PRIORITY QUEUES[1]

By Rupert G. Miller, Jr.

Stanford University

1. Introduction and summary. In a priority queue different types of items (individuals or elements) arrive at a service mechanism and each item has a relative priority for order of service. Let there be K classes of items, $1, 2, \cdots, K$. If the service mechanism is to select an item for service, a type i item will be selected in preference to a type j item for $i < j$ even if the type j item arrived before the type i item, and within each class the "first come, first served" policy determines the order of service. When a type j item is in service and a type i item arrives $(i < j)$, there are two primary disciplines for handling the priority demand. The "head-of-the-line" discipline allows the type j item to complete service but places the type i item ahead of any other lower priority items. The "preemptive" discipline withdraws the type j item from service and replaces it by the type i item. Under the preemptive scheme the only time at which a type j item $(1 < j)$ can be in service is when there are no items of types $1, \cdots, j - 1$ in the queue. When a lower priority item which has been preempted returns to service, the preemptive discipline must distinguish two cases. The "preemptive resume" policy allows the preempted item to resume service at the point at which it was preempted so that its service time upon reentry has been reduced by the amount of time the item has already spent in service. The "preemptive repeat" policy requires the preempted item to commence service again at the beginning. A priority queue with an indifferent server is of course a special case of the preemptive resume discipline.

In the special case $K = 2$ the type 1 items will be referred to as priority items and the type 2 items as non-priority items.

It will be assumed throughout this paper that the input process for type i items, $i = 1, \cdots, K$, is Poisson with arrival rate λ_i and the input processes operate independently. The service time distribution for a type i item (in isolation) will be denoted by F_{S_i} and unless explicitly stated to the contrary will be assumed to be general subject only to the restrictions $F_{S_i}(0+) = 0$ and $E(S_i) < \infty$. Let $\rho_i = \lambda_i E(S_i)$, and let $\tilde{S}_i$ be the Laplace-Stieltjes transform of F_{S_i}. The service mechanism consists of a single channel or server.

A. Cobham [1], [2] introduced the head-of-the-line priority queue and derived equilibrium expected waiting times. Subsequent contributions have been made by Holley [3], Kesten and Runnenburg [4], [5], and Morse [6]. The first published results for the preemptive discipline were by H. White and L. S. Christie [7], and additional results have been presented by Stephan [8]. Koenigsberg [9] has

Received October 31, 1958; revised June 2, 1959.

[1] Technical Report No. 6, prepared under contract Nonr-225(28) for the Office of Naval Research, Reference No. NR-047-019.

generalized the priority model to a continuous number of priority types with application to machine breakdown problems.

Under various assumptions in this paper the following quantities have either been obtained explicitly or characterized as the unique (subject to regularity conditions) solution to a functional equation: the generating function for the stationary probabilities on the number of priority and non-priority items ($K = 2$) in the queue, the Laplace-Stieltjes transforms of the waiting time distributions, the Laplace-Stieltjes transform of the distribution of a busy period, and the generating function for the probabilities on the number of items serviced during a busy period. For most of the distributions mentioned the first two moments are computed.

2. Stationary distributions of the number of items in the queue ($K = 2$)**.** For completeness the results of White and Christie [7] and Morse [6] are summarized briefly.

For Poisson arrivals and exponential service the queue process is a continuous time parameter Markov process. If P is a stationary distribution of the queue process, it must be a solution to the forward steady state equations which, symbolically, can be represented as $PA = 0$ where A is the infinitesimal matrix of the process. If the system of equations $PA = 0$ has a unique solution (subject to the condition it be a probability distribution) and a stationary distribution is assumed to exist, then algebraic manipulation of the equations $PA = 0$ will yield a characterization of the stationary distribution or its generating function.

For the preemptive priority queue with $K = 2$ White and Christie employ this method to obtain the generating function and thereby the first and second moments. Let μ_1 and μ_2 denote the service rate parameters for the priority and non-priority items, respectively. Justification of a non-priority Poisson service process with the parameter μ_2 from assumptions on the non priority service process in isolation is discussed in detail in [7] with regard to the resume and repeat disciplines and the indifferent server queue.

If $\rho_1 + \rho_2 > 1$, the queue will become saturated with items and no stationary distribution will exist so the equilibrium condition $\rho_1 + \rho_2 < 1$ is assumed.

Let P_{nm} be the stationary probability that there are n priority and m non-priority items in the queue and $P(s, t) = \sum_{n,m} P_{n,m}s^n t^m$. That $P_{n,m}$ is uniquely defined is evident from inspection of the equations.

$$P(s, t) = \frac{(1 - \rho_1 - \rho_2)\mu_2(t^{-1} - 1)}{[\mu_1 \alpha(t) - \lambda_1 - \lambda_2(1 - t) - \mu_2(1 - t^{-1})][1 - \alpha(t)s]}, \tag{2.1}$$

where

$$\alpha(t) = \frac{\lambda_1 + \lambda_2(1 - t) + \mu_1 - \sqrt{(\lambda_1 + \lambda_2(1 - t) + \mu_1)^2 - 4\lambda_1\mu_1}}{2\mu_1}. \tag{2.2}$$

The moments of the number of priority items in the queue are the same as those for priority items in isolation; e.g.,

$$E(n) = \frac{\rho_1}{1 - \rho_1}; \qquad E(n^2) = \frac{2\rho_1^2}{(1 - \rho_1)^2} + \frac{\rho_1}{(1 - \rho_1)}. \tag{2.3}$$

The moments of the number of non-priority items in the queue can be evaluated from (2.1).

$$E(m) = \left[\frac{\rho_2}{1 - \rho_2 - \rho_2}\right]\left[\frac{\mu_2}{\mu_1}\left(\frac{\rho_1}{1 - \rho_1}\right) + 1\right];$$

$$E(m^2) = \frac{2\rho_1(\lambda_2/\mu_1)^2}{(1 - \rho_1)^3(1 - \rho_1 - \rho_2)} + \frac{\rho_2^2}{(1 - \rho_1 - \rho_2)^2}\left[1 + \frac{\mu_2}{\mu_1}\left(\frac{\rho_1}{1 - \rho_1}\right)\right]^2 + \frac{\rho_2(1 - \rho_1)^3 + \rho_1^2(\lambda_2/\mu_1)^2 + \rho_1(1 - \rho_1)(1 - \rho_1 + \rho_2)(\lambda_2/\mu_1)}{(1 - \rho_1)^2(1 - \rho_1 - \rho_2)^2}. \tag{2.4}$$

For a head-of-the-line priority queue with exponential service (μ_1, μ_2) Morse ([6], Ch. 9) has derived the generating function and first moments of the stationary probabilities through the same technique.

