Polling on a Line: General Service Times

E. G. Coffman, Jr. 1, A. L. Stolyar²

 1 AT&T Bell Laboratories, Murray Hill, NJ 07974 2 Institute of Control Sciences, GSP-312 Moscow, USSR

1 Introduction

Continuous polling systems on a line were investigated in [1, 2]. Customers arrive by a Poisson process on a closed path at independent locations having a continuous distribution. A server scans the closed path at a constant rate stopping to perform services wherever customers are encountered.

Continuous systems correspond to discrete systems in which the server visits N queues cyclically; the sequence of visits in a cycle is fixed. The server's transit time between two consecutive queues is a fixed nonzero constant. Whenever the server finds a nonempty queue, it serves all customers in this queue exhaustively (including arrivals to the queue while the server is there), and then moves on to the next queue. Arrivals at each queue are Poisson at a fixed rate. If as $N \to \infty$, the transit times and input rates are O(1/N), then we obtain the continuous system in the limit. The literature on discrete polling systems is large and growing (see [6] for an extensive bibliography).

The following more general continuous system is studied in this paper. Customers arrive on the interval $[0, \alpha]$, $\alpha > 0$, according to a Poisson process with constant rate λ . The coordinates of arriving customers are i.i.d. random variables. We assume that the distribution of customer coordinates is absolutely continuous with a density f(x), $x \in [0, \alpha]$.

The server scans the interval at a constant speed which we normalize to 1. The scan follows a fixed path $\mathcal{M} = \{(a_{11}, a_{12}), \dots, (a_{\nu 1}, a_{\nu 2})\}$, where ν is a positive integer, and $a_{i1}, a_{i2} \in [0, \alpha]$ are initial and final points of the *i*th segment, $i = 1, 2, \dots, \nu$, with $a_{11} = 0$. The server scans each segment (a_{i1}, a_{i2}) from the initial point to the final point, and then jumps instantaneously from the final point a_{i2} to the initial point $a_{i+1,1}$ of the next segment (and from $a_{\nu 2}$ to a_{11}). The interval $[0, \alpha]$ is entirely covered by the set of segments; this is the only constraint on the path \mathcal{M} .

Wherever the server encounters a customer, it stops and performs a service, continuing its motion just after service completion. Customer service times are i.i.d. samples of the random variable B,

with mean $\beta_1 = E(B)$. We assume that the following intuitive stability condition is satisfied:

$$\rho = \lambda \beta_1 < 1 \ . \tag{1}$$

In addition, we impose a technical requirement of absolute continuity on the service time distribution:

$$P\{B \le y\} = \int_0^y b(x)dx, \quad y \ge 0.$$
 (2)

Note that we are effectively modeling a system in which a server scans an arbitrary finite graph and the customers arrive randomly on the edges of the graph. During one cycle the server can scan certain branches several times in either direction. The server can also jump instantaneously from one point to another. The aim of the present work is an investigation of such a system in steady-state.

Special cases of our system were considered in [1, 2]. The steady-state behavior of the system with parameters $\mathcal{M} = \{(0, \alpha)\}$, $B = \beta_1 = const$, $f(x) = 1/\alpha = const$, was investigated in detail in [1]. The server scans in a fixed direction around the closed tour of length α . Service times are constant, and coordinates of arriving customers are uniformly distributed on the tour. However, the ergodicity of the corresponding stochastic process was not proved.

The system with parameters $\mathcal{M} = \{(0, \alpha), (\alpha, 0)\}, B = \beta_1 = const, f(x) = 1/\alpha = const,$ was considered in [2]. The server scans endlessly back and forth across the interval $[0, \alpha]$. Other parameters are the same as those in [1]. In [2] the limit case

$$\lambda \to \infty, \quad \beta_1 \to 0, \quad \lambda \beta_1 = \rho = const$$
 (3)

was analyzed. The limit system is deterministic; a deterministic continuous flow of work replaces the stochastic flow of individual customers. The problem of finding the characteristics of this "snow plow" system as a limit of our more general system will be an essential part of this paper. The snow plow system will be called the SP-system.

Another crucial concept is that of an embedded busy period (EBP). The EBP will be interpreted as a busy period structure for a certain single-server queue derived from the original system.

The main results of the paper are:

1) It is proved that condition (1) is sufficient for the ergodicity of the stochastic process describing system behavior.

- 2) The stationary distribution of this process is derived and expressed in terms of an EBP distribution.
 - 3) An explicit expression for the mean waiting time of customers in steady-state is obtained.

The mean waiting time turns out to be a sum of the steady-state mean waiting times in two limiting cases of the original system, namely, the SP-system (limit (3)), and the M/GI/1 system that results when the server speed is taken to be infinite. In particular, it is shown that, for the system with a server scanning a closed tour (the special case $\mathcal{M} = \{(0, \alpha)\}$), the mean waiting time is invariant to the form of the density f(x) of arrival coordinates and is given by

$$\bar{W} = \frac{\alpha + \lambda E(B^2)}{2(1 - \rho)} \ .$$

The paper is organized as follows. In Sections 2-4 we analyze the special case $\mathcal{M} = \{(0, \alpha)\}$ which differs from that considered in [1] only in the assumption of general service times. The proof of all results for this system are given in detail. Section 5 contains the results for the general case. Proofs of these results are analogous to those for the special case, so they are given only briefly. In Section 6 explicit formulas for mean waiting times in two interesting special cases are presented. An appendix contains the proof of the ergodicity theorem.

2 Description of the model

Consider the special case $\mathcal{M} = \{(0, \alpha)\}$, $f(x) = 1/\alpha = const$; the server scans, say clockwise, around the closed tour of length α , the distribution of arrival coordinates is uniform, and the service time distribution is general.

The analysis is simplified somewhat by considering the following equivalent open system. The server moves on the real axis $\mathbb{R} = (-\infty, \infty)$ in the positive direction at constant (unit) speed when not serving. Each arriving customer has a random label h independent of other customer labels and uniformly distributed on $[0, \alpha]$. An arrival occurs at the point $x + h \in \mathbb{R}$ ahead of the server, where x is the server's location at the time of arrival and h is the arrival's label. In other respects, the new system is the same as the original one; assumptions and notation are preserved in the open system.

