Adv. Appl. Prob. **16,** 887–905 (1984)
Printed in N. Ireland

SINGLE-SERVER QUEUES WITH IMPATIENT CUSTOMERS

F. BACCELLI,* *INRIA*
P. BOYER,** *CNET*
G. HEBUTERNE,** *CNET*

Abstract

We consider a single-server queueing system in which a customer gives up whenever his waiting time is larger than a random threshold, his patience time. In the case of a *GI*/*GI*/1 queue with i.i.d. patience times, we establish the extensions of the classical *GI*/*GI*/1 formulae concerning the stability condition and the relation between actual and virtual waiting-time distribution functions. We also prove that these last two distribution functions coincide in the case of a Poisson input process and determine their common law.

QUEUEING THEORY; LIMITED WAITING TIMES; ERGODIC MARKOV CHAIN; ACTUAL WAITING TIMES; VIRTUAL WAITING TIMES; REGENERATIVE PROCESSES; INVARIANT MEASURE; FUNCTIONAL EQUATION

1. Introduction

In most studies of queueing systems, little attention is paid to practical limitations such as finiteness of queue length (limited capacities) or finiteness of waiting times (time out, or limited patience); however, such phenomena are often encountered in telecommunication systems:

– In a telecommunication network, a subscriber may give up due to impatience before the connexion he asks for is completely established, resulting in inefficient use of resources.

– In a packet switching network, the switching nodes have limited buffer capacities. Hence, an arriving customer is accepted only if his size added to the sizes of the packets already present in the node is smaller than the total capacity. Since the output rate is constant, this is equivalent to a limitation of its waiting time.

Systems with limited waiting times can be classified as follows:

– The limitation acts only on waiting time *or* on sojourn time (waiting + service).

Received 15 November 1982; revision received 1 November 1983.

* Postal address: INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay Cedex, France.

** Postal address: CNET, LAA-SLC-EVP, Route de Tregastel, 22301 Lannion A, France.

– The customer can calculate his prospective waiting time at the arrival epoch and balks if this exceeds his patience *or* he joins the queue regardless, leaving the system if and when his patience expires.

Combining these two distinctions gives four queueing systems with 'impatient customers':

(a) limitation on sojourn time, aware customers. The entering customer leaves immediately if he knows that his total sojourn time is beyond his patience (in such a system, all server work is useful). Ergodicity conditions for general single-server queues are given in [6]. Some special cases are solved in [12], [15].

(b) limitation on sojourn time, unaware customers. This is the case if customers do not know anything about the system and are unaware of the beginning of service (e.g. a calling subscriber waiting for a dialing tone). In this case service may be interrupted by discouragement, so that some server work may be useless. Some special cases can be found in [8], [10], [22].

(c) limitation on waiting time, aware customers. The same as (a) above, with the impatience acting only on waiting time.

(d) limitation on waiting time, unaware customers. The same as (b) above with the impatience acting only on waiting time.

The study of systems (c) and (d) can be unified through the following remark. As long as we are concerned with rejection probabilities, or with the waiting time distributions of successful customers, the finally discouraged customers (of Case (d)) do not influence the system and can be discarded on arrival (as in (c)). The correctness of this statement will be made clear if one realizes that (supposing service in order of arrivals), the fate of an arriving customer depends only on the *unfinished work* of the server, which is clearly not modified by customers who finally leave impatiently, even if they stay in queue (see the remark on virtual waiting times in Section 2.1).

The present paper is devoted to the analytical characterization of waiting times in system (c) (and hence (d)). For this, we use the notation $G/G/m+G$: the first three symbols have the same meaning as in Kendall's notation and the last one specifies the impatience law. Section 2 is concerned with $GI/GI/1+GI$ queues. Some functional equations are established for the distribution functions of the waiting times offered to customers. This approach was investigated by Pollaczek [19], who reduced the problem to the resolution of a set of (unsolved) integral equations. Our contribution concerning these general queues consists in determining the condition assuring stability, by means of probabilistic methods. In Section 3, we establish some relations concerning the probabilities of rejection and of an empty system, by means of the limit theorems on regenerative processes. We also study the virtual waiting time and relate it to the actual offered waiting time.

In Section 4 we limit ourselves to $M/GI/1+GI$ queues. The stationary distribution functions of actual and virtual offered waiting times are shown to coincide and are given by means of the resolvent of a Volterra equation. In the special cases of exponential and Erlangian impatience distribution functions, series-form solutions are given, generalizing the results obtained by Barrer [2] on $M/M/1+D$ and Gnedenko and Kovalenko [13] on $M/M/m+D$ and $M/M/m+M$. For the $M/M/m+GI$ system, see [1] and [14].

2. On $GI/GI/1+GI$ queues

2.1 *Assumptions and notation.* In this section, we consider a first-in–first-out single-server queueing system in which customers are subject to impatience. More precisely, let T_n, $n\in\mathbb{N}$ be the arrival epoch of the nth customer ($T_0\equiv 0$). For $n>0$, let

$t_n \triangleq T_n - T_{n-1}$: the nth interarrival time ($t_n\in\mathbb{R}^+$)
s_n: the service time of the nth customer ($s_n\in\mathbb{R}^+$)
g_n: the patience time of the nth customer ($g_n\in\mathbb{R}^+$).

Let w_n, $n\in\mathbb{N}$ be the workload just before T_n (unfinished work). We assume the system to be of type (c) of Section 1: the nth customer enters the system only if the time to wait for accessing the server does not exceed his own patience. That is:
If $g_n \leqq w_n$ the nth customer is impatient and does not enter.
If $g_n > w_n$ the nth customer stays in queue.

Remark. For system (d), where all customers enter the queue, we ought to say:

If $g_n \leqq w_n$, the nth customer does not modify w_n, the workload of the server.
If $g_n > w_n$, the nth customer will be served and thus modify w_n.

