

## DUALITY AND OTHER RESULTS FOR $M/G/1$ AND $GI/M/1$ QUEUES, VIA A NEW BALLOT THEOREM\*

SHUN-CHEN NIU<sup>†</sup> AND ROBERT B. COOPER<sup>‡</sup>

We generalize the classical ballot theorem and use it to obtain direct probabilistic derivations of some well-known and some new results relating to busy and idle periods and waiting times in  $M/G/1$  and  $GI/M/1$  queues. In particular, we uncover a duality relation between the joint distribution of several variables associated with the busy cycle in  $M/G/1$  and the corresponding joint distribution in  $GI/M/1$ . In contrast with the classical derivations of queueing theory, our arguments avoid the use of transforms, and thereby provide insight and term-by-term "explanations" for the remarkable forms of some of these results.

**0. Introduction.** Many classical results in queueing theory were obtained originally by formal mathematical analysis based on skillful use of transforms and inversion techniques, as exemplified by Takács (1962). It is not uncommon for such a result to be somewhat mysterious in that it is composed of familiar-looking terms, but a term-by-term interpretation is hidden by the opacity of the transform analysis. Hence, direct probabilistic or combinatorial derivations are enlightening.

In his treatise *Combinatorial Methods in the Theory of Stochastic Processes*, Takács (1967) generalized the classical ballot theorem and used it to give direct probabilistic and combinatorial derivations of numerous classical results of queueing theory, many of which had been previously obtained by transform methods in Takács (1962). In this paper, we continue in this spirit: We further generalize a ballot theorem of Takács, and we use it to give direct probabilistic derivations of generalizations of several well known (and somewhat mysterious) formulas relating to busy periods and waiting times in  $M/G/1$  and  $GI/M/1$  queues. In the process, we "explain" and interpret these classical formulas, and we obtain other new results as well, including, in particular, a duality relation between the joint distribution of several variables associated with the busy cycle in  $M/G/1$  and the corresponding joint distribution in  $GI/M/1$ .

The outline of our paper is as follows. In §1, we state all of our results and summarize their significance and how they relate to each other; and in §2, we supply the proofs of the assertions of §1.

**1. Results.** The following lemma, which we prove in §2.1, is a slight generalization of a ballot theorem of Takács (1962, p. 231):

**LEMMA.** *Let  $\{v_i, i = 1, 2, \dots, n\}$  be a set of  $n \geq 1$  (possibly dependent) nonnegative integer-valued random variables. If the joint distribution of  $(v_1, \dots, v_{n-1}, v_n)$  is indepen-*

\*Received February 2, 1987; revised December 9, 1987.

AMS 1980 subject classification. Primary: 90B22; Secondary: 60K25.

IAOR 1973 subject classification. Main: Queues.

OR/MS Index 1978 subject classification. Primary: 681 Queues.

Key words. Ballot theorems, duality, busy and idle periods, waiting times,  $M/G/1$  and  $GI/M/1$  queues.

<sup>†</sup>University of Texas at Dallas.

<sup>‡</sup>Florida Atlantic University.

dent of the ordering of  $v_1, \dots, v_{n-1}$ , then for  $0 \leq k \leq n-1$ ,

$$(1) \quad P\{v_1 < 1, v_1 + v_2 < 2, \dots, v_1 + \dots + v_n < n \mid v_1 + \dots + v_n = k\} \\ = \frac{n - k - 1 + E(v_n \mid v_1 + \dots + v_n = k)}{n - 1}$$

(where  $0/0$  is defined to be 1).

We note that since (1) is a conditional probability involving conditioning events that could have probability 0, its right-hand side is only one version of the conditional probability (see, for example, Chapter 4 of Breiman 1968). Similar qualifications apply to other conditional probabilities in this paper.

In §2.2, we apply our lemma to prove that, in  $M/G/1$ , conditional on the events (i) there are  $j$  customers (who have not received any service) present at time 0 when the server begins to work, (ii)  $n - j$  customers arrive during  $(0, x + y)$ , (iii) the first service time has length  $x$ , and (iv) the sum of the other  $n - 1$  service times is equal to  $y$ , then the probability  $p(j, n, x, y)$  that the arrivals and departures occur in such a way that the server first becomes idle at time  $x + y$  (i.e., the interval  $[0, x + y]$  constitutes a  $j$ -busy period) is given by

$$(2) \quad p(j, n, x, y) = \frac{(j - 1) + (n - j)x/(x + y)}{n - 1}.$$

Observe that this conditional probability is insensitive to the form of the service-time distribution function, and does not depend on even its mean value.

It follows easily from (2) that the joint distribution of the number  $K_j$  of customers served during the  $j$ -busy period and its duration  $B_j$ , conditional on the duration  $S_1$  of the initial service, is given by

$$(3) \quad P\{K_j = n, B_j \leq t \mid S_1 = x\} \\ = \int_0^{t-x} \frac{[\lambda(x+y)]^{n-j}}{(n-j)!} e^{-\lambda(x+y)} p(j, n, x, y) dG^{[n-1]}(y),$$

where  $\lambda$  is the arrival rate,  $G$  is the service-time distribution function, and  $G^{[m]}$  is its  $m$ -fold self-convolution.

