is the mean and $\eta_i(\cdot)$ the Laplace–Stieltjes transform of the service time distribution function for a unit at queue i.

We proceed to calculate $\bar{m}_i(1, 1)$ in the same way we calculated $\bar{m}_i(1)$ in Section VII of Ref. 1. Differentiating twice through equation (2) and setting $x_1 = \cdots = x_N = 1$, we obtain the three-dimensional set of linear equations

$$\begin{aligned}\bar{m}_i(j, k) = {} & \frac{\lambda(1 - \rho)}{P(0)} \bar{m}_{i-1}(1)\bar{\beta}_i(j, k) + (1 - \rho)\lambda_i\bar{\beta}_i(j, k) \\ & + \bar{\beta}_i(j)\bar{\beta}_i(k)\bar{m}_{i-1}(1, 1) + (1 - \delta(N - j))\bar{\beta}_i(k)\bar{m}_{i-1}(1, j + 1) \\ & + (1 - \delta(N - k))\bar{\beta}_i(j)\bar{m}_{i-1}(1, k + 1) \\ & + (1 - \delta(N - j))(1 - \delta(N - k))\bar{m}_{i-1}(j + 1, k + 1) \\ & (i = 0, 1, \cdots, N; \quad j = 1, \cdots, N; \quad k = 1, \cdots, N)\end{aligned} \tag{38}$$

where $\delta(x) = 1$ if $x = 0$ and $\delta(x) = 0$ if $x \neq 0$. Using equation (3) and combining the first two terms on the right side of equation (38), we can write

$$\begin{aligned}\bar{m}_i(j, k) = {} & \lambda_i(1 - \rho_i)\bar{\beta}_i(j, k) + \bar{\beta}_i(j)\bar{\beta}_i(k)\bar{m}_{i-1}(1, 1) \\ & + \bar{\beta}_i(k)\bar{m}_{i-1}(1, j + 1) + \bar{\beta}_i(j)\bar{m}_{i-1}(1, k + 1) + \bar{m}_{i-1}(j + 1, k + 1) \\ & (i = 0, 1, \cdots, N; \quad j = 1, \cdots, N; \quad k = 1, \cdots, N)\end{aligned} \tag{39}$$

where all undefined terms are taken to be zero. (The functions $\bar{m}_i(j, k)$ are defined only for $j, k = 1, \cdots, N$.) It is required to solve this set of $\frac{1}{2}N(N + 1)^2$ independent linear equations for the $\bar{m}_i(1, 1)$. [Note that $\bar{m}_i(j, k) = \bar{m}_i(k, j)$.]

Successive substitution into the last term on the right side of equation (39) gives

$$\begin{aligned}\bar{m}_i(1, k) = {} & \sum_{\nu=0} \lambda_{i-\nu}(1 - \rho_{i-\nu})\bar{\beta}_{i-\nu}(1 + \nu, k + \nu) \\ & + \sum_{\nu=0} \bar{\beta}_{i-\nu}(1 + \nu)\bar{\beta}_{i-\nu}(k + \nu)\bar{m}_{i-1-\nu}(1, 1)\end{aligned}$$

$$+ \sum_{\nu=0} \bar{\beta}_{i-\nu}(k + \nu)\bar{m}_{i-1-\nu}(1, 2 + \nu)$$

$$+ \sum_{\nu=0} \bar{\beta}_{i-\nu}(1 + \nu)\bar{m}_{i-1-\nu}(1, k + 1 + \nu)$$

$$(i = 0, 1, \cdots, N; \quad k = 1, \cdots, N) \qquad (40)$$

where each sum is continued as long as the terms are defined.

The set (40) consists of $N(N + 1)$ independent linear equations. Unfortunately, it does not appear that further algebraic simplification is likely. However, for particular values of the parameters and reasonable values of N, numerical solution should not be difficult.

