The Annals of Applied Probability
1994, Vol. 4, No. 2, 591–606

ASYMPTOTIC BEHAVIOR OF LARGE DISCRETE-TIME CYCLIC QUEUEING NETWORKS

BY VICTOR PESTIEN AND S. RAMAKRISHNAN

University of Miami

Assume that k jobs circulate clockwise through a cyclic network of n single-server queues, where at each integer time instant the job at the head of each queue moves with probability p to the next queue, independent of the other jobs. The equilibrium distribution for the associated Markov chain is determined, and an exact expression for the expected number of busy servers is obtained. If n and k are large, a simple approximation for the proportion of busy servers is derived. In a second model, where the queues have no waiting room and where movement of a job occurs with probability p only if the next queue was empty, a similar, simple asymptotic representation for the proportion of busy servers is deduced. This representation readily yields a simple expression for the asymptotic cycle time for a single job.

1. Introduction. Suppose n nodes, labelled $0, 1, \dots, n-1$, are arranged in a circle and that at each node there is a service queue. Suppose there are k jobs, which move clockwise around the circle, waiting for and receiving service at each node in succession. There might be more than one job at a given node, but each node has only one server.

Jobs can move only at discrete times t, where t is a positive integer. At every time instant t each job is at one of the nodes, and the following movements occur simultaneously at each node m, independently of other nodes: If the queue at node m is nonempty, then the job at the head of that queue moves with probability p $(0 < p < 1)$ to the tail of the queue at node $m \oplus 1$; with probability $1 - p$ there is no movement of jobs from node m to node $m \oplus 1$. (Here, $\oplus$ means addition modulo n.)

Thus the n queues have independent geometric service times, each with expectation $1/p$, there is ample waiting room in each queue and the cyclic network of queues is closed.

In Section 2 we model the behavior of this queueing network by means of a discrete-time Markov chain where each state is an n-tuple describing the number of jobs present at each of the n nodes. We derive the equilibrium distribution for the chain and find that the equilibrium probabilities of the states are weighted according to how many nodes are occupied (that is, according to how many servers are busy). This unequal weighting is sharply different from the continuous-time analogue of this model—a Gordon and Newell network—where all states have equal weight.

Received February 1993; revised September 1993.
AMS 1991 *subject classifications*. Primary 60K25; secondary 60K35, 90B12, 90B22.
Key words and phrases. Discrete-time queue, cyclic queueing network, equilibrium distribution.

In Section 3 we study the model of Section 2 in the case where n and k are large, and we estimate the asymptotic proportion of occupied nodes. If k and n approach infinity in such a way that the ratio k/n approaches a constant α, we prove that the proportion of occupied nodes converges in probability to the constant

$$\frac{1+\alpha-\sqrt{(1+\alpha)^2-4p\alpha}}{2p}.$$

This proportion serves as a measure of progress for the system at a given time instant, and thus it can be used to estimate the average cycle time for a job in the network.

In Section 4 we study a second model, where only one job is permitted at a node at a given time instant. A job moves ahead to the next node with probability p only if the next node was unoccupied. The equilibrium distribution for this model has been given in Pestien, Ramakrishnan and Sarkar (1993), hereafter abbreviated PRS (1993), but here the methods of Section 3 are used to determine the asymptotic proportion of occupied nodes in the system.

Although cyclic networks may at first glance appear rather special in structure, they can be used to model other linearly-arranged systems where the number of jobs in the system is constant over a long period and where a job is always waiting to enter whenever another job leaves the system. [For example, see Onvural (1990).] Most of the literature on queueing networks pertains to continuous time, but there are synchronous communication systems that are more naturally described in a discrete-time setting. The survey article by Kobayashi and Konheim (1977) gives an account of such models. The paper by Boxma and Groenendijk (1988) is one of the more recent studies in this direction.

2. Equilibrium distribution in the ample-waiting-room model. To describe the behavior of the cyclic queueing network given in the introduction, we use a discrete-time Markov chain with stationary transition probabilities. The state space S of the Markov chain will consist of n-tuples $s=(s_0, s_1, \dots, s_{n-1})$, where s_j is the number of jobs present (length of the queue) at node j. To be precise, let

$$S=\left\{(s_0, s_1, \dots, s_{n-1}) \in \{0,1,\dots k\}^n \colon \sum_{j=0}^{n-1} s_j = k\right\}.$$

For each s in S, let

$$\mathrm{OCC}(s)=\{j\colon 0 \le j \le n-1 \text{ and } s_j>0\},$$

so that $\mathrm{OCC}(s)$ is the set of occupied nodes for the state s.

If $s \in S$ and $A \subseteq \mathrm{OCC}(s)$, define the n-tuples $v^A = (v_0^A, \dots, v_{n-1}^A)$ and $w^A = (w_0^A, \dots, w_{n-1}^A)$ by

$$v_j^A = \begin{cases} 1, & \text{if } j \in A, \\ 0, & \text{if } j \notin A, \end{cases}$$

and

$$w_j^A = \begin{cases} 1, & \text{if } j \ominus 1 \in A, \\ 0, & \text{if } j \ominus 1 \notin A, \end{cases}$$

where $\ominus$ is subtraction modulo n. Further define the n-tuple $s^A = (s_0^A, \dots, s_{n-1}^A)$ by

$$s_j^A = s_j - v_j^A + w_j^A.$$

Then s^A is the state that will result from s when A is the set of nodes at which service has just been completed.

