A LEVEL CROSSING ANALYSIS OF THE *MAP/G/1* QUEUE

Qi-Ming HE and E. M. JEWKES Department of Management Sciences, University of Waterloo, Waterloo, Canada, N2L 3G1

ABSTRACT A level crossing analysis is used to characterize the virtual waiting-time process in a *MAP/G/1* queue. The basic relation between the number of upcrossings and downcrossings over level x is established. Based on this basic relation, a generalized Pollaczek-Khinchin formula for the virtual waiting time is derived.

1. INTRODUCTION

The *MAP/G/1* queue was introduced by Lucantoni, Meier-Hellstern and Neuts (1990) (also see Lucantoni (1991), Neuts (1989b) and Ramaswami (1980)). This is an algorithmically tractable queueing model which can be used to model a wide variety of systems due to the rich class of input processes the Markov arrival process (*MAP)* can represent. The *Poisson* process, *Erlang* process, *PH*-renewal process and *Markov-Modulated Poisson process* are all special cases of a Markov arrival process. Existing methods for analysing the *MAP/G/1* queue rely on Markov renewal theory, via matrix generalizations of embedded Markov chains. In this paper, a level crossing analysis of the virtual waiting time process is used to analyze the *MAP/G/1* queue from a different perspective.

The level crossing method was pioneered in the mid 1970's by Brill (1975). The same level crossing idea was later applied to cycles in regenerative processes (see, for example, Cohen (1977) and Shanthikumar (1980)). Level crossing analysis has been used successfully in analysing waiting-time processes in various queueing models (e.g. Brill and Posner (1977), Shanthikumar (1980), Jewkes and Buzacott (1991), and Doshi (1992)). However, no work has been published so far on the level crossing analysis as applied to the *MAP/G/1* queue, though related ideas can be found in Asmussen (1989,1992).

The main contributions of this paper are twofold: first, a level crossing analysis is used to characterize the virtual waiting-time process in the *MAP/G/1* queue. Though the *MAP/G/1* queue has been studied by others, the level crossing analysis provides a much simpler and more compact method for deriving the basic results for the *MAP/G/1* queue. The second contribution is that the level crossing analysis yields a new generalized Pollaczek-Khinchin formula (see Neuts (1986) and

Ramaswami (1980)).

The rest of the paper is organized as follows: Section 2 introduces the *MAP/G/1* queue. In Section 3, the basic relation between the number of upcrossings and downcrossings over level *x* of the virtual waiting-time process will be established. Section 4 will show how the basic relation can be used in the analysis of the *MAP/G/1* queue by deriving a new generalized Pollaczek-Khinchin formula for the virtual waiting time.

2. The *MAP/G/1* queue

We consider a single server queueing system characterized by a Markov arrival process (*MAP*), general independent service times and a "first come first served" service discipline. The Markov arrival process was first introduced by Neuts (1979) as a generalization of a phase-type renewal process (also see Lucantoni, Meier-Hellstern and Neuts (1990)). It is defined on a finite Markov process (called the underlying Markov process) which is irreducible and has *m* states and an irreducible infinitesimal generator *D*. In the *MAP*, the sojourn time in state *i* is exponentially distributed with parameter D_{ii}. At the end of the sojourn time in state *i*, a transition occurs to state *j*, $1 \leq j \leq m$, where the transition may or may not represent an arrival. Let D_0 be the (matrix) rate of transitions without an arrival and D_1 be the rate of transitions with an arrival. D_0 and D_1 are $m \times m$ matrices where D_0 has negative diagonal elements and non-negative off-diagonal elements. D_1 is a non-negative matrix and $D=D_0+D_1$.

Using *?* to symbolize the stationary probability vector of the Markov process with the generator *D*, *?* will satisfy the equations *?D=0* and *?e=1*, where *e* is a column vector with all components *1*. The stationary arrival rate of the Markov arrival process is then *?= ?D₁e*.

The service time of a customer is a random variable with finite mean and variance. It will be denoted by t. Further, let *F(x)* be the distribution function of t and denote the Laplace-Stieltjes transform (*LST*) of *F(x)* by $f^*(s)$. In addition, t_n will be used to represent the service time of the *n*th customer, where $\{t_n, n \geq 1\}$ are i.i.d. random variables independent of the input process. Throughout this paper, we assume that the queueing system is in steady state. We use the terms *phase* and *state* of the *MAP* synonymously.

