# PREEMPTIVE PRIORITY QUEUES

G. F. YEO

(received 11 June 1962)

## Summary

In this paper priority queues with $K$ classes of customers with a preemptive repeat and a preemptive resume policy are considered. Customers arrive in independent Poisson processes, are served, within classes, in order of arrival, and have general requirements for service. Transforms of stationary waiting time and queue size distributions and busy period distributions are obtained for individual classes and for the system; the moments of the distributions are considered.

## 1. Introduction

Let us consider a preemptive priority queueing system where there are $K$ classes $1, 2, \ldots, K$ of customers arriving at a counter with a single server. Within each class there is a first-come first-served queue discipline, and between classes there is a relative priority of service such that a class $i$ customer is served in preference to a class $j$ customer whenever $i < j$. When a class $j$ customer is being served and a class $i (< j)$ customer arrives, then service of the class $j$ customer ceases immediately in favour of the class $i$ customer, and it resumes only when the queueing system is next cleared of all customers of all classes $1, 2, \cdots, j-1$. When a preempted customer returns to actual service we consider two possible cases; a preemptive resume policy allows the customer to re-enter service at the point where it was preempted; a preemptive repeat policy means that the customer has to begin service again. When $K = 2$ class 1 customers are called priority and class 2 non-priority customers.

We assume that class $i$ $(1 \leqq i \leqq K)$ customers arrive independently in a Poisson process with parameter $\lambda_i$, and independently of customers in other classes. Class $i$ customers occupy the server for a time, which is total time for a resume policy and uninterrupted time for a repeat policy, which has a distribution function (d.f.) $F_i(x)$ $(0 \leqq x < \infty)$ with Laplace-Stieltjes transform (L.T.) $\psi_i(\theta) = \int_0^\infty e^{-\theta x} dF_i(x) (\theta \geqq 0)$ and finite mean $\mu_i = -\psi_i'(0) < \infty$. We suppose also that $\psi_i(\theta)$ exists and is finite for all real $\theta \geqq -R_i$, where $R_i$ is a real non-negative constant; when $R_i > 0$ this



https://doi.org/10.1017/S1446788700039124 Published online by Cambridge University Press

is equivalent to the d.f. $F_i(x)$ having an analytic characteristic function. As a special case we consider a constant service time of $b_i$ where $F_i(x) = 1$ for $x \geqq b_i$ and $F_i(x) = 0$ for $x < b_i$.

The preemptive resume priority system has been considered by several authors such as Miller [6], and some of our results for this case overlap with those of [6]. Although the results obtained for the preemptive resume system are more complete than for the preemptive repeat system, we shall consider the latter in greater detail, as fewer results have previously been obtained for this system.

We are concerned with the stationary waiting time and queue size distributions and busy period distributions, where they exist, for the queueing system as a whole and for individual classes of customers. We shall consider the case $K = 2$, and then extend the results for $K \geqq 2$, but first we shall give the general formulae to be used for this purpose.

## 2. General formulae

We require a number of results of models which are generalisations of the $M/G/1$ single server queueing system with a negative exponential inter-arrival time distribution and a general service time distribution. Firstly let us consider $K$ classes of customers arriving independently in Poisson processes with parameters $\lambda_i (1 \leqq i \leqq K)$, and being served in order of arrival, i.e. no priority, with service times having d.f.'s $C_i(x)$ $(0 \leqq x < \infty)$ with L.T.'s $\xi_i(\theta)$ and means $\bar{\xi}_i = -\xi'(0) < \infty$. For $\sum_{i=1}^K \lambda_i \bar{\xi}_i < 1$ the L.T. $\zeta_K(\theta)$ of the stationary waiting time distribution and the probability generating function (p.g.f.) $q_K(z)$ of the stationary queue size distribution may be obtained, by the same method as for a single class of customers [7], as

$$\zeta_K(\theta) = \left\{1 - \sum_{i=1}^K \lambda_i \bar{\xi}_i\right\} \Big/ \left\{1 - \theta^{-1} \sum_{i=1}^K \lambda_i (1 - \xi_i(\theta))\right\} \qquad \theta \geqq 0, \tag{2.1}$$

$$q_K(z) = \frac{\left(1 - \sum_{i=1}^K \lambda_i \bar{\xi}_i\right)(1-z) \sum_{i=1}^K \lambda_i \xi_i(\nu_K(1-z))}{\sum_{i=1}^K \lambda_j \xi_j(\lambda(1-z)) - \nu_K z} \qquad |z| \leqq 1, \tag{2.2}$$

where $\nu_K = \lambda_1 + \lambda_2 + \cdots + \lambda_K$. The mean waiting time is

$$\bar{\zeta}_K = -\zeta_K'(0) = \sum_{i=1}^K \lambda_i \xi_i''(0) \Big/ 2 \left(1 - \sum_{i=1}^K \lambda_i \bar{\xi}_i\right). \tag{2.3}$$

