On queues with impatient customers

François Baccelli, G. Hebuterne

INRIA

CENTRE DE ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél. 954 90 20

Rapports de Recherche

N° 94

ON QUEUES WITH IMPATIENT CUSTOMERS

François BACCELLI
Gérard HEBUTERNE

Septembre 1981

ON QUEUES WITH IMPATIENT CUSTOMERS

François BACCELLI
INRIA
Domaine de Voluceau
BP 105 - Rocquencourt
78153 Le Chesnay, France

Gérard HEBUTERNE
CNET
Route de Trégastel
22301 Lannion,
France

Résumé

On considère les systèmes de files d'attente dans lesquels les clients, sujets à un phénomène d'impatience, peuvent éventuellement quitter le système lors de leur période d'attente. Les conditions de stabilité d'un tel système sont établies dans le cas GI/GI/1 pour une loi d'impatience de répartition générale. Dans le cas particulier où les arrivées forment un processus de Poisson on caractérise analytiquement la distribution stationnaire du temps d'attente virtuel ; on montre aussi que cette dernière est égale à la distribution stationnaire du temps d'attente actuel. On analyse enfin un cas de système multiserveurs (avec une loi d'impatience de répartition générale) dont l'intérêt pratique est illustré par l'évaluation des performances d'un système de télécommunications.

Abstract

This paper is concerned with the analysis of queueing systems in which customers may leave due to impatience. In the single server case, we derive the stability condition of a GI/GI/1 queue when the impatience has an arbitrary distribution function (d.f.). For the case of Poisson arrivals, we also determine analytically the stationary d.f. of the virtual waiting time which is shown to coincide with the stationary d.f. of the actual waiting time. Concerning multi server queues, we analyze some special cases [involving impatience with a general d.f.] of practical interest for modelling telecommunication systems.

ON QUEUES WITH IMPATIENT CUSTOMERS

I. INTRODUCTION

In most studies of queueing systems, little attention is paid to practical limitations such as finiteness of queue length (limited buffer capacities) or finiteness of waiting times (times out, or limited patience). However such a phenomenon is 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 its 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 on its waiting time

Systems with limited waiting times can be classified as follows :

- the limitation acts only on waiting time or only on sojourn time (waiting + service) ;
- 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 immediatly if he knows that his total sojourn time is above his patience (in such a system, all server work is useful). Ergodicity conditions for general single server queues are given in [ChP 79]. Some special cases are solved in [Ga 77], [Ho 79].

b) limitation on sojourn time, unaware customers :

This is be 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 not be useful. Some special cases can be found in [Da 64], [Co 69], [Ta 74].

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 c)) do not influence the system and can be discarded on (arrival (as in d)). 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 of 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 section 2.1).

The present paper is devoted to the analytical caracterization of waiting times in system c) (and hence d)). For this, we use the notation G/G/m+G : the three first symbols have the same meaning as in Kendall's notation. 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 F. Pollaczek in [Po 62], who reduces the problem to the resolution of a set of (unsolved) integral equations. Our contribution, concerning these general queues, consists in determinating the condition assuring stability, by means of probabilistic methods. In section 3 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 ([Ba 57]) on M/M/1+D and Gnedenko ([GnK 68]) on M/M/m+D and M/M/m+M. Multi server queues with general impatience distribution function are considered in section 4. Section 5 contains general relations (probability of rejection...) as well as mean values for waiting times. Lastly an application involving the evaluation of some major features of a telecommunication system is described in section 6.

II. ON GI/GI/1+GI QUEUES

II.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 n-th customer ($T_0=o$). We define :

$t_n \triangleq T_n - T_{n-1}$: The n-th inter arrival time ($t_n \in \mathbb{R}^+$)

s_n : the service time of the n-th customer ($s_n \in \mathbb{R}^+$)

g_n : the patience time of the n-th customer ($g_n \in \mathbb{R}^+$).

Let w_n, $n \in \mathbb{N}$ be the work load just before T_n (unfinished work). We assume the system to be of type a) of section I : the n-th customer enters the system only if the time to wait for accessing the server does not exceed his own patience. That is :

If $g_n \leq w_n$ the n-th customer is impatient and does not enter

If $g_n > w_n$ the n-th customer stays in queue

Remark : For system c), where all customers enter the queue, we ought to say : If $g_n \leq w_n$, the n-th customer does not modify w_n, the work load of the server.

If $g_n > w_n$, the n^{th} 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 : $\{t_n, n\in N\}$ (resp. $\{s_n, n\in N\}$) is a sequence of independant and identicaly distributed random variables on $\mathbb{R}^+$ with distribution function A (x) (resp. 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. C (x) may be defective (ie, we may have $\lim_{x\to\infty} C(x) \neq 1$) but we assume that $C(0)=0$.

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

II.2. Recursive equations for the offered waiting times

We derive now a recursive equation for the sequence $\{w_n, n \in \mathbb{N}\}$ generalizing Lindley's equation ([Bo 76]). Notice that w_n is the time that the n-th customer would have to wait for accessing the server if he were sufficently patient. Hence, we call it the actual offered waiting time. Let $w_0 \in \mathbb{R}^+$ be some initial condition, we have for $n \geq 0$:

$$(2.1) \qquad \begin{cases} w_{n+1} = [w_n + s_n - t_{n+1}]^+ & \text{if } y_n > w_n \\ w_{n+1} = [w_n - t_{n+1}] & \text{otherwise} \end{cases}$$

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

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

given by :

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

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 :

$$(2.3) \qquad \begin{cases} W_{n+1}(x) = \int_0^\infty G(u)\, D(x-u)\, dW_n(u) \\ + \int_0^\infty (1 - G(u))(1 - A(u-x))dW_n(u), \quad x \in \mathbb{R}^+ \\ \text{where } D(y) \triangleq P\,[s_n - t_{n+1} \leq y] = \int_0^\infty B(t+y)dA(t) \end{cases}$$

II.3. Stability condition

II.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 [LPT 78]. Let :

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

Lemma 1 :

