The Annals of Applied Probability
1992, Vol. 2, No. 4, 906–927

A CONTINUOUS POLLING SYSTEM WITH GENERAL SERVICE TIMES

BY DIRK P. KROESE AND VOLKER SCHMIDT

University of Twente and University of Ulm

Consider a ring on which a server travels at constant speed. Customers arrive on the ring according to a Poisson process, at locations independently and uniformly distributed over the circle. Whenever the server encounters a customer, he stops and serves the client according to a general service time distribution. After the service is completed, the server removes the customer from the ring and resumes his round.

The model is analyzed by means of point processes and regenerative processes in combination with some stochastic integration theory. This approach clarifies the analysis of the continuous polling model and provides the means for further generalizations.

For every time t, the locations of customers that are waiting for service and the positions of clients that have been served during the last tour of the server are represented by random counting measures. These measures converge in distribution as $t \to \infty$. A recursive expression for the Laplace functionals of the limiting random measures is found, from which the corresponding kth moment measures can be derived.

1. Introduction. Queuing systems in which clients are served in a certain cyclic order are usually referred to as *cyclic service systems* or *polling systems*. Over the last few years a wide variety of cyclic service models has found application in telecommunication and reliability. Usually such polling systems are formulated in discrete time; see, for example [6], [12] and [14]. Although a discrete setting of the problem provides a rigorous basis for analysis, on the other hand, it often severely obscures the analysis due to combinatorial difficulties. Moreover, results obtained in this way are usually hard to generalize.

In [3] and [9] it was recognized that some discrete models could be fruitfully approximated by their continuous counterparts. The results for the continuous models are usually much more transparent. However, the existing theory on continuous models seems to lack a rigorous mathematical basis, which forms an obstacle for generalizations and further development of the theory. In this paper we generalize the continuous polling model of [3] and [9], using a new approach which makes use of random measure theory and some basic stochastic integration theory. This provides a convenient, clarifying and general way to describe such queuing models and leads to several new results. The main

Received April 1991; revised November 1991.

AMS 1980 *subject classifications*. Primary 60K25, 90B22, 60G55; secondary 60G57, 60H05.

Key words and phrases. Cyclic server system, cluster representation, random counting measures, limit distributions, Laplace functionals, moment formulas, stochastic decomposition, stochastic integration, regenerative processes.

results are Theorems 3.3 and 4.1. Given the abundance of different polling models, we remark that our model deals with variants of the zero-buffer case (see, e.g., [14]).

2. The model. Throughout this paper $(\Omega, \mathscr{H}, P)$ denotes the probability space in the background. For any topological space E, $\mathscr{B}(E)$ denotes either the Borel σ-algebra on E or the set of nonnegative measurable functions on E. The indicator function corresponding to a set A is written as I_A. The Lebesgue measure of a Borel set A of $\mathbb{R}^m$ $(m \in \mathbb{N})$ is denoted by $l(A)$. We will frequently write μf for the integral of a function f with respect to a (random) measure μ. For basic definitions and results on stochastic integration, random measures and point processes, we refer to the Appendix.

Let C be a circle with circumference one. Starting from an empty system, customers arrive according to a Poisson process with intensity a and drop independently of each other on the circle, according to a uniform distribution. A server moves on C with constant speed α^{-1} and stops to serve a customer whenever he encounters one on his way. The consecutive service times are i.i.d. random variables with distribution function (d.f.) F and ith moment e_i, $i = 1, 2, \ldots$. When a service has been completed, the server resumes his journey without changing direction. We assume that the service times are independent of the arrival process and the locations of the clients on C. For convenience we fix an origin O on C and assume that the server is in O at time $t = 0$. We denote by W_t the random counting measure on $[0, 1)$ corresponding to the customers that are waiting on C at time t, relative to the location of the server. Specifically (see Figure 1), if for a realization $\omega \in \Omega$, the server is in s at time t, and if n denotes the number of customers on the circle at that time, at locations $p_1, \ldots, p_n$ (for $n \geq 1$), then $W_t(\omega, \cdot)$ is defined by

$$W_t(\omega, A) = \begin{cases} \sum_i I_A \circ v_i, & \text{if } n \geq 1, \\ 0, & \text{if } n = 0, \end{cases} \qquad \text{for all } A \in \mathscr{B}[0, 1),$$

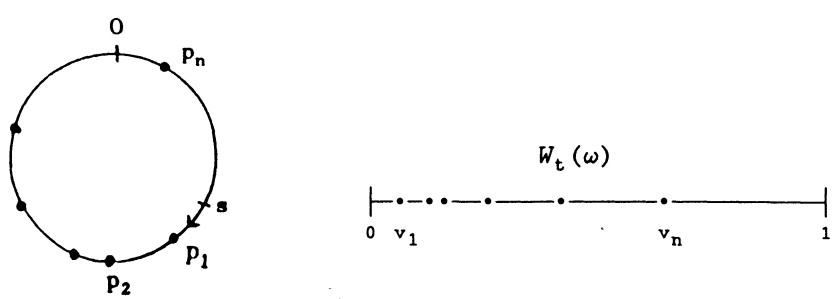

FIG. 1. *The random measure W_t represents the positions of customers that are waiting for service, relative to the position of the server, at time t (≥ 0).*

where

$$v_i = \begin{cases} p_i - s, & \text{if } p_i \geq s, \\ 1 - (s - p_i), & \text{if } p_i < s, \end{cases} \qquad i = 1, \ldots, n, \text{ for } n \geq 1.$$

Denote by τ_i the starting time of the ith service and let Z_i be its duration, $i = 1, 2, \ldots$. Furthermore, let U_1 be the time that the first service commences and denote the time between service completion of the $(i-1)$th and the start of the ith service by U_i, $i = 2, 3, \ldots$. Finally, let T_i be the arrival time of the ith customer, and let $S_i = U_1 + \cdots + U_i$, $i = 2, 3, \ldots$ and $S_1 = U_1$.

Since in the definition of W_t, we regard the locations of the customers on C *relative* to the location of the server, it is more convenient to view the polling system from the point of view of the server, as is done in Figure 2. Now the location of the server is fixed at 0, while the circle rotates with constant speed α^{-1}, provided that the server is not busy. For every t, the atoms of W_t are the locations of the customers on the circle at time t.

By continuing the customer paths *below* the X-axis in the same way as *above* the X-axis (see Figure 2), we construct a random counting measure H_t on $(0, 1]$, whose atoms are formed by the y-coordinates of the intersections of the customer paths and the line $x = t$. This measure describes the relative locations of the customers that have been served during the last tour of the server.

In the next section, we show that the stochastic process (W_t, H_t) is regenerative, and that the Laplace functionals of W_t and H_t can be analyzed more easily after applying a random time change. In order to accomplish this we define the "clock process" A_t as follows: For any realization $(U_i(\omega), Z_i(\omega))$, let

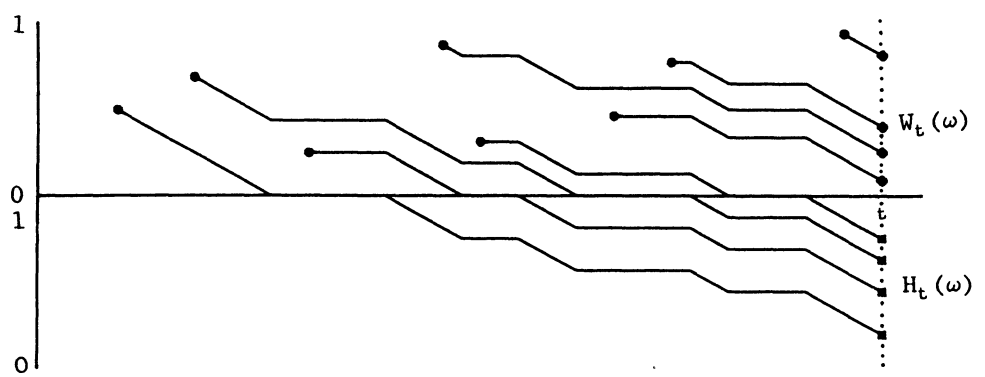

FIG. 2. *From the point of view of the server, customers arrive according to a homogeneous Poisson random measure on* $\mathbb{R}_+ \times [0, 1)$ *with intensity* a, *and move towards* 0 *with constant speed* α^{-1}, *unless the server is busy. Whenever a customer reaches the origin, all customers stop moving for a certain service period. After the service has been completed, the customer at the origin is removed and the other customers resume their journeys. The atoms of* $W_t(\omega)$ *and* $H_t(\omega)$ *are given by* (●) *and* (■), *respectively.*