$$\begin{aligned} P(s, t) &= \frac{\mu_1(1 - s^{-1})(1 - \rho_1 - \rho_2)}{\lambda_1(1 - s) + \lambda_2(1 - t) + \mu_1(1 - s^{-1})} \\ &\quad + \frac{P_{20}(t)[\lambda_1 + \lambda_2(1 - t) + \mu_2][\mu_1(1 - s^{-1}) - \mu_2(1 - t^{-1})]}{[\lambda_1(1 - s) + \lambda_2(1 - t) + \mu_2][\lambda_1(1 - s) + \lambda_2(1 - t) + \mu_1(1 - s^{-1})]}, \end{aligned} \tag{2.5}$$

where

$$P_{20}(t) = \frac{t[1 - \rho_1 - \rho_2][\lambda_1 + \lambda_2(1 - t) - \mu_1\alpha(t)][(\mu_1 - \mu_2)(\lambda_1 + \lambda_2(1 - t) + \mu_2) - \lambda_1 \mu_1]}{[\lambda_1 + \lambda_2(1 - t) + \mu_2][\lambda_1 \mu_1 t - (\mu_1 - \mu_2)t(\lambda_1 + \lambda_2(1 - t) + \mu_2) + \mu_2(\mu_1(1 - \alpha(t)) - \mu_2)]}. \tag{2.6}$$

$$E(n) = \frac{\lambda_1}{\mu_1 - \lambda_1}\left[1 + \rho_2 \frac{\mu_1}{\mu_2}\right]. \tag{2.7}$$

$$E(m) = \rho_2 + \frac{\lambda_2}{\mu_1 - \lambda_1}\left[\frac{\rho_1 + \rho_2 \mu_1/\mu_2}{1 - \rho_1 - \rho_2}\right]. \tag{2.8}$$

The previous technique is not applicable for a head-of-the-line priority queue with Poisson arrivals but non-exponential service since the number of items in the queue no longer has the Markov property. The Markov property can be restored by reducing the continuous time parameter process to discrete time. This technique was introduced by D. G. Kendall ([10], [11]) and has been utilized by others (cf. [12], [13]). A discrete time Markov process is generated if the queue process is observed only at those points in time which are the termination points of a service period—priority or non-priority. The state of the queue is (n, m) where n is the number of priority and m the number of non-priority items in the queue (at the end of the service period). Since both priority and non-priority arrivals are Poisson the discrete time process has the Markov property.

The behavior of the Markov chain "imbedded" in the continuous time process is determined by the transition probability matrix which is expressible in terms of p_{ij} = probability that i priority and j non-priority items arrive during a priority service period and q_{ij} = probability that i priority and j non-priority items arrive during a non-priority service period.

$$p_{ij} = \int_0^\infty e^{-(\lambda_1+\lambda_2)t} \frac{(\lambda_1 t)^i}{i!} \frac{(\lambda_2 t)^j}{j!} dF_{S_1}(t)$$

so

$$P(s, t) = \sum_{i,j} p_{ij}s^it^j = \tilde{S}_1(\lambda_1(1 - s) + \lambda_2(1 - t)),$$

and

$$q_{ij} = \int_0^\infty e^{-(\lambda_1+\lambda_2)t} \frac{(\lambda_1 t)^i}{i!} \frac{(\lambda_2 t)^j}{j!} dF_{S_2}(t)$$

so

$$Q(s, t) = \sum_{i,j} q_{ij}s^it^j = \tilde{S}_2(\lambda_1(1 - s) + \lambda_2(1 - t)).$$

Let $P\{(n, m) \rightarrow (n', m')\}$ be the probability the queue moves from state (n, m) to state (n', m') in one transition. In terms of the p_{ij}, q_{ij}, the

$$P\{(n, m) \rightarrow (n', m')\}$$

are

$$\begin{aligned}
&(1)\ P\{(n,m) \rightarrow (n',m')\} = 0 \text{ for } n' < n - 1, n > 1, \text{all } m, m',\\
&(2)\ P\{(n,m) \rightarrow (n',m')\} = 0 \text{ for } m' < m, n \geqq 1, \text{all } n',\\
(2.9)\quad &(3)\ P\{(n,m) \rightarrow (n - 1 + i, m + j)\} = p_{ij} \text{ for } i,j \geqq 0, n \geqq 1, \text{all } m,\\
&(4)\ P\{(0,m) \rightarrow (n,m')\} = 0 \text{ for } m' < m - 1, \text{all } n,\\
&(5)\ P\{(0,m) \rightarrow (i, m - 1 + j)\} = q_{ij} \text{ for } i,j \geqq 0, m > 0,\\
&(6)\ P\{(0,0) \rightarrow (i,j)\} = r_1p_{ij} + r_2q_{ij} \text{ for } i,j \geqq 0,
\end{aligned}$$

where $r_1 = \lambda_1/(\lambda_1 + \lambda_2)$ and $r_2 = \lambda_2/(\lambda_1 + \lambda_2)$. The transition probabilities under (6) have their special form because if the state is (0, 0) the queue is next observed at the end of the service period for the first arrival so the probability of the new state (n', m') depends on which type of item was first to arrive.

The state (0, 0) is ergodic if the equilibrium condition $\rho_1 + \rho_2 < 1$ is satisfied. Since the proof of this is analogous to those in [13] and [14] for different queues, it will be omitted. The ergodicity of the state (0, 0) guarantees the existence of the stationary distribution for the imbedded Markov chain.

The stationary probability of there being n priority, m non-priority items in the queue will be denoted by π_{nm}. By definition, the stationary distribution $\pi = \{\pi_{nm}\}$ must satisfy the system of equations

$$\pi_{n'm'} = \sum_{n,m} \pi_{nm}P\{(n, m) \rightarrow (n', m')\}, \qquad \text{all } n', m'. \tag{2.10}$$

From (2.10)

$$\pi(s, t) = \sum_{n',m'} \sum_{n,m} \pi_{nm} P\{(n, m) \rightarrow (n', m')\} s^{n'} t^{m'} \tag{2.11}$$

which, when simplified, gives the following expression for $\pi(s, t)$:

$$\begin{aligned} \pi(s, t) = {} & [\pi_{00}(\tau_1 P(s, t) + \tau_2 Q(s, t) - t^{-1}Q(s, t)) \\ & + \pi_0(t)(t^{-1}Q(s, t) - s^{-1}P(s, t))][1 - s^{-1}P(s, t)]^{-1}, \end{aligned} \tag{2.12}$$

where $\pi_0(t) = \sum_m \pi_{0m} t^m$ is analogous to $P_{20}(t)$ of (2.5).

To determine $\pi(s, t)$ it is necessary to determine $\pi_0(t)$. This can be accomplished by imbedding a second Markov chain within the original imbedded Markov chain. The second Markov chain is defined by taking cognizance of the state of the process only at those time points which are the termination points of a service period leaving 0 priority items in the queue. The state of the queue is (m), the number of non-priority items in the queue. This Markov chain is imbedded within the original chain since a trial for the second chain occurs at the end of a service period only if there are 0 priority items left in line whereas, previously, the termination of any service period constituted a trial.