The L.T. $\varphi_K(\theta)$ of the busy period distribution is the unique solution, with $\lim_{\theta \to \infty} \varphi_K(\theta) = 0$, of the functional equation

$$\varphi_K(\theta) = \nu_K^{-1} \sum_{i=1}^K \lambda_i \xi_i(\nu_K + \theta - \nu_K \varphi_K(\theta)). \tag{2.4}$$

https://doi.org/10.1017/S1446788700039124 Published online by Cambridge University Press

Another generalisation is where customers, arriving in a Poisson process with parameter $\lambda$, who arrive to find the server idle wait a time whose d.f. $V(x)(0 \leqq x < \infty)$ has L.T. $\vartheta(\theta)$ with mean $\bar{v} = -\vartheta'(0) < \infty$ before commencing service, which has a d.f. $A(x)$ with L.T. $\alpha(\theta)$ and mean $\bar{\alpha} = -\alpha'(0) < \infty$, and which is the same for all customers. Finch [2] has shown that the L.T. $\Omega(\theta)$ of the stationary waiting time distribution, which exists for $\lambda\bar{\alpha} < 1$, is given by

$$(2.5) \qquad \Omega(\theta) = \frac{(1-\lambda\bar{\alpha})\{\theta\vartheta(\theta)+\lambda-\lambda\vartheta(\theta)\}}{(1+\lambda\bar{v})\{\theta-\lambda+\lambda\alpha(\theta)\}},$$

with the mean waiting time being

$$(2.6) \qquad \bar{\Omega} = -\Omega'(0) = \frac{(1-\lambda\bar{\alpha})(2\bar{v}+\lambda\vartheta''(0))+\lambda\alpha''(0)(1+\lambda\bar{v})}{2(1+\lambda\bar{v})(1-\lambda\bar{\alpha})}.$$

A similar case is where customers joining non-empty and empty queues have different service time d.f.'s $A(x)$ and $B(x)$ $(0 \leqq x < \infty)$ with L.T.'s $\alpha(\theta)$ and $\beta(\theta)$ with finite means $\bar{\alpha} = -\alpha'(0)$ and $\bar{\beta} = -\beta'(0)$ respectively. The L.T. $\Phi(\theta)$ of the stationary waiting time distribution and the p.g.f. $r(z)$ of the stationary queue size distribution, which exist for $\lambda\bar{\alpha} < 1$, have been obtained [9] as

$$(2.7) \qquad \Phi(\theta) = \frac{(1-\lambda\bar{\alpha})\{\theta+\lambda\alpha(\theta)-\lambda\beta(\theta)\}}{(1-\lambda\bar{\alpha}+\lambda\bar{\beta})\{\theta-\lambda+\lambda\alpha(\theta)\}},$$

$$(2.8) \qquad r(z) = \frac{(1-\lambda\bar{\alpha})\{\alpha(\lambda(1-z))-z\beta(\lambda(1-z))\}}{(1-\lambda\bar{\alpha}+\lambda\bar{\beta})\{\alpha(\lambda(1-z))-z\}}.$$

The stationary queue size distribution in the previous case is given by (2.8) with $\beta(\theta) = \alpha(\theta)\vartheta(\theta)$. The mean waiting time is

$$(2.9) \qquad \bar{\Phi} = -\Phi'(0) = \frac{\lambda\beta''(0)-\lambda^2\bar{\alpha}\beta''(0)+\lambda^2\bar{\beta}\alpha''(0)}{2(1-\lambda\bar{\alpha})(1-\lambda\bar{\alpha}+\lambda\bar{\beta})}.$$

The L.T. $\delta(\theta)$ of the busy period distribution is given by

$$(2.10) \qquad \delta(\theta) = \beta(\eta(\theta)),$$

where $\eta(\theta)$ is the unique solution with $\eta(0) = 0$ of the functional equation

$$(2.11) \qquad \eta(\theta) = \lambda+\theta-\lambda\alpha(\eta(\theta)).$$

## 3. Waiting time distributions for $K = 2$

We consider the following stationary waiting time distributions for both preemptive repeat and resume policies: for non-priority customers, waiting time is defined as the length of time a non-priority customer spends

https://doi.org/10.1017/S1446788700039124 Published online by Cambridge University Press

from the time of arrival to the time (i) he reaches the counter for the first time and (ii) he reaches the head of the non-priority queue; for the priority non-priority system, waiting time is the time a customer takes from arrival to reaching the counter for the first time. Once a customer has commenced service for the first time, he is no longer waiting, even if he has been pre-empted, but is considered to be "in service". Case (i) is the usual problem considered; however, it is no more difficult to also include the second case, which has some applications. A similar model is used by Yeo and Weesakul [8] for a traffic problem of the delay to vehicles at an intersection.