Assume $b - a < 0$. Then the Markov chain $\{w_n \quad n \in \mathbb{N}\}$ 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

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

when comparing with (2.1), we get :

(2.4) $$w_n \geq z_n \qquad \forall n \in \mathbb{N}$$

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

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

Let $x \in \mathbb{R}^+$ and $k = [\frac{x}{b-a}]$ (where $[y]$, $y \in \mathbb{R}^+$ denotes the smallest integer greater than y) consider the event :

$$E = \bigcap_{0 \leq i \leq k} \{b-a \leq s_i - t_{i+1} \leq b-a+\varepsilon\}$$

We have :

$$E \subset \{z_k = 0 |\ z_0 = x\}$$

Hence :

(2.5) $$P[z_k = 0 \mid z_0 = x] > p^k > 0$$

From (2.4) and (2.5), we derive completing the proof :

$$P[\bigcup_{n \geq 0} \{w_n = 0 \mid w_0 = x\}] > 0$$

Lemma 2 :

Assume $b-a < 0$. Let $\rho = \frac{\lambda}{\mu}$. When $0 < 1-\rho G(\infty)$ the Markov chain $\{w_n,\ n \in \mathbb{N}\}$ is ergodic.

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 set is a bounded function, so that the chain is proved to be ergodic (see [LPT 78]).

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 :

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

(see theorem 3.2. in [LPT 78]). Consider the sequence $\{z_n, n \in \mathbb{N}\}$ defined in the proof of lemma 1. $\forall y \in B$, we have :

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

Thus any $N \ge \frac{\beta}{a-b}$ matches.

Second part of the proof

Let T_B be the hitting time of B :

$$T_B = \inf \{n \ge 0 \,/\, w_n \in B\}$$

We have to show that :

$$\sup_{x \in B} E[T_B \mid w_0 = x] < \infty \tag{2.6}$$

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

$$E\,[w_1 \mid w_0 = x] \le x - \varepsilon \quad \forall x \in B^c \tag{2.7}$$

$$E\,[w_1 \mid w_0 = x] \le M \qquad \forall x \in B \tag{2.8}$$

(see th-2.2 in [LPT 78]). We first derive from (2.2) :

$$E[w_1 \mid w_0 = x] = G(x) \int_{\mathbb{R}^+ \times \mathbb{R}^+} [x + s - t]^+ \; dA(t)\, dF(s)$$

$$+ (1 - G(x)) \int_{\mathbb{R}^+} [x-t]^+ \, dA(t)$$

$$= \int_0^x A(t)dt + g(x) \int_x^\infty (1 - F(t-x))\, A(t)dt$$

Hence for $x \le \beta$, (2.8) is satisfied since

$$E\,[w_1 \mid w_0 = x] \le x + \int_0^\infty (1 - F(u))du \le \beta + \frac{1}{\mu}$$

Concerning (2.7), we write :

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

$$\leq x - \frac{1}{\lambda} + \int_x^\infty (1 - A(t))dt - G(x) \cdot \frac{1}{\mu}$$

Thus, $\forall \varepsilon > 0 \; \exists x_0 \,/\, \forall x > x_0$

$$E[w_1 \mid w_0 = x] \leq x - \frac{1 - \rho G(x)}{\lambda} + \varepsilon$$

Assume now that the condition $0 < 1 - \rho G(\infty)$ is fulfilled. Then there exists $x_1 \in \mathbb{R}^+$ such that for $x > x_1$, $1 - \rho G(x) > 2\varepsilon$. Hence, if $\beta > \max(x_0, x_1)$ $[0,\beta]$ is a test set with bounded mean hitting time, completing the proof.

II.3.2. Necessary condition

Lemma 3

If w_n is ergodic, then $0 \leq 1 - \rho G(\infty)$

Proof : From equation (2.1) we get :

$$w_{n+1} \geq w_n + s_n\, 1_{(y_n = \infty)} - t_{n+1}, \quad n \geq 0$$

Hence :

$$(2.9) \qquad \begin{cases} w_{n+1} \geq w_0 + \sum_{i=0}^{n} b_i \\ \text{where } b_i \triangleq s_i\, 1_{(g_i = \infty)} - t_{i+1} \end{cases}$$

If $0 > 1 - \rho G(\infty)$, $E[b_i] > 0$. Thus the R.H.S. converges a.s. to infinity (strong law of large numbers) so that w_n is not ergodic.

□

III. ON M/GI/1+GI QUEUES

Throughout this section, we assume that

$$A(x) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x \leq 0 \end{cases}$$

III.1. Functional equation for the work load distribution function

We define the virtual offered waiting time (v.o.w.t.) process for M/GI/1+GI. An integral equation generalizing Takacs equation ([Ta 62]) is then established. The v.o.w.t. at time t, $\eta(t)$, is the time a test customer of infinite patience would have to wait before service if entering the queue at time t. $\eta(t)$ will be also the "unfinished work" of the server, and will only be modified by successful customers. Let $V(t,x)$ be the distribution function of $\eta(t)$ and $\psi(t,s)$ the Laplace stieljes transform (L.S.) of $V(t,x)$:

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

$$\psi(t,s) = \int_{0^-}^{\infty} e^{-sx}\, d_x V(t,x) \quad t \in \mathbb{R}^+, \quad s \in \mathbb{C},\ R_e(s) \geq 0$$

We proceed now as for Takacs equation, outlining the most important steps. Using the Markovian property of $\eta(t)$, we get :

$$(3.1) \qquad V(t+\Delta,x) = V(t,\ x+\Delta) + \lambda\Delta \int_0^{x+\Delta} G(u)(1 - F(x-u))\, D_u V(t,u) + 0\,(\Delta)$$

We multiply both sides by e^{-su} and sum for $x \in [0,\infty]$. Using the relation :