This formulation is clearly equivalent to the previous one: the evolution of w_n in the two cases will be the same, establishing the equivalence of systems (c) and (d). We make the following assumptions: $\{t_n, n\in\mathbb{N}\}$ (or $\{s_n, n\in\mathbb{N}\}$, $\{g_n, n\in\mathbb{N}\}$) is a sequence of independent and identically distributed random variables on $\mathbb{R}^+$ with distribution function $A(x)$ (or $B(x)$, $C(x)$, $x\in\mathbb{R}^+$). $A(x)$ and $B(x)$ are supposed to have finite first moments denoted as $1/\lambda$ and $1/\mu$ respectively. As usual, we set $\rho=\lambda/\mu$. $C(x)$ may be defective (i.e., we may have $\lim_{x\to\infty} C(x)\neq 1$) but we assume that $C(0)=0$.

Throughout the paper, we mainly use $G(x)\triangleq 1-C(x)$.

2.2. *Recursive equations for the offered waiting times.* We now derive a

recursive equation for the sequence $\{w_n, n \in \mathbb{N}\}$ generalizing the Loynes equation [3]. Notice that w_n is the time that the nth customer would have to wait for accessing the server if he were sufficiently patient. Hence, we call it the actual offered waiting time. Let $w_0 \in \mathbb{R}^+$ be some initial condition. We have for $n \geqq 0$:

$$\begin{cases} w_{n+1} = [w_n + s_n - t_{n+1}]^+ & \text{if} \quad g_n > w_n \\ w_{n+1} = [w_n - t_{n+1}]^+ & \text{otherwise.} \end{cases} \tag{2.1}$$

With our assumptions, $\{w_n, n \in \mathbb{N}\}$ is a Markov chain with state space $\mathbb{R}^+$ and transition kernel

$$\begin{cases} P(x, A) \triangleq P[w_{n+1} \in A \mid w_n = x] \\ x \in \mathbb{R}^+, \qquad A \in \mathcal{B}(\mathbb{R}^+) \end{cases}$$

given by

$$\begin{cases} P(x, A) = G(x) \displaystyle\int_{\mathbb{R}^+ \times \mathbb{R}^+} 1_A([x + y - z]^+)\, dB(y)\, dA(z) \\ \qquad\qquad + (1 - G(x)) \displaystyle\int_{\mathbb{R}^+} 1_A([x - z]^+)\, dA(z) \\ \text{where } 1_A(u) = 1 \quad \text{iff} \quad u \in A. \end{cases} \tag{2.2}$$

Let $W_n(x)$, $x \in \mathbb{R}^+$ be the distribution function of w_n. We therefore have the following integral equation for the W_n's:

$$W_{n+1}(x) = \int_0^\infty dA(t) \left[W_n(t + x) - \int_{u=0}^{t+x} G(u)[1 - B(t + x - u)]\, dW_n(u) \right]. \tag{2.3}$$

This relation is derived from (2.2) after some manipulations.

2.3. *Stability condition*

2.3.1. *Sufficient condition.* This section is devoted to the determination of a sufficient condition for w_n to be an ergodic Markov chain (and hence for (2.3) to have a unique stationary solution). It is based on the method proposed by Laslett, Pollard and Tweedie in [16]. Let

$$\begin{cases} a \triangleq \inf\,(t / A(t) = 1) \\ b \triangleq \sup\,(t / B(t) = 0). \end{cases}$$

Lemma 1. Assume $b - a < 0$. Then the Markov chain $\{w_n, n \in \mathbb{N}\}$ with transition kernel $P(x, A)$ (2.2) is ε_0-irreducible (where ε_0 is a measure on $\mathbb{R}^+$ concentrated in $\{0\}$).

Proof. Consider

$$\begin{cases} z_0 = w_0 \\ z_{n+1} = [z_n + s_n - t_{n+1}]^+ \quad n \geqq 0. \end{cases}$$

When comparing with (2.1), we get by induction

$$w_n \leqq z_n \qquad \forall n \in \mathbb{N}. \tag{2.4}$$

We have furthermore, $\forall \varepsilon > 0$, $\exists p > 0$,

$$P[b - a \leqq s_n - t_{n+1} \leqq b - a + \varepsilon] = p > 0.$$

Let $x \in \mathbb{R}^+$ and $k = [x/(a-b)]$ (where $[y]$, $y \in \mathbb{R}^+$ denotes the greatest integer smaller than y); we have

$$\begin{aligned} P\left[\bigcup_{n \geqq 1} \{w_n = 0\} \mid w_0 = x\right] &\geqq P\left[\bigcup_{n \geqq 1} \{z_n = 0\} \mid w_0 = x\right] \\ &\geqq P\left[\bigcap_{0 \leqq i \leqq k} \{b - a \leqq s_i - t_{i+1} \leqq b - a + \varepsilon\}\right] > p^k. \end{aligned} \tag{2.5}$$

Lemma 2. Assume $b - a < 0$. Let $\rho = \lambda/\mu$. The Markov chain $\{w_n, n \in \mathbb{N}\}$ is ergodic under the condition

$$1 - \rho G(\infty) > 0. \tag{2.6}$$

Proof. For any $\beta \in \mathbb{R}^+$, let B be the interval $[0, \beta]$. We first prove that B is a test set for the Markov chain. Then, we show that when the condition $0 < 1 - \rho G(\infty)$ is fulfilled, and for a sufficiently large β, the mean hitting time of this test is a bounded function, so that the chain is proved to be ergodic (see [16]).

First part of the proof. Since w_n is ε_0-irreducible, B will be proved to be a test set if one can find $N > 0 \in \mathbb{N}$ and $\delta > 0 \in \mathbb{R}^+$ such that

$$\max_{0 \leqq n \leqq N} P^n(y, \{0\}) \geqq \delta, \qquad \forall y \in B$$

(see Theorem 3.2 in [16]). Consider the sequence $\{z_n, n \in \mathbb{N}\}$ defined in the proof of Lemma 1. $\forall y \in B$, we have

$$P^n(y, \{0\}) = P[w_n = 0 \mid w_0 = y] \geqq P[z_n = 0 \mid w_0 = y].$$

Thus any $N \geqq \beta/(a-b)$ matches.

