

ON FINITE MOMENTS OF FULL BUSY PERIODS OF GI/G/c QUEUES

SAEED GHAHRAMANI,* *Western New England College*

RONALD W. WOLFF,** *University of California, Berkeley*

Abstract

For a GI/G/c queue, a full busy period is an interval that begins when an arrival finds $c - 1$ customers in the system, and ends when, for the first time after that, a departure leaves behind $c - 1$ customers in the system. We present a probabilistic proof of conditions for full busy periods to have finite moments. For queues that empty, this result may be deduced from results in the literature, but our proof is much easier. For queues that do not empty, our proof still applies, and this result is new.

Keywords: Full busy period; remaining busy period; fast single-server; equilibrium distribution

2000 Mathematics Subject Classification: Primary 60K25

1. Introduction

Consider the GI/G/c queue, which has c servers in parallel, and interarrival and service times that are independent sequences of independent and identically distributed (i.i.d.) random variables. Customers are served in the order of their arrival.

Let S_n be the service time of arrival n and T_n be the time between arrivals n and $n + 1$, where $E(T_n) = 1/\lambda$ and $E(S_n) = 1/\mu$, $n \geq 1$. Let $\rho = \lambda/c\mu$. We assume that $\rho \in (0, 1)$, and call the queue *stable* when this holds.

For the GI/G/c queue, Kiefer and Wolfowitz defined in [4] an ordered vector of work at each server, as found by each arrival, and, for $\rho \in (0, 1)$, showed that this sequence of vectors converges in distribution to a unique stationary distribution, independent of initial conditions. This implies that there is a unique stationary distribution of delay in a queue; let D be a random variable with this distribution. For $\rho \in (0, 1)$, it was shown in [5] that, for a generic service time S ,

$$E(S^{r+1}) < \infty \implies E(D^r) < \infty \quad \text{for } r > 0. \quad (1)$$

In [5], Kiefer and Wolfowitz introduced a *full busy period*, i.e. an interval that begins when an arrival finds $c - 1$ customers in a system and ends when, for the first time after that, a departure leaves behind $c - 1$ customers in a system. A conventional *busy period* begins when an arrival finds the system empty and ends when, for the first time after that, a departure leaves the system empty.

These definitions may be ambiguous when multiple events (arrivals and departures) occur at the same time, for example when the interarrival and service times are lattice with the same span. One way to remove ambiguity is to say that a full busy period begins at some t if the

Received 8 May 2003; revision received 2 September 2004.

* Postal address: School of Arts and Sciences, Western New England College, Springfield, MA 01119-2684, USA.

** Postal address: Department of Industrial Engineering and Operations Research, University of California, Berkeley, CA 94720, USA. Email address: wolff@ieor.berkeley.edu

number of customers in a system is less than or equal to $c - 1$ at $t -$, and greater than or equal to c at $t +$, and ends when the reverse transition occurs. This eliminates full busy periods of zero length.

For $c = 1$, it has long been known that the converse of (1) is true; for an elementary proof of (1) and its converse, see [11]. For $c > 1$, the converse does not hold. Remarkably, this was not resolved until 1997 (see [6]).

The Kiefer and Wolfowitz work-vector process is defined at arrival times. For an arrival n , the sum of the work-vector components is the total work in a system (or just work) found by that arrival. For the corresponding continuous-time process, we define *work* at time t , $V(t)$, as the sum of the remaining service times of all customers in a system at time t .

We consider two cases, as follows.

Case A. A stable queue with $c \geq 2$ and $P(T_n > S_n) > 0$.

For Case A, Whitt [10] showed that discrete- and continuous-time processes, such as the process $\{V(t)\}$, are regenerative in the classical sense, where a *cycle* begins each time an arrival finds the system empty. Cycles and cycle lengths are each i.i.d. Furthermore, he showed that these processes are *positive recurrent*, which means that cycle lengths have finite mean.

Case B. A stable queue with $c \geq 2$ and $P(T_n > S_n) = 0$.

For Case B, the system never empties. Sigman [8] showed that these processes are regenerative, but in a weaker sense. In particular, $\{V(t)\}$ is a *one-dependent* regenerative process. Cycles and cycle lengths are identically distributed; *adjacent* cycles and cycle lengths may be dependent, but otherwise they are independent. The process $\{V(t)\}$ is positive recurrent, with a unique stationary version, as in Case A.

