IMBEDDED MARKOV CHAIN ANALYSIS OF A WAITING-LINE PROCESS IN CONTINUOUS TIME

DONALD P. GAVER, JR.

Westinghouse Research Laboratories

1. Summary. Bunches of individual customers approach a single servicing facility according to a stationary compound Poisson process. The resulting waiting line process is studied in continuous time by the method of the imbedded Markov chain, cf. Kendall [7], [8], and of renewal theory, cf. Blackwell [3], Feller [5], and Smith [9]. Busy period phenomena are discussed, cf. Theorem 1, in which the transform of the joint d.f. of busy period duration and the number of departures in that duration is expressed as the root $x_1(s, z)$ of a functional equation, a generalization of a result of Takács [12]. In terms of $x_1(s, z)$ "zero-avoiding" transition probabilities are characterized. A simple model for "instantaneous defection" is analyzed. Using renewal theory, ergodic properties of waiting line lengths and waiting times are discussed for the "general" process, in which idle and busy periods recur.

2. Mathematical formulation. In this section arrivals to, and departures from, the system (single servicing facility plus waiting line) are characterized, and basic random variables and probabilities associated with the resulting waiting-line process are described.

For clarity the following terminology will be used: "time" will refer to an "instant", specified by a real parameter $t (0 \leqq t < \infty)$, and is measured from some initial instant taken as origin; the time axis will also be the range space of certain random variables; "period" will mean a time interval, such as (t_1, t_2). Deviations from common terminology, as in the case of "service times", will be pointed out when they occur.

Arrivals. Arrivals at the system occur in accordance with a stationary compound Poisson process. Such a process can be described in terms of the following random variables:

(a) $\{\Delta_k\}$ is a sequence of positive independent random variables, where

$$\Pr[\Delta_k \leqq x] = 1 - e^{-\lambda x} \ (\lambda > 0); \tag{2.1}$$

Δ can be interpreted as the period that elapses between the times of arrival of two successive *bunches* of customers.

(b) $\{\alpha_k\}$ is a sequence of random variables such that

$$\alpha_k = \sum_{i=1}^{k} \Delta_i ; \tag{2.2}$$

α_k can be interpreted as the random time of arrival of the kth bunch of customers.

Received March 3, 1958; revised March 26, 1959.

(c) $\{B_k\}$ is a sequence of positive random variables, mutually independent and independent of $\{\Delta_k\}$, where

$$\Pr[B_k = j] = b_j \qquad (j = 1, 2, 3, \cdots); \tag{2.3}$$

B_k represents the number of customers in the bunch that arrives at time α_k.

For a fixed value of the real parameter $t(0 \leqq t < \infty)$ let

$$A(0, t) = \sum_{0<\alpha_i<t} B_i . \tag{2.4}$$

$A(0, t)$ is the total number of customers that arrive in the time interval $(0, t)$. We have from (2.1), (2.2), and (2.3),

$$\Pr[A(0,t) = k] = a_k(t) = \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n}{n!} b_k^{n*} \tag{2.5}$$

where * denotes the convolution operation. The generating function (G.F.) of $\{a_k(t)\}$ is

$$a(z, t) = \sum_{k=0}^{\infty} z^k a_k(t) = \exp[-\lambda t\{1 - b(x)\}]. \tag{2.6}$$

Thus the distribution $\{a_k(t)\}$ is infinitely divisible, cf. Doob [4]. Note that the numbers of arrivals in any non-overlapping time intervals are independent random variables, each with distribution (2.5), t being the length of an interval.

It will be convenient for later developments to let $\{\beta_n\}$ be a sequence such that

$$\begin{aligned} \beta_n &= 0 && (n = 1, 2, \cdots, i) \\ \beta_n &= \alpha_1 && (n = i + 1, i + 2, \cdots, i + B_1) \\ \beta_n &= \alpha_{k+1} && (n = i + B_k + 1, i + B_k + 2, \cdots, i + B_k + B_{k+1}). \end{aligned} \tag{2.7}$$

Thus β_n is the time of arrival of the (bunch containing the) nth customer to arrive; $\beta_n = 0$ is the time of "arrival" of the $i \geqq 0$ customers present at the initial instant. We shall assume that the customers in a bunch are assigned numbers in the ranges given by (2.7), and that they receive service in the order specified by those numbers.

Departures. Single customers depart from the system, with service completed, at random times. The departure process can be described in terms of

(d) $\{S_n\}$ $(n = 1, 2, 3, \cdots)$, a sequence of mutually independent random variables with

$$\Pr[S_n \leqq x] = U(x), \tag{2.8}$$

where $U(x)$ is a d.f. with $U(0) = 0$.

Put

$$u(s) = \int_0^{\infty} e^{-sx}\, dU(x) \tag{2.9}$$

for the Laplace-Stieltjes transform (L.S.T.) of $U(x)$.

The random variables of the sequence $\{S_n\}$ are independent of those of $\{\Delta_k\}$ and $\{B_k\}$. Associate S_n with β_n, i.e. S_n is to be interpreted as the *service-time* (service period) of the nth customer to arrive, and consequently of the nth customer to be served.

Definition (*Departure Times*): Let the sequence $\{T_n\}(n = 1, 2, 3, \cdots)$ where

$$T_0 = t_0 \geqq 0$$

and

$$T_n = S_n + \max(\beta_n, T_{n-1}) \tag{2.10}$$

be called the sequence of *departure times* from the system. T_n is the time of departure of the nth customer to receive service after time t_0.

If, at the initial instant, $i = 0$ in (2.7), then $T_1 = S_1 + \beta_1$. This follows from well-known properties of the exponential d.f. If initially $i > 0$, and service of the first customer is just commencing, then $T_1 = S_1$; if service of the first customer has proceeded for a time x', then the d.f. of S_1 is $[U(x) - (U(x')]/[1 - U(x')]$.

Definition (*Number of Customers in the System*): The random variable $N(T_n^+)$, (hereafter denoted by $N(T_n)$), defined as

$$N(T_n^+) = i + A(t_0, T_n) - n, \qquad \text{where } T_0 = t_0 \geqq 0 \text{ and } N(T_0) = i, \tag{2.11}$$

is the *number of customers in the system at the time of the nth departure* (*service completion*) *after time* t_0. Initially $T_0 = t_0$, and $N(T_0) = i(i = 0, 1, \cdots)$. For a fixed time, say $t_0 + t$, the random variable

$$N(t_0 + t) = i + A(t_0, T_n) + A(T_n, t) - n \qquad \text{if } n\varepsilon\{n \mid t_0 < T_n \leqq t_0 + t < T_{n+1}\} \tag{2.12}$$

is the *number of customers in the system at time* $t_0 + t$. We have

$$N(t_0 + t) = N(T_n) + A(T_n, t) \qquad \text{if } n\varepsilon\{n \mid t_0 < T_n \leqq t_0 + t < T_{n+1}\}; \tag{2.13}$$

clearly $N(T_n)$ and $N(t)$ are non-negative integer-valued random variables, and $N(t)$ is continuous on the right.

Definition (*Number of Departures*). The random variable $M(t)$ defined by

$$M(t_0 + t) = n \quad \text{if} \quad t_0 < T_n \leqq t_0 + t < T_{n+1}, \qquad \text{where } T_0 = t_0, \tag{2.14}$$

is the *number of customers who have departed* (*received service*) *in the period* $(t_0, t_0 + t)$.

Definition (*Idle Period*): An idle period is a time interval (T_n, β_{n+1}), where $\beta_{n+1} > T_n$. The length of an idle period is the random variable $\tau_I = \beta_{n+1} - T_n$. Let $t_0 = 0$ in (2.11). Then since $\beta_n < T_n$ from (2.10), $N(t) > 0$ for t such that $(\beta_n \leqq t < T_n)$ and $N(t) = 0$ for $(T_n \leqq t < \beta_{n+1})$, i.e. there are no customers in the system during an idle period, and it is preceded (and followed) by periods during which customers are always present, or "busy periods" (definition next). From familiar properties of the exponential d.f. and from (2.1), we have

$$\Pr[\tau_I \leqq x] = I(x) = 1 - e^{-\lambda x}.$$

If (T_n, β_{n+1}) and $(T_m, \beta_{m+1})(n \neq m)$ are any two idle periods, their lengths are independent random variables, each with d.f. $I(x)$.