A customer is distinguished by the state of the server at the time of the customer's arrival. A customer arriving while the server is moving is called a *root* customer. If a customer arrives while the server is busy serving another customer, then the arriving customer is called a descendant of the first generation, or 1-descendant, of the customer being served. Descendants of the *i*th generation

are defined in the obvious way. Each customer is either a root customer or some *i*-descendant of a root customer, i.e., it arrived while some (i-1)-descendant of a root customer was being served.

Note that the probability that the server encounters a *root* customer in any small interval $(x, x + \Delta x)$ is $\lambda \Delta x + o(\Delta x)$ and is independent of the history of the process before the server reached point x. Thus, the points at which root customers are encountered by the server form a Poisson process with constant rate λ on the axis \mathbb{R} .

Therefore, we may consider that root customers arise in the system only at the instants when they are encountered by the server. We will use this convention hereafter.

We say that a root customer together with all its descendants forms an *embedded busy period* (EBP). By condition (1), the number of customers in an EBP is finite with probability 1.

For any customer C in an EBP, the conditional distribution of the number k_C of C's 1-descendants, given C's service time τ_C , is the Poisson distribution

$$P\{k_C = i\} = \frac{(\lambda \tau_C)^i}{i!} e^{-\lambda \tau_C}, \quad i \ge 0.$$

With k_C given, the label samples h_1, \ldots, h_{k_C} of these 1-descendants are taken from the uniform density $f(x) = 1/\alpha$, $x \in [0, \alpha]$. These 1-descendants are placed at locations denoted by $\zeta(x, h_m) = x + h_m$, $1 \le m \le k_C$.

Let k denote the (random) total number of descendants of some root customer. Let these descendants be numbered in order of increasing distance y_i from the root customer, i.e., if x is the location of the root customer, then $0 \le y_1 \le \cdots \le y_k \le \infty$, where $y_i \equiv x_i - x$, and x_i is the location of the ith descendant. Let τ_0 and τ_i , $1 \le i \le k$, denote the root-customer and the ith descendant service times, respectively. Then an EBP can be represented as a random vector $\xi = [\tau_0, \ldots, \tau_k; y_1, \ldots, y_k]$, with a dimension 2k - 1 that is also random.

It is obvious that all embedded busy periods are independent identically distributed random elements in a measurable space $(G, \mathcal{B}(G))$, where $G \subseteq \bigcup_{k=1}^{\infty} \mathbb{R}^{2k-1}_+$, \mathbb{R}_+ is the set of real nonnegative numbers, and $\mathcal{B}(G)$ is a σ -algebra of Borel subsets of G.

Thus, we can adopt the following convention.

(A) The entire sample of an EBP is taken just at the instant when its root customer is encountered; i.e., at such an instant, locations and service times of all customers of the EBP are fixed instantaneously.

Except when noted otherwise, we will use convention (A) hereafter. Let Q denote the distribution of a generic EBP ξ , and let Λ denote Lebesgue measure on the space G. The assumed absolute continuity of the service time distribution implies the existence of a density $q(\xi) = dQ(\xi)/d\Lambda$, $\xi \in G$. Let us also define $\tau(\xi) \equiv \sum_{i=0}^k \tau_k$ and $l(\xi) \equiv y_k$.

3 Interpretation of an embedded busy period

Consider the limit of the original system (and its open equivalent on \mathbb{R}) as the server speed becomes infinite (the server spends zero time in advancing to the next waiting customer). This limiting system is an M/GI/1 queue with random priorities determining the order of service; we call it the RP-system.

The process, customer service times and labels in the RP-system are the same as those in the original system. An arriving customer that finds the system empty originates a new busy period and is called a root customer. A root customer receives the priority y = 0 and begins service immediately. A customer that arrives while the server is busy enters a queue and receives a priority $y = y_C + h$, where h is the label of the customer, and y_C is the priority of the customer being served at the time of arrival. Service is nonpreemptive. After each service completion, the waiting customer having minimum priority leaves the queue and begins service immediately. A busy period is finished when the queue is found to be empty just after a service completion.

A busy period of the RP-system can be represented by the structure

$$\xi = [\tau_0, \ldots, \tau_k; \ y_1, \ldots, y_k] ,$$

where k is the number of customers, excluding the root customer, of the busy period, and where y_i and τ_i , $0 \le i \le k$, are customer priorities and service times, respectively. Customers are indexed beginning with the root customer, i.e., $0 \equiv y_0 \le y_1 \le \cdots \le y_k$.

It is clear that a busy period ξ of the RP-system and an EBP of the original system have identical studients \mathcal{B} and \mathcal{B} is the customer than a busy period \mathcal{B} in \mathcal{B} is \mathcal{B} diate. The S S S of the RP-system and \mathcal{B} is \mathcal{B} diate.

$$\bar{l} \equiv E(l(\xi)) < \infty . \tag{5}$$

$$\bar{W}^{RP} = \bar{W}^{MG1} = \frac{\lambda \beta_2}{2(1-\rho)} ,$$
 (6)

where $\beta_2 \equiv E(B^2)$, \bar{W}^{RP} is the steady-state mean waiting time in the RP-system, and \bar{W}^{MG1} is the Pollachek-Khinchine formula for the mean waiting time in an M/GI/1 system with a conservative and nonpreemptive service discipline.

We will adopt a convention similar to (A) for the RP-system. Namely, the entire sample of an RP-system's busy period is taken just at the instant of its root customer's arrival. A state of the RP-system is defined as the pair (ξ, τ) , where ξ is the structure of the current busy period, and τ is the time that has elapsed since its beginning, $0 \le \tau \le \tau(\xi)$. If the system is empty then the corresponding state is defined to be $(\xi, \tau) = *$. Thus, the state (ξ, τ) of the RP-system at any given time is a random element in the measurable space $(\tilde{G}, \mathcal{B}(\tilde{G}))$, where $\tilde{G} \subseteq (G \times \mathbb{R}_+) \cup \{*\}$. A stationary state distribution \tilde{Q} exists by (4) and renewal theory arguments. Let Λ denote Lebesgue measure on \tilde{G} with $\Lambda(\{*\}) = 1$ by definition. Then the density corresponding to \tilde{Q} exists and has the form

$$\tilde{q}(\xi,\tau) \equiv \frac{d\tilde{Q}}{d\Lambda}(\xi,\tau) = \begin{cases} 1 - \rho, & \text{if } (\xi,\tau) = *, \\ \rho \frac{q(\xi)\tau(\xi)}{\tau} \frac{1}{\tau(\xi)}, & \text{if } (\xi,\tau) \neq *. \end{cases}$$
 (7)

4 Results

To define a state it is convenient to normalize the server's location to 0. Then coordinates of customers already served are negative and those of customers not yet served are positive. The customer being served always has coordinate 0.