A slightly more general model, which may be considered by the same method as used here, is to suppose that instead of there being priority customers there are interruptions to the system. These interruptions have a duration which has a general d.f. $G_1(x)$, not necessarily that of a busy period, and the distribution of time between interruptions is negative exponential.

Priority customers have a general service time distribution with L.T. $\psi_1(\theta)$. The distribution of priority customers is independent of non-priority customers, and is given by the $M/G/1$ queueing system; the transforms $\Omega_1(\theta)$, $q_1(z)$ and $\varphi_1(\theta)$ of the stationary waiting time, stationary queue size and busy period distributions are given by (2.1), (2.2) and (2.4) with $\lambda = \lambda_1$, $K = 1$, and $\xi_1(\theta) = \psi_1(\theta)$.

We now consider the preemptive repeat priority queueing system with non-priority customers requiring uninterrupted occupation of the server for a fixed time $b (> 0)$ before being able to depart. If there is a priority customer arriving at time zero we require the L.T. $\chi_1(\theta, b)$ of the d.f. $K_1(x, b)$ of the time until the first gap of at least $b$, including $b$, appears in the priority stream; this is the continuous analogue of a generalisation of the success run problem of Feller [1] (p. 299), and we obtain (or by renewal theory)

$$
\begin{aligned}
\chi_1(\theta, b) = & \int_{z=0}^{\infty} dG_1(z) \Big\{ e^{-(\lambda_1+\theta)b-\theta z} + \int_{y_1=0}^{b} \lambda_1 e^{-\lambda_1 y_1} dy_1 \int_{z_1=y_1}^{\infty} dG_1(z_1-y_1) \\
& \Big\{ e^{-(\lambda_1+\theta)b-\theta(z+z_1)} + \int_{y_2=0}^{b} \lambda_1 e^{-\lambda_1 y_2} dy_2 \int_{z_2=y_2}^{\infty} dG_2(z_2-y_2) \\
& \Big\{ e^{-(\lambda_1+\theta)b-(z+z_1+z_2)} + \cdots \Big\}\Big\}\Big\} \\
= & \frac{(\lambda_1+\theta)\varphi_1(\theta)e^{-(\lambda_1+\theta)b}}{\lambda_1+\theta-\lambda_1\varphi_1(\theta)\{1-e^{-(\lambda_1+\theta)b}\}}.
\end{aligned}
\tag{3.1}
$$

When a non-priority customer reaches the counter for the first time he has a service time of length $b$ if no priority customers arrive in this period; otherwise he waits until the server is free of priority customers for at least $b$ before he departs from the system. The L.T. $\alpha_2(\theta, b) = \int_0^\infty e^{-\theta x} dA_2(x, b)$

https://doi.org/10.1017/S1446788700039124 Published online by Cambridge University Press

of the length of time a non-priority customer spends from his arrival at the counter to his departure is

$$\alpha_2(\theta, b) = e^{-(\lambda_1+\theta)b} + \int_{y=0}^{b} \lambda_1 e^{-\lambda_1 y} dy \int_{z=y}^{\infty} e^{-\theta z} dK_1(z-y, b) \tag{3.2}$$
$$= e^{-(\lambda_1+\theta)b} + \frac{\lambda_1 \chi_1(\theta, b)}{\lambda_1+\theta}\{1-e^{-(\lambda_1+\theta)b}\}.$$

A non-priority customer may arrive to find the queue free of other non-priority customers, but not of priority customers, so that he must wait until the priority stream is cleared before he commences service. This waiting time is zero if there are no priority customers in the queue, and lasts until the end of the priority busy period if there is at least one priority customer in the queue when he arrives. This depends on the length of time since the departure of the last non-priority customer. From the point of view of non-priority customers the server appears to be alternately present, for a length of time which has a negative exponential distribution with mean $\lambda_1^{-1}$, and absent, for a length of time which as the busy period d.f. $G_1(x)$ of priority customers. During the first of the present-absent periods of the server since the departure of the last non-priority customer, the L.T. ${}_1\vartheta_2(\theta, b)$ of the wait of the next non-priority customer is

$${}_1\vartheta_2(\theta, b) \equiv {}_1\vartheta_2(\theta) = \int_{v=0}^{\infty} \lambda_2 e^{-\lambda_2 v-\lambda_1 v} dv \tag{3.3}$$
$$+ \int_{y=0}^{\infty}\int_{z=y}^{\infty}\int_{v=y}^{z} \lambda_1\lambda_2 e^{-\lambda_2 v-\lambda_1 y-\theta(z-v)} dy\, dG_1(z-y) dv$$
$$= \frac{\lambda_2(\lambda_2-\theta+\lambda_1\varphi_1(\theta)-\lambda_1\varphi_1(\lambda_2))}{(\lambda_1+\lambda_2)(\lambda_2-\theta)}.$$