Definition (*Busy Period*): A busy period is a time interval (β_n, T_{n+r}) $(r = 0, 1, 2, \cdots)$ such that

$$\text{(2.15i)} \quad \text{(i)} \qquad \beta_n > T_{n-1}$$

(ii) For each T_i satisfying $(\beta_n < T_i \leqq T_{n+r})$ we have

$$\text{(2.15ii)} \qquad T_i = S_i + T_{i-1}$$

$$\text{(2.15iii)} \quad \text{(iii)} \quad T_{n+r+1} = S_{n+r+1} + \beta_{n+r+1} > S_{n+r+1} + T_{n+r}$$

The length of a busy period is the random variable

$$\tau = T_{n+r} - \beta_n .$$

Put $t_0 = 0$ in (2.11) and let the conditions (2.15) hold. Then $N(t) = 0$ for t satisfying $(T_{n-1} \leqq t < \beta_n)$, and $N(t) > 0$ for t satisfying $(\beta_n \leqq t < T_{n+r})$. That is, customers are always present in the system during a busy period, and busy periods are preceded and followed by idle periods. It will be shown later that, under some circumstances, busy periods are prolonged indefinitely with positive probability.

Suppose (β_n, T_{n+r}) and $(\beta_m, \beta_{m+q})(m > n + r)$ are any two (non-overlapping) busy periods. Then it follows from the assumed arrival and departure processes that their lengths are independent random variables.

Definition. (*State of the System*): The pair of random variables $[N(T_n), T_n]$ will be called *the state of the system at the time of the nth departure*. In words, the state of the system is the number of customers left behind by the nth departing customer, and the time at which this departure takes place.

Remark: $\{[N(T_n), T_n]\}$ forms an imbedded Markov chain. $\{[N(T_n), T_n]\}$ is essentially the imbedded Markov chain of Kendall [7], [8], but is a somewhat more comprehensive description of the system. Referring to (2.5), (2.8), and (2.11), and recalling the independence of arrivals and departures, the one-step transition probabilities for the chain are, when the initial state $i > 0$,

$$\text{(2.16)} \quad \underset{i,i+h}{P(t)} \equiv P_h(t) = \Pr[N(T_{n+1}) = i + h,\ T_{n+1} \leqq t_0 + t \mid N(T_n) = i,\ T_n = t_0]$$

$$\text{(2.17)} \qquad = \int_0^t a_{h+1}(t')\, dU(t'), \qquad \begin{pmatrix} i = 1, 2, 3, \cdots \\ h = -1, 0, 1, 2, \cdots \end{pmatrix}$$

These transition probabilities are stationary; they do not apply when $N(T_n) = 0$, for when this event occurs the system remains idle until a new bunch of customers arrives and service of the $n + 1$st customer can begin (see (2.15)). We shall therefore first (Section 3) study the chain during a busy period, obtaining an expression for the *n-step zero-avoiding transition probabilities*,

$$
\begin{aligned}
P_{ij}^{(n)}(t) &= \Pr[N(T_n) = j, T_n \leqq t_0 + t, N(T_k) > 0 (k = 1, 2, \cdots, n) \mid \\
&\qquad \cdot N(T_0) = i, T_0 = t_0] \\
&\equiv \Pr[N(T_n) = j, T_n \leqq t_0 + t, N(t') > 0 (t_0 \leqq t' < T_n) \mid \\
&\qquad \cdot N(T_0) = i, T_0 = t_0]
\end{aligned}
\tag{2.18}
$$

In words, $P_{ij}^{(n)}(t)$ is the probability that the number of customers in the system passes from $i > 0$ at time t_0 to $j > 0$ immediately after the nth departure, the latter occurring before time $t_0 + t$, without having passed to zero in the meantime. Equivalently, the transition occurs "during a busy period". We call the process whose transition probabilities are (2.18) the *busy-period process.* An expression for the probabilities (2.18) are derived in Section 3. Making use of this, an expression for the joint probability distribution of $N(t + t_0)$ and $M(t_0 + t)$ is obtained. An explicit expression for the d.f. of the duration of a busy period results as a by-product.

In general, transition from $N(t_0) = i > 0$ to $N(t_0 + t) = j > 0$ can occur with $N(t') = 0$, where t' satisfies $(t_0 < t' < t)$, i.e. transitions may occur with intermediate passages to zero. We call the process that permits such transitions the *general process,* and discuss it further in Section 5 and 6, using methods of renewal theory. In Section 7 the d.f. of waiting times (waiting period durations, in our terminology) in the general process is discussed.

3. The busy period. In (2.18) t_0 can be interpreted as the instant at which a busy period commences, and $T_0 = t_0$ as the time at which service of the first customer to receive service during the period begins. Because the transition probabilities (2.17) and (2.18) are stationary, we shall consider t to be time measured from t_0 in this section. Since $U(0) = 0$

$$P_{ij}^{(n)}(0-) = 0 \qquad (n \geqq 0) \tag{3.1}$$

and

$$P_{ij}^{(0)}(t) = \delta_{ij} U_0(t), \tag{3.2}$$

the Kronecker delta multiplied by the unit step at $t = 0$. Prescription of other initial conditions is straightforward.

To derive the transition probabilities (2.18) observe that $N(T_{n+1}) = j > 1$ if $N(T_n) = j - h$ and exactly $h + 1$ customers arrive at the system in (T_n, T_{n+1}). Thus by direct enumeration the $P_{ij}^{(n)}(t)$ satisfy the forward Chapman-Kolmogorov equations

$$P_{ij}^{(n+1)}(t) = \sum_{h=-1}^{j-1} \int_0^t P_{i,j-h}^{(n)}(t - t') a_{h+1}(t') \, dU(t') \tag{3.3}$$

$$(3.4) \qquad = \sum_{h=-1}^{j-1} P_{i,j-h}^{(n)}(t) * P_h(t).$$

To obtain a formal solution to (3.4), first introduce the Laplace-Stieltjes transforms (L.S.T.)

$$p_{ij}^{(n)}(s) = \int_0^\infty e^{-st}\, dP_{ij}^{(n)}(t)$$

and

$$(3.5) \qquad p_h(s) = \int_0^\infty e^{-st}\, dP_h(t)$$

converging at least for $s > 0$, and the generating function

$$(3.6) \qquad G_{ij}(z; s) = \sum_{n=0}^{\infty} z^n p_{ij}^{(n)}(s)$$

the latter converging at least for $|z| < 1$. Then (3.4) becomes, using familiar properties of the L.S.T. and G.F.,

$$(3.7) \qquad G_{ij}(z; s) - \delta_{ij} = z \sum_{h=-1}^{j-1} G_{i,j-h}(z; s) p_h(s).$$

Next introduce the generating function

$$(3.8) \qquad g_i(z, x; s) = \sum_{j=1}^{\infty} x^j G_{i,j}(z; s)$$

again assumed to converge at least for $|x| < 1$. After some simple manipulations we have

$$(3.9) \qquad g_i(z, x; s) = \frac{x^i - zG_{i1}(z; s)p_{-1}(s)}{1 - z\pi(s, x)}$$

where the function $\pi(s, x)$ is the generating function of the transforms of the one-step probabilities,

$$(3.10) \qquad \begin{aligned} \pi(s, x) = \sum_{h=-1}^{\infty} x^h p_h(s) &= \frac{1}{x} \int_0^\infty \exp\left[-st - \lambda t\{1 - b(x)\}\right] dU(t), \\ &= \frac{1}{x} u[s + \lambda\{1 - b(x)\}] \end{aligned}$$

where $u(s)$ is the L.S.T. (2.9). Formula (3.9) depends upon the unknown function $G_{i1}(z; s)$ which must now be determined. To do this we make use of the result of

LEMMA 3.1. *For $s > 0$ and $0 < z \leqq 1$ there exists a unique root, $0 < x_1(s, z) < 1$ of the equation*

$$(3.11) \qquad x = zu[s + \lambda\{1 - b(x)\}].$$

PROOF: For fixed $s > 0$, $0 < z \leqq 1$, $I_1(x, z, s) \equiv zu[s + \lambda\{1 - b(x)\}]$ is a convex continuous function of x. Furthermore,

$$0 < I_1(0, z, s) < I_1(1, z, s) < 1.$$

Therefore there is exactly one root in the interval (0, 1).