The pair (ξ_0, τ) denotes a state of the server, where ξ_0 is the structure of the active EBP (i.e., the EBP containing the customer currently being served); τ is the total time spent so far on serving the active EBP; and $(\xi_0, \tau) = *$ by definition if the server is moving and hence not serving. EBPs other than the active one, which have been started but not finished by the server, are called passive EBPs.

The structure of the model along with simple intuitive reasoning anticipates the following proposition.

Proposition 1 In statistical equilibrium, the state (ξ_0, τ) of the server has the marginal distribution \tilde{Q} , which is the stationary distribution of the RP-system's state. Locations and structures of all

EBPs, excluding the active one, on the entire axis (i.e. relating to both the past and the future) are independent of the server's state and are described as follows. Root customer locations of all EBPs form a Poisson process with constant rate λ , on the entire axis, and their structures are i.i.d. random elements with the distribution Q.

Proposition 1 can be formulated and proved rigorously for a suitably defined Markov process with a sufficiently general phase space. However, for simplicity we shall consider the Markov process with the less general phase space given below. And it will be seen that all steady-state characteristics of this process (and the original system) can be derived from Proposition 1.

Formally, a state of the system at time $t \geq 0$ is defined as

$$s(t) = \{(\xi_0, \tau); \ \bar{x} \equiv (x_j : 1 \le j \le n), \ \bar{\xi} \equiv (\xi_j : 1 \le j \le n)\}\ ,$$

where (ξ_0, τ) is the state of the server at time t; $n \geq 0$ is the number of passive EBPs; $0 \leq x_1 \leq x_2 \leq \cdots \leq x_n < \infty$ are the ordered distances from the server to the customers of the passive EBPs that are to be served next; and ξ_j , $1 \leq j \leq n$, is the structure of the *j*th passive EBP. By this definition we have $l(\xi_j) > x_j$, $1 \leq j \leq n$.

It is obvious that s(t), $t \geq 0$, is a homogeneous Markov process with phase space $(S, \mathcal{B}(S))$, where

$$S \subseteq \tilde{G} \times \left[\bigcup_{m=0}^{\infty} \mathbb{R}_{+}^{m} \times G^{m} \right]$$
.

Again, Λ denotes Lebesgue measure on the space S.

Theorem 1 If condition (1) is satisfied then the process s(t), $t \ge 0$, is ergodic.

(A proof of Theorem 1 is in the appendix.)

Theorem 2 If condition (1) and absolute continuity in (2) are satisfied, then the ergodic distribution of the process s(t), $t \ge 0$, has the density

$$\rho(s) = \tilde{q}(\xi_0, \tau) \exp(-\lambda \bar{l}) \lambda^n \prod_{j=1}^n q(\xi_j), \quad s \in S .$$
 (8)

Remark. It is easy to see that the ergodic distribution of the process s(t) given in Theorem 2 is a projection of the distribution (in the more general phase space) given in Proposition 1. Thus,

all steady-state characteristics of the system depending on the ergodic distribution of s(t) can be derived from Proposition 1.

Proof of Theorem 2. It will be convenient to rewrite (8) using (7) for $\tilde{q}(\cdot)$:

$$\rho(s) = \begin{cases} \rho_1(s) & \text{if } s \in S_1, \\ \rho_2(s) & \text{if } s \in S_2, \end{cases}$$

$$(9)$$

where $s = \{(\xi_0, \tau); \ \bar{x} \equiv (x_j : 1 \le j \le n), \bar{\xi} \equiv (\xi_j : 1 \le j \le n)\} \in S$

$$S_1 = \{ s \in S : (\xi_0, \tau) = * \}, S_2 = \{ s \in S : (\xi_0, \tau) \neq * \},$$

$$\rho_1(s) = (1 - \rho) \exp(-\lambda \bar{l}) \lambda^n \prod_{j=1}^n q(\xi_j),$$

$$\rho_2(s) = \rho \frac{q(\xi_0)}{\bar{\tau}} \exp(-\lambda \bar{l}) \lambda^n \prod_{j=1}^n q(\xi_j) .$$

Next, we derive equations giving necessary and sufficient conditions for the stationarity of the density $\rho(s)$. We use implicitly the properties that s(t) is continuous from the right and that $\rho(s)$ in (9) is continuous and differentiable.

Consider the evolution of the system state in a small time interval $[t - \Delta t, t]$. We must consider the following classes of states s at the end t of the interval.

1)
$$s \in S_1$$
 and
$$y_{ij} \neq x_j \text{ for all } 0 \le i \le k_j - 1, \quad 1 \le j \le n , \tag{10}$$

where k_j is the number of customers in EBP ξ_j , and y_{ij} is defined to be y_i of EBP ξ_j , with $y_{0j} = 0$. Then

$$\rho_1(s) = \rho_1(\{*; (x_1 - \Delta t, \dots, x_n - \Delta t), \bar{\xi}\})(1 - \lambda \Delta t) + \int_G \rho_2(\{(\xi_0, \tau(\xi_0)); \bar{x}, \bar{\xi}\}) \Delta t d\xi_0 + o(\Delta t) .$$

The left-hand side of the equation refers to time t. The first term in the right-hand side corresponds to the event "the server did not serve in the interval $[t - \Delta t, t]$," while the second term corresponds to the event "the service of some passive EBP finished in the interval $[t - \Delta t, t]$." In the limit $\Delta t \to 0$

we obtain from the above equation

$$\sum_{i=1}^{n} \frac{\partial}{\partial x_{j}} \rho_{1}(s) = -\lambda \rho_{1}(s) + \int_{G} \rho_{2}(\{(\xi_{0}, \tau(\xi_{0})); \ \bar{x}, \bar{\xi}\}) d\xi_{0} \ . \tag{11}$$

1') $s \in S - 1$ and condition (10) does not hold, i.e., $y_{ij} = x_j$ for some $j \in \{1, ..., n\}$ and $i \in \{0, ..., k_j - 1\}$. In such a state the server has just finished the service of the *i*th customer of the *j*th passive EBP. We obtain the boundary condition

$$\rho_1(s) = \rho_2(\{(\xi_j, \tau'_{ij}); (x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n),$$

$$(\xi_1, \dots, \xi_{j-1}, \xi_{j+1}, \dots, x_n)\}),$$
(12)

with $\tau'_{ij} = \sum_{m=0}^{i} \tau_m$ for EBP ξ_j .