Second part of the proof. We have to show that T_B, the hitting time of B, satisfies

$$\sup_{x \in B} E[T_B \mid w_0 = x] < \infty. \tag{2.7}$$

This will be proved if one can find $\varepsilon > 0$ and $M < \infty$ such that

$$E[w_1 \mid w_0 = x] \leqq x - \varepsilon \qquad \forall x \in B^c \tag{2.8}$$

$$E[w_1 \mid w_0 = x] \leqq M \qquad \forall x \in B \tag{2.9}$$

(see Theorem 2.2 in [16]). We first derive from (2.2)

$$E[w_1 \mid w_0 = x] = \int_0^x A(t)\,dt + G(x)\int_x^\infty (1 - B(t-x))A(t)\,dt.$$

Hence for $x \leqq \beta$, $E[w_1, w_0 = u] \leqq \beta + 1/\mu$, proving (2.9). Concerning (2.8) we get

$$E[w_1 \mid w_0 = x] \leqq x - \frac{1 - \rho G(x)}{\lambda} + \int_x^\infty (1 - A(t))\,dt.$$

Assume now that the condition $0 < 1 - \rho G(\infty)$ is fulfilled. Then there exists $\varepsilon > 0$ $x_1 \in \mathbb{R}^+$ such that for $x > x_1$, $E[w_1 \mid w_0 = x] < \varepsilon$. Hence, if $\beta > x_1$, $[0, \beta]$ is a test set with bounded mean hitting time, completing the proof.

2.3.2. *Necessary condition*

Lemma 3. If $\rho G(\infty) > 1$, $w_n \to \infty$ a.s.

Proof. From Equation (2.1) we get

$$w_{n+1} \geqq w_n + s_n \quad 1_{(g_n = \infty)} - t_{n+1}, \qquad n \geqq 0.$$

Hence

$$\begin{cases} w_{n+1} \geqq w_0 + \sum_{i=1}^{n} u_i \\ \text{where} \quad u_i \triangleq s_i \; 1_{\{g_i = \infty\}} - t_{i+1}. \end{cases} \tag{2.10}$$

If $0 > 1 - \rho G(\infty)$, $E[u_i] > 0$. Thus the right-hand side converges almost surely to ∞ (strong law of large numbers).

2.3.3. *Remark.* In Section 3.2 we shall also derive from the irreducibility of w_n that the busy periods are integrable. Under the former two assumptions ($b - a < 0$ and $1 - \rho G(\infty) < 0$) sufficient conditions for the integrability of busy periods in such state-dependent queues have already been obtained in [23]. Our condition on the support of A and B ($b - a < 0$) is however finer than the condition (iii) of Theorem 3 in [23].

Nevertheless, it is still possible to have $b - a \geqq 0$ and w_n ergodic (infinite busy period and stable w_n can be obtained for instance when $C(t) = 1$ for $t > T < \infty$ and $b > a$).

3. The use of regenerative processes for the $GI/GI/1+GI$ queue

3.1. *Introduction and notation.* In this section we apply the regenerative approach (see [7], [9], [20]) to the $GI/GI/1+GI$ queue in order to derive general relationships.

Let $\{C_i, i \geqq 1\}$ be the sequence of successive busy cycles (a busy cycle is the duration between two successive arrival epochs of customers finding an idle server). The C_i's are i.i.d. random variables on $\mathbb{R}^+$.

Let $\{P_i, i \geqq 1\}$ be the sequence of busy periods (the part of the busy cycle during which the server continuously works). The P_i's are i.i.d. random variables on $\mathbb{R}^+$.

Let $\{l_i, i \geqq 1\}$ (or $\{m_i, i \geqq 1\}$) be the number of customers arrived in the system (or successfully served) during the ith busy period, including the one initializing it. The l_i's (or m_i's) are i.i.d. on $\mathbb{N}$.

Let σ_n, $n \geqq 1$ be the σ-field generated by the random variables $(t_1, \cdots, t_n, s_0, \cdots . s_{n-1}, g_0, \cdots, g_{n-1})$.

In what follows we use the following renewal processes:

$$\begin{cases} \tau_0 = 0 \\ \tau_i = \tau_{i-1} + C_i \qquad i \geqq 1 \text{ on } \mathbb{R}^+ \end{cases}$$

$$\begin{cases} \Lambda_0 = 0 \\ \Lambda_i = \Lambda_{i-1} + l_i \qquad i \geqq 1 \text{ on } \mathbb{N} \end{cases}$$

$$\begin{cases} K_0 = 0 \\ K_i = K_{i-1} + m_i, \qquad i \geqq 1 \text{ on } \mathbb{N}. \end{cases}$$

We essentially study the *virtual offered waiting time*: let v_t be the time a test customer of infinite patience arriving at t would have to wait under FIFO discipline, or equivalently the unfinished work at time t. In order to use regenerative limit theorems one has first to prove finiteness of busy cycles.

Throughout this section, we assume that (2.6) holds, (i.e. the ergodicity of $\{w_n\}$).

3.2. *The finiteness of busy cycles.* Our aim is to prove that, under the conditions of Lemma 2, then

$$EC_1 < \infty. \tag{3.1}$$

Using the regenerative property of the l_i's, one gets

$$\exists \lim_{n\to\infty} P[W_n = 0] = (E[l_1])^{-1} \tag{3.2}$$

both members being here either positive or 0. Furthermore, since W_n is an

ε_0-irreducible Markov chain, $\limsup_{n\to\infty} P[W_n = 0] < 0$, for any initial condition ([23], Proposition 4.2). Hence $E[l_1] < \infty$ as well as $E[C_1] < \infty$.