Next, the rules of motion given in the introduction lead us to define the transition matrix P for the Markov chain as follows: If s and r are elements of S, let

$$P(s,r) = \sum p^{|A|}(1-p)^{|\mathrm{OCC}(s)|-|A|}, \tag{2.1}$$

where the summation is taken over all subsets A of $\mathrm{OCC}(s)$ for which $r = s^A$. If there are no such subsets, let $P(s,r) = 0$. ($|\cdot|$ denotes cardinality.)

Each addend in expression (2.1) reflects the fact that in the transition from s to r there has been movement at $|A|$ of the occupied nodes and no movement at $|\mathrm{OCC}(s)| - |A|$ of the occupied nodes. It is easy to see that if $k < n$, then for each r and s there is at most one set A such that $A \subseteq \mathrm{OCC}(s)$ and $r = s^A$. However, if $k \geq n$, there could be more than one such set A. For example, if $k = 4$ and $n = 3$ and if $r = s = (2,1,1)$, then both $A = \{1,2,3\}$ and $A = \varnothing$ satisfy $r = s^A$. That is, movement could have occurred at all nodes or at none of the nodes to transform s into r.

It is evident that the finite Markov chain with transition matrix P is irreducible and positive recurrent. Therefore, as is well known, there is a unique equilibrium probability distribution for P. That is, there is a unique vector π on S such that

$$\pi(s) = \sum_{r \in S} \pi(r) P(r,s) \quad \text{for all } s \text{ in } S, \tag{2.2}$$

$$\pi(s) > 0 \quad \text{for all } s \text{ in } S \tag{2.3}$$

and

$$\sum_{s \in S} \pi(s) = 1. \tag{2.4}$$

Thus the equilibrium probability distribution for P can be obtained by first finding a vector γ that satisfies (2.2) and (2.3) and then normalizing it to satisfy (2.4).

THEOREM 2.1. *Let γ be the vector on S defined by*

$$\gamma(s) = (1-p)^{-|\mathrm{OCC}(s)|}. \tag{2.5}$$

Then

$$\gamma(s) = \sum_{r \in S} \gamma(r) P(r,s) \quad \text{for all } s \text{ in } S.$$

PROOF. For each s in S, let

$$\mathrm{PREOCC}(s) = \{j \ominus 1 : j \in \mathrm{OCC}(s)\}.$$

So PREOCC(s) is the set of labels of nodes from which jobs might have just moved at the previous time instant to create the state s. Certainly

$$|\mathrm{PREOCC}(s)| = |\mathrm{OCC}(s)|. \tag{2.6}$$

By (2.1) and (2.5),

$$\begin{aligned}
&\sum_{r \in S} \gamma(r) P(r,s) \\
&\quad = \sum_{r \in S} (1-p)^{-|\mathrm{OCC}(r)|} \sum_{\{A:\, A \subseteq \mathrm{OCC}(r) \text{ and } s = r^A\}} p^{|A|} (1-p)^{|\mathrm{OCC}(r)| - |A|} \\
&\quad = \sum_{r \in S} \sum_{\{A:\, A \subseteq \mathrm{OCC}(r) \text{ and } s = r^A\}} \left(\frac{p}{1-p} \right)^{|A|}.
\end{aligned}$$

For a fixed $s \in S$, note that there is a one-to-one correspondence between the sets

$$\{(r, A) : r \in S,\ A \subseteq \mathrm{OCC}(r) \text{ and } r^A = s\} \quad \text{and} \quad \{A : A \subseteq \mathrm{PREOCC}(s)\}.$$

Therefore

$$\sum_{r \in S} \gamma(r) P(r,s) = \sum_{\{A:\, A \subseteq \mathrm{PREOCC}(s)\}} \left(\frac{p}{1-p} \right)^{|A|}.$$

Now use the binomial theorem to conclude that

$$\sum_{r \in S} \gamma(r) P(r,s) = \left(1 + \frac{p}{1-p} \right)^{|\mathrm{PREOCC}(s)|} = (1-p)^{-|\mathrm{PREOCC}(s)|},$$

which by (2.6) equals $(1-p)^{-|\mathrm{OCC}(s)|}$. □

COROLLARY 2.2. *Define the vector π on S by*

$$\pi(s) = \frac{(1-p)^{-|\mathrm{OCC}(s)|}}{\sum_{s' \in S} (1-p)^{-|\mathrm{OCC}(s')|}}.$$

Then π is the equilibrium probability distribution for the Markov chain whose transition matrix is P.

Of course, *continuous-time* cyclic queueing networks have been studied thoroughly. In particular, the continuous-time analogue of our discrete-time large-waiting-room model was investigated by Koenigsberg (1958) and generalized by Gordon and Newell (1967). The equilibrium distribution in such networks is of so-called product form, and since all service times are exponential with the same parameter, all states have equal probability under the equilibrium distribution. Thus there is a distinct difference between the respective equilibrium distributions in the continuous-time and discrete-time models. In our discrete-time model, Theorem 2.1 says that the equilibrium probabilities of the states are far from being equal, as they are weighted exponentially according to the number of occupied nodes. Some other differences between continuous-time and discrete-time models are mentioned by Kobayashi and Konheim (1977) and by Walrand (1983).