In particular, we can calculate the unconditional probability  $P\{K_j = n, B_j \leq t\}$  easily from the new result (3) by imagining the addition of a fictitious customer who enters service at time 0 and whose service time has duration 0; that is, the joint probability  $P\{K_j = n, B_j \leq t\}$  is given by the right-hand side of (3) with  $j$  replaced by  $j + 1$ ,  $n$  replaced by  $n + 1$ , and  $x = 0$ :

$$(4) \quad P\{K_j = n, B_j \leq t\} = \frac{j}{n} \int_0^t \frac{(\lambda y)^{n-j}}{(n-j)!} e^{-\lambda y} dG^{[n]}(y).$$

Formula (4), which of course can also be derived directly by evaluating the integral obtained from (3) by removing the condition  $S_1 = x$ , is the well known result given, for example, in Takács (1967, Problem 2, p. 125; solution, p. 230) or Bhat (1968, equation (1.5)).

In §2.3, we apply our lemma to obtain easily (i.e., by elementary probabilistic arguments that use no transforms) the following two results for  $GI/M/1$ . Formula (5)

gives the joint distribution of the number  $K$  served during an ordinary busy period, its duration  $B$ , and the duration  $I$  of the idle period that follows; and (6) gives the joint distribution of  $K$  and the duration of the busy cycle  $B + I$ :

$$(5) \quad P\{K = n, B \leq t, I \leq z\}$$

$$= \iint_{\substack{x+y \leq t \\ x, y \geq 0}} \frac{[\mu(x+y)]^{n-1}}{(n-1)!} e^{-\mu(x+y)} \left(\frac{y}{x+y}\right) [F(y+z) - F(y)] \mu \, dy \, dF^{[n-1]}(x),$$

and

$$(6) \quad P\{K = n, B + I \leq t\}$$

$$= \iiint_{\substack{x+y+z \leq t \\ x, y, z \geq 0}} \frac{[\mu(x+y)]^{n-1}}{(n-1)!} e^{-\mu(x+y)} \left(\frac{y}{x+y}\right) d_z F(y+z) \mu \, dy \, dF^{[n-1]}(x),$$

where  $\mu^{-1}$  is the mean service time and  $F$  is the interarrival-time distribution function, and  $d_z$  indicates that the variable of integration is  $z$ . (Note that whereas  $B$  and  $I$  are independent in  $M/G/1$ , they are dependent in  $GI/M/1$ .)

The application of our lemma to  $GI/M/1$  is justified by the following argument: Imagine first that the interarrival-time distribution and the service-time distribution are interchanged in an  $M/G/1$  queue (the resulting  $GI/M/1$  queue is called the *dual* or the *inverse*; see Bhat 1968, Prabhu 1965, or Takács 1967); then, in any busy period, relabel each arrival epoch as a departure epoch and vice versa, look backward in time, and apply formula (2) to the resulting  $GI/M/1$  busy period. (We note that the idea of duality and time-reversal was used in a similar way by Bhat 1968, pp. 45–46.)

Formulas (5) and (6), which appear to be new, can of course be used to calculate the marginal distributions of  $K$ ,  $B$ ,  $I$ , and  $B + I$ . In particular, setting  $z = \infty$  in (5) yields

$$(7) \quad P\{K = n, B \leq t\}$$

$$= \iint_{\substack{x+y \leq t \\ x, y \geq 0}} \frac{[\mu(x+y)]^{n-1}}{(n-1)!} e^{-\mu(x+y)} \left(\frac{y}{x+y}\right) [1 - F(y)] \mu \, dy \, dF^{[n-1]}(x).$$

(Formula (7) agrees with equation (16) in Bhat 1967; it is stated, but not derived, in the solution, p. 236, for Problem 11, p. 125, of Takács 1967, where it corrects two typographical errors in equation (30), p. 122, of Takács 1962.) Also, letting  $t$  go to infinity in (5) yields the joint distribution of the number served during a busy period and the duration  $I$  of the following idle period, a result that appears to be new:

$$(8) \quad P\{K = n, I \leq z\}$$

$$= \int_0^\infty \int_0^\infty \frac{[\mu(x+y)]^{n-1}}{(n-1)!} e^{-\mu(x+y)} \left(\frac{y}{x+y}\right) [F(y+z) - F(y)] \mu \, dy \, dF^{[n-1]}(x).$$

Summing over  $n$  in (6) and (8) yields expressions, apparently new, for the distributions of  $B + I$  and  $I$ . (We will shortly, in (15) and (16) below, give much simpler formulas for the distribution of  $I$ .) Interestingly, it is known (Takács 1962, p. 124;

Takács 1967, p. 236) that the distribution of  $B$  is given by

$$(9) \quad P\{B \leq t\} = \sum_{n=1}^{\infty} \frac{(\mu t)^n}{n!} e^{-\mu t} \frac{1}{t} \int_0^t [1 - F^{[n]}(y)] dy;$$

in principle, one can obtain (9) by summing over  $n$  in (7) (but it's not easy). In §2.4, we derive (9) by an argument similar to our derivation of (5) and (6).

In §2.5, we consider a duality or inverse relation between  $M/G/1$  and  $GI/M/1$ : Let  $\tilde{K}_{M/G/1}$  and  $\tilde{B}_{M/G/1}$  be, respectively, the number of customers served during an ordinary busy period and its duration in  $M/G/1$  when the first service time  $\tilde{S}_1$  of the busy period has the distribution of the forward recurrence time of a service time, i.e.,

$$(10) \quad P\{\tilde{S}_1 \leq t\} = \mu \int_0^t [1 - G(y)] dy;$$

and let  $K_{GI/M/1}$  and  $B_{GI/M/1}$  be the corresponding random variables for the dual  $GI/M/1$  queue. Then, we show that, remarkably,

$$(11) \quad \lambda P\{\tilde{K}_{M/G/1} = n, \tilde{B}_{M/G/1} \leq t\} = \mu P\{K_{GI/M/1} = n, B_{GI/M/1} \leq t\},$$

a result that is valid even when  $\lambda/\mu \geq 1$ .

