DYNAMIC, TRANSIENT AND STATIONARY BEHAVIOR OF THE M/GI/1 QUEUE VIA MARTINGALES

By François Baccelli¹ and Armand M. Makowski²

Institut National de Recherche en Informatique et en Automatique and University of Maryland

An exponential martingale is associated with the Markov chain of the number of customers in the M/GI/1 queue. With the help of arguments from renewal theory, this martingale provides a unified probabilistic framework for deriving several well-known generating functions for the M/GI/1 queue, such as the Pollaczek-Khintchine formula, the transient generating function of the number of customers at departure epochs and the generating function of the number of customers served in a busy period.

Introduction. An exponential martingale is associated with the Markov chain that describes the number of customers at departure epochs in the M/GI/1 queue. Basic regularity properties of this martingale and standard arguments from renewal theory are shown to provide a unified probabilistic framework for deriving three well-known analytical formulas which, respectively, characterize the dynamic, transient and stationary behavior of this queue. The main theoretical ingredient of this new approach lies in a general equivalence relationship between the law of the embedded Markov chain and the law of the forward recurrence time of a discrete-time renewal process associated with this chain (Theorem 3). This equivalence produces several results which appear to be new, at least to the best of the authors' knowledge. For instance, a new probabilistic representation is established for the generating function of the number of customers at the nth departure epoch (Corollary 6). The usual representation of the transient generating function of this quantity, in terms of its double generating function, both in time and space, is derived from this probabilistic representation.

The paper is organized as follows. The exponential martingale is introduced in Section 1 and its definition is followed by a summary of its key regularity properties, already established by the authors in a previous paper [1]. Section 2 contains the derivation of the analytical results mentioned above, with some of the calculations collected in Section 3. The generating function of the number of customers served in a busy period is obtained in Section 2.1. The generating

Received June 1988; revised November 1988.

¹Work supported partially through a grant from AT & T Bell Laboratories and partially through a grant from the Minta Martin Aeronautical Research Fund, College of Engineering, University of Maryland, College Park.

²Work supported partially through NSF Engineering Research Centers Program Grant NSFD CDR-88-03012, partially through NSF Grant ECS-83-51836 and partially through ONR Contract N00014-84-K-0614.

AMS 1980 subject classifications. 60E10, 60F05, 60G17, 60G40, 60G42, 60J05, 60K05, 60K25.

Key words and phrases. Martingales, Doob's optional sampling theorem, renewal theory, generating functions, queueing theory, Pollaczek-Khintchine formula, busy period.

function of the number of customers served in a busy period is obtained in Section 2.1. The proof of this first result is very much in the spirit of the derivation of the Laplace transform of the length of the busy period via martingales given in [7]. The full power of the equivalence relationship mentioned above becomes apparent in Sections 2.2 and 2.3, where the Pollaczek–Khintchine formula and the transient generating function of the number of customers are shown to be mere rephrasings of basic formulas from renewal theory.

It should be emphasized that the authors' aim was not to give here a comprehensive list of the applications of this new approach, but rather to illustrate its usefulness and versatility. For instance, the results on the transient generating function are limited to the analysis of a simple particular case. Similarly, only the hitting time to the zero state (i.e., the number of customers served in a busy period) was considered in the study of the dynamic properties of the chain. However, it is easy to see that most known formulas concerning the transient dynamic behavior can be obtained through similar arguments. These techniques are of independent interest and apply to other queueing systems, like, for instance, queues in random environments [2].

1. Preliminaries.

1.1. Notation. All the random variables (RV) and stochastic elements occurring in this paper are defined on some fixed underlying probability triple $(\Omega, \mathcal{F}, \mathcal{P})$. Throughout, the characteristic function of any event A in \mathcal{F} is denoted by I[A]. The collection of all nonnegative integers is denoted by \mathbb{N} and \mathbb{R} (resp. \mathbb{R}_+) denotes the set of (resp. nonnegative) real numbers.

Consider an M/GI/1 queue with Poissonian arrival pattern of intensity λ . The consecutive service times form a sequence of i.i.d. RV's independent of the arrival process. Throughout, the common probability distribution of the service times and its Laplace–Stieltjes transform are denoted by S and S^* , respectively. The initial queue size is given by an N-valued RV Ξ which is independent of both the arrival and service processes.

