ON QUEUES IN TANDEM[1]

By Gregory E. Masterson[2] and Seymour Sherman[3]

University of Pennsylvania

1. Introduction and summary. It was in 1952 that D. V. Lindley [1] obtained the steady-state distribution function of the waiting-time in a single-server queue for the case when the interarrival times are independent random variables with identical probability distributions having a finite mean. He applied the same restrictions to the service times. The resulting waiting time distribution was shown to be the solution of an equation of the Wiener-Hopf type.

Queues in tandem have only recently been studied. In 1957, E. Reich [2] found that in "equilibrium" whereas a non-saturating exponential distribution of interarrival times together with an exponential distribution of service times yields a stationary exponential distribution of interdeparture times, "no such simple behaviour can be expected when the service time distributions are even slightly more general." More recently, J. Sacks [3] has found criteria similar to Lindley's for the existence of steady-state distributions of waiting-times in a finite number of single-server queues in tandem.

The motivation for the work reported in this paper originated in a talk given on April 15, 1958, before the Operations Research Seminar of the University of Michigan by G. D. Camp, who made the following intuitive assertion. "Suppose that we imagine an infinite number of identical servers connected in series, and inject any non-saturating input into the first one. Then we expect the statistics of the outputs to change progressively from server to server and since we are dealing with a diffusion process, it seems intuitively obvious that some equilibrium statistics will be approached (the proof is here left to professional mathematicians)." Also, in a talk given on October 16, 1958, before the Institute of Management Science in Philadelphia, he asserted that in this same queueing system, the probability that the time between the ith and the $i + 1$st customers from the nth service point is less than x approaches, as $n \to \infty$, a probability distribution function $F(x)$, i.e. $F(x)$ is monotone increasing, $F(+\infty) = 1$ and $F(-\infty) = 0$.

It is shown below that these assertions are not true, at least as far as the interdeparture time of the first and second customers is concerned. However, in the unique case of constant service time, the assertions are true and statistical equilibrium is achieved by the output from the first server.

2. Glossary of terms and symbols. *Customer*—An object, animate or inanimate, which enters a queueing system requiring service. *Service*—An operation

Received June 6, 1960; revised May 26, 1962.

[1] Revised with support (S. S.) by NSF G 14093.

[2] Now with Sperry-Rand Corporation.

[3] Now with Department of Mathematics, Wayne State University.

performed on a customer. *Server*—An object, animate or inanimate, which services a customer. *Queue*—Backlog of customers waiting for service. A customer enters a queue when he requests service and leaves when the service is complete. *Queue Discipline*—The procedure by which customers are chosen for service from a queue. *Waiting Time*—The interval between the time a customer requests service and the time service commences. *Steady State*—The property that the probabilities are independent of time. *Non-Saturating Input*—An input such that the expected value of the service time is less than the expected value of the interarrival time. $t_i^{(n)}$ is the time at which customer i arrives at the nth service point. $g_i^{(n)} = t_{i+1}^{(n)} - t_i^{(n)}$. $T_i^{(n)}$ is the time at which customer i leaves the nth service point. $S_i^{(n)} = T_{i+1}^{(n)} - T_i^{(n)}$. $T_i^{(n-1)} = t_i^{(n)}$. $R_i^{(n)}$ is the service time for the ith customer at the nth service point. $w_i^{(n)}$ is the time of completion of service for the $(i - 1)$th customer at the nth service point minus the time of arrival of the ith customer at the nth service point. $x \vee 0$ is the maximum of x and 0. $x \wedge 0$ is the minimum of x and 0. $w_i^{(n)} \vee 0$ is the waiting-time for the ith customer at the nth service point.

3. Defining equations of the process. The various possibilities which exist at the nth service point for the ith and $i + 1$st customer are illustrated in the figure together with the defining equations.