3.3. *Some consequences of Wald's lemma.* Clearly, one can write

$$C_1 = t_1 + t_2 + \cdots + t_{l_1} \tag{3.3a}$$

$$P_1 = s_{i_1} + s_{i_2} + \cdots + s_{i_{m_1}} \tag{3.3b}$$

where $(i_1, i_2, \cdots, i_{m_1})$ are successful customer's numbers. Wald's lemma holds for C_1 as well as for P_1 [17] yielding:

$$EC_1 = El_1 \cdot Et_1 = (1/\lambda) El_1 \tag{3.4}$$

$$EP_1 = Em_1 \,.\, Es_1. \tag{3.5}$$

Notice that $l_1 = \inf(j \geqq 1, t_1 + \cdots + t_j > r_0 + \cdots + r_{j-1})$ is a stopping time for the sequence σ_n. The proof of (3.4) follows directly from Wald's lemma.

The proof of (3.5) uses Wald's argument (see [4]). We have

$$i_1 = \inf(j \geqq 0,\ w_j < g_j)$$

$$i_k = \inf(j > i_{k-1},\ w_j < g_j).$$

First, note that the event $\{i_k = n\}$ is $\sigma(t_1, \cdots, t_n, g_0, \cdots, g_n, s_0, \cdots, s_{n-1})$ measurable. As a consequence, $\{i_k = n\}$ is independent of s_n.

Then, note that the event $\{m_1 \geqq k, i_k = n\}$ is independent of s_n. $P_1 = \sum_{k=1}^{m_1} s_{i_k}$ can also be written as

$$\sum_{k \geqq 1} \sum_{j \geqq k} s_j \,.\, 1_{\{m_1 \geqq k, i_k = j\}}$$

and the proof follows immediately.

3.4. *The virtual offered waiting time.* We have defined, v_t, the virtual offered waiting time. This stochastic process is regenerative with regeneration points $(\tau_i)_{i \geqq 0}$.

Assuming the finiteness of busy cycles, v_t converges in distribution as t goes to ∞:

$$\lim_{t\to\infty} P(v_t \leqq x) = \frac{1}{EC_1} E\left\{\int_0^{c_1} 1_{(v_t \leqq x)}\, dt\right\} \tag{3.6}$$

and the limit theorem for regenerative processes states the existence of the left-hand side ([7], Chapter IX, Theorem 2.25).

Equation (2.6) gives a necessary and sufficient condition for this convergence.

Let us call $V(t)$ this limiting distribution function and V_∞ a random variable which has this limiting distribution.

3.5. *The probability of rejection.* Let Z_n be a random variable such that the

event $\{Z_n = 1\}$ means that the nth customer is rejected:

$$Z_n = 1_{\{g_n < w_n\}}.$$

$\{Z_n, n \geqq 0\}$ is regenerative relative to the discrete renewal process $\{\Lambda_i\}$; since $E[l_1] < \infty$, the limit theorem [20] states that

$$\pi \triangleq \lim_{n\to\infty} \Pr\{Z_n = 1\} = \frac{1}{El_1} E\left[\sum_{j=0}^{l_1-1} 1\{Z_j = 1\}\right] \tag{3.7}$$
$$= 1 - Em_1/El_1,$$

where π is defined as the steady-state probability of being rejected. Now, (3.6) taken at $x = 0$ yields

$$V(0) = \frac{1}{EC_1} E\left[\int_0^{C_1} 1_{\{V_t = 0\}}\, dt\right]$$
$$= \frac{1}{EC_1} E[C_1 - P_1].$$

Comparing with (3.7), where we insert (3.4), (3.5) leads to

$$(1-\pi)\rho = 1 - V(0). \tag{3.8}$$

3.6. *A relation between actual and virtual waiting time.* For a $GI/GI/1$ queue with FIFO discipline, previous work ([9], formulae 6·15–6·17) has established a relation between the limiting distribution of v_t and w_n, that is

$$E\exp(-sV_\infty) = 1 - \rho + a(s)\,.\,E\exp(-sW_\infty)$$

where

$$a(s) = \begin{cases} \rho & \text{if} \quad s = 0 \\ \lambda \dfrac{1 - E\exp(-ss_1)}{s} & s > 0. \end{cases}$$

As will be seen in this section, a similar relation holds for a $GI/GI/1+GI$ queue with FIFO discipline.

We go back to (3.6):

$$E\exp(-sV_\infty) = \frac{1}{EC_1} E\left\{\int_0^{C_1} \exp(-sv_t)\, dt\right\}$$

using (3.8)

$$= P(V_\infty = 0) + \frac{1}{EC_1} E\left\{\int_0^{P_1} \exp(-sv_t)\, dt\right\}.$$

Then, we perform integration truncating as follows:

$$\int_0^{P_1} = \sum_{k=1}^{l_1} \int_{T_{k-1}}^{T_k} + \int_{T_{l_1}}^{P_1}.$$

For $0 \leqq t < P_1$, w_n and v_t are related as follows:

$$1 \leqq k \leqq l_1 - 1 \qquad T_{k-1} \leqq t \leqq T_k$$
$$v_t = w_{k-1} + b_{k-1} - (t - T_{k-1})$$
$$T_{l_1-1} \leqq t < P_1$$
$$v_t = w_{l_1-1} + b_{l_1-1} - (t - T_{l_1-1}).$$

The use of Lindley's relation leads to

$$E \exp(-sV_\infty) = P(V_\infty = 0) + \frac{1}{EC_1} E\left\{\sum_{k=0}^{l_1-1} \exp(-sw_k) \frac{1 - \exp(-sb_k)}{s}\right\} \tag{3.9}$$

for $k \geqq 0$ $b_k = \mathbf{s}_k \,.\, 1_{(g_k > w_k)}$. Then

$$E \exp(-sV_\infty) = P(V_\infty = 0) + \frac{1}{EC_1} E\left\{\sum_{k=0}^{l_1-1} \exp(-sw_k) \,.\, 1_{(g_k > w_k)} \frac{1 - \exp(-s\mathbf{s}_k)}{s}\right\}.$$