Now if $g_i(z, x; s)$ is to generate (transforms of) probabilities, it must be bounded at least for all $0 \leqq x \leqq 1$. Since Lemma 3.1 shows that the denominator of (3.9) has a zero in this interval, we determine the numerator so as to keep the expression bounded. Thus we have

$$zG_{i1}(z; s)p_{-1}(s) = x_1^i(s, z) \tag{3.12}$$

and substituting this into (3.9) completely determines $g_i(z, x; s)$. For a similar argument see Bailey [1] and Beneš [2].

The expression (3.12) has an interesting probabilistic interpretation in its own right. Let $\tau_m^{(i)}$ be a random variable such that

$$\tau_m^{(i)} = T_m - T_0 \qquad (i, m > 0) \tag{3.13}$$

where $N(T_0) = i$ and $N(T_m) = 0$ *for the first time* thereafter. Putting this more informally, $\tau_m^{(i)}$ is the length of a busy period that begins with exactly i customers present, the first just commencing service, and ends with the departure of the mth customer to receive service. Let $M^{(i)} \geqq 1$ be a random variable denoting the number of customers discharged by the server during a busy period that begins with i customers present. If we let

$$F_m(t; i) = \Pr[\tau_m^{(i)} \leqq t, M^{(i)} = m \mid N(T_0) = i],$$

then, because of independence,

$$F_m(t; i) = P_{i1}^{(m-1)}(t) * P_{-1}(t). \tag{3.14}$$

If we now introduce the L.S.T. and G.F. we obtain from (3.14)

$$\begin{aligned} f_z(s; i) = \int_0^\infty e^{-st} \sum_{m=1}^\infty z^m \, dF_m(t; i) &= zG_{i1}(z; s)p_{-1}(s) \\ &= x_1^i(s, z) \end{aligned} \tag{3.15}$$

We have proved

THEOREM 1. *The G.F. of the L.S.T. of the joint distribution of busy period duration and the corresponding number of departures in that duration is given by*

$$f_z(s; i) = x_1(s, z)$$

where $x_1(s, z)$ *is the root of* (3.11) *discussed in Lemma* 3.1.

In Theorem 1 it is assumed that the busy period begins with a single customer in the system. If it begins with $i \geqq 1$ present, the appropriate transform is $x_1^i(s, z)$; if it begins with the entry of a bunch of customers the transform is $b[x_1(s, z)$, $b(x)$ being the G.F. of bunch size.

An explicit solution of (3.12) will now be given as

THEOREM 2. *For* $s > 0$ *and* $|z| < 1$ *the root* $x_1(s, z)$ *can be written as*

$$x_1(s, z) = \sum_{m=1}^{\infty} z^m \int_0^{\infty} e^{-st} \frac{e^{-\lambda t}}{m} \sum_{n=0}^{m-1} \frac{(\lambda t)^n}{n!} b_{m-1}^{n*} \, dU^{m*}(t); \tag{3.16a}$$

consequently

$$F_m(t; 1) = \int_0^t \frac{e^{-\gamma t'}}{m} \sum_{n=0}^{m-1} \frac{(\lambda t')^n}{n!} b_{m-1}^{n*} \, dU^{m*}(t') \tag{3.16b}$$

PROOF: Since $u[s + \lambda\{1 - b(x)\}]$ is an analytic function of x at least for $|x| < 1$ we can apply Lagrange's Theorem for series reversion [13] to (3.11); making use of the fact that

$$u^n[s + \lambda\{1 - b(z)\}] = \int_0^{\infty} \exp[-st - \lambda t\{1 - b(z)\}] \, dU^{n*}(t)$$

and expanding around the origin in powers of z gives (3.16). Observe that $F_m(t; 1)$ can be interpreted as the d.f. of the first-passage time (period) to zero from an arbitrary instant at which one customer is present, the latter just commencing service, and the number of customers receiving service in that time (period).

Expansion (3.16a), together with (3.11), can be used to verify that the root $x_1(s, z)$ is the transform of a bivariate d.f. With the aid of this fact it can be shown by direct series expansion that the functions whose transforms are given by (3.9) and (3.12) are the transition probabilities of a Markov process. The uniqueness of the solutions to (3.4) is guaranteed by properties of the transforms. We state

THEOREM 3. *The G.F. of the L.S.T. of the transition probabilities* $P_{ij}^{(n)}(t)$ *is given by* (3.9) *in terms of the root* $x_1(s, z)$ *of Lemma* 3.1:

Finally, we can obtain an expression for the joint d.f. of $N(t)$, the number of customers present in the system at a fixed time t measured from the beginning of a busy period, and $M(t)$, the number of customers who have been serviced by that time. Let

$$\begin{aligned} \mathbf{P}_{ij}^{(n)}(t) &= \Pr[N(t) = j, \quad T_n \leqq t < T_{n+1}, \\ &\qquad N(t') > 0 \ (0 \leqq t' \leqq t) \mid N(T_0) = i, \quad T_0 = 0], \\ &= \Pr[N(t) = j, \quad M(t) = n, \\ &\qquad N(t') > 0 \ (0 \leqq t' \leqq t) \mid N(T_0) = 0, \quad T_0 = 0] \end{aligned} \tag{3.17}$$

be that joint distribution. For the marginal d.f. of $N(t)$ alone, with the condition that $N(t') > 0$ $(0 \leqq t' \leqq t)$, we write

$$\mathbf{P}_{ij}(t) = \Pr[N(t) = j, \quad N(t') > 0 \ (0 \leqq t' \leqq t) \mid N(T_0) = i, \quad T_0 = 0], \tag{3.18}$$

then, by simple enumeration and independence,

$$\mathbf{P}_{ij}^{(n)}(t) = \sum_{j'=1}^{j} P_{ij'}^{(n)}(t)*[\{1 - U(t)\}a_{j-j'}(t)] \tag{3.19}$$

Summation on n gives a corresponding expression for (3.18).

Now introduce the Laplace transform of $\mathbf{P}_{ij}^{(n)}(t)$,

$$\mathbf{p}_{ij}^{(n)}(s) = \int_0^\infty e^{-st}\mathbf{P}_{ij}^{(n)}(t)\,dt$$

and the G.F. with respect to n and j, denoting the result by $\mathbf{g}_s(z, x; i)$. Then a few manipulations, using (3.19) and the properties of transforms of convolutions, show that

$$\mathbf{g}_i(z, x; s) = g_i(z, x; s)\left\{\frac{1 - u[s + \lambda - \lambda b(x)]}{s + \lambda - \lambda b(x)}\right\}, \tag{3.20}$$

the transform of $\mathbf{P}_{ij}(t)$ is obtained by letting z tend to one in (3.20).

4. Busy periods with instantaneous defections. The functional equation for the joint distribution (3.11) of busy period length and number of departures will now be derived directly, and certain ergodic properties of the process deduced. Our method is basically the same as that used by Takács [12] for deriving the marginal distributions of this joint distribution when arrivals are Poisson. We shall, however, generalize the arrival process slightly to allow for "instantaneous defection". By instantaneous defection we shall mean that when a customer arrives and finds that he cannot be served immediately he joins the queue with binomial probability p, independent of the state of the system, and departs immediately without waiting for service with probability $q = 1 - p$. Such an assumption is reasonable when the arriving customer can only discover whether or not he must wait, and not how long. This state of affairs is not uncommon in actual congestion situations. We wish to find the joint distribution of a busy period length, the number of discharged customers, and the number of defecting customers in the busy period.

Let $A_1(t)$ be the number of customers who arrive in a time interval of length t who wait, and $A_2(t)$ the number who arrive but immediately defect. Then by our assumption about the defection process

$$\Pr[A_1(t) = k_1, A_2(t) = k_2 \mid A(t) = k] = \binom{k}{k_1} p^{k_1}q^{k_2}.$$

Since $A(t)$ has distribution $\{a_k(t)\}$, we have for the joint distribution of waiting and defecting arrivals

$$a_{k_1,k_2}(t) = a_k(t)\binom{k}{k_1} p^{k_1}q^{k_2}, \qquad k_1 + k_2 = k. \tag{4.1}$$

The generating function of this joint distribution is, from (2.6),

$$(4.2) \qquad a(z_1, z_2; t) = \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{\infty} z_1^{k_1} z_2^{k_2} a_{k_1,k_2}(t) = a(pz_1 + qz_2, t)$$

Let us define the random variables $\tau_{m_1}^{(i)}$, $M_1^{(i)}$, and $M_2^{(i)}$ just as we did $\tau_m^{(i)}$ and $M^{(i)}$: $\tau_{m_1}^{(i)}$ is the length of a busy period that begins with i customers present and ends with the discharge of the m_1st customer to receive service, $M_1^{(i)}$ is the number of customers receiving service, and $M_2^{(i)}$ the number who come to the system but immediately defect.