Let $P\{m \rightarrow m'\}$ denote the probability of moving from state m to state m' in one transition. For $m > 0$, $j \geqq 0$,

$$\begin{aligned} P\{m \rightarrow m - 1 + j\} &= P\{m \rightarrow m - 1 + j, \text{ no priority arrivals in interim}\} \\ & \quad + P\{m \rightarrow m - 1 + j, \text{ priority arrivals in interim}\} \\ &= q_{0j} + \sum_{l=0}^{j} \sum_{n=1}^{\infty} \left[\int_0^\infty e^{-(\lambda_1+\lambda_2)u} \frac{(\lambda_1 u)^n}{n!} \frac{(\lambda_2 u)^l}{l!} \, dF_{S_2}(u) \right] \\ & \qquad \cdot \left[\int_0^\infty e^{-\lambda_2 v} \frac{(\lambda_2 v)^{j-l}}{(j-l)!} \, dF_B^{(n)}(v) \right]. \end{aligned}$$

F_B is the distribution of the busy period (see [15] or Section 4) for priority items in isolation and $F_B^{(n)}$ is its n-fold convolution. If

$$P(t) = \sum_{j=0}^{\infty} P\{m \rightarrow m - 1 + j\} t^j$$

for $m > 0$, it is readily verified that

$$P(t) = \tilde{S}_2(\lambda_1(1 - \tilde{B}(\lambda_2(1 - t))) + \lambda_2(1 - t)) \tag{2.13}$$

where $\tilde{B}$ is the Laplace-Stieltjes transform of F_B.

For $m = 0$, $j \geqq 0$,

$$\begin{aligned} P\{0 \rightarrow j\} &= P\{0 \rightarrow j, \text{ first arrival is priority}\} \\ & \quad + P\{0 \rightarrow j, \text{ first arrival is non-priority}\} = \tau_1 \int_0^\infty e^{-\lambda_2 u} \frac{(\lambda_2 u)^j}{j!} \, dF_B(u) \end{aligned}$$

$$+ \tau_2 \sum_{l=0}^{j} \sum_{n=0}^{\infty} \left[\int_0^\infty e^{-(\lambda_1+\lambda_2)u} \frac{(\lambda_1 u)^n}{n!} \frac{(\lambda_2 u)^l}{l!} \, dF_{S_2}(u) \right] \cdot \left[\int_0^\infty e^{-\lambda_2 v} \frac{(\lambda_2 v)^{j-l}}{(j-l)!} \, dF_B^{(n)}(v) \right],$$

where $F^{(0)}$ is the distribution which concentrates its total mass at 0. If

$$Q(t) = \sum_{j=0}^{\infty} P\{0 \to j\} t^j,$$

then

$$(2.14) \quad Q(t) = \tau_1 \tilde{B}(\lambda_2(1-t)) + \tau_2 \tilde{S}_2(\lambda_1(1 - \tilde{B}(\lambda_2(1-t))) + \lambda_2(1-t)).$$

Let π_m^0 be the stationary probability of there being m non-priority items in the queue (for the second imbedded Markov chain). Algebraic manipulation of the system of equations

$$(2.15) \qquad \pi_{m'}^0 = \sum_m \pi_m^0 P\{m \to m'\}, \qquad \text{all } m'$$

yields the following expression for $\pi^0(t) = \sum_{m=0}^{\infty} \pi_m^0 t^m$:

$$(2.16) \qquad \pi^0(t) = \pi_0^0[Q(t) - t^{-1}P(t)][1 - t^{-1}P(t)]^{-1}.$$

π_0^0 determines the normalization for the distribution π^0. If the second Markov chain is to be viewed as imbedded within the first, the proper normalization is $\pi_0^0 = \pi_{00}$ which implies $\pi_0^0 = \pi_{0m}$ for all m. Hence,

$$(2.17) \qquad \pi_0(t) = \pi_{00}[Q(t) - t^{-1}P(t)][1 - t^{-1}P(t)]^{-1}.$$

In conjunction with (2.12), (2.17) yields

$$(2.18) \qquad \begin{aligned} \pi(s, t) = {} & \pi_{00}[1 - s^{-1}P(s, t)]^{-1}\{\tau_1 P(s, t) + \tau_2 Q(s, t) - t^{-1}Q(s, t) \\ & + (t^{-1}Q(s, t) - s^{-1}P(s, t))[1 - t^{-1}P(t)]^{-1}[Q(t) - t^{-1}P(t)]\}. \end{aligned}$$

π_{00} is determined by the restraint $\pi(1, 1) = 1$; $\pi_{00} = 1 - \rho_1 - \rho_2$.

The first moments of n and m (and also higher moments) can be calculated from (2.18).

$$(2.19) \qquad E(n) = \tau_1(\rho_1 + \rho_2) + \frac{\lambda_1^2[\tau_1 E(S_1^2) + \tau_2 E(S_2^2)]}{2(1 - \rho_1)}.$$

$$(2.20) \qquad \begin{aligned} E(m) = {} & \tau_2(\rho_1 + \rho_2) \\ & + \frac{\lambda_2[\tau_1 E(S_1^2) + \tau_2 E(S_2^2)]}{2\mu_1} \left[\frac{\lambda_2(\mu_1 + \mu_2) + \rho_1(1 - \rho_1 - \rho_2)}{(1 - \rho_1 - \rho_2)(1 - \rho_1)} \right]. \end{aligned}$$

From (2.18)–(2.20) it is apparent that (2.18) does not agree with (2.5) when the service times are exponentially distributed. As more complex queues are studied, it becomes clear that the stationary distributions for the imbedded Markov chain and general time t are identical only for the simpler queues. For example, a similar discrepancy is noted in [13].

It might be hoped to duplicate this analysis for the preemptive priority queue. However, there does not exist a natural imbedding procedure for the preemptive queue. The only method of avoiding incorporation of an additional time quantity into the definition of the state would be to observe the process just at the termination of service of a non-priority item. But this is a one-dimensional queue and is of no significance.

For those one-dimensional queues in which the arrival distribution is general and the service exponential the natural imbedding considers those times at which a new arrival enters the queue. For a priority queue with general arrival distributions this imbedding pattern leads to a non-Markov process unless the time to the last arrival of the other type item is incorporated into the definition of the state. The addition of this extra time variable prohibits any simple analysis.

3. Waiting time distributions. The waiting time of an item is defined to be the length of time the item must wait in the queue before it is taken into service. The time an item spends in service is not included in the waiting time. For a priority queue with head-of-the-line discipline the time in service of an item is just the length of its service period, but for the preemptive discipline the term "time in service" will mean the total time from the moment the item first enters service to the moment it completes service including those periods of time in which it is waiting for reentry into service after having been preempted.

The equilibrium condition $1 - \rho_1 - \cdots - \rho_K > 0$ will be assumed throughout this section so that it is meaningful to discuss stationary distributions. The discussion for general time t also applies to the transient case.

The method introduced by D. G. Kendall for the single class queue can be applied to derive the Laplace-Stieltjes transform of the steady state waiting time distribution for a priority item in a head-of-the-line priority queue ($K = 2$). Suppose an item has just completed service. Since the queue is assumed to be operating in a state of equilibrium, with probability τ_1 the item was a priority item and with probability τ_2 the item was non-priority. If the item was a priority item, the number of priority items remaining in the queue must be the number which arrived during its waiting time and service period. If the item was non-priority, the number of priority items in the queue is just the number which arrived during its service period. The probability there are n priority items remaining in the queue is $\sum_{m=0}^{\infty} \pi_{nm}$ so

$$\begin{aligned}\sum_{m=0}^{\infty} \pi_{nm} = \tau_1 \int_0^{\infty} e^{-\lambda_1 t} \frac{(\lambda_1 t)^n}{n!} \, dF_{W_1} * F_{S_1}(t) \\ + \tau_2 \int_0^{\infty} e^{-\lambda_1 t} \frac{(\lambda_1 t)^n}{n!} \, dF_{S_2}(t),\end{aligned} \tag{3.1}$$

where F_{W_1} is the waiting time distribution for a priority item and $F_{W_1} * F_{S_1}$ denotes the convolution of F_{W_1} and F_{S_1}. From (3.1)

$$\tilde{W}_1(s) = \frac{\pi((\lambda_1 - s)/\lambda_1, 1) - \tau_2 \tilde{S}_2(s)}{\tau_1 \tilde{S}_1(s)}, \tag{3.2}$$

where $\tilde{W}_1$ is the Laplace-Stieltjes transform of F_{W_1}. Substitution of the value of $\pi((\lambda_1 - s)/\lambda_1, 1)$ as given by (2.18) gives

$$\tilde{W}_1(s) = \frac{-(1 - \rho_1 - \rho_2)s - \lambda_2[1 - \tilde{S}_2(s)]}{\lambda_1 - s - \lambda_1 \tilde{S}_1(s)}. \tag{3.3}$$