We have ${}_1\vartheta_2(0) = 1-g_1 = 1-\lambda_1\nu_2^{-1}\varphi_1(\lambda_2)$, $g_1$ being the probability of no non-priority customers arriving during a server present-absent period. If the first non-priority customer arrives during a later server present-absent period, then the distribution of his wait is of the same form, so that the L.T. $\vartheta_2(\theta)$ of the distribution of the wait of a non-priority customer before reaching the counter is

$$\vartheta_2(\theta) = {}_1\vartheta_2(\theta)\{1+g_1+g_1^2+\cdots\} \tag{3.4}$$
$$= \frac{\lambda_2\{\lambda_2-\theta+\lambda_1\varphi_1(\theta)-\lambda_1\varphi_1(\lambda_2)\}}{\nu_2(\lambda_2-\theta)(1-g_1)}.$$

For the type (ii) of waiting time we have defined the service time distribution of non-priority customers joining a non-empty queue is readily seen to have the L.T $\alpha_2(\theta, b)$ given by (3.2). For a non-priority customer joining a queue free of other non-priority customers his service time distribution, with L.T. $\beta_2(\theta, b) = \int_{x=0}^{\infty} e^{-\theta x} dB_2(x, b)$, is the convolution of the

https://doi.org/10.1017/S1446788700039124 Published online by Cambridge University Press

distribution of his wait to reach the counter and of the distribution of the time he spends from reaching the counter to departing from the system; thus

$$\beta_2(\theta, b) = \alpha_2(\theta, b)\eth_2(\theta). \tag{3.5}$$

Let us now suppose that non-priority customers require an uninterrupted service with d.f $F_2(x)$ having L.T. $\psi_2(\theta)$. The L.T.'s of the distributions of time in service are given by

$$\begin{aligned} \int_{x=0}^{\infty} \alpha_2(\theta, x) dF_2(x) &= \int_{x=0}^{\infty} \int_{y=0}^{\infty} e^{-\theta y} dA_2(y, x) dF_2(x) \\ &= \int_{y=0}^{\infty} e^{-\theta y} \int_{x=0}^{\infty} dA_2(y, x) dF_2(x) \\ &= \int_{y=0}^{\infty} e^{-\theta y} dA_2(y) = \alpha_2(\theta) \\ \beta_2(\theta) &= \int_0^{\infty} e^{-\theta x} dB_2(x) = \alpha_2(\theta)\eth(\theta), \end{aligned} \tag{3.6}$$

where the inversion of the order of integration may be justified by Fubini's theorem [5]. Integration of (3.2), (3.5) and (3.6) does not in general yield explicit results, although the results may formally be given. For definitions (i) and (ii) of waiting time the L.T.'s $\Omega_2(\theta)$ and $\Phi_2(\theta)$ of the stationary waiting time distributions for non-priority customers are given by substitution in (2.5) and (2.7) respectively with $\Omega(\theta) = \Omega_2(\theta)$, $\Phi(\theta) = \Phi_2(\theta)$, $\lambda = \lambda_2$, $\eth(\theta) = \eth_2(\theta)$, $\alpha(\theta) = \alpha_2(\theta)$, $\beta(\theta) = \beta_2(\theta)$. The p.g.f. $r_2(z)$ of the stationary queue size distribution, which is identical with both definitions, is given by (2.8) with $r(z) = r_2(z)$. We have assumed the existence of stationary distributions for which the condition has been given in § 2 as $\lambda_2 \bar{\alpha}_2 < 1$, where $\bar{\alpha}_2 = -\alpha_2'(0)$.