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

(3.1) becomes :

$$\begin{cases} \frac{1}{s}\,\psi(t+\Delta,s) = e^{s\Delta}\left(\frac{\psi(t,s)}{s} - \Delta V(t,\Delta)\right) + o(\Delta) - \\ - \lambda\Delta \int_{0^-}^{\infty} dx\, e^{-sx} \int_0^x G(u)\,[1 - F(x,u)]\, dV(t,u) \end{cases}$$

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

$$(3.2) \qquad \psi_G(t,s) = \int_{0^-}^{\infty} e^{-su}\, G(x)\, d_x V(t,x)$$

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

$$(3.3) \qquad \begin{cases} \frac{1}{s}\frac{\partial\psi}{\partial t} = \psi(t,s) - V(t,0) - a(s)\,\psi_G(t,s) \\ \text{where } \begin{cases} a(s) \triangleq \lambda\,\frac{1-B^*(s)}{s} \\ B^*(s) \text{ is the L.S. transform of B.} \end{cases} \end{cases}$$

III.2. A necessary and sufficient condition for the complete convergence of the v.o.w.t. process

III.2.1. Necessary part

Assume the existence of a limit

(3.4) $$V(x) = \lim_{t\to\infty} V(t,x)$$

$\psi(s)$, the L.S. transform of $V(x)$ will be solution of :

(3.5) $$\psi(s) = V(0) + a(s)\ \psi_G(s),\ R\,e(s) \geq 0$$

with :

$$a(s) = \lambda\ \frac{1-B^*(s)}{s}$$

We restrict s to take real values. In this case ψ and ψ_G are real, and :

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

or :

(3.6) $$a(s)\ \psi(s)\ G(\infty) \leq \psi(s) - V(0) \leq a(s)\ \psi(s)$$

The first inequality of (3.6) implies $V(0) > 0$ (if not, as $a(s) < 1$ for some $s > s_0$, it would mean $\psi(s) = 0$ for $s > s_0$ and thus for all s). The second inequality gives :

$$0 < V(0) \leq \psi(s)\ [1 - a(s)\ G(\infty)]$$

We must have $0 < \psi(s) \leq 1$ and thus for all s : $1 - a(s)\ G(\infty) > 0$.

For $s \to 0$, $a(s) \to \rho$ and the following inequality must then hold :

(3.7) $$1 - G(\infty) > 0$$

Note that the inequalities (3.6) also imply the uniqueness of the solution of (3.5)

III.2.2. Sufficient part

The discrete time Markov chain $\{w_n,\ n \in \mathbb{N}\}$ is imbedded in the continuous time Markov process $\{\eta(t),\ t \in \mathbb{R}^+\}$:

(3.8) $$w_n = \eta(T_n^-)$$

We use this property to derive the complete convergence of $\eta(t)$ as a consequence of the limit theorems on semi-regenerative processes. Let :

$$\begin{cases} K_t(x,B) = P[\eta(t) \in B,\ T_1 > t\ |\eta(o^+) = x],\quad B \in \mathcal{B}(\mathbb{R}^+) \\ \qquad\qquad = P[\eta(t) \in B|\ T_1 > t,\ \eta(o^+) = x]\ .\ P[T_1 > t\ |\eta(o^+) = x] \\ \qquad\qquad = \delta_{(x,t)_+}\ (B)\ e^{-\lambda t} \end{cases}$$

where :

$$\delta_u(B) = 1 \text{ if } u \in B \qquad u \in \mathbb{R}^+$$
$$\quad\; 0 \text{ if } u \notin B, \qquad B \in \mathcal{B}(\mathbb{R}^+)$$

When $\rho G(\infty) < 1$, w_n is an ergodic Markov chain. Therefore, if (3.7) holds, there exists a non defective distribution function $W(x)$ on $\mathbb{R}^+$ such that :

$$\begin{cases} \exists \lim_{n\to\infty} W_n(x) \triangleq W(x) \qquad \forall x \in \mathbb{R}^+ \\ \text{with } \lim_{x\to\infty} W(x) = 1 \end{cases}$$

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

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

The limit theorem on semi-regenerative processes yields :

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

i.e. :

$$(3.9) \qquad V(x) \begin{cases} = \lambda \int_0^\infty (1 - A(t))\, J(t+x)\, dt \\ = \lambda \int_0^\infty (1 - A(t))\, [W(t+x) - \int_0^{t+x} G(u)(1 - B(x+t-u))\, dW(u)]\, dt. \end{cases}$$

To prove the complete convergence of $\eta(t)$, we remark that because $W(x)$ is non defective, $V(x)$ is also a proper distribution.

III.3. <u>On the stationary distributions of actual and virtual offered waiting time</u>

In this section, we extend Khinchin's theorem in proving analytically that the stationary distribution fonctions of w_n and $\eta(t)$, W and V, coincide in M/GI/1+GI also. If (3.7) is fulfilled, w_n is ergodic. Hence $W(x) \triangleq \lim_{n\to\infty} W_n(x)$ is the unique solution of the invariant measure equation derived from (2.3) :

$$(3.10) \qquad \begin{cases} W(x) = \int_0^\infty G(u)\, D(x-u)\, dW(u) \\ \qquad\quad + \int_0^\infty (1 - G(u))(e^{-\lambda(u-x)_+})\, dW(u) \\ \text{with } D(y) = \lambda \int_0^\infty B(t+y) e^{-\lambda t} \end{cases}$$

Furthermore, we get from the semi-regenerative approach (equation (3.9) in which we take $A(x) = 1 - e^{-\lambda x})$:

$$V(x) = \lambda \int_x^\infty e^{-\lambda(u-x)} W(u)du - \lambda \int_x^\infty e^{-\lambda(u-x)} \int_0^u G(t)(1 - B(u-t))dW(t) \qquad (3.11)$$

Some inversions in the right hand side of (3.10) show that the right hand sides of equations (3.10) and (3.11) coincide, completing the proof.

III.4. Resolution of the functional equation

III.4.1. The density of the stationary v.o.w.t. distribution function

Let us assume the existence of a stationary solution. To prove the existence of a probability density function for the v.o.w.t. distribution function, we use the following lemma (see for instance [Fe 71]).