2) $s \in S_2$ and the following holds

$$\begin{cases}
\tau \neq \tau'_{0i}, \ 0 \le i \le k_0 - 1, \\
\tau \neq 0.
\end{cases}$$
(13)

In this case

$$\rho_2(s) = \rho_2(\{(\xi_0, \tau - \Delta t); \ \bar{x}, \bar{\xi}\})$$

which implies

$$\frac{\partial}{\partial \tau} \rho_2(s) = 0 \ . \tag{14}$$

2') $s \in S_2$ and $\tau = \tau'_{i0}$ for some $i \in \{0, ..., k_0 - 1\}$. In such a state the server just begins the service of the (i + 1)st customer of the active EBP ξ_0 . Then the following boundary condition is derived:

$$\rho_2(s) = \rho_1(\{*; (x_1, \dots, x_j, y_{i+1,0}, x_{j+1}, \dots, x_n), (\xi_1, \dots, \xi_j, \xi_0, \xi_{j+1}, \dots, \xi_n)\}),$$
(15)

where $x_j \leq y_{i+1,0} \leq x_{j+1}$.

2'') $s \in S-2$ and $\tau=0$. The root customer of the active EBP has just been encountered. In this case we get

$$\rho_2(s) = \rho_1(\{*; (x_1 - \Delta t, \dots, x_n - \Delta t), \bar{\xi}\})\lambda \exp(-\lambda \Delta t)q(\xi_0) + o(\Delta t),$$

Polling on a Line: General Service Times

10

implying the boundary condition

$$\rho_2(s) = \rho_1(\{*; \bar{x}, \bar{\xi}\}) \lambda q(\xi_0) . \tag{16}$$

Immediate substitution of $\rho(s)$ as given by (9), into (11), (12), (14)-(16) shows that (14) is satisfied trivially and that the other equations are reduced to the identity $(1 - \rho)\lambda = \rho/\bar{\tau}$, which is equivalent to (4).

Thus, (9) gives the density of a stationary distribution of s(t). Since this stationary distribution is unique and ergodic by Theorem 1, the proof is complete.

Theorem 2 yields insight into the structure of customer waiting times in steady-state. Moreover, the theorem makes it possible to derive an explicit expression for the mean waiting time without knowing an explicit form for the EBP distribution Q.

The following notation will be used:

 W_h , $h \in [0, \alpha]$, is the (random) steady-state waiting time for a customer having label h;

 $W_{\scriptscriptstyle h}^{RP}$ is the corresponding waiting time in the RP-system;

 V_y , $y \ge 0$, is the total service time of all customers located in interval [0, y], under the assumptions that the EBP root customers form a Poisson process with rate λ on the entire axis $\mathbb R$ and that the EBP structures are i.i.d. random elements with distributions Q. The model structure shows that V_y can be interpreted as the time spent serving in any fixed interval of length y in steady-state.

Lemma 1 The following equality in distribution holds:

$$W_h \stackrel{\mathrm{d}}{=} h + V_h + W_h^{RP} , \quad 0 \le h \le \alpha ,$$

where V_h and W_h^{RP} are independent.

Proof. Since the input process is Poisson, it is sufficient to consider the waiting time of a customer with label h arriving in steady-state. This customer is located at point h. (Recall that the server location is 0 by definition.) Then $W_h = h + W_1 + W_2$, where h is the time taken by the server to move from point 0 to point h; W_2 is the total service time of the active EBP's customers located in [0, h), with the remaining service time being taken for the customer being served, and with $W_2 = 0$

if the server is moving; and W_1 is the total service time of all other customers served in [0, h). The independence of W_1 and W_2 , and the relations $W_1 \stackrel{\text{d}}{=} V_h$ and $W_2 \stackrel{\text{d}}{=} W_h^{RP}$ all follow immediately from Proposition 1 (see the remark after Theorem 2).

Lemma 2 The expected value of V_y is

$$\bar{V}_y = \frac{\rho}{1-\rho} y, \quad y \ge 0 \ . \tag{17}$$

This fact was pointed out in [1] and is almost obvious, since the server in steady-state spends a fraction ρ of its time serving and a fraction $1 - \rho$ moving. The proof is obvious, so we omit it.

Theorem 3 The expected customer waiting time in steady-state is

$$\bar{W} = \frac{\alpha}{2(1-\rho)} + \frac{\lambda \beta_2}{2(1-\rho)} \ . \tag{18}$$

Proof. Since a customer's label h is uniformly distributed on $[0, \alpha]$, Lemma 1 implies

$$\bar{W} = \int_0^\alpha \frac{dh}{\alpha} \left(h + \bar{V}_h + E(W_h^{RP}) \right) .$$

Application of Lemma 2 and the fact that (see (6))

$$\int_0^\alpha \frac{dh}{\alpha} E(W_h^{RP}) = \bar{W}^{RP} = \bar{W}^{MG1} = \frac{\lambda \beta_2}{2(1-\rho)} \ , \label{eq:energy_energy}$$

then completes the proof.

Theorem 3 can be formulated as

$$\bar{W} = \bar{W}^{SP} + \bar{W}^{MG1} \,, \tag{19}$$

where \bar{W}^{SP} is the mean waiting time in the (snow plow) SP-system, obtained from the original system in the limit (3): $\lambda \to \infty$, $\beta_1 \to 0$, $\lambda \beta_1 = \rho = const$. (A formal definition of the SP-system for the general model will be given in the next section.) Indeed, in the SP-system a continuous flow of

work arrives with rate ρ , and is distributed uniformly on the tour. By symmetry, the density of work at any point of the tour just before it is reached by the server in steady-state is some constant ϕ . Because the server spends a fraction ρ of its time serving (i.e. performing the work), and a fraction $1 - \rho$ moving, we have $\phi = \rho/(1 - \rho)$. Thus taking into account the time spent for service, a time $(1 + \phi)y = y/(1 - \rho)$ is required for the server to traverse any interval of length y. Because the mean distance from the server to the customer at its time of arrival is $\alpha/2$, we obtain $\bar{W}^{SP} = \alpha/(2(1 - \rho))$.

