Probability in the Engineering and Informational Sciences, 7, 1993, 209–226. Printed in the U.S.A.

CONTINUOUS POLLING ON GRAPHS

E. G. COFFMAN, JR.
AT&T Bell Laboratories
Murray Hill, New Jersey 07974

ALEKSANDR STOLYAR
Institute of Control Sciences
Profsoyuznaya ul. 65, GSP-312 Moscow, Russia

Past research on polling systems has been quite restricted in the form of the paths followed by the server. This paper formulates a general, continuous model of such paths that includes closed walks on graphs. Customers arrive by a Poisson process and have general service times. The distribution of arrivals over the path is governed by an absolutely continuous, but otherwise arbitrary, distribution. The main results include a characterization of the stationary state distribution and explicit formulas for expected waiting times. The formulas reveal an interesting decomposition of the system into two components: a fluid limit and an M/G/1 queue.

1. INTRODUCTION

Consider a server moving at constant speed along the edges of a given graph G in a circuit that traverses all edges. Edges may be traversed more than once and in either direction, and they have varying lengths. To move from one isolated component to another the server is allowed to jump instantaneously from a vertex in one to a vertex in the other. The simple examples shown in Figure 1 will be considered in more detail later.

Customers arrive by a Poisson process at rate λ per unit time and are placed at random locations on the edges of G according to a given, absolutely contin-

 0269-9648/93 $5.00 + .00

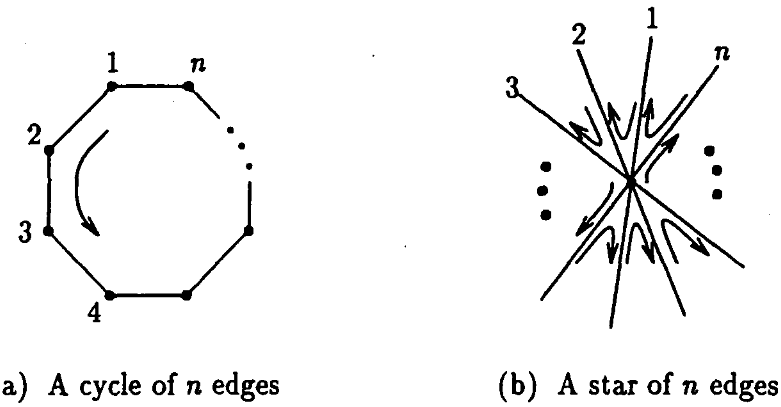

FIGURE 1. Examples of servers. (a) The server moves counterclockwise around the cycle. (b) The server scans the edges in a counterclockwise sequence; each edge is scanned by going from the center out to the end and then back to the center again (there are no instantaneous jumps).

uous distribution function. Whenever the server encounters a waiting customer, it stops, serves the customer (whereupon the customer departs), and then resumes its motion. At the end of a circuit the server immediately begins a new one. Service times s are independent samples from a given distribution.

The objectives of this paper are (1) the steady-state probability density describing the server's position in G, (2) the expected customer waiting time, and (3) a fluid limit in which $\lambda \to \infty$, $E[s] \to 0$, with the intensity $\rho = \lambda E[s]$ held constant. As a by-product of the analysis, we obtain useful decomposition results.

For convenience the preceding graph problems are reduced to the following general problem defined on sets of line segments. In a Poisson stream at rate λ customers arrive on the interval $[0, a]$, $a > 0$, at random locations determined by the probability density $f(x)$, $0 \le x \le a$. As before, the server scans the interval at constant speed, stopping to serve customers where they are encountered. The server's scan cycles through a fixed path $\mathcal{P} = \{(a_{11}, a_{12}), \ldots, (a_{r1}, a_{r2})\}$, $r \ge 1$, where $a_{i1}, a_{i2} \in [0, a]$ are the initial and final points of the ith path segment, $1 \le i \le r$, with $a_{11} = 0$. In scanning the segment (a_{i1}, a_{i2}) the server is moving from left to right over the subinterval $[a_{i1}, a_{i2}]$ if $a_{i1} < a_{i2}$, and from right to left over the subinterval $[a_{i2}, a_{i1}]$ if $a_{i1} > a_{i2}$. After scanning the segment (a_{i1}, a_{i2}) the server moves from $a_{i,2}$ to $a_{i+1,1}$ if $i < r$, jumping instantaneously if $a_{i,2} \ne a_{i+1,1}$. The server moves similarly from a_{r2} to a_{11}, whereupon a new cycle is started. Without loss of generality we assume that the entire interval $[0, a]$ is spanned by $\mathcal{P}$.

It is trivial to map any graph problem into an instance of this interval problem. Note that, because jumping, scan directions, and the number of times a subinterval is scanned are all unrestricted in general, graph structure is transparent in the latter model. The remainder of this paper analyzes the interval problem. For convenience we assume that the server moves at unit speed, but the formulas are easily modified to accommodate any fixed but arbitrary speed.

Two special cases of our system were considered by Coffman and Gilbert [1,2], Fuhrmann and Cooper [3,4], and Kroese and Schmidt [6]. In one the server scans in a fixed direction around a closed tour of length a. Coordinates of arriving customers are independent and uniformly distributed around the tour. Figure 1(a) illustrates a system of this type. However, vertices other than the first and last have no special significance, so the simplest equivalent instance of the interval problem is $\mathcal{P} = \{(0,a)\}, f(x) = 1/a, 0 \leq x \leq a$; the server repeatedly scans the interval $[0,a]$ left to right, jumping instantaneously back to 0 when it reaches a. In Coffman and Gilbert's [1] work steady-state distributions of waiting times and the number-in-system process were derived for constant service times. An explicit expression for the mean waiting time with a general service time distribution was first obtained by Fuhrmann and Cooper [3]; however, Kroese and Schmidt's [6] work contains a complete analysis of the steady-state regime of this system, including higher moments of waiting times.

The system with parameters $\mathcal{P} = \{(0,a),(a,0)\}, f(x) = 1/a$, was studied by Coffman and Gilbert [2]. (It specializes the star in Fig. 1(b) to one branch.) The server oscillates back and forth across the interval $[0,a]$. The fluid limit $\lambda \to \infty$, $E[s] \to 0$, with $\rho = \lambda E[s]$ held constant was analyzed by Coffman and Gilbert [2]. In the fluid limit a deterministic continuous flow of work replaces the stochastic flow of individual customers. The problem of characterizing this fluid limit in our more general model is an essential part of this paper,

The approach of this paper is similar to that of Kroese and Schmidt [6] in that we also consider an equivalent open system in which the server moves on the infinite axis. The set of points where customers arrive forms a Poisson cluster process. A crucial concept of this paper is that of an *embedded busy period* (EBP). The EBP is a structure of the Poisson cluster and can be interpreted as a busy period in a certain M/G/1 queue.