The moments of W_1 can be calculated from (3.3). In particular,

$$E(W_1) = \frac{\lambda_1 E(S_1^2) + \lambda_2 E(S_2^2)}{2(1 - \rho_1)}; \tag{3.4}$$

$$E(W_1^2) = \frac{\lambda_1 E(S_1^3) + \lambda_2 E(S_2^3)}{3(1 - \rho_1)} + \frac{\lambda_1 E(S_1^2)[\lambda_1 E(S_1^2) + \lambda_2 E(S_2^2)]}{2(1 - \rho_1)^2}. \tag{3.5}$$

The transform of the non-priority waiting time distribution can be obtained by adaptation of the results of L. Takács [15]. For a simple, single class queue with Poisson arrivals (λ) and general service distribution F_S Takács established that the Laplace-Stieltjes transform $\tilde{W}(s; t)$ of the waiting time distribution at time t is given by

$$\tilde{W}(s; t) = e^{t[s-\lambda(1-\tilde{S}(s))]}\left[1 - s\int_0^t e^{-u[s-\lambda(1-\tilde{S}(s))]}F_W(0^+; u)\,du\right], \tag{3.6}$$

where $\tilde{S}(s)$ is the Laplace-Stieltjes transform of F_S and $F_W(0+; u)$ is the probability the queue is empty at time u. The Laplace transform of $F_W(0^+; u)$ is related to $\tilde{B}(s)$, the Laplace-Stieltjes transform of the busy period distribution, by

$$\int_0^\infty e^{-su}F_W(0^+; u)\,du = \frac{1}{s + \lambda(1 - \tilde{B}(s))}. \tag{3.7}$$

The transform of the steady state waiting time distribution is

$$\tilde{W}(s) = \lim_{t\to\infty} \tilde{W}(s; t) = \frac{1 - \lambda E(S)}{1 - \lambda(1 - \tilde{S}(s))/s}. \tag{3.8}$$

The waiting time of a non-priority item is the sum of two waiting times, W_2^* and W_2^{**}. W_2^* is the time required to service all priority and non-priority items already in the queue at the arrival of the non-priority item, and W_2^{**} is the time consumed in servicing all subsequent priority arrivals which precede the entrance into service of the non-priority item. As far as the waiting time of the non-priority item is concerned, the following queue discipline could be in effect at its arrival. Service all priority and non-priority items in the queue ahead of the non-priority item at its arrival. Any priority arrivals occurring during this time interval are refused service until the items initially in the queue have been serviced—even if this means servicing a non-priority item in preference to a priority item. After the initial group has been serviced, commence service on the by-passed priority items and continue service until the queue has been emptied of priority items. At this moment the non-priority item whose waiting time is in question may enter service. W_2^{**} is defined to be the service time for the by-passed priority items.

If n priority items arrive during the W_2^* units of time, the distribution of W_2^{**} is the same as the distribution of B_n where B_n is the length of a busy period (see [15] and Section 4) for a one-dimensional queue with only priority items in which there are n priority items initially. Thus,

$$P\{W_2 \leqq x\} = \int_0^x \left[\sum_{n=0}^{\infty} e^{-\lambda_1 y} \frac{(\lambda_1 y)^n}{n!} P\{B_n \leqq x - y\} \right] dP\{W_2^* \leqq y\} , \tag{3.9}$$

and

$$\tilde{W}_2(s) = \tilde{W}_2^*(s + \lambda_1(1 - \tilde{B}_1(s))). \tag{3.10}$$

$\tilde{W}_2^*$ is obtainable from (3.8) with the identifications $\lambda = \Lambda_2 = \lambda_1 + \lambda_2$ and $\tilde{S}(s) = \tilde{S}_2^*(s) = \tau_1 \tilde{S}_1(s) + \tau_2 \tilde{S}_2(s)$. To a non-priority item arriving at the queue the distinction between previously arrived priority and non-priority items is immaterial. All are serviced ahead of the non-priority item and could just as well be viewed as having the average service time distribution $\tilde{S}_2^*(s)$. Hence,

$$\tilde{W}_2(s) = \frac{1 - \rho_1 - \rho_2}{1 - \dfrac{\sum_{i=1}^{2} \lambda_i [1 - \tilde{S}_i(s + \lambda_1(1 - \tilde{B}_1(s)))]}{s + \lambda_1(1 - \tilde{B}_1(s))}} . \tag{3.11}$$

Moments of W_2 can be determined from (3.11) and the results for $\tilde{B}_1(s)$ in [15] or the next section.

$$E(W_2) = \frac{\lambda_1 E(S_1^2) + \lambda_2 E(S_2^2)}{2(1 - \rho_1)(1 - \rho_1 - \rho_2)} ; \tag{3.12}$$

$$\begin{aligned} E(W_2^2) = {} & \frac{\lambda_1 E(S_1^3) + \lambda_2 E(S_2^3)}{3(1 - \rho_1)^2(1 - \rho_1 - \rho_2)} + \frac{[\lambda_1 E(S_1^2) + \lambda_2 E(S_2^2)]^2}{2(1 - \rho_1)^2(1 - \rho_1 - \rho_2)^2} \\ & + \frac{\lambda_1 E(S_1^2)[\lambda_1 E(S_1^2) + \lambda_2 E(S_2^2)]}{2(1 - \rho_1)^3(1 - \rho_1 - \rho_2)} . \end{aligned} \tag{3.13}$$

Kesten and Runnenburg ([4], [5]) have obtained an alternative characterization of $\tilde{W}_1$ and $\tilde{W}_2$. The first two moments as computed by their method agree with (3.4)-(3.5) and (3.12)-(3.13). In addition, Kesten and Runnenburg have derived a characterization for the transform of the steady state waiting time distribution of any type j item for a priority queue with general K.