Lemma 4

For ψ to be of the form

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

it is sufficient and necessary that :

$$0 \le \frac{(-s)^n}{n!} \psi^{(n)}(s) \le \frac{A}{s} \qquad (3.12)$$

for all $s > 0$ and all n, where $\psi^{(n)}(s)$ denotes the n-th derivative of $\psi(s)$.

We apply the criterion to $\psi(s) - V(0)$. Let $\phi(s)$ be the L.S. transform of $V(x) - V(o)$: we have :

$$\phi(s) = \psi(s) - V(0) = a(s)\, \psi_G(s) \text{ from (3.5).} \qquad (3.13)$$

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

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

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 :

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

and so :

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

satisfies (3.12) for $n = 0$. For $n \ge 1$, (3.13) above gives :

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

a(s) is the L.S. transform of a density :

$$0 \le (-1)^{\ell} a^{(\ell)}(s) \le \frac{D\ell!}{s^{\ell+1}}$$

Therefore :

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

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

From the definition of ψ_G, we write :

$$\delta \triangleq \sum_{j=0}^{n} \frac{(-s)^j}{j!} \psi_G^{(j)}(s) = \sum_{j=0}^{n} \int_0^{\infty} \frac{(sx)^j}{j!} e^{-sx} G(x) dV(x)$$

$$\le \int_0^{\infty} [\sum_{j=0}^{n} \frac{(sx)^j}{j!}] e^{-sx} G(x) dV(x)$$

and finally :

$$\delta \le \int_0^{\infty} G(x) dV(x) \le 1$$

which puts $\phi^{(n)}(s)$ above in the desired form (it is easy to verify directly the positiveness. This completes the proof. So, when it exists, the limiting distribution function may be written as :

$$\begin{cases} \psi(s) = V(0) + \int_0^{\infty} v(u) e^{-su} du \\ 0 \le v(u) \le D \end{cases}$$

Hence V(x) is composed of an absolutely continuous part and a mass at the origin. In this case (3.5) may be inverted as follows (with v(x) the unknown density function) :

$$v(x) = \lambda V(0) [1 - B(x)] + \int_0^{x} v(u) G(u) [1 - B(x-u)] du$$

(3.14)

$$V(0) + \int^{\infty} v(x) dx = 1$$

III.4.2. <u>Resolution of the functional equation for M/GI/1+GI</u>

In this section we derive the general solution of equation (3.14) when (3.7) is satisfied so that V(0) > 0. (3.14) 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-Stieljes transforms.

Consider the following functions :

(3.15)

$$\begin{cases} f(s) = \lambda[1 - F(s)] \\ K(s,t) = G(t) [1 - B(s-t)] \quad \begin{matrix} s \ge 0 \\ t \ge 0 \end{matrix} \\ \hat{v}(s) = v(s)/V(0). \end{cases}$$

From equation (3.14), $\tilde{V}(s)$ is solution of :

(3.16)
$$\begin{cases} \tilde{V}(s) = f(s) + \lambda \int_0^s J(s,t)\, \tilde{V}(t)dt \\ s \geq 0 \end{cases}$$

Due to our assumptions concerning the existence of the first moment of F, f(s) and K(s,t) are both square integrable functions ; thus (3.16) is a Volterra equation for which the method of the resolvent applies. Let :

(3.17)
$$\begin{cases} K_m(s,t) = \int_t^s K(s,x)\, K_{m-1}(x,t)dx \\ m \geq 2 \\ K_1(s,t) = K(s,t) \qquad s \geq 0, \quad t \geq 0 \end{cases}$$

An induction yields the following expression :

(3.18)
$$\begin{cases} K_m(s,t) = G(t) \int_t^s G(x_{m-1}) \,.\, (1 - F(s-x_{m-1})). \\ dx_{m-1} \int_t^{x_{m-1}} G(x_{m-2}) \,.\, (1 - F(x_{m-1} - x_{m-2})). \\ dx_{m-2} \int_t^{x_{m-2}} \ldots \\ . \\ . \\ . \\ dx_2 \int_t^{x_2} G(x_1) \,.\, (1 - F(x_2 - x_1)) \,.\, (1 - F(x_1-t))dx_1. \end{cases}$$

In this case, the solution of (3.16) always exists, is unique and is given by (see [Mi 57]) :

(3.19)
$$\tilde{V}(s)_{s \to 0} = f(s) + \sum_{m=1}^{\infty} \lambda^m \int_0^s K_m(s,t)\, f(t)dt$$

The results of section III.2.1. yield the following : when $\rho G(\infty) < 1$: we have necessarily $V(0) > 0$ and $\int_0^\infty v(s) + V(0) = 1$ (equation (3.14)), so that the unknown constant $V(0)$ is

(3.20)
$$V(0) = (1 + \int_0^\infty \tilde{V}(s)\, ds)^{-1}$$

III.4.3. 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 L.S. transform of the stationary v.o.w.t. distribution function is given by :

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

$$\mathrm{Re}(s) \geq 0$$

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

(3.21)
$$R_j^m(s) = \sum_{k=[j+1-N]^+}^{(m-1).(N-1)} R_k^{m-1} M_{k,j}^{m-1}(s) \quad m \geq 2$$

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

$a^{(\ell)}(r)$ denotes the ℓ-th derivative of $a(s)$ at point r, $Re(r) \geq 0$.

$$V(0) = [1 + \rho \sum_{\ell=1}^{\infty} R_0^{\ell}(0)]^{-1}$$

The assertions concerning stability are obtained from the results of section III.2 and from $G(\infty) = 0$. The series are obtained either by direct transformation of (3.19) or by self-iteration of equation (3.5). The convergence of the series is assured because (3.7) is satisfied.