Therefore, to calculate the mean wait $\bar{W}_i$ for service at queue i ($i = 0, 1, \cdots, N$) for any particular values of the basic parameters, simply solve the set (40) of $N(N + 1)$ linear equations numerically, and use the resulting value of $\bar{m}_{i-1}$ (1, 1) in equation (33). Note that these calculations for the mean waiting times require no iteration, since neither generating functions nor state probabilities appear. This last observation is remarkable in light of the complicated iteration process (discussed in Ref. 1) underlying the derivation of these results. Thus, despite the complicated derivation, calculations do not seem impractical.

In the particular case of two queues ($N = 1$), only $N(N + 1) = 2$ simultaneous equations must be solved to find $\bar{m}_i(1, 1)$, and an algebraic solution is easily obtained. For $N = 1$ equation (40) gives

$$\bar{m}_{i-1}(1, 1) = \lambda_i^2 \frac{\lambda_{i-1}\eta_{i-1}''(0)(1 - \rho_i)^2 + \lambda_i\eta_i''(0)\rho_{i-1}^2}{(1 - \rho_{i-1})^2(1 - \rho_i)^2 - \rho_{i-1}^2\rho_i^2}$$

and hence, for two queues,

$$\bar{W}_i = \frac{\lambda_i\eta_i''(0)}{2(1 - \rho_i)} + \frac{\lambda_{i-1}\eta_{i-1}''(0)(1 - \rho_i)^2 + \lambda_i\eta_i''(0)\rho_{i-1}^2}{2(1 - \rho_i)[(1 - \rho_{i-1})^2(1 - \rho_i)^2 - \rho_{i-1}^2\rho_i^2]}$$

$$(i = 0, 1). \qquad (41)$$

Our result (41) is in agreement with previous results of Takács,[2] Avi-Itzhak, Maxwell and Miller,[3] and Eisenberg.[4]

Although service in order of arrival was assumed throughout, the results for the mean waiting time are valid for any order of service that is independent of the service times.

V. LAPLACE–STIELTJES TRANSFORM OF WAITING TIME DISTRIBUTION FUNCTION FOR GATING MODEL

Turning to the gating model, we now study briefly the distribution of waiting times for units served in order of arrival at the ith queue.

As with the exhaustive service model, we first calculate the probability generating function of the number of units left in queue i by an arbitrary unit departing from queue i; and using the same arguments, we obtain from this generating function the Laplace–Stieltjes transform of the waiting time distribution function. As in Ref. 1, the notation for the gating model is the same as for the exhaustive service model; $g_{i-1}(x_1, \cdots, x_N)$ and related probabilities are defined and calculated as described in Section IX of Ref. 1.

Let $\pi_i(j)$ be the conditional probability that an arbitrary departure from queue i leaves behind it j units in queue i, given that this departure did not arrive when the system was completely empty. Then

$$\pi_i(j) = \sum_{n=1}^{\infty} \frac{P_{i-1}(n)}{1 - P_{i-1}(0)} \frac{1}{n} \sum_{k=1}^{n} \int_0^{\infty} \frac{(\lambda_i \xi)^{j-n+k}}{(j-n+k)!} \exp(-\lambda_i \xi) \, dH_i^{*k}(\xi) \qquad (j = 0, 1, \cdots) \tag{42}$$

where $P_{i-1}(n)/[1 - P_{i-1}(0)]$ is the conditional probability that $n \geqq 1$ units are waiting in queue i when the gate closes, given that at least one unit is waiting; and $1/n$ is the probability that a departing unit is kth in line for service ($k = 1, 2, \cdots, n$) given that n units are present at the closing of the gate. The integrand in equation (42) is taken to be zero when $n - k > j$.