Actually, our primary duality relation is stronger than (11): Let  $\tilde{A}_{M/G/1}$  be the "age" of the interval sampled when the equilibrium renewal process whose interval lengths have distribution function  $G$  is interrupted at random (i.e., if  $\tilde{A}_{M/G/1} + \tilde{S}_1$  has distribution function  $\hat{G}$ , then  $\hat{G}(t) = \mu \int_0^t x dG(x)$ ), then

$$(12) \quad \begin{aligned} \lambda P\{\tilde{K}_{M/G/1} = n, \tilde{B}_{M/G/1} \leq t, \tilde{A}_{M/G/1} \leq z\} \\ = \mu P\{K_{GI/M/1} = n, B_{GI/M/1} \leq t, I_{GI/M/1} \leq z\}, \end{aligned}$$

of which (11) is the special case with  $z = \infty$ . Apart from its startling symmetry, the duality relation (12) has the practical value that, in any particular application, it allows one to work with either  $M/G/1$  or  $GI/M/1$ , whichever is more convenient.

An easy consequence of (11), which we demonstrate in §2.6, is that in the equilibrium  $M/G/1$  queue with nonpreemptive LIFO queue discipline, the waiting time  $W$  has distribution function

$$(13) \quad P\{W \leq t\} = 1 - \rho + \sum_{n=1}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} \frac{1}{t} \int_0^t [1 - G^{[n]}(y)] dy$$

(where  $\rho \equiv \lambda/\mu < 1$ ), a well-known (see, e.g., Cooper 1981, p. 235) but now less mysterious result. In particular, this "explains" the startling similarity of form between (9), which describes busy periods in  $GI/M/1$ , and (13), which describes waiting times in nonpreemptive LIFO  $M/G/1$ .

In §2.7, we show that (12) implies

$$(14) \quad \begin{aligned} P\{\tilde{K}_{M/G/1} = n, \tilde{B}_{M/G/1} \leq t, \tilde{A}_{M/G/1} \leq z | \tilde{B}_{M/G/1} < \infty\} \\ = P\{K_{GI/M/1} = n, B_{GI/M/1} \leq t, I_{GI/M/1} \leq z | B_{GI/M/1} < \infty\}. \end{aligned}$$

According to (14), whenever the busy periods are finite, then the busy cycle in

GI/M/1 and its reversed M/G/1 dual (with exceptional first service time, described by (10)) are stochastically identical. When  $\lambda = \mu$ , the conditions  $\tilde{B}_{M/G/1} < \infty$  and  $B_{GI/M/1} < \infty$  are both vacuous; but when  $\lambda > \mu$  ( $\lambda < \mu$ ), then the condition on the left-hand (right-hand) side of (14) is not vacuous.

In §2.8, we show that the duality relation (12) allows us to describe completely the idle periods in GI/M/1: If the GI/M/1 queue is unstable (i.e., if the arrival rate  $\mu$  in the dual GI/M/1 queue is not less than the service rate  $\lambda$ ) then

$$(15) \quad P\{I_{GI/M/1} \leq z \mid B_{GI/M/1} < \infty\} = P\{\tilde{A}_{M/G/1} \leq z\},$$

where  $\tilde{A}_{M/G/1}$  has the same distribution as  $\tilde{S}_1$ , given by (10); on the other hand, if  $\lambda \geq \mu$ , then we have the interesting formula, whose explicit statement apparently is new,

$$(16) \quad P\{I_{GI/M/1} \leq z\} = \lambda \int_0^\infty e^{-(1-\omega)\lambda x} [G(x+z) - G(x)] dx,$$

where  $\omega$  is the smallest root of the equation

$$(17) \quad \omega = \int_0^\infty e^{-(1-\omega)\lambda x} dG(x)$$

(see, e.g., (14.11), p. 270 of Cooper 1981). Of course, (15) and (16) can be shown to be in agreement when  $\lambda = \mu$ .

Also in §2.8, we observe that (16) can be obtained by an intuitive argument that relates the idle period to the waiting times, and which, it turns out, is a specialization of a result obtained by Marshall (1968) for GI/G/1, and discussed further by Evans (1968).

Finally, in §2.9, we show that (12) can serve as the starting point for a new constructive proof of the remarkable inversion of the Pollaczek-Khintchine transform, first found by Beneš (1957), for the waiting time  $W$  in M/G/1 FIFO:

$$(18) \quad P\{W \leq t\} = \sum_{j=0}^\infty (1-\rho)\rho^j \tilde{G}^{[j]}(t),$$

where  $\rho = \lambda/\mu < 1$  and  $\tilde{G}$  is given by the right-hand side of (10). This complements work of Cooper and Niu (1986), Fakinos (1981, 1986, 1987), Kelly (1976, 1979), Niu (1988), Shanthikumar and Sumita (1986), and Yamazaki (1982, 1984).

**2. Proofs.**

2.1. *Proof of lemma.* We give an inductive proof that is a minor modification of one given by Takács (1962, pp. 231-232).

