TWO QUEUES IN SERIES

TAKEJI SUZUKI

Hosei University

(*Received May* 31, 1963)

I. INTRODUCTION

In this paper we consider the following queueing system consisting of two queues in series: Customers arrive at the first counter at the instants $\tau_0, \tau_1, \ldots, \tau_n, \ldots$ The interarrival times $\vartheta_n=\tau_{n+1}-\tau_n$ ($n=0, 1, 2, \ldots$; $\tau_0=0$) are identically distributed, mutually independent random variables with distribution function $P[\vartheta \leqq x]=F(x)$ where

$$F(x)=\begin{cases} 1-e^{-\lambda x} & \text{if } x\geqq 0, \\ 0 & \text{if } x<0. \end{cases}$$

At the first counter the customers will be served by a single server in the order of arrival and the service times are identically distributed, mutually independent, positive random variables with the distribution function $G(x)$. Put $\varphi(s)=\int_0^\infty e^{-sx}dG(x)$ and $\frac{1}{\mu_1}=\int_0^\infty x dG(x)$. Define $\rho_1=\lambda/\mu_1$.

After finishing service at the first counter the customers move on to the second counter and they are served in the order of arrival to the second counter. The service times at the second counter are identically distributed, mutually independent random variables with the distribution function $H(x)$ where

$$H(x)=\begin{cases} 1-e^{-\mu_2 x} & \text{if } x\geqq 0, \\ 0 & \text{if } x<0. \end{cases}$$

Define $\rho_2=\lambda/\mu_2$.

The sequences of $\{\tau_n\}$, the service times at the first and the second counter are mutually independent. The number of customers in front of each counter is unlimited, losses are not allowed.

We are interested here in the investigation of the probability law of the queue size and the waiting time at the second counter in the steady state. In the special case where

$$G(x)=\begin{cases} 1-e^{-\mu_1 x} & \text{if } x\geqq 0, \\ 0 & \text{if } x<0, \end{cases}$$

the distribution of the queue size has been given by R. R. P. Jackson [1]. In his paper he notes that although the size of the second queue would seem to be dependent upon the output from the first counter, in the steady state each queue settles down in apparent independence. The validity of his note has been proved by P. J. Burke [2], E. Reich [3], T. Homma [4] and P. D. Finch [5].

J. Sacks [6] has characterized ergodicity of queues in series on the basis of general assumptions as to interarrival time and service time at each counter.

Under ergodic condition, max $(\rho_1, \rho_2)<1$, as given by Sacks we shall derive the stationary distribution $\{\pi_n, n=0, 1, 2,\ldots\}$ of the second queue size at instants just before arrivals to the second conter and also we shall determine the distribution $W(t)(t\geqq 0)$ of the waiting time for an arbitrary customer arriving at the second counter (see Theorem 1 and Theorem 2). The analysis in this paper has referenc to the imbedded Markov chain.

I am indebted to Dr. Morimura for his encouragement.

II. THE QUEUE SIZE DISTRIBUTION AT THE SECOND COUNTER

Define

$$a_n=(1-\rho_1)\int_0^\infty e^{-\mu_2 x}\frac{(\mu_2 x)^n}{n!}dG(x)*F(x)+\rho_1\int_0^\infty e^{-\mu_2 x}\frac{(\mu_2 x)^n}{n!}dG(x),$$

and put $A(z)=\sum_{n=0}^{\infty}a_n z^n$ for $|z|\leqq 1$. Then we have that

$$A(z)=\int_0^\infty e^{-\mu_2 x(1-z)}d[(1-\rho_1)G(x)*F(x)+\rho_1 G(x)]$$

$$=\varphi[\mu_2(1-z)]\frac{\lambda+\mu_2\rho_1(1-z)}{\lambda+\mu_2(1-z)}.$$

We assume max $(\rho_1,\rho_2)<1$. Then it follows from Rouche's thorem that the equation

$$\mathrm{A}(z)=z$$

has exactly one root in the interval $0<z<1$. Denote the root by $z=\omega$.

Theorem 1.

$$\pi_n=(1-\omega)\omega^n \quad (n=0,\ 1,\ 2,\ldots).$$

Proof. In the steady state the distribution function of an interdeparture time is given by

$$(1-\rho_1)G(x)*F(x)+\rho_1 G(x).$$

Then the sequence of the second queue size at instants just before arrivals to the second counter forms an aperiodic, irreducible Markov chain with transition probabilities

$$P_{ij}=\begin{cases}1-\sum_{n=0}^{i}a_n & \text{if } j=0,\\ a_{i+1-j} & \text{if } 0<j\leqq i+1,\\ 0 & \text{if } i+1<j, \qquad (i,\ j=0,\ 1,\ 2,\ \ldots)\end{cases}$$

By the well-konown method, being similar to that in Kendall [7], we obtain the required result.