Following the argument for the exhaustive service model we see that the (conditional) probability generating function of the number of arrivals at queue i that occur during the waiting time of a departing unit (given that the departing unit did not find the system empty on arrival) is

$$\sum_{j=0}^{\infty} \pi_i(j) x^j / \eta_i(\lambda_i - \lambda_i x).$$

A simple calculation gives

$$\frac{\sum_{j=0}^{\infty} \pi_i(j) x^j}{\eta_i(\lambda_i - \lambda_i x)} = \sum_{n=1}^{\infty} \frac{P_{i-1}(n)}{1 - P_{i-1}(0)} \frac{1}{n} \sum_{k=1}^{n} x^{n-k} \eta_i^{k-1}(\lambda_i - \lambda_i x). \tag{43}$$

The probability that an arbitrary departing unit did not find the system completely empty on arrival is $\rho = \sum_{i=0}^{N-1} \lambda_i h_i$. Thus, after summing the geometric series in equation (43), the unconditional order-of-arrival waiting time distribution function for units served at queue i has Laplace–Stieltjes transform $\omega_i(\cdot)$ given by

$$\omega_i(s) = (1 - \rho) + \rho \sum_{n=1}^{\infty} \frac{P_{i-1}(n)}{1 - P_{i-1}(0)} \frac{1}{n\lambda_i^{n-1}} \frac{[\lambda_i \eta_i(s)]^n - (\lambda_i - s)^n}{s - \lambda_i + \lambda_i \eta_i(s)}$$
$$(i = 0, 1, \cdots, N - 1). \qquad (44)$$

Equation (44) allows numerical calculation (and hence numerical inversion) of the transform $\omega_i(\cdot)$. Unfortunately, this procedure requires knowledge of the distribution $\{P_{i-1}(n)\}$, which is specified only through its generating function g_{i-1} $(x, 1, \cdots, 1)$. Thus, to obtain numerical results for the gating model one must solve two distinct problems in numerical analysis, numerical calculation of the $\{P_{i-1}(n)\}$ and then numerical inversion of the transform. Note that the first of these numerical calculations is not required for the exhaustive service model. The subject of numerical inversion of Laplace–Stieltjes transforms and probability generating functions (the latter being, in fact, a special case of the former) is important for the reduction to practice of these cyclic queuing models. However, it is a subject best treated separately, without regard to the particular applications at hand, and will not be discussed further here.

VI. SUMMARY AND PROPOSALS FOR FUTURE WORK

We have extended our previous study of cyclic queues to obtain waiting time results. In particular we have obtained, for both the exhaustive service model and the gating model, the Laplace–Stieltjes transform of the waiting time distribution function for units arriving at the ith queue, when units at that queue are served in order of arrival. These transforms are given by equation (27) for the exhaustive service model and equation (44) for the gating model. Also, we have obtained for the exhaustive service model a formula (33) for the mean waiting time for units arriving at the ith queue. Use of equation (33) requires calculation of the value $\bar{m}_{i-1}$ (1, 1), which can be obtained in any particular case by numerical solution of the $N(N + 1)$ linear equations (40). It is noteworthy that the calculation of the mean waiting time requires no iteration.

The techniques used in this and our previous study might be useful in the analyses of priority queuing models and other cyclic queuing models that have important practical applications. Examples of the latter are: extensions of the present models to include arbitrary switching times and/or set up times; systems of queues served in arbitrary periodic order (of which cyclic order is a special case); and within-queue disciplines other than service in random order.

REFERENCES

1. Cooper, R. B., and Murray, G., "Queues Served in Cyclic Order," BSTJ, *48*, No. 3 (March 1969), pp. 675–689.
2. Takács, L., "Two Queues Attended by a Single Server," Operations Research, *16*, No. 3 (May-June 1968), pp. 639–650.
3. Avi-Itzhak, B., Maxwell, W. L., and Miller, L. W., "Queuing with Alternating Priorities," Operations Research, *13*, No. 2 (March-April 1965), pp. 306–318.
4. Eisenberg, M., "Multiqueues with Changeover Times," MIT Doctoral Dissertation, September 1967.