We will now determine the expected number of occupied nodes (i.e., the number of busy servers) under the equilibrium distribution.

LEMMA 2.3. *For* $1 \le l \le \min\{k, n\}$, *let*

$$S_l = \{s \in S: |\mathrm{OCC}(s)| = l\}.$$

Then

$$|S_l| = \binom{n}{l}\binom{k-1}{l-1}. \tag{2.7}$$

PROOF. Observe that

$$S_l = \left\{(j_0, \dots, j_{n-1}): \text{each } j_i \ge 0 \text{ and } \sum_{i=0}^{n-1} j_i = k \right.$$
$$\left. \text{and exactly } l \text{ of the } j_i\text{'s are positive}\right\}$$

and use a standard counting argument [for example, see Johnson and Kotz (1977)]. □

THEOREM 2.4. *Let*

$$E(|\mathrm{OCC}|) = \sum_{s \in S} |\mathrm{OCC}(s)| \pi(s),$$

the expected number of occupied nodes under the equilibrium distribution π. *Then*

$$E(|\mathrm{OCC}|) = \frac{\sum_{l=1}^{\min\{k,n\}} l \binom{n}{l}\binom{k-1}{l-1}(1-p)^{-l}}{\sum_{l=1}^{\min\{k,n\}} \binom{n}{l}\binom{k-1}{l-1}(1-p)^{-l}}. \tag{2.8}$$

PROOF. Use (2.7) and (2.5). □

3. Asymptotics in the ample-waiting-room model. Because of (2.8), $E(|\mathrm{OCC}|)$ is the expected value of the random variable X_n^k, where X_n^k has probability function f_n^k given by

$$f_n^k(l) = P[X_n^k = l] = \frac{\binom{n}{l}\binom{k-1}{l-1}(1-p)^{-l}}{\sum_{l'=1}^{\min\{k,n\}}\binom{n}{l'}\binom{k-1}{l'-1}(1-p)^{-l'}} \tag{3.1}$$

for l such that $1 \le l \le \min\{k, n\}$. We can think of X_n^k as the number of occupied nodes (busy servers) in the network.

The purpose of this section is to estimate the proportion X_n^k/n of occupied nodes in the case where the cyclic network is very large. At a given time instant, the average number of nodes at which service is successfully completed is p times the number of occupied nodes. Thus the proportion of occupied nodes gives a measure of progress for the system at that instant. We establish the following asymptotic result:

THEOREM 3.1. *Suppose $k = k_n$ varies with n in such a way that*

$$\lim_{n\to\infty} \frac{k_n}{n} = \alpha,$$

where α is a constant, and let

$$\varphi = \frac{1 + \alpha - \sqrt{(1+\alpha)^2 - 4p\alpha}}{2p}. \tag{3.2}$$

Then $X_n^{k_n}/n$ converges in probability to the constant φ. That is, for each $\zeta > 0$,

$$\lim_{n\to\infty} P\left[\left|\frac{X_n^{k_n}}{n} - \varphi\right| > \zeta\right] = 0.$$

Informally, Theorem 3.1 says that for very large networks in equilibrium, the proportion of busy servers stays approximately constant, and the theorem exhibits that constant φ in (3.2). It is interesting to note, for fixed α, that φ is an increasing function of p.

To prepare for the argument for Theorem 3.1, define for each pair of positive integers k and n,

$$l_n^k := \frac{n + k + 1 - p - \sqrt{(n+k+1-p)^2 - 4pnk}}{2p}. \tag{3.3}$$

Notice that if

$$\lim_{n\to\infty} \frac{k_n}{n} = \alpha,$$

then

$$\lim_{n\to\infty} \frac{l_n^{k_n}}{n} = \varphi.$$

Also notice that l_n^k is a solution to the following equation in the variable l:

$$pl^2 + (-n - k - 1 + p)l + nk = 0. \tag{3.4}$$

Now (3.4) is equivalent to

$$\frac{l+1}{n-l} \cdot \frac{l}{k-l} = \frac{1}{1-p}. \tag{3.5}$$

In the special case where l is a positive integer, (3.5) is in turn equivalent to

$$\binom{n}{l}\binom{k-1}{l-1}(1-p)^{-l} = \binom{n}{l+1}\binom{k-1}{l}(1-p)^{-l-1} \tag{3.6}$$

and to

$$f_n^k(l) = f_n^k(l+1).$$

Further, define $\hat{l}_n^k$ as the unique positive integer such that

$$l_n^k \leq \hat{l}_n^k < l_n^k + 1.$$

The key idea in the proof of Theorem 3.1 is that the probability function f_n^k of the random variable X_n^k is increasing for $l \leq \hat{l}_n^k$ and decreasing for $l > \hat{l}_n^k$. Moreover, for large n, the probability histogram of X_n^k is sufficiently steep on either side of $\hat{l}_n^k$.