Let

$$(4.3) \qquad F_{m_1,m_2}(t) = \Pr\{\tau_{m_1}^{(1)} \leqq t, \quad M_1^{(1)} = m_1, \quad M_2^{(1)} = m_2\}$$

We can now write down directly the equations to be satisfied by $F_{m_1,m_2}(t)$. Suppose that at some initial instant the system is in the state $[N(T_0) = i, T_0 = 0]$, i.e. $i \geqq 1$ customers are present and one has just commenced service. Then in order to pass to the zero state after exactly $m_1 + 1$ customers depart, the last departure occurring at some time not later than t, the system must (a) pass from $[N(T_0) = i, T_0 = 0]$ to $[N(T_1) = j, T_1 = t']$, where $t' < t$, and then (b) from $[N(T_1) = j, T_1 = t']$ to $[N(T_{m_1}) = 0, T_{m_1} \leqq t]$. These events are independent. Furthermore, the events of passing from $[N(T_k) = j, T_k = 0]$ to $[N(T_{k+m}) = j - 1, T_{k+m} \leqq t]$ $(j > 1)$ for the first time are independent, with the same probability, and do not depend upon j and k. Introduction of the defections does not materially alter the above observations. The following equations result:

$$(4.4) \qquad \begin{aligned} F_{1,m_2}(t) &= \int_0^t a_{0,m_2}(\tau)\, dU(\tau) \\ F_{m_1+1,m_2}(t) &= \sum_{k_1=0}^{\infty} \sum_{k_2=0}^{m_2} \int_0^t a_{k_1,k_2}(\tau)\, dU(\tau) * F_{m_1,m_2-k_2}^{k_1*}(t) \end{aligned}$$

Introducing the G.F.

$$(4.5) \qquad \sum_{m_1=1}^{\infty} \sum_{m_2=0}^{\infty} z_1^{m_1} z_2^{m_2} F_{m_1,m_2}(t)$$

convergent at least for $|z_1|$, $|z_2| \leqq 1$, and the L.S.T.

$$(4.6) \qquad f_{z_1,z_2}(s) = \int_0^{\infty} e^{-st}\, d_t\, F(z_1, z_2; t),$$

we obtain, using (4.2) and (3.10) and the properties of generating functions, the functional equation

$$(4.7) \qquad f_{z_1,z_2}(s) = z_1 u[s + \lambda + \lambda b\{p f_{z_1,z_2}(s) + q z_2\}]$$

For $s > 0$, $|z_1| \leqq 1$, $|z_2| \leqq 1$ equation (4.7) has a single root less than unity. This root is the transform of the joint d.f. (4.3). An explicit solution can be given to (4.7) using a Lagrange expansion, but we shall content ourselves with the information obtainable directly from (4.7).

Assume that the service times have expectations

$$m_i = \int_0^\infty t^i \, dU(t), \qquad i = 1, 2, \cdots, \tag{4.8a}$$

and let $\sigma^2 = m_2 - m_1^2$. Let

$$\delta_i = E[B^i], \qquad i = 1, 2, \cdots, \tag{4.8b}$$

denote the expectations of the bunch-size. It is convenient to define $\rho = \lambda\delta_1 m_1$, the *traffic intensity parameter.*

Then

LEMMA 4.1 (a) $\frac{d}{dx} u[\lambda - \lambda b(px + q)] \mid_{x=1} = p\rho$

(b) $f_{1,1}(0) = 1, \qquad p\rho \leqq 1$

(c) $f_{1,1}(0) = \epsilon < 1, \qquad p\rho > 1$

PROOF: Direct differentiation establishes (a). Using Abel's theorem, $f_{1,1}(0)$ satisfies (4.7) with $s = 0$, $z_1 = z_2 = 1$. Then, using (a) and the continuity and convexity of $u[\lambda - \lambda b(px + q)]$, (b) and (c) follow.

THEOREM 4. *When $p\rho < 1$, busy periods end in finite time with probability one. When $p\rho > 1$, busy periods last indefinitely long with probability $1 - \epsilon$, where ϵ is the root (less than unity) of*

$$x = u[\lambda - \lambda b(px + q)], \qquad 0 \leqq p \leqq 1. \tag{4.9}$$

In many practical situations busy periods alternate with idle periods; from Section 2 the latter are independent, exponentially distributed random variables. We call this the "general" process, and discuss it further in Section 5. Several properties of such processes are apparent from Theorem 4. Since, when $p\rho < 1$, return to $N = 0$ from any state will always occur in finite time, this event is "persistent" in the language of recurrent events [6]. It follows from the theory of recurrent events that $N = j$ will occur infinitely often in any realization of the general process. When $\rho > 1$ return to $N = 0$ is "transient", and the event that $N = j$, j finite, will occur only finitely many times during a process realization; eventually the number of customers in the system grows indefinitely.

From (4.9) we can deduce other properties of "transient" systems, i.e. those for which $p\rho > 1$. Suppose two queueing systems are confronted with identical arrival patterns but have different service time distributions, $U_A(t)$ and $U_B(t)$. Suppose further that $p\rho_A = p\rho_B > 1$. Then, *if* the L.S.T.'s of $U_A(t)$ and $U_B(t)$ satisfy

$$u_A[\lambda - \lambda b(px + q)] > u_B[\lambda - \lambda b(px + q)],$$

for all x in the interval (0, 1) we observe immediately that $\epsilon_A > \epsilon_B$. Hence, at least in some cases, the relative tendency for busy periods of systems to be indefinitely prolonged when traffic intensity exceeds unity can be deduced directly from the L.S.T.'s of the respective service time distributions. As an example, if $U_A(t) = 1 - e^{-t}$, and $U_B(t)$ is the distribution of "constant" service times, both having unit means, we see from their transforms that $\epsilon_A > \epsilon_B$, and can conclude that a saturated ($p\rho > 1$) "exponential" system is likely to have a greater number of finite busy periods before becoming permanently busy than is a saturated "constant" system with the same mean service time and arrival process.

Returning now to (4.7), we discuss the moments of busy period length, τ, and of the number of discharges, M, during a busy period, when we suppose no defections occur ($p = 1$). Thus, first setting $z_1 = 1$ in (4.7) we have

$$(4.10)\qquad \begin{aligned} E[\tau^{(1)}] &= \frac{\rho}{\lambda\delta_1(1-\rho)}, \qquad \rho < 1,\\ \operatorname{Var}[\tau^{(1)}] &= \frac{\sigma^2 + \rho m_2(\delta_2/\delta_1)}{(1-\rho)^3}. \end{aligned}$$

We observe that *mean* busy period length depends only upon the mean number of arrivals in a unit time interval and mean service time, while the *variance* of busy period length in general depends upon the second moment δ_2 of the distribution $\{b_j\}$ as well as mean and variances of arrivals and service times.

The moments of the distribution of the number of customers discharged during a busy period are, similarly,

$$(4.11)\qquad \begin{aligned} E[M^{(1)}] &= \frac{1}{1-\rho}\\ \operatorname{Var}[M^{(1)}] &= \frac{\rho\left(\frac{\delta_2}{\delta_1}\right) + \rho^2\left(\frac{\sigma^2}{m_1^2}\right)}{(1-\rho)^3} \end{aligned}$$

Again, the mean number of discharges depends only on mean arrivals and the mean service time. The variance depends upon second moments, both of arrivals and of service times, in much the same way as did Var $[\tau^{(1)}]$.