Now, using Wald's argument,

$$E\left\{\sum_{k=1}^{l_1} \exp(-sw_k) 1_{(g_k > w_k)} \frac{1 - \exp(-s\mathbf{s}_k)}{s}\right\}$$
$$= E\left\{\sum_{j \geqq 1} 1(l_1 - 1 = j) \sum_{k=0}^{j} \exp(-sw_k) \,.\, 1_{(g_k > w_k)} \frac{1 - \exp(-s\mathbf{s}_k)}{s}\right\}$$
$$= E\left\{\sum_{k \geqq 0} \exp(-sw_k) 1_{(g_k > w_k)} \,.\, 1_{(l_1 \geqq k)} \frac{1 - \exp(-s\mathbf{s}_k)}{s}\right\}.$$

$\mathbf{s}_k$ is independent of g_k, w_k and $1_{(l_1 \geqq k)}$ which are $\tau(s_0, \cdots, s_{k-1}, t_1, \cdots, t_k, g_0, \cdots g_k)$-measurable, $k \geqq 1$. Going back to (3.9) this leads to

$$E \exp(-sV_\infty) = P(V_\infty = 0) + \frac{1}{EC_1} E\left\{\frac{1 - \exp(-s\mathbf{s}_k)}{s}\right\} E\left\{\sum_{k=0}^{l_1-1} \exp(-sw_k) \,.\, 1_{(g_k > w_k)}\right\}. \tag{3.10}$$

Then we notice that

$$\sum_{k=0}^{l_1-1} \exp(-sw_k) \,.\, 1_{(g_k > w_k)} = \sum_{k=1}^{m_1} \exp(-sw_{i_k})$$

where $(i_k)_{1 \leqq k \leqq m_1}$ are successive successful customer's numbers in a busy cycle.

The stochastic process $(w_{i_k})_{k \geqq 1}$ is regenerative with regeneration points $(m_i)_{i \geqq 1}$. As k goes to ∞, it converges in distribution. We denote by W_∞ a random variable which has this limiting distribution:

$$E \exp(-sW_\infty) = \frac{1}{Em_1} E\left\{\sum_{k=1}^{m_1} \exp(-sw_{i_k})\right\}.$$

(3.10) leads to

$$E \exp(-sV_\infty) = P(V_\infty = 0) + (1 - \pi) E\left\{\frac{1 - \exp(-s\mathbf{s}_1)}{s}\right\} E \exp(-sW_\infty).$$

We define $a(s)$:

$$a(s)=\begin{cases}\rho & \text{if} \quad s=0\\ \lambda E\left[\dfrac{1-\exp(-s s_1)}{s}\right] & s>0.\end{cases}$$

One obtains the following relation between V_∞ and W_∞:

$$E\exp(-sV_\infty)=P(V_\infty=0)+(1-\pi)a(s)E\exp(-sW_\infty). \tag{3.11}$$

Remarks

1. As an immediate consequence, here is a kind of Pollaczek–Kintchine formula:

$$EV_\infty=(1-\pi)\rho\left[EW_\infty+\frac{\lambda}{2}\frac{Es_1^2}{Es_1}\right].$$

Dropping the $(1-\pi)$ factor, one obtains an equivalent relation for a $GI/GI/1$ queue, as established by Cohen ([9], expression 6.17).

2. Recall that (w_n) is the actual offered waiting time (see Section 2.2) (w_n) is an ergodic Markovian chain. Let $W(u)$ be the limiting distribution function of (w_n). Let $W_\infty(u)$ be the limiting distribution function of (w_{i_k}) as $k\to\infty$. We have $dW(u)=1/(1-\pi)G(u)\,dW(u)$ and the equation (3.11) can be read

$$\int e^{-su}\,dV(u)=V(0)+a(s)\int_0^\infty e^{-su}G(u)\,dW(u). \tag{3.11′}$$

3. (3.11′) implies the complete convergence of $\{V_t\}$: as $s\to 0$:

$$\int_0^\infty dV(u)=V(0)+\rho\int_0^\infty G(u)\,dW(u).$$

From (3.8)

$$1-\pi=\int_0^\infty G(u)\,dW(u)=\frac{1-V(0)}{\rho}$$

and thus

$$\int_0^\infty dV(u)=V(0)+[1-V(0)]=1.$$

4. (3.11) proves the necessity of (2.6) for V_t to converge:

$$\int_0^\infty e^{-su}\,dV(u)-V(0)=a(s)\int_0^\infty e^{-su}G(u)\,dW(u)$$

$$\geqq a(s)G(\infty)\int_0^\infty e^{-su}\,dW(u).$$

If v_t converges properly, one has as $s\to 0$

$$1-V(0)\geqq\rho G(\infty),$$

from which (2.6) can be derived, since $V(0)>0$.

3.7 *A semi-regenerative equation.* The discrete-time Markov chain $\{w_n, n \in \mathbb{N}\}$ is imbedded in the continuous-time Markov process $\{v_t, t \in \mathbb{R}^+\}$:

$$w_n = v_{T_n}. \tag{3.12}$$

This property allows to establish the complete convergence of v_t as a consequence of the limit theorems on semi-regenerative processes. Let

$$\begin{cases} K_t(x, B) = P[v_t \in B,\ T_1 > t \mid v_{0^+} = x] B \in \mathcal{B}(\mathbb{R}^+) \\ \qquad\qquad = P[v_t \in B \mid T_1 > t,\ v_{0^+} = x] \,.\, P[T_1 > t \mid v_{0^+} = x] \\ \qquad\qquad = \delta_{(x-t)^+}(B)[1 - A(t)], \end{cases}$$

where $\delta_u(B) = 1$ iff $u \in B$. When (2.6) holds, $\{W_n\}$ is an ergodic Markov chain. The limiting distribution exists, and is non-defective.