Section 2 introduces notation and discusses the EBP structures. The stationary regime and the fluid limit are characterized in Sections 3 and 4, respectively. Section 5 derives formulas for the steady-state density of the server position on $\mathcal{P}$ and the expected customer waiting time. New results for special cases are proved in Section 6 as corollaries to the results in Section 5.

2. PRELIMINARIES

Let $T = \sum_{i=1}^{r} |a_{i2} - a_{i1}|$ be the length of the path $\mathcal{P}$. Since the server moves at unit speed, T also denotes the time required by a nonstop traversal of $\mathcal{P}$. We say that the server is at a point $y \in [0, T)$ of $\mathcal{P}$ if it has moved a distance y from the beginning of the current cycle. The function $g(y) \in [0,a], y \in [0, T)$, gives the server's actual location in the interval $[0,a]$. This piecewise linear function is given by

$$g(y) = \frac{(y - \hat{a}_{i-1})a_{i2} + (\hat{a}_i - y)a_{i1}}{|a_{i2} - a_{i1}|}, \qquad y \in [\hat{a}_{i-1}, \hat{a}_i),\ 1 \le i \le r,$$

where $\hat{a}_0 = 0$, $\hat{a}_i = \sum_{j=1}^{i} |a_{j2} - a_{j1}|$, $i = 1, \ldots, r$, are the initial points of the path segments in $\mathcal{P}$. For reasons made clear later $g(y)$ is continued periodically with period T to the entire real axis $\mathbb{R} = (-\infty, \infty)$.

In analogy with the approach of Kroese and Schmidt [6] it is convenient to analyze an equivalent *open* system in which the path $\mathcal{P}$ is laid out on the real axis in consecutive intervals. Because of our interest in stationary distributions, this mapping will be on the entire axis $\mathbb{R}$. In this open system the server moves along $\mathbb{R}$ in the positive direction at unit speed when not serving. Thus, consecutive intervals $[iT, (i+1)T)$, $i = 0, \pm 1, \pm 2, \ldots$, correspond to consecutive cycles through $\mathcal{P}$. Each point $z \in \mathbb{R}$ corresponds to a point $y = z (\text{mod } T)$ of $\mathcal{P}$ and a point $x = g(z)$ of the "physical" interval $[0, a]$.

To map the arrival mechanism onto the open system, we consider arriving customers as having i.i.d. location parameters, called *labels*, distributed on $[0, a]$ with density f. If the server is at location z on $\mathbb{R}$ at the instant a customer having label x arrives, then the customer is placed at the point $\zeta(x, z)$ on $\mathbb{R}$, where

$$\zeta(x, z) = \inf\{z' \ge z : g(z') = x\}, \qquad z \in \mathbb{R},\ x \in [0, a],$$

is the point z' closest to and to the right of z which corresponds to the point x of the physical interval $[0, a]$. The function $\zeta(x, z) - z$ is periodic in z with period T. Figure 2 illustrates the functions $g(y)$ and $\zeta(x, z)$ for the star system in Figure 1(b).

A customer that arrives while the server is moving is called a *root* customer. If a customer arrives while the server is busy, then it is called a *first-generation descendant* of the customer being served. Descendants of the ith generation are defined in the obvious way; a customer is either a root customer or in the ith generation of a root customer, for some $i \ge 1$.

Let $\eta(z)$ denote the nearest point in $\mathbb{R}$ to the left of z which corresponds to the same point $g(z) \in [0, a]$, i.e., $\eta(z) = \sup\{z' < z : g(z') = g(z)\}$. The function $z - \eta(z)$ is periodic with period T, as illustrated in Figure 2. Now consider the probability that the server encounters a root customer in a small interval $[z, z + \Delta z]$ of $\mathbb{R}$. Because root customers arrive only while the server is moving, this probability must be proportional to the time spent moving from $\eta(z)$ to z (not including intervening service times). Clearly, it must also be proportional to the arrival rate λ and the density $f(g(z))$. Thus, because λT is the average number of arrivals during the time spent moving in one cycle, the probability that the server encounters a root customer in $[z, z + \Delta z]$ is $\lambda(z)\Delta z + o(\Delta z)$, where

$$\lambda(z) = [z - \eta(z)]\lambda f(g(z)), \tag{2.1}$$

is periodic with period T and satisfies $\int_0^T \lambda(z)\, dz = \lambda T$. Thus, the points on the axis $\mathbb{R}$ where root customers are encountered by the server form a Poisson process with the variable space rate $\lambda(z)$:

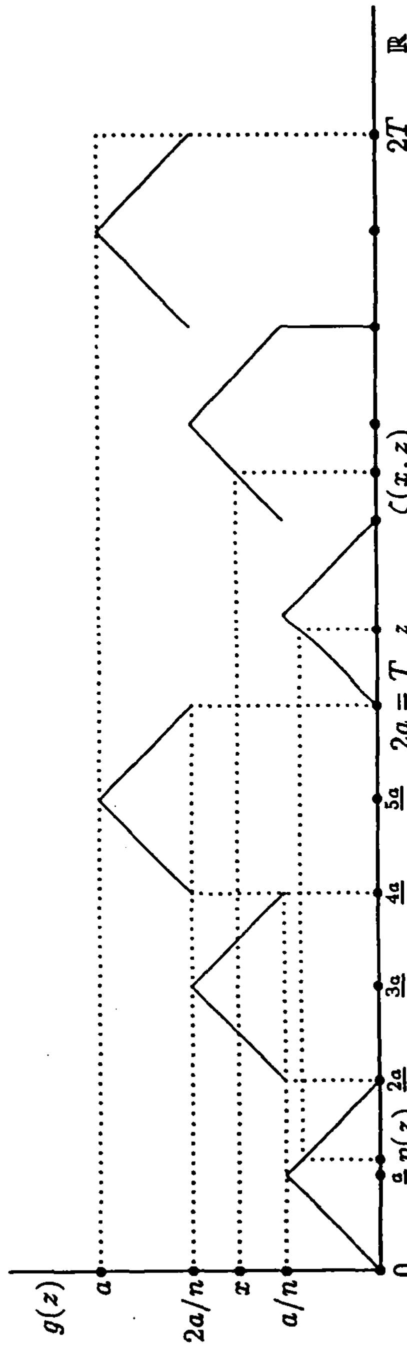

FIGURE 2. Functions $g(z)$, $\eta(z)$, and $\zeta(x,z)$ for the star system, $n = 3$.