The next theorem is really a technical reformulation of Theorem 3.1:

THEOREM 3.2. *Assume that* $\lim_{n\to\infty}(k_n/n) = \alpha$, *where* $\alpha > 0$, *and suppose* $\varepsilon > 0$ *and* $\Delta > 0$. *Then there exists a positive integer* N *such that if* $n \geq N$ *and if*

$$|J| > n^{2/3}\Delta,$$

then

$$\frac{f_n^{k_n}\left(\hat{l}_n^{k_n} + J\right)}{f_n^{k_n}\left(\hat{l}_n^{k_n}\right)} < \frac{\varepsilon}{n}.$$

PROOF OF THEOREM 3.1 USING THEOREM 3.2. If $\alpha = 0$, then $\varphi = 0$ and $X_n^{k_n}/n$ is bounded above by k_n/n, which converges to 0. Next assume $\alpha > 0$, and fix $\varepsilon > 0$ and $\zeta > 0$. Clearly n can be made large enough so

$$\frac{\hat{l}_n^{k_n} - l_n^{k_n}}{n} < \frac{\zeta}{3}$$

and

$$\varphi - \frac{l_n^{k_n}}{n} < \frac{\zeta}{3}.$$

Let $\Delta = \zeta/3$ and let

$$\mathscr{J} = \left\{j\colon 0 \leq \hat{l}_n^{k_n} + j \leq \min\{k, n\} \text{ and } |j| > n\Delta\right\}.$$

Then for n sufficiently large, Theorem 3.2 gives

$$\begin{aligned} P\left[\left|\frac{X_n^{k_n}}{n} - \varphi\right| > \zeta\right] &\le P\left[\left|\frac{X_n^{k_n}}{n} - \frac{\hat{l}_n^{k_n}}{n}\right| > \Delta\right] \\ &= \sum_{j \in \mathscr{J}} f_n^{k_n}\left(\hat{l}_n^{k_n} + j\right) \\ &\le \sum_{j \in \mathscr{J}} \frac{\varepsilon}{n} \cdot f_n^{k_n}\left(\hat{l}_n^{k_n}\right). \end{aligned}$$

Finally, because the summation contains at most n addends and because $f_n^{k_n}(\hat{l}_n^{k_n}) \le 1$, we have

$$P\left[\left|\frac{X_n^{k_n}}{n} - \varphi\right| > \zeta\right] \le \varepsilon. \qquad \square$$

Before proceeding with the proof of Theorem 3.2, we need the following two lemmas. The first lemma checks that l_n^k is in the correct range.

LEMMA 3.3. *For each $k \ge 1$ and each $n \ge 1$, $l_n^k < \min\{k, n\}$.*

PROOF.

$$\begin{aligned} l_n^k &= \frac{n + k + 1 - p - \sqrt{(n + k + 1 - p)^2 - 4pnk}}{2p} \\ &= \frac{2nk}{n + k + 1 - p} \cdot \left(1 + \sqrt{1 - \frac{4pnk}{(n + k + 1 - p)^2}}\right)^{-1} \\ &< \frac{2nk}{n + k} \cdot \left(1 + \sqrt{1 - \frac{4nk}{(n + k)^2}}\right)^{-1} \\ &= \frac{n + k - \sqrt{(n - k)^2}}{2} \\ &= \min\{k, n\}. \end{aligned} \qquad \square$$

LEMMA 3.4. *Suppose $1 \le k \le n$. The integer $\hat{l}_n^k$ is a "mode" for the random variable X_n^k in the sense that*

$$f_n^k(l) > f_n^k(l + 1)$$

for each positive integer l such that $\hat{l}_n^k < l < \min\{k, n\}$ and

$$f_n^k(l) < f_n^k(l + 1)$$

for each positive integer l such that $1 \le l < \hat{l}_n^k$.

PROOF. First, for any positive integer l such that $1 \le l \le \min\{k, n\}$,

$$\frac{f_n^k(l+1)}{f_n^k(l)} = \frac{(n-l)(k-l)}{(l+1)l} \cdot \frac{1}{1-p}.$$

Now use (3.4) and (3.5) to get

$$\frac{f_n^k(l+1)}{f_n^k(l)} = \left[\frac{n-l}{l+1} \cdot \frac{l_n^k+1}{n-l_n^k}\right]\left[\frac{k-l}{l} \cdot \frac{l_n^k}{k-l_n^k}\right]. \tag{3.7}$$

It is easily verified that

$$\frac{n-l}{l+1} \cdot \frac{l_n^k+1}{n-l_n^k} > 1 \quad \text{iff} \quad l_n^k > l \tag{3.8}$$

and that

$$\frac{k-l}{l} \cdot \frac{l_n^k}{k-l_n^k} > 1 \quad \text{iff} \quad l_n^k > l. \tag{3.9}$$

Therefore,

$$\frac{f_n^k(l+1)}{f_n^k(l)} > 1 \quad \text{iff} \quad l_n^k > l.$$

Also, observe that $l_n^k = l$ if and only if $f_n^k(l+1) = f_n^k(l)$. Finally, note that by the definition of $\hat{l}_n^k$, we have $l_n^k > l$ if and only if $\hat{l}_n^k > l$, for any positive integer l. □