3. The Level Crossing Analysis

First, we consider a busy cycle for the *MAP/G/1* queue. For simplicity, we assume that the busy cycle begins at time *t = 0* with the first arrival, ends at time *t = C* and that N_c customers are served in the busy cycle. Other definitions used are as follows:

t_n: the arrival epoch of the nth customer;
w_n: the actual waiting time of the *n*th customer;
v_t: the virtual waiting time of the *MAP/G/1* queue at time *t*;
J(t):   the phase of the input *MAP* at time *t*.

ntainsData

Figure 1: A Sample Path of the Virtual Waiting Time Process for a Type i_1 Busy Cycle. In Figure 1, i_n and j_n are used to denote the phase of the input *MAP* just prior to and just after the arrival of the *n*th customer, respectively.

Generally, the total number of upcrossings and downcrossings over level x by the virtual waiting time process in a busy cycle are equal (see (1.7) in Cohen (1977)). It is through this observation that the distribution functions for the virtual and actual waiting time can be related. To do so, it is useful to distinguish those upcrossings and downcrossings of v_t over level x by the phase of the *MAP* when they occur. It is also useful to characterize a busy cycle by the state of the *MAP* just prior to the first arrival in the busy cycle: a type i busy cycle is one in which the *MAP* just prior to the start of a busy cycle is in phase i. Accordingly, for a type i busy cycle, the number of downcrossings over the level x during the busy period when the *MAP* is in phase j (for $1 \le i, j \le m$) is defined by:

$$\Omega_{i,j}(x) = \#\{\ t : v_t = x, J(t) = j \,|\, 0 \le t < C, v_{0-} = 0, v_{0+} > 0, J(0-) = i\}, \quad x > 0. \qquad \mathbf{1}$$

Throughout this paper, the downcrossing (upcrossing) level x is assumed to be positive.

Similarly, we define the upcrossings in terms of the phase of the *MAP* just after the arrival (j) and the type of the busy cycle (i):

$$U_{i,j}(x, y) = \#\{\ n : w_n < x \le w_n + \tau_n < x + y, J(t_n+) = j \,|\, n = 1, 2, ..., N_c, J(0-) = i\}. \qquad \mathbf{2}$$

Let $\mathcal{O}(x)$ be an $m \times m$ matrix with the elements $\mathcal{O}_{i,j}(x)$ and $U(x, y)$ an $m \times m$ matrix with the elements $U_{i,j}(x, y)$. Thus $\mathcal{O}(x)$ records the number of down-crossings over level x in a busy cycle and $U(x, y)$ records the number of up-crossings over level x and under level $x+y$ in a busy cycle.

The first task in relating the virtual and actual waiting time distributions is to establish the relationship between $\mathcal{O}(x)$ and $U(x, y)$. To do so, we introduce the matrices G and $F(y)$. We define the $m \times m$ matrix G with elements $G_{i,j}$ ($1 \le i, j \le m$) in the following way,

$G_{i,j} =$ **P**{ A busy period ends when the *MAP* is in phase j, given that the phase of the *MAP* just after the start of the busy period was i. },

where **P**{···} denotes probability. It is known that G is the minimal non-negative solution to the equation (see Neuts (1989a) and (1989b) and Lucantoni (1991)):

$$G = \int_0^\infty \exp\{ \ y(\ D_0 + D_1 G) \ \} F(dy).$$

3

Further, we define the $m \times m$ matrix $F(y)$ with the elements $F_{i,j}(y)$ $(1 \leq i, j \leq m)$ as:

$F_{i,j}(y) = I(\{$ A busy period begins with an amount of work y and ends when the *MAP* is in phase j, given that the phase of the *MAP* just after the start of the busy period was i. }),

where $I(\cdot)$ is an indicator function, i.e., $I(\cdot) = 1$ or 0 depending on whether the statement is true or not. By Theorem 2 in Neuts (1989a) (or see Lucantoni (1991)), the expectation of $F(y)$ is given by:

$$\mathbf{E} \ \Phi(\mathbf{y}) = \exp\{\mathbf{y}(\mathbf{D_0} + \mathbf{D_1} \mathbf{G})\}.$$

4

where **E** means mathematical expectation. With (4), the relation between $\Omega(x)$ and $U(x, y)$ can now be defined.