From inspection of the figure we may immediately write the following equations:

$$w_{i+1}^{(n)} = -g_i^{(n)} + S_{i-1}^{(n)} + w_i^{(n)}, \tag{1}$$

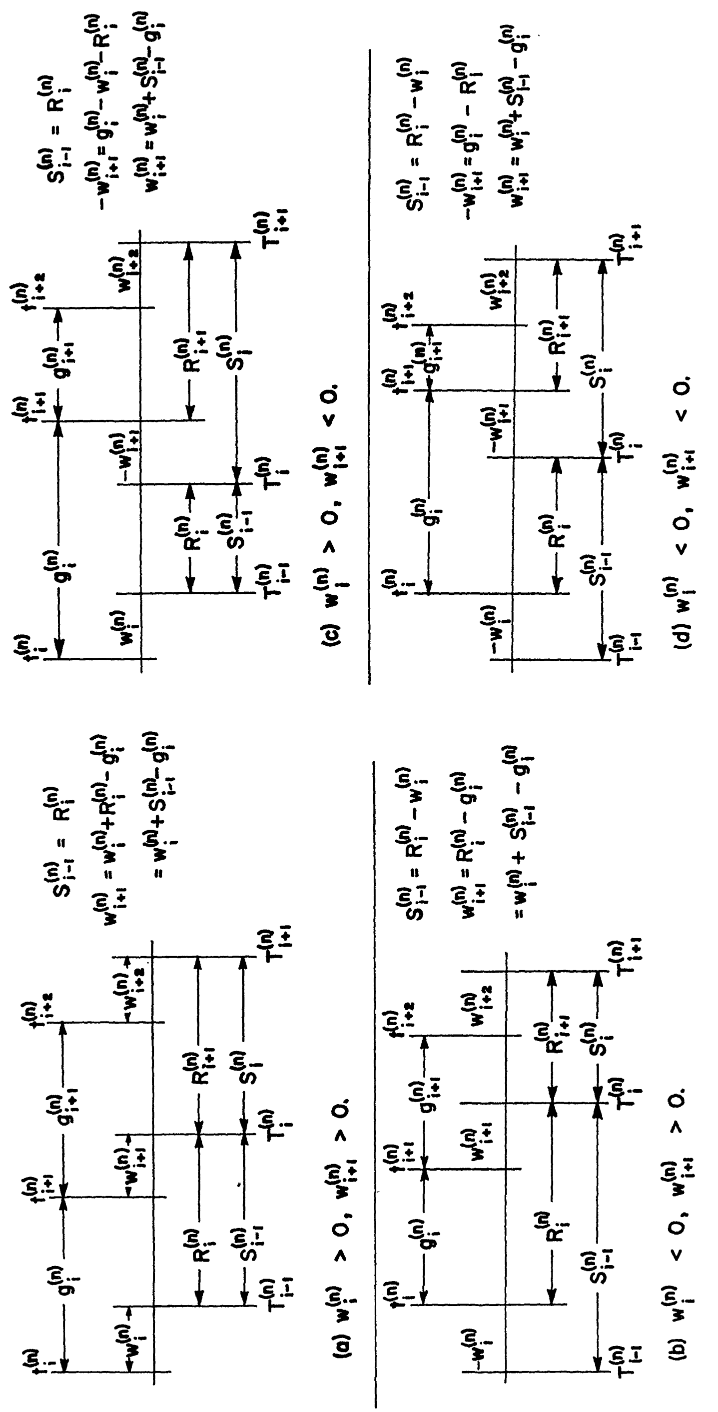
$$S_{i-1}^{(n)} = R_i^{(n)} - (w_i^{(n)} \wedge 0), \tag{2}$$

$$g_i^{(n)} = S_i^{(n-1)}. \tag{3}$$

In addition, the system is empty before the arrival of the first customer. Hence,

$$w_1^{(n)} = 0. \tag{4}$$

The queueing system to be considered consists of an infinite number of identical servers in tandem. The service times for all customers and all servers are independent random variables with identical probability distributions. The distribution is arbitrary except that it has a finite mean. The interarrival times of customers at the input to the system, i.e., at the first server, are also independent random variables with identical probability distributions. Again, the distribution is arbitrary except that it has a finite mean. When a customer has been served, he immediately proceeds to the next server, where he may have to join a queue if that server has not yet finished serving the previous customers. Customers may, of course have to queue at the input to the system. The service discipline is "first come, first served."