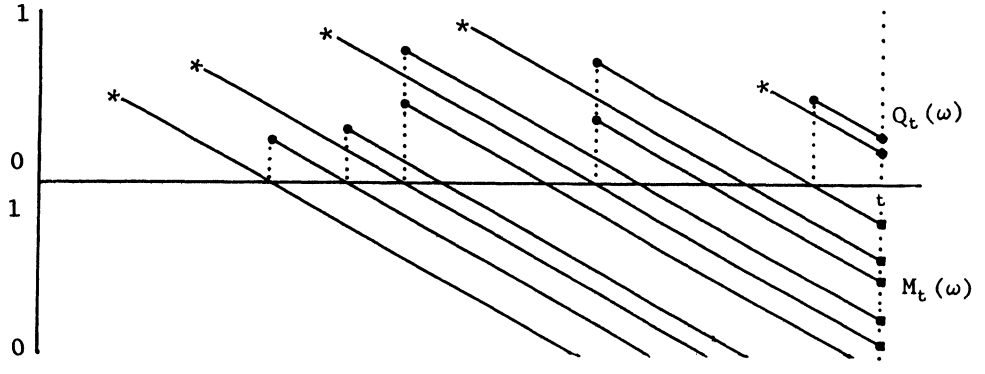

FIG. 3. *First generation particles are born via a Poisson random measure* (∗). *Whenever a particle hits the X-axis new particles can be born* (•).

$B = B(\omega)$ denote the set of times that the server is busy, and let

$$A_t(\omega) = \int_0^t dx\, I_{\bar{B}}(x), \qquad t \geq 0.$$

That is, we stop our clock when the server is busy. Let (ν_t) be the right-continuous functional inverse of (A_t) and define

$$Q_t = W_{\nu_t} \quad \text{and} \quad M_t = H_{\nu_t}, \qquad t \geq 0.$$

Obviously, Q_t and M_t are again random counting measures on $[0, 1)$ and $(0, 1]$, respectively. In the next section we show that, for $t \to \infty$, Q_t can be interpreted as the random measure of the locations of customers (relative to the server) on the circle at the new time t *given* that the server is not busy. A similar interpretation holds for M_t. A typical realization of Q_t and M_t is constructed from Figure 2 by first cutting away the customer paths on B and afterwards squeezing the remaining parts together. This leads to another way of looking at M_t and Q_t, namely through the following particle system (see Figure 3).

Particles emerge in $\mathbb{R}_+ \times [0, 1)$ in two ways. The first way is through a homogeneous Poisson random measure on $\mathbb{R}_+ \times [0, 1)$ with intensity a. We call these particles *first generation particles*. Immediately after their birth, all particles move downwards in a straight line with slope $-\alpha^{-1}$. A second type of birth occurs when a particle hits the X-axis. It then has the possibility to generate a number of new particles. Specifically, if the parent particle hits the X-axis at x, new points can be born at locations $(x, \xi_1), \ldots, (x, \xi_N)$, where $\xi_1, \xi_2, \ldots$ are i.i.d. r.v.'s, uniformly distributed over $[0, 1)$ and $P(N = n) = E \exp(-aZ)(aZ)^n/n!$, $n \in \mathbb{N}$. Here Z is a random variable with d.f. F. At time t, the atoms of Q_t are formed by the locations of the particles at time t in the interval $[0, 1)$ above the X-axis and similarly for M_t below the X-axis.

In the context of our polling system (of Figure 2), the first type of birth of a particle corresponds to the arrival of customers when the server is not busy

serving. The second way that particles are born corresponds to the arrival of customers when the server is busy serving. During a service period of length z, the number of clients N that arrive has a Poisson distribution with parameter az.

Note that by the definition of ν_t, the times that particles hit the X-axis in the particle system of Figure 3 are distributed as the times $S_1, S_2, \ldots$ of the original polling system. These points form the atoms of a *Poisson cluster process* PCP (cf. the Appendix). The cluster centers are the times that first generation particles hit the X-axis. The cluster belonging to a particular center is formed by the times that the descendants of the corresponding first generation particle hit the X-axis.

In the following sections, considering the processes (W_t), (H_t), (Q_t) and (M_t), we are concerned with the stochastic behavior of the customer *positions* on the circle. The investigation of customer waiting times, which is of equal interest, could be the subject of future research. For partial results on this subject we refer to [3].

3. Processes (W_t), (Q_t), (H_t) and (M_t). In this section we show that the processes (W_t), (H_t), (Q_t) and (M_t) are regenerative when the traffic intensity $ae_1 < 1$. Hence, they converge in distribution to limiting random measures W, H, Q and M, respectively. An intuitive interpretation of Q and M is given, and the connection between the Laplace functionals of W and Q (H and M) is established as a kind of stochastic decomposition result. This result is proved using some basic techniques from stochastic integration theory.

THEOREM 3.1. *If $ae_1 < 1$, then (Q_t, M_t) and (W_t, H_t) are regenerative processes, the regeneration cycles of which have an absolutely continuous distribution and finite expectation.*

PROOF. Denote the cluster centers of PCP (see previous section) by (X_n) and let L_i be the length of the ith cluster. Obviously, L_i is independent of (X_n) and $(L_n, n \neq i)$. Moreover, because $ae_1 < 1$, the expected cluster length EL_i is finite. This follows from the fact that in this case the total number of descendants produced by a single first generation particle (see Figure 3) is finite and that the distance on the X-axis between two consecutive particles of the same generation does not exceed α. As regeneration epochs of (Q_t, M_t), we take the arrival times of those first generation particles that arrive after intervals of length greater than or equal to α during which Q_t is the zero measure. The time Y between two such consecutive regeneration times has the same distribution as the time between two consecutive beginnings of busy periods in an M/GI/∞ queue [with arrival intensity a and service times (L_i)] which follow after an idle period of length α. Hence the regeneration cycle Y of (Q_t, M_t) has an absolutely continuous distribution and finite expectation; see, for example, Theorem 2.2 of [10]. The process (W_t, H_t) has a similar cluster structure as (Q_t, M_t). From Wald's lemma, we infer that the expected cluster

length is again finite. Thus, using analogous arguments as for (Q_t, M_t), we can show that (W_t, H_t) is regenerative and that the regeneration cycles have an absolutely continuous distribution and finite expectation. □

Using Theorem 3.1 and the key renewal theorem (cf. Theorem 9.2.8 of [2]), we get the existence of limiting random counting measures M, Q, W and H such that

$$M_t \to_{\mathscr{D}} M, \qquad Q_t \to_{\mathscr{D}} Q$$

and

$$W_t \to_{\mathscr{D}} W, \qquad H_t \to_{\mathscr{D}} H$$

as $t \to \infty$, where $\to_{\mathscr{D}}$ denotes convergence in distribution. The next theorem gives an interpretation of the limit random measure Q as the random measure of locations (relative to the server) of customers that are waiting, in the stationary situation, *given* that the server is *not* busy. Furthermore, a similar interpretation for M is given.