THEOREM 1. For the *MAP/G/1* queue:

$$\Omega(x) = \int_0^\infty [\ d_y \ U(x, y)] \ \Phi(y);$$

5

$$\mathbf{E} \ \Omega(x) = \int_0^\infty \mathbf{E} \ [\mathbf{d_y} \ \mathbf{U}(\mathbf{x}, \mathbf{y})] \exp\{ \ \mathbf{y}(\mathbf{D_0} + \mathbf{D_1} \mathbf{G}) \ \}.$$

where "d_y" means differentiation with respect to y.

Proof. First, observe that during a busy period, a downcrossing over level x must be preceded by an up-crossing over level x (see Figure 1). In particular, we are concerned with the phase of the *MAP* just prior to the upcrossing and later when the downcrossing occurs. Consider an upcrossing over level x which brings the virtual waiting time to $x+y$ and the phase of the *MAP* from phase i to phase k. If the phase of the *MAP* at the next downcrossing of level x is j, then we refer to this downcrossing of level x as one which started from phase i and concluded in phase j. The total number of downcrossings over level x is found by summing over all possible intermediate phases k $(1 \leq k \leq m)$ and all values of y $(y > 0)$. Conditioning on the phase of the *MAP* just after the upcrossings over level x and the magnitude of the upcrossing over x gives the following basic relation:

$$\Omega_{i,j}(x) = \sum_{k=1}^m \int_0^\infty [\ d_y U_{i,k}(x, \ y)] \ \Phi_{k,j}(y).$$

6

The matrix form of this equation is the first equation in (5).

By the Markovian property of the input *MAP*, $F_{k,j}(y)$ is independent of $U_{i,k}(x, y)$ for $1 \le i, k, j \le m$. Taking mathematical expectation on both sides of (6), the second equation in (5) is obtained by (4). ¦

It is noteworthy that the equations in (5) can also be proved by using a ladder height analysis (some applications of ladder height analysis in queueing theory can be found, for example, in Asmussen (1989,1992)). The derivation starts by defining the matrix Q as:

$$Q = D_0 + D_1 G. \qquad \qquad \mathbf{7}$$

Then (4) can also be written as $\mathbf{E}F(y)=exp\{yQ\}$ and (3) as $G=\int exp\{yQ\}dF(y)$. In Asmussen (1989, 1992), it was shown that Q is an irreducible infinitesimal generator satisfying

$$Q = D_0 + D_1 \int_0^\infty exp\{ yQ \} dF(y) \qquad \qquad \mathbf{8}$$

by using a ladder heights analysis for the *Markov-Modulated M/G/1* queue. The proof is also valid for the *MAP/G/1* queue. A probabilistic interpretation of Q is given in those papers.

Asmussen (1989) also gave a probabilistic interpretation of the matrix Q: Let $\{ S(t) \}$ be a process which starts at zero, decreases linearly between arrivals and takes an upwards jump upon the arrival of a customer with an amount equal to the service time of that customer. The Markov jump process $m(x)$ is obtained by observing $J(t)$ when $S(t)$ is at its minimum, i.e., $m(x)=j$ when for some t we have $S(t)=-x$, $J(t)=j$, $S(t)<S(u)$ for $u<t$. The matrix Q is then the intensity matrix (or infinitesimal generator) of the Markov jump process $m(x)$.

4. The Virtual and Actual Waiting Times

Now, as for an application of Theorem 1, we consider the waiting times in the *MAP/G/1* queue. This leads to a generalization of the classical Pollaczek-Khinchin formula. We will use the results of the previous section to define the relationship between the virtual and actual waiting time distributions. For this purpose, we define (for $1 \le i, j \le m$):

$V_{i,j}(x) = \qquad \mathbf{P}\{$ The virtual waiting time at an arbitrary time is no more than x and the phase is j, given that the current busy cycle is type i. $\}$

Figure 2 A Sample Path of a Type i_1 Regeneration Cycle

$W_{i,j}(x) =$ **P**{ The waiting time of an arbitrary customer is no more than x and the phase just after the arrival of the customer is j, given that the current busy cycle is type i.}

Let $V(x)$ be the $m \times m$ matrix with elements $V_{i,j}(x)$ and $W(x)$ be the $m \times m$ matrix with elements $W_{i,j}(x)$. Also, let $V^*(s)$ and $W^*(s)$ denote the *LST*s of $V(x)$ and $W(x)$, respectively. Now, let $\{ t_n(i) \}$ be the epochs at which type i busy cycles begin. For a given i, it is clear that $\{ t_n(i) \}$ are regeneration points of the virtual waiting-time process and so define what we will refer to as type i regeneration cycles (see Figure 2 for a type i_1 regeneration cycle). Let $C(i) = (t_{n+1}(i)\text{-}t_n(i))$ be the length of a type i regeneration period and $N(i)$ the number of customers served during the time interval $[t_n(i), t_{n+1}(i))$.