It will be shown in the next section that Theorem 3 in the form (19) is valid also for the general system, with a more general expression for \bar{W}^{SP} .

5 Results for the general case

We first introduce some notation. Let T be the length of the path

$$T = \sum_{i=1}^{\nu} |a_{i2} - a_{i1}| \ .$$

We say that the server is sited at point $x \in [0, T)$ of the path \mathcal{M} , if it has moved a distance x from the beginning of the current cycle. The function $g(x) \in [0, \alpha)$, $x \in [0, T)$, gives the actual point of the interval $[0, \alpha]$, where the server sited at point x of \mathcal{M} is located. This function is piecewise linear and has the form

$$g(x) = \frac{(x - \hat{a}_{i-1})a_{i2} + (\hat{a}_i - x)a_{i1}}{|a_{i2} - a_{i1}|}, \quad x \in [\hat{a}_{i-1}, \hat{a}_i), \quad i = 1 + \nu,$$

where $\hat{a}_0 = 0$, $\hat{a}_i = \sum_{j=1}^i |a_{j\,2} - a_{j\,1}|$, $1 \leq i \leq \nu$, are the initial points of the path segments in \mathcal{M} . Continue the function g(x) periodically with period T to the entire axis \mathbb{R} . As in the earlier special case, we consider an open system which is equivalent to the original one. In the open system the server moves along the real axis $\mathbb{R} \equiv (-\infty, \infty)$ in the positive direction at unit speed when not serving. However, points on the axis are not homogeneous; a point $x \pmod{T}$ of \mathcal{M} and a point g(x) of $[0,\alpha]$ correspond to each point $x \in \mathbb{R}$. Each arriving customer has a random label h in $[0,\alpha]$ with density f(h). If the server has location x at the instant a customer having label h arrives, then the customer is placed at the point $\zeta(x,h)$, where

$$\zeta(x,h) \equiv \inf \{x' > x : q(x') = h\}, x \in \mathbb{R}, h \in [0,\alpha],$$

is the point x' nearest to x (from the right) corresponding to point h of the physical interval $[0, \alpha]$. The function $\zeta(x, h)$ is periodic in x with period T. The definitions of root customer and embedded busy period are the same as before. Then the points on the axis at which customers are encountered by the server form a Poisson process with variable rate

$$\lambda(x) = (x - \eta(x))f(g(x))\lambda, \quad x \in \mathbb{R}$$

where

$$\eta(x) \equiv \sup\{x' < x : g(x') = g(x)\}$$

is the point x' nearest to x (from the left) corresponding to the same point in $[0, \alpha]$ as x.

The functions $x - \eta(x)$ and $\lambda(x)$ are periodic with period T. Furthermore,

$$\int_0^T \lambda(x)dx = \lambda T .$$

The distribution of an EBP now depends on the location x of its root customer. Given this location x, an EBP and its distribution will be denoted by EBP^(x) and $Q^{(x)}$ respectively. Clearly, $Q^{(x)}$ is also periodic with period T. The density $q^{(x)}(\xi)$, $\xi \in G$, of distribution $Q^{(x)}$ exists because the density $f(\cdot)$ exists and condition (2) holds.

It is obvious that all EBPs are independent, given the locations of their root customers. Thus, convention (A) can again be adopted:

(A) The entire sample of an EBP is taken just at the instant when its root customer is encountered.

Given x, it is easy to see that $\mathrm{EBP}^{(x)}$ can be interpreted as a busy period structure of the following M/GI/1 system, to be called the $\mathrm{RP}^{(x)}$ -system. The $\mathrm{RP}^{(x)}$ -system is the same as an RP-system, except as follows. Customer labels have a density f(h) on $[0,\alpha]$. Any customer arriving to find an empty system (i.e. a root customer) receives a fixed priority x independent of its label. A customer having label h and finding a nonempty system on arrival receives priority $\zeta(y,h)$ where y is the priority of the customer being served at the instant of arrival. In the busy period structure $\xi = [\tau_0, \ldots, \tau_k; y_1, \ldots, y_k]$ of the $\mathrm{RP}^{(x)}$ -system, y_i is the priority of the ith customer reduced by x.

For any fixed x, all other definitions and properties given in Section 3 for the RP-system are also valid for the RP^(x)-system, with the notation RP-system, EBP, $Q(\cdot)$, and $q(\cdot)$ replaced by RP^(x)-system, EBP^(x), $Q^{(x)}(\cdot)$, and $q^{(x)}(\cdot)$ respectively.

In defining a state of the general system we will always assign coordinate 0 to the last point in \mathbb{R} , already passed by the server, which corresponds to the initial point of \mathcal{M} . The state of the server

is defined as a vector $[(\xi_0, \tau, z_0), z]$, where ξ_0 and τ have the same meaning as before; $z_0 \in [0, T)$ is the point on \mathcal{M} corresponding to the location of the root customer of the active EBP; $(\xi_0, \tau, z_0) = *$ by definition, if the server is moving; and $z \in [0, T)$ is the current server location. Note that if the server is serving, then z is a deterministic function of (ξ_0, τ, z_0) , namely $z = (z_0 + y(\xi_0, \tau)) \pmod{T}$, where $y(\xi, \tau)$ is the priority of the customer of EBP ξ in service, and τ is the elapsed service time of ξ .

For the general model we are led to the following generalization of Proposition 1.

Proposition 2 In statistical equilibrium the server state $[(\xi_0, \tau, z_0), z]$ has the density

$$\tilde{q}[(\xi_0, \tau, z_0), z] = \begin{cases} (1 - \rho)/T, & \text{if } (\xi_0, \tau, z_0) = *, \\ \rho \frac{\lambda(z_0)}{\lambda T} \frac{q^{(z_0)}(\xi_0)}{\tau}, & \text{if } (\xi_0, \tau, z_0) \neq *. \end{cases}$$

The server is traveling with probability $(1 - \rho)$. Conditioned on that event, the server's location z is distributed uniformly on [0,T). Given that the server is serving, an event of probability ρ , z_0 is distributed on [0,T) with density $\lambda(x)/\lambda T$. Given z_0 , (ξ_0,τ) has a conditional distribution equal to the conditional steady-state distribution of the $RP^{(z_0)}$ -system, given that the server is busy. The server location z is a deterministic function of (ξ_0,τ,z_0) .