We can obtain the moments of the service time distributions by differentiation as

$$\bar{\alpha}_2 = -\int_0^{\infty} \alpha'(0, x) dF_2(x), \; \bar{\beta}_2 = \bar{\alpha}_2 + \bar{v}_2, \tag{3.7}$$

where a sufficient condition for these means to be finite is that $R_2 \geqq \lambda_1$ as may be observed by differentiating (3.2) and (3.5); we obtain

$$\begin{aligned} \bar{\alpha}_2 &= \lambda_1^{-1}(1+\lambda_1\bar{\varphi}_1)\{\psi_2(-\lambda_1)-1\} \\ \bar{v}_2 &= \nu_2^{-1}(1+\lambda_1\bar{\varphi}_1)/(1-g_1)-\lambda_2^{-1} \\ \bar{\beta}_2 &= \bar{\alpha}_2+\bar{v}_2. \end{aligned} \tag{3.8}$$

Similarly higher order moments may be determined; the second moments, which are finite if $R_2 \geqq 2\lambda_1$, are

https://doi.org/10.1017/S1446788700039124 Published online by Cambridge University Press

$$\begin{aligned}
\alpha_2''(0) &= \varphi_1''(0)\{\psi_2(-\lambda_1)-1\}+2\lambda_1^{-2}(1+\lambda_1\bar\varphi_1)\left[(1+\lambda_1\bar\varphi_1)(\psi_2(-2\lambda_1)\right.\\
&\qquad \left.-\psi_2(-\lambda_1))-\lambda_1\bar\varphi_1\{\psi_2(-\lambda_1)-1\}+\lambda_1\psi_2'(-\lambda_1)\right]\\
\text{(3.9)}\qquad \mathfrak{d}_2''(0) &= \lambda_1\nu_2^{-1}\varphi_1''(0)/(1-g_1)-2\lambda_2^{-1}\nu_2^{-1}(1+\lambda_1\bar\varphi_1)/(1-g_1)+2\lambda_2^{-2}\\
\beta_2''(0) &= \alpha_2''(0)+2\bar\alpha_2\bar v_2+\mathfrak{d}_2''(0),
\end{aligned}$$

where $\psi_2'(-\lambda_1)=\lim_{\theta\to-\lambda_1} d/d\theta\,\psi_2(\theta)$. The mean stationary waiting times are obtained by substitution of (3.8) and (3.9) in (2.6) and (2.9).

We define the delay as the total time a customer spends in the system; its distribution is the convolution of the distribution of wait and the distribution of service time. For non-priority items this is the same for both definitions (i) and (ii) and its L.T. $D_2(\theta)$ is given by

$$\text{(3.10)}\qquad D_2(\theta)=\frac{(1-\lambda\bar\alpha)\alpha(\theta)\{\theta\mathfrak{d}(\theta)+\lambda-\lambda\mathfrak{d}(\theta)\}}{(1+\lambda\bar v)\{(\theta-\lambda+\lambda\alpha(\theta)\}}.$$

The mean delay is

$$\text{(3.11)}\qquad -D_2'(0)=\frac{(1+\lambda\bar v)\{\lambda\alpha''(0)+2\bar\alpha(1-\lambda\bar\alpha)\}+(1-\lambda\bar\alpha)(\lambda\mathfrak{d}''(0)+2\bar v)}{2(1-\lambda\bar\alpha)(1+\lambda\bar v)}.$$

For a preemptive resume priority queueing system the same method as used above may be employed; however, the problem is simpler. If non-priority customers occupy the server for a total time $b$, we obtain

$$\begin{aligned}
\alpha_2(\theta,b) &= e^{-(\lambda_1+\theta)b}+\int_{y_1=0}^{b}\lambda_1e^{-\lambda_1y_1}dy_1\int_{z_1=y_1}^{\infty}dG_1(z_1-y_1)\\
&\qquad \Big\{e^{-\lambda_1(b-y_1)-\theta(z_1-y_1+b)}\\
&\qquad +\int_{y_2=0}^{b-y_1}\lambda_1e^{-\lambda_1y_2}dy_2\int_{z_2=y_2}^{\infty}dG_1(z_2-y_2)\\
&\qquad \Big\{e^{-\lambda_1(b-y_1-y_2)-\theta(z_1+z_2+b-y_1-y_2)}\\
\text{(3.12)}\qquad &\qquad +\int_{y_3=0}^{b-y_1-y_2}\lambda_1e^{-\lambda_1y_3}dy_3\int_{z_3=y_3}^{\infty}dG_1(z_3-y_3)\\
&\qquad \Big\{e^{-\lambda_1(b-y_1-y_2-y_3)}e^{-\theta(z_1+z_2+z_3+b-y_1-y_2-y_3)}+\cdots\Big\}\Big\}\Big\}\\
&= e^{-(\lambda_1+\theta-\lambda_1\varphi_1(\theta))b}\\
\mathfrak{d}_2(\theta,b) &\equiv \mathfrak{d}_2(\theta)=\frac{\lambda_2\{\lambda_2-\theta+\lambda_1\varphi_1(\theta)-\lambda_1\varphi_1(\lambda_2)\}}{\nu_2(\lambda_2-\theta)(1-g_1)}\\
\beta_2(\theta,b) &= \alpha_2(\theta,b)\mathfrak{d}_2(\theta).
\end{aligned}$$