A root customer together with all of its descendants forms an EBP. A sample EBP with a root customer located at $z \in \mathbb{R}$ is a random structure $\xi^{(z)} = (s_0, \ldots, s_k; d_1, \ldots, d_k)$, where s_0 is the service time of the root customer, s_i, $1 \le i \le k$, is the service time of the ith descendant, and d_i, $1 \le i \le k$, is the distance between the root customer and the ith descendant. For convenience suppose the descendants are numbered so that $d_1 \le \cdots \le d_k$. Let $G \subseteq \bigcup_{k=1}^{\infty} \mathbb{R}_+^{2k-1}$ denote the EBP sample space, and let $\beta(\xi^{(z)}) = \sum_{i=0}^{k} s_i$ denote the duration of $\xi^{(z)}$. Hereafter, the term *location of an EBP* refers to the location of its root customer.

The study of busy period structures in polling systems is not new. Recently, an analysis of such structures was given by Konheim and Levy [5] for the purpose of deriving improved numerical methods for discrete polling systems.

For simplicity we assume that the service-time random variable is absolutely continuous. Then a standard busy period analysis shows that, if $\rho = \lambda E[s] < 1$, then $\xi^{(z)}$ has a distribution $Q^{(z)}$ with a density $q^{(z)}$ such that $\beta(\xi^{(z)})$ has finite moments independent of z, the first being

$$\bar{\beta} \equiv E[\beta(\xi^{(z)})] = \frac{E[s]}{1-\rho}. \tag{2.2}$$

Both $Q^{(z)}$ and $q^{(z)}$ are periodic in z with period T. If a customer at z is a first-generation descendant of a customer at z', then $z - z' < T$. Thus, the expected length of a smallest interval $[z, z']$ in $\mathbb{R}$ containing the locations of all of the customers in an EBP $\xi^{(z)}$ is bounded by $\lambda \bar{\beta} T < \infty$ for all z.

A history of the arrival and service processes can be represented as a sequence of EBPs at locations $\ldots, \tilde{z}_{-2}, \tilde{z}_{-1}, \tilde{z}_0, \tilde{z}_1, \tilde{z}_2, \ldots$ as shown in Figure 3, where vertical segments are placed at the customer locations on $\mathbb{R}$ with lengths equal to the corresponding service times. The definitions and observations of this section are summarized as follows.

PROPOSITION 2.1: *A sample on $\mathbb{R}$ of customer service times and locations is given by a sequence of EBPs, $\ldots, \xi^{(\tilde{z}_{-1})}, \xi^{(\tilde{z}_0)}, \xi^{(\tilde{z}_1)}, \ldots$ with the following properties: (i) the locations $\ldots, \tilde{z}_{-1}, \tilde{z}_0, \tilde{z}_1, \ldots$ are the epochs of a Poisson process on $\mathbb{R}$ with the rate parameter $\lambda(z) = [z - \eta(z)]\lambda f(g(z))$, and (ii) given their locations, the EBPs are independent samples from the respective distributions $\ldots, Q^{(\tilde{z}_{-1})}, Q^{(\tilde{z}_0)}, Q^{(\tilde{z}_1)}, \ldots$.*

3. STATIONARY BEHAVIOR

It will be useful to modify the arrival mechanism, keeping the rate parameter fixed, so that customers arrive only at those times when root customers are encountered by the server. Thus, root customers are placed on $\mathbb{R}$ only when their positions are reached by the server; when a root customer location $z \in \mathbb{R}$ is encountered, an entire sample $\xi^{(z)}$ is drawn from $Q^{(z)}$ and placed on $\mathbb{R}$. Although arrival times have been shifted, samples of the process expressed in terms of EBPs remain the same. Thus, Proposition 2.1 still applies to the new system,