First, as induction base, note that (1) obviously holds for  $n = 1$  and  $k = 0$ . Next, for any  $n > 1$  and  $1 \leq k \leq n - 1$  (the case for  $k = 0$  is trivial), the left-hand side of (1), by conditioning on  $\sum_{i=1}^k v_i$ , can be written as

$$(19) \quad P\{v_1 < 1, v_1 + v_2 < 2, \dots, v_1 + \dots + v_n < n \mid v_1 + \dots + v_n = k\}$$

$$= \sum_{j=0}^k P\left\{ \sum_{i=1}^r v_i < r \text{ for } r = 1, 2, \dots, n \mid \sum_{i=1}^k v_i = j, \sum_{i=1}^n v_i = k \right\} P\left\{ \sum_{i=1}^k v_i = j \mid \sum_{i=1}^n v_i = k \right\}.$$

To evaluate (19) further, we consider two cases:

*Case 1:*  $0 \leq j < k$ . Observe that, conditional on the event  $\sum_{i=1}^k v_i = j$  and  $\sum_{i=1}^n v_i = k$ , the joint distribution of  $(v_1, v_2, \dots, v_k)$  is independent of the ordering of  $v_1, v_2, \dots, v_k$ , i.e., the vector is fully exchangeable. Hence,

$$\begin{aligned} & P\left\{\sum_{i=1}^r v_i < r \text{ for } r = 1, 2, \dots, n \mid \sum_{i=1}^k v_i = j, \sum_{i=1}^n v_i = k\right\} \\ &= P\left\{\sum_{i=1}^r v_i < r \text{ for } r = 1, 2, \dots, k \mid \sum_{i=1}^k v_i = j, \sum_{i=1}^n v_i = k\right\} \\ &= \left[ k - j - 1 + E\left(v_k \mid \sum_{i=1}^k v_i = j, \sum_{i=1}^n v_i = k\right) \right] / (k - 1) \\ &= [k - j - 1 + (j/k)] / (k - 1) \\ &= 1 - \frac{j}{k}, \end{aligned}$$

where the second equality is due to the induction hypothesis (note that  $j < k < n$ ), and the third equality follows from

$$(20) \quad E\left(v_k \mid \sum_{i=1}^k v_i = j, \sum_{i=1}^n v_i = k\right) = \frac{1}{k} E\left(\sum_{i=1}^k v_i \mid \sum_{i=1}^k v_i = j, \sum_{i=1}^n v_i = k\right) = \frac{j}{k}.$$

*Case 2:*  $j = k$ . We have, as in Case 1,

$$P\left\{\sum_{i=1}^r v_i < r \text{ for } r = 1, 2, \dots, n \mid \sum_{i=1}^k v_i = j, \sum_{i=1}^n v_i = k\right\} = 0 = 1 - \frac{j}{k},$$

since the event  $\sum_{i=1}^k v_i < k$  cannot occur under the stated condition.

Combining Cases 1 and 2, we see that (19) simplifies to

$$\begin{aligned} (21) \quad & P\{v_1 < 1, v_1 + v_2 < 2, \dots, v_1 + \dots + v_n < n \mid v_1 + \dots + v_n = k\} \\ &= \sum_{j=0}^k \left(1 - \frac{j}{k}\right) P\left\{\sum_{i=1}^k v_i = j \mid \sum_{i=1}^n v_i = k\right\} \\ &= 1 - \frac{1}{k} E\left(\sum_{i=1}^k v_i \mid \sum_{i=1}^n v_i = k\right). \end{aligned}$$

Finally, to evaluate the right-hand side of (21), observe that, by conditioning on  $v_n$ , we

have

$$\begin{aligned} E\left(\sum_{i=1}^k v_i \mid \sum_{i=1}^n v_i = k\right) &= \sum_{j=0}^k E\left(\sum_{i=1}^k v_i \mid v_n = j, \sum_{i=1}^n v_i = k\right) P\left\{v_n = j \mid \sum_{i=1}^n v_i = k\right\} \\ &= \sum_{j=0}^k k \left(\frac{k-j}{n-1}\right) P\left\{v_n = j \mid \sum_{i=1}^n v_i = k\right\} \\ &= \frac{k}{n-1} \left[ k - E\left(v_n \mid \sum_{i=1}^n v_i = k\right) \right], \end{aligned}$$

where the second equality is due to the fact that, conditional on  $v_n = j$  and  $\sum_{i=1}^n v_i = k$ , the vector  $(v_1, v_2, \dots, v_{n-1})$  is fully exchangeable (similar to (20)); and therefore (21) reduces to (1), completing the proof.

2.2. *Proof of (2).* Let  $S_1$  be the duration of the initial service after time 0, and  $S_i$  for  $i = 2, \dots, n$  be the durations of the subsequent service times (iid with distribution  $G$ ). Denote by  $N(t)$ ,  $t \geq 0$ , the number of arrivals in the interval  $[0, t]$ ; and define for  $m = 1, 2, \dots, n$ ,

$$(22) \quad v_m = \begin{cases} \left[ N(S_1 + \dots + S_{n-m+1}) - N(S_1 + \dots + S_{n-m}) \mid \right. \\ \left. S_1 = x, \sum_{i=2}^n S_i = y, N(x+y) = n-j \right] & \text{if } m \neq n, \\ \left[ N(S_1) \mid S_1 = x, \sum_{i=2}^n S_i = y, N(x+y) = n-j \right] & \text{if } m = n. \end{cases}$$

In other words,  $v_1$  is the number of arrivals during the last service interval,  $v_2$  is the number of arrivals during the service interval immediately before the last, and so on, given that  $S_1 = x$ ,  $\sum_{i=2}^n S_i = y$ , and  $N(x+y) = n-j$ . Since the arrival process is Poisson, then, conditional on the number of arrivals in  $(0, x+y)$  being fixed, the arrival times are distributed as if they were selected independently from a population uniformly distributed over  $(0, x+y)$  (see, for example, Cooper 1981, p. 54). It follows that the vector  $(v_1, \dots, v_{n-1}, v_n)$  defined by (22) satisfies the condition stated in the lemma; and in particular,  $v_n$  has the binomial distribution with parameters  $n-j$  and  $x/(x+y)$ . Finally, note that the interval  $[0, x+y]$  constitutes a busy period if and only if  $\sum_{i=1}^r v_i < r$  for all  $r = 1, 2, \dots, n$  (arrivals must occur fast enough). Hence, (2) follows from (1) by substituting  $k = n-j$  and  $E(v_n \mid v_1 + \dots + v_n = n-j) = (n-j)x/(x+y)$ , and the proof is complete.