In Case B, cycles are more complicated. For $c = 2$, let (V_{n1}, V_{n2}) be the work vector found by arrival n . For some fixed $\delta > 0$, a cycle begins when arrival $n + 1$ occurs if $V_{n1} = 0$, $V_{n2} \leq \delta$, and $T_n > \delta$. When this occurs, the work vector found by arrival $n + 1$ is of the form $(0, (S - T \mid T > \delta))$; the work found by arrival n is gone when arrival $n + 1$ occurs.

For $c \geq 3$ under Case B, see [7], where Lemma 1 gives explicit regeneration points, and Equation (4.6) is an exact representation of the work vector when regeneration occurs. This vector always has at least one zero component.

These details are not important in our analysis, except to note that when a cycle begins, an arrival finds at least one idle server. Thus, a full busy period begins and ends within the same cycle. In our analysis, B and B_f denote a busy period and a full busy period respectively, and $c \geq 2$.

The basic properties of $\{V(t)\}$ that we need are the same in Cases A and B. In particular, for a generic cycle length C , we have

$$E(C) < \infty. \quad (2)$$

For either case, let V be the stationary work in a system. It is readily shown from similar conditions for finite delay moments, for example in [2], that when the queue is stable,

$$E(S^{r+1}) < \infty \implies E(V^r) < \infty \quad \text{for } r > 0. \quad (3)$$

Thorisson [9] established conditions for finite busy period and busy cycle moments, but only under Case A. He obtained results not only for ordinary moments but also for more general geometric moments and other functional forms. By a much easier analysis, but restricted to

$c = 1$, Ghahramani and Wolff [3] showed that when the queue is stable,

$$E(S^r) < \infty \implies E(B^r) < \infty \quad \text{for } r \geq 1. \tag{4}$$

While a special case of results in [9], (4) was extended in [3] to an *associated* queue: (S_n, T_n) , $n \geq 1$, are i.i.d. vectors, but S_n and T_n may be dependent.

Every cycle contains a random number (possibly zero) of full busy periods. Let N be a generic number of full busy periods in a cycle of length C . We assume that they occur; that is, $P(N > 0) > 0$. Sufficient conditions for this are either $P(T_n > \varepsilon) > 0$ for every $\varepsilon > 0$, or $P(S_n \leq x) < 1$ for every $x > 0$, where the second condition implies Case A. These conditions are independent of c . For fixed $c \geq 2$, the condition $P(S > T_1 + \dots + T_{c-1}) > 0$, where the T_i are i.i.d., is necessary and sufficient for $P(N > 0) > 0$.

In Section 2, we present a probabilistic proof of the following result.

Theorem 1. *When the queue is stable and $P(N > 0) > 0$,*

$$E(S^r) < \infty \implies E(B_f^r) < \infty \quad \text{for } r \geq 1.$$

Under Case A, Theorem 1 may be deduced from results in [9], but our proof is much easier, and rests on results that are much easier to prove. As our proof also holds in Case B, Theorem 1 is new for that case.

2. Proof of Theorem 1

Even in Case A, full busy periods are not i.i.d., because full busy periods within the same cycle are dependent. In Case B, the situation is complicated further by having one-dependent cycles. With one-dependent cycles, the even-numbered and odd-numbered cycles are i.i.d., so that even in this case we can define the distribution of B_f as a long-run average, as is done for various quantities in [12]. Thus, B_f has some distribution function F with this property: for every $x \geq 0$, $F(x)$ is the long-run *fraction* of full busy periods with lengths not exceeding x .

Let Z_f be the sum of the lengths of the full busy periods that occur within a cycle of length C , where, from (2), $E(Z_f) < \infty$. From this, we have that $E(Z_f \mid Z_f > 0) < \infty$ and

$$E(B_f) < \infty, \tag{5}$$

which proves Theorem 1 for $r = 1$.