The method employed to characterize $\tilde{W}_2(s)$ above can be extended to characterize the waiting time transform of the lowest priority, type K item for arbitrary time t and in equilibrium. Let $W_K(t)$ denote the waiting time for a lowest priority item if it were to arrive at time t. $W_K(t)$ is the sum of two components, $W_K^*(t)$ and $W_K^{**}(t)$, which are defined analogously to W_2^* and W_2^{**}. The same argument verifies that

$$\begin{aligned} P\{W_K(t) \leqq x\} = {} & \int_0^x \left[\sum_{n_1, \cdots, n_{K-1}} e^{-(\lambda_1 + \cdots + \lambda_{K-1}) y} \frac{(\lambda_1 y)^{n_1}}{n_1!} \right. \\ & \left. \cdot \cdots \frac{(\lambda_{K-1} y)^{n_{K-1}}}{n_{K-1}!} \cdot P\{B_{K-1:n_1 \cdots n_{K-1}} \leqq x - y\} \right] dP\{W_K^*(t) \leqq y\} , \end{aligned} \tag{3.14}$$

where $B_{K-1:n_1\cdots n_{K-1}}$ is the length of a busy period for a priority queue with just $K-1$ types $1, \cdots, K-1$ which commences with n_i type i items, $i = 1, \cdots, K-1$, in line initially (see Section 4). In terms of Laplace-Stieltjes transforms (3.14) becomes

$$(3.15)\qquad \tilde{W}_K(s;t) = \tilde{W}_K^*(s + \Lambda_{K-1}(1 - \tilde{B}_{K-1}^*(s));t),$$

where $\tilde{B}_{K-1,i}(s)$ is the transform of the busy period distribution for the $K-1$ dimensional priority queue which commences with a single type i item in line and $\tilde{B}_{K-1}^*(s) = \sum_{i=1}^{K-1} \tau_i \tilde{B}_{K-1,i}(s)$. $\Lambda_{K-1} = \lambda_1 + \cdots + \lambda_{K-1}$ and $\tau_i = \lambda_i/\Lambda_{K-1}$, $i = 1, \cdots, K-1$.

$\tilde{B}_{K-1,i}(s)$ and $B_{K-1}^*(s)$ will be characterized in the next section, and $\tilde{W}_K^*(s;t)$ can be obtained from (3.6) with the identifications

$$(3.16)\qquad \begin{aligned} \lambda &= \Lambda_K = \lambda_1 + \cdots + \lambda_K, \\ \tilde{S}(s) &= \tilde{S}_K^*(s) = \tau_1\tilde{S}_1(s) + \cdots + \tau_K\tilde{S}_K(s), \\ \tilde{B}(s) &= \tilde{B}_K^*(s) = \tau_1\tilde{B}_{K1}(s) + \cdots + \tau_K\tilde{B}_{KK}(s). \end{aligned}$$

In the limit as $t \to \infty$

$$(3.17)\qquad \tilde{W}_K(s) = \lim_{t\to\infty} \tilde{W}_K(s;t) = \tilde{W}_K^*(s + \Lambda_{K-1}(1 - \tilde{B}_{K-1}^*(s))),$$

where

$$(3.18)\qquad \tilde{W}_K^*(\alpha) = \frac{1 - \rho_1 - \cdots - \rho_K}{1 - \dfrac{\sum_{i=1}^{K} \lambda_i[1 - \tilde{S}_i(\alpha)]}{\alpha}}.$$

Moments of the steady state waiting time can be computed from (3.17)-(3.18) and the results of the next section.

$$(3.19)\qquad E(W_K) = \frac{\sum_1^K \lambda_i E(S_i^2)}{2\left(1 - \sum_1^{K-1} \rho_i\right)\left(1 - \sum_1^K \rho_i\right)};$$

$$(3.20)\qquad \begin{aligned} E(W_K^2) = {}& \frac{\sum_1^K \lambda_i E(S_i^3)}{3\left(1 - \sum_1^{K-1} \rho_i\right)^2\left(1 - \sum_1^K \rho_i\right)} \\ &+ \frac{\left[\sum_1^K \lambda_i E(S_i^2)\right]^2}{2\left(1 - \sum_1^{K-1} \rho_i\right)^2\left(1 - \sum_1^K \rho_i\right)^2} + \frac{\left[\sum_1^K \lambda_i E(S_i^2)\right]\left[\sum_1^{K-1} \lambda_i E(S_i^2)\right]}{2\left(1 - \sum_1^{K-1} \rho_i\right)^3\left(1 - \sum_1^K \rho_i\right)}. \end{aligned}$$

This technique can also be applied to the preemptive "resume" priority queue to characterize the waiting time distribution for any type item at general time

t and in equilibrium. Let there be K priority classes and $W_j(t)$ be the waiting time for an item in the jth class if it were to arrive at time t. The distribution of $W_1(t)$ is the same as for type one items in isolation since priority items preempt any lower class items in service. The waiting time $W_j(t)$, $j > 1$, consists of two components, $W_j^*(t)$ and $W_j^{**}(t)$. $W_j^*(t)$ is the time required to service all items of priority $\leqq j$ which are in the queue at time t, and its Laplace-Stieltjes transform is given by (3.6) and (3.7) with $\lambda = \Lambda_j$, $\tilde{S}(s) = \tilde{S}_j^*(s)$, and

$$\tilde{B}(s) = \tilde{B}_j^*(s) = \tau_1 \tilde{B}_{j1}(s) + \cdots + \tau_j \tilde{B}_{jj}(s).$$

$\tilde{B}_{ji}(s)$ and $\tilde{B}_j^*(s)$ for the preemptive resume discipline will be characterized in Section 4. The distribution of $W_j^*(t)$ is unaffected by the presence of lower priority items because of the preemptive discipline, and since an item "resumes" service after preemption, the priority discipline among the items of types $1, \cdots, j$ could just as well be abandoned as far as the distribution of $W_j^*(t)$ is concerned. $W_j^{**}(t)$ is the time required to service all arrivals of priority $<j$ which arrive after t but before the type j item can enter service, and it is given by a convolution of busy periods $B_{j-1,i}$, $i = 1, \cdots, j-1$, where the degree of the convolution is determined by the number of arrivals in the time interval $(0, W_j^*(t))$. Hence,

$$(3.21)\quad \begin{aligned} P\{W_j(t) \leqq x\} = \int_0^x \Bigg[\sum_{n_1,\cdots,n_{j-1}} e^{-\Lambda_{j-1}y} \frac{(\lambda_1 y)^{n_1}}{n_1!} \cdots \frac{(\lambda_{j-1} y)^{n_{j-1}}}{n_{j-1}!} \\ \cdot P\{B_{j-1:n_1\cdots n_{j-1}} \leqq x - y\} \Bigg] dP\{W_j^*(t) \leqq y\}, \end{aligned}$$

and

$$(3.22)\qquad \tilde{W}_j(s;t) = \tilde{W}_j^*(s + \Lambda_{j-1}(1 - \tilde{B}_{j-1}^*(s));t).$$

For the stationary case

$$(3.23)\qquad \tilde{W}_j(s) = \lim_{t\to\infty} \tilde{W}_j(s;t) = \tilde{W}_j^*(s + \Lambda_{j-1}(1 - \tilde{B}_{j-1}^*(s))),$$

where

$$(3.24)\qquad \tilde{W}_j^*(\alpha) = \frac{1 - \Lambda_j E(S_j^*)}{1 - \Lambda_j\left(\dfrac{1 - \tilde{S}_j^*(\alpha)}{\alpha}\right)}.$$

The first two moments of W_j are given by (3.19) and (3.20) with K replaced by j.