From these we readily obtain for a general d.f. $F_2(x)$ for the time a non-priority customer occupies the server that

https://doi.org/10.1017/S1446788700039124 Published online by Cambridge University Press

$$\alpha_2(\theta) = \psi_2(\lambda_1+\theta-\lambda_1\varphi_1(\theta))$$

(3.13)
$$\mathfrak{d}_2(\theta) = \frac{\lambda_2\{\lambda_2-\theta+\lambda_1\varphi_1(\theta)-\lambda_1\varphi_1(\lambda_2)\}}{\nu_2(\lambda_2-\theta)(1-g_1)}$$

$$\beta_2(\theta) = \alpha_2(\theta)\mathfrak{d}_2(\theta).$$

We substitute (3.13) in (2.5) and (2.7) to obtain the L.T.'s of the stationary waiting time distributions; these exist for $\lambda_1\mu_1+\lambda_2\mu_2 < 1$. The moments may be found by differentiation. The stationary queue size distribution for non-priority customers is given by (2.8). Miller [6] has obtained $\alpha_2(\theta)$ and $\Omega_2(\theta)$ ((3.26) and (3.23) with $j = 2$) by a slightly different method; he has also obtained (4.1) for a preemptive resume policy.

The L.T. of the distribution of the time required to service all the customers in the queue at a particular time, neglecting further arrivals, is given by (2.1) with $\xi_i(\theta) = \psi_i(\theta)$ $(i = 1, 2)$ and $K = 2$ (c.f. Miller [6], (3.24)). This is identical with that for a (non-preemptive) postponable priority policy for two classes of customers, as here the order of service is irrelevant if further arrivals are not considered. For this reason the busy period has the same distribution for the two cases, and so has the waiting time distribution of non-priority (and in general for the lowest class $K$) customers. However, the waiting time distribution for priority customers is different (greater) for a postponable priority policy as priority customers may be delayed by non-priority customers.

## 4. Busy period distributions for $K = 2$

If we define the L.T.'s $\alpha_2(\theta)$ and $\beta_2(\theta)$ of the service time distributions by (3.6) for a preemptive repeat policy and by (3.13) for a preemptive resume policy then the busy period distributions for these two problems may be considered together.

A busy period which commences with a non-priority customer has L.T. $\varphi_{22}(\theta) = \int_{x=0}^{\infty} e^{-\theta x} dG_{22}(x)$ satisfying

(4.1)
$$\varphi_{22}(\theta) = \int_{y=0}^{\infty} \sum_{n=0}^{\infty} \int_{z=0}^{\infty} e^{-\lambda_2 y} \frac{(\lambda_2 y)^n}{n!} e^{-\theta(y+z)} dA_2(y) dG_{22}^{(n)}(z)$$
$$= \alpha_2(\lambda_2+\theta-\lambda_{22}\varphi_2(\theta)),$$

where $G_{22}^{(n)}(z)$ is the $n$-fold convolution of $G_{22}(z)$. The L.T. $\varphi_{12}(\theta)$ of the distribution of the length of a busy period starting with a priority customer is

(4.2)
$$\varphi_{12}(\theta) = \int_{y=0}^{\infty} \sum_{n=0}^{\infty} \int_{z=0}^{\infty} e^{-\lambda_2 y} \frac{(\lambda_2 y)^n}{n!} e^{-\theta(y+z)} dG_1(y) dG_{22}^{(n)}(z)$$
$$= \varphi_1(\lambda_2+\theta-\lambda_2\varphi_{22}(\theta)).$$

https://doi.org/10.1017/S1446788700039124 Published online by Cambridge University Press

It is apparent from the definitions of $\varphi_{22}(\theta)$ and $\varphi_{12}(\theta)$ that the L.T. $\varphi_2^*(\theta) = \int_0^\infty e^{-\theta x} dG_2^*(x)$ of the busy period distribution for the priority non-priority system is given by

$$\varphi_2^*(\theta) = \lambda_1 \nu_2^{-1} \varphi_{12}(\theta) + \lambda_2 \nu_2^{-1} \varphi_{22}(\theta). \tag{4.3}$$

The busy period distribution for non-priority customers has L.T. $\delta_2(\theta)$ given by the relations (2.10) and (2.11) with $\lambda = \lambda_2$, $\alpha(\theta) = \alpha_2(\theta)$, $\beta(\theta) = \beta_2(\theta)$, $\delta(\theta) = \delta_2(\theta)$.