The locations and structures of all passive EBPs on the entire axis \mathbb{R} are independent of the server state and described as follows. Root customer locations form a Poisson process with periodic rate $\lambda(x)$. Given a set $\{z_j\}$ of root-customer locations, their EBP structures are independent random elements with corresponding distributions $Q^{(z_j)}$.

We will not formulate and prove rigorously Proposition 2, although that can be done. The following observations and results for a process with a less general state space will justify the application of Proposition 2 to the derivation of all steady-state characteristics of the original system, in analogy with Section 3 (see the remarks after Proposition 1 and Theorem 2).

Let a formal state of the system at time $t \geq 0$ be

$$s(t) = \{(\xi_0, \tau, z_0), z; \ \bar{x} \equiv (x_i : 1 \le j \le n), \ \bar{\xi} \equiv (\xi_i : 1 \le j \le n)\}$$
.

where $[(\xi_0, \tau, z_0), z]$ is the server state at time t, and \bar{x} , $\bar{\xi}$ have the same meanings as in Section 3. The following ergodic theorem generalizes Theorem 1 to the general model. **Theorem 4** If condition (1) is satisfied then the process s(t), $t \ge 0$, is ergodic.

With a few obvious modifications, the proof is the same as that of Theorem 1 (see the appendix).

Theorem 5 If condition (1) and absolute continuity in (2) are satisfied then the ergodic distribution of the process s(t), $t \ge 0$, has the density

$$\rho(s) = \tilde{q}[(\xi_0, \tau, z_0), z] \exp(-c(z)) \prod_{j=1}^n \lambda(z - x_j) q^{(z - x_j)}(\xi_j), \ s \in S,$$

where S is the phase space of the process s(t), and

$$c(z) = \int_{-\infty}^{z} \lambda(x) Q^{(x)} \{l(\xi) > z - x\} dx.$$

Thus, the ergodic distribution of s(t) is a projection of the distribution described in Proposition 2. The proof of Theorem 5 is analogous to that of Theorem 2 and is omitted.

Let us formally define the deterministic SP-system corresponding to the original stochastic system. (Such a system can be considered as the limit (3): $\lambda \to \infty$, $\beta_1 \to 0$, $\lambda \beta_1 = \rho = const$. But we will not need any result about convergence.) A deterministic continuous flow of work arrives on the physical interval $[0, \alpha]$ with rate ρ . The work $\rho \Delta t$ that arrives during time Δt , is distributed with density $\rho \Delta t f(y)$, $y \in [0, \alpha]$. The server scans the interval $[0, \alpha]$ according to the path \mathcal{M} . In the absence of work the server would move at unit speed. But in each small piece $[x, x + \Delta x]$ of the path \mathcal{M} the server must perform work $\phi(x)\Delta x$, where $\phi(x)$, $x \in [0, T)$, is the density of work at point x of the path \mathcal{M} just before the server reaches x. Thus, the actual time spent by the server in passing through $[x, x + \Delta x]$ is $\Delta x + \phi(x)\Delta x$, i.e., the actual speed at point x is $(1 + \phi(x))^{-1}$.

The existence and uniqueness of the density function $\phi(x)$ in the steady-state regime of the SP-system is given by the following lemma. The proof is straightforward and left to the reader.

Lemma 3 There exists a unique nonnegative periodic function $\phi(x)$ that has period T and satisfies

$$\phi(x) = \rho f(g(x)) \int_{\eta(x)}^{x} (1 + \phi(u)) du, \quad x \in \mathbb{R} .$$
(20)

The periodicity of $\phi(x)$ and the balance of arriving work and work performed implies the condition

$$\int_{0}^{T} \phi(x)dx = \rho T/(1-\rho) \ . \tag{21}$$

The mean waiting time in the SP-system is defined naturally by the formula

$$\bar{W}^{SP} = \int_{0}^{T} \theta(x) dx \int_{0}^{\alpha} f(y) dy \int_{x}^{\zeta(x,y)} (1 + \phi(u)) du , \qquad (22)$$

where

$$\theta(x) = \frac{1 + \phi(x)}{\int_0^t (1 + \phi(u)) du} = \frac{(1 - \rho)(1 + \phi(x))}{T} , \qquad (23)$$

with $\phi(x)$ defined by Lemma 3.

Theorem 6 The mean customer waiting time in the original system in steady-state is

$$\bar{W} = \bar{W}^{SP} + \bar{W}^{MG1}$$

where \bar{W}^{SP} is the mean waiting time in the corresponding SP-system, and $\bar{W}^{MG1} = \lambda \beta_2/(2(1-\rho))$ is the mean waiting time in the usual M/GI/1 system.

Proof. Consider a state in the stationary regime. We may assume that the state is described by Proposition 2.

Consider a set of all EBPs, excluding the active one, on the entire axis \mathbb{R} . Let $\Phi(x, \mu; \gamma)$, $x \leq \mu$, be the mean total service time of all customers of EBP^(\gamma) located in $[x, \mu]$; and let $\Phi(x, \mu)$ be the mean total service time of all customers (of all EBPs under consideration) located in $[x, \mu]$. It is obvious that

$$\Phi(x,\mu) = \int_{-\infty}^{\mu} \lambda(\gamma) d\gamma \Phi(x,\mu;\gamma) \ .$$

It is easy to see that the density $\frac{\partial}{\partial \mu} \Phi(x, \mu) \Big|_{\mu=x+0}$ is a periodic nonnegative function that must satisfy (20). Thus

$$\frac{\partial}{\partial \mu} \Phi(x,\mu) \Big|_{\mu=x+0} = \phi(x), \quad x \in \mathbb{R}.$$

We now derive an expression for $\hat{\theta}(x)$, $x \in [0, T)$, the density of the conditional distribution of the server location z in steady-state, given that the server is serving. We can write

$$\hat{\theta}(x)\Delta x = \int_{-\infty}^{x+\Delta x} \frac{\lambda(z_0)}{\lambda T} dz_0 \int_{\xi_0 \in G} \frac{1}{\bar{\tau}} q^{(z_0)}(\xi_0) d\xi_0 \\ \times \left[\sum_{i=0}^{k_0} \tau_{i0} I\{z_0 + y_{i0} \in [x, x + \Delta x]\} \right] + o(\Delta x) ,$$

where k_0 , τ_{i0} , y_{i0} are the parameters k, τ_i , y_i of EBP ξ_0 , and $I\{\cdot\}$ is the set indicator function.