The quantity T_j, the "time in service" of a type j priority item, is S_j only for $j = 1$ under the preemptive resume discipline. For $j > 1$

$$(3.25)\quad \begin{aligned} P\{T_j \leqq x\} = \int_0^x \Bigg[\sum_{n_1,\cdots,n_{j-1}} e^{-\Lambda_{j-1}y} \frac{(\lambda_1 y)^{n_1}}{n_1!} \cdots \frac{(\lambda_{j-1} y)^{n_{j-1}}}{n_{j-1}!} \\ \cdot P\{B_{j-1:n_1\cdots n_{j-1}} \leqq x - y\} \Bigg] dF_{S_j}(y) \end{aligned}$$

so

$$\tilde{T}_j(s) = \tilde{S}_j(s + \Lambda_{j-1}(1 - \tilde{B}^*_{j-1}(s))). \tag{3.26}$$

The first two moments of T_j are

$$E(T_j) = \frac{E(S_j)}{1 - \sum_{i=1}^{j-1} \rho_i}, \tag{3.27}$$

$$E(T_j^2) = \frac{E(S_j^2)}{\left[1 - \sum_{i=1}^{j-1} \rho_i\right]^2} + \frac{E(S_j)[\lambda_1 E(S_1^2) + \cdots + \lambda_{j-1} E(S_{j-1}^2)]}{\left[1 - \sum_{i=1}^{j-1} \rho_i\right]^3}. \tag{3.28}$$

The preemptive priority queue with indifferent server is a special case of the preemptive resume priority queue so the above results apply as well to the indifferent server queue. The waiting time questions for the preemptive repeat priority queue are for the most part still unsolved.

4. Busy period distributions. A queue is said to be "busy" or "empty" depending upon whether or not there is an item in service. The length of a busy period is the length of time between the arrival of an item at the empty queue and the first subsequent moment at which the queue is again empty. The technique which will be used to characterize the busy period is an adaptation of that introduced by Takács [15] to solve the busy period problem for the simple queue with a single class, Poisson arrivals (λ), and a general service distribution F_S. Takács established that $\tilde{B}(s)$, the Laplace-Stieltjes transform of the busy period distribution, satisfies the functional equation

$$f(s) = \tilde{S}(s + \lambda(1 - f(s))) \tag{4.1}$$

and is in fact the unique solution to (4.1) which satisfies in addition

$$\text{(i) } f(s) \text{ analytic for } \operatorname{Re}\{s\} > 0, \qquad \text{(ii) } \lim_{\substack{s \to \infty \\ s \text{ real}}} f(s) = 0. \tag{4.2}$$

Consider a priority queue with K priority classes and head-of-the-line discipline. B_{Ki} is the length of a busy period which commences with the arrival of a type i item, $i = 1, \cdots, K$. $F_{B_{Ki}}$ will denote the distribution function of B_{Ki} and $\tilde{B}_{Ki}(s)$ the corresponding Laplace-Stieltjes transform. B_K^* is the average busy period in which the priority class of the initial arrival is not specified. $F_{B_K^*} = \tau_1 F_{B_{K1}} + \cdots + \tau_K F_{B_{KK}}$, and $\tilde{B}_K^*(s) = \tau_1 \tilde{B}_{K1}(s) + \cdots + \tau_K \tilde{B}_{KK}(s)$.

The equilibrium condition $1 - \rho_1 - \cdots - \rho_K > 0$ will be assumed so that $F_{B_{Ki}}$ is a bona fide distribution. With modification the discussion applies as well to the transient case.

Arrivals at the queue constitute a Poisson process with parameter Λ_K, and given that an arrival has occurred the probability it belongs to priority class j is τ_j. At the end of the service period of the initial arrival there will be $n_1, \cdots, n_K$ items of types $1, \cdots, K$, respectively, in the queue. The busy period will be prolonged by the amount $B_{K:n_1\cdots n_K}$ which denotes a busy period commencing with $n_1, \cdots, n_K$ items of types $1, \cdots, K$, respectively, in line initially. However, the distribution of $B_{K:n_1\cdots n_K}$ is just a convolution of the distributions $F_{B_{K1}}^{(n_1)}, \cdots, F_{B_{KK}}^{(n_K)}$ where $F_{B_{Ki}}^{(n_i)}$ denotes an n_i-fold convolution of $F_{B_{Ki}}$. Hence,

$$P\{B_K^* \leqq x\} = \int_0^x \Bigg[\sum_{n=0}^{\infty} e^{-\Lambda_K y} \frac{(\Lambda_K y)^n}{n!} \sum_{n_1,\cdots,n_K} \frac{n!}{n_1!\cdots n_K!} \cdot (\tau_1)^{n_1} \cdots (\tau_K)^{n_K} P\{B_{K:n_1\cdots n_K} \leqq x - y\} \Bigg] dF_{S_K^*}(y), \tag{4.3}$$

and

$$\tilde{B}_K^*(s) = \tilde{S}_K^*(s + \Lambda_K(1 - \tilde{B}_K^*(s))), \tag{4.4}$$

where $F_{S_K^*} = \tau_1 F_{S_1} + \cdots + \tau_K F_{S_K}$ and $\tilde{S}_K^*$ is its transform.

$\tilde{B}_K^*(s)$ is the unique solution to the functional equation

$$f(s) = \tilde{S}_K^*(s + \Lambda_K(1 - f(s))) \tag{4.5}$$

which satisfies the regularity conditions (4.2). The proof of this assertion can be obtained from the proof for the simple queue [15] with the appropriate identifications, but an alternative proof is presented below. This proof represents a simpler proof of the result for the single class queue.

It is sufficient to show that (4.5) and (ii) determine $f(s)$ uniquely for real $s > 0$ since by (i) this determines $f(s)$ in the whole half-plane. Suppose there exist two functions $f_1(s)$ and $f_2(s)$ which satisfy (4.5), (i), (ii) but $f_1(s) \not\equiv f_2(s)$ for real $s > 0$. Let $s_0 > 0$ be a point for which $f_1(s_0) > f_2(s_0) > 0$. Since $\lim_{s\to\infty} f_1(s) = 0$ and f_1 is continuous, there must exist an $s_1 > s_0$ such that $f_1(s_1) = f_2(s_0) = c$. But this implies that there exist two different values, namely s_0 and s_1, which satisfy $c = \tilde{S}_K^*(s + \Lambda_K(1 - c))$ which is impossible since the right-hand side is a strictly decreasing function of s.

This proof can be extended to characterize $F_{B_K^*}(+\infty)$ in the transient case. Moments of B_K^* can be computed from (4.4).

$$E(B_K^*) = \frac{E(S_K^*)}{1 - \Lambda_K E(S_K^*)} = \frac{\tau_1 E(S_1) + \cdots + \tau_K E(S_K)}{1 - \rho_1 - \cdots - \rho_K}; \tag{4.6}$$

$$E(B_K^{*^2}) = \frac{E(S_K^{*^2})}{(1 - \Lambda_K E(S_K^*))^3} = \frac{\tau_1 E(S_1^2) + \cdots + \tau_K E(S_K^2)}{(1 - \rho_1 - \cdots - \rho_K)^3}. \tag{4.7}$$