PROOF OF THEOREM 3.2. Use the notation $b(x; m, \theta)$ for the binomial probability $\binom{m}{x}\theta^x(1-\theta)^{m-x}$ and write (3.7) equivalently as

$$\frac{f_n^k(l+1)}{f_n^k(l)} = \left[\frac{b(l+1; n, (l_n^k+1)/(n+1))}{b(l; n, (l_n^k+1)/(n+1))}\right]\left[\frac{b(l; k-1, l_n^k/k)}{b(l-1; k-1, l_n^k/k)}\right],$$

where, moreover, by (3.8) and (3.9), each expression within square brackets on the right is less than or equal to 1 if and only if $l \ge \hat{l}_n^k$.

Thus for any integer J satisfying $1 \le \hat{l}_n^k + J \le \min\{k, n\}$,

$$\begin{aligned} \frac{f_n^k(\hat{l}_n^k+J)}{f_n^k(\hat{l}_n^k)} &= \left[\frac{b(\hat{l}_n^k+J; n, (l_n^k+1)/(n+1))}{b(\hat{l}_n^k; n, (l_n^k+1)/(n+1))}\right] \\ &\quad \times \left[\frac{b(\hat{l}_n^k-1+J; k-1, l_n^k/k)}{b(\hat{l}_n^k-1; k-1, l_n^k/k)}\right]. \end{aligned} \tag{3.10}$$

Moreover, each expression within square brackets on the right side is at most 1, is a nondecreasing function of J for $J < 0$ and is a nonincreasing function of J for $J > 0$. In particular,

$$\frac{f_n^k(\hat{l}_n^k + J)}{f_n^k(\hat{l}_n^k)} \le \frac{b(\hat{l}_n^k + J; n, (l_n^k + 1)/(n+1))}{b(\hat{l}_n^k; n, (l_n^k + 1)/(n+1))}. \tag{3.11}$$

The existence of upper bounds, exponentially decreasing in $|J|$, is well known for the ratio of binomial probabilities on the right of (3.11). For example, it follows from relation (VII.3.9) in Feller [(1968), page 183], together with the remark after (VII.3.2) in the same reference, that

$$\frac{b(\hat{l}_n^k + J; n, (l_n^k + 1)/(n+1))}{b(\hat{l}_n^k; n, (l_n^k + 1)/(n+1))} = \exp\left(-\frac{J^2}{2np_nq_n} + \xi_J\right), \tag{3.12}$$

where $p_n = (l_n^k + 1)/(n+1) = 1 - q_n$ and where

$$|\xi_J| < \frac{|J|^3}{(np_nq_n)^2} + \frac{2|J|}{np_nq_n}.$$

Therefore, we have

$$\begin{aligned} &\frac{b(\hat{l}_n^k + J; n, (l_n^k + 1)/(n+1))}{b(\hat{l}_n^k; n, (l_n^k + 1)/(n+1))} \\ &\quad \le \exp\left(-\frac{J^2}{2np_nq_n} + \frac{|J|^3}{(np_nq_n)^2} + \frac{2|J|}{np_nq_n}\right). \end{aligned} \tag{3.13}$$

It follows from (3.11) and (3.13) that for $\Delta > 0$, if $|J| > n^{2/3}\Delta$, we have

$$\frac{f_n^k(\hat{l}_n^k + J)}{f_n^k(\hat{l}_n^k)} \le \exp\left\{-\frac{(\lfloor n^{2/3}\Delta\rfloor)^2}{2np_nq_n} + \frac{(n^{2/3}\Delta)^3}{(np_nq_n)^2} + \frac{2(n^{2/3}\Delta)}{np_nq_n}\right\}.$$

The hypothesis $\lim_{n\to\infty}(k/n) = \alpha > 0$ implies $\lim_{n\to\infty} p_n = \varphi > 0$. So given $\varepsilon > 0$, the right side is bounded above by ε/n for sufficiently large n. □

REMARK 1. The foregoing arguments can be refined to establish the asymptotic normality of the sequence X_n^k. More precisely, it can be shown that $(X_n^k - \hat{l}_n^k)/\sqrt{n}$ converges in law to the normal distribution with mean 0 and variance $[1/(\varphi(1-\varphi)) + \alpha/(\varphi(\alpha - \varphi))]^{-1}$. (The hypothesis $\alpha > 0$ guarantees that $0 < \varphi < \min\{1, \alpha\}$.)