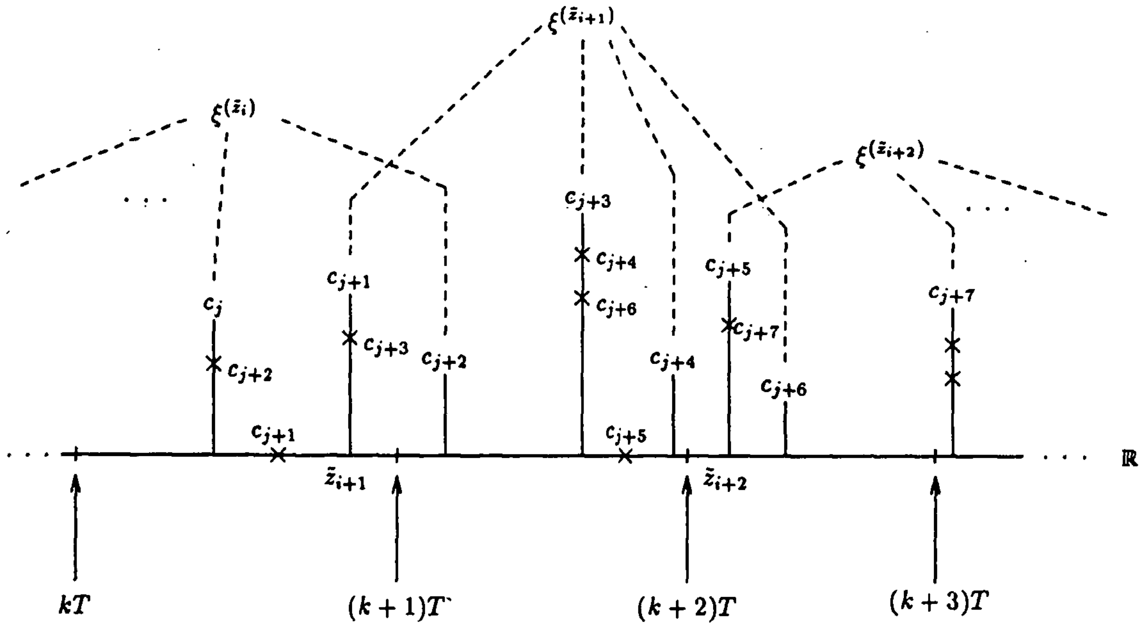

FIGURE 3. A sample function. The axis $\mathbb{R}$ denotes distance traveled, and the vertical segments denote customer service times on the same scale, since the server moves at unit speed. The server moves across the horizontal segments and stops at the vertical segments for times equal to the segment heights. Thus, the union of horizontal and vertical segments measures elapsed time. The locations of successive EBP root customers are denoted $\tilde{z}_i, \tilde{z}_{i+1}, \ldots$. Vertical segments are labeled on top by customer names c_j in order of their locations on $\mathbb{R}$. Crosses denote arrival times, where an arrival time is the sum of the lengths of all complete vertical and horizontal segments preceding the position of the cross plus the length of the partial (vertical or horizontal) segment extending up to the cross. Corresponding customer names are given next to each cross. On the axis $\mathbb{R}$ the coordinates of the crosses indicate the server locations at the times of arrival.

and the fluid limit of Section 4 remains the same for the new system. The fluid limit, Proposition 2.1, and the properties of the stationary distribution in Proposition 3.1 provide the entire basis for the results in Sections 5 and 6. Thus, the latter results will apply to both the modified and the original systems. Hereafter, the open system is assumed to have the modified arrival mechanism, unless noted otherwise.

To simplify the description of the stationary regime, the following "normalization" of $\mathbb{R}$ will be very helpful. At a random time t in steady state, if the server is moving, then the origin of $\mathbb{R}$ is assumed to coincide with the point where the server last visited the origin of the path $\mathcal{P}$; thus, the position of a moving server is always a point in $[0, T)$. If the server is busy at time t, then the origin of $\mathbb{R}$ is chosen so that the location of the *active* EBP, i.e., the EBP containing the customer being served, is in $[0, T)$. (In this case, the server position is not necessarily in $[0, T)$.) This convention will be discussed further after we define the system state.

We first define the *server state* at time t as a random structure $\sigma_t = ((\xi, \tau)_t, z_t)$ interpreted as follows. If the server is busy, the component $(\xi, \tau)_t$ consists of the entire sample of the active EBP ξ along with the total service time τ already received by customers of ξ; the component z_t gives ξ's location on $[0, T)$, consistent with the preceding normalization. If the server is idle and hence moving, then the first component takes on the null value $(\xi, \tau)_t = *$, and the component z_t denotes the server's location on $[0, T)$, again consistent with the preceding normalization. In a busy state $((\xi, \tau)_t, z_t)$, τ and the parameters of ξ uniquely determine the customer being served and the point where it is located relative to the location z_t of ξ. Thus, a busy state $((\xi, \tau)_t, z_t)$ uniquely determines the server location on $\mathbb{R}_+$ at time t. To define a complete state we adjoin to the server state the full sequence $\vec{\xi}_t = ((z_i, \xi^{(z_i)}),\ i = 1, 2, \dots)$ of *passive* EBP structures and their locations $z_1 \geq z_2 \geq \cdots$ on $\mathbb{R}$. Here, a passive EBP at time t is any EBP that arrived before time t, except for the active one, if any, in σ_t.

Consider now the transitions of the process $(\sigma_t, \vec{\xi}_t)_{t \geq 0}$ within the earlier normalization. While customers are being served, τ is the only state parameter to change. Suppose a customer completes service at time t with the server located at $z \in \mathbb{R}_+$. The new, normalized server state will be $(*, z(\text{mod } T))$. The new sequence of passive EBPs is formed by first inserting the previously active EBP along with its location z_t as a new pair $(z_i, \xi^{(z_i)})$ in the old sequence of EBPs, the index i being chosen so as to retain the ordering $z_1 \geq z_2 \geq \cdots$. Then, to be consistent with the normalized server state, $z - z(\text{mod } T)$ is subtracted from all of the passive EBP locations, including the one just inserted; this defines the sequence of passive EBP structures and locations in the new state.

While the server is moving, z_t increases and is the only state parameter to change, except when the server passes the end of $\mathcal{P}$, at which times z_t resets to 0 and all passive EBP locations jump by $-T$, thus preserving relative locations in the normalized state. Suppose that at time t the server stops in order to begin serving a customer C at point $y \in [0, T)$. (Note that z_t along with $\vec{\xi}_t$

uniquely specifies the locations of customers still waiting while the server is moving.) If C is a root customer of a new EBP, say ξ_0, that has just arrived, then the new (busy) server state is $((\xi_0,0),y)$ and the sequence of passive EBPs and their locations remain unchanged in the new state. If C is in a passive EBP, say $\xi^{(z)}$ at location $z \in \mathbb{R}$, then the pair $(z,\xi^{(z)})$ is deleted from the sequence of passive EBPs and the new server state is $((\xi^{(z)},\tau), z(\text{mod } T))$, where τ is the sum of the service times of the customers already served in $\xi^{(z)}$; these are all closer to the root customer than C is. The remaining passive EPBs all have their locations increased by $|z - z(\text{mod } T)|$ in the new state, in order to retain their proper locations relative to the normalized location of the active EBP.

From the preceding discussion it is clear that $(\sigma_t, \vec{\xi}_t)_{t\geq 0}$ is a Markov process. Although it is not difficult to formulate more compact Markov processes for our polling problem, the process $(\sigma_t, \vec{\xi}_t)_{t\geq 0}$ has the advantage that we can prove a strong yet simple characterization of its stationary distribution. As part of this characterization, we identify a decomposition property similar to those of Fuhrmann and Cooper [4]. Before giving this result we comment briefly on the stability condition.

The analysis of M/G/1 queues suggests that $\rho < 1$ must hold if $(\sigma_t, \vec{\xi}_t)_{t\geq 0}$ is to have a stationary distribution; but the server spends time moving between customers, so at first glance it might appear that $\rho < 1$ is not sufficient. However, as noted by Coffman and Gilbert [1] for a special case, with any fixed server speed, the fraction of time the server spends moving tends to 0 as $\rho \to 1$. On this basis one expects $\rho < 1$ to be the stability condition independent of server speed.