(a) $w_i^{(n)} > 0,\ w_{i+1}^{(n)} > 0.$

(b) $w_i^{(n)} < 0,\ w_{i+1}^{(n)} > 0.$

(c) $w_i^{(n)} > 0,\ w_{i+1}^{(n)} < 0.$

(d) $w_i^{(n)} < 0,\ w_{i+1}^{(n)} < 0.$

From (1), (2) and (3) we have

$$w_{i+1}^{(n)} = -S_i^{(n-1)} + R_i^{(n)} + (w_i^{(n)} \vee 0), \tag{5}$$

and (2) and (5) give

$$S_i^{(n)} = R_{i+1}^{(n)} + [S_i^{(n-1)} - R_i^{(n)} - (w_i^{(n)} \vee 0)] \vee 0. \tag{6}$$

4. Limiting distribution for $S_1^{(n)}$. Equations (4) and (6) give

$$S_1^{(n)} = R_2^{(n)} + (S_1^{(n-1)} - R_1^{(n)}) \vee 0. \tag{7}$$

The shortest way to obtain the desired result is as follows. Consider

$$S_1'^{(n)} = (R_2^{(n)} + S_1^{(n-1)} - R_1^{(n)}) \vee 0 \leqq S_1^{(n)} \tag{8}$$

as $R_2^{(n)} \geqq 0$. Therefore

$$S_1'^{(n)} \geqq (S_1'^{(n-1)} + R_2^{(n)} - R_1^{(n)}) \vee 0, \tag{9}$$

where $E(R_2^{(n)} - R_1^{(n)}) = 0$ for $n = 1, 2$, etc., and $S_1^{(0)} = g_1^{(1)} \geqq 0$. Now Lindley ([1], page 278) has shown that for the process

$$u^{(n)} = (u^{(n-1)} + v^{(n)}) \vee 0, \tag{10}$$

where $E(v^{(n)}) = 0$ and $u^{(0)} = 0$, that the probability that $u^{(n)}$ is not more than x tends to zero as n tends to infinity, for any x. Remembering that (10) is identifiable with a random walk we have, a fortiori, from (8) and (9) that $\lim_{n\to\infty} P(S_1^{(n)} \leqq x) = 0$ for all x.

Lindley's result does not apply to the unique case when the variables $v^{(n)}$ equal zero certainly. In our context, this only occurs with constant service time. We treat this case by examining the original equations.

Let $R_i^{(n)} = R$ (a constant), for all i and n. Then (2) gives $S_1^{(n)} \geqq R$ and (7) may be written

$$\begin{aligned} S_1^{(n)} &= R + (S_1^{(n-1)} - R) \\ &= S_1^{(n-1)}. \end{aligned}$$

Thus, in this unique case, a limiting distribution does exist. It is interesting to note that the result following from (9) does not depend on the distribution of interarrival times at the first service point and hence we may remove the restrictions on this random variable, i.e., on $g_1^{(1)}$.

The same problem may be formulated in a different way, which while being more cumbersome does lend itself better to the generalization to the case of $S_i^{(n)}$. Let $M_1^{(n)} = R_2^{(n)} - R_1^{(n)}$; then equation (7) may be written

$$S_1^{(n)} = R_2^{(n)} \vee (S_1^{(n-1)} + M_1^{(n)}).$$

Iterating, we have that

$$\begin{aligned} S_1^{(n)} &= R_2^{(n)} \vee ([R_2^{(n-1)} \vee (S_1^{(n-2)} + M_1^{(n-1)})] + M_1^{(n)}) \\ &= R_2^{(n)} \vee (R_2^{(n-1)} + M_1^{(n)}) \vee (S_1^{(n-2)} + M_1^{(n-1)} + M_1^{(n)}) \\ &= R_2^{(n)} \vee (R_2^{(n-1)} + M_1^{(n)}) \vee ([R_2^{(n-2)} \vee (S_1^{(n-3)} + M_1^{(n-2)})] \\ &\quad + M_1^{(n-1)} + M_1^{(n)}) \\ &= R_2^{(n)} \vee (R_2^{(n-1)} + M_1^{(n)}) \vee (R_2^{(n-2)} + M_1^{(n-1)} + M_1^{(n)}) \\ &\quad \vee (S_1^{(n-3)} + M_1^{(n-2)} + M_1^{(n-1)} + M_1^{(n)}) \\ &= \text{etc.} \end{aligned}$$