Let $J(x)$, $x \in \mathbb{R}^+$, be the distribution function of $V_{T_n^+}$:

$$J(x) = \int_0^x (G(u)B(x-u) + 1 - G(u))\, dW(u).$$

The limit theorem on semi-regenerative processes yields

$$\exists \lim_{t \to \infty} P[V_t \in B] = \begin{cases} \dfrac{1}{E[T_1]} \displaystyle\int_0^\infty J(dx) \int_0^\infty K_t(x, B)\, dt \\ \lambda \displaystyle\int_0^\infty J(dx) \int_0^t (1 - A(t)) \delta_{(x-t)^+}(B)\, dt, \end{cases}$$

i.e.

$$V(x) = \begin{cases} \lambda \displaystyle\int_0^\infty (1 - A(t)) J(t + x)\, dt \\ \lambda \displaystyle\int_0^\infty (1 - A(t)) \left[W(t + x) - \int_0^{t+x} G(u)(1 - B(x + t - u))\, dW(u) \right] dt. \end{cases} \tag{3.13}$$

4. On $M/GI/1 + GI$ queues

Throughout this section, we assume that

$$A(x) = \begin{cases} 1 - \exp(-\lambda x) & x \geqq 0 \\ 0 & x \leqq 0. \end{cases}$$

4.1. *Functional equation for the workload distribution function.*† We study the virtual offered waiting time v_t for the $M/GI/1 + GI$ queue. An integral equation

† Equation (4.3) could have been derived from (3.11) and the equality property of Section 4.2: $V = W$. The simplicity of a direct proof justifies this paragraph.

generalizing Takács's equation [21] is then established. We make use of the properties of our system, namely that v_t is modified only by successful customers.

Let $V(t, x)$ be the distribution function of v_t, and $\psi(t, s)$ the Laplace–Stieltjes transform of $V(t, x)$:

$$V(t, x) = P[V_t \leqq x] \qquad t, x \in \mathbb{R}^+$$

$$\psi(t, s) = \int_{0^-}^{\infty} e^{-sx}\, dV(t, x) \qquad t \in \mathbb{R}^+,\ s \in \mathbb{C},\ \mathrm{Re}\,(s) \geqq 0.$$

We proceed now as for Takács's equation, outlining the most important steps. Using the Markovian property of v_t, we get:

$$(4.1) \quad V(t+\Delta, x) = V(t, x+\Delta) + \lambda\Delta \int_0^{x+\Delta} G(u)(1-B(x-u))\, d_u V(t, u) + O(\Delta).$$

We multiply both sides by $\exp(-sx)$ and sum up for $x \in [0, \infty]$. Using the relation

$$\psi(t, s) = s \int_{0^-}^{\infty} e^{-sx} V(t, x)\, dx,$$

(4.1) becomes

$$\frac{1}{s}\psi(t+\Delta, s) = \exp(s\Delta)\left(\frac{\psi(t, s)}{s} - \Delta V(t, \Delta)\right) + O(\Delta)$$

$$-\lambda\Delta \int_{-0}^{\infty} dx \exp(-sx) \int_0^x G(u)[1-B(x-u)]\, dV(t, u).$$

By analogy to $\psi(t, s)$, define

$$(4.2) \qquad \psi_G(t, s) = \int_{0^-}^{\infty} \exp(-su) G(x)\, d_x V(t, x).$$

After reversing summations and letting $\Delta \to 0$, the above equation finally becomes

$$\frac{1}{s}\frac{\partial \psi}{\partial t} = \psi(t, s) - V(t, 0) - a(s)\psi_G(t, s)$$

(4.3)

$$\text{where} \begin{cases} a(s) \triangleq \lambda \dfrac{1-B^*(s)}{s} \\ B^*(s) \text{ is the Laplace–Stieltjes transform of } B. \end{cases}$$

4.2. *On the stationary distributions of actual and virtual offered waiting time.* The equations (2.3) and (4.3) will always have a stationary solution if (2.6) holds (the conditions of Lemma 2 being fulfilled).

On one hand, $W(x) \triangleq \lim_{n\to\infty} W_n(x)$ exists and is the unique solution of (2.3):

$$W(x) = \int_0^\infty dA(t)\left[W(t+x) - \int_0^{t+x} G(u)[1-B(t+x-u)\,dW(u)]\right]. \tag{2.3}$$

On the other hand, (3.13) yields

$$V(x) = \int_0^\infty [1-A(t)]\left[W(t+x) - \int_0^{t+x} G(u)[1-B(t+x-u)]\,dW(u)\right] dt. \tag{4.4}$$

We remark that for $A(t) = 1-\exp(-\lambda t)$, $\lambda(1-A(t))$ and dA/dt coincide, which means that $V = W$.

This result extends Kintchine's theorem, proving the equality of the stationary distribution functions of w_n and v_t.

Remarks. Making $V = W$ reduces (3.11) to the stationary version of (4.3), as needed:

$$\psi(s) = V(0) + a(s)\psi_G(s). \tag{4.5}$$

4.3. *Resolution of the functional equation.*

4.3.1. *The density of the stationary virtual offered waiting time distribution function.* Let us assume the existence of a stationary solution. To prove the existence of a probability density function for the virtual offered waiting time distribution function, we use the following lemma (see for instance [11]).