2.3. *Proofs of (5) and (6).* We begin with (5). For any  $x \geq 0$ ,  $y \geq 0$ , and  $x+y \leq t$ , interpret  $x$  as the time of the  $(n-1)$ th arrival and  $x+y$  as the time of the  $n$ th departure after time 0, assuming that the busy period starts with one initial customer; note that the first of these two events occurs with "probability"  $dF^{[n-1]}(x)$ , and that the second occurs with conditional "probability"

$$\frac{[\mu(x+y)]^{n-1}}{(n-1)!} e^{-\mu(x+y)} \mu \, dy,$$

given (explaining the appearance of  $dy$ , rather than  $d(x+y)$ , in the above expression)

that the first occurs. With these interpretations, the event  $\{K = n, B \leq t, I \leq z\}$  occurs if and only if (i) the  $n$ th interarrival time  $T_n$  satisfies  $y < T_n \leq y + z$  ( $z \geq 0$ ), and (ii) the conditional arrival and departure times in the interval  $[0, x + y]$  are such that the interval constitutes a busy period. Since event (i) has probability  $F(y + z) - F(y)$ , we see that (5) will follow if event (ii) has probability  $y/(x + y)$ . Now, look backward in time in the interval  $[0, x + y]$ , relabel each arrival epoch as a departure epoch and each departure epoch as an arrival epoch, and then interpret the *reversed* interval as an  $M/G/1$  busy period with first service time  $S_1 = y$ . Then (2) applies with  $j = 1$  and  $x$  and  $y$  reversed; and (5) is proved.

The argument for (6) is very similar: Interpret  $x + y + z$  ( $\leq t$ ) as the end of the busy cycle, and note that the conditional "probability" of this event is given by  $d_z F(y + z)$  whenever  $x$  and  $x + y$  have the same interpretations as in the proof of (5).

2.4. *Proof of (9).* Let  $D(t)$ ,  $t \geq 0$ , be the number of Poisson events, each of which is a *potential* departure epoch, at rate  $\mu$  in  $(0, t)$ ; and observe that the right-hand side of (9) can be interpreted as  $E\{\min[\sum_{i=1}^{D(t)} T_i, t]\}/t$ , where  $T_i$  denotes the  $i$ th ( $i \geq 1$ ) interarrival interval after time 0. It follows that (9) is equivalent to

$$\begin{aligned} (23) \quad P\{B > t\} &= 1 - \frac{1}{t} E \left\{ \min \left[ \sum_{i=1}^{D(t)} T_i, t \right] \right\} \\ &= \frac{1}{t} E \left\{ t - \min \left[ \sum_{i=1}^{D(t)} T_i, t \right] \right\} \\ &= \frac{1}{t} E \left\{ \left( t - \sum_{i=1}^{D(t)} T_i \right)^+ \right\}, \end{aligned}$$

where  $x^+ \equiv \max(x, 0)$  for any real number  $x$ . We shall show that the right-hand side of (23) is indeed the probability that the busy period does *not* end by time  $t$ . More precisely, we will prove that

$$(24) \quad P \left\{ B > t \mid D(t), \sum_{i=1}^{D(t)} T_i \right\} = \frac{1}{t} \left( t - \sum_{i=1}^{D(t)} T_i \right)^+,$$

from which (23) follows by taking expectations.

To prove (24), we consider two cases:

*Case 1:*  $\sum_{i=1}^{D(t)} T_i \geq t$ . Since the total number of customers to be served in  $[0, t)$  is no more than the number  $D(t)$  of potential departures, the busy period must be over by time  $t$  (in fact, it ends no later than the last potential departure epoch). Hence,  $P\{B > t \mid D(t), \sum_{i=1}^{D(t)} T_i\} = 0$ , and since  $(t - \sum_{i=1}^{D(t)} T_i)^+ = 0$  when  $\sum_{i=1}^{D(t)} T_i \geq t$ , therefore (24) is true for Case 1.

*Case 2:*  $\sum_{i=1}^{D(t)} T_i < t$ . Let  $t + \delta$  be the first potential departure epoch *after* time  $t$ , where the random variable  $\delta$  is exponentially distributed with rate  $\mu$ . If the busy period is to last beyond time  $t$ , then  $t + \delta$  would be the departure epoch assigned to the  $D(t)$ th arrival after time 0; the *probability* of the event  $\{B > t\}$  is, however, independent of the value of  $\delta$ . We could, therefore, for convenience, let  $\delta = 0$ ; that is, schedule a (fictitious) departure at time  $t$ . Now, look backward in time in the interval  $[0, t]$ , relabel arrivals as departures and departures as arrivals, and then interpret the reversed interval as an  $M/G/1$  busy period with first service time  $S_1 = t - \sum_{i=1}^{D(t)} T_i$ .

Hence,

$$P\left\{B > t \mid D(t), \sum_{i=1}^{D(t)} T_i\right\} = p\left(1, D(t) + 1, t - \sum_{i=1}^{D(t)} T_i, \sum_{i=1}^{D(t)} T_i\right),$$

where the right-hand side is given by (2), which reduces to (24). Thus, (24) is true also for Case 2, and the proof is complete.