1.2. The embedding. At time t=0, a dummy customer is assumed to complete service and by leaving the system, generate the 0th departure. For $n=0,1,\ldots$, denote by X_n the number of customers in the system as seen by the nth departing customer and by A_{n+1} the number of arrivals during the (n+1)st service period. With these definitions, the queue size sequence $\{X_n, n=0,1,\ldots\}$ is readily seen to satisfy the recursion

$$X_{n+1} = X_n + A_{n+1} - I[X_n \neq 0], \qquad n = 0, 1, ...,$$

$$X_0 = \Xi.$$

Under the enforced assumptions, the RV's $\{A_{n+1}, n=0,1,\ldots\}$ are i.i.d. RV's independent of the initial queue size Ξ and its underlying probability generating

function a is given by

$$(1.2) a(z) = E[z^{A_n}] = S^*(\lambda(1-z)), 0 \le z \le 1, n = 0, 1, \dots$$

It is then clear from (1.1) that the N-valued process $\{X_n, n = 0, 1, ...\}$ is an *irreducible* Markov chain with countable state space [4,5].

1.3. The martingale. For all $n=0,1,\ldots$, the RV's $\{\Xi,A_k,\ 0< k\leq n\}$ generate the σ -field of events \mathscr{F}_n and set $\mathscr{F}_\infty=\bigvee_n\mathscr{F}_n$ with the standard notation. In view of (1.1) the RV's $\{X_0,\ldots,X_n\}$ are all \mathscr{F}_n -measurable. Given an arbitrary \mathscr{F}_n -stopping time σ , define the N-valued RV $\nu(\sigma)$ by

(1.3)
$$\nu(\sigma) = \begin{cases} \inf\{n \ge 1 \colon X_{\sigma+n} = 0\}, & \text{if } \sigma < \infty \text{ and this set is nonempty,} \\ \infty, & \text{otherwise.} \end{cases}$$

For $0 < z \le 1$, the \mathscr{F}_n -measurable \mathbb{R}_+ -valued RV's $\{M_n(z), n = 0, 1, \dots\}$ are defined by

(1.4)
$$M_n(z) = \begin{cases} z^{X_0}, & \text{if } n = 0, \\ z^{X_n} \frac{z^{\sum_{k=0}^{n-1} I[X_k \neq 0]}}{a(z)^n}, & \text{for } n = 1, 2, \dots. \end{cases}$$

The following result was established in [1], Theorem 2, pages 181-186.

THEOREM 1. For all 0 < z < 1, the RV's $\{M_n(z), n = 0, 1, ...\}$ are integrable and form a positive \mathcal{F}_n -martingale. If $\rho \leq 1$, the stopping time $\nu(\sigma)$ is regular for this martingale and the relation

$$(1.5) E\left[I\left[\sigma<\infty,\nu(\sigma)<\infty\right]\left[\frac{z}{a(z)}\right]^{\nu(\sigma)}\middle|\mathscr{F}_{\sigma}\right]=I\left[\sigma<\infty\right]z^{I\left[X_{\sigma}=0\right]}z^{X_{\sigma}}\quad a.s.$$

holds for all $0 < z \le 1$.

The relation (1.5) is a simple consequence of the regularity of the stopping time σ and of Doob's optional sampling theorem ([6], Corollary IV-2-6, page 67). Moreover, under the condition $\rho \leq 1$, z < a(z) for 0 < z < 1 by Takács's lemma ([8], page 46) and letting $z \uparrow 1$ in (1.5) yields

(1.6)
$$P[\sigma < \infty, \nu(\sigma) < \infty | \mathscr{F}_{\sigma}] = I[\sigma < \infty] \quad \text{a.s.}$$

as an immediate consequence of the bounded convergence theorem.