Lemma 4. For ψ to be of the form

$$\psi(s) = \int_0^\infty \exp(-sx) f(x)\,dx \quad \text{where} \quad 0 \leqq f \leqq A$$

it is sufficient and necessary that

$$0 \leqq \frac{(-s)^n}{n!}\psi^{(n)}(s) \leqq \frac{A}{s} \tag{4.6}$$

for all $s>0$ and all n, where $\psi^{(n)}(s)$ denotes the nth derivative of $\psi(s)$. We apply the criterion to $\psi(s) - V(0)$. Let $\phi(s)$ be the Laplace–Stieltjes transform of $V(x) - V(0)$:

$$\phi(s) = \psi(s) - V(0) = a(s)\psi_G(s) \quad \text{from} \quad (4.5). \tag{4.7}$$

At this point we note that $a(s)$ is the Laplace transform of the 'unfinished work', and as such has a density:

$$a(s) = \int_0^\infty \exp(-sx)\alpha(x)\,dx, \quad \alpha(x) = \lambda[1-B(x)].$$

Thus $a(s)$ satisfies the conditions of the above lemma; let D be the maximum

of its density. For $n=0$ we use the simple bound

$$\psi_G(x)=\int_0^\infty e^{-sx}G(x)\,dV(x)\leqq\psi(s)\leqq 1$$

and so

$$0\leqq\phi(s)\leqq\frac{D}{s}$$

satisfies (4.6) for $n=0$. For $n\geqq 1$, (4.7) above gives

$$\phi^{(n)}(s)=\sum_{j=0}^{n}C_n^j\psi_G^{(j)}(s)a^{(n-j)}(s).$$

$a(s)$ is the Laplace–Stieltjes transform of a density

$$0\leqq(-1)^l a^{(l)}(s)\leqq\frac{Dl!}{s^{l+1}}.$$

Therefore

$$(-1)^n\phi^{(n)}(s)\leqq\frac{Dn!}{s^{n+1}}\sum_{j=0}^{n}\frac{(-s)^j}{j!}\psi_G^{(j)}(s).$$

From the definition of ψ_G, we write

$$\delta\triangleq\sum_{j=0}^{n}\frac{(-s)^j}{j!}\psi_G^{(j)}(s)\leqq\int_0^\infty\left[\sum_{j=0}^{n}\frac{(sx)^j}{j!}\right]\exp(-sx)G(x)\,dV(x)\leqq 1.$$

We now complete the proof. Denote by $V(0)$ the mass at the origin and by $v(x)$ the density. (4.5) may be inverted as follows:

$$\begin{aligned}v(x)&=\lambda V(0)[1-B(x)]+\lambda\int_0^x v(u)G(u)[1-B(x-u)]\,du\\ V(0)&+\int_0^\infty v(x)\,dx=1.\end{aligned}\tag{4.8}$$

4.3.2. *Resolution of the functional equation for M/GI/1+GI.* In this section we derive the general solution of Equation (4.8) when (2.6) is satisfied and B and G are continuous. (4.8) is shown to be a Fredholm integral equation of the second kind. The method of the resolvent yields integral series for the desired density function. In the following sections, further results are obtained concerning the special cases of Poisson and Erlang impatience distribution functions, in terms of series for the Laplace–Stieltjes transforms.

Consider the following functions:

$$\begin{cases}f(s)=\lambda[1-B(s)]\\ K(s,t)=G(t)[1-B(s-t)] & s\geqq 0,\ t\geqq 0\\ \hat{v}(s)=v(s)/V(0).\end{cases}\tag{4.9}$$

From Equation (3.14), $\hat{v}(s)$ is the solution of

$$\begin{cases} \hat{v}(s) = f(s) + \lambda \int_0^s K(s,t)\hat{v}(t)\,dt \\ s \geqq 0. \end{cases} \tag{4.10}$$

Due to our assumptions concerning the existence of the first moment of B, $f(s)$ is an absolutely integrable function; thus (4.10) is a Volterra equation for which the method of the resolvent applies. Let

$$\begin{cases} K_m(s,t) = \int_t^s K(s,x)K_{m-1}(x,t)\,dx \\ m \geqq 2 \\ K_1(s,t) = K(s,t) \qquad s \geqq 0,\ t \geqq 0. \end{cases} \tag{4.11}$$

An induction yields the following expression:

$$\begin{cases} K_m(s,t) = G(t) \int_t^s G(x_{m-1}) \,.\, (1 - F(s - x_{m-1})). \\ dx_{m-1} \int_t^{x_{m-1}} G(x_{m-2}) \,.\, (1 - F(x_{m-1} - x_{m-2})). \\ dx_{m-2} \int_t^{x_{m-2}} \cdots \\ dx_2 \int_t^{x_2} G(x_1) \,.\, (1 - F(x_2 - x_1)) \,.\, (1 - F(x_1 - t))\,dx_1. \end{cases} \tag{4.12}$$

In this case, the solution of (4.10) always exists, is unique and is given by (see [18]):

$$\hat{v}(s) = f(s) + \sum_{m=1}^{\infty} \lambda^m \int_0^s K_m(s,t)f(t)\,dt. \tag{4.13}$$

The results of Section 3.2.1. yield the following: when $\rho G(\infty) < 1$, we necessarily have $V(0) > 0$ and $\int_0^\infty dV(s) + V(0) = 1$ (Equation (4.8)), so that the unknown constant $V(0)$ is

$$V(0) = \left(1 + \int_0^\infty \hat{v}(s)\,ds\right)^{-1}. \tag{4.14}$$

4.3.3. *Resolution for $M/GI/1 + Er$.* The following results are directly obtained from the preceding section. The $M/G/1$ queue with Erlang (N, γ) impatience distribution function always has a steady state and the Laplace–Stieltjes transform of the stationary virtual offered waiting time distribution function is given

by:

$$\psi(s) = V(0)\left[1 + a(s)\sum_{l=1}^{\infty} R_0^l(s)\right]$$

$$\mathrm{Re}\,(s) \geqq 0$$

$$\text{where} \quad R_j^1 = \frac{(-\gamma)^j}{j!}$$

(4.15)
$$R_j^m(s) = \sum_{K=[j+1-N]^+}^{(m-1)(N-1)} R_K^{m-1} M_{K,j}^{m-1}(s) \qquad m \geqq 2$$

$$M_{K,j}^m(s) = \sum_{i=[j+1-N]^+}^{\min(K,\,j)} C_K^i a^{(K-i)}(s + m\gamma)\frac{(-\gamma)^{j-i}}{(j-i)!}$$

$a^{(l)}(r)$ denotes the lth derivative of $a(s)$ at point r, $\mathrm{Re}\,(r) \geqq 0$.

$$V(0) = \left[1 + \rho\sum_{l=1}^{\infty} R_0^l(0)\right]^{-1}.$$

The assertions concerning stability are obtained from the results of Section 4.2 and from $G(\infty)=0$. The series are obtained either by direct transformation of (4.13) or by self-iteration of Equation (4.5). The convergence of the series is assured because (2.6) is satisfied.