The iteration is continued until the last term is reached, namely,

$$(S_1^{(0)} + \sum_{\sigma=1}^{n} M_1^{(\sigma)}) = [(g_1^{(1)} - R_1^{(1)}) + (R_2^{(1)} - R_1^{(2)}) + (R_2^{(2)} - R_1^{(3)}) + \cdots + (R_2^{(n-1)} - R_1^{(n)}) + R_2^{(n)}].$$

Hence,

$$S_1^{(n)} - R_2^{(n)} = \max\nolimits_{0 \leq j \leq n} [A_1^{(n)} - A_1^{(j)}], \tag{11}$$

where $A_1^{(j)} = \sum_{\sigma=1}^{j} p_1^{(\sigma)}$, $A_1^{(0)} = 0$, $p_1^{(\sigma)} = R_2^{(\sigma-1)} - R_1^{(\sigma)}$, $\sigma > 1$, $p_1^{(1)} = g_1^{(1)} - R_1^{(1)}$. Clearly, equation (11) may also be written

$$\begin{aligned} S_1^{(n)} - R_2^{(n)} &= \max\nolimits_{0 \leq j \leq n} \sum_{\sigma=j+1}^{n} p_1^{(\sigma)} \\ &= 0 \vee p_1^{(n)} \vee (p_1^{(n)} + p_1^{(n-1)}) \vee \cdots \vee (p_1^{(n)} + \cdots + p_1^{(1)}) \\ &= Y_1^{(n)} \vee (p_1^{(n)} + p_1^{(n-1)} + \cdots + p_1^{(1)}). \end{aligned} \tag{12}$$

Now $Y_1^{(n)}$ does not contain $p_1^{(1)}$ and hence it is the maximum of partial sums of independent identically distributed random variables of mean zero. As such, it is well known (see for instance [1], page 281), that in the limit as n tends to infinity, $Y_1^{(n)}$ tends to infinity in probability.

5. Limiting distribution for $S_i^{(n)}$, $i > 1$. Writing $M_i^{(n)} = R_{i+1}^{(n)} - R_i^{(n)} - (w_i^{(n)} \vee 0)$, (6) becomes $S_i^{(n)} = R_{i+1}^{(n)} \vee (S_i^{(n-1)} + M_i^{(n)})$. Iterating exactly as before, we obtain

$$S_i^{(n)} - R_{i+1}^{(n)} = \max\nolimits_{0 \leq j \leq n} [A_i^{(n)} - A_i^{(j)}], \tag{13}$$

where $A_i^{(j)} = \sum_{\sigma=1}^{j} [p_i^{(\sigma)} - (w_i^{(\sigma)} \vee 0)]$, $A_i^{(0)} = 0$, $p_i^{(\sigma)} = R_{i+1}^{(\sigma-1)} - R_i^{(\sigma)}$, $\sigma > 1$, $p_i^{(1)} = g_i^{(1)} - R_i^{(1)}$.

Now it has been shown by J. Sacks [3] that

$$(w_i^{(\sigma)} \vee 0) = \max\nolimits_{0 \leq r \leq i-1} [B_{i-1}^{(\sigma)} - B_r^{(\sigma)} + D_r^{(\sigma-1)} - D_{i-1}^{(\sigma-1)}], \tag{14}$$

where $B_r^{(\sigma)} = -\sum_{k=1}^{r} p_k^{(\sigma)}$ and

$$D_r^{(\sigma)} = \max_{0 \leqq j_1 \leqq j_2 \leqq \cdots \leqq j_\sigma \leqq r} [-B_{j_1}^{(1)} - B_{j_2}^{(2)} - \cdots - B_{j_\sigma}^{(\sigma)}].$$