For a preemptive priority resume policy, and for a postponable priority policy, the L.T. $\varphi_2^*(\theta)$ of the busy period distribution for the system satisfies (2.4) with $\varphi_2(\theta) = \varphi_2^*(\theta)$, $\xi_i(\theta) = \psi_i(\theta)$ $(i = 1, 2)$ and $K = 2$; this is equivalent to (4.3).

The above properties of the busy period distributions are sufficient to extend our results to $K(\geqq 2)$ classes; however, for the non-priority customers let us consider the more general problem of the joint d.f. $G_{22}(n, t)$ of the length $t$ of a busy period and the number $n$ of non-priority customers served in this period. Using the method of Gaver [3] we can show that the transform

$$\gamma_2(\theta, z) = \sum_{n=1}^{\infty} \int_{t=0}^{\infty} z^n e^{-\theta t} dG_{22}(n, t) \qquad (0 < z \leqq 1, \text{ real } \theta \geqq 0)$$

is given by

$$\gamma_2(\theta, z) = z\beta_2(\lambda_2 + \theta - \lambda_2 x) \tag{4.4}$$

where $x$ is the unique root for real $\theta > 0$ and $0 < z \leqq 1$ of the equation

$$x = z\alpha_2(\lambda_2 + \theta - \lambda_2 x).$$

## 5. The general case $K \geqq 2$

We now generalise the results of § 3 and § 4 to the case where there are $K \geqq 2$ priority classes of customers. Class $i$ $(1 \leqq i \leqq K-1)$ customers have a service requirement (uninterrupted and total for preemptive repeat and resume policies respectively) d.f. $F_i(x)$ $(0 \leqq x < \infty)$, while class $K$ customers have fixed service time requirement $b$. We extend the previous definitions in a natural manner so that $\alpha_K(\theta, b)$ is the L.T. of the distribution of the time a class $K$ customer takes from arriving at the counter to his departure from the system; $\varphi_{ij}(\theta)$ $(1 \leqq i \leqq j \leqq K)$ is the L.T. of the busy period distribution for the first $j$ classes given that it commences with the arrival of a class $i$ customer at a free counter; and so on.

Since the method is only a simple extension of that for two classes, we shall not give a detailed derivation of the service time distributions. For a preemptive repeat policy we obtain

https://doi.org/10.1017/S1446788700039124 Published online by Cambridge University Press

$$\alpha_K(\theta, b) = e^{-(\nu_{K-1}+\theta)b} \Big/ \left\{1 - \sum_{i=1}^{K-1} \frac{\lambda_i}{\nu_{K-1}+\theta} \varphi_{iK-1}(\theta)\{1-e^{-(\nu_{K-1}+\theta)b}\}\right\}$$

$$(5.1) \qquad \ni_K(\theta, b) \equiv \ni_K(\theta) = \lambda_K \frac{\left[\lambda_K - \theta + \sum_{i=1}^{K} \lambda_i\{\varphi_{iK-1}(\theta) - \varphi_{iK-1}(\lambda_K)\}\right]}{\nu_K(\lambda_K - \theta)(1-g_{K-1})}$$

$$\beta_K(\theta, b) = \alpha_K(\theta, b)\ni_K(\theta),$$

where

$$g_{K-1} = \nu_K^{-1} \sum_{i=1}^{K-1} \lambda_i \varphi_{iK-1}(\lambda_K), \quad \nu_K = \sum_{i=1}^{K} \lambda_i,$$

and the $\varphi_{iK-1}(\theta)$ $(1 \leqq i \leqq K-1)$ are obtained below in (5.4).

If class $K$ customers require uninterrupted service which has d.f. $F_K(x)$ $(0 \leqq x < \infty)$ we can integrate the expressions of (5.1) to yield

$$(5.2) \qquad \begin{aligned} \alpha_K(\theta) &= \int_{x=0}^{\infty} \alpha_K(\theta, x) dF_K(x) \\ \beta_K(\theta) &= \alpha_K(\theta)\ni_K(\theta). \end{aligned}$$

The L.T.'s $\Omega_K(\theta)$ and $\Phi_K(\theta)$ of the stationary waiting time distributions for class $K$ customers with waiting time definitions (i) and (ii) are obtained by substitution of (5.2) in the appropriate formulae of § 2 with $\Omega(\theta) = \Omega_K(\theta)$, $\Phi(\theta) = \Phi_K(\theta)$, $\lambda = \lambda_K$, $\alpha(\theta) = \alpha_K(\theta)$, $\ni(\theta) = \ni_K(\theta)$, $\beta(\theta) = \beta_K(\theta)$. The mean stationary waiting times may be obtained and are finite if $R_K \geqq 2\nu_{K-1}$ (and $R_j \geqq 2\nu_{j-1}$, $j = 2, 3, \cdots, K-1$). The stationary queue size distributions may be obtained from (2.2) and (2.8). We have assumed the existence of proper stationary distributions for which a necessary and sufficient condition is $\lambda_K \bar{\alpha}_K < 1$.