The following lemma gives the relations between $U(x, y)$ and $W(x)$ as well as between $\Omega(x)$ and $V(x)$. Similar relations for the *GI/G/1* queue are given in Cohen [**Error! Reference source not found.**].

LEMMA 2. For the *MAP/G/1* queue,

$$\int_0^\infty E[d_y\, U(x, y)] \exp\{yQ\} = \mathbf{diag}(\mathbf{EN(1)},\ ...,\ \mathbf{EN(m)}) \int_{y=0}^\infty \int_{u=0}^x [\mathbf{dW(u)}] \exp\{\, yQ\} d_y\, F(x - u + y);$$

9

$$E\,\Omega(x) = diag(\,\mathbf{EC(1)},\ ...,\ \mathbf{EC(m)})\, \frac{\mathbf{d}}{\mathbf{dx}}\, \mathbf{V(x)},$$

where *diag()* means diagonal matrix.

Proof. Let w be the stationary waiting time of an arbitrary customer and v the stationary virtual waiting time at an arbitrary time. Also let $J(t, 0\text{-})$ denote the phase just prior to the current busy cycle (including time t), $J_{n, 0\text{-}}$ the phase just prior to the start of the busy cycle which includes the nth arrival, and

$$\hat{J} = \lim_{n \to \infty} J(t_n+) \quad \text{and} \quad \hat{j}_{0-} = \lim_{n \to \infty} J_{n,0-}.$$ **10**

$\hat{J}$ 11 is the phase just after the arrival of an arbitrary customer and $\hat{J}_{0-}$ 12 is the phase just prior the busy cycle which includes an arbitrary customer.

For a fixed i, we consider type i regeneration cycles. The number of up-crossings $U_{i,j}(x, y)$ is the same for a type i busy cycle and for a type i regeneration cycle. By definition, we have

$$\sum_{k=1}^{m} \int_{y=0}^{\infty} [\, d_y U_{i,k}(x, y)] \, \Phi_{k,j}(y) = \sum_{n=1}^{N(i)} (\sum_{k=1}^{m} \int_{y=0}^{\infty} [\, d_y \, I(\{\, w_n < x \le w_n + \tau_n < x + y, J(t_n+) = k \,|\, J_{n,0-} = i\}]) \Phi_{k,j}(y,$$

(11)

From the theory of regenerative processes (see Cohen (1976)), we have

$$\mathbf{E}(\mathbf{I}(\{\mathbf{w} < \mathbf{x} < \mathbf{w} + \tau < \mathbf{x} + \mathbf{y}, \hat{\mathbf{J}} = \mathbf{k} \,|\, \hat{\mathbf{j}}_{0-} = \mathbf{i}\})) = \dfrac{\mathbf{E}(\sum_{n=1}^{N(i)} \mathbf{I}(\{\mathbf{w_n} < \mathbf{x} \le \mathbf{w_n} + \tau_n < \mathbf{x} + \mathbf{y}, \mathbf{J(t_n}+) = \mathbf{k} \,|\, \mathbf{J_{n,0-}} = \mathbf{i}\}}{\mathbf{E} \, \mathbf{N(i)}}$$

(12)

and

$$\mathbf{E}(\int_{\mathbf{y=0}}^{\infty} [\mathbf{d_y} \mathbf{I}(\{\mathbf{w} < \mathbf{x} < \mathbf{w} + \tau < \mathbf{x} + \mathbf{y}, \hat{\mathbf{J}} = \mathbf{k} \,|\, \hat{\mathbf{j}}_{0-} = \mathbf{i}\})]) = \int_{\mathbf{y=0}}^{\infty} \int_{\mathbf{u=0}}^{\mathbf{x}} [\mathbf{dW_