THE ERGODICITY OF SERIES QUEUES WITH GENERAL PRIORITIES[1]

By Bruce McK. Johnson

Northwestern University

1. Introduction. In this paper we study the structure of and obtain certain limit theorems for a rather general and complex queue process. Specifically, the model considered is the following: Let t_j, $j = 1, \cdots$, denote arrival epochs and set $\tau_j = t_j - t_{j-1}$. The $\{\tau_j\}$ are assumed independent and identically distributed with finite expectation. At each arrival epoch the item type $(i = 1, \cdots, r)$ of the current arrival is determined by a multinomial experiment. That is, α_i, $i = 1, \cdots, r$, is the probability that an arriving item is of type i. Each arriving item is to be serviced first by a facility 1, secondly by a facility 2, $\cdots$, and finally by a facility q (q queues in series). The service time of type i items at facility k, $S_{i,k}$, are assumed nonnegative and independent and identically distributed with finite expectation. The service times of the jth arrival, $S_1^j, \cdots, S_q^j$, are taken to be independent. Also, all service times and interarrival times, the τ, are assumed independent.

The physical operation of the system is restricted as follows. No item is allowed to wait in a queue if the corresponding facility is idle. Also, after the servicing of an item is started it may not be displaced at the facility by another item. At those epochs when a facility becomes idle (completion epochs) a priority structure is employed to make a decision as to which of the items waiting in the corresponding queue, if any, is to be serviced next. The admissible priority structures (APS) are general in that they may depend on the state of the system at the time of such a decision. Although there are other possibilities, the APS are assumed to depend on: the elapsed time since the last arrival (V), the composition of the queues ($(L_1, \cdots, L_q)$, where $L_j = (L_{1,j}, \cdots, L_{q,j})$ and $L_{i,j}$ is the number of type i items waiting in queue j and being serviced by facility j), the types of items being serviced at the q queues $(K_1, \cdots, K_q)$, and the elapsed service time of the items being serviced $(U_1, \cdots, U_q)$. For simplicity we assume that items of the same type waiting in the same queue are not distinguishable to the priority structure. Thus, the decision made at a completion epoch is the type of item to be processed next. The order of servicing items of the same type at a facility is taken to be the order in which they joined the queue. Finally, we assume that the decision made at facility j, $j = 1, \cdots, q$, is a measurable function of the above variables with range, in $1, \cdots, r$.

Models of this generality have not evidently been studied. Sacks [10] (see also Loynes [9]) gives necessary and sufficient conditions for the existence of a bona fide limit distribution of the total time an item waits in q series queues. However,

Received 17 September 1964; revised 10 August 1965.

[1] Work supported in part by the National Science Foundation, GP-87, at Northwestern University. Based on a section of the author's Ph.D. dissertation written at Cornell University.

this analysis does not generalize to accommodate priority structures. In the case of a single queue, priority structures have been investigated by several workers. The model generally considered differs from the above for $q = 1$ in that the τ are assumed exponentially distributed and the priority structure is a constant ordering of item types. For this model, Kesten and Runnenburg [7] establish necessary and sufficient conditions for the existence of bona fide limit distributions of the waiting time of each item type and the queue composition at completion epochs.

The general method of analysis we employ consists of exhibiting a regenerative structure for the processes of interest and appealing to certain renewal limit theorems. In Section 3 a single queue is considered. Although the results obtained are superficial and likely known, they have not evidently been stated. In Section 4 the general case of q queues in series is studied. In all cases sufficient (also necessary for a single queue) conditions are given for the existence of bona fide limit distributions of queue composition, for both the continuous time and imbedded (on arrival epochs) processes. These conditions are also sufficient for the existence of bona fide limit distributions of the total waiting time for each item type. For $q > 1$ more is required than the usual condition

$$E\tau > \sum_{i=1}^{r} \alpha_i ES_{i,k}, \qquad k = 1, \cdots, q. \tag{1}$$

However, if τ is unbounded or $S_{i,k} = S_k$, $i = 1, \cdots, r$ and $k = 1, \cdots, q$, these conditions reduce to (1). In Section 5 a few extensions of the model for which the present analysis remains valid are considered.

2. Preliminaries. Let Ω be the space of elements ω which are sequences of the form $\{(\tau_n, i_n, s_1^n, \cdots, s_q^n)\}$ where $i_n = 1, \cdots, r$ and $\tau_n, s_1^n, \cdots, s_q^n > 0$. (Although not entirely necessary, this is the simplest way to avoid ambiguity in the random times t_j.) The entries in the nth term of ω are interpreted as the values taken on at ω by τ_n, the type of the nth arrival, and the S_j^n respectively. The measure space $\{\Omega, \mathcal{B}, P\}$, where $\mathcal{B}$ is the Borel field generated by the above variables and P is the corresponding measure, will be denoted simply by Ω. We shall investigate the Ω measurable Markov processes $\{Z_t = [V \times (L_1, K_1, U_1) \times \cdots \times (L_q, K_q, U_q)]_t, 0 \leqq t < \infty\}$ (with some obvious conventions the paths $Z_t(\omega)$ may be taken to be right continuous) and $\{Z_{t_j} = [(L, K, U) \times \cdots \times (L, K, U)]_{t_j}, 1 \leqq j < \infty\}$ taking values respectively in $\mathfrak{X} = \{[v \times (l_1, k_1, u_1) \times \cdots \times (l_q, k_q, u_q)]\}$ and $\mathfrak{X}' = \{[(l_1, k_1, u_1) \times \cdots \times (l_1, k_1, u_1)]\}$, where $l_j = (l_{1,j}, \cdots, l_{r,j})$ and $l_{i,j} = 0, 1, \cdots, k_j = 0, 1, \cdots, r$, and $v, u_j \geqq 0$. The definitions of the Z_t and Z_{t_j} processes entail an APS which remains fixed throughout the discussion. Let $\mathcal{A}[\mathcal{A}']$ be the class of left semi-closed intervals in $\mathfrak{X}[\mathfrak{X}']$ and $\mathcal{B}(\mathcal{A})[\mathcal{B}(\mathcal{A}')]$ the generated Borel field. Let $\mathcal{F}(\mathcal{A})$ be the field of finite disjoint unions of $\mathcal{A}$ sets.