To see normality, first note that an argument similar to the one for (3.12) yields

$$\frac{b(\hat{l}_n^k - 1 + J; k-1, l_n^k/k)}{b(\hat{l}_n^k - 1; k-1, l_n^k/k)} = \exp\left\{-\frac{J^2}{2(k-1)u_nv_n} + \zeta_J\right\}, \tag{3.14}$$

where $u_n = l_n^k/k = 1 - v_n$ and where

$$|\zeta_J| < \frac{|J|^3}{[(k-1)u_n v_n]^2} + \frac{2|J|}{(k-1)u_n v_n}.$$

So by (3.10), (3.12) and (3.14), we have

$$\frac{f_n^k(\hat{l}_n^k + J)}{f_n^k(\hat{l}_n^k)} = \exp\left\{-\frac{J^2}{2n}\left[\frac{1}{p_n q_n} + \frac{n}{(k-1)u_n v_n}\right]\right\}\exp\{\xi_J + \zeta_J\}. \tag{3.15}$$

Because $\lim_{n\to\infty}(k/n) = \alpha$, we have $\lim_{n\to\infty} p_n = \varphi$ and $\lim_{n\to\infty} u_n = \varphi/\alpha$. Moreover, for any pair of real numbers a and b such that $a < b$, the upper bounds on $|\xi_J|$ and $|\zeta_J|$ imply that

$$\lim_{n\to\infty} \sup_{a\sqrt{n} \le J \le b\sqrt{n}} |\xi_J + \zeta_J| = 0.$$

As a consequence, (3.15) yields

$$\lim_{n\to\infty} \sum_{a\sqrt{n} \le J \le b\sqrt{n}} \frac{1}{\sqrt{n}} \frac{f_n^k(\hat{l}_n^k + J)}{f_n^k(\hat{l}_n^k)} = \lim_{n\to\infty} \sum_{a\sqrt{n} \le J \le b\sqrt{n}} \frac{1}{\sqrt{n}} \exp\left\{-\frac{J^2}{2n}\left[\frac{1}{p_n q_n} + \frac{n}{(k-1)u_n v_n}\right]\right\}.$$

Equivalently,

$$\lim_{n\to\infty} \sum_{a\sqrt{n} \le J \le b\sqrt{n}} \frac{1}{\sqrt{n}} \frac{f_n^k(\hat{l}_n^k + J)}{f_n^k(\hat{l}_n^k)} = \int_a^b \exp\left(-\frac{x^2}{2\sigma^2}\right) dx, \tag{3.16}$$

where

$$\sigma^2 = \left[\frac{1}{\varphi(1-\varphi)} + \frac{\alpha}{\varphi(\alpha-\varphi)}\right]^{-1}.$$

Thus the assertion of asymptotic normality of X_n^k would be proved if we show that

$$\lim_{n\to\infty} \frac{1}{\sqrt{n}\, f_n^k(\hat{l}_n^k)} = \sigma\sqrt{2\pi}. \tag{3.17}$$

Of course, since $\sum_J f_n^k(\hat{l}_n^k + J) = 1$, it easily follows from (3.16) that

$$\liminf_{n\to\infty} \frac{1}{\sqrt{n}\, f_n^k(\hat{l}_n^k)} \ge \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \sigma\sqrt{2\pi}. \tag{3.18}$$

On the other hand, for each $\Delta > 0$, by Theorem 3.2,

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}\, f_n^k(\hat{l}_n^k)} = \limsup_{n\to\infty} \sum_{|J| \le n^{2/3}\Delta} \frac{f_n^k(\hat{l}_n^k + J)}{\sqrt{n}\, f_n^k(\hat{l}_n^k)},$$

which, by an argument similar to that of (3.16), implies that

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}\, f_n^k(\hat{l}_n^k)} \le \left[\int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx\right] \times \limsup_{n\to\infty}\left[\sup_{|J|\le n^{2/3}\Delta} \exp(|\xi_J + \zeta_J|)\right].$$

Finally, verify from the upper bounds of $|\xi_J|$ and $|\zeta_J|$ that

$$\lim_{\Delta\downarrow 0}\left\{\limsup_{n\to\infty}\left[\sup_{|J|\le n^{2/3}\Delta} \exp(|\xi_J + \zeta_J|)\right]\right\} = 1,$$

so that

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}\, f_n^k(\hat{l}_n^k)} \le \sigma\sqrt{2\pi}. \tag{3.19}$$

The assertion (3.17) follows from (3.18) and (3.19).

REMARK 2. In the theory of continuous-time interacting particle systems, as pointed out in Liggett [(1985), page 415], "Results..., which give the asymptotic profile of the disturbance in a system, are sometimes referred to as being hydrodynamical in nature." Thus the conclusions of Theorem 1 and Remark 1 may be regarded as hydrodynamical. For the continuous-time analogue of our discrete-time ample-waiting-room model, since all states have equal probability under the equilibrium distribution, the probability function of X_n^k is given by (3.1) with $p = 0$, which is a hypergeometric distribution with

$$E(X_n^k) = \frac{kn}{n+k-1} \quad \text{and} \quad \operatorname{Var}(X_n^k) = \frac{kn(k-1)(n-1)}{(n+k-1)^2(n+k-2)}.$$

Therefore, in continuous-time, X_n^k/n converges in probability to $\alpha/(\alpha+1)$ if $k/n \to \alpha$ as $n \to \infty$. The constant $\alpha/(\alpha+1)$ is easily seen from (3.2) to be the limit of φ as p decreases to 0. In this sense, our results for the discrete-time model may be viewed as refinements of corresponding calculations for the continuous-time model.