2. Transforms. Let $\{\tau_n, n = 0, 1, ...\}$ be the sequence of \mathscr{F}_n -stopping times defined by the recursion

(2.1)
$$\tau_{n+1} = \tau_n + \nu(\tau_n), \quad n = 0, 1, \dots,$$

with $\tau_0 = 0$. With $\sigma = 0$, (1.6) specializes to

(2.2)
$$P[\tau_1 < \infty | \mathscr{F}_0] = 1 \quad \text{a.s.}$$

LEMMA 2. If $\rho \leq 1$, the RV's $\{\tau_n, n = 1, 2, ...\}$ form a possibly delayed recurrent renewal process.

PROOF. It is plain from (1.5) that the RV $\nu(\tau_n)$ and the σ -field \mathscr{F}_{τ_n} are independent for all $n=0,1,\ldots$, so that the RV's $\{\tau_n,\ n=1,2,\ldots\}$ form a renewal process if $\Xi=0$ a.s. and a delayed renewal process otherwise. For $\rho\leq 1$, the recurrence property is immediately obtained from (1.6). \square

The forward recurrence times $\{\mu(n), n = 0, 1, ...\}$ of the recurrent renewal process $\{\tau_n, n = 1, 2, ...\}$ are defined by

(2.3)
$$\mu(n) = \begin{cases} \inf\{m \geq 0 \colon X_{n+m} = 0\}, & \text{if this set is nonempty,} \\ \infty, & \text{otherwise,} \end{cases}$$
 $n = 0, 1, \dots$

It turns out that the generating function of the number X_n of customers at the nth service completion is very simply related to the generating function of the forward recurrence time $\mu(n)$. The key relationship is provided in the next theorem. For every $0 < z \le 1$, it is convenient to introduce the quantity $\xi(z)$ as the ratio

(2.4)
$$\xi(z) = \frac{z}{a(z)}.$$

THEOREM 3. Assume $\rho < 1$. For all $0 < z \le 1$, the relation

(2.5)
$$E[z^{X_n}] = E[\xi(z)^{\mu(n)}], \qquad n = 0, 1, ...,$$

holds.

PROOF. Note from (1.2) and (2.3) the easy facts

$$[X_n = 0] = [\mu(n) = 0], \qquad n = 0, 1, \dots,$$

and

(2.7)
$$\nu(n) = \mu(n) \text{ on } [X_n \neq 0], \quad n = 0, 1, \dots$$

It is then plain that for each y > 0, the relations

(2.8)
$$y^{\mu(n)}I[\mu(n) \neq 0] = y^{\mu(n)}I[X_n \neq 0] = y^{\nu(n)}I[X_n \neq 0], \quad n = 0, 1, ...,$$
 hold.

Specialize (1.5) to $\sigma = n$ and multiply both ides of the resulting equation by $I[X_n \neq 0]$. For all $0 < z \le 1$, the relation

$$(2.9) E\left[I\left[\nu(n)<\infty, X_n\neq 0\right]\xi(z)^{\nu(n)}|\mathscr{F}_n\right]=I\left[X_n\neq 0\right]z^{X_n} \text{ a.s.}$$

readily follows since the RV X_n is \mathcal{F}_n -measurable and, therefore,

(2.10)
$$E\left[I[\nu(n) < \infty, X_n \neq 0]\xi(z)^{\nu(n)}\right] = E\left[I[X_n \neq 0]z^{X_n}\right]$$

after taking the mathematical expectation of both sides of (2.9).

Starting with a simple decomposition argument, it is now plain that for all $n = 0, 1, \ldots$, the relations

$$E[z^{X_n}] = P[X_n = 0] + E[z^{X_n}I[X_n \neq 0]]$$

$$= P[\mu(n) = 0] + E[I[\nu(n) < \infty, X_n \neq 0]\xi(z)^{\nu(n)}]$$

$$= P[\mu(n) = 0] + E[I[\nu(n) < \infty, \mu(n) \neq 0]\xi(z)^{\mu(n)}]$$

hold. The second equality is a consequence of (2.6) and (2.10), while the last equality follows from (2.8). Under the foregoing assumptions, $P[\nu(n) < \infty | \mathcal{F}_n] = 1$ a.s. by virtue of (1.6), whence $\mu(n) < \infty$ a.s. since $0 \le \mu(n) \le \nu(n)$ as a consequence of (2.6) and (2.7). This remark readily validates the passage from (2.11) to (2.5). \square

The aim of the remainder of the section is to recover various known transforms from (1.5) and (2.5), thus avoiding the usual analytical calculations.