Equation (14) reduces to

$$(15) \qquad (w_i^{(\sigma)} \vee 0) = B_{i-1}^{(\sigma)} - D_{i-1}^{(\sigma-1)} + D_{i-1}^{(\sigma)}.$$

Summing we obtain

$$(16) \qquad \sum_{\sigma=1}^{j} (w_i^{(\sigma)} \vee 0) = D_{i-1}^{(j)} + \sum_{\sigma=1}^{j} B_{i-1}^{(\sigma)}.$$

Now remembering that $p_i^{(\sigma)} = B_{i-1}^{(\sigma)} - B_i^{(\sigma)}$, we have

$$(17) \qquad A_i^{(j)} = \sum_{\sigma=1}^{j} B_{i-1}^{(\sigma)} - \sum_{\sigma=1}^{j} B_i^{(\sigma)} - \sum_{\sigma=1}^{j} (w_i^{(\sigma)} \vee 0).$$

Equations (16) and (17) give

$$(18) \qquad A_i^{(j)} = -D_{i-1}^{(j)} - \sum_{\sigma=1}^{j} B_i^{(\sigma)},$$

which with (13) yield

$$(19) \qquad S_i^{(n)} - R_{i+1}^{(n)} = \max_{0 \leqq j \leqq n} \left[D_{i-1}^{(j)} - \sum_{\sigma=j+1}^{n} B_i^{(\sigma)} - D_{i-1}^{(n)} \right].$$

Equation (19) is explicit, but unfortunately unmanageable. However, it may be rewritten as

$$\begin{aligned}
S_i^{(n)} - R_{i+1}^{(n)} &= \max_{0 \leqq j \leqq n} \left[D_{i-1}^{(j)} - \sum_{\sigma=j+1}^{n} B_i^{(\sigma)} - D_{i-1}^{(n)} \right] \\
&= \max_{0 \leqq j \leqq n} \left[D_{i-1}^{(j)} + \sum_{\sigma=1}^{j} B_i^{(\sigma)} \right] + A_i^{(n)} \\
&= \max_{0 \leqq j \leqq n} \left[D_{i-1}^{(j)} + \sum_{\sigma=1}^{j} B_i^{(\sigma)} \right] + \sum_{\sigma=1}^{n} B_{i-1}^{(\sigma)} - \sum_{\sigma=1}^{n} B_i^{(\sigma)} \\
(20) \qquad & \qquad - \sum_{\sigma=1}^{n} (w_i^{(\sigma)} \vee 0) \\
&= \max_{0 \leqq j \leqq n} \left[D_{i-1}^{(j)} - \sum_{\sigma=j+1}^{n} B_i^{(\sigma)} + \sum_{\sigma=1}^{n} B_{i-1}^{(\sigma)} \right] - \sum_{\sigma=1}^{n} (w_i^{(\sigma)} \vee 0) \\
&= \max_{0 \leqq j \leqq n} \left[\sum_{\sigma=j+1}^{n} p_i^{(\sigma)} + D_{i-1}^{(j)} + \sum_{\sigma=1}^{j} B_{i-1}^{(\sigma)} \right] - \sum_{\sigma=1}^{n} (w_i^{(\sigma)} \vee 0) \\
&\geqq \max_{0 \leqq j \leqq n} \sum_{\sigma=j+1}^{n} p_i^{(\sigma)} - \sum_{\sigma=1}^{n} (w_i^{(\sigma)} \vee 0),
\end{aligned}$$

because $D_{i-1}^{(j)} \geqq -\sum_{\sigma=1}^{j} B_{i-1}^{(\sigma)}$.

Now the first term on the right hand side of (20) goes to infinity in probability as n goes to infinity for the same reasons as before. We would have the same result as in Section 9.1 if we knew that $\lim_{n\to\infty} \sum_{\sigma=1}^{n} (w_i^{(\sigma)} \vee 0) < \infty$. This is not known, although it remains a distinct possibility.

It is logical to examine $S_i^{(n)}$ for the case of constant service times. From (4) and (5) we have $w_2^{(n)} = -S_i^{(n-1)} + R \leqq 0$ from (2). Using (5) repeatedly, we find that $w_i^{(n)} \leqq 0$ for all i and n.