2.5. *Proof of (12).* First, observe that the vector  $(\tilde{K}_{M/G/1}, \tilde{B}_{M/G/1})$  and the variable  $\tilde{A}_{M/G/1}$  are conditionally independent for any given fixed value of  $\tilde{S}_1$ . Our strategy is to compute the probability  $P\{\tilde{K}_{M/G/1} = n, \tilde{B}_{M/G/1} \leq t, \tilde{A}_{M/G/1} \leq z\}$  by conditioning on  $\tilde{S}_1$ . Formally, for  $n \geq 1, t \geq 0, z \geq 0$ , and any positive  $\lambda$  and  $\mu$ , we have

$$\begin{aligned} &P\{\tilde{K}_{M/G/1} = n, \tilde{B}_{M/G/1} \leq t, \tilde{A}_{M/G/1} \leq z\} \\ &= \int_0^t P\{\tilde{K}_{M/G/1} = n, \tilde{B}_{M/G/1} \leq t \mid \tilde{S}_1 = x\} P\{\tilde{A}_{M/G/1} \leq z \mid \tilde{S}_1 = x\} dP\{\tilde{S}_1 \leq x\} \\ &= \int_0^t \int_0^{t-x} \frac{[\lambda(x+y)]^{n-1}}{(n-1)!} e^{-\lambda(x+y)} \left(\frac{x}{x+y}\right) \\ &\quad \times dG^{[n-1]}(y) \frac{G(x+z) - G(x)}{1 - G(x)} \mu [1 - G(x)] dx \\ &= \frac{\mu}{\lambda} \iint_{\substack{x+y \leq t \\ x, y \geq 0}} \frac{[\lambda(x+y)]^{n-1}}{(n-1)!} e^{-\lambda(x+y)} \left(\frac{x}{x+y}\right) [G(x+z) - G(x)] dG^{[n-1]}(y) \lambda dx, \end{aligned}$$

where the second equality is due to (3) (with  $j = 1$ ) and (10). Comparison of the last integral with (5) (replacing  $\mu$  by  $\lambda$  and  $F$  by  $G$ , and interchanging  $x$  and  $y$ ) yields (12), completing the proof.

2.6. *Proof of (13).* In nonpreemptive LIFO  $M/G/1$ , the waiting time  $W$  for any blocked customer is the same as the busy period whose first service time is the remainder of the service time in progress when the blocked customer arrives. This remainder has the distribution (10) (Takács 1963; see also Cooper 1981, p. 225, equation (8)), and therefore  $P\{W \leq t \mid W > 0\} = P\{\tilde{B}_{M/G/1} \leq t\}$ . Since  $P\{W \leq t\} = 1 - \rho + \rho P\{W \leq t \mid W > 0\}$ , therefore,

$$\begin{aligned} P\{W \leq t\} &= 1 - \rho + \rho P\{\tilde{B}_{M/G/1} \leq t\} \\ &= 1 - \rho + P\{B_{GI/M/1} \leq t\}, \end{aligned}$$

where the second equality is due to the duality relation (11) (after summing over  $n$ ). Substituting (9), with  $\lambda$  replacing  $\mu$  and  $G$  replacing  $F$ , into the last expression yields (13).

2.7. *Proof of (14).* After summing over  $n$  in (11) and then letting  $t$  go to infinity, we have, for any positive  $\lambda$  and  $\mu$ ,

$$\lambda P\{\tilde{B}_{M/G/\lambda} < \infty\} = \mu P\{B_{GI/M/\lambda} < \infty\},$$

which, together with a well-known result from random-walk theory (Feller 1971, pp. 396–397, Theorem 2), implies that, if  $\rho = \lambda/\mu \leq 1$ , then

$$(25) \quad 1 = P\{\tilde{B}_{M/G/\lambda} < \infty\} = \frac{1}{\rho} P\{B_{GI/M/\lambda} < \infty\};$$

and if  $\rho \geq 1$ , then

$$(26) \quad \rho P\{\tilde{B}_{M/G/\lambda} < \infty\} = P\{B_{GI/M/\lambda} < \infty\} = 1.$$

Combining (25), (26), and (12) yields, for all  $\rho > 0$ ,

$$\begin{aligned} & P\{\tilde{K}_{M/G/\lambda} = n, \tilde{B}_{M/G/\lambda} \leq t, \tilde{A}_{M/G/\lambda} \leq z\} / P\{\tilde{B}_{M/G/\lambda} < \infty\} \\ &= P\{K_{GI/M/\lambda} = n, B_{GI/M/\lambda} \leq t, I_{GI/M/\lambda} \leq z\} / P\{B_{GI/M/\lambda} < \infty\}, \end{aligned}$$

which is equivalent to (14), completing the proof.

2.8. *Proofs of (15) and (16).* Summing over  $n$  in (12) and letting  $t$  go to infinity, we have

$$(27) \quad \lambda P\{\tilde{B}_{M/G/\lambda} < \infty, \tilde{A}_{M/G/\lambda} \leq z\} = \mu P\{B_{GI/M/\lambda} < \infty, I_{GI/M/\lambda} \leq z\}.$$

To prove (15), note that when  $\rho \leq 1$  the event  $\{\tilde{B}_{M/G/\lambda} < \infty\}$  has probability 1 (see (25)), and hence (27) yields

$$\begin{aligned} P\{\tilde{A}_{M/G/\lambda} \leq z\} &= \frac{1}{\rho} P\{B_{GI/M/\lambda} < \infty, I_{GI/M/\lambda} \leq z\} \\ &= P\{B_{GI/M/\lambda} < \infty, I_{GI/M/\lambda} \leq z\} / P\{B_{GI/M/\lambda} < \infty\}, \end{aligned}$$

where the second equality follows from (25); and this is equivalent to (15).