4.3.4. *Resolution for $M/GI/1+M$.* The $M/G/1$ queue with exponentially distributed impatience of parameter γ always has a steady distribution and the Laplace–Stieltjes transform of the stationary distribution function is given by:

(4.16)
$$\begin{cases} \psi(s) = V(0)\left[1 + a(s)\sum_{i=j}^{\infty} b_i(s)\right] \\ \mathrm{Re}\,(s) \geqq 0 \\ \text{with } b_i(s) = \prod_{j=1}^{i} a(s + j\gamma). \\ \text{Furthermore} \\ V(0) = \left[1 + \rho\sum_{i=1}^{\infty} b_i(0)\right]^{-1}. \end{cases}$$

5. Conclusion

We summarize the results we have obtained for the $GI/GI/1+GI$ queue:
– Conditions ensuring both the ergodicity of the Markov chain of waiting times and the integrability of the busy period have been given.
– The $GI/GI/1$ relationship between the distribution functions of actual and virtual waiting times has been continued to the case with impatience, providing

a relation of practical interest between the probability of rejection and the probability of an empty system.

– In the case of a Poisson input, an integral equation has been established for the waiting-time distribution function. This equation has been shown to be a Volterra equation and solved in terms of series when the impatience law has a rational Laplace transform.

References

[1] BACCELLI, F. AND HEBUTERNE, G. (1981) On queues with impatient customers. In *Performance 81*, ed. F. J. Kylstra. Elsevier, Amsterdam.

[2] BARRER, D. Y. (1957) Queueing with impatient customers and ordered service. *Operat. Res.* **5,** 650–656.

[3] BOROVKOV, A. A. (1976) *Stochastic Processes in Queueing Theory*. Springer-Verlag, Berlin.

[4] BOYER, P. AND HEBUTERNE, G. (1983) Relations de conservation pour une file d'attente avec clients impatients. *Ann. Telecomm.* **38** (5–6), 226–230.

[5] CALLAHAN, J. R. (1973) A queue with waiting time dependent service time. *Naval Res. Logist. Quart.* **20,** 321–324.

[6] CHARLOT, F. AND PUJOLLE, G. (1978) Recurrence in single server queues with impatient customers. *Ann. Inst. H. Poincaré* B**14,** 399–410.

[7] ÇINLAR, E. (1975) *Introduction to Stochastic Processes*. Prentice-Hall, Englewood Cliffs, NJ.

[8] COHEN, J. W. (1969) Single server queues with restricted accessibility. *J. Engineering Math.* **3,** 265–284.

[9] COHEN, J. W. (1976) *On Regenerative Processes in Queueing Theory*. Springer-Verlag, Berlin.

[10] DALEY, D. J. (1964) Single server queueing systems with uniformly limited queueing time. *J. Austral. Math. Soc.* **4,** 489–505.

[11] FELLER, W. (1971) *An Introduction to Probability Theory and its Applications*, Vol. 2. Wiley, New York.

[12] GAVISH, B. AND SCHWEITZER, P. J. (1977) The Markovian queue with bounded waiting time. *Management Sci.* **23,** 1349–1357.

[13] GNEDENKO, B. V. AND KOVALENKO, I. N. (1968) *Introduction to Queueing Theory*. Israel Program for Scientific Translations, Jerusalem.

[14] HAUGEN, R. AND SKOGAN, E. (1980) Queueing systems with stochastic time out. *IEEE Trans. Comm.* **28,** 1984–1989.

[15] HOKSTAD, P. A. (1979) Single server queue with constant service time and restricted accessibility. *Management Sci.* **25,** 205–208.

[16] LASLETT, G. M., POLLARD, D. B. AND TWEEDIE, R. L. (1978) Techniques for establishing ergodic and recurrence properties of continuous valued Markov chains. *Naval Res. Logist. Quart.* **25,** 455–472.

[17] LOÈVE, M. (1977) *Probability Theory*, 4th edn. Springer-Verlag, Berlin.

[18] MIKHLIN, S. G. (1957) *Integral Equations*. Pergamon Press, Oxford.

[19] POLLACZEK, F. (1962) Sur une théorie unifiée des problèmes stochastiques soulevés par l'encombrement d'un faisceau parfait de lignes téléphoniques. *C.R. Acad. Sci. Paris* A **254,** 3965–3967.

[20] STIDHAM, S. (1972) Regenerative processes in the theory of queues. *Adv. Appl. Prob.* **4,** 532–577.

[21] TAKÁCS, L. (1962) *Introduction to the Theory of Queues*. Oxford University Press, New York.

[22] TAKÁCS, L. (1974) A single server queue with limited virtual waiting time. *J. Appl. Prob.* **11,** 612–617.

[23] TWEEDIE, R. L. (1976) Criteria for classifying general Markov chains. *Adv. Appl. Prob.* **8,** 737–771.

[24] TWEEDIE, R. L. (1977) Hitting times of Markov chains with application to state-dependent queues. *Bull. Austral. Math. Soc.* **17,** 97–107.

A bibliography on related topics may be found in *Fundamentals of Queueing Theory* by D. Gross and M. T. Harris (Wiley, New York, 1974).