Hence from (6),

$$\begin{aligned} S_i^{(n)} &= R + (S_i^{(n-1)} - R) \vee 0 \\ &= S_i^{(n-1)} \end{aligned}$$

from (2).

It is interesting to note that in the case of constant service times, not only does a limiting distribution exist which is identical with the distribution of interdepartures from the first service point, but also that the interdeparture distribution is a bona fide probability distribution whether the first queue saturates or not.

In conclusion, the following is significant. The technique adopted in the above analysis was to set $i = 1$, and then to examine $\lim_{n\to\infty} S_1^{(n)}$. It was then found that this random variable went to infinity in probability. Unfortunately, this method was not found to be extensible to $i > 1$. Another approach is possible.

Sacks [3] has shown that

$$\begin{aligned} (w_{i+1}^{(n)} \vee 0) = 0 \vee [(w_i^{(n)} \vee 0) - p_i^{(n)} \\ + \sum_{\sigma=1}^{n-1} \{(w_i^{(\sigma)} \vee 0) - (w_{i+1}^{(\sigma)} \vee 0) - p_i^{(\sigma)}\}]. \end{aligned}$$

But, from (3) and (5), we have

$$(w_{i+1}^{(n)} \vee 0) = 0 \vee [(w_i^{(n)} \vee 0) + R_i^{(n)} - S_i^{(n-1)}].$$

Hence,

$$S_i^{(n-1)} = R_{i+1}^{(n-1)} - \sum_{\sigma=1}^{n-1} [(w_i^{(\sigma)} \vee 0) - (w_{i+1}^{(\sigma)} \vee 0) - p_i^{(\sigma)}].$$

Now in the interesting case of an ergodic system, i.e., when

$$E(g_1^{(1)}) > \max_{1\leqq\sigma} E(R_1^{(\sigma)}),$$

we have that a limiting probability distribution function $F^{(\sigma)}$ exists, as $i \to \infty$ for all σ.

If we further assume that, for all σ,

(a) $\int x\, dF^{(\sigma)}(x) < \infty$,

(b) $\lim_{i\to\infty} E[(w_i^{(\sigma)} \vee 0)] = \int x\, dF^{(\sigma)}(x)$, then we have

$$\lim_{i\to\infty} E(S_i^{(n-1)}) = E(R_{i+1}^{(n-1)}) + \sum_{\sigma=1}^{n-1} E(p_i^{(\sigma)}) + \sum_{\sigma=1}^{n-1} \lim_{i\to\infty} E[(w_{i+1}^{(\sigma)} \vee 0) - (w_i^{(\sigma)} \vee 0)]$$

$$= \sum_{\sigma=1}^{n} E(R_{i+1}^{(\sigma-1)}) - \sum_{\sigma=1}^{n-1} E(R_i^{(\sigma)})$$

for all n.

Now the quantities $R_i^{(\sigma)}$, $\sigma \geqq 1$ are independent identically distributed random variables, as are the quantities $R_i^{(0)} = g_i^{(1)}$. Hence,

$$\lim_{i\to\infty} E(S_i^{(n)}) = E(g_1^{(1)}) \text{ for all } n.$$

i.e.,

$$\lim_{n\to\infty} \lim_{i\to\infty} E(S_i^{(n)}) = E(g_1^{(1)}) < \infty$$

while

$$\lim_{n\to\infty} E(S_1^{(n)}) = \infty.$$

Thus renewed interest exists in the behaviour of $\lim_{i\to\infty} \lim_{n\to\infty} E(S_i^{(n)})$.

REFERENCES

[1] Lindley, D. V. (1952). The theory of queues with a single server. *Proc. Cambridge Philos. Soc.* **48** 277–289.

[2] Reich, Edgar (1957). Waiting times when queues are in tandem. *Ann. Math. Statist.* **28** 768–773.

[3] Sacks, J. (1960). Ergodicity of queues in series. Columbia University Memorandum. Work sponsored by the Office of Naval Research, Contract Number Nonr-266(33), Project Number 042-034.