III.4.4. Resolution for M/GI/1+M

The M/G/1 queue with exponentially distributed impatience of parameter γ always has a steady and the L.S. transform of the stationary distribution function is given by :

(3.22)
$$\begin{cases} \psi(s) = V(0) \, [1 + a(s) \sum_{i=j}^{\infty} b_i(s)] \\ Re(s) \geq 0 \\ \text{with } b_i(s) = \prod_{j=1}^{i} a(s + j\gamma) \\ \text{Furthermore :} \\ V(0) = [1 + \rho \sum_{i=1}^{\infty} b_i(0)]^{-1} \end{cases}$$

□

IV. ON M/M/m+GI QUEUES

IV.1. Introduction

For the m-servers case, we consider exponential service times and Poisson arrivals ($B(x) = 1 - e^{-\mu x}$, $A(x) = 1 - e^{-\lambda x}$). The system is defined as in section 2. Consider the following process : $\{N(t),\ t \in \mathbb{R}^+\}$ is equal to n when the number of customers in the system at time t is n and $0 \leq n \leq m-1$. $N(t)$ is equal to L when the number of customers at time t is greater than m-1. The v.o.w.t. $\eta(t)$, is equal to zero when $N(t) \neq L$ and is strictly positive otherwise. Clearly, $\{(N(t), \eta(t)),\ t \in \mathbb{R}^+\}$ is a Markov process with state space $[\{0\}, \{1\}, \ldots, \{m-1\}, L] \times \mathbb{R}^+$. Consider the following functions (when they exist) :

(4.1)
$$\begin{cases} \underset{x \geq 0}{v(x)} = \lim\limits_{t \to \infty} \lim\limits_{dx \to 0} P[N(t) = L,\ x < \eta(t) \leq x + dx] \\ \underset{0 \leq j \leq m-1}{P_j} = \lim\limits_{t \to \infty} P[N(t) = j,\ \eta(t) = 0]. \end{cases}$$

Kolmogorov's equations for (N, η) at steady state yield the following relations between these functions :

(4.2)
$$\begin{cases} \lambda P_0 = \mu P_1 \\ (\lambda + \mu j)P_j = \lambda P_{j-1} + (j+1)\mu P_{j+1} \qquad 0 < j < m-1 \\ v(0) = (\lambda + (m-1)\mu)P_{m-1} - \lambda P_{m-2} \\ \underset{x>0}{v(x)} = \lambda P_{m-1}\, e^{-m\mu x} + \lambda \int_0^x G(u)\, v(u)\, e^{-m\mu(x-u)} du \end{cases}$$

We obtain from this :

(4.3)
$$\begin{cases} P_j = \left(\frac{\lambda}{\mu}\right)^j \frac{1}{j!} P_0 \qquad j = 0, \ldots, m-1 \\ v(0) = \lambda P_{m-1} \end{cases}$$

Furthermore, $H(x) = e^{m\mu x} v(x)$ is the solution of the following equation :

$$\underset{x>0}{H(x)} = \lambda P_{m-1} + \lambda \int_0^x G(u)\, H(u)\, du$$

which yields :

(4.4)
$$v(x) = \lambda P_{m-1} \exp\{\lambda \int_0^x G(u)du - m\mu x\}.$$

The normalizing condition is :

(4.5)
$$\sum_{j=0}^{m-1} P_j + \int_0^\infty v(x)dx = 1.$$

That is :

(4.6)
$$\begin{cases} P_0 = [1 + \frac{\lambda}{\mu} + \ldots + \left(\frac{\lambda}{\mu}\right)^{m-1} \frac{1}{(m-1)!} [1 + \lambda J]]^{-1} \\ J = \int_0^\infty \exp\{\lambda \int_0^x (G(u) - \frac{m\mu}{\lambda})du\} \end{cases}$$

For stability, one can check that the normalization is possible if and only if the integral in (4.6) converges, which is equivalent to the condition :

(4.7) $$\lambda G(\infty) < m\mu$$

Lastly, the result of section III.3. can be extended to the considered multi server queue.

V. PROBABILITY OF REJECTION, MEAN VALUES

V.1. Single server queues

In this section, we derive further relations between quantities of practical interest. Let Π be the probability of rejection of customers at steady, i.e., the probability that an arriving customer decides not to enter the system. For GI/GI/1+GI queues we have :

(5.1) $$\Pi = \int_{0^-}^{\infty} (1 - G(u))dW(u)$$

where $W(x)$ is the limiting distribution of W_n which exists when $0 < 1 - \rho G(\infty)$. In the M/GI/1+GI case, we obtain from the result of section III.3. and III.4.1. :

(5.2) $$\Pi = \int_0^\infty v(x)\ [1 - G(x)]\ dx$$

From (3.5) we get, for $s = 0$:

$$1 = V(0) + \rho \int_0^\infty v(x)\ G(x)\ dx$$

Thus :

(5.3) $$(1 - \Pi)\rho = 1 - V(0)$$

It is interesting to compare this relation with the one obtained for the M/G/1 queue with a limited capacity N :

$$(1 - P_N)\rho = 1 - P_0$$

(In this system P_N is the probability of rejection).

We now derive some relations between mean waiting time and mean queue length. Little's formula applies (since the beginning of busy periods are regeneration points for all of the defined stochastic processes). Let $\bar{W}_1$ be the mean waiting time spent in the queue by patient customers and $\bar{L}_1$ be the mean number of patient customers in the queue. For the case d) (defined in the introduction), let $\bar{W}_2$ be the mean time spent in queue by all customers (those rejected after their time out, the impatient ones, and those served the patient ones), and $\bar{L}_2$ be the mean number of customers in the queue.

For GI/GI/1+GI, we have :

(5.4) $$\bar{W}_1 = \int_0^\infty x\ G(x)\ dW(x)$$

(5.5) $$\bar{W}_2 = \int_0^\infty dW(u) \int_0^u G(t)dt$$

Let γ be the first moment of $C(x)$. We get for instance $\bar{W}_2 = \bar{W}_1 + \gamma\Pi$ for GI/GI/1+D and $\bar{W}_2 = \gamma\Pi$ for GI/GI/1+M. We also obtain from Little's formula :

(5.6) $$\bar{L}_1 = \lambda(1 - \Pi)\ \bar{W}_1$$

(5.7) $$\bar{L}_2 = \lambda\bar{W}_2$$

Consider now $\bar{v}$, the mean v.o.w.t. in the M/GI/1+GI case

$$\bar{v} = \int_0^\infty x\ v(x)\ dx$$

Differentiating equation (3.5) at point s = 0 yields :

(5.8) $$\bar{v} = \rho\bar{W}_1 + \frac{\lambda}{2}(1 - \Pi)\ E[S_1^2]$$

which is a Pollaczek-Khinchin mean value formula for queues with impatient customers.