Corollary. The expected number L in the second queue is given by

$$L=\frac{\omega}{1-\omega}.$$

Example 1. If the service-time distribution at the first counter is negative exponential, then $\varphi(s)=\mu_1/(\mu_1+s)$. Therefore $A(z)=\mu_1[\lambda+\mu_2\rho_1(1-z)]/[\mu_1+\mu_2(1-z)][\lambda+\mu_2(1-z)]$. Therefore the root ω of the equation $A(z)=z$ is seen to be $\omega=\rho_2$. Thus we have $\pi_n=(1-\rho_2)\rho_2{}^n$ $(n=0,\ 1,\ 2,\ldots)$ and $L=\frac{\rho_2}{1-\rho_2}$. The result is coincide with that obtained already by Jackson [1].

Example 2. If the service-time distribution at the first counter is Erlangian E_k, that is

$$G(x)=1-\sum_{j=0}^{k-1}\frac{(\mu_1 kx)^j}{j!}e^{-\mu_1 kx} \quad (x\geqq 0),$$

then $\varphi(s)=\left(\frac{\mu_1 k}{\mu_1 k+s}\right)^k$. Therefore ω is the root of the equation

$$\left[\frac{\mu_1 k}{\mu_1 k+\mu_2(1-z)}\right]^k\frac{\lambda+\mu_2\rho_1(1-z)}{\lambda+\mu_2(1-z)}=z$$

in the interval $0<z<1$.

Example 3. If the service-time distribution at the first counter is regular, that is

$$G(x)=\begin{cases}1 & \text{if } x=\frac{1}{\mu_1},\\ 0 & \text{otherwise,}\end{cases}$$

then $\varphi(s)=e^{-\frac{s}{\mu_1}}$. Therefore ω is the root of the equation

$$e^{-\frac{\mu_2}{\mu_1}(1-z)} \cdot \frac{\lambda+\mu_2\rho_1(1-z)}{\lambda+\mu_2(1-z)}=z$$

in the interval $0<z<1$.

In the last section we shall consider the relationship existing between the above three cases.

III. THE WAITING TIME DISTRIBUTION AT THE SECOND COUNTER

We consider the waiting time of the customer arriving at the second counter in the steady state. The waiting time is the time which elapses between the instant at which the customer arrives at the second counter and the instant at which its service begins. Dentoe by $W(x)$ the distribution function of the waiting time.

Thorem 2.

$$W(x)=1-\omega \exp [-(1-\omega)\mu_2 x] \quad (x\geqq 0).$$

Proof. Put $\Omega(s)=\int_0^\infty e^{-sx}dW(x)$. Each customer has a service time whose distribution is exponential with the Laplace transform $\mu_2/(\mu_2+s)$. If an arriving customer finds n customers in the second queue, the waiting time of the customer will have Laplace transform $[\mu_2/(\mu_2+s)]^n$. Therefore

$$\Omega(s)=\sum_{n=1}^{\infty}\pi_n\left(\frac{\mu_2}{\mu_2+s}\right)^n=\frac{\mu_2\omega(1-\omega)}{s+\mu_2(1-\omega)}.$$

Therefore we have the required result from $\Omega(s)$ by inversion.

Corollary. The expected waiting time W^* of the arriving customer at the second counter is given by

$$W^*=\frac{\omega}{\mu_2(1-\omega)}.$$

Proof. $$W^*=-\lim_{s\to 0}\Omega'(s)=\frac{\omega}{\mu_2(1-\omega)}.$$

Thus we have the identity

$$\mu_2 W^*=L.$$

Denote the expectation of waiting time elapsed in the second system by W. Then

$$W=W^*+\frac{1}{\mu_2}=\frac{1}{\mu_2(1-\omega)}.$$

Hence we get

$$\omega\mu_2 W = L.$$

We are interested to compare the above fact with the results due to Morimura [8] and Little [9].