The expected number of defections ($p < 1$) during a busy period comes directly from (4.7) by differentiation. We have

$$E[M_2^{(1)}] = \frac{(1-p)\rho}{1-p\rho} \qquad p\rho < 1$$

The corresponding expected number of discharges following service is

$$E[M_1^{(1)}] = \frac{1}{1-p\rho} \qquad p\rho < 1$$

We can thus conclude that busy periods end in finite time and the expected number of defecting customers equals or exceeds the expected number actually served if, and only if,

$$(4.12)\qquad \frac{1}{1-p} \leqq \rho < \frac{1}{p}.$$

In order for this to be true, at least half of the arriving customers must defect, on the average.

5. The general process. The formulation of Section 2 implies that transitions between states may occur with the intervention of one or more idle periods. We call the process permitting such transitions the *general process*, and in this section relate it to the busy period process discussed earlier.

In a realization of the general process new busy periods begin at the sequence of random times $(0 < \mathbf{t}_1 < \mathbf{t}_2 < \mathbf{t}_3 \cdots)$, where time is measured from an arbitrary initial instant. We shall call $\mathbf{t}_k$ the *time of the beginning of the* kth *busy period*. From (2.15) we have that

$$\mathbf{t}_k \in \{\beta_n \mid \beta_n > T_{n-1}\}$$

Definition. The interval $(\mathbf{t}_{k-1}, \mathbf{t}_k)$ will be called a (the kth) *renewal period*. The random variable $\tau_R(k) = \mathbf{t}_k - \mathbf{t}_{k-1}$ is the length of the kth renewal period. In the general process

$$\tau_R(k) = \tau(k) + \tau_I(k) \qquad k \geqq 2 \tag{5.1}$$

where $\tau(k)$ is the length of the busy period that immediately follows the arrival at $\mathbf{t}_{k-1}$, and $\tau_I(k)$ is the length of the idle period immediately following the latter busy period, and preceding the arrival at $\mathbf{t}_k$. We have, from Section 2,

$$\Pr[\tau_I(k) \leqq x] = 1 - \exp(-\lambda x). \tag{5.2}$$

Since the busy periods beginning at $\mathbf{t}_k (k \geqq 1)$ each commence with the arrival of a bunch of customers, the latter having bunch size distribution $\{b_i\}$, we have from Section 3,

$$F(t) = \Pr[\tau(k) \leqq t] = \sum_{i=1}^{\infty} b_i F(t; i), \tag{5.3}$$

where

$$F(t; i) = \Pr[\tau^{(i)} \leqq t] = \sum_{m=1}^{\infty} F_m(t; i). \tag{5.4}$$

From Section 2 and the above $\{\tau_I(k)\} (k \geqq 2)$ is a sequence of independent random variables, each having the exponential d.f. (5.2), and $\{\tau(k)\} (k \geqq 2)$ is a sequence of independent random variables with d.f. (5.3). It follows from (5.1) that the sequence $\{\tau_R(k)\} (k \geqq 2)$ of renewal periods form a sequence of independent, identically distributed random variables with d.f.

$$R(t) = \Pr[\tau_R(k) \leqq t] = F(t) * I(t). \tag{5.5}$$

Such a sequence of random variables is called a *renewal process*, cf. Blackwell [3], Feller [5], Smith [9], [10]. We can therefore state

THEOREM 5. *The sequence of renewal periods* $\{\tau_R(k)\} (k \geqq 2)$ *constitutes a renewal process.*

Note that because of the imposition of initial conditions, the d.f. of $\tau_R(1) = \mathbf{t}_1$, the time of the beginning of the first busy period, is

$$(5.6)\qquad \begin{aligned} R(t; i) &= \Pr[\tau_R^{(i)} \leqq t] = F(t; i) * I(t), && i > 0 \\ &= I(t), && i = 0 \end{aligned}$$

where we take the initial state to be $[N(T_0) = i,\ T_0 = 0]$ when $i > 0$ and a service is beginning at $T_0 = 0$. From the above considerations the d.f. of $\mathbf{t}_k$, the time to the beginning of the kth busy period, is given by

$$(5.7)\qquad \begin{aligned} r_k(t; i) &= \Pr[\mathbf{t}_k \leqq t \mid N(T_0) = i, \quad T_0 = 0] \\ &= R(t; i) * R^{(k-2)*}(t) && (k \geqq 2; i > 0), \end{aligned}$$

and, when $N(T_0) = 0$, by

$$(5.8)\qquad r_k(t; 0) = I(t) * R^{(k-1)*}(t) \qquad (k \geqq 1; i = 0),$$

and we adopt the usual convention that $R^{0*}(t) = U_0(t)$, the unit step at the origin. We shall call $r_k(t; i)$ the *renewal distribution.*

Now in order to obtain the probability that $N(t) = j > 0$ in the general process, given that at $t = 0[N(T_0) = i,\ T_0 = 0](i > 0)$ i.e. initially there are i customers in the system, and one is just commencing service, then either (a) $N(t) = j > 0$ and $N(t') > 0(0 \leqq t' \leqq t)$, i.e. that at time t the number of customers present is j and the first busy period has not terminated, or (b) $N(t) = j > 0$ and $N(t') = 0$ at least once in $(0 \leqq t' \leqq t)$; i.e. that at time t the number of customers present is j and at least one busy period has elapsed. The probability of the event (a) is $\mathbf{P}_{ij}(t)$, as given by (3.18). The probability of the event (b) is easily seen to be

$$(5.9)\qquad \sum_{k=2}^{\infty} \mathbf{P}_j(t) * r_k(t; i) \qquad (i, j \geqq 1),$$

where

$$(5.10)\qquad \mathbf{P}_j(t) = \sum_{i=1}^{\infty} b_i \mathbf{P}_{ij}(t).$$

Summing the probabilities of the mutually exclusive events (a) and (b) we obtain

THEOREM 6. *Let*

$$(5.11)\qquad \mathbf{Q}_{ij}(t) = \Pr[N(t) = j \mid N(T_0) = i, \quad T_0 = 0]$$

be the distribution of $N(t)$ in the general process described. Then $\mathbf{Q}_{ij}(t)$ is expressible in terms of $\mathbf{P}_{ij}(t)$, the d.f. of $N(t)$ for the busy period process, and the renewal distribution $r_k(t; i)$: when $i > 0$

$$(5.12)\qquad \mathbf{Q}_{ij}(t) = \mathbf{P}_{ij}(t) + \sum_{k=2}^{\infty} \mathbf{P}_j(t) * r_k(t; i) \qquad (i, j \geqq 1),$$

and

$$\mathbf{Q}_{i0}(t) = F(t; i)*[1 - I(t)] + \sum_{k=2}^{\infty} F(t)*[1 - I(t)]*r_k(t; i) \tag{5.13}$$

when $i = 0$

$$\mathbf{Q}_{0j}(t) = \sum_{k=1}^{\infty} \mathbf{P}_j(t)*r_k(t; 0) \tag{5.14}$$

and

$$\mathbf{Q}_{00}(t) = [1 - I(t)] + \sum_{k=1}^{\infty} F(t)*[1 - I(t)]*r_k(t; 0). \tag{5.15}$$

Lastly we sketch the derivation of the joint distribution of the random variables $N(t)$ and $M(t)$ in the general process. Observe that the joint d.f. of $\tau_R(k)$, the kth ($k \geqq 2$) renewal period, and $M(k)$, the number of customers receiving service in $\tau_R(k)$, is given by

$$R_n(t) = \Pr[\tau_R(k) \leqq t, \quad M(k) = n] = \sum_{i=1}^{\infty} b_i F_n(t; i)*I(t). \tag{5.16}$$

It follows from independence considerations that the joint d.f. of $\mathbf{t}_k$, the time to the beginning of the kth busy period, and $M(\mathbf{t}_k)$, the number of service completions in that time, is given by

$$r_k^{(n)}(t; i) = R_n(t; i)*R_n^{(k-2)*}(t) \qquad (k \geqq 2; i > 0), \tag{5.17}$$

and, when $N(T_0) = 0$, by

$$r_k^{(n)}(t; 0) = I(t)*R_n^{(k-1)*}(t), \qquad (k \geqq 1; i = 0). \tag{5.18}$$

The convolution operation is to be understood as applying to both n and t.