It will be seen that there is a regenerative event **R**, Ω measurable, with the properties to be enumerated below.

(i) **R** occurs wp 1. Let $X_j[M_j]$ be the time [number of arrivals] between the jth and $j - 1$st occurrences of **R** and $n(t)[n'(j)]$ the number of occurrences of **R** in $(0, t]$ [the first j arrivals].

(ii) The $\{X_j\}$ and $\{M_j\}$ are sequences of independent and, for $j > 1$, identically distributed random variables with EX and EM finite. Let $T_n = X_1 + \cdots + X_n$ and $N_n = M_1 + \cdots + M_n$. Let $H(t) = En(t)$ and $H'(j) = En'(j)$.

(iii) There exist $\phi(\cdot, t)$ and $\theta(\cdot, j)$, completely additive on $\mathcal{B}(\mathcal{A})$ and $\mathcal{B}(\mathcal{A}')$ for fixed t and j, for which

$$\phi(A, t - x) = \Pr \{Z_t \varepsilon A \mid n(t) \geqq 1, T_{n(t)} = x\} \text{ a.e. } [H], \tag{2}$$

$$\theta(A', j - i) = \Pr \{Z_{t_j} \varepsilon A' \mid n'(j) \geqq 1, N_{n'(j)} = i\} \text{ a.e. } [H'], \tag{3}$$

when $A \varepsilon \mathcal{B}(\mathcal{A})$ and $A' \varepsilon \mathcal{B}(\mathcal{A}')$. Then it is well known (Beneš [1] and Smith [12], [13]) that

$$P\{Z_t \varepsilon A\} = \Pr \{Z_t \varepsilon A, n(t) = 0\} + \int_0^t \phi(A, t - x)[1 - F_X(t - x)]\, dH(x). \tag{4}$$

and

$$P\{Z_{t_j} \varepsilon A'\} = \Pr \{Z_{t_j} \varepsilon A', n'(j) = 0\} + \textstyle\sum_{i=1}^{j} \theta(A', j - i)[1 - F_M(j - i)][H'(i) - H'(i - 1)] \tag{5}$$

are valid representations. In (4) and (5), as well as below, F_Y is the df of Y.

The following versions of the so called Key renewal theorem (Beneš [2] and Smith [11], [14]) will be employed in the study of the limit behavior of (4) and (5).

THEOREM A. *If* **R** *is certain, X is not a lattice variable, ψ is continuous* a.e. [*Lebesgue measure*], *and* $\sum_{j=1}^{\infty} \sup_{j-1 \leqq t < j} |\psi(t)| < \infty$, *then*

$$\lim_{t\to\infty} \int_0^t \psi(t - x)\, dH(x) = (EX)^{-1} \int_0^\infty \psi(x)\, dx.$$

THEOREM B. *If* **R** *is certain and M has period* 1, *then*

$$\lim_{j\to\infty} \textstyle\sum_{i=1}^{j} \psi(j - i)[H'(i) - H'(i - 1)] = (EM)_i^{-1} \sum_{i=1}^{\infty} \psi(i),$$

whenever the sum on the right is convergent.

If EX or EM is infinite the corresponding limit is zero.

3. A single priority queue. In the case of a single queue the regenerative structure is easily identified (Kendall [6] and Smith [12]). Let **R** be the event "an arriving item finds the facility idle." In view of the general independence of the model and the fact that the APS depend only on the state of the system at the time a decision is affected, it is evident that there exist ϕ and θ with the stipulated properties. Also, it is clear that the $\{X_j\}$ and $\{M_j\}$ are independent and identically distributed, if they exist. To simplify statements the uninteresting case of degenerate τ and S_i with $\Pr \{\tau = S_i\} = 1$, for $i = 1, \cdots, r$, is excluded.

LEMMA 1. *If* $\infty > E\tau \geqq \sum_{i=1}^{r} \alpha_i ES_i$, *then* **R** *is certain.* (1) *is necessary and sufficient for EX and EM to be finite.*

PROOF. We observe that the distributions of X and M are invariant with respect to the APS. This follows from the fact that the APS can only achieve reorderings of the queue. Then we may, and do, assume that the df of the service time of each arrival is $\sum_{i=1}^{r} \alpha_i F_{S_i}$. The lemma is thus reduced to an assertion for an ordinary queue: the proof of which is essentially contained in the result and discussion of Lindley [8].

The discrete parameter process $\{Z_{t_j}, 1 \leqq j < \infty\}$ will be considered first. Let $W_{i,j}$ be the waiting time of the jth arrival conditional on its being a type i item. It follows from the nature of the APS that $W_{i,j}$ is conditionally independent of the history of the process prior to t_j given Z_{t_j}. Let $F_{W_i}(\cdot : l, k, u)$ be the corresponding wide sense conditional df: a Baire function of l, k, and u. The F_{W_i} determined by different j are equal wp 1 due to the assumptions on τ, S, and the APS.

THEOREM 1–d. *If* (1) *is satisfied, then for* $A' \varepsilon \mathcal{B}(\mathcal{A}')$ *and* $1 \leqq i \leqq r$,

$$\lim_{j\to\infty} P\{Z_{t_j} \varepsilon A'\} = (EM)^{-1} \sum_{p=1}^{\infty} \theta(A', p)[1 - F_M(p)], \tag{6}$$

and

$$\lim_{j\to\infty} \Pr\{W_{i,j} < w\} = \sum_k \sum_l \int_0^\infty F(w{:}l, k, u)\, d_u\sigma(l, k, u), \tag{7}$$

where σ *is the limit* (6). *If* (1) *is not satisfied and* A' *is bounded in all* l_i, *the limit* (6) *is zero.*

PROOF. The first assertion is an immediate consequence of Lemma 1 and Theorem B. Let $A' = \{l, k, [0, u)\}$. Then if the system is initially empty, and by Lemma 1 we may as well make this assumption,

$$\Pr\{W_{i,j} < w\} = \sum_k \sum_l \int_0^\infty F_{W_i}(w{:}l, k, u)\, d_u \Pr\{Z_t \varepsilon A'\}. \tag{8}$$