Now let B_{fe} be a corresponding *equilibrium* full busy period, with distribution function F_e that has density $f_e(x) = [1 - F(x)]/E(B_f)$. It is well known that, when finite, the moments of B_f and B_{fe} are related, i.e.

$$E(B_{fe}^{r-1}) = \frac{E(B_f^r)}{r E(B_f)} \quad \text{for } r \geq 1. \tag{6}$$

To complete the proof, consider the stationary version of the stable continuous-time work-vector process, and suppose that $E(S^r) < \infty$, $r > 1$. At arbitrary time t , (total) work in system V is also stationary.

We condition on event A : ‘a full busy period is in progress at time t ’. Given that A occurs, the *remaining* full busy period has distribution F_e , and we denote it by B_{fe} ; see [12]. From (3) and $P(A) > 0$, we have

$$E[(V \mid A)^{r-1}] < \infty. \tag{7}$$

From time t onwards, we couple the c -server work-vector process with the work process of the corresponding fast single-server queue, first introduced by Brumelle in [1]. This is a GI/G/1 queue with the same arrival process, where, now, a customer that has service time S is served in time S/c . At time t , we give this queue the same amount of (total) work, i.e. $(V | A)$. For this queue, work decreases at rate c , where positive. For our c -server queue, work decreases at rate c when a full busy period is in progress. After time t , both queues are fed the same arrival process and service times. We now compare sample paths of these processes from t onwards.

Let B_r be the duration of the remaining busy period for this single-server queue under these conditions. As the work processes of these queues are identical while B_{fe} is in progress, we must have $B_{fe} \leq B_r$.

We now perform another modification of (only) the single-server queue. Shift every arrival time after t to the left, by the same amount, until the next arrival occurs at t . Call the queue with this modification *shifted*, and let B_s be the duration of the remaining busy period for this shifted queue. This shift can only make the remaining busy period longer. So $B_r \leq B_s$, and we have

$$B_{fe} \leq B_s. \quad (8)$$

For the shifted queue, B_s may also be viewed as an *exceptional first service* busy period, with exceptional first service $S + (V | A)$, where S is a service time independent of all else.

Now $E(S^r) < \infty$ and (7) implies that $E[S + (V | A)]^{r-1} < \infty$ and, hence, by Lemma 2 of [3], $E(B_s^{r-1}) < \infty$. Combining this with (6) and (8), we have $E(B_{fe}^r) < \infty$, which completes the proof.

References

- [1] BRUMELLE, S. L. (1971). Some inequalities for parallel-server queues. *Operat. Res.* **19**, 402–413.
- [2] GHARAMANI, S. (1989). Finiteness of moments of virtual work for GI/G/c queues. *J. Appl. Prob.* **26**, 423–425.
- [3] GHARAMANI, S. AND WOLFF, R. W. (1989). A new proof of finite moment conditions for GI/G/1 busy periods. *Queueing Systems* **4**, 171–178.
- [4] KIEFER, J. AND WOLFOWITZ, J. (1955). On the theory of queues with many servers. *Trans. Amer. Math. Soc.* **78**, 1–18.
- [5] KIEFER, J. AND WOLFOWITZ, J. (1956). On the characteristics of the general queueing process with applications to random walks. *Ann. Math. Statist.* **27**, 147–161.
- [6] SCHELLER-WOLF, A. AND SIGMAN, K. (1997). Delay moments for FIFO GI/G/1 queues. *Queueing Systems* **25**, 77–95.
- [7] SIGMAN, K. (1988). Regeneration in tandem queues with multi-server stations. *J. Appl. Prob.* **25**, 391–403.
- [8] SIGMAN, K. (1990). One-dependent regenerative processes and queues in continuous time. *Math. Operat. Res.* **15**, 175–189.
- [9] THORISSON, H. (1985). The queue GI/G/k: finite moments of the cycle variables and uniform rates of convergence. *Commun. Statist. Stoch. Models* **1**, 221–238.
- [10] WHITT, W. (1972). Embedded renewal processes in the GI/G/1 queues. *J. Appl. Prob.* **9**, 650–658.
- [11] WOLFF, R. W. (1984). Conditions for finite ladder height and delay moments. *Operat. Res.* **32**, 906–916.
- [12] WOLFF, R. W. (1988). Sample-path derivations of the excess, age, and spread distributions. *J. Appl. Prob.* **25**, 432–436.