Next let

$$\mathbf{Q}_{ij}^{(n)}(t) = \Pr[N(t) = j, \quad M(t) = n \mid N(T_0) = i, \quad T_0 = 0] \tag{5.19}$$

be the joint distribution of $N(t)$ and $M(t)$. Then, by an argument analogous to that giving Theorem 6, we obtain

$$\mathbf{Q}_{ij}^{(n)}(t) = \mathbf{P}_{ij}^{(n)}(t) + \sum_{k=2}^{\infty} \mathbf{P}_j^{(n)}(t)*r_k^{(n)}(t; i) \qquad (i > 0), \tag{5.20}$$

where again convolution applies to both n and t. The marginal distribution of $M(t)$ alone is obtained by summing on j and adding $\mathbf{Q}_{i0}^{(n)}(t)$. We observe that it is only necessary to omit the summation on k in (5.20) to obtain the joint distribution of $N(t)$, $M(t)$, and $K(t)$, where the latter is a random variable denoting the number of busy periods that have terminated in time t.

6. Ergodic properties of the general process. We shall now investigate the ergodic properties of the random variable $N(t)$ in the general process with the aid of a result in renewal theory.

Definition: The expression

$$r'(t; i) = \sum_{k=2}^{\infty} \frac{d}{dt} r_k(t; i) \tag{6.1}$$

will be called the *renewal density*.

In words, $r'(t; i)\, dt + o(dt)$ is the probability that a busy period begins in the time interval $(t, t + dt)$. We observe that $r'(t; i)$ exists for all t since $R(t)$ is convolution of $F(t)$ with $I(t)$, the latter being absolutely continuous.

Under broad conditions the renewal density converges to a constant as $t \to \infty$. We make use of a result of W. L. Smith [9], [10]. For similar results cf. Feller [5], and Blackwell [3].

THEOREM 7 (W. L. Smith): *If*

(i) *The renewal periods* $\{\tau_R(k)\}$ *are non-negative and* $E[\tau_R(k)] \leqq \infty$,

(ii) $\dfrac{dR(t)}{dt} \in L_{1+\delta}$ *for some* $\delta > 0$,

(iii) $\dfrac{dR(t)}{dt}$ *tends to zero as* t *tends to infinity,*

then

$$\lim_{t\to\infty} r'(t; i) = \frac{1}{E[\tau_R]}. \tag{6.2}$$

Referring to the definition of renewal periods, the expression (5.5), and Theorem 4, it is easy to verify that the conditions of Theorem 7 are satisfied. From (5.1), (5.2), (5.3), (4.10), and Theorem 4,

$$(6.3) \qquad \begin{aligned} E[\tau_R] &= \frac{1}{\lambda(1 - \rho)}, && \rho < 1, \\ &= \infty, && \rho \geqq 1. \end{aligned}$$

We have, then,

THEOREM 8: *The renewal density* $r'(t)$ *tends to a constant as* t *tends to infinity*:

$$(6.4) \qquad \begin{aligned} \lim_{t\to\infty} r'(t; i) &= \lambda(1 - \rho), && \rho < 1 \\ &= 0, && \rho \geqq 1. \end{aligned}$$

Now from Theorem 6, Theorem 8, (6.1), and a simple lemma (cf. Smith [10], p. 14) we have

THEOREM 9: *The distribution of* $N(t)$ *in the general process tends to a limit independent of the initial conditions as* t *tends to infinity*:

$$(6.5) \qquad \begin{aligned} \lim_{t\to\infty} Q_{i0}(t) &= 1 - \rho, && \rho < 1, \\ &= 0, && \rho \geqq 1, \end{aligned}$$

$$\lim_{t\to\infty} \mathbf{Q}_{ij}(t) = \frac{1}{E[\tau_R]} \int_0^\infty \mathbf{P}_j(t)\,dt, \qquad \rho < 1, \tag{6.6}$$
$$= 0, \qquad \rho \geqq 1.$$

We put

$$q_j = \lim_{t\to\infty} \mathbf{Q}_{ij}(t).$$

When $\rho < 1$ the generating function of $\{q_j\}$ is obtained from (6.5), (6.6), the developments of Section 3, and (6.3). Letting s tend to zero in (3.20), referring to (3.9), and recalling that $x_1(0, 1) = 1(\rho \leqq 1)$, there results the following expression for $q(x)$, the generating function of $\{q_j\}$:

$$q(x) = \sum_{j=0}^{\infty} x^j q_j = \frac{(1-\rho)(1-x)u[\lambda\{1-b(x)\}]}{u[\lambda\{1-b(x)\}] - x}. \tag{6.7}$$

The classical Pollaczek-Khintchine-Kendall formula for pure Poisson single arrivals, cf. Kendall [7], [8], comes from (6.7) by setting $b(x) = x$. Formula (6.7) can also be derived using the matrix methods of Kendall. From (6.7) it can be verified that $\{q_j\}$ forms a bona-fide probability distribution when $\rho < 1$; we shall term this distribution the long-run distribution of $N(t)$.

Moments of the distribution $\{q_j\}$ are available by differentiating (6.7). Thus for example

$$E(N) = \sum_{j=0}^{\infty} jq_j = \rho + \frac{\rho}{2(1-\rho)}\left[\left(\frac{\delta_2}{\delta_1} - 1\right) + \rho\left(1 + \frac{\sigma^2}{m_1^2}\right)\right]. \tag{6.8}$$

For a fixed value of ρ the effect of a departure from pure Poisson arrivals is to increase the average number of customers waiting. This increase is more pronounced for ρ close to unity than for ρ small.

Expression (6.7) can be expanded to yield the probabilities q_j explicitly. A useful approximation to these probabilities can frequently be obtained by making use of

LEMMA 6.1: *Suppose $\rho < 1$. If, for complex x, $b(x)$ and $u[\lambda\{1 - b(x)\}]$ converge or $1 < |x| < L$, L real and greater than unity, then*

$$x - u[\lambda\{1 - b(x)\}] = 0 \tag{6.9}$$

has two real roots: $x_1 = 1$ *and* $x_2 > 1$. The magnitudes of x_1 and x_2 are smaller than those of any other roots of (6.9). Note that the assumption that $u[\lambda\{1 - b(x)\}]$ converges for $1 < |x| < L$ implies that $1 - U(t) = o(e^{ct})(t \to \infty)$, where c is real and negative, cf. Widder [14]. This assumption is seldom restrictive in practice.

PROOF: For $x < L$, $u[\lambda\{1 - b(x)\}]$ is a continuous convex function of x. Clearly $x_1 = 1$ satisfies (6.9), and

$$\frac{d}{dx} u[\lambda\{1 - b(x)\}]\,|_{x=1} = \rho < 1.$$

Thus there exists exactly one more real root $x_2 > 1$. On the circle C_1: $|x| = x_2 - \epsilon > 1$, $u[\lambda\{1 - b(x)\}] < |x|$, so by Rouché's theorem there is exactly one root of (6.9) inside C_1. This must be $x_1 = 1$. Since ϵ can be made arbitrarily small, there are no complex roots of smaller magnitude than x_2, and there is only the root x_2 on $|x| = x_2$. Both x_1 and x_2 are simple.

Now from (6.7) and familiar properties of generating functions,

$$\sum_{n=0}^{\infty} x^n \sum_{j=0}^{n} q_j = \frac{q(x)}{1-x} \tag{6.10}$$

and

$$\sum_{j=0}^{n} q_j = \frac{1}{n!} \frac{d^n}{dx^n} \frac{q(x)}{1-x}\bigg|_{x=0}. \tag{6.11}$$