THEOREM 3.2. *For any $f \in \mathscr{B}[0, 1)$ and $g \in \mathscr{B}(0, 1]$,. we have*

$$\lim_{t\to\infty} E\left(e^{-W_t f} | W_t\{0\} = 0\right) = Ee^{-Qf}$$

and

$$\lim_{t\to\infty} E\left(e^{-H_t g} | W_t\{0\} = 0\right) = Ee^{-Mg}.$$

PROOF. Let Y and Y' denote the length of the first regeneration cycle of (Q_t, M_t) and (W_t, H_t), respectively. Note that since the system starts empty, we can take T_1 as the first regeneration time for both processes. Since Y' is also the length of the first regeneration cycle of $(\exp\{-W_t f)I_{\{W_t\{0\}=0\}}$, we get from the key renewal theorem,

$$\begin{aligned}\lim_{t\to\infty} E\left(e^{-W_t f} I_{\{W_t\{0\}=0\}}\right) &= \frac{1}{EY'} E\int_{T_1}^{T_1+Y'} ds\, e^{-W_s f} I_{\{W_s\{0\}=0\}} \\ &= \frac{1}{EY'} E\int_{T_1}^{T_1+Y} ds\, e^{-Q_s f} = \frac{EY}{EY'} Ee^{-Qf}.\end{aligned}$$

Similarly, we get

$$\lim_{t\to\infty} EI_{\{W_t\{0\}=0\}} = \frac{EY}{EY'}, \tag{3.1}$$

and the first part of the theorem follows. The second part is proved analogously. □

COROLLARY 3.1. *The limiting probability of being idle is given by*

$$\lim_{t\to\infty} P(W_t\{0\} = 0) = 1 - ae_1.$$

PROOF. Let Y and Y' be defined as in the proof of Theorem 3.2 and let $A_i = T_i - T_{i-1}$, $i = 2, 3, \ldots$ denote the interarrival times of customers. Let K denote the number of customers served during $[T_1, T_1 + Y')$. Obviously, $K + 1$ is the first regeneration time for the regenerative process (A_i). Hence by the key renewal theorem, we have $a^{-1} = E(A_2 + \cdots + A_{K+1})/EK = EY'/EK$ and similarly, $e_1 = E(Z_1 + \cdots + Z_K)/EK = (EY' - EY)/EK$. The theorem therefore follows from (3.1). □

Motivated by Theorem 3.2, we define measures Q^0 and M^0 by

$$\lim_{t\to\infty} E\big(e^{-W_t f} | W_t\{0\} = 1\big) = Ee^{-Q^0 f}$$

and

$$\lim_{t\to\infty} E\big(e^{-H_t g} | W_t\{0\} = 1\big) = Ee^{-M^0 g},$$

which means that we can view Q^0 as the random measures corresponding to customers that are waiting to be served (in the stationary situation), *given* that the server is *busy*. A similar interpretation (in terms of customers that have been served during the last cycle) holds for M^0. The next theorem shows how the Laplace functionals of Q^0 and M^0 can be determined from those of Q and M, respectively. Note that this theorem can be considered a generalization of the stochastic decomposition result of [8], Proposition 3 (called 5).

THEOREM 3.3. *For all continuously-differentiable functions f on $[0, 1)$, we have*

$$\begin{aligned} 0 = {} & (1 - ae_1) Ee^{-Qf}(\alpha^{-1} Qf' - \beta) \\ & + ae_1 Ee^{-Q^0 f}\left((e^{f(0)} - 1)\frac{\beta L_F(\beta)}{1 - L_F(\beta)} - \beta\right), \end{aligned} \tag{3.2}$$

and for all continuously-differentiable functions f on $(0, 1]$, with $f(0+) = 0$, the following holds:

$$0 = (1 - ae_1)\alpha^{-1} Ee^{-Mf} Mf' + ae_1(e^{-f(1)} - 1)\frac{\beta L_F(\beta)}{1 - L_F(\beta)} Ee^{-M^0 f}, \tag{3.3}$$

where $\beta = a\int_0^1 dx(1 - e^{-f(x)})$ and L_F is the Laplace–Stieltjes transform of F.

In order to prove Theorem 3.3 we need some preliminaries. Let $f \in C_+^1[0, 1)$ (the set of positive, continuously-differentiable functions on $[0, 1)$), and consider the stochastic process $(W_t f)$, starting at 0. Upward jumps occur via a Poisson process with rate a on $\mathbb{R}_+$. The size $f(\xi)$ of a jump is independent of everything else, where ξ is uniformly distributed on $[0, 1)$. Downward jumps have size $f(0)$. Let (A_t) denote the arrival counting process and (D_t) the departure counting process. Let $(\tilde{A}_t)$ be the compound Poisson process that jumps at arrival times (T_i), with jump sizes $(f(\xi_i))$, where the ξ_i's are uniformly distributed on $[0, 1)$ and independent of everything else. Let

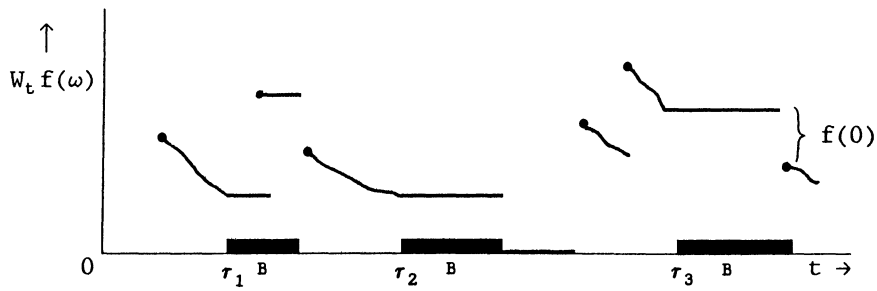

FIG. 4. *A realization of* $(W_t f)$. *For every* ω, $B(\omega)$ *denotes the set of times where the server is busy. Upward jumps correspond to arrivals of customers and have size* $f(\xi_i)$, *where* (ξ_i) *is an i.i.d. sequence of uniformly distributed r.v.'s on* $[0, 1)$. *Downward jumps correspond to departures of customers and have size* $f(0)$.

$(C_t) = (W_t f - \Sigma_{0<s\le t}\Delta W_s f)$ denote the continuous part of $(W_t f)$. Finally, let $B = B(\omega)$ denote the set of times that the server is busy, that is, $B(\omega) = \{t \ge 0\colon W_t(\omega, \{0\}) = 1\}$. A typical realization of $(W_t f)$ is given in Figure 4. For all $t \ge 0$,

$$W_t f = C_t - D_t f(0) + \tilde{A}_t. \tag{3.4}$$

LEMMA 3.1. *Let* (C_t) *be the continuous part of* $(W_t f)$ *and let* $(D_c(t))$ *be the compensator of* (D_t), *w.r.t. the filtration generated by* (W_t). *Then almost surely,*

$$\frac{d}{dt}C_t = -\alpha^{-1} I_{\bar{B}}(t) W_t f' \tag{3.5}$$

and

$$D_c(t) = \int_0^t I_B(s) \sum_i I_{\{\tau_i \le s < \tau_{i+1}\}} \frac{dF(s-\tau_i)}{1 - F(s-\tau_i)}. \tag{3.6}$$

PROOF. First observe that $C_t(\omega)$ is constant on $B(\omega)$. Next, suppose that $|W_t(\omega)| > 0$ on $\bar{B}(\omega)$. Let $x_1, \ldots, x_n$ be the atoms of $W_t(\omega)$; then

$$W_t(\omega) f = \sum_{i=1}^{n} f(x_i),$$

and, for sufficiently small $h > 0$,

$$W_{t+h}(\omega) f = \sum_{i=1}^{n} f(x_i - \alpha^{-1}h) = \sum_{i=1}^{n} f(x_i) - h\alpha^{-1} \sum_{i=1}^{n} f'(x_i) + o(h)$$

$$= W_t(\omega) f - h\alpha^{-1} W_t(\omega) f' + o(h).$$

Therefore,

$$\lim_{h \downarrow 0} \frac{W_{t+h} f - W_t f}{h} = -\alpha^{-1} W_t f' \quad \text{on } \bar{B}.$$

This is valid if $|W_t(\omega)| > 0$, but trivially also for $|W_t(\omega)| = 0$, so that (3.5) follows. As for the second part: (3.6) is a direct consequence of a well-known result in the theory of point processes (see, e.g., Theorem 13.2 III in [5]). □

PROOF OF THEOREM 3.3. We only prove the first part of the theorem. The second part is proved analogously. Throughout the proof, stochastic intensities, compensators, and so on, are always w.r.t. the filtration generated by (W_t). By (3.4) and Theorem A.2, we have

$$e^{-W_t f} = e^{-W_0 f} - \int_0^t e^{-W_s f}\, dC_s + \sum_{0<s\le t} [e^{-W_s f} - e^{-W_{s-} f}]. \tag{3.7}$$