4. Asymptotics in the no-waiting-room model. We now discuss a second discrete-time model, where again there is a closed cyclic network with n nodes and k jobs which circulate the network in a clockwise direction. This time, however, assume that $k \le n$ and that the maximum queue length at each node is constrained to be 1. That is, there is no "waiting room" for a job at a node. We assume that if a job would complete service at a node but the next node is occupied at that instant, it must repeat service at the previous node at the next instant. In continuous-time networks, such a constraint is called "communication blocking;" see, for example, Tsoucas and Walrand (1989). A survey of various blocking protocols is given by Onvural (1990). A discrete-time model with a blocking protocol different from ours is studied by Chen and Shepp (1991).

We assume here that at each time instant $t = 1, 2, \dots$ a job located at node j can move only if node $j \oplus 1$ does not have a job; if the job *can* move, it *will* move, independently of other jobs that can move, with probability p $(0 < p < 1)$ to node $j + 1$, and it will stay at station j with probability $1 - p$. This model is introduced in a computing setting by Berman and Simon (1988), who use the colorful analogy of frogs leaping from stone to stone in a pond where the stones are arranged in a circle. In PRS (1993), the model is formalized as a Markov chain with state space the set of n-tuples of 0's and 1's for which exactly k entries are 1. For a state s, the set OPP(s) is the set of nodes j such that the jth entry of s is 1 and the ($j \oplus 1$)st entry of s is 0. At each node of OPP(s) there is an "opportunity" for the job at that node to move ahead. The equilibrium distribution for this model is given [PRS (1993), Theorem 2.1] by

$$\pi(s) = \frac{(1-p)^{-|\mathrm{OPP}(s)|}}{\sum_{r \in S}(1-p)^{-|\mathrm{OPP}(r)|}}.$$

In PRS [(1993), Proposition 3.1], the following expression for the expected number of "opportunities" under the equilibrium distribution π is derived:

$$E(|\mathrm{OPP}|) = \frac{\sum_{l=1}^{\min(k,n-k)} \binom{k-1}{l-1}\binom{n-k-1}{l-1}(1-p)^{-l}}{\sum_{l'=1}^{\min(k,n-k)} \frac{1}{l'}\binom{k-1}{l'-1}\binom{n-k-1}{l'-1}(1-p)^{-l'}}.$$

We can rewrite E(|OPP|) equivalently as

$$(4.1)\qquad E(|\mathrm{OPP}|) = \frac{\sum_{l=1}^{\min(k,n-k)} l\binom{k}{l}\binom{n-k-1}{l-1}(1-p)^{-l}}{\sum_{l'=1}^{\min(k,n-k)} \binom{k}{l'}\binom{n-k-1}{l'-1}(1-p)^{-l'}}.$$

Thus $E(|\mathrm{OPP}|)$ is the expected value of the random variable Y_n^k, where Y_n^k has probability function g_n^k given by

$$(4.2)\qquad g_n^k(l) = P\left[Y_n^k = l\right] = \frac{\binom{k}{l}\binom{n-k-1}{l-1}(1-p)^{-l}}{\sum_{l'=1}^{\min(k,n-k)} \binom{k}{l'}\binom{n-k-1}{l'-1}(1-p)^{-l'}}$$

for l such that $1 \le l \le \min(k, n-k)$.

The similarity between this probability function g_n^k and the probability function f_n^k of (3.1) allows us to get the following asymptotic result:

THEOREM 4.1. *Suppose $k = k_n$ varies with n in such a way that*

$$\lim_{n \to \infty} \frac{k_n}{n} = \beta,$$

where β is a constant, and let

$$\psi = \frac{1 - \sqrt{1 - 4p\beta(1-\beta)}}{2p}. \tag{4.3}$$

Then $Y_n^{k_n}/n$ converges in probability to the constant ψ. That is, for each $\zeta > 0$,

$$\lim_{n \to \infty} P\left[\left|\frac{Y_n^{k_n}}{n} - \psi\right| > \zeta\right] = 0.$$

PROOF. Use Theorem 3.1, replacing n of that theorem by k and replacing k of that theorem by $n - k$. □

One of the main results in PRS (1993) obtains an exact expression for the long-run average cycle time for a job. This answered a question that was posed and left open in Berman and Simon (1988). The progress of the system at a given instant can be measured by how many jobs successfully move from one node to the next at that instant. On the average, the number of successful moves is $pE(|\mathrm{OPP}|)$. Not surprisingly, it follows [see PRS (1993) for details] that the long-run average cycle time $T(n, k, p)$ for a job is the long-run average time for nk units of progress of the system, namely,

$$T(n, k, p) = nk/(pE(|\mathrm{OPP}|)). \tag{4.4}$$

Thus for the model of this section, (4.1) together with (4.4) gives a complete characterization of long-run average cycle time.

On the other hand, for the large-waiting-room model of the previous sections, the same reasoning given in PRS (1993) for (4.4) can be applied together with (2.8) to describe explicitly the long-run average cycle time. In ring networks, the existence of long-run average cycle time as a constant depending only on the initial configuration has been proved under very general hypotheses by Bambos (1992). (See also references contained there.)

In the continuous-time case, without blocking, characterizations of the cycle time for cyclic exponential queueing networks have been obtained. See Chow (1980), Schassberger and Daduna (1983) or the survey by Boxma and Daduna (1990). In certain continuous-time cyclic networks with blocking, results on cycle time have also been given [for example, see Balsamo and Donatiello (1989)].