If the measure in (8) is rewritten as $\sum_{n=1}^{j} \theta(A', j - n)[1 - F_M(j - n)] \cdot [H'(n+) - H'(n-)]$, the assertion follows with the interchange of the limit and integration operations. This, being routinely justified in view of the non-negativity, (6), and $H'(n) - H'(n - 1) \to (EM)^{-1} < \infty$. To prove the final assertion it will suffice to consider the boundry case $E\tau = \sum_{i=1}^{r} \alpha_i ES_i$. Suppose that for some A', bounded in the l_i, $\limsup_{j\to\infty} P\{Z_{t_j} \varepsilon A'\} > 0$. Then as there is clearly an n for which $\Pr\{\mathbf{R} \text{ occurs at } t_{j+n} \mid Z_{t_j} A'\} > 0$ for every j, it follows that $\limsup_{j\to\infty} \Pr\{\mathbf{R} \text{ occurs at } t_j\} > 0$. This however contradicts Feller's renewal theorem [5] as $EM = \infty$ by Lemma 1. In this regard, (1) is not necessary for (7) to exist as a bona fide probability distribution for some i (e.g. Kesten and Runnenburg [7]). In general, this question depends intimately on the details of the APS.

In considering the continuous parameter process $\{Z_t, 0 \leqq t < \infty\}$ two cases are identified depending on whether τ, and hence X, is or is not a lattice variable. The result in an obvious notation is

THEOREM 1–c. *If* (1) *is satisfied, and if* X *is not a lattice variable,*

$$\lim_{t\to\infty} P\{Z_t \varepsilon A\} = (EX)^{-1} \int_0^\infty \phi(A, x)[1 - F_X(x)]\, dx, \quad \text{for} \quad A \varepsilon \mathcal{F}(\mathcal{A}), \tag{9}$$

if X *has period* λ *and* $0 \leqq \xi < \lambda$

(10) $\lim_{j\to\infty} P\{Z_{j\lambda+\xi} \varepsilon A\} = \lambda(EX)^{-1} \sum_{i=1}^{\infty} \phi(A, i\lambda + \xi)[1 - F_X(i\lambda + \xi)]$,

for $A \varepsilon \mathcal{B}(\mathcal{A})$. If (1) is not satisfied and A is bounded in all l_i, the appropriate limit is zero.

PROOF. The second assertion is an immediate consequence of Lemma 1 and an obvious modification of Theorem B. The final assertion is proved in exactly the manner the corresponding part of Theorem 1–d was proved, Blackwell's theorem [3] replacing Feller's theorem.

To prove (9) we first note that as $EX < \infty$ the hypothesis of Theorem A will be satisfied once the continuity property of the kernel $\psi(\cdot) = \phi(A, \cdot)[1 - F_X(\cdot)]$ is established. Let $A \varepsilon \mathcal{A}$. We employ Lemma 2 of Smith [14] to show the kernel is of bounded total variation in each finite interval $I = (t', t'']$, $0 \leqq t' < t'' < \infty$. Define

$$\begin{aligned} \delta_{I,A} &= 0 \quad \text{for} \quad Z_{t'} \quad \text{and} \quad Z_{t''} \,\varepsilon\, A \\ &= -1 \quad \text{for} \quad Z_{t'} \not\varepsilon A \quad \text{and} \quad Z_{t''} \,\varepsilon\, A \\ &= 1 \quad \text{for} \quad Z_{t'} \,\varepsilon\, A \quad \text{and} \quad Z_{t''} \not\varepsilon A. \end{aligned}$$

Then it is easily verified that $Y_{I,A} = 2$ [number of arrivals in $I+$ number of completions in I] $+ \delta_{I,A}$ satisfies Smith's lemma. Thus, from the representation (4) and Theorem A, (9) is valid for $A \varepsilon \mathcal{A}$. The extension of (9) to $\mathcal{F}(\mathcal{A})$ is immediate.

The validity of (9) for $A \varepsilon \mathcal{B}(\mathcal{A})$ is evidently a more delicate matter. It appears that additional information is required. For example, if τ is absolutely continuous it may be seen from Theorem 1 of [14] that (9) is valid for $\mathcal{B}(\mathcal{A})$.

4. Priority queues in series. A crucial problem in the analysis of a finite number of queues in series is the identification of an appropriate regenerative event. While we could take "an arriving item finds all q facilities idle" as the regenerative event, it would then be necessary to impose the strong requirement $\Pr\{\tau \geqq \sum_{k=1}^{q} S_{i,k}\} > 0$ for at least one i. Otherwise the event is not possible. Although we do not find a regenerative structure without making some auxiliary assumption on τ and $S_{i,k}$, it is possible to do considerably better.

Let $c = (c_1, \cdots, c_q)$ with $0 = c_1 < c_2 < \cdots < c_q < \infty$. Suppose that the following are satisfied in $[t_n, t_n + c_q]$ at ω:

(i) all preceding arrivals, $1, \cdots, n-1$, are completed by the kth facility not later than $t_n + c_k$, $k = 1, \cdots, q$, regardless of the order in which they are serviced after t_n if the nth and subsequent arrivals do not join the system,

(ii) the service times of the nth arrival satisfy $\sum_{i=1}^{k} S_i^n \geqq c_{k+1}$ for $k = 1, \cdots, q-1$, and

(iii) the order of service in the interval $[t_n, t_n + c_q]$ of the nth and subsequent arrivals is the order of arrival, independent of the servicing of arrivals $1, \cdots, n-1$.