Now

$$\begin{aligned} &\sum_{0<s\le t} [e^{-W_s f} - e^{-W_{s-} f}] \\ &\quad = \sum_{i=1}^{\infty} (e^{-f(\xi_i)} - 1) e^{-W_{T_i-} f} I_{(0,t]}(T_i) + (e^{f(0)} - 1) \int_0^t e^{-W_{s-} f}\, dD_s, \end{aligned} \tag{3.8}$$

where ξ_i is the distance that the server has to travel before he can serve the ith customer, $i = 1, 2 \ldots$. Note that ξ_i is independent of T_i and W_{T_i-}, and is uniformly distributed on $[0, 1)$. Taking expectations on both sides of (3.7) yields by (3.8), Lemma 3.1 and Theorem A.3 and the fact that the stochastic intensity of (A_t) is a,

$$\begin{aligned} Ee^{-W_t f} &= 1 + \alpha^{-1} \int_0^t ds\, Ee^{-W_s f} I_{\bar{B}}(s) W_s f' + a \int_0^1 dx (e^{-f(x)} - 1) \int_0^t ds\, Ee^{-W_s f} \\ &\quad + (e^{f(0)} - 1) E \int_0^t e^{-W_{s-} f} I_B(s) \sum_i I_{\{\tau_i \le s < \tau_{i+1}\}} \frac{dF(s - \tau_i)}{1 - F(s - \tau_i)}. \end{aligned} \tag{3.9}$$

If we subtract 1 from both sides of (3.9), divide by t and let $t \to \infty$, we obtain

$$\lim_{t\to\infty} \frac{1}{t} E(e^{-W_t f} - 1) \tag{3.10}$$

$$= \lim_{t\to\infty} \frac{1}{t} \int_0^t ds\, Ee^{-W_s f} I_{\bar{B}}(s) (\alpha^{-1} W_s f' - \beta) \tag{3.11}$$

$$\begin{aligned} &+ \lim_{t\to\infty} E \frac{1}{t} \int_0^t e^{-W_{s-} f} I_B(s) \\ &\quad \times \left((e^{f(0)} - 1) \sum_i I_{\{\tau_i \le s < \tau_{i+1}\}} \frac{dF(s - \tau_i)}{1 - F(s - \tau_i)} - ds\, \beta \right). \end{aligned} \tag{3.12}$$

Since $Ee^{-W_t f} - 1$ is bounded, (3.10) vanishes. Let Y, Y' and K be defined as in the proof of Corollary 3.1. Observe that the processes $(e^{-W_s f} I_{\bar{B}}(s) W_s f')$ and

$(e^{-W_s f} I_{\overline{B}}(s))$ are regenerative. Therefore, by the well-known time-average properties of regenerative processes (cf. Theorem V.3.1 of [1]) and Corollary 3.1, (3.11) becomes

$$\begin{aligned}(3.13)\qquad &\frac{1}{EY'} E\int_{T_1}^{Y'+T_1} ds\, e^{-W_s f} I_{\overline{B}}(s)\left(\alpha^{-1} W_s f' - \beta\right)\\ &= \frac{EY}{EY'}\frac{1}{EY} E\int_{T_1}^{Y+T_1} ds\, e^{-Q_s f}\left(\alpha^{-1} Q_s f' - \beta\right)\\ &= (1 - ae_1) Ee^{-Qf}\left(\alpha^{-1} Qf' - \beta\right).\end{aligned}$$

Analogously, (3.12) equals

$$(3.14)\qquad \frac{1}{EY'} E \sum_{i=1}^{K} \int_0^{Z_i} e^{-W_{\tau_i+s} f}\left((e^{f(0)} - 1)\frac{dF(s)}{1 - F(s)} - ds\,\beta\right).$$

Note that W_{τ_i+s} is distributed as $W_{\tau_i} + R_i(s)$, where $R_i(s)$ is the random measure of customers that arrive during $(\tau_i, \tau_i + s]$, $i = 1, 2, \ldots$. The R_i's are independent of the corresponding W_{τ_i} and have Laplace functional $Ee^{-R_i(s) f} = e^{-s\beta}$, $s \geq 0$. Define two regenerative processes (X_i) and $(\tilde{X}_i)$ by

$$X_i = \int_0^{Z_i} e^{-W_{\tau_i+s} f} \frac{dF(s)}{1 - F(s)} \quad \text{and} \quad \tilde{X}_i = \frac{\beta L_F(\beta)}{1 - L_F(\beta)} \int_0^{Z_i} ds\, e^{-W_{\tau_i+s} f},$$

$$i = 1, 2, \ldots.$$

It is easy to see that $EX_i = E\tilde{X}_i$, $i = 1, 2, \ldots$. Moreover, $K + 1$ is a regeneration time for both (X_i) and $(\tilde{X}_i)$. Hence, by the key renewal theorem, we have

$$\begin{aligned}(3.15)\qquad \frac{1}{EK} E \sum_{i=1}^{K} X_i &= \lim_{i\to\infty} EX_i = \lim_{i\to\infty} E\tilde{X}_i = \frac{1}{EK} E \sum_{i=1}^{K} \tilde{X}_i\\ &= \frac{1}{EK}\frac{\beta L_F(\beta)}{1 - L_F(\beta)}(EY' - EY) Ee^{-Q^0 f}.\end{aligned}$$

Therefore (3.2) follows from (3.13)–(3.15) and Corollary 3.1. □

4. The Laplace functionals of Q and M. In the previous section we have seen that the laws of W and H are completely specified by those of Q and M, respectively. It therefore suffices to focus on the distributions of these random measures. In this section, we derive an expression for the Laplace functionals of Q and M, from which several characteristics of these measures can be found.

As is explained in Section 2, we can construct the measures Q_t and M_t from the particle system of Figure 3. Note that, there, the initial locations of all particles form a two-dimensional cluster process N, say, in $\mathbb{R}_+ \times [0, 1)$. The cluster centers (the first generation particles) form a homogeneous Poisson random measure on $\mathbb{R}_+ \times [0, 1)$ with intensity a. Now for any $f \in \mathscr{B}[0, 1)$ and

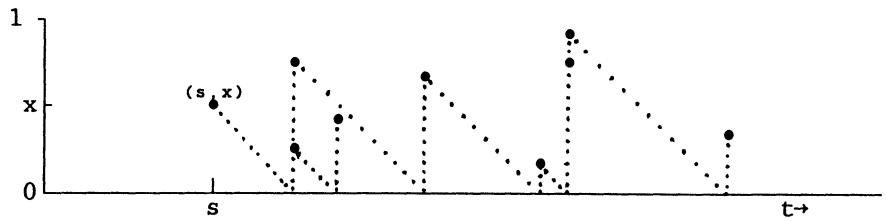

FIG. 5. *The points* • *form the atoms of a realization of the cluster* $L(s, x)$. *The atoms are constructed as in the particle system of Figure* 3.

$g \in \mathscr{B}(0, 1]$, we have $Q_t f = Nf_t$ and $M_t g = Ng_t$, where

$$f_t(s, x) = f(x - \alpha^{-1}(t - s))I_{[0,x]}(\alpha^{-1}(t - s))$$

and

$$g_t(s, x) = g(x + 1 - \alpha^{-1}(t - s))I_{[x,x+1]}(\alpha^{-1}(t - s)).$$

In particular,

$$Ee^{-Qf} = \lim_{t\to\infty} Ee^{-Nf_t} \quad \text{and} \quad Ee^{-Mg} = \lim_{t\to\infty} Ee^{-Ng_t}, \tag{4.1}$$

which is well defined by Theorem 3.1. The distributions of Q and M are therefore determined by the Laplace functional of N, for which the following result holds:

LEMMA 4.1. *For an initial (first generation) point* $(s, x) \in \mathbb{R}_+ \times [0, 1)$ *of the cluster process of Figure* 3, *let* $L(s, x)$ *denote the random counting measure on* $\mathbb{R}_+ \times [0, 1)$ *generated by this point (see Figure* 5). *Let*

$$G_f(s, x) = Ee^{-L(s,x)f} \quad \text{and} \quad H_f(s) = \int_0^1 dx\, G_f(s, x).$$

Then for every $f \in \mathscr{B}(\mathbb{R}_+ \times [0, 1))$,

$$Ee^{-Nf} = \exp - a\int_0^\infty ds\big(1 - H_f(s)\big) \tag{4.2}$$

and

$$G_f(s, x) = e^{-f(s,x)}G\big(H_f(\alpha x + s)\big), \tag{4.3}$$

where

$$G(z) = Ee^{-a(1-z)X}, \qquad z \in [0, 1], \tag{4.4}$$

X *being a random variable with d.f.* F.