PROPOSITION 3.1: *If $\rho < 1$, then there exists a unique stationary distribution of the process $(\sigma_t, \vec{\xi}_t)_{t\geq 0}$, described as follows. The stationary distribution of the server state has the density*

$$p(\xi,\tau;u) = \begin{cases} (1-\rho)/T, & \text{if } (\xi,\tau) = * \\ \rho \dfrac{\lambda(u)}{\lambda T} \dfrac{q^{(u)}(\xi)}{\bar{\beta}}, & \text{if } (\xi,\tau) \neq *, \end{cases} \tag{3.1}$$

periodic in u with period T. The sequence $\vec{\xi}_t$ of passive EBPs at a random time t in steady state, together with the EBPs that arrive after time t, forms a Poisson process on $\mathbb{R}$ with rate parameter $\lambda(u)$. Given their locations, these EBPs are mutually independent, and each is independent of the server state σ_t.

PROOF: The density p can be explained by standard informal arguments. Consider a random time t in statistical equilibrium and sample paths of EBPs as shown in Figure 3. The time instant t can be interpreted as a point on a vertical segment or a point on the axis $\mathbb{R}$. The mean total duration of all EBPs with root customers located in an interval $[iT,(i+1)T]$ is

$$\int_0^T \bar{\beta}\lambda(u)\,du = \bar{\beta}\lambda T = \frac{\rho}{1-\rho}\,T,$$

so by the law of large numbers, vertical segments comprise a fraction ρ of the sample measure (union of vertical and horizontal segments). Then with probability $1-\rho$, t is sited on the axis (i.e., the server is idle and $(\xi,\tau)=*$); and with probability ρ, t is sited on a vertical segment (i.e., the server is busy).

Conditioned on the server being idle, the server is moving at constant speed, so its normalized position on $[0,T)$ has the uniform density $1/T$. Conditioned on the server being busy, the joint density of ξ and its normalized location in $[0,T)$ are proportional to $\lambda(u)q^{(u)}(\xi)\beta(\xi)$ and, hence, equal to $(\lambda(u)/\lambda T)\cdot(q^{(u)}(\xi)\beta(\xi)/\bar{\beta})$, because

$$\int_{(\xi,u)}\lambda(u)q^{(u)}(\xi)\beta(\xi)\,du\,d\xi=\bar{\beta}\lambda T.$$

Given ξ and its location, the variable τ is distributed uniformly on the interval $[0,\beta(\xi)]$, so its conditional density is $1/\beta(\xi)$. Formula (3.1) follows.

A rigorous proof can be streamlined by analyzing an even simpler modified system, which we will call the *reduced system*. In the reduced system all of the customers of an EBP $\xi^{(z)}$ are served while the server is stopped at point z. Sample paths analogous to Figure 3 now have vertical segments with heights equal to the duration of entire EBPs. The overall service order has changed, although the ordering of customer service within EBPs is preserved. The state definition remains unchanged in the reduced system, but the state transitions are much simpler. Note that z_t in σ_t always gives the normalized server location, whether σ_t is busy or not, and that once an EBP becomes passive it remains so. It is easy to verify that Propositions 2.1 and 3.1 hold for the reduced system if and only if they hold for the original modified system. For the reduced system the earlier arguments apply without change to the density in Formula (3.1). By Proposition 2.1 and elementary properties of the Poisson process, the remaining assertions are easily seen to apply to the reduced system and, hence, to the original modified system.

Based on the reduced system, a rigorous proof that $\rho<1$ is a necessary and sufficient condition for the existence of a unique stationary distribution has been formulated using classical techniques. The procedure has yielded no new insights into the polling problem, so we omit the details. ∎

4. THE FLUID LIMIT

In the fluid limit, also called the *snowplow* limit by Coffman and Gilbert [2], we let $\lambda\to\infty$, $E[s]\to 0$, with $\rho=\lambda E[s]$ held constant. Convergence of the underlying stochastic process will be of no concern in what follows, only the convergence of mean waiting times. As we will see in Section 5, certain properties of the system are preserved in the fluid limit; these properties depend on the arrival rate and service times only through the constant ρ.

The fluid limit consists of a deterministic continuous flow of work (cumulative service time) arriving on the physical interval $[0,a]$ with rate ρ. The distribution of the work $\rho\Delta t$ that arrives during time Δt has the density $\rho\Delta t f(x)$,

$x \in [0,a]$. Let $\varphi(y)$, $y \in [0,T]$, denote the density of work encountered by the server at point y of the path $\mathcal{P}$ in steady state. Besides having to perform the work $\varphi(y)\Delta y$ in moving through the interval $[y, y+\Delta y] \subset [0,T]$, the server also incurs a delay Δy because of its motion at unit speed; i.e., the actual speed at point y is $[1+\varphi(y)]^{-1}$. The existence and uniqueness of the density $\varphi(y)$ in the fluid limit is given by the following result, which will be needed in Section 5.

THEOREM 4.1: *If $\rho < 1$, then there exists a unique nonnegative function $\varphi(z)$ that is periodic with period T and satisfies*

$$\varphi(z) = \rho f(g(z)) \int_{\eta(z)}^{z} [1+\varphi(y)]\, dy, \qquad z \in \mathbb{R}. \tag{4.1}$$

This function satisfies the normalization condition

$$\int_0^T \varphi(z)\, dz = \rho T/(1-\rho). \tag{4.2}$$

Remark: Note that Eq. (4.1) can be interpreted as the density of work accumulated at z since the server was last at a point on $\mathbb{R}$ corresponding to z (i.e., at the point $\eta(z)$).

PROOF: Consider the set $\mathcal{F}$ of functions on $\mathbb{R}$ that are periodic with period T and satisfy

$$\|\alpha\| = \int_0^T |\alpha(y)|\, dy < \infty, \qquad \alpha \in \mathcal{F}.$$

(Thus, we are actually considering the space $L_1([0,T))$.) Define the operator $A: \mathcal{F} \to \mathcal{F}$ by

$$(A\alpha)(z) = \rho f(g(z)) \int_{\eta(z)}^{z} \alpha(y)\, dy, \qquad z \in \mathbb{R}.$$

The following property holds:

$$\int_0^T (A\alpha)(z)\, dz = \rho \int_0^T \alpha(z)\, dz, \qquad \alpha \in \mathcal{F}. \tag{4.3}$$

To see this, substitute into the left-hand side of Eq. (4.3) and write

$$\begin{aligned}\int_0^T \rho f(g(z))\, dz \int_{\eta(z)}^{z} \alpha(y)\, dy &= \rho \int_{-T}^{T} \alpha(y)\, dy \int_{\{z \in [0,T): z=\zeta(x,y), x \in [0,a]\}} f(g(z))\, dz \\ &= \rho \int_0^T \alpha(y)\, dy \int_{\{z=\zeta(x,y), x \in [0,a]\}} f(g(z))\, dz \\ &= \rho \int_0^T \alpha(y)\, dy \int_{x \in [0,a]} f(x)\, dx = \rho \int_0^T \alpha(y)\, dy.\end{aligned}$$

Note that if α is a nonnegative function, then $A\alpha$ is also nonnegative, and $\|A\alpha\| = \rho\|\alpha\|$.