Applying the Cauchy Integral Formula, we have

$$\sum_{j=0}^{n} q_j = \frac{1}{2\pi i} \int_C \frac{q(w)}{(1-w)w^{n+1}}\, dw$$

where C is a circle in the x-plane, centered at the origin and with radius less than unity. Enlarge the contour C to C_2, a circle with center the origin and radius $x_2 + \delta$, where $\delta > 0$ is chosen so that $x_2 < x_2 + \delta < |x_3|$, x_3 being the third root of (6.9), if it exists, in order of increasing magnitude. This circle surrounds the simple poles of the integrand at $x_1 = 1$ and x_2, so

$$\sum_{j=0}^{n} q_j = 1 - r_2 + \frac{1}{2\pi i} \int_{C_2} \frac{q(w)\, dw}{(1-w)w^{n+1}} \tag{6.12}$$

where r_2 is the residue of $q(x) / ((1 - x)x^{n+1})$, evaluated at $x = x_2$:

$$r_2 = (1-\rho) \cdot \left\{\frac{-1}{u'[\lambda\{1 - b(x_2)\}]\lambda b'(x_2) + 1}\right\} \frac{1}{x_2^n}, \tag{6.13}$$

since $q(x)$ is bounded on C_2, we have finally

$$\sum_{j=0}^{n} q_j = 1 - r_2 + 0\left(\frac{1}{(z_2 + \delta)^n}\right).$$

We state this result as

THEOREM 10. *If $b(x)$, and $u[\lambda\{1 - b(x)\}]$ converge for $|x| < L$, where $L > 1$, then*

$$\sum_{j=n+1}^{\infty} q_j \sim (1-\rho)\left\{\frac{-1}{u'[\lambda\{1 - b(x_2)\}]\lambda b'(x_2) + 1}\right\} \cdot \frac{1}{x_2^n}, \tag{6.14}$$

where x_2 is the second real root of (6.9) *in order of increasing magnitude.*

In other words, the long-run probability distribution of long waiting lines is asymptotic to the geometric. From (6.14) it is apparent that in order to reduce the probability of long lines, x_2 should be increased, if this is possible in practice. Because of the convexity of $u[\lambda\{1 - b(x)\}]$, x_2 is increased if $u[\lambda\{1 - b(x)\}]$ is decreased for each $x \geqq 1$. On the basis of these observations we can state the following simple result:

THEOREM 11. *A and B are two waiting-line systems. A is characterized by an arrival process with parameter* λ_A, *generating function* $b_A(x)$, *and service time d.f.* $U_A(t)$; *B, by* λ_B, $b_B(x)$, *and* $U_B(t)$. *Then if, for all* $x \geqq 1$,

$$u_A[\lambda_A\{1 - b_A(x)\}] < u_B[\lambda_B\{1 - b_B(x)\}],$$

the long-run probability of long waiting lines, as given asymptotically by (6.14), *is smaller for system A than for B.*

Although the criterion given by Theorem 11 is crude it allows some interesting comparisons to be made. For example, if two systems have identical distributions of arrivals in any time interval ($\lambda_A = \lambda_B$; $b_A(x) = b_B(x)$), and the same mean service time (e.g. of unit length), but $U_A(t) = 1 - e^{-t}$, while $U_B(t)$ is a degenerate d.f. concentrating at unity, the probability of lines exceeding n in length is greater for system A than B, asymptotically as n tends to infinity. This result is not surprising when we compare the corresponding means and other moments of long-run line length. Similarly, if the kth member of a (hypothetical) sequence of systems has service time d.f. with LST $[u(s/k)]^k (k = 1, 2, 3, \cdots)$, and each member of the sequence has identically distributed arrivals during a time interval, then the probability of lines longer than n decreases as k increases, asymptotically as n approaches infinity. These results can be compared to those of Smith [11].

7. Waiting times. Suppose a customer arrives at the system (line plus server) at time $t \geqq 0$. "First-come, first served" dictates the order of service. Then in order to reach the server he must wait a time equal to the unelapsed service time of the customer currently being served, plus the service times of those customers ahead of him in line.

Let $X(t)$ be the unelapsed service time of the customer occupying the server at time t. Then, given that the last previous departure from the system occurred at $t - \tau$,

$$\Pr[X(t) \leqq \alpha] = \frac{U(\tau + \alpha) - U(\tau)}{1 - U(\tau)}. \tag{7.1}$$

Referring to Section 3, in particular to the developments leading to (3.17)–(3.19), we can derive an expression for the joint d.f. of $N(t)$ and $X(t)$ at time t after the beginning of a busy period. The result is easily seen to be

$$\begin{aligned} \mathbf{P}_{ij}(\alpha, t) &= \Pr[0 < X(t) \leqq \alpha, \quad N(t) = j, \\ &\qquad N(t') > 0 (0 \leqq t' \leqq t) \mid N(T_0) = i, \quad T_0 = 0] \\ &= \sum_{j'=1}^{j} P_{ij'}(t) * [\{U(t + \alpha) - U(t)\} a_{j-j'}(t)], \end{aligned} \tag{7.2}$$

We again find the introduction of transforms useful. It can be verified that

$$\int_0^\infty \int_0^\infty e^{-st} e^{-\zeta\alpha} \, d_\alpha\{U(t + \alpha) - U(t)\} \, dt = \frac{u(s) - u(\zeta)}{\zeta - s}, \tag{7.3}$$

assumed to converge at least for ζ, $s \geqq 0$. Since the first moment of $U(t)$ is assumed finite, the limit of (7.3) exists when $\zeta - s$ tends to zero. Let

$$\mathbf{p}_{ij}(\zeta, s) = \int_0^\infty \int_0^\infty e^{-st} e^{-\zeta\alpha} \, d_\alpha \mathbf{P}_{ij}(\alpha, t) \, dt \tag{7.4}$$

and

$$\mathbf{g}_i(x; \zeta, s) = \sum_{j=1}^{\infty} x^j \mathbf{p}_{ij}(\zeta, s) \tag{7.5}$$

Transforming throughout (7.2), we obtain, using (2.2), (7.3), and the familiar properties of L.S.T.'s and G.F.'s of convolutions,

$$\mathbf{g}_i(x; \zeta, s) = g_i(x; s) \frac{u[s + \lambda\{1 - b(x)\}] - u(\zeta)}{\zeta - s - \lambda\{1 - b(x)\}}, \tag{7.6}$$

where $g_i(x; s)$ is given by (3.9), (3.10,) and (3.12) after setting $z = 1$.

The total unelapsed service time at t, $W(t)$, is $X(t)$ plus the sum of the service times of the $N(t) - 1$ customers in line:

$$\begin{aligned} \mathbf{P}_i(\alpha, t) &= \Pr[0 < W(t) \leqq \alpha, \\ &\qquad N(t') > 0 (0 \leqq t' \leqq t) \mid N(T_0) = i, T_0 = 0] \\ &= \sum_{j=1}^{\infty} \int_0^\alpha U^{(j-1)*}(\alpha - y) \, d_y \mathbf{P}_{ij}(y, t). \end{aligned} \tag{7.7}$$

An expression for the transform

$$\mathbf{p}_i(\zeta, s) = \int_0^\infty \int_0^\infty e^{-st} e^{-\zeta\alpha} \, d_\alpha \mathbf{P}_i(\alpha, t) \, dt \tag{7.8}$$

comes directly from (7.6). After transformation with respect to α (L.S.T.) and t(L. T.), the right hand side of (7.7) is seen to be the generating function $\mathbf{g}_i(x; \zeta, s)$ with x replaced by $u(\zeta)$, the whole then divided by $u(\zeta)$. After a little simplification we have

$$\mathbf{p}_i(\zeta, s) = \frac{[u(\zeta)]^i - [x_1(s; 1)]^i}{s - \zeta + \lambda[1 - b\{u(\zeta)\}]}. \tag{7.9}$$

The transform of the more involved joint d.f. of $M(t)$, the number of customers who have received service by time t; $N(t)$, the number present in the system at t; and $W(t)$, for the busy period process is easly obtained. It is only necessary to replace $g_i(x; s)$ by $g_i(z, x; s)$ in (7.6) to include $M(t)$, and to replace x by $xu(\zeta)$ in the same expression, afterwards dividing by $u(\zeta)$, to account for $N(t)$.

In practice interest frequently centers around $W(t)$ for the general process described in Section 6. Let

$$\mathbf{Q}_1(\alpha, t) = \Pr[0 < W(t) \leqq \alpha \mid N(T_0) = i, \quad T_0 = 0]. \tag{7.10}$$

The enumerative argument leading to Theorem 6 can be used to show

THEOREM 12. *The d.f. of $W(t)$ for a general process is expressible in terms of the d.f. of $W(t)$ for the busy period process and the renewal distribution*:

$$\mathbf{Q}_i(\alpha, t) = \mathbf{P}_i(\alpha, t) + \sum_{k=2}^{\infty} \mathbf{P}(\alpha, t) * r_k(t; i), \qquad \alpha > 0, \tag{7.11}$$

where

$$\mathbf{P}(\alpha, t) = \sum_{i=1}^{\infty} b_i \mathbf{P}_i(\alpha, t).$$

Since $W(t) = 0$ if and only if $N(t) = 0$, we have $\mathbf{Q}_i(0, t) = \mathbf{Q}_{i0}(t)$; see (5.15).