PROOF. See Figure 5 for an illustration of random measure $L(s, x)$. From the recursive way the cluster is constructed, we obtain (4.3). (Note that G is the generating function of a compound Poisson distribution.) Denote the Poisson random measure of cluster centers by A, with atoms (S_i, X_i). Then

the conditional expectation of e^{-Nf} *given* A satisfies

$$E_A e^{-Nf} = E_A \exp - \sum_i L(S_i, X_i) f = \prod_i e^{-f(S_i, X_i)} G\big(H_f(\alpha X_i + S_i)\big)$$

$$= \exp - \int A(ds, dx)\big[f(s,x) - \log G\big(H_f(\alpha x + s)\big)\big].$$

Hence, since A is Poisson,

$$Ee^{-Nf} = \exp - a\int_0^\infty ds \int_0^1 dx\{1 - e^{-f(s,x) + \log G[H_f(\alpha x + s)]}\}$$

$$= \exp - a\int_0^\infty ds\{1 - H_f(s)\},$$

which proves the theorem. □

The following theorem is the main result of this section.

THEOREM 4.1. *Let* G *be defined by* (4.4). *Then*

$$Ee^{-Qf} = \exp - a\alpha\int_0^\infty ds\{1 - K(s)\}, \tag{4.5}$$

where K *is the unique solution of*

$$K(t) = \begin{cases} \int_0^t dx\, G(K(t-x)) + \int_t^1 dx\, e^{-f(x-t)}, & \text{for } 0 \le t \le 1, \\ \int_0^1 dx\, G(K(t-x)), & \text{for } t \ge 1, \end{cases} \tag{4.6}$$

and, moreover,

$$Ee^{-Mg} = \exp - a\alpha\int_0^\infty ds\{1 - V(s)\}, \tag{4.7}$$

where V *is the unique solution of*

$$V(t) = \begin{cases} \int_0^t dx\, e^{-g(x+1-t)} G(V(t-x)) + 1 - t, & \text{for } 0 \le t \le 1, \\ \int_0^{t-1} dx\, G(V(t-x)) & \\ \quad + \int_{t-1}^1 dx\, e^{-g(x+1-t)} G(V(t-x)), & \text{for } 1 \le t \le 2, \\ \int_0^1 dx\, G(V(t-x)), & \text{for } t \ge 2. \end{cases} \tag{4.8}$$

PROOF. By (4.1) and Lemma 4.1, we have

$$Ee^{-Qf} = \lim_{t\to\infty} \exp - a\int_0^\infty ds\big(1 - H_{f_t}(s)\big). \tag{4.9}$$

But if we write $H_{f_t}(u)$ as $E\exp - L(u, X)f_t$, where X is uniformly distributed on $[0, 1)$, then it is easy to see that

$$H_{f_t}(u) = \begin{cases} 1, & \text{for } u > t, \\ H_{f_{t-u}}(0), & \text{for } u \le t. \end{cases} \tag{4.10}$$

Hence, if we let $K(\alpha^{-1}t) = H_{f_t}(0)$, for all $t \ge 0$, then (4.5) follows from (4.9) and (4.10). Moreover, by (4.3), we have that

$$K(t) = \int_0^1 dx \exp(-f_{t\alpha}(0, x))G\big(H_{f_{t\alpha}}(\alpha x)\big); \qquad t \ge 0,$$

from which (4.6) follows straightforwardly. (4.7) and (4.8) are proved in the same fashion. □

5. The moment measures of Q and M. In this section we show how the moment measures of Q and M can be determined and give an explicit solution for the first and second moment measures. The proofs and intermediate results in this section mainly concern measure Q, since similar results for M follow by analogy.

By writing (4.6) in differential form, we obtain

$$K'(t) = \begin{cases} G(K(t)) - e^{-f(1-t)}, & \text{for } 0 \le t < 1, \\ G(K(t)) - G(K(t-1)), & \text{for } t > 1, \end{cases} \tag{5.1}$$

where K is a continuous function, almost everywhere differentiable except possibly at 1 and $K(0) = \int_0^1 dx\, e^{-f(x)}$. Since

$$\frac{d^n}{dp^n} Ee^{-pQf}\bigg|_{p=0} = (-1)^n E(Qf)^n$$

for $f \in \mathscr{B}[0, 1)$ and $n \in \mathbb{N}$, it follows by Theorem 4.1 that

$$E(Qf)^n = (-1)^n \frac{d^n}{dp^n} \exp - a\alpha\int_0^\infty ds\big\{1 - K_p(s)\big\}\bigg|_{p=0}, \tag{5.2}$$

where $K_p(s)$ is defined by (5.1) if pf is substituted for f. In particular, if we denote $(-1)^n(d^n/dp^n)K_p(t)|_{p=0}$ by $k_n(t)$, $n = 1, 2, \ldots$, we find

$$EQf = a\alpha\int_0^\infty ds\, k_1(s), \tag{5.3}$$

$$E(Qf)^2 = \left(a\alpha\int_0^\infty ds\, k_1(s)\right)^2 + a\alpha\int_0^\infty ds\, k_2(s), \tag{5.4}$$

$$\begin{aligned} E(Qf)^3 = {} & \left(a\alpha\int_0^\infty ds\, k_1(s)\right)^3 \\ & + 2\left(a\alpha\int_0^\infty ds\, k_1(s)\right)\left(a\alpha\int_0^\infty ds\, k_2(s)\right) + a\alpha\int_0^\infty ds\, k_3(s), \end{aligned} \tag{5.5}$$

and similar equations for higher order moments. From (5.1) we can derive differential equations for k_n, $n = 1, 2 \ldots$. In particular [remember (4.4)],

$$k_1'(t) = \begin{cases} ae_1 k_1(t) - f(1-t), & \text{for } 0 \le t < 1, \\ ae_1(k_1(t) - k_1(t-1)), & \text{for } t > 1 \end{cases} \tag{5.6}$$

and

$$k_2'(t) = \begin{cases} -f^2(1-t) + a^2 e_2 k_1^2(t) + ae_1 k_2(t), & \text{for } 0 \le t < 1, \\ a^2 e_2(k_1^2(t) - k_1^2(t-1)) & \\ \quad + ae_1(k_2(t) - k_2(t-1)), & \text{for } t > 1, \end{cases} \tag{5.7}$$

where the k_n's are continuous with $k_n(0) = \int_0^1 dt f^n(t)$, $n = 1, 2 \ldots$. Note that from the procedure indicated in (5.3)–(5.7), it follows that the nth moment measure of Q only depends on a, α and $e_1, \ldots, e_n$, $n = 1, 2, \ldots$. In order to calculate $E(Qf)^n$, it takes the (numerical) evaluation of a system of n ordinary differential equations, which can be done fairly easily. We do not know if it is possible to give an analytic solution for the nth moment measures of Q and M ($n \ge 3$), other than in the constant service time case (cf. Section 6). However, the first and second moment measures of Q and M can be explicitly given, as is done in the next two theorems. The first moment measure of Q can be easily derived from (5.3) and (5.6), using a Laplace transform of k_1. An equally simple argument, however, is provided by Theorem 3.3, which moreover gives us directly the mean measure of W. In a way we use a similar argument as in [9], where a stochastic decomposition result is used in the derivation of expected waiting times for several cyclic server models.