V.2. Multi server queues

For the queueing system of section IV, we get :

(5.9) $$\Pi = \left(1 - \frac{m\mu}{\lambda}\right)\left[1 - \sum_{j=0}^{m-1} P_j\right] + P_{m-1}$$

The equivalent of relation (5.8) is :

(5.10) $$\bar{v} = \frac{\rho}{m}\ \bar{W}_1 + \frac{1}{\mu m}[1 - V(0)]\text{, where } V(0) \triangleq \sum_{j=0}^{m-1} P_j$$

For M/M/m+D, we obtain the following :

(5.11) $$\begin{cases} P_0 = \left[1 + \sum_{k=0}^{m-1} \rho^k/k! + \frac{\rho^m}{(\rho-m)m!}\left(\rho e^{(\lambda-m\mu)\gamma} - m\right)\right]^{-1} \\ \Pi = \frac{\rho^m}{m!} P_0 \exp(\lambda - m\mu)\gamma \\ \bar{W}_1 = P_0 \cdot \frac{\rho^m}{m!} \cdot \frac{m/\mu}{\rho-m}\{1 + [\gamma\rho(\rho - m) - 1]e^{\lambda\gamma - m\mu\gamma}\} \\ \bar{W}_2 = \bar{W}_1 + \gamma\Pi \end{cases}$$

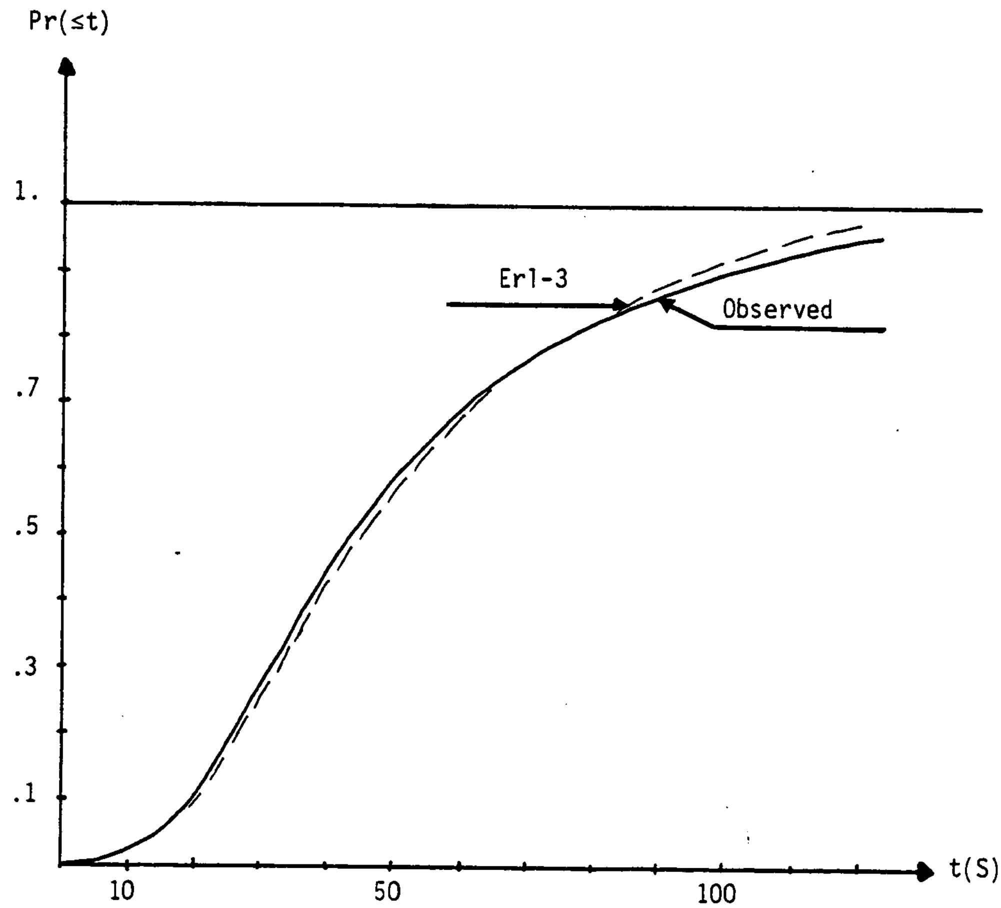
For M/M/m+M, let $\alpha = \frac{1}{m\gamma\mu}$. We have :

(5.12) $$\begin{cases} P_0 = \{1 + \rho + \dots \rho^{m-1}/(m-1)! + \frac{\rho^m}{m!}\left(1 + \frac{\rho/m}{1+\alpha} + \dots\right)\}^{-1} \\ \Pi = P_0 \frac{\rho^{m-1}}{(m-1)!}\left[1 + \left(\frac{\rho}{m} - 1\right)\left(1 + \frac{\rho/m}{1+\alpha} + \dots\right)\right] \\ \bar{W}_2 = \gamma\Pi \end{cases}$$

VI. APPLICATION TO THE MODELLING OF A TELECOMMUNICATION SYSTEM

We consider the behaviour of customers and operators in a PABX : In such a system, operators receive calls from the public network, and switch these calls to the appropriate called subscriber. Thus, from the telephone network, a PABX can be seen as a GI/GI/m queueing system, m being the number of operators. In fact, calling customers waiting in the queue can leave impatiently. Hence the real model to be used is our GI/GI/m+GI queue. The phenomenon can be of great importance, for both calling and called subscribers (these last cannot be reached, without beeing aware of it), and also for telephone administrations (the network successfully routes the demand using resources for this, but an unanswered call gives rise to no revenue).