Then we say that $\mathbf{R}_c$ occurs at $t_n + c_q$. That the occurrences of $\mathbf{R}_c$ will in fact provide a regenerative structure is easily seen. The following statements are relative to $\{\mathbf{R}_c$ occurs at $t_n + c_q\}$. First, it is evident that the nth and subsequent arrivals do not wait in any queue because one of the first $n - 1$ arrivals is being serviced. Secondly, by (iii) there are no nontrivial decisions made relative to the servicing of the nth and subsequent arrivals in $[t_n, t_n + c_q]$ which might be influenced through the APS by the first $n - 1$ arrivals. Finally, (i) insures that the conditions imposed on the τ_{j+1} and $S_1^j, \cdots, S_q^j, j \geqq n$, by an occurrence of $\mathbf{R}_c$ at $t_n + c_q$ are independent of the history of the process up to t_n. It follows that the $Z_t[Z_{t_j}]$ has a common law of evolution after each occurrence of $\mathbf{R}_c$. Let X_j^c be the time between the jth and $j - 1$st occurrences of $\mathbf{R}_c$. If $\mathbf{R}_c$ occurs for the jth and $j - 1$st times at $t_{n_1} + c_q$ and $t_{n_2} + c_q$, then $M_j^c = n_2 - n_1$. It is evident that the $\{X_j^c\}$ and $\{M_j^c\}$ are independent and, for $j > 1$, identically distributed, if they exist.

The determination of a necessary and sufficient condition on the τ and $S_{i,k}$ for $\mathbf{R}_c$ to be possible for some c and all APS is apparently difficult. An obvious necessary condition is that there exist item types i' and i'' and a c for which $\Pr\{\sum_{j=1}^k S_{i',j} - \tau \leqq c_k,\ k = 1, \cdots, q\} > 0$ and $\Pr\{\sum_{j=1}^k S_{i'',j} \geqq c_{k+1},\ k = 1, \cdots, q - 1\} > 0$. That this condition is not also sufficient is easily seen by examining the example: $\tau = 10$, $S_{1,1} = 11$, $S_{1,2} = 8$, $S_{2,1} = 5$, $S_{2,2} = 15$, all wp 1, and $\alpha_1 = \frac{3}{4}$. Sufficient conditions are numerous, but tend to depend on the specific properties of the APS under consideration. We shall be content with the following simple, yet reasonably general, condition:

(11) There exists an item type, say,

$$i^* \quad \text{for which} \quad \Pr\{\tau > S_{i^*,k}\} > 0 \quad \text{for} \quad k = 1, \cdots, q.$$

Before establishing the sufficiency of (11), we note that the condition is not necessary. For example, $\mathbf{R}_c$ is possible with $c_1 = 0$ and $c_2 = 6$ if $\tau = 10$, $S_{1,1} = 11$, $S_{1,2} = 5$, $S_{2,1} = 5$, $S_{2,2} = 11$, all wp 1, and $\alpha_1 = \frac{1}{2}$.

Let $c_0', c_1', \cdots, c_q'$ be positive constants with $c_0' > c_j'$, $j = 1, \cdots, q$, which satisfy $\Pr\{S_{i^*,k} \geqq c_k'\} > 0$, $\Pr\{S_{i^*,k} \leqq c_k'\} > 0$, $k = 1, \cdots, q$, and $\Pr\{\tau \geqq c_0'\} > 0$. It may easily be seen that such c' exist by (11). We shall say that an i^* arrival occurs at t_n if $\tau_n \geqq c_0'$, the arriving item is of type i^*, and $S_k^n \leqq c_k'$, $k = 1, \cdots, q$. Let $i^*(n', n'')$ denote the event "arrivals n' through n'' are i^* arrivals." Let $D_k(t)$ be the least upper bound, over all orders of servicing, of the time required to complete all items in the system at time $t-$ at facility k if arrivals at t and later are not allowed to join the system. Let n_0 be fixed and $0 \leqq d_k^0 < \infty$, $k = 1, \cdots, q$, and define $\Lambda(n_0, n) = \{D_k(t_{n_0}) \leqq d_k^0,\ k = 1, \cdots, q,\ i^*(n_0, n)\}$. If arrivals n_0 and $n_0 + 1$ are i^* arrivals, then $D_1(t_{n_0+1}) \leqq \max(0, d_1^0 - (c_0' - c_1'))$. Thus, there are finite n_1 and d_2^1 for which $D_1(t_n) = 0$ and $D_2(t_{n_1}) \leqq d_2^1$ on $\Lambda(n_0, n_1)$. If arrival $n_1 + 1$ is an i^* arrival, then $D_1(t_{n_1+1}) = 0$ and $D_2(t_{n_1+1}) \leqq \max(0, d_2^1 - (c_0' - c_2'))$ on $\Lambda(n_0, n_1 + 1)\{c_1' \leqq D_2(t_{n_1+1}) \leqq d_2^1\}$ and $D_1(t_{n_1+1}) = 0$ and $D_2(t_{n_1+1}) \leqq \max(0, c_2' - (c_0' - c_1'))$ on $\Lambda(n_0, n_1 + 1) \cdot \{D_2(t_{n_1}) \leqq c_1'\}$. Thus, there are finite n_2 and d_3^2 such that $D_1(t_{n_2}) = 0$, $D_2(t_{n_2}) \leqq c_1'$ (the second of the two possibilities eventually holds), $D_3(t_{n_2}) \leqq d_3^2$ on $\Lambda(n_0, n_2)$. Continuing this construction there is an n such that $D_1(t_n) = 0$, $D_k(t_n) \leqq \sum_{j=1}^{k-1} c_j'$, $k = 2, \cdots, q$ on $\Lambda(n_0, n - 1)$. Finally, if $(m - n)c_0' \geqq$

$\sum_{k=1}^{q-1} c_k'$ and $\tau_j \geqq c_0'$, arrival j is type i^*, $S_k^j \geqq c_k'$ for $j = n, \cdots, m$ and $k = 1, \cdots, q$, then $\mathbf{R}_c$ occurs at $t_n + c_q$ with $c_1 = 0$ and $c_k = \sum_{i=1}^{k-1} c_i'$, $k = 2 \cdots, q$. As each of these sequences of arrivals has positive probability, there follows

LEMMA 2. *If* (11) *is satisfied there exists a c for which* $\mathbf{R}_c$ *is possible for every* APS.