THEOREM 5.1. *The mean measures of Q and M are given by*

$$EQ(dx) = \frac{\alpha a}{1 - ae_1}(1 - x)\, dx \tag{5.8}$$

and

$$EM(dx) = \frac{\alpha a}{1 - ae_1}\, dx. \tag{5.9}$$

The expected number of customers on the circle is

$$E|W| = ae_1 + \frac{a(\alpha + ae_2)}{2(1 - ae_1)}. \tag{5.10}$$

PROOF. Let $c = \alpha a/(1 - ae_1)$. Since (3.2) holds for *any* f (arbitrarily "small"), we have in particular the functional equation

$$-a\int_0^1 dx f(x) + f(0)a + \alpha^{-1}(1 - ae_1) EQf' = 0, \qquad f \in C^1[0,1).$$

Therefore,

$$EQf' = c\int_0^1 dx(f(x) - f(0)) = c\int_0^1 dx \int_0^x du f'(u) = c\int_0^1 du(1-u)f'(u)$$

for all $f' \in C[0,1)$, which proves (5.8). Equation (5.9) is proved analogously. Finally, let $f(x) \equiv p$, $p > 0$, in (3.2). Then, for all $p > 0$,

$$a(1-ae_1)Ee^{-p|Q|} + ae_1\left(\frac{a(1-e^p)L_F(a(1-e^{-p}))}{1-L_F(a(1-e^{-p}))} + a\right)Ee^{-p|Q^0|} = 0,$$

where $|Q| = Q[0,1)$ and $|Q^0| = Q^0[0,1)$. Hence

$$\begin{aligned}(5.11)\qquad &-a(1-ae_1)E|Q| + ae_1(e_1^{-1} - a)E|Q^0| \\ &-\tfrac{1}{2}a^2e_1^{-1}e_2 + (a - e_1^{-1})ae_1 = 0.\end{aligned}$$

Therefore, (5.10) follows from (5.11), since $E|Q| = \frac{1}{2}c$ and $E|W| = (1-ae_1)E|Q| + ae_1E|Q^0|$. □

In the next theorem we give explicit results for the variances of Qf and Mf. Covariance measures of Q and M can be derived from (5.2) in a standard way (cf. [5]), but yield much more complicated expressions.

THEOREM 5.2. *For $f \in \mathscr{B}[0,1)$ and $g \in \mathscr{B}(0,1]$, we have*

$$\begin{aligned}\operatorname{var} Qf = {} & \frac{\alpha a}{1-ae_1}\int_0^1 dx(1-x)f^2(x) \\ & + \frac{\alpha a^3 e_2}{1-ae_1}\int_0^1 dy\left(\int_0^y dx f(x)\right)^2 + \frac{\alpha a^4 e_1 e_2}{(1-ae_1)^2}\left(\int_0^1 dx(1-x)f(x)\right)^2\end{aligned}$$

and

$$\begin{aligned}\operatorname{var} Mg = {} & \frac{\alpha a^2 e_2}{e_1(1-ae_1)^2}\left(\int_0^1 dt\, g(t)\right)^2 \\ & + \frac{2\alpha a(1-e_1^{-2}e_2)}{1-ae_1}\int_0^1 dt\, g(1-t)(e^{ae_1 t} - 1) \\ & + \frac{\alpha a}{1-ae_1}\int_0^1 dt\, g^2(t).\end{aligned}$$

PROOF. First observe that $\operatorname{var} Qf = \alpha a\int_0^\infty dt\, k_2(t)$. Define the Laplace transforms

$$h_i(s) = \int_0^\infty dt\, k_i(t)e^{-st}, \qquad i = 1,2,$$

and

$$h_1^{(2)}(s) = \int_0^\infty dt\, k_1^2(t) e^{-st} \quad \text{for } s \geq 0.$$

After taking Laplace transforms on both sides of (5.7), we obtain for all $s \geq 0$,

$$sh_2(s) - \int_0^1 dt f^2(t) = a^2 e_2 (1 - e^{-s}) h_1^{(2)}(s) + ae_1(1 - e^{-s}) h_2(s)$$

$$- \int_0^1 dt f^2(1 - t) e^{-st}.$$

Hence

$$h_2(0) = a^2 e_2 h_1^{(2)}(0) + ae_1 h_2(0) + \int_0^1 dt(1 - t) f^2(t). \tag{5.12}$$

Since $h_1^{(2)}(0)$ only depends on a and e_1, we find from the constant service time case ($e_2 = e_1^2$, cf. Corollary 6.1) that

$$h_1^{(2)}(0) = \int_0^1 dy \left(\int_0^y dx f(x) \right)^2 + \frac{ae_1}{1 - ae_1} \left(\int_0^1 dx (1 - x) f(x) \right)^2 \tag{5.13}$$

and the first part of the theorem follows from (5.12) and (5.13). The second part is proved in the same way. □

REMARK 5.1. For exponentially distributed service times, the calculation of higher order moment measures of Q and M seems to remain difficult, in contrast to the situation usually appearing in queuing theory. However, for constant service times, this problem considerably simplifies, and explicit expressions for the Laplace functionals of M and Q can be found, as is shown in the next section. Moreover, in the general case the functions k_i, $i = 1, 2 \ldots$, can in principle be obtained by the following method. Suppose, without loss of generality, that $e_1 = 1$. Assume that the k_i's can be expanded in terms of a, that is, $k_i(t) = k_{i0}(t) + ak_{i1}(t) + a^2 k_{i2}(t) + \cdots$, $i = 1, 2, \ldots$. We can derive differential equations for the k_{ij}'s by using (5.1), similarly to the derivation of (5.6) and (5.7). These differential equations can be analytically solved in successive order. But the bookkeeping involved becomes very messy and no clear structure in the procedure has been found yet.

6. The constant service time case. In this section we briefly discuss the case where the service time is of fixed length e_1. This case has been studied in [3], where the analysis was based on a modification of the analysis for the corresponding discrete model in [12]. This, however, gives rise to several problems concerning convergence. Here we give a proof based on insensitivity results for queues (see also Remark 6.1). Moreover, some extra results are given.

THEOREM 6.1. *The Laplace functionals of M and Q are given by*

$$Ee^{-Mg} = \left(\frac{1 - ae_1}{1 - ae_1 \int_0^1 dx\, e^{-g(x)}} \right)^{\alpha/e_1} \tag{6.1}$$

and

$$Ee^{-Qf} = e^{-c_f} \left(\frac{1 - ae_1}{1 - ae_1 \int_0^1 dy\, e^{-h(y)}} \right)^{\alpha/e_1}, \tag{6.2}$$

where $c_f = \alpha a \int_0^1 dx(1 - x)(1 - e^{-f(x)})$, $h(y) = ae_1 \int_0^y dx(1 - e^{-f(x)})$ *and* f *and* g *are positive measurable functions on* $[0, 1)$ *and* $(0, 1]$, *respectively.*

REMARK 6.1. In principle, Theorem 6.1 is a special case of Theorem 4.1. But even in the case of constant service times it seems to be difficult to solve the integral equations (4.6) and (4.8) directly. Thus, in the present section we give a separate proof of Theorem 6.1.

REMARK 6.2. From (6.1) it follows that the number of atoms of M is negative binomially distributed and given $|M| = m$, the m points are independent and uniformly distributed over $(0, 1]$.

PROOF OF THEOREM 6.1. Without loss of generality, we can take $e_1 = 1$.

(First part.) Consider an auxiliary $\tilde{\mathrm{M}}/\mathrm{D}/\infty$-queue with a conditionally Poisson arrival process, services times 1 and service speeds α^{-1}. Let the state-dependent arrival intensity be $a(1 + \alpha^{-1}N_t)$, where N_t is the number of customers in this auxiliary system at time t. Note that a conditionally Poisson process with such a state-dependent intensity is called a *Hawkes* or *self-exiting* process (cf. page 367 in [5]). Denote the vector of residual service times at time t (arranged in descending order) by R_t (cf. Chapter 2 of [7] for terminology). The process (M_t) in Figure 3 can be described in terms of this queuing system. Namely, N_t corresponds to $|M_t|$ and R_t corresponds to the distances from the zero level in Figure 3 (below the X-axis) to the atoms of M_t. The insensitivity results of Chapter 6 of [7] cannot be directly applied to the $\tilde{\mathrm{M}}/\mathrm{D}/\infty$ system. We therefore first consider the corresponding loss system $\tilde{\mathrm{M}}/\mathrm{D}/\mathrm{s}/0$ with the same input process, service times and service speeds, the arrival intensity at time t being now $a(1 + \alpha^{-1}N_t^{(s)})$, where $N_t^{(s)}$ denotes the number of customers in the loss system. The corresponding residual service time vector at time t is denoted by $R_t^{(s)}$, where only the first s entries of $R_t^{(s)} \in \mathbb{R}_+^\infty$ are possibly nonzero.