Some measurements, performed on various PABX'S have shown that, for such a system, the impatience distribution (1-G(t)) is approximately Erlang-3 (the curve in Figure 1 may be found in [ROB 79]).



Estimated customers patience d.f., fitted with Erl-3

Figure 1

Accordingly, we take :

$$G(t) = e^{-t/\theta} \{1 + \frac{t}{\theta} + \frac{t^2}{2\theta^2}\}, \theta = 17 \text{ sec.}$$

Assuming Poisson arrivals and exponentially distributed service for the system, we can then use the results of Section 4 (eq 4.4)

$$w(x) = \lambda \frac{\rho^{m-1}}{(m-1)!} P_0 \ \text{Exp} \{\lambda \int_0^x G(u)du - m\mu x\}$$

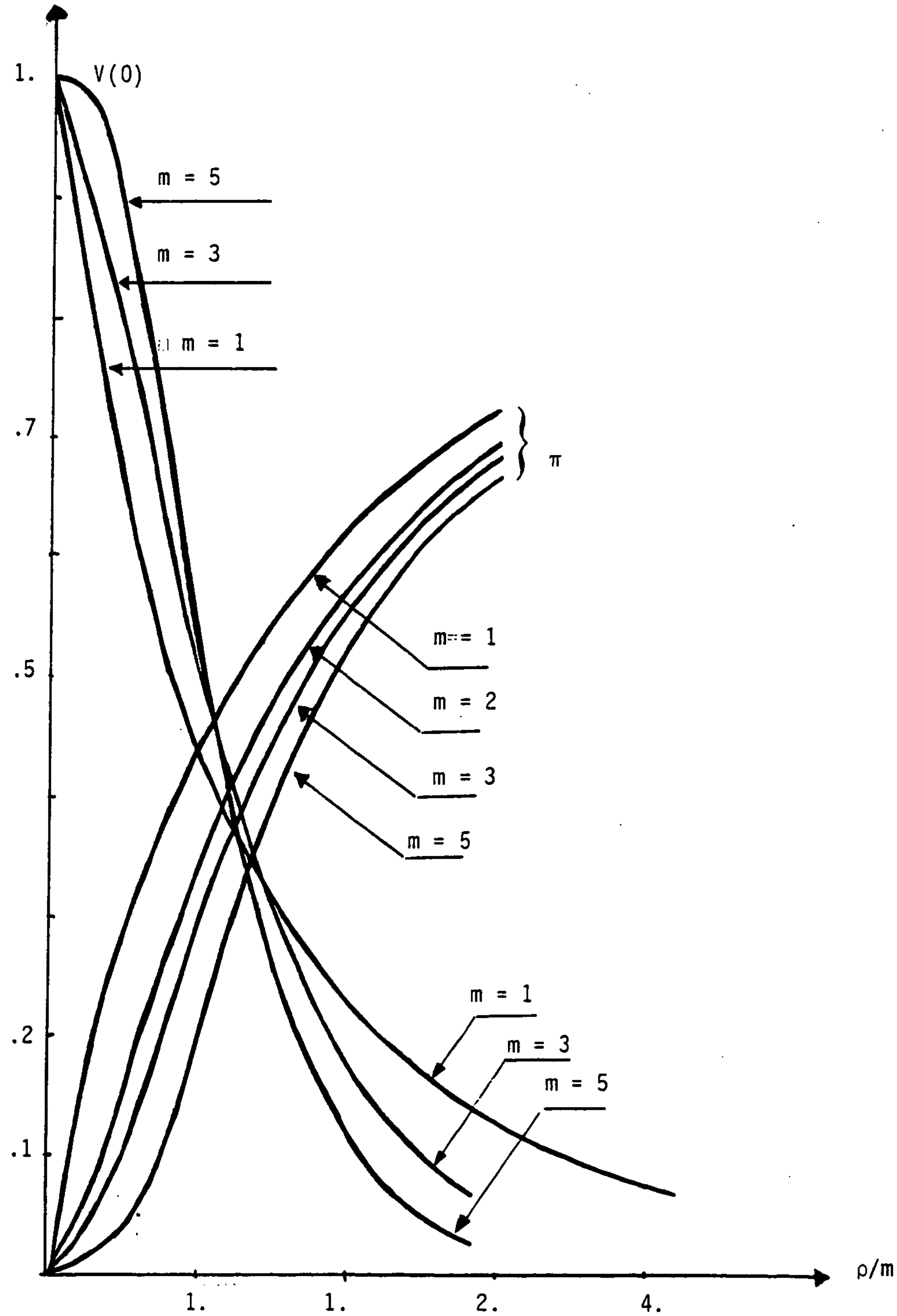
$$P_0 = [1 + \rho + \ldots \frac{\rho^{m-1}}{(m-1)!} (1 + Le^{3L} \int_0^\infty \text{Exp} \{-(3+2u+u^2/2)Le^{-u} + $$