2.1. Generating function of the number of customers served in a busy period. Denote by $\{f(n), n=0,1,\dots\}$ the point mass distribution function of the number of customers served in a busy period of the M/GI/1 queue. On the event $[X_0=0]$, the number of customers served in the first busy period coincides with $\nu(0)$, so that the generating function F^* of $\{f(n), n=0,1,\dots\}$ is given by the relation

(2.12)
$$F^*(y) = E[y^{\nu(0)}|X_0 = 0], \quad 0 \le y \le 1.$$

In other words, F^* is also the generating function of the interevent times of the discrete-time renewal process $\{\tau_n, n=1,2,\ldots\}$. It is now shown that F^* can be obtained as an immediate consequence of (1.5).

LEMMA 4. Assume $\rho \leq 1$. For each $0 < \xi \leq 1$, the equation in the unknown variable z,

$$(2.13) z = \xi a(z),$$

has a unique solution $Z(\xi)$ in the interval [0,1]. The generating function F^* of the number of customers served in a busy period, or equivalently of the interevent times of the discrete-time renewal process $\{\tau_n, n = 1, 2, \ldots\}$, is given by

$$(2.14) F^*(y) = Z(y), 0 \le y \le 1,$$

and the mean value m of this distribution function is given by

(2.15)
$$m = E[\nu(0)|X_0 = 0] = \begin{cases} \frac{1}{1-\rho}, & \text{if } \rho < 1, \\ \infty, & \text{if } \rho = 1. \end{cases}$$

PROOF. The first statement concerning the solutions of (2.13) follows from classical convexity arguments ([8], Lemma 1, page 47) and its proof is therefore

omitted for the sake of brevity. Specializing again (1.5) to $\sigma = 0$, it is plain that for all $0 < z \le 1$,

(2.16)
$$E\left[\xi(z)^{\nu(0)}|X_0=0\right]=z \text{ a.s.}$$

Let ξ be a real number such that $0 \le \xi \le 1$ and let $Z(\xi)$ denote the unique solution to (2.13) in [0,1]. It is now immediate from (2.16) and from the definition of $Z(\xi)$ that

(2.17)
$$F^*(\xi) = E\left[\xi^{\nu(0)}|X_0 = 0\right] = Z(\xi) \quad \text{a.s.}$$

and (2.14) is obtained. Equation (2.15) is now obtained by differentiating (2.16) with respect to z in a left neighborhood of 1 and by letting $z \uparrow 1$ in the resulting expression. The differentiation step is validated by well-known properties of uniform convergence for generating functions. \Box

2.2. Generating function of the number of customers at steady state. The next step consists in establishing the Pollaczek-Khintchine transform.

COROLLARY 5. Assume $\rho < 1$. For all $0 < z \le 1$, $\lim_n E[z^{X_n}]$ exists (when n goes to ∞) and is given by

(2.18)
$$\lim_{n} E[z^{X_{n}}] = (1 - \rho) \frac{(1 - z)a(z)}{a(z) - z}.$$

PROOF. From the key renewal theorem ([9], Theorem 2.3, page 18), the forward recurrence times $\{\mu(n), n = 0, 1, ...\}$ of a discrete-time renewal process with interevent generating function F^* and finite mean m satisfy the convergence property

(2.19)
$$\lim_{n} E[y^{\mu(n)}] = \frac{1}{m} \frac{1 - F^{*}(y)}{1 - y},$$

this independently of the initial condition. Using this fact and (2.5), it is now plain that $\lim_{n} \{E[z^{X_n}] \text{ exists, is independent of the initial condition and is given by$

(2.20)
$$\lim_{n} E[z^{X_{n}}] = \lim_{n} E[\xi(z)^{\mu(n)}] = (1 - \rho) \frac{1 - Z(\xi(z))}{1 - \xi(z)}.$$

Equation (2.18) follows immediately from (2.20) since $Z(\xi(z)) = z$. \square

2.3. Transient generating function for the number of customers. For the sake of brevity, the discussion is limited to the case of a stable queue, namely $\rho < 1$, when the initial queue size is zero, namely $\Xi = 0$ a.s. The proposed approach extends easily to the other cases by following arguments very similar to the ones reported below. Details are left to the interested reader.