For a preemptive resume policy the results of the previous section may be generalised to give

$$(5.3) \qquad \begin{aligned} \alpha_K(\theta) &= \psi_K\left(\nu_{K-1} + \theta - \sum_{i=1}^{K-1} \lambda_i \varphi_{iK-1}(\theta)\right) \\ \ni_K(\theta) &= \lambda_K \frac{\left[\lambda_K - \theta + \sum_{i=1}^{K-1} \lambda_i\{\varphi_{iK-1}(\theta) - \varphi_{iK-1}(\lambda_K)\}\right]}{\nu_K(\lambda_K - \theta)(1-g_{K-1})} \\ \beta_K(\theta) &= \alpha_K(\theta)\ni_K(\theta). \end{aligned}$$

The L.T.'s of the stationary waiting time and queue size distributions, which exist for $\sum_{i=1}^{K} \lambda_i \mu_i < 1$, may be obtained from the formulae of § 2 with $\lambda = \lambda_K$, $\alpha(\theta) = \alpha_K(\theta)$, $\ni(\theta) = \ni_K(\theta)$, $\beta(\theta) = \beta_K(\theta)$.

https://doi.org/10.1017/S1446788700039124 Published online by Cambridge University Press

For both a repeat and a resume policy the L.T.'s $\varphi_{iK}(\theta)$ of the distribution of the length of a busy period starting with a class $i$ customer are

$$\varphi_{KK}(\theta) = \alpha_K(\lambda_K+\theta-\lambda_K\varphi_{KK}(\theta))$$

(5.4)

$$\varphi_{iK}(\theta) = \varphi_{ii}(\nu_K-\nu_i+\theta-\sum_{j=i+1}^{K}\lambda_j\varphi_{jK}(\theta)),$$

so that the L.T. $\varphi_K^*(\theta)$ of the busy period distribution for the system is

$$\varphi_K^*(\theta) = \sum_{i=1}^{K}\lambda_i\varphi_{iK}(\theta). \tag{5.5}$$

For a preemptive resume policy $\varphi_K^*(\theta)$ is also given by

$$\varphi_K^*(\theta) = \nu_K^{-1}\sum_{i=1}^{K}\lambda_i\psi_i(\nu_K+\theta-\nu_K\varphi_K^*(\theta)),$$

which is equivalent to (5.5); for this policy Miller has obtained $\varphi_{KK}(\theta)$ and $\Omega_K(\theta)$.

We see that from the solutions obtained for the service time and busy period distributions, we can find the service time distributions of the $j$th priority class in terms of the busy period distributions for the $j$-1th class. We may then obtain the busy period distributions for the $j$th class from the service time distributions for the $j$th class. Thus the service time, and hence the waiting time distributions, and busy period distributions can be constructed by iteration from one class to the next.

Since this paper was prepared Gaver [4] has published a paper in which he obtains some of our results by a slightly different method, and includes some results for the preemptive repeat priority system, e.g. (3.2) and (4.1). Gaver [4] has considered as well a preemptive repeat different priority system, which may also be solved by our method.

## References

[1] Feller, W., "An Introduction to Probability Theory and its Applications", 2nd Ed. Vol. 1, John Wiley and Sons, (1957).

[2] Finch, P. D., "A probability limit theorem with application to a generalisation of queueing theory, "Acta Math. Acad. Sci. Hung. (1959) **10**, 113—122.

[3] Gaver, D. P., "Imbedded Markov chain analysis of a waiting-line process in continuous time", Ann. Math. Statist. (1959) **30**, 698—720.

[4] Gaver, D. P., "A waiting line with interrupted service, including priorities", J. Roy. Statist. Soc. (1962) **B 24**, 73—90.

[5] Lukács, E., "Characteristic Functions", Charles Griffen (1960).

[6] Miller, R. G., "Priority queues", Ann. Math. Statist. (1960) **31**, 86—103.

https://doi.org/10.1017/S1446788700039124 Published online by Cambridge University Press

[7] Takács, L., "Investigation of waiting time problems by reduction to Markov processes", Acta Math. Acad. Sci. Hung. (1955) **6**, 101—129.
[8] Yeo, G. F. and Weesakul, B., "Distribution of delay to traffic at an intersection", (1963) (submitted for publication).
[9] Yeo, G. F., "Single server queues with modified service mechanisms", J. Aust. Math. Soc. (1962) **2**, 499—507.

Australian National University,
Canberra.

https://doi.org/10.1017/S1446788700039124 Published online by Cambridge University Press