We next show that $\mathbf{R}_c$ is certain and that EX^c and EM^c are finite under (1) and (11). These conditions are not of course unrelated. If $S_{i,k} = S_k$, $i = 1, \cdots, r$ and $k = 1, \cdots, q$, and (1) is satisfied, then (11) is satisfied. Also, if τ is not bounded and (1) is satisfied, then (11) is satisfied. However, the counterexample to the necessity of (11) satisfies (1).

THEOREM 2. *If* (1) *and* (11) *are satisfied, then* $\mathbf{R}_c$ *is certain. Condition* (1) *is necessary, and under* (11), *sufficient for* EX^c *and* EM^c *to be finite.*

PROOF. To fix ideas the case of $q = 2$ will be considered first. Let $\mathbf{R}$ again be defined as "an arriving item finds the first facility idle." Let X, M, T, and N be defined as above for this event. Let $\hat{S}_n$ be the sum of the S_2^j for arrivals from the $n - 1$st to the nth occurrence of $\mathbf{R}$, exclusive of the arrival at the nth occurrence. Then the $\hat{S}_n$ are clearly independent and, for $n > 1$, identically distributed. As $N_j \geqq j$ wp 1 and

$$|\sum_{j=1}^{N_n} \tau_j - S_2^{j-1}/N_n - E\tau + \sum_{i=1}^{r} \alpha_i ES_{i,2}|$$
$$\leqq \sup_{k \geqq n} |\sum_{j=1}^{k} (\tau_j - S_2^{j-1})/k - E\tau + \sum_{i=1}^{r} \alpha_i ES_{i,2}| \text{ wp 1},$$

it follows by an application of the strong law of large numbers (SLLN) that

$$\sum_{j=1}^{N_n} (\tau_j - S_2^{j-1})/N_n \to E\tau - \sum_{i=1}^{r} \alpha_i ES_{i,2} \quad \text{wp 1}.$$

Also, $N_n/n \to EM < \infty$ wp 1. Then as $\sum_{j=1}^{n} (X_j - S_j)/n = (N_n/n) \cdot \sum_{j=1}^{N_n} (\tau_j - S_2^{j-1})/N_n$, it is concluded that

$$E\hat{S} < EX < \infty. \tag{12}$$

A dominating process is now constructed for the second queue. We consider an ordinary queue process with interarrival times X and service times $\hat{S}$. The only difference between this process and the one usually studied is that the X_n and $\hat{S}_n$ are not independent (Lindley [8] assumes but does not use this independence). Let $D(T_n)$ be the total uncompleted service time of the composite items waiting and being serviced in the new system at time T_n, including the contribution of the arrival at T_n, $\hat{S}_n$. Then it is evident that $D(T_n) \geqq D_2(T_n)$ wp 1. First, $D_2(T_n)$ is, as the first facility is idle, the total uncompleted service time of the items waiting in the second queue and being serviced by the second facility. Then, as the arrivals to the dominating process through the epoch T_n are composed of precisely those items which arrive to the original process up to T_n, and as no composite item arrives to the dominating process before any of the corresponding items reach the second queue, the inequality follows. It is next shown that $D(T_n) \leqq d < \infty$ for infinitely many values of n wp 1.

It is evident that

$$D(T_n) = \max \hat{S}_n, \hat{S}_{n-1} - X_n + \hat{S}_n, \cdots, \hat{S}_1 - X_2 + \cdots - X_n + \hat{S}_n. \tag{13}$$

Let $\epsilon > 0$ and such that $E(\hat{S} - X) + \epsilon < 0$. Consider the following Ω sets:

$C_1 = \{\hat{S}_n \leqq d, \hat{S}_{n-1} + \hat{S}_n - X_n \leqq d, \cdots, \sum_{j=m}^{n} (\hat{S}_j - X_j) + \hat{S}_{m-1} \leqq d\}$,

$C_2 = \{\max_{1 \leqq k \leqq m} \hat{S}_k/(n-k) \leqq E(X - \hat{S}) - \epsilon\}$,

and

$C_3 = \{\sum_{j=k+1}^{n} (\hat{S}_j - X_j)/(n-k) \leqq E(\hat{S} - X) + \epsilon, k = 1, \cdots, m\}$.

In view of (12) the measure of C_3 can be made arbitrarily close to 1 by choosing $n - m$ and m sufficiently large. For then an application of the SLLN will suffice. The measure of C_2 can also be made close to 1. For by independence

$$\begin{aligned} \Pr\{\max_{k \leqq m} \hat{S}_k/(n-k) &\leqq E(X - \hat{S}) - \epsilon\} \\ &\geqq \Pr\{\sup_{k \geqq n-m} \hat{S}_k/k \leqq E(X - \hat{S}) - \epsilon\} \\ &= \prod_{k=n-m}^{\infty} [1 - \Pr\{\hat{S}_k/k > E(X - \hat{S}) - \epsilon\}], \end{aligned} \tag{14}$$

and as $\sum_{k=n-m}^{\infty} \Pr\{\hat{S} > k(E(X - \hat{S}) - \epsilon)\} \leqq E\hat{S}(E(X - \hat{S}) - \epsilon)^{-1} < \infty$, the infinite product converges. That is, is close to 1 for m large. Finally, with m and n as above, the measure of C_1 may be made close to 1 by choosing d sufficiently large. It follows then that

$$\Pr\{D(T_n) \leqq d\} \geqq P\{C_1 \cap C_2 \cap C_3\} > 0 \tag{15}$$

for n sufficiently large.

Let d be fixed. As the process $\{D(T_n), 1 \leqq n < \infty\}$ has stationary transition probabilities, it makes sense to set $h(d, d') = \Pr\{D(T_n) > d, n > n' \mid D(T_{n'}) = d'\}$. From (13) if $d'' \geqq d'$, then $h(d, d'') \geqq h(d, d')$. If $\Pr\{X \geqq \hat{S}\} = 1$, it is easily seen that $h(d, d') = 0$ for any d with $\Pr\{\hat{S} < d\} > 0$. Otherwise, if $h(d, d') > 0$ for some d', then $h(d, 0) > 0$ as a transition from 0 to a state larger than d' has positive probability. Then $\Pr\{D(T_n) \leqq d$ for at least m values of $n\} \leqq (1 - h(d, 0))^m$. Thus $h(d, d') > 0$ contradicts (15) and $D(T_n) \leqq d$ for infinitely many values of n wp 1.