Next, with reference to Eq. (4.1), let $B: \mathcal{F} \to \mathcal{F}$ denote the nonlinear operator

$$(B\alpha)(z) = \rho f(g(z)) \int_{\eta(z)}^{z} [1 + \alpha(y)]\, dy, \qquad \alpha \in \mathcal{F},$$

or, equivalently, $B\alpha = A\omega_1 + A\alpha$, where $\omega_1(z) \equiv 1$, $z \in \mathbb{R}$. This operator is contractive because $\rho < 1$ and

$$\|B\alpha_1 - B\alpha_2\| = \|A(\alpha_1 - \alpha_2)\| \le \|A(|\alpha_1 - \alpha_2|)\| = \rho\|\alpha_1 - \alpha_2\|.$$

Thus, operator B has a unique fixed point φ.

To show that this fixed point must satisfy Eq. (4.2), consider the increasing sequence of nonnegative functions

$$B^n\omega_0 = \sum_{i=1}^{n} A^i\omega_1, \qquad n = 1,2,\ldots,$$

where $\omega_0(z) \equiv 0$, $y \in \mathbb{R}$. We have already shown that $\|B^n\omega_0 - \varphi\| \to 0$, $n \to \infty$, and consequently that $\|\varphi\| = \lim_{n\to\infty}\|B^n\omega_0\|$. By Eq. (4.3) we have $\|A^i\omega_1\| = \rho^i T$, $i = 1,2,\ldots$, so $\|B^n\omega_0\| = \sum_{i=1}^{n} \rho^i T \to \rho T/(1-\rho) = \|\varphi\|$, $n \to \infty$. ■

Let $\theta^{FL}(y)$ denote the density describing the server's position on the path $\mathcal{P}$ in the fluid limit. By Eq. (4.2)

$$\theta^{\mathrm{FL}}(y) = \frac{1 + \varphi(y)}{\int_0^T [1 + \varphi(u)]\, du} = \frac{(1-\rho)[1 + \varphi(y)]}{T}. \tag{4.4}$$

Suppose a small "element" of fluid with label x arrives when the server is at z. Then, the time for the server to move from z to $\zeta(x,z)$ is $\int_x^{\zeta(x,z)} [1 + \varphi(u)]\, du$. Averaging over the densities $f(x)$ and $\theta^{\mathrm{FL}}(y)$, $y = z(\mathrm{mod}\ T)$, we get the average waiting time in the fluid limit,

$$\overline{W}^{\mathrm{FL}} = \frac{1-\rho}{T} \int_0^T [1 + \varphi(y)]\, dy \int_0^a f(x)\, dx \int_x^{\zeta(x,y)} [1 + \varphi(u)]\, du. \tag{4.5}$$

For the moment we take Eq. (4.5) as a natural definition of $\overline{W}^{\mathrm{FL}}$; Section 5 confirms that $\overline{W}^{\mathrm{FL}}$ is in fact the limit of the expected steady-state waiting time as $\lambda \to \infty$, $E[s] \to 0$, with $\rho = \lambda E[s]$ held constant.

A useful companion function to $\varphi(y)$ is defined as

$$\psi(x) = \sum_{\{y \in [0,T]: g(x) = y\}} \varphi(y), \qquad x \in [0,a]. \tag{4.6}$$

Thus, $\psi(x)\Delta x + o(\Delta x)$ is the total work performed by the server in the piece $[x, x + \Delta x]$ of the physical interval during one cycle through the path $\mathcal{P}$. Then,

$$\psi(x) = \frac{\rho T}{1-\rho} f(x), \qquad x \in [0,a], \tag{4.7}$$

because $\psi(x)$ is proportional to $f(x)$ and $\int_0^a \psi(x)\,dx = \int_0^T \varphi(y)\,dy$.

5. SERVER POSITION AND EXPECTED WAITING TIMES

Consider a random time t in steady state. Let $\xi^{(z)}$ be any EBP other than the active one, if any, at time t; i.e., $\xi^{(z)}$ is either a passive EBP in $\vec{\xi}_t$ or an EBP that arrives after time t. Then, $\Phi(u_1,u_2;z)$ is defined as the expected total work contributed by $\xi^{(z)}$ to the interval $[u_1,u_2]$; i.e.,

$$\Phi(u_1,u_2;z) = \int_{\xi\in G} \sum_{i=0}^{k} s_i I(z+d_i \in [u_1,u_2]) q^{(z)}(\xi)\,d\xi, \tag{5.1}$$

where k, $s_i (0 \le i \le k)$ and $d_i (1 \le i \le k)$ are the parameters of ξ, $d_0 = 0$ by definition, and I is the set indicator function. By Proposition 3.1 the EBPs for which Φ is defined form a Poisson process on $\mathbb{R}$ with rate parameter $\lambda(z)$. Then,

$$\Phi(u_1,u_2) = \int_{-\infty}^{u_2} \lambda(z)\Phi(u_1,u_2;z)\,dz \tag{5.2}$$

gives the expected total work contributed by all such EBPs to the interval $[u_1,u_2]$.

It is easy to see that the density $\lim_{\Delta z\to 0}(\Phi(z,z+\Delta z))/\Delta z$ is a periodic nonnegative function that must satisfy Eq. (4.1). Then, by Theorem 4.1 we have

$$\lim_{\Delta z\to 0} \frac{\Phi(z,z+\Delta z)}{\Delta z} = \varphi(z) \tag{5.3}$$

with $\varphi(z)$ defined as the density of arriving work at point z in the fluid limit. This connection with the fluid limit establishes the following result.