$$+ Lmu/\rho \ du)\}]^{-1}$$

with $\rho = \lambda/\mu, L = \lambda\Theta$

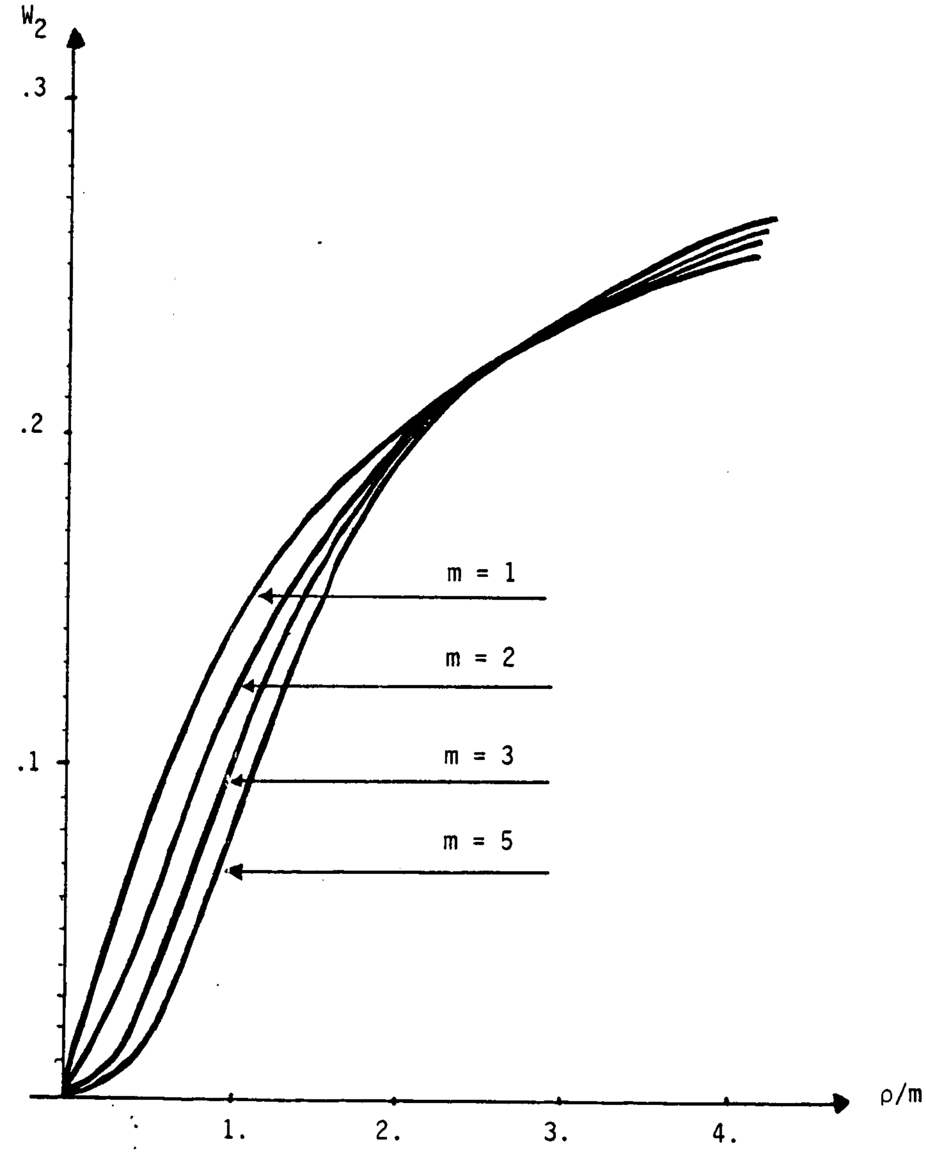
In Figures 2 and 3, we plot Π (probability of rejection), $V(0)$, probability of no waiting, and $\bar{W}_2$, mean waiting time of all customers. The latter measures the ineffective occupation of telephone lines. The curves show that a fast degradation when ρ/m approaches 1 : For $\frac{\rho}{m} = 0.8$, $m = 2$, $V(0) = 0.6$, the operators are not overloaded but $\Pi = 0.25$.

□



Evolution of Π and V(o) with ρ and m

Figure 2



Mean waiting time (in units of service time)

Figure 3

BIBLIOGRAPHY

[Ba 57] BARRER D.Y., Queueing with impatient customers and ordered service, Operations Research, Vol. 5, (1957), pp. 650-656.

[Bo 76] BOROVKOV A.A., Stochastic processes in queueing theory, Springer Verlag 1976.

[Ca 73] CALLAHAN J.R., A queue with waiting time dependant service time, Naval Research Logistics Quaterly, Vol. 20 (1973), pp. 321-324.

[ChP 79] CHARLOT F., PUJOLLE G., Recurrence in single server queues with impatient customers, Annales de l'Institut Henri Poincaré, Vol. 4, (1978), n° 4.

[Co 69] COHEN J.W., Single server queues with restricted accessibility, Journal of Engineering Mathematics, Vol. 3, (1969), pp. 265-284.

[Da 64] DALEY D.J., Single server queueing systems with uniformly limited queuing time, Jal Austral. Math. Soc., Vol. 4, (1964), pp. 489-505.

[Fe 71] FELLER W., An introduction to probability theory and its applications, Vol. II, J. Wiley and Sons, (1971).

[Ga 77] GAVISH B., SCHWEITZER P.J., The markovian queue with bounded waiting time, Management Science, Vol. 23, n° 12, Aug. 1977.

[GnK 68] GNEDENKO, KOVALENKO, Introduction to queueing theory, Israël Program for Scientific Translations, Jerusalem (1968).

[Ho 79] HOKSTAD P.A., Single server queue with constant service time and restricted accessibility, Management Science, Vol. 25, n° 2, Feb. 1979.

[LPT 78] LASLETT G.M., POLLARD D.B., TWEEDIE R.E., Techniques for establishing ergodic and recurrence properties of continuous valued Markov chains, Naval Research Logistics Quaterly, Sept. 78, Vol. 25, n° 3, pp. 455-472.

[Mi 57] MIKHLIN S.G., Integral equations, Pergamon Press (1957).

[Po 62] POLLACZEK F., Sur une théorie unifiée des problèmes stochastiques soulevés par l'encombrement d'un faisceau parfait de lignes téléphoniques, C.R. Académie des Sciences, 1962, 1er semestre (T. 254, n° 23), Paris.

[Rob 79] ROBERTS J.W., Recent observations of subscriber behaviour, 9th International Teletraffe Congress, Torremolinos 1979.

[Ta 62] TAKACS L., Introduction to the theory of queues, Oxford University Press (1962).

[Ta 74] TAKACS L., A single server queue with limited virtual waiting time, Jal Applied Prob., Vol. 11, pp.612-617, (1974).

A bibliography for related topics may be found in GROSS D., HARRIS M.T. Fundamentals of Queueing Theory, John Willey and Sons, 1974.