The renewal kernel $R: \mathbb{N} \to \mathbb{R}$ associated with the discrete-time, nondelayed renewal process $\{\tau_n, n=1,2,\ldots\}$ is given by

(2.21)
$$R(n) = \sum_{k=0}^{\infty} P[\tau_k = n], \qquad n = 0, 1, \dots.$$

Moreover, let the function $g: [0,1] \times \mathbb{N} \to \mathbb{R}$ be defined by

$$(2.22) \quad g(y,n) = E\left[y^{\mu(n)}I\left[\tau_1 > n\right]\right] = \begin{cases} E\left[y^{\tau_1 - n}I\left[\tau_1 > n\right]\right], & \text{if } n \geq 1, \\ 1, & \text{if } n = 0. \end{cases}$$

As indicated in the next corollary, the *transient* behavior of the queue size process can be fully characterized by means of these two functions.

COROLLARY 6. Assume $\rho < 1$ and $\Xi = 0$. For all $0 < z \le 1$, the relation

(2.23)
$$E[z^{X_n}] = \sum_{m=0}^n R(m)g(\xi(z), n-m), \qquad n=0,1,\ldots,$$

holds with $\xi(z)$ defined by (2.4).

PROOF. The forward recurrence times process $\{\mu(n), n = 0, 1, ...\}$ of the discrete-time, nondelayed renewal process $\{\tau_n, n = 1, 2, ...\}$ is also a discrete-time N-valued regenerative process ([3], Chapter 9). As is well known ([3], Chapter 9), this regeneration property translates into the renewal equation

(2.24)
$$E[y^{\mu(n)}] = E[y^{\mu(n)}I[\tau_1 > n]] + \sum_{m=0}^{n} f(m)E[y^{\mu(n-m)}],$$

valid for all $0 \le y \le 1$, where $\{f(n), n = 0, 1, ...\}$ denotes the point mass distribution characterizing the nondelayed renewal process $\{\tau_n, n = 1, 2, ...\}$. The solution of the renewal equation (2.24) is given by [3], Theorem 2.3, page 294,

(2.25)
$$E[y^{\mu(n)}] = \sum_{m=0}^{n} R(m)g(y, n-m), \qquad n=0,1,\ldots,$$

and the representation (2.23) follows immediately from (2.5) and (2.25). \Box

Classically ([8], equation (59), page 70), the transient behavior of the queue is given in terms of the double transform N^* defined as

(2.26)
$$N^*(z,t) = \sum_{n=0}^{\infty} t^n E[z^{X_n}], \quad 0 \le z \le 1, 0 \le t < 1.$$

In order to recover the classical expression for this function from (2.23), it is convenient to introduce the generating functions

(2.27)
$$R^*(t) = \sum_{n=0}^{\infty} R(n)t^n$$

and

(2.28)
$$G^*(y,t) = \sum_{n=0}^{\infty} g(y,n)t^n.$$

These generating functions are expressed in term of the root function $\xi \to Z(\xi)$ of Lemma 4 through the relations

(2.29)
$$R^*(t) = \frac{1}{1 - Z(t)}$$

and

(2.30)
$$G^*(y,t) = 1 + \frac{tZ(y) - yZ(t)}{y - t}$$

defined for $0 \le t < 1$. The derivation of (2.29) is immediate from (2.21), while the derivation of (2.30) from (2.22) is established in Section 3. The well-known formula for $N^*(z,t)$ given in [8], equation (59), page 70 can now be directly recovered from these relations.

COROLLARY 7. Assume $\rho < 1$ and $\Xi = 0$. For all $0 < z \le 1$, the relation

(2.31)
$$N^*(z,t) = \frac{z(1-Z(t))-(1-z)ta(z)}{(1-Z(t))(z-ta(z))}$$

holds with Z as defined in Lemma 4.