Thus,

$$\hat{\theta}(x)\Delta x = \int_{-\infty}^{x+\Delta x} \frac{\lambda(z_0)}{\lambda T} \frac{dz_0}{\bar{\tau}} \Phi(x, x + \Delta x; z_0) + o(\Delta x)$$
$$= \frac{1}{\lambda T \bar{\tau}} \phi(x) \Delta x + o(\Delta x) .$$

Finally, using the identity $\lambda \bar{\tau} = \rho/(1-\rho)$, we have

$$\hat{\theta}(x) = \frac{\phi(x)}{\lambda \bar{\tau} T} = \frac{\phi(x)}{\rho T / (1 - \rho)}, \quad x \in [0, T) ,$$

implying in particular equation (22).

Then the density $\theta(x)$, $x \in [0, T)$, of the unconditional distribution of the server location z in steady-state is

$$\theta(x) = \rho \hat{\theta}(x) + (1 - \rho) \frac{1}{T} = \frac{(1 - \rho)(1 + \phi(x))}{T}$$
,

which coincides with (23). Consider a customer entering the system in steady-state. Since the input process is Poisson, in order to derive \bar{W} , it is sufficient to calculate such a customer's mean waiting time. Its random label y has density f(y) in $[0, \alpha]$ and its waiting time is

$$W = \zeta(z, y) - z + W_1 + W_2 ,$$

where $\zeta(z, y) - z$ is the server's moving time to the location $\zeta(z, y)$ of the customer; W_2 is the total service time of the active EBP's customers located in $[z, \zeta(z, y))$, with the remaining service time taken for the customer being served and with $W_2 = 0$ if the server is traveling; and W_1 is the total service time of all other customers to be served in $[z, \zeta(z, y))$.

Using our preliminary results we obtain

$$E[\zeta(z,y) - z + W_1] = \bar{W}^{SP}.$$

The form of the server-state distribution shows that we can write

$$EW_2 = \rho \int_0^T \frac{\lambda(z_0)}{\lambda T} \bar{W}_b^{RP^{(z_0)}} dz_0 ,$$

where $\bar{W}_b^{RP^{(z_0)}}$ is the conditional steady-state mean waiting time of a customer in the $\mathrm{RP}^{(z_0)}$ system in steady-state, given that the system is nonempty at its time of arrival. But $\bar{W}_b^{RP^{(z_0)}} = \bar{W}^{RP^{(z_0)}}/\rho = \bar{W}^{MG^1}/\rho$ for any z_0 . Then $EW_2 = \bar{W}^{MG^1}$ and the proof is complete.

6 Examples

In this section we obtain two interesting consequences of Theorem 6.

Corollary 1 The mean waiting time in the system with a server scanning continuously around a closed tour is invariant to the form of the coordinate density f and is given by (18).

Proof. We have the special case $\mathcal{M} = \{(0, \alpha)\}$, with the density f(y) arbitrary. Then $T = \alpha$, g(x) = x, $\eta(x) = x - \alpha$, and

$$\zeta(x,y) = \begin{cases} y, & \text{if } y \ge x, \\ y + \alpha, & \text{if } y < x, \end{cases} \quad x, y \in [0, \alpha) .$$

It is convenient to continue the function f(x) periodically with period α to the entire axis \mathbb{R} .

It follows from (20) and (21) that $\phi(x) = cf(x), c = \rho\alpha/(1-\rho)$. Thus,

$$\theta(x) = \frac{(1-\rho)(1+cf(x))}{\alpha} ,$$

and

$$\bar{W}^{SP} = \int_0^\alpha \frac{(1-\rho)(1+cf(x))}{\alpha} dx \int_0^\alpha f(y) dy \int_x^{\zeta(x,y)} (1+cf(u)) du \ .$$

Note that

$$\int_0^\alpha \frac{dx}{\alpha} (\zeta(x, y) - x) = \int_0^y \frac{dx}{\alpha} (y - x) + \int_y^\alpha \frac{dx}{\alpha} (\alpha + y - x)$$

$$= \int_{y - \alpha}^y \frac{dx}{\alpha} (y - x) = \frac{\alpha}{2} \text{ for arbitrary } y \in [0, \alpha) ;$$

$$\begin{split} \int_0^\alpha f(y)dy \int_x^{\zeta(x,y)} f(u)du &= \int_0^x f(y)dy \int_x^{y+\alpha} f(u)du + \int_x^\alpha f(y)dy \int_x^y f(u)du \\ &= \int_x^{\alpha+x} f(y)dy \int_x^y f(u)du = \frac{1}{2} \text{ for arbitrary } x \in [0,\alpha) \ ; \end{split}$$

and

$$\int_0^\alpha \int_0^\alpha f(x)f(y)dxdy(\zeta(x,y)-x) = \frac{\alpha}{2} ,$$

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

$$\int_0^\alpha \int_0^\alpha f(x)f(y)dxdy[(\zeta(x,y)-x)+(\zeta(y,x)-y)] = \alpha .$$

From these properties we obtain

$$\bar{W}^{SP} = \frac{\alpha}{2(1-\rho)} \ .$$

Application of Theorem 6 completes the proof.

As a second consequence, consider a star system with n branches of length α/n . The server scans the branches cyclically in a fixed sequence, moving out to the end of each branch then returning to the center of the star. Arrival locations are distributed uniformly over the star.

Corollary 2 The steady-state mean waiting time in the star system is

$$\bar{W} = \frac{\alpha}{(1-\rho)} \left(1 - \frac{1}{n} + \frac{2}{3n^2} \right) + \frac{\lambda \beta_2}{(1-\rho)}, \quad n = 1, 2, \dots$$

The proof consists of a straightforward calculation of \bar{W}^{SP} according to formula (22), and then an application of Theorem 6.

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

$$\bar{W} = \frac{\alpha}{1 - \rho} + \frac{\lambda \beta_2}{2(1 - \rho)} \ .$$

This is not surprising since the star system with $n = \infty$ is equivalent to the system with a server scanning a closed tour of length 2α .