The characterization of $\tilde{B}_{Ki}$ will be constructed recursively. Assume that $F_{B_{K-1,i}}$, $i = 1, \cdots, K - 1$, and the corresponding $\tilde{B}_{K-1,i}$ have been determined. As far as the distribution of the busy period is concerned, the priority discipline can be disregarded. If the busy period commences with the arrival of a type i item, the queue discipline could just as well stipulate that after this item has been serviced no other type i items will be serviced until the queue no longer contains other type items. Let H_{Ki} be the distribution of the time required to empty the queue of items other than type i (including the service time of the initial i item). If in the time required to clear the queue of non-type i items

n type i items arrive, the busy period is prolonged by an n-fold convolution of busy periods B_{Ki}. Hence,

$$F_{B_{Ki}}(x) = \int_0^x \left[\sum_{n=0}^{\infty} e^{-\lambda_i y} \frac{(\lambda_i y)^n}{n!} F_{B_{Ki}}^{(n)}(x - y) \right] dH_{Ki}(y), \tag{4.8}$$

or

$$\tilde{B}_{Ki}(s) = \tilde{H}_{Ki}(s + \lambda_i(1 - \tilde{B}_{Ki}(s))). \tag{4.9}$$

But H_{Ki} is given by

$$\begin{aligned} H_{Ki}(y) = \int_0^y \Bigg[\sum_{n_1} \cdots \sum_{n_{i-1}} \sum_{n_{i+1}} \cdots \sum_{n_K} & e^{-(\lambda_1 + \cdots + \lambda_{i-1} + \lambda_{i+1} + \cdots + \lambda_K) y} \\ & \cdot \frac{(\lambda_1 y)^{n_1}}{n_1!} \cdots \frac{(\lambda_{i-1} y)^{n_{i-1}}}{n_{i-1}!} \frac{(\lambda_{i+1} y)^{n_{i+1}}}{n_{i+1}!} \cdots \frac{(\lambda_K y)^{n_K}}{n_K!} \\ & \cdot F_{{}_iB_{K-1,1}}^{(n_1)} * \cdots * F_{{}_iB_{K-1,i-1}}^{(n_{i-1})} * F_{{}_iB_{K-1,i+1}}^{(n_{i+1})} * \cdots * F_{{}_iB_{K-1,K}}^{(n_K)}(y - z) \Bigg] \cdot dF_{S_i}(z), \end{aligned} \tag{4.10}$$

where ${}_iB_{K-1,j}$ denotes a busy period for a queue with $K - 1$ priority classes, the ith class of the original K classes being absent, and a type j item in line initially. (4.10) implies

$$\tilde{H}_{Ki}(s) = \tilde{S}_i(s + {}_i\Lambda_{K-1}(1 - {}_i\tilde{B}_{K-1}^*(s))), \tag{4.11}$$

where ${}_i\Lambda_{K-1} = \Lambda_K - \lambda_i$, ${}_i\tilde{B}_{K-1}^* = \sum_{j \neq i} \lambda_j \, {}_i\tilde{B}_{K-1,j} / {}_i\Lambda_{K-1}$. (4.9) and (4.11) together yield

$$\begin{aligned} \tilde{B}_{Ki}(s) = \tilde{S}_i(s + \lambda_i(1 - \tilde{B}_{Ki}(s)) & \\ + {}_i\Lambda_{K-1}(1 - {}_i\tilde{B}_{K-1}^*(s + \lambda_i(1 - \tilde{B}_{Ki}(s))))). & \end{aligned} \tag{4.12}$$

$\tilde{B}_{Ki}$ is in fact the unique solution to the functional equation (4.12) subject to the regularity conditions (4.2). The proof of uniqueness is analogous to the previous proof for $\tilde{B}_K^*$.

The moments of B_{Ki} are derivable from (4.12). In particular,

$$E(B_{Ki}) = \frac{E(S_i)}{1 - \rho_1 - \cdots - \rho_K}; \tag{4.13}$$

$$E(B_{Ki}^2) = \frac{E(S_i^2)[1 - \sum_{j \neq i} \rho_j] + E(S_i)[\sum_{j \neq i} \lambda_j E(S_j^2))]}{\left[1 - \sum_1^K \rho_j\right]^3}. \tag{4.14}$$

The distribution of the busy period for a preemptive resume priority queue is identical to the busy period distribution for the same queue with head-of-the-line discipline. The order of service is immaterial to the busy period as long as preemption does not increase the time spent in the service mechanism which is the case for a resume discipline. Hence, all the previous results for head-of-the-line discipline apply as well to the preemptive resume queue. The indifferent server queue is included as a special case of the resume discipline. A similar identification cannot be made for a preemptive repeat priority queue since the length of time a lower priority item spends in the service mechanism is no longer equal to the service time in isolation. To date no characterization has been obtained for the busy period of a preemptive repeat priority queue.

5. Distribution of number of items serviced during a busy period. Takács in [15] characterized the distribution of the number of items serviced during a busy period for the simple queue, and this method can be adapted in a fashion analogous to Section 4. As before, the equilibrium condition $1 - \rho_1 - \cdots - \rho_K > 0$ will be assumed so that all the distributions which will be discussed have total variation one. The reader can easily modify the discussion to cover the transient case.

Consider a priority queue with K priority classes and head-of-the-line discipline. Let $f^*_{Ki}(j)$ be the probability that a total of j items, irrespective of class, are serviced during a busy period commencing with a single type i item in the queue, and let $f^*_K(j) = \tau_1 f^*_{K1}(j) + \cdots + \tau_K f^*_{KK}(j)$ be the probability of servicing a total of j items where the class of the initial item is unspecified. $\tilde{f}^*_{Ki}(s)$ and $\tilde{f}^*_K(s)$ will denote the generating functions of $\{f^*_{Ki}(j)\}$ and $\{f^*_K(j)\}$, respectively. For a specific class i let $f_{Ki}(j)$ be the probability of servicing j type i items in a service period which commences with a type i item in line initially. $\tilde{f}_{Ki}(s)$ will denote the generating function of $\{f_{Ki}(j)\}$.

The determination of $\tilde{f}^*_K(s)$ will be treated first. Let $p_{K:n_1 \ldots n_K}$ be the probability that during the service period for the initial unspecified item n_j type j items, $j = 1, \cdots, K$, arrive. Since the initial item is type j with probability τ_j,

$$p_{K:n_1 \cdots n_K} = \int_0^\infty \left[\sum_{n_1, \cdots, n_K} e^{-\Lambda_K t} \frac{(\Lambda_K t)^{\Sigma n_i}}{n_1! \cdots n_K!} (\tau_1)^{n_1} \cdots (\tau_K)^{n_K} \right] dF_{s^*_K}(t), \tag{5.1}$$

and

$$\begin{aligned} P_K(s_1, \cdots, s_K) &= \sum_{n_1, \cdots, n_K} p_{K:n_1 \cdots n_K} s_1^{n_1} \cdots s_K^{n_K} \\ &= \tilde{S}^*_K \left(\Lambda_K \left(1 - \sum_1^K \tau_i s_i \right) \right). \end{aligned} \tag{5.2}$$

By an argument analogous to that employed in Section 4, the $f^*_K(j)$ and $f^*_{Ki}(j)$ can be shown to satisfy the relations

$$\begin{aligned} f^*_K(1) &= p_{K:0 \ldots 0}, \\ f^*_K(j) &= \sum_{\substack{n_1, \cdots, n_K \\ 0 < \Sigma n_i \leqq j-1}} p_{K:n_1 \ldots n_K} \sum_{\substack{j_{11}, \cdots j_{Kn_K} \\ j_{11} + \cdots + j_{Kn_K} = j-1}} f^*_{K1}(j_{11}) \cdots \\ &\qquad \cdot f^*_{K1}(j_{1n_1}) \cdots f^*_{KK}(j_{K1}) \cdots f^*_{KK}(j_{Kn_K}), \qquad j \geqq 2. \end{aligned} \tag{5.3}$$