THEOREM 5.1: *Let $\theta(y)$, $y \in [0,T]$, denote the density of the server's position on $\mathcal{P}$ in steady state. Then $\theta(y) = \theta^{\mathrm{FL}}(y)$, as given by Eq. (4.4).*

PROOF: By Proposition 3.1 $(\lambda(u)/\lambda T)\cdot(q^{(u)}(\xi)/\bar{\beta})$ is the conditional joint density of the active EBP structure and its location $y \in [0,T)$, given that the server is busy. Then, by Eq. (5.1) the conditional probability, $\hat{\theta}(y)\Delta y$, that the server's position is in $[y,y+\Delta y]$, given that the server is busy, can be expressed as

$$\hat{\theta}(y)\Delta y = \int_{-\infty}^{y+\Delta y} \frac{\lambda(z)}{\lambda T\bar{\beta}} \Phi(y,y+\Delta y;z)\,dz + o(\Delta y),$$

so by Eqs. (5.2) and (5.3)

$$\hat{\theta}(y)\Delta y = \frac{1}{\lambda T\bar{\beta}} \varphi(y)\Delta y + o(\Delta y).$$

The identity $\lambda\bar{\beta} = \rho/(1-\rho)$ then gives

$$\hat{\theta}(y) = \frac{\varphi(y)}{\rho T/(1-\rho)}, \qquad y \in [0, T], \tag{5.4}$$

as the conditional density of the server's position, given that the server is busy. Note that Eq. (5.4) implies Eq. (4.2). By Proposition 3.1 the unconditional density is therefore $\theta(y) = \rho\hat{\theta}(y) + (1-\rho)/T = (1-\rho)[1+\varphi(y)]/T$, as given in Eq. (4.4). ■

The following decomposition result shows that derivations of the mean waiting time reduce to the calculation of fluid limits. (Throughout this section waiting time excludes the time in service.) The result also reduces the problem of finding those paths through a given set of line segments that minimize expected waiting time to the deterministic problem of comparing the behavior of alternative fluid limits.

THEOREM 5.2: *The steady-state mean waiting time is*

$$\overline{W} = \overline{W}^{\mathrm{FL}} + \overline{W}^{\mathrm{MG1}} \tag{5.5}$$

where $\overline{W}^{\mathrm{FL}}$ *is given by Eq. (4.5) and*

$$\overline{W}^{\mathrm{MG1}} = \frac{\lambda E[s^2]}{2(1-\rho)}. \tag{5.6}$$

PROOF: Consider a customer arriving at time t in steady state. Since arrivals are Poisson, the state seen by arrivals is described by the stationary density in Proposition 3.1. Let x denote the customer's label, and let z be the server's location. Then the customer's waiting time can be broken down into three components.

$$W = \zeta(x,z) - z + W_1 + W_2, \tag{5.7}$$

where $\zeta(x,z) - z$ is the server's moving time to the location $\zeta(x,z)$ of the customer; W_1 is the delay created by customers of the active EBP with locations in $[z, \zeta(x,z)]$ ($W_1 = 0$ if the server is idle at time t); and W_2 is the total service time of all other customers to be served in $[z, \zeta(x,z)]$. If the server is busy at time t, then the customer being served contributes only its remaining time to W_1.

Consider W_2 first and note that it is the total work placed in the interval $[z, \zeta(x,z)]$ by all EBPs except the active one, if any, at time t. Then, by the definition of the function $\Phi(u_1, u_2)$ we have for fixed x and z $E[W_2(x,z)] = \Phi(z, \zeta(x,z))$. By Eq. (5.2) we can write

$$\zeta(x,z) - z + \Phi(z, \zeta(x,z)) = \int_z^{\zeta(x,z)} [1 + \varphi(u)]\, du. \tag{5.8}$$

Recall that f is the density of x, that x and z are independent, and that by Theorem 5.1 $\theta(y)$ is given by Eq. (4.4) as the density of $z(\text{mod } T)$. Thus, taking expected values in Eq. (5.8) gives Eq. (4.5) and, hence,

$$E[\zeta(x,z) - z + E[W_2(x,z)]] = \overline{W}^{\text{FL}}. \tag{5.9}$$

Then $\overline{W} = \overline{W}^{\text{FL}} + E[W_1]$ by Eqs. (5.7) and (5.9), so it remains to show that $E[W_1] = \overline{W}^{\text{MG1}}$.

To determine $E[W_1]$ let z_t denote the root customer location on $\mathcal{P}$, as specified in σ_t, given that the server is busy at time t. By Proposition 3.1 the conditional density governing z_t is $\lambda(u)/(\lambda T)$. Consider the M/G/1 queue that results when the server speed is taken to be infinite and when the customers that start the M/G/1 busy periods are always placed at location z_t of the path, irrespective of their labels; apart from the customers starting busy periods, sequencing is random and controlled as before by the path $\mathcal{P}$ and the label density $f(x)$. With $z_t = u$ fixed let $\overline{W}(u)$ denote the mean steady-state waiting time in this M/G/1 queue. The conditional distribution of the server state in this queue, given that the server is busy, is the same as the conditional distribution of the server state in the original system, given that the server is busy serving a customer in an EBP $\xi^{(u)}$. It follows from the definition of W_1 that the unconditional expectation of W_1 is given by

$$E[W_1] = \int_0^T \frac{\lambda(u)}{\lambda T}\, \overline{W}(u)\, du.$$

But $\overline{W}(u)$ is the expected waiting time in a work-conserving M/G/1 queue, so it is independent of u and the service order. This mean waiting time is given by $\overline{W}^{\text{MG1}}$ in Eq. (5.6), so $E[W_1] = \overline{W}^{\text{MG1}}$ and the proof is complete. ■

6. SPECIAL CASES

This section presents two consequences of Theorem 5.2. The first deals with the system in which a server cycles around a simple closed tour, the case $\mathcal{P} = \{(0,a)\}$ in our formulation. Generalizing a result of Fuhrmann and Cooper [3] we have the following corollary.