Similarly to Section 3, we show that the processes $(X_t^{(s)}) = (N_t^{(s)}, R_t^{(s)})$ and $(X_t) = (N_t, R_t)$ are regenerative. Note that it is possible to take the same regeneration epochs in both systems, because the input of the first one can be seen as a thinning of the second one. Hence by the key renewal theorem, we

have for all $A \in \mathscr{B}(\mathbb{N} \times \mathbb{R}_+^\infty)$,

$$\lim_{t\to\infty} P\big(X_t^{(s)} \in A\big) = \frac{1}{EC}\int_0^\infty P\big(C > t, X_t^{(s)} \in A\big)\, dt$$

and

$$\lim_{t\to\infty} P(X_t \in A) = \frac{1}{EC}\int_0^\infty P(C > t, X_t \in A)\, dt,$$

where C denotes the length of the first regeneration cycle of (X_t). Let N_C denote the number of customers that are served in the $\tilde{M}/D/\infty$-system during the first regeneration cycle. By arguments used in the proof of Theorem 3.1, it is clear that $P(N_C < \infty) = 1$. Now since

$$\{C > t, X_t^{(s)} \in A, N_C < s\} = \{C > t, X_t \in A, N_C < s\},$$

we have

$$\left|\lim_{t\to\infty} P\big(X_t^{(s)} \in A\big) - \lim_{t\to\infty} P(X_t \in A)\right| \le \frac{1}{EC}\int_0^\infty P(C > t, N_C \ge s)\, dt. \tag{6.3}$$

We are now ready to use insensitivity. Namely, when an arriving customer in the loss system is assigned to one of the idle servers (if there are any) with equal probability, then this system is insensitive to the service time distribution. Let $(N_t^{(s)}, R_t^{(s)})$ converge in distribution to $(N^{(s)}, R^{(s)})$, say. It follows from Theorem 6.7.1 of [7] that the distribution of $(N^{(s)}, R^{(s)})$ has the following form: With $p_k = P(N^{(s)} = k)$, we have

$$kp_k = \alpha(\alpha + k - 1)p_{k-1} \quad \text{for all } k = 1, 2, \ldots, s. \tag{6.4}$$

Moreover, given $\{N^{(s)} = k\}$, the k residual service times are distributed as the order statistics of k independent uniformly distributed random variables on $[0, 1]$ (cf. (6.4.1) of [7]). Note that in particular, given $N^{(s)}$, the conditional distribution of the residual service times does not depend on s. And, moreover, from (6.4), it follows that the distribution (p_k) converges to a negative binomial distribution as $s \to \infty$. And hence (6.2) follows from the observation that the difference in (6.3) converges to zero as $s \to \infty$ by the dominated convergence theorem.

(Second part.) Fix $t \ge 0$, let $T = |H_t| + \alpha$ and define the stochastic process C on $[0, 1)$ by

$$C(x) = t - \left(\int_0^x H_t(dy) + \alpha x\right), \qquad x \in [0, 1).$$

Denote the functional inverse of C on $[t - T, t]$ by A. (The graph of a realization A is given by the solid line in Figure 6.) Let Ω_t be the set of all $\omega \in \Omega$ such that $t - T(\omega) \ge 0$. Obviously, $\Omega_t \uparrow \Omega$ as $t \to \infty$. By Theorem 3.2

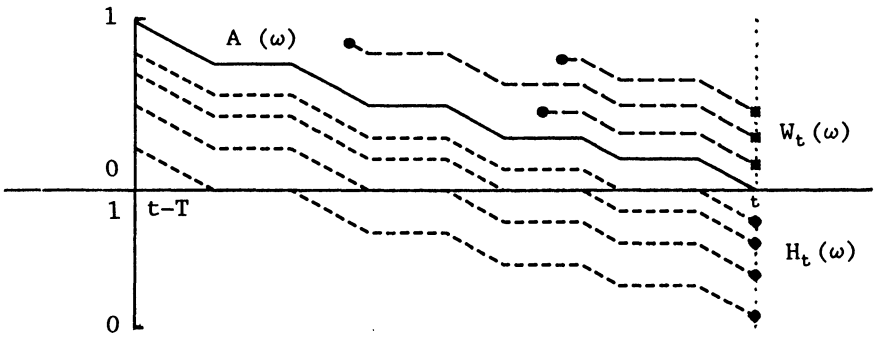

FIG. 6. *H_t determines the law of W_t. The atoms corresponding to H_t must have followed (possibly partially) the dashed lines beneath the graph of $A(\omega)$. Points occur above this graph according to a homogeneous Poisson process and lead to atoms of W_t.*

and the monotone convergence theorem we have for all $f \in \mathscr{B}[0, 1)$,

$$Ee^{-Qf} = (1 - a)^{-1} \lim_{t \to \infty} Ee^{-W_t f} I_{\{W_t\{0\}=0\}} I_{\Omega_t}. \tag{6.5}$$

Now observe that clients that are waiting at time t all have arrived during the interval $[t - T, t]$ "above the graph of A" (see Figure 6). Specifically, for every $f \in \mathscr{B}[0, 1)$, we can take as a version of the conditional distribution of $W_t f$ given H_t, the distribution of Nq, where N is a Poisson process on $[0, 1) \times \mathbb{R}_+$ with intensity a, independent of H_t, and $q(x, s) = f(x - A(s)) I_{[A(s),1]}(x) I_{[t-T,t]}(s)$ for all $(x, s) \in [0, 1) \times \mathbb{R}_+$. Therefore, the conditional expectation of $\exp(-W_t f) I_{\Omega_t}$ given H_t can be expressed in terms of A and T as

$$E_{H_t} e^{-W_t f} I_{\Omega_t} = \exp\left(-a \int_0^1 dx \int_{t-T}^t ds (1 - e^{-f(x - A(s))}) I_{[A(s),1]}(x) \right) I_{\Omega_t}.$$

Hence, by conditioning on H_t, we have for all $f \in \mathscr{B}[0, 1)$,

$$\begin{aligned}
&Ee^{-W_t f} I_{\{W_t\{0\}=0\}} I_{\Omega_t} \\
&\quad = E \exp\left(-a \int_0^1 dx \int_{t-T}^t ds (1 - e^{-f(x - A(s))}) I_{[A(s),1]}(x) \right) I_{\{W_t\{0\}=0\}} I_{\Omega_t} \\
&\quad = E \exp\left(-a \int_0^1 dx \int_0^1 \{\alpha\, du + H_t(du)\} (1 - e^{-f(x-u)}) I_{[u,1]}(x) \right) I_{\{W_t\{0\}=0\}} I_{\Omega_t} \\
&\quad = E \exp\left(-a \int_0^1 dx (1 - e^{-f(x)}) \int_0^{1-x} \{\alpha\, du + H_t(du)\} \right) I_{\{W_t\{0\}=0\}} I_{\Omega_t} \\
&\quad = e^{-c_f} E \exp\left(-a \int_0^1 dx (1 - e^{-f(x)}) H_t[0, 1 - x] \right) I_{\{W_t\{0\}=0\}} I_{\Omega_t}.
\end{aligned}$$

If we let $t \to \infty$, the last expression converges, by Theorem 3.2 and the

monotone convergence theorem, to

$$(6.6)\qquad \begin{aligned}(1-a)e^{-c_f}E\exp\left(-a\int_0^1 dx(1-e^{-f(x)})M[0,1-x]\right)\\ =(1-a)e^{-c_f}Ee^{-Mh},\end{aligned}$$

and (6.2) follows from (6.1), (6.5) and (6.6). □

COROLLARY 6.1. *For $f\in\mathscr{B}[0,1)$ and $g\in\mathscr{B}(0,1]$, we have*

$$\operatorname{var} Qf=\frac{\alpha a}{1-ae_1}\int_0^1 dx(1-x)f^2(x)+\frac{\alpha a^3e_1^2}{1-ae_1}\int_0^1 dy\left(\int_0^y dx f(x)\right)^2$$
$$+\frac{\alpha a^4e_1^3}{(1-ae_1)^2}\left(\int_0^1 dx\,(1-x)f(x)\right)^2$$

and

$$\operatorname{var} Mg=\frac{\alpha a^2e_1}{(1-ae_1)^2}\left(\int_0^1 dx\,g(x)\right)^2+\frac{\alpha a}{1-ae_1}\int_0^1 dx\,g^2(x).$$

PROOF. This follows straightforwardly from Theorem 6.1. □

APPENDIX

First, we give some basic definitions and results on random measures. References are, for example, [4] and [5].