PROOF. Taking the generating function in n of both sides of (2.23) readily yields the relation

$$(2.32) N^*(z,t) = R^*(t)G^*(\xi(z),t), n = 0,1,\ldots,$$

and (2.31) is a now a direct consequence of (2.29) and (2.30). \square

3. Derivation of (2.30). It is plain from (2.22) that the relation

(3.1)
$$g(y,n) = E[y^{\tau_1-n}I[\tau_1 > n]] = \sum_{k=n+1}^{\infty} f(k)y^{k-n}$$

holds for all n = 1, 2, ... and $0 < y \le 1$. With C denoting the unit circle in the complex plane, the integral representation

(3.2)
$$f(k) = \frac{1}{2i\pi} \int_C \frac{F^*(u)}{u^{k+1}} du, \qquad k = 0, 1, \dots,$$

can be used in (3.1) to readily yield the relation

(3.3)
$$g(y,n) = \frac{y}{2i\pi} \int_C \frac{F^*(u)}{u-y} \frac{du}{u^{n+1}}, \qquad n = 1, 2, \dots,$$

after standard algebraic manipulations. It is now plain from (2.22), (2.28) and (3.3) that

(3.4)
$$G^{*}(y,t) = 1 + \frac{y}{2i\pi} \int_{C} \frac{F^{*}(u)}{u - y} \sum_{n=1}^{\infty} \left(\frac{t}{u}\right)^{n} \frac{du}{u}$$
$$= 1 + \frac{yt}{2i\pi} \int_{C} \frac{F^{*}(u)}{(u - y)u(u - t)} du,$$

where the passage from the first to the second equality follows from standard facts on geometric series. Finally, substitute the decomposition

(3.5)
$$\frac{1}{(u-y)u(u-t)} = \frac{1}{(u-y)}\frac{1}{y(y-t)} + \frac{1}{(u-t)}\frac{1}{t(t-y)} + \frac{1}{uty}$$

in the last expression of (3.4) and note by a straightforward application of Cauchy's formula that

(3.6)
$$G^*(y,t) = 1 + \frac{tF^*(y) - yF^*(t)}{y-t} + F^*(0).$$

The identity (2.14) and the fact that Z(0) = 0 imply that

(3.7)
$$G^*(y,t) = 1 + \frac{tZ(y) - yZ(t)}{y - t}$$

and (2.30) is obtained by some elementary algebra after using (3.7) in (2.32). □

Observe that $N^*(z, t)$ also admits an integral representation in the form

(3.8)
$$N^*(z,t) = \frac{1}{1-Z(t)} \left[1 + \frac{\xi(z)t}{2i\pi} \int_C \frac{Z(u)}{(u-\xi(z))u(u-t)} du \right],$$

which follows from (2.32) and (3.4).

REFERENCES

- BACCELLI, F. and MAKOWSKI, A. M. (1986). Martingale arguments for stability: The M/GI/1 case. Systems Control Lett. 6 181-186.
- [2] BACCELLI, F. and MAKOWSKI, A. M. (1988). Exponential martingales for queues in random environment: The M/GI/1 case. Unpublished.
- [3] CINLAR, E. (1975). Introduction to Stochastic Processes. Prentice-Hall, Englewood Cliffs, N.J.
- [4] COHEN, J. W. (1969). The Single Server Queue. North-Holland, Amsterdam.
- [5] KLEINROCK, L. (1976). Queueing Systems I: Theory. Wiley, New York.
- [6] NEVEU, J. (1975). Discrete-Time Martingales. North-Holland, Asmterdam.
- [7] ROSENKRANTZ, W. A. (1983). Calculation of the Laplace transform of the length of the busy period for the M/GI/1 queue via martingales. Ann. Probab. 11 817-818.
- [8] TAKACS, L. (1962). Introduction to the Theory of Queues. Oxford Univ. Press, Oxford.
- [9] WOODROOFE, M. (1982). Nonlinear Renewal Theory in Sequential Analysis. SIAM, Philadel-

Institut National de Recherche en Informatique et en Automatique Unite de Recherche Sophia—Antipolis 2004 Route des Lucioles 06565 Valbonne Cedex France DEPARTMENT OF ELECTRICAL ENGINEERING AND SYSTEMS RESEARCH CENTER UNIVERSITY OF MARYLAND COLLEGE PARK, MARYLAND 20742