To prove (16), a similar argument (when  $\rho \geq 1$ ) based on (27) and (26) yields

$$(28) \quad P\{I_{GI/M/\lambda} \leq z\} = \rho P\{\tilde{B}_{M/G/\lambda} < \infty, \tilde{A}_{M/G/\lambda} \leq z\}.$$

As in the proof of (12), the right-hand side of (28) can be explicitly evaluated as follows:

$$\begin{aligned} & \rho P\{\tilde{B}_{M/G/\lambda} < \infty, \tilde{A}_{M/G/\lambda} \leq z\} \\ &= \rho \int_0^\infty P\{\tilde{B}_{M/G/\lambda} < \infty | \tilde{S}_1 = x\} P\{\tilde{A}_{M/G/\lambda} \leq z | \tilde{S}_1 = x\} dP\{\tilde{S}_1 \leq x\} \\ &= \rho \int_0^\infty P\{\tilde{B}_{M/G/\lambda} < \infty | \tilde{S}_1 = x\} \frac{G(x+z) - G(x)}{1 - G(x)} \mu [1 - G(x)] dx \\ &= \lambda \int_0^\infty P\{\tilde{B}_{M/G/\lambda} < \infty | \tilde{S}_1 = x\} [G(x+z) - G(x)] dx. \end{aligned}$$

Therefore, (16) will follow if we can show that

$$(29) \quad P\{\tilde{B}_{M/G/1} < \infty \mid \tilde{S}_1 = x\} = e^{-(1-\omega)\lambda x}.$$

Denote by  $B_{M/G/1}$  a standard (i.e., the first service time is not exceptional)  $M/G/1$  busy period, and let  $\gamma \equiv P\{B_{M/G/1} < \infty\}$ . (Note that, since  $\rho \geq 1$ ,  $0 < \gamma \leq 1$ .) It is well known (Takács 1962, p. 60, Theorem 4) that, with  $=^d$  denoting equality in distribution,  $(\tilde{B}_{M/G/1} \mid \tilde{S}_1 = x)$  has the representation

$$(\tilde{B}_{M/G/1} \mid \tilde{S}_1 = x) =^d x + \sum_{i=1}^N \Theta_i,$$

where  $N$  is the number of arrivals during the first service interval  $(0, x)$  (the empty sum is defined to be zero) and the  $\Theta_i$ s are independent and identically distributed versions of  $B_{M/G/1}$ . Therefore, conditional on  $N$  being fixed,  $(\tilde{B}_{M/G/1} \mid N, \tilde{S}_1 = x)$  is finite if and only if  $\Theta_1, \Theta_2, \dots, \Theta_N$  are all finite. It follows that

$$(30) \quad P\{\tilde{B}_{M/G/1} < \infty \mid \tilde{S}_1 = x\} = \sum_{n=0}^{\infty} \frac{(\lambda x)^n}{n!} e^{-\lambda x} \gamma^n = e^{-(1-\gamma)\lambda x}.$$

Since

$$(31) \quad \gamma = P\{B_{M/G/1} < \infty\} = \int_0^{\infty} P\{\tilde{B}_{M/G/1} < \infty \mid \tilde{S}_1 = x\} dG(x),$$

it also follows, by substituting (30) into (31), that

$$\gamma = \int_0^{\infty} e^{-(1-\gamma)\lambda x} dG(x),$$

which is (17); this establishes (29), and the proof is complete.

We can obtain (16) also by an intuitive argument if we are willing to assume as known the equilibrium distribution  $\{\pi_j\}$  of the number of customers found by an arbitrary arrival,  $\pi_j = (1 - \omega)\omega^j$  ( $j = 0, 1, 2, \dots$ ), where  $\omega$  is defined by (17) (and  $\lambda > \mu$ ):

Let  $C$  be the event that a randomly selected customer (the *test customer*) is the last arrival of his busy period, and let  $I$  be the idle time at the end of his busy period. Then

$$P\{C, I \leq z\} = \sum_{j=0}^{\infty} \int_0^{\infty} \pi_j [G(x+z) - G(x)] dP\{R_{j+1} \leq x\},$$

where  $R_{j+1}$  is the sum of the  $j + 1$  independent exponential service times that remain in the busy period when the test customer arrives (i.e.,  $dP\{R_{j+1} \leq x\} = [e^{-\lambda x}(\lambda x)^j/j!]\lambda dx$ ). Since the events  $C$  and  $\{I \leq z\}$  are (intuitively) independent, hence  $P\{I \leq z\} = P\{C, I \leq z\}/P\{C\}$ ; and furthermore,  $P\{C\} = \pi_0 = 1 - \omega$  because for every customer who is last to arrive in his busy period, there is exactly one who is first. Therefore,

$$P\{I \leq z\} = \frac{1}{1 - \omega} \sum_{j=0}^{\infty} \int_0^{\infty} (1 - \omega)\omega^j [G(x+z) - G(x)] \frac{(\lambda x)^j}{j!} e^{-\lambda x} \lambda dx,$$

which reduces to (16). (For further discussion of this argument in the context of  $GI/G/1$ , see Marshall 1968 and Evans 1968.)

2.9. *Proof of (18).* It is well known (see, e.g., Feller 1971, Chapter XII) that in the stable  $GI/G/1$  queue, the waiting time  $W$  has the representation

$$(32) \quad W = \sum_{i=1}^J H_i,$$

where  $J$  is the number of strict ascending ladder epochs that occur in an associated random walk whose step sizes have the same distribution as the difference of a service time and an interarrival time,  $W \equiv 0$  when  $J = 0$ , and the  $H_i$ s are the corresponding independent and identically distributed ladder heights. The distribution of  $J$  is given by

$$(33) \quad P\{J = j\} = (1 - \sigma)\sigma^j \quad (j = 0, 1, 2, \dots),$$

where  $\sigma$  ( $0 < \sigma < 1$ ) is the probability that a strict ascending ladder epoch occurs.