Accordingly, $D_2(T_n) \leqq d$ for infinitely many values of n wp 1. After each occurrence of this event there is, by the construction of Lemma 2, a finite sequence of i^* arrivals, with probability, say, $\epsilon > 0$ and not depending on the value of Z, which results in an occurrence of $\mathbf{R}_{(c_1, c_2)}$. Then wp 1 there are infinitely many trials for $\mathbf{R}_{(c_1, c_2)}$, each with probability of success at least ϵ. $\mathbf{R}_{(c_1, c_2)}$ is certain.

To prove the sufficiency part of the second assertion, it will suffice to consider EM^c. For the finiteness of EX^c is then a consequence of the finiteness of EX and, for example, an easy adaptation of a standard martingale result (Doob [3], Theorem 2.2, pp. 302). Define $n'(j)$ to be the number of occurrences of $\mathbf{R}_{(c_1, c_2)}$ in

$t_1 + c_q, \cdots, t_j + c_q$ and $H'(j) = En'(j)$. By the construction of Lemma 2, if $d_1, d_2 < \infty$ there exist m, finite, and $\epsilon > 0$ such that

$$\Pr\{n'(j+m) > n'(j) \mid D_k(t_j) \leqq d_k, k = 1, 2\} \geqq \epsilon > 0 \quad \text{for all} \quad j.$$

Then

$$\epsilon \Pr\{D_k(t_j) \leqq d_k, k = 1, 2\} \leqq \Pr\{n'(j+m) > n'(j)\} \leqq H'(j+m) - H'(j). \tag{16}$$

In view of Feller's renewal theorem it will suffice to show that

$$\limsup_{j\to\infty} \Pr\{D_k(t_j) \leqq d_k, k = 1, 2\} > 0. \tag{17}$$

While (17) may be established directly rather easily, it will be convenient below to introduce the process

$$\hat{D}(T_n) = \max(\hat{S}_n, d_2), \hat{S}_n - X_n + \max(\hat{S}_{n-1}, d_2), \cdots, \hat{S}_n - \cdots - X_2 + \max(\hat{S}_1, d_2), \quad n \geqq 1. \tag{18}$$

It is evident that $\hat{D}$ dominates D wp 1. An argument like that used to establish (15) results in the conclusion that $\liminf_{n\to\infty} \Pr\{\hat{D}(T_n) \leqq d\} > 0$ for all $d \geqq d_2$. The event $\hat{D} = d_2$ is a regenerative event for this process. Then the expected number of occurrences of $\mathbf{R}$ between successive occurrences of $\hat{D} = d_2$ is finite, Feller's theorem. As $EM < \infty$ it follows from the above mentioned martingale result that the expected number of arrivals between successive events $\hat{D} = d_2$ is finite. A final application of Feller's theorem yields

$$\limsup_{j\to\infty} \sum_{k=1}^{\infty} \Pr\{\hat{D}(T_k) = d_2, T_k = t_j\} > 0.$$

As $\hat{D}(T_n) \geqq D(T_n) \geqq D_2(T_n)$ wp 1 (17) is established.

The proof that $\mathbf{R}_c$ is certain and $\mathrm{E}X^c$ and EM^c are finite under (1) and (11) is by induction for $q > 2$. We first suppose that the APS is such that a decision made in the kth queue, $k = 1, \cdots, q$, depends only on $[V \times (L_1, K_1, U_1) \times \cdots \times (L_k, K_k, U_k)]$. Thus the evolution of the first k queues is not affected by queues $k + 1, \cdots, q$. We assume that $\mathbf{R}_{(c_1,\cdots,c_k)}$ is a regenerative event for the first k queues, is certain, and that $EX^{(c_1,\cdots,c_k)}$ and $EM^{(c_1,\cdots,c_k)}$ are finite. If $\mathbf{R}_{(c_1,\cdots,c_k)}$ occurs for the jth and $j - 1$st times at $t_{n_2+c_k}$ and $t_{n_1+c_k}$ respectively, then $X^{(c_1,\cdots,c_k)} = t_{n_2} - t_{n_1}$ and $M^{(c_1,\cdots,c_k)} = n_2 - n_1$. Also, in this case define $T_j^k = t_{n_2}$. A dominating process may now be defined for queue $k + 1$. Let $\hat{S}_j = S_{k+1}^{n_1} + \cdots + S_{k+1}^{n_2-1}$. By the induction hypothesis the $\hat{S}_j$ are independent and, for $j > 1$, identically distributed. The arrival epochs for these composite items are $\{T_j^k + c_k\}$. Let D be defined for this process as above. Then $D_{k+1}(T_j^k) \leqq D(T_j^k + c_k) + c_k$ wp 1. For the arrivals to the dominating process through $T_j^k + c_k$ include all arrivals to the original process up to T_j^k. Also, it is clear that no arrival to the original process reaches the $k + 1$st queue after the composite item of which it is a component arrives. Finally, by the definition of

$\mathbf{R}_{(c_1,\cdots,c_k)}$ all items in queues 2, $\cdots$, k or being serviced by facilities 2, $\cdots$, k at T_j^k reach queue $k + 1$ not later than $T_j^k + c_k$. From this point the proof for $q = 2$ may be repeated.