Throughout the following, let $(E,\mathscr{E})$ be a measurable space; for definiteness we assume that E is Polish and that $\mathscr{E}$ is the Borel σ-algebra on E. A mapping M from $\Omega\times\mathscr{E}$ into $\overline{\mathbb{R}}_+$ is called a *random measure* on $(E,\mathscr{E})$ if: (a) $B\to M(\omega,B)$ is a measure on $(E,\mathscr{E})$ for every $\omega\in\Omega$; and (b) $\omega\to M(\omega,B)$ is a random variable for every $B\in\mathscr{E}$. According to Fubini's theorem,

$$Mf(\omega)=\int_E M(\omega,dx)f(x),\qquad \omega\in\Omega,$$

defines a positive random variable Mf for each positive $\mathscr{E}$-measurable function and

$$\mu(A)=EM(A)=\int P(d\omega)M(\omega,A),\qquad A\in\mathscr{E},$$

defines a measure μ on $(E,\mathscr{E})$, which is called the *mean measure* or *expectation measure* of M.

M is called a *random counting measure* if for almost every ω, there exists a countable set $D(\omega)$ such that

$$M(\omega,A)=\sum_{x\in D(\omega)}\delta_x(A),$$

where δ_x denotes the usual Dirac measure at $x \in E$. When the sets $D(\omega)$ are locally finite, M is called a *point process*.

A random measure M is said to be a *Poisson random measure* [on $(E, \mathscr{E})$] with *mean measure* (or *intensity measure*) μ if: (a) $M(A)$ has the Poisson distribution with mean $\mu(A)$ for all $A \in \mathscr{E}$; and (b) $M(A_1), \ldots, M(A_n)$ are independent whenever $A_1, \ldots, A_n \in \mathscr{E}$ are disjoint, this being true for every $n \geq 2$.

Frequently in applications, a random counting measure (point process) M satisfies $M(A) = \int M_x(A)K(dx)$, where K is a Poisson random measure, $M_x(\cdot)$ is a point process (for all $x \in E$) and $M_{\cdot}(A)$ is a measurable function (for all $A \in \mathscr{E}$). Such a point process is called a *Poisson cluster process* (on E) and the atoms of K are called *cluster centers*. A Poisson random measure on a Borel subset E of $\mathbb{R}^n$ with mean measure al_E (a times the trace of the Lebesgue measure on E) is called a *homogeneous* Poisson process with *intensity* a.

THEOREM A.1. *The probability law of random measure M is completely specified by its Laplace functional L defined by*

$$Lf = Ee^{-Mf}.$$

Moreover, the Laplace functional of a Poisson random measure on $(E, \mathscr{E})$ with mean measure μ is given by

$$Lf = \exp - \int_E \mu(dx)(1 - e^{-f(x)}) \quad \text{for all } f \in \mathscr{E}.$$

We further restrict ourselves to the case where $E = \mathbb{R}_+$. Let $\mathscr{F} = (\mathscr{F}_t)_{t \geq 0}$ be an augmented and right-continuous filtration (cf. [13]). Adaptedness, stopping times, martingales, compensators and so on are always with respect to this filtration. Let $\mathbb{D}$ denote the collection of all real-valued adapted processes on $\mathbb{R}_+$ whose every path $t \to X_t(\omega)$ is right-continuous and has left limits. Let $\mathbb{L}$ denote the collection of all adapted real-valued processes on $\mathbb{R}_+$ whose every path is left-continuous and has right-limits.

We give some basic results on stochastic integration. The definitions and proofs can be found, for example, in [11] and [13].

THEOREM A.2. *Let $X \in \mathbb{D}$ have locally finite variation. Let $f \in C^1$; then*

$$f(X_t) = f(X_0) + \int_0^t f'(X_{s-})\, dX_s + \sum_{0 < s \leq t} (f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s),$$

where $\Delta X_s = X_s - X_{s-}$ for $s > 0$ and $\Delta X_0 = 0$.

The next theorem is one of the main theorems of stochastic integration w.r.t. point processes, where now a point process N is considered as a counting process instead of a random counting measure.

THEOREM A.3. *Let $N \in \mathbb{D}$ be a point process with compensator A. Then for all $F \in \mathbb{L}$,*

$$E\int F_u \, dN_u = E\int F_u \, dA_u.$$

In many cases of practical interest A is given by

$$A_t = \int_0^t ds \, \lambda_s, \qquad t \geq 0,$$

where (λ_s) is called the stochastic intensity of N. If N is a renewal (counting) process, then A is given by Theorem 13.2 III *of* [5].

Acknowledgments. This work was begun when the first author was visiting the Department of Civil Engineering and Operations Research of Princeton University, supported by the Netherlands Organization for Scientific Research. We are most grateful to Professor Erhan Çinlar for his kind hospitality and his inspiring comments.

REFERENCES

[1] ASMUSSEN, S. (1987). *Applied Probability and Queues*. Wiley, New York.
[2] ÇINLAR, E. (1975). *Introduction to Stochastic Processes*. Prentice-Hall, Englewood Cliffs, N.J.
[3] COFFMAN, JR., E. G. and GILBERT, E. N. (1986). A continuous polling system with constant service times. *IEEE Trans. Inform. Theory* **32** 584–591.
[4] COX, D. R. and ISHAM, V. (1980). *Point Processes*. Chapman and Hall, London.
[5] DALEY, D. J. and VERE-JONES, D. (1988). *An Introduction to the Theory of Point Processes*. Springer, New York.
[6] EISENBERG, M. (1972). Queues with periodic service and changeover times. *Oper. Res.* **20** 440–451.
[7] FRANKEN, P., KÖNIG, D., ARNDT, U. and SCHMIDT, V. (1982). *Queues and Point Processes*. Wiley, Chichester
[8] FUHRMANN, S. and COOPER, R. B. (1985). Stochastic decomposition in the M/G/1 queue with generalized vacations. *Oper. Res.* **33** 1117–1129.
[9] FUHRMANN, S. and COOPER, R. B. (1985). Application of decomposition principle in M/G/1 vacation model to two continuum cyclic queueing models—especially token-ring LANs. *AT&T Tech. J.* **64** 1091–1098.
[10] HALL, P. (1988). *Introduction to the Theory of Coverage Processes*. Wiley, New York.
[11] LIPTSER, R. S. and SHIRYAYEV, A. N. (1978). *Statistics of Random Processes II: Applications*. Springer, New York.
[12] MACK, C., MURPHY, T. and WEBB, N. L. (1957). The efficiency of N machines uni-directionally patrolled by one operative when walking time and repair times are constant. *J. Roy. Statist. Soc. Ser. B* **19** 166–172.
[13] MEYER, P. A. (1966). *Probability and Potentials*. Blaisdell, San Francisco.
[14] TAKAGI, H. (1986). *Analysis of Polling Systems*. MIT Press.

DEPARTMENT OF APPLIED MATHEMATICS
UNIVERSITY OF TWENTE
P.O. BOX 217
7500 AE ENSCHEDE
THE NETHERLANDS

DEPARTMENT OF STOCHASTICS
UNIVERSITY OF ULM
HELMHOLTZSTRASSE 18
D-7900 ULM
GERMANY