COROLLARY 6.1: *The mean waiting time in the system $\mathcal{P} = \{(0,a)\}$ is invariant to the form of the density $f(x)$ and given by*

$$\overline{W} = \frac{a + \lambda E[s^2]}{2(1-\rho)}. \tag{6.1}$$

PROOF: We have $T = a$, $g(y) = y(\text{mod } a)$, $\eta(y) = y - a$, and for $x, y \in [0,a]$

$$\zeta(x,y) = \begin{cases} x, & \text{if } x \geq y \\ x + a, & \text{if } y < x. \end{cases} \tag{6.2}$$

By Eq. (4.6) we have $\psi(y) = \varphi(y)$, $y \in [0,a]$, so from Eq. (4.7) we obtain $\varphi(y) = cf(y)$, $c = \rho a/(1-\rho)$. Thus, $\theta(y) = (1-\rho)[1 + cf(y)]/a$, and by Eq. (4.5)

$$\overline{W}^{\mathrm{FL}} = \frac{1-\rho}{a}\int_0^a [1 + cf(y)]\, dy \int_0^a f(x)\, dx \int_x^{\zeta(x,y)} [1 + cf(u)]\, du. \quad \textbf{(6.3)}$$

We have the following preliminary calculations. From Eq. (6.2)

$$\int_0^a \frac{dy}{a}(\zeta(x,y) - y) = \int_0^x \frac{dy}{a}(x-y) + \int_x^a \frac{dy}{a}(a + x - y)$$
$$= \int_{x-a}^x \frac{dy}{a}(x-y) = \frac{a}{2}$$
$$\text{for arbitrary } x \in [0,a) \quad \textbf{(6.4)}$$

and

$$\int_0^a f(x)\, dz \int_y^{\zeta(x,y)} f(u)\, du = \int_0^y f(x)\, dx \int_y^{x+a} f(u)\, du$$
$$+ \int_y^a f(x)\, dx \int_y^x f(u)\, du$$
$$= \int_y^{a+y} f(x)\, dx \int_y^x f(u)\, du = \frac{1}{2}$$
$$\text{for arbitrary } y \in [0,a). \quad \textbf{(6.5)}$$

Note also that

$$\int_0^a \int_0^a f(x)f(y)\, dx\, dy(\zeta(x,y) - y) = \frac{a}{2}, \quad \textbf{(6.6)}$$

since $(\zeta(y,x) - x) + (\zeta(x,y) - y) = \alpha$; therefore,

$$\int_0^a \int_0^a f(x)f(y)\, dx\, dy[(\zeta(y,x) - x) + \zeta(x,y) - y)] = a.$$

With the help of Eqs. (6.4)–(6.6), $\overline{W}^{\mathrm{FL}}$ in Eq. (6.3) evaluates to

$$\overline{W}^{\mathrm{FL}} = \frac{a}{2(1-\rho)}.$$

An application of Theorem 5.2 completes the proof. ■

The second result concerns the star system in Figure 1.

COROLLARY 6.2: *The steady-state mean waiting time in the star system with n branches of length a/n, and $f(x) = 1/a$, $x \in [0,a]$, is*

$$\overline{W} = \frac{a}{1-\rho}\left(1 - \frac{1}{n} + \frac{2}{3n^2}\right) + \frac{\lambda E[s^2]}{(1-\rho)}. \tag{6.7}$$

Remark: In the limit $n \to \infty$ we obtain

$$\overline{W} = \frac{\alpha}{1-\rho} + \frac{\lambda E[s^2]}{2(1-\rho)}.$$

This is to be expected because the star system with $n = \infty$ is equivalent to the system with a server scanning a closed tour of length $2a$ (see Corollary 6.1). For the case $n = 1$, Eq. (6.7) shows that $\overline{W}^{FL} = (2a/3)/(1-\rho)$, which checks with the result of Coffman and Gilbert [2]. Note also that, as one might expect, the results for $n = 1$ and $n = 2$ are the same.

PROOF: We have $T = 2a$. Note that, by symmetry, in addition to being periodic in y with period T, $\varphi(y)$ and $y - \eta(y)$ are periodic with period $T/n = 2a/n$. From Eq. (4.7) we get $\psi(y) = \rho T f(y)/(1-\rho) = 2\rho/(1-\rho)$, a constant over $[0,a)$. For any $x \in [0,a/n)$ the set $\{y \in [0,T) : g(y) = x\}$ consists of the two elements x and $2a/n - x$. Therefore, by Eq. (4.6) $\varphi(y) + \varphi(2a/n - y) = \psi(y)$. Then,

$$\varphi(y) + \varphi(2a/n - y) = 2\rho/(1-\rho), \qquad y \in [0,2a/n).$$

The function $\eta(y)$ for $y \in [0,2a/n]$ is given by

$$\eta(y) = \begin{cases} 2a/n - y - 2a, & y \in [0,a/n) \\ 2a/n - y, & y \in [a/n,2a/n). \end{cases}$$

Then, according to Eq. (4.1) we obtain for $y \in [a/n,2a/n)$

$$\begin{aligned}\varphi(y) &= \frac{\rho}{a}\int_{2a/n-y}^{y} [1 + \varphi(u)]\, du \\ &= \frac{\rho}{a}\int_{a/n}^{y} [2 + \varphi(u) + \varphi(2a/n - u)]\, du = \frac{2\rho(y - a/n)}{a(1-\rho)}.\end{aligned}$$

Thus, for $y \in [0,a/n)$ we can write

$$\varphi(y) = \frac{2\rho}{1-\rho} - \varphi(2a/n - y) = \frac{2\rho}{1-\rho} + \frac{2\rho(y - a/n)}{a(1-\rho)}.$$

This determines $\varphi(y)$ on $[0,2a/n)$ and, hence, on the entire axis $\mathbb{R}$ by periodicity.

The remainder of the proof consists of determining $\zeta(x,y)$ and then calculating $\overline{W}^{FL}$ by Eq. (4.5). The details are straightforward and left to the interested reader. ■

References

1. Coffman, E.G., Jr. & Gilbert, E.N. (1986). A continuous polling system with constant service times. *IEEE Transactions on Information Theory* IT-32: 584–591.

2. Coffman, E.G., Jr. & Gilbert, E.N. (1987). Polling and greedy servers on the line. *Queueing Systems* 2: 115–145.
3. Fuhrmann, S.W. & Cooper, R.B. (1985). Application of the decomposition principle in M/G/1 vacation models to two continuum cyclic queueing models—Especially token-ring LANs. *AT&T Technology Journal* 64: 1091–1098.
4. Fuhrmann, S.W. & Cooper, R.B. (1985). Stochastic decompositions in the M/G/1 queue with generalized vacations. *Operations Research* 33: 1117–1129.
5. Konheim, A.G. & Levy, H. (1992). Descendant set: An efficient approach for the analysis of polling systems. Preprint. Department of Computer Science, Tel Aviv University, Israel.
6. Kroese, D.P. & Schmidt, V. (1992). A continuous polling system with general service times. *Annals of Applied Probability* 2: 906–927.