For the stable  $M/G/1$  queue, it follows easily from duality (interchange the interarrival times and the service times) that

$$(34) \quad \sigma = P\{B_{GI/M/1} < \infty\},$$

and that

$$(35) \quad H_i = \sum_{j=1}^i (I_{GI/M/1} | B_{GI/M/1} < \infty)$$

for all  $i \geq 1$  (see, e.g., Kleinrock 1975, pp. 309–311). Comparisons of (34) with (25) and (35) with (15) show, respectively, that (i)  $\sigma = \rho$  and (ii) each of the  $H_i$ s has the distribution (10). Combining (i), (ii), (32), and (33) yields (18), and the proof is complete.

Finally, as an aside, we observe that for the stable  $GI/M/1$  queue, a similar argument also yields the classical result that the waiting time  $W$  has the representation  $W = \sum_{i=1}^J S_i$ , where the  $S_i$ s denote independent and identically distributed service times and  $P\{J = j\} = (1 - \omega)\omega^j$  for  $j \geq 0$ : Note that  $\sigma = P\{B_{M/G/1} < \infty\} = \omega$  (see the proof of (16)), and that, by the memoryless property of the exponential distribution,  $H_i = \sum_{j=1}^i (I_{M/G/1} | B_{M/G/1} < \infty) = \sum_{j=1}^i S_j$  for all  $i \geq 1$ .

### References

- Beneš, V. E. (1957). On Queues with Poisson Arrivals. *Ann. Math. Statist.* **28** 670–677.  
 Bhat, U. N. (1967). Some Explicit Results for the Queue  $GI/M/1$  with Group Service. *Sankhya* **A29** 199–206.  
 ——— (1968). *A Study of the Queueing Systems  $M/G/1$  and  $GI/M/1$* . Springer-Verlag, Berlin and New York.  
 Breiman, L. (1968). *Probability*. Addison-Wesley, Reading, MA.  
 Cooper, R. B. (1981). *Introduction to Queueing Theory*, second edition. North Holland (Elsevier), New York.  
 ——— and Niu, S.-C. (1986). Beneš's Formula for  $M/G/1$ -FIFO "Explained" by Preemptive-Resume LIFO. *J. Appl. Probab.* **23** 550–554.  
 Evans, R. V. (1968). Comments on "Some Inequalities in Queuing" by K. T. Marshall. *Oper. Res.* **16** 666–668.  
 Fakinos, D. (1981). The  $G/G/1$  Queueing System with a Particular Queue Discipline. *J. Roy. Statist. Soc.* **B43** 190–196.  
 ——— (1986). On the Single-Server Queue with the Preemptive-Resume Last-Come-First-Served Queue Discipline. *J. Appl. Probab.* **23** 243–248.

- \_\_\_\_\_ (1987). The Single-Server Queue with Service Depending on Queue Size and with the Preemptive-Resume Last-Come-First-Served Queue Discipline. *J. Appl. Probab.* **24** 758-767.
- Feller, W. (1971). *An Introduction to Probability Theory and Its Applications, Vol. II*, second edition. Wiley, New York.
- Kelly, F. P. (1976). The Departure Process from a Queueing System. *Math. Proc. Cambridge Philos. Soc.* **80** 283-285.
- \_\_\_\_\_ (1979). *Reversibility and Stochastic Networks*. Wiley, Chichester.
- Kleinrock, L. (1975). *Queueing Systems, Vol. I: Theory*. Wiley, New York.
- Marshall, K. T. (1968). Some Inequalities in Queuing. *Oper. Res.* **16** 651-665.
- Niu, S.-C. (1988). Representing Workloads in GI/G/1 Queues Through the Preemptive-Resume LIFO Queue Discipline. *Queueing Systems* **3** 157-178.
- Prabhu, N. U. (1965). *Queues and Inventories: A Study of Their Basic Stochastic Processes*. Wiley, New York.
- Shanthikumar, J. G. and Sumita, U. (1986). On G/G/1 Queues with LIFO-P Service Discipline. *J. Oper. Res. Soc. Japan* **29** 220-231.
- Takács, L. (1962). *Introduction to the Theory of Queues*. Oxford University Press, New York.
- \_\_\_\_\_ (1963). Delay Distributions for One Line with Poisson Input, General Holding Times, and Various Orders of Service. *Bell Systems Tech. J.* **42** 487-503.
- \_\_\_\_\_ (1967). *Combinatorial Methods in the Theory of Stochastic Processes*. Wiley, New York.
- Yamazaki, G. (1982). The GI/G/1 Queue with Last-Come-First-Served. *Ann. Inst. Statist. Math.* **A34** 599-604.
- \_\_\_\_\_ (1984). Invariance Relations of GI/G/1 Queueing Systems with Preemptive-Resume Last-Come-First-Served Queue Discipline. *J. Oper. Res. Soc. Japan* **27** 338-346.

NIU: SCHOOL OF MANAGEMENT, UNIVERSITY OF TEXAS AT DALLAS, POB 830688,  
RICHARDSON, TEXAS 75083-0688

COOPER: DEPARTMENT OF COMPUTER SCIENCE, FLORIDA ATLANTIC UNIVERSITY, POB  
3091, BOCA RATON, FLORIDA 33431-0991