Now we consider the case where the service-time distribution at the first counter is Erlangian, E_k. Let us denote the root ω of the equation in Example 2 by ω_k. For fixed μ_1, μ_2 and z, the function $\left[\frac{\mu_1 k}{\mu_1 k+\mu_2(1-z)}\right]^k$ is a strictly montone decreasing function of $k (k=1, 2,\dots)$. Also for fixed μ_1, μ_2 and k the function is a strictly monotone increasing function of z $(0\leqq z\leqq 1)$. By these monotone prpoerties we can see that $1>\omega_1>\omega_2>\cdots>0$, which yields $L_1>L_2\cdots>0$, where L_k is the expected number corresponding to ω_k in the second queue. Let us denote the root ω of the equation in Example 3 by ω_D. Then we have $\omega_D=\lim_{k\to\infty}\omega_k$. Thus we have that $1>\omega_1>\omega_2>\cdots w_D$ and $L_1>L_2>\cdots L_D$, where L_D is the expected number corresponding to ω_D in the second queue.

The explicit form of the root ω of the equation $A(z)=z$ can be obtained by Lagrange's expansion, that is,

$$\omega=\sum_{j=1}^{\infty}\frac{(-\mu_2)^{j-1}}{j!}\left(\frac{d^{j-1}}{dy^{j-1}}[A(y)]^j\right)_{y=0}.$$

In the case where $\lambda=1$ and $\rho_2=0.7$, the calculated values of the pair (ω_1, ω_D) and (L_1, L_D) are shown in Table I and II.

Table I
Calculated values of ω_1 and ω_D

μ_1	ω_1	ω_D	ω_2
1	0.7	0.5	0.6
2	0.7	0.6	
3	0.7	0.63	
10	0.7	0.67	

Table II
Calculated values of L_1 and L_D

μ_1	L_1	L_D	L_2
1	2.3	1	1.5
2	2.3	1.5	
3	2.3	1.7	
10	2.3	2.1	

Since $\omega_D\to\omega_1$ as $\mu_1\to\infty$, then $L_D\to L_1$ as $\mu_1\to\infty$.

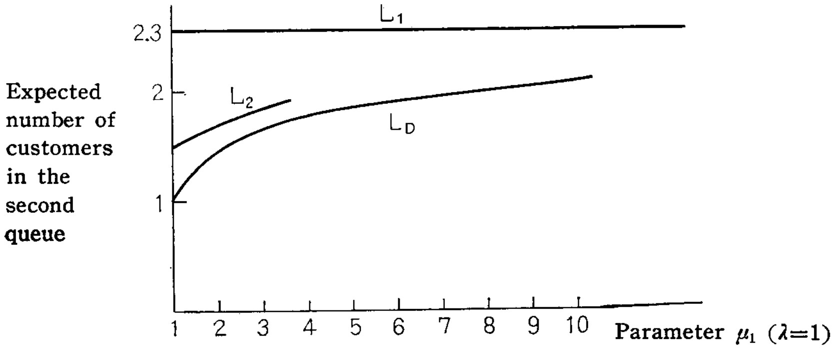

Fig. The expected number of customers in the second queue present as a function of μ_1 for fixed $\lambda=1$ and $\rho_2=0.7$.

As shown in Fig, in the case where ρ_1 is heavy, the value of L_D nearly equals to $\frac{1}{2}L_1$ and in the case where ρ_1 is light. the value of L_D nearly equals to L_1. In view of the graph, the regular servicetime at the first counter seems to be the most effectual one in order to minimize the expected waiting time at the second counter for Erlang's class of the service-time distributions.

REFERENCES

1. R. R. P. Jackson, "Queueing systems with phase type service", O. R. Quart. 5, 109—120 (1954).
2. P. J. Burke, "The output of a queueing system", O. R., Vol. 4, No. 6, 699—704 (1956).
3. E. Reich, "Waiting times when queues are in tandem", A. M. S. Vol. 28, No. 3, 699—704 (1956).
4. T. Homma, "A certain queueing problem of two service stages and the efflux distribution", J. O. R. S. (Japan) Vol. 1, No. 1, 25—36 (1957).
5. P. D. Finch, "The output process of the queueing system M/G/1", J. R. S. S., 375—380 (1960).
6. J. Sacks, "Ergodicity of queues in series", A. M. S., Vol. 31, 579—588 (1960).
7. D. G. Kendall, "Stochastic Processes Ocurring in the Theory of Queues and Their Analysis by Means of the Imbedded Markov Chain", A. M. S., Vol. 24, 338—354 (1954).

8. H. Morimura, "On the relation between the distributions of the queue size and the waiting time", Kōdai Math. Sem. Rep. (Japan) Vol. 14, No. 1, 6—19 (1962).
9. J. D. C. Little, "A proof for the queueing formula: $L=\lambda W$", O. R. Vol. 9, 383—387 (1961).