One drawback to our representations for cycle time based on (4.1) and (2.8) is that the combinatorial formulas in (4.1) and (2.8) are quite complicated. Using moment-generating functions, a somewhat unsatisfactory attempt to approximate the expression in (4.1) was made in PRS (1993). In particular, the paper left open the question of obtaining a *simple* expression in the no-waiting-room model for the limit of the long-run progress of a job in unit time, namely,

$$\frac{n}{T(n,k,p)}, \tag{4.5}$$

as $n \to \infty$, $k \to \infty$ and $k/n \to \beta$. Earlier, in the special case where $\beta = \frac{1}{2}$, Piotr Berman (personal communication) has studied the problem empirically and has conjectured that the limit of (4.5) was $1 - \sqrt{1-p}$.

Here, Theorem 4.1 answers the question and says that the limit of (4.5) is

$$\frac{p\psi}{\beta} = \frac{1 - \sqrt{1 - 4p\beta(1-\beta)}}{2\beta}.$$

In the special case where $\beta = \frac{1}{2}$, this limit is $1 - \sqrt{1-p}$, which confirms Berman's conjecture.

The no-waiting-room model can be viewed as a discrete-time version of the "exclusion process" in the theory of interacting particle systems. See Liggett [(1985), Chapter VIII] for an account of the exclusion process. In the absence of any "interaction" with other jobs, the long-run progress of a job in unit time would be p. Thus, the presence of interaction slows down the system by a factor of ψ/β. Reasoning similar to that in Remark 2 of Section 3 allows us to view this result as a refinement of a corresponding result in continuous time. The "slow-down factor" for the continuous-time analogue would, therefore, be given by

$$\lim_{p \to 0^+} \frac{\psi}{\beta} = 1 - \beta.$$

This answer agrees with the conclusions of theorems of Spitzer and of Kesten [see Liggett (1985), Corollaries VIII.4.6 and VIII.4.9]. These are results for exclusion processes on the lattice of integers and may be regarded as variations of the continuous-time analogue of the no-waiting-room model.

Acknowledgment. We are grateful to the referee for providing detailed comments, proposing a more appropriate title for the paper and suggesting that we present the relation between our results and the theory of interacting particle systems.

REFERENCES

BALSAMO, S. and DONATIELLO, L. (1989). On the cycle time distribution in a two-stage cyclic network with blocking. *IEEE Trans. Software Engrg.* **15** 1206–1216.

BAMBOS, N. (1992). On closed ring queueing networks. *J. Appl. Probab.* **29** 975–995.

BERMAN, P. and SIMON, J. (1988). Investigations of fault-tolerant networks of computers (preliminary version). In *Proceedings 20th ACM Symposium on Theory of Computing* 66–77. ACM Press, New York.

BOXMA, O. and DADUNA, H. (1990). Sojourn times in queueing networks. In *Stochastic Analysis of Computer Communication Systems* (H. Takagi, ed.). North-Holland, Amsterdam.

BOXMA, O. and GROENENDIJK, W. (1988). Waiting times in discrete-time cyclic-service systems. *IEEE Trans. Comm.* **36** 164–170.

CHEN, R. and SHEPP, L. (1991). Synchronous service on a circle. *Adv. in Appl. Probab.* **23** 140–151.

CHOW, W.-M. (1980). The cycle time distribution of exponential cyclic queues. *J. Assoc. Comp. Mach.* **27** 281–286.

FELLER, W. (1968). *An Introduction to Probability Theory and Its Applications* **1**, 3rd ed. Wiley, New York.

GORDON, W. and NEWELL, G. (1967). Closed queueing systems with exponential servers. *Operat. Res.* **15** 254–265.

JOHNSON, N. and KOTZ, S. (1977). *Urn Models and Their Application.* Wiley, New York.

KOBAYASHI, H. and KONHEIM, A. (1977). Queueing models for communications system analysis. *IEEE Trans. Commun.* **25** 2–29.

KOENIGSBERG, E. (1958). Cyclic queues. *Operat. Res. Quart.* **9** 22–35.

LIGGETT, T. (1985). *Interacting Particle Systems.* Springer, New York.

ONVURAL, R. (1990). Survey of closed queueing networks with blocking. *Comput. Surveys* **22** 83–121.

PESTIEN, V., RAMAKRISHNAN, S. and SARKAR, D. (1993). Packet transmission in a noisy-channel ring network. *SIAM J. Comput.* To appear.

SCHASSBERGER, R. and DADUNA, H. (1983). The time for a round trip in a cycle of exponential queues. *J. Assoc. Comput. Mach.* **30** 146–150.

TSOUCAS, P. and WALRAND, J. (1989). Monotonicity of throughput in non-Markovian networks. *J. Appl. Probab.* **26** 134–141.

WALRAND, J. (1983). A discrete-time queueing network. *J. Appl. Probab.* **20** 903–909.

DEPARTMENT OF MATHEMATICS
AND COMPUTER SCIENCE
UNIVERSITY OF MIAMI
CORAL GABLES, FLORIDA 33124-4250