This yields for the generating function

$$\tilde{f}^*_K(s) = s \tilde{S}^*_K(\Lambda_K(1 - \tilde{f}^*_K(s))). \tag{5.4}$$

$\tilde{f}^*_K(s)$ is the unique solution to the functional equation

$$f(s) = s\tilde{S}^*_K(\Lambda_K(1 - f(s))), \qquad |s| \leqq 1, \tag{5.5}$$

subject to the regularity conditions

$$\begin{aligned} &\text{(i) } f(s) \text{ analytic for } |s| \leqq 1, \\ &\text{(ii) } f(0) = 0. \end{aligned} \tag{5.6}$$

The proof of uniqueness is omitted since it follows either from the simple queue proof [15] or from an argument similar to that for the busy period in the previous section.

The moments of N^*_K, the total number of items serviced during a busy period, are obtainable from (5.4). In particular,

$$\begin{aligned} E(N^*_K) &= \frac{1}{1 - \Lambda_K E(S^*_K)} = \frac{1}{1 - \rho_1 - \cdots - \rho_K}; \\ E(N^{*2}_K) &= \frac{\Lambda^2_K E(S^{*2}_K)}{(1 - \Lambda_K E(S^*_K))^3} + \frac{2\Lambda_K E(S^*_K)}{(1 - \Lambda_K E(S^*_K))^2} + \frac{1}{1 - \Lambda_K E(S^*_K)}. \end{aligned} \tag{5.7}$$

The property that the number of type i items serviced during a busy period is independent of the priority discipline makes it feasible to obtain a functional relation for $\tilde{f}_{Ki}(s)$. Let $p_{Ki:n}$ be the probability that n type i items arrive during the time it takes to service the initial type i item and then clear the queue of other type items without admitting any type i items into service.

$$p_{Ki:n} = \int_0^\infty e^{-\lambda_i t} \frac{(\lambda_i t)^n}{n!} \, dH_{Ki}(t), \tag{5.8}$$

and

$$P_{Ki}(s) = \sum_{n=0}^{\infty} p_{Ki:n} s^n = \tilde{H}_{Ki}(\lambda_i(1 - s)), \tag{5.9}$$

where $H_{Ki}(t)$ and $\tilde{H}_{Ki}(s)$ were defined in (4.10) and (4.11). Under the discipline of servicing the initial type i item and then clearing the queue of the other class items the $f_{Ki}(j)$ must satisfy

$$\begin{aligned} f_{Ki}(1) &= p_{Ki:0}, \\ f_{Ki}(j) &= \sum_{n=1}^{j-1} p_{Ki:n} \sum_{j_1 + \cdots + j_n = j-1} f_{Ki}(j_1) \cdots f_{Ki}(j_n), \qquad j \geqq 2, \end{aligned} \tag{5.10}$$

so

$$\tilde{f}_{Ki}(s) = s\tilde{H}_{Ki}(\lambda_i(1 - \tilde{f}_{Ki}(s))). \tag{5.11}$$

The proof that $\tilde{f}_{Ki}(s)$ is the unique solution to (5.11) subject to the regularity conditions (5.6) is omitted.

The first two moments of N_{Ki}, the number of type i items serviced during a busy period commencing with a type i item, are determinable from (5.11).

$$E(N_{Ki}) = \frac{1 - \sum_{j \neq i} \rho_j}{1 - \sum_{j} \rho_j};$$

$$(5.12) \quad E(N_{Ki}^2) = \frac{\lambda_i^2 E(S_i^2)[1 - \sum_{j \neq i} \rho_j] + \lambda_i^2 E(S_i)\,[\sum_{j \neq i} \lambda_j E(S_j^2))]}{[1 - \sum_{j} \rho_j]^3} + \frac{2\rho_i[1 - \sum_{j \neq i} \rho_j]}{[1 - \sum_{j} \rho_j]^2} + \frac{1 - \sum_{j \neq i} \rho_j}{1 - \sum_{j} \rho_j}.$$

The distributions of the total number of items and the number of type i items serviced when the initial arrival is of type j can be determined by forming the appropriate convolutions of service periods and busy periods with the distributions already determined in this section.

As in the case of the busy period distributions the above results for head-of-the-line discipline apply equally as well to the preemptive resume priority queue. This also includes as a special case the priority queue with indifferent server. The corresponding distributions for the preemptive repeat priority queue still remain to be determined.

6. Acknowledgments. I would like to thank Professor Samuel Karlin for suggesting this research topic and for generously contributing his help and guidance. My thanks also to Mr. J. H. Kullback for checking part of the algebraic acrobatics.

REFERENCES

[1] Alan Cobham, "Priority assignment in waiting line problems," *J. Opns. Res. Soc. Am.*, Vol. 2 (1954), pp. 70–76.

[2] Alan Cobham, "Priority assignment—a correction," *J. Opns. Res. Soc. Am.*, Vol. 3 (1955), p. 547.

[3] Julian L. Holley, "Waiting line subject to priorities," *J. Opns. Res. Soc. Am.*, Vol. 2 (1954), pp. 341–343.

[4] H. Kesten and J. Th. Runnenburg, "Priority in waiting line problems," *Proc. Akad. Wet. Amst. A*, Vol. 60 (1957), pp. 312–336.

[5] Thomas L. Saaty, "Résumé of useful formulas in queueing theory," *J. Opns. Res. Soc. Am.*, Vol. 5 (1957), pp. 161–200.

[6] Philip M. Morse, *Queues, Inventories, and Maintenance*, John Wiley and Sons, New York, 1958.

[7] Harrison White and Lee S. Christie, "Queueing with preemptive priorities or with breakdown," *J. Opns. Res. Soc. Am.*, Vol. 6 (1958), pp. 79–95.

[8] Frederick F. Stephan, "Two queues under preemptive priority with Poisson arrival and service rates," *J. Opns. Res. Soc. Am.*, Vol. 6 (1958), pp. 399–418.

[9] Ernest Koenigsberg, "Queueing with special service," *J. Opns. Res. Soc. Am.*, Vol. 4 (1956), pp. 213–220.

[10] David G. Kendall, "Some problems in the theory of queues," *J. Roy. Stat. Soc. B.*, Vol. 13 (1951), pp. 151–173.

[11] David 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.

[12] DAVID M. G. WISHART, "A queueing system with χ^2 service-time distribution," *Ann. Math. Stat.*, Vol. 27 (1956), pp. 768–779.

[13] RUPERT G. MILLER, JR., "A contribution to the theory of bulk queues," *J. Roy. Stat. Soc. B.*, to be published.

[14] D. V. LINDLEY, "The theory of queues with a single server," *Proc. Cambridge Philos. Soc.*, Vol. 48 (1952), pp. 277–289.

[15] LAJOS TAKÁCS, "Investigation of waiting time problems by reduction to Markov processes," *Acta Math., Acad. Sci. Hung.*, Vol. 6 (1955), pp. 101–128.