The ergodic properties of $W(t)$ for the general process follow from those of $N(t)$. The limiting d.f. of $W(t)$ comes from (7.11), and we have

$$\begin{aligned} q(\alpha) = \lim_{t\to\infty} \mathbf{Q}_i(\alpha, t) &= \frac{1}{E(\tau_B)} \int_0^\infty \mathbf{P}(\alpha, t)\, dt, \qquad \rho < 1, \\ &= 0, \qquad \rho \geqq 1. \end{aligned} \tag{7.12}$$

The L.S.T., $\mathfrak{q}(\zeta)$, of the limiting d.f. $q(\alpha)$ comes from (7.9) by letting s tend to zero, and dividing by $E(\tau_B)$. Justification follows almost exactly that for (6.7). We obtain, after adding the long-run probability that $W(t) = 0$,

$$\begin{aligned} w(\zeta) = \mathfrak{q}(\zeta) + q_0 &= (1-\rho)\left[1 + \lambda \frac{1 - b\{u(\zeta)\}}{\zeta - \lambda[1 - b\{u(\zeta)\}]}\right] \\ &= (1-\rho)\left[1 + \frac{\rho C(\zeta)}{1 - \rho C(\zeta)}\right], \end{aligned} \tag{7.13}$$

where

$$C(\zeta) = \frac{1}{m_1 \delta_1}\left[\frac{1 - b\{u(\zeta)\}}{\zeta}\right] \tag{7.14}$$

is the L.S.T. of an (absolutely continuous) d.f. The expression (7.13) may be written as

$$w(\zeta) = (1-\rho) \sum_{n=0}^{\infty} \rho^n C^n(\zeta) \tag{7.15}$$

which shows that $q(\alpha)$ has a single jump at the origin, equal to $(1 - \rho)$, and is absolutely continuous elsewhere, cf. Beneš [2]. Notice that if all departures from pure Poisson arrivals are due to bunches arriving together,

$$w(\zeta) = \frac{(1-\rho)}{1 - \lambda\left[\dfrac{1 - b\{u(\zeta)\}}{\zeta}\right]} = \frac{1-\rho}{1 - \rho C(\zeta)} \tag{7.16}$$

which is essentially the Pollaczek-Khintchine formula, cf. Kendall [2, 3], with "customers" now made up of the bunches of individuals arriving simultaneously.

The moments of the d.f. $q(\alpha)$ come from (7.16) by differentiation with respect to ζ, the derivatives evaluated at $\zeta = 0$. For example,

$$E(W) = \frac{\rho}{1-\rho} C_1 = \frac{\rho}{2(1-\rho)} \left(\frac{\delta_2}{\delta_1} + \frac{\sigma^2}{m_1^2}\right) m_1 \tag{7.17}$$

and

$$\text{Var}\,[W] = \frac{\rho}{(1-\rho)^2} [\rho C_1^2 + (1-\rho)C_2] \tag{7.18}$$

where $\{C_i\}$ is the sequence of moments about the origin of the d.f. whose transform is $C(\zeta)$.

An approximation to $1 - \{q_0 + q(\alpha)\} = \Pr[W > \alpha]$, valid asymptotically as α increases, can be obtained by methods similar to those of Lemma 6.1 and Theorem 10. We state,

LEMMA 7.1: *Suppose $\rho < 1$. If, for complex ζ, $b\{u(\zeta)\}$ and $u(\zeta)$ converge for $-L < \text{Re}(\zeta) < \infty$ $(L > 0)$, then the denominator of (7.13), $D(\zeta) = \zeta - \lambda[1 - b\{u(\zeta)\}]$, has two real zeros, $\zeta_1 = 0$ and $\zeta_2 < 0$, such that if ζ_3 is any other real zero of $D(\zeta)$, $\text{Re}(\zeta_3) < \zeta_2$.* The proof is omitted. Note that our assumptions imply a restriction on $U(t)$; see Lemma 6.1.

We now apply the complex inversion formula for Laplace transforms, cf. Widder [14],

$$q(0) + q(\alpha) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} e^{\zeta\alpha} \frac{w(\zeta)}{\zeta}\, d\zeta, \qquad -L < c < 0 \tag{7.19}$$

Consider the rectangle in the ζ-plane with corners at $c \pm iT$ and

$$\alpha_2 - \delta \pm iT(c < \alpha_3 < \alpha_2 - \delta < 0).$$

Integrate $w(\zeta) / \zeta$ around this contour and let T tend to infinity. From Cauchy's theorem,

$$q_0 + q(\alpha) = a_1 + a_2 e^{\zeta_2\alpha} + \frac{1}{2\pi i} \int_{\alpha_2-\delta-i\infty}^{\alpha_2-\delta+i\infty} e^{\zeta\alpha} \frac{w(\zeta)}{\zeta}\, d\zeta \tag{7.20}$$

where a_1 and a_2 are the residues of $(w(\zeta) / \zeta$ at the poles $\zeta_1 = 0$ and $\zeta_2 < 0$. We have $a_1 = 1$, and

$$a_2 = (1-\rho) \frac{1}{1 + \lambda u'(\zeta_2) b'\{u(\zeta_2)\}}.$$

Since $w(\zeta)$ is bounded on the line of integration in (7.19), we have

$$q_0 + q(\alpha) = a_1 + a_2 e^{\zeta_2\alpha} + 0(e^{(\zeta_2-\delta)\alpha}), \tag{7.21}$$

thus we have

THEOREM 12. *The probability of waiting times exceeding α, when the d.f. (7.12) can be assumed to apply, is asymptotically exponential*

$$1 - \{q_0 + q(\alpha)\} \sim a_2 e^{\zeta_2\alpha}.$$

Compare Theorem 10 and the results of Smith [11].

8. Acknowledgements. The writer is indebted to Professor William Feller, Dr. Robert Hooke, and the referee for comments leading to clarification of the original presentation.

REFERENCES

[1] N. T. J. BAILEY, "A continuous time treatment of a simple queue using generating functions", *J. Royal Stat. Soc., Series B*, Vol. 15, (1954), pp. 288–291.

[2] V. BENEŠ, "On Queues with Poisson Arrivals", *Ann. Math. Stat.*, Vol. 28 (1957), pp. 670–677.

[3] D. BLACKWELL, "A renewal theorem", *Duke Math. J.*, Vol. 15 (1948), pp. 145–150.

[4] J. L. DOOB, *Stochastic Processes*, John Wiley and Sons,Inc., New York, 1953.

[5] W. FELLER, "On the integral equation of renewal theory," *Ann. Math. Stat.*, Vol. 12 (1941), pp. 243–267.

[6] W. FELLER, *An Introduction to Probability Theory and Its Applications* (Second Edition), John Wiley and Sons, Inc., New York, 1957.

[7] D. G. KENDALL, "Stochastic Processes Occurring in the Theory of Queues and Their Analysis by the Method of the Imbedded Markov Chain", *Ann. Math. Stat.*, Vol. 24 (1953), pp. 338–354.

[8] D. G. KENDALL, "Some problems in the theory of queues", *J. Royal Stat. Soc. Series B*, Vol. XIII (1951); (with references and discussion), pp. 151–185.

[9] W. L. SMITH, "Extensions of a renewal theorem", *Proc. Cambridge Phil. Soc.*, Vol. 51 (1955), pp. 629–638.

[10] W. L. SMITH, "Asymptotic Renewal Theorems", *Proc. Royal Soc. Edinburgh, Sec. A*, Vol. LXIV (1954), pp. 9–48.

[11] W. L. SMITH, "On the distribution of queueing times", *Proc. Cambridge Phil. Soc.*, Vol. 49 (1953), pp. 449–461.

[12] L. TAKÁCS, "Investigation of waiting time problems by reduction to Markov processes", *Acta Math.* (Budapest) Vol. 6 (1955), pp. 101–129.

[13] E. T. WHITTAKER AND G. N. WATSON, *Modern Analysis*, Cambridge, 1946.

[14] D. V. WIDDER, *The Laplace Transform*, Princeton, 1941.