Appendix

Proof of Theorem 1. The process s(t) is regenerative, with regeneration epochs at the transitions into state $s_0 = \{*\}$ (the server is traveling and the system is empty). The duration of a regeneration cycle is

$$T_s = T_s^f + T_s^b ,$$

where T_s^f is the sojourn time of the process in state s_0 (this random variable has an exponential distribution with parameter λ); and T_s^b is the duration of the process busy period. Clearly, T^f and T_s^b are independent and therefore T_s has an absolutely continuous distribution. Thus it is sufficient to show that $E(T_s^b) < \infty$.

To proceed, it is convenient to revoke convention (A), i.e., let the 1-descendants of a customer C arise on the axis \mathbb{R} at just those times when they arrive during C's service in the original open system. In this version of the system, $\sigma(t)$ counts the number of customers that have not yet completed service by time t. Root customers sited ahead of the server are not counted; the server has not yet encountered them, so they are not yet considered to be present in the system.

The process $\sigma(t)$ is also regenerative at transitions into the state $\sigma_0 = 0$ corresponding to $s_0 = \{*\}$. Let T^b_σ denote the duration of a busy period in $\sigma(t)$. Clearly, T^b_σ and T^b_s are equal in distribution.

For convenience, suppose a busy period of $\sigma(t)$ starts at t=0 with the server located at point 0 on the axis. We compare $\sigma(t)$, $t \geq 0$, with a process $\hat{\sigma}(t)$, $t \geq 0$, counting the number of unfinished customers in the following modified system. Customer services are performed only at the points $x = \alpha n$, $n = 0, 1, 2, \ldots$ In each interval $(n\alpha, (n+1)\alpha)$ the server moves without stopping to serve any customers. The server collects all customers found in this interval and serves them at point $(n+1)\alpha$. By definition, $\sigma(0) = \hat{\sigma}(0) = 1$.

Let t_n , n = 0, 1, 2, ..., be the times that the server reaches points $n\alpha$, and let $T^b_{\hat{\sigma}} = \min\{t_n : \hat{\sigma}(t_n) = 0\}$, $\hat{n} = \min\{n \geq 0 : \hat{\sigma}(t_n) = 0\}$. Then it is easy to see that $T^b_{\hat{\sigma}}$ is stochastically no larger than $T^b_{\hat{\sigma}}$. Therefore, $E(T^b_{\hat{\sigma}}) \leq E(T^b_{\hat{\sigma}})$. For $E(T^b_{\hat{\sigma}})$ we have

$$E(T_{\hat{\sigma}}^b) = dE(\hat{n}) + \beta_1 E(\hat{k}) ,$$

where \hat{k} is the total number of customers served in $[0, T_{\hat{\sigma}}^b]$. Thus, it is sufficient to show that

$$E(\hat{n}) < \infty$$
 , (A.1)

$$E(\hat{k}) < \infty$$
 . (A.2)

To prove (A.1) and (A.2) consider the discrete-time Markov chain $Y(n) \equiv \hat{\sigma}(t_n)$, n = 0, 1, 2, ...The initial state of the chain is Y(0) = 1. Then

$$\hat{n} = \min\{n \ge 0 : Y(n) = 0\}$$
,

$$\hat{k} = \sum_{n=0}^{\hat{n}} Y(n) .$$

Let $F_n(z)$ be the generating function of the distribution of Y_n , and let $\beta(u) = E(e^{-uB})$ be the transform of the service time distribution. The Markov chain Y_n is described by

$$F_{n+1}(z) = F_n(\beta(\lambda - \lambda z))e^{-\alpha\lambda(1-z)}, \quad |z| \le 1, \quad n \ge 0.$$

Then the chain Y_n is ergodic by the Moustafa-Foster criteria [5], so (A.1) follows.

To prove (A.2) it is sufficient to show that

$$\sum_{i=0}^{\infty} i\rho_i < \infty \tag{A.3}$$

where $\{\rho_i; i=0,1,\ldots\}$ is the ergodic distribution of Y_n . To see this, note that Y_n can be considered regenerative in the 0 states. Then the expected value of the sum of the Y_n values during a busy period is equal to $\sum i\rho_i$ times the mean duration of a regeneration cycle. This observation and (A.3) imply (A.2).

Finally, (A.3) holds by the variant of the Moustafa-Foster criteria given below; the association $v_j = j$ then completes the proof of the theorem.

Lemma 4 Let Y(n), n = 0, 1, ..., be an irreducible Markov chain in the space $\mathbb{Z}_+ = \{0, 1, 2, ...\}$ with the transition matrix (ρ_{ij}) , $i, j \in \mathbb{Z}_+$. Let there exist a function $\nu_j \geq 0$, $j \in \mathbb{Z}_+$, a finite subset $A \subset \mathbb{Z}_+$, and a real number γ , $0 \leq \gamma < 1$, such that

- a) $\forall i \in A, \quad \sum_{j} \rho_{ij} \nu_{j} < \infty \; ;$
- b) $\forall i \in \mathbb{Z}_+ \setminus A$, $\sum_j \rho_{ij} \nu_j < \gamma \nu_i$;

c)
$$inf\{\nu_i: i \in \mathbb{Z}_+ \setminus A\} > 0$$
.

Then the chain Y(n) is ergodic. Moreover, its stationary distribution (ρ_i) , $i \in \mathbb{Z}_+$, satisfies the condition

$$\sum_{i\in\mathbb{Z}_+} \rho_i \nu_i < \infty .$$

We omit the proof since it is very similar to that of the usual Moustafa-Foster criteria [5].

References

- [1] Coffman, E. G., Jr., Gilbert, E. N., "A continuous polling system with constant service times," IEEE Trans. Inf. Theory, vol. IT-32, pp. 584-591, 1986.
- [2] Coffman, E. G., Jr., Gilbert, E. N., "Polling and greedy servers on a line," Queueing Systems

 Theory and Applications, 1987, no. 8.
- [3] Mack, C., Murphy, T., Webb, N. L., "The efficiency of N machines unidirectionally patrolled by one operative when walking time and repair times are constant," J. Roy. Statist. Soc. Series B, vol. 19, pp. 166-172 1957.
- [4] Kleinrock, L., Queueing Systems, vol. 1. New York: Wiley, 1975.
- [5] Moustafa, M. D., "Input-output Markov processes," Proc. Konjukijke Netherlands Acad. Wetenschappen, vol. 60, pp. 112-118, 1957.
- [6] Takagi, H., Analysis of Polling Systems, The MIT Press, Cambridge, MA, 1986.