When the APS does not have the special structure assumed above, the $\hat{S}_j$ of the dominating process may not be independent. However, it will be shown that $\mathbf{R}_{(c_1,\cdots,c_k)}$ can be suitably redefined so as to preserve this independence. It was observed that the occurrences of **R**, "an arriving item finds the first facility idle," are independent of the APS. Consider the process $\{\hat{D}(T_j),\ 1 \leqq j < \infty\}$ defined above. As this process is defined in terms of $X_j = T_j - T_{j-1}$ and $\hat{S}_j = S_2^{N_{j-1}} + \cdots + S_2^{N_j-1}$, the distributions of which are invariant with respect to the APS, it follows that the evolution of the $\hat{D}$ process is the same for every APS. Let $\hat{T}_j$ be the time of the jth occurrence of $\hat{D} = d_2$. It was shown that $E(\hat{T}_j - \hat{T}_{j-1}) < \infty$. Suppose $\hat{T}_j = t_p$. Then $D_1(t_p) = 0$ and $D_2(t_p) \leqq d_2$ for each APS. By Lemma 2 there is a sequence of, say, $n + m$ arrivals, with positive probability, which results in an occurrence of $\mathbf{R}_{(c_1,c)}$ at $t_{p+n} + c_2$; again, for all APS. If only occurrences of $\mathbf{R}_{(c_1,c_2)}$ which result from an occurrence of $\hat{D} = d_2$ followed by such a sequence of arrivals are counted, then $\mathbf{R}_{(c_1,c_2)}$ is a recurrent event for the first two queues. It follows trivially that this event is certain and that the expected number of arrivals between successive occurrences is finite. A dominating process may now be constructed for the third queue. It is concluded that

$$\limsup_{p\to\infty} \Pr\{D_i(t_p) \leqq d_i,\ i = 1, 2, 3\} > 0.$$

Also, a new $\hat{D}$ process results which gives rise to a regenerative event for the first three queues. After $q - 1$ such steps, a regenerative event with the asserted properties is obtained for the q queues. As the number of arrivals between its successive occurrences clearly dominates the number of arrivals between successive occurrences of $\mathbf{R}_c$, as originally defined, the assertion is established for general APS.

To prove the necessity of (1), suppose that there is equality in (1) for $k = k'$. Consider the single queue with arrival epochs $\{t_j\}$ and service times $\{S_{k'}^j\}$. Let $D'(t)$ be the uncompleted service time for this queue at time t, exclusive of any arrival at t. Then wp 1 $D_{k'}(t_j) \geqq D'(t_j)$. The necessity follows from E {first j with $D'(t_{n+j}) \leqq c_{k'} \mid D'(t_n)\} = \infty$ wp 1. At this point it is noted that Theorem 2 does not assert that $\mathbf{R}_c$ is certain if equality holds in (1) for some k. Evidently, a more delicate analysis is required.

With the mechanism provided by Theorem 2 limit theorems corresponding to 1–d and 1–c are easily obtained. Let $A' \varepsilon \mathcal{B}(\mathcal{A}')$ and $n(t_j)$ the number of occurrences of $\mathbf{R}_c$ in $(0, t_j]$. Set

$$\theta(A', j - i) = \Pr\{Z_{t_j} \varepsilon A' \mid n(t_j) \geqq 1, N_{n(t_j)} = i\}.$$

The existence of θ, and ϕ below, is a consequence of the common evolution of the process after each occurrence of $\mathbf{R}_c$. The representation (5) takes the form

$$P\{Z_{t_j} \varepsilon A'\} = \Pr\{Z_{t_j} \varepsilon A', n(t_j) = 0\}$$

$$(19) \qquad + \sum_{i=1}^{j} \theta(A', j - i) \Pr\{n(t_j) - n(t_i + c_q)$$

$$= 0 \mid \mathbf{R}_c \text{ occurs at } t_i + c_q\}[H'(i) - H'(i-1)].$$

Let $W_{i,j}$ denote the total waiting time in the q queues of the jth arrival conditional on its being a type i item. $W_{i,j}$ is clearly conditionally independent of the history of the process prior to t_j given Z_{t_j}. Let $F_{W_i}(\cdot : l_1, k_1, u_1, \cdots, u_q)$ be the corresponding wide sense conditional df: a Baire function of the l, k, and u.

THEOREM 3–d. *If* (1) *and* (11) *are satisfied, then for* $A' \varepsilon \mathcal{B}(\mathcal{A}')$ *and* $1 \leqq i \leqq r$,

$$(20) \quad \lim_{j\to\infty} P\{Z_{t_j} \varepsilon A'\} = (EM^c)^{-1} \sum_{p=1}^{\infty} \theta(A', p) \Pr\{n(t_p) - n(c_q)$$

$$= 0 \mid \mathbf{R}_c \text{ occurs at } t = 0\}$$

and

$$(21) \quad \lim_{j\to\infty} \Pr\{W_{i,j} < w\} = \sum_{k_1 \cdots k_q} \sum_{l_1 \cdots l_q} \int_0^\infty F_{W_i}(w : l_1, \cdots, u_q)$$

$$\cdot d_{u_1, \cdots, u_q} \sigma(l_1, \cdots, u_q),$$

where σ *is the limit* (20). *If* (1) *is not satisfied and* A' *is bounded in all* $l_{i,k}$, *the limit* (20) *is zero.*

PROOF. The first term of (19) has limit zero as $\mathbf{R}_c$ is certain. Also, it is evident that $\mathbf{R}_c$ has period 1. (20) follows from Theorem 2 and Theorem A once the sum is shown to converge. However, if θ is replaced by 1, the sum may be seen to equal EM^c. The proof of the remainder of the theorem is identical to the proof of the corresponding parts of Theorem 1–d.

THEOREM 3–c. *If* (1) *and* (11) *are satisfied, and if* X^c *is not a lattice variable*

$$(22) \quad \lim_{t\to\infty} P\{Z_t \varepsilon A\} = (EX^c)^{-1} \int_0^\infty \phi(A, x)[1 - F_X(x)]\, dx, \quad \text{for} \quad A \varepsilon \mathfrak{F}(\mathcal{A}),$$

if X^c *has period* λ *and* $0 \leqq \xi < \lambda$

$$(23) \quad \lim_{j\to\infty} P\{Z_{j\lambda+\xi} \varepsilon A\} = \lambda(EX^c)^{-1} \sum_{i=1}^{\infty} \phi(A, i\lambda + \xi)[1 - F_X(i\lambda + \xi)],$$

for $A \varepsilon \mathcal{B}(\mathcal{A})$. *If* (1) *is not satisfied and* A *is bounded in all* $l_{i,k}$, *the appropriate limit is zero.*

PROOF. The proof is identical in all details to that given for Theorem 1–c.

5. Extensions of the model. The event $\mathbf{R}_c$ remains a regenerative event and the conclusions of Theorem 2 are valid for several modifications in the model. Rather than make an attempt at being exhaustive, a few of the possibly more interesting ones will be examined briefly.

A serious limitation of the model is the requirement that a facility not idle when items are available for servicing. One would expect that in certain situations it would be desirable to allow a facility to idle when items are available in anticipation of some item yet to join the queue. However, if the restriction that the servicing of an item not be interrupted once started is relaxed, much of the objection to this restriction would seem overcome. Suppose that interruptions in the kth queue can only occur at completion epochs of the $k - 1$st facility, or at arrival epochs if $k = 1$, and that items of the same type do not interrupt each other. In this case, it may be seen that the proof of Theorem 2 remains valid so long as the interrupted item eventually resumes service where it was interrupted.

A second possibility involves a more complex priority structure. For simplicity suppose that τ has period 1 and that the nth arrival is assigned a due date, time by which it should be completed by the qth facility, of $t_n + Y_n$, where Y_n is a positive integer valued random variable. Suppose that the Y are independent, and for fixed item type identically distributed. If the nth arrival is a type i item and Y_n takes the value y, we shall say that a type (i, y) item arrived. At time $t_n + 1$ this item becomes a type $(i, y - 1)$ item. Thus, we have a countable class of item types. However, no modification in the proof of Theorem 2 is required and with a minor amount of additional work limit theorems for such functionals as the deviation of actual completion from due date are obtained.

The final extension to be considered here is relative to more general queue systems. The most general case being when the specification of an item type includes a subset of the facilities on which the item is to be serviced and the order in which the facilities are to be visited. While our model will not support this generality, it is easily seen that items may be allowed to skip facilities (and the associated queues). Suppose that (11) is satisfied by some i^* and that items of type i are not to be serviced at facility k. For the purpose of establishing Theorem 2, it is clear that the dominating process may be constructed as though type i items are in fact serviced at the kth facility ($S_{i,k} = 0$ wp 1). Next, suppose that there is no item type satisfying (11) which is serviced at each facility, but that item types i' and i'' satisfy (11) for the facilities at which they are serviced and that the union of these facilities is all q facilities. In this case two events $\mathbf{R}_{c'}$ and $\mathbf{R}_{c''}$, corresponding to item types i' and i'', are defined. Suppose that the facility k is the uth facility on which i' items are serviced, i' items are not serviced on facility $k + 1$, and facility $k + 1$ is the vth facility on which i'' items are serviced. If a regenerative event $\mathbf{R}_{(c_1', \cdots, c_u')}$ with the properties of Theorem 2 is available for the first u queues, one constructs a dominating process for the $k + 1$st queue as in the proof of Theorem 2. It then easily follows that a regenerative event $\mathbf{R}_{(c_1'', \cdots, c_v'')}$ may be defined for the first $k + 1$ queues, which also will have the properties of Theorem 2. Once it has been demonstrated that no queue grows unboundedly large wp 1, either of $\mathbf{R}_{c'}$ and $\mathbf{R}_{c''}$ may be used as the regenerative event in Theorems 3–d and 3–c.

6. Conclusions. We shall say that a system of q series queues has capacity for an imput $(\tau: \alpha_1, \cdots, \alpha_r: S_{i,k}, i = 1, \cdots, r \text{ and } k = 1, \cdots, q)$ with a specific APS if $\limsup_{t\to\infty} P\{Z_t \varepsilon A\} > 0$ for some bounded $A \varepsilon \mathcal{B}(\mathcal{A})$. It is of interest to know what role the APS actually plays in the determination of the capacity of a system of series queues. For inputs satisfying (1) and (11) the answer is none. However, as the condition (11) is not necessary, this is only a partial answer. Whether or not there is an input satisfying (1) for which a given system of series queues has sufficient capacity with one APS but not with some other APS is an open question.

7. Acknowledgments. It is a pleasure to acknowledge the helpful suggestions of Professor Peter Ney during the conduct of this research. Also, the author is indebted to the referee for pointing out some slip-ups and for several constructive suggestions.

REFERENCES

[1] BENEŠ, V. (1961). A renewal limit theorem for general stochastic processes. *Ann. Math. Statist.* **33** 93–113.
[2] BENEŠ, V. (1961). Extensions of Wiener's Tauberian theorem for positive measures. *J. Math. Anal. Appl.* **2** 1–20.
[3] BLACKWELL, D. (1953). A renewal theorem. *Duke Math. J.* **15** 145–150.
[4] DOOB, J. (1953). *Stochastic Processes.* Wiley, New York.
[5] FELLER, W. (1941). Fluctuation theory of recurrent events. *Ann. Math. Statist.* **12** 234–267.
[6] KENDALL, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of imbedded Markov chain. *Ann. Math. Statist.* **24** 338–354.
[7] KESTEN, H. and RUNNENBURG, J. TH. (1957). Priority in waiting-line problems. *Proc. Akad. Wet. Amst. A* **60** 161–200.
[8] LINDLEY, D. (1952). The theory of queues with a single server. *Proc. Cambridge Philos. Soc.* **48** 277–289.
[9] LOYNES, R. M. (1964). The stability of a system of queues in series. *Proc. Cambridge Philos. Soc.* **60** 569–574.
[10] SACKS, J. (1960). Ergodicity of queues in series. *Ann. Math. Statist.* **31** 569–583.
[11] SMITH, W. (1954). Asymptotic renewal theorems. *Proc. Roy. Soc. Edinburgh. Sect. A* **64** 9–48.
[12] SMITH, W. (1953). On the distribution of queuing times. *Proc. Cambridge Philos. Soc.* **49** 449–461.
[13] SMITH, W. (1958). Renewal theory and its ramifications. *J. Roy. Statist. Soc. Ser. B* **20** 243–302.
[14] SMITH, W. (1955). Regenerative stochastic processes. *Proc. Roy. Soc. London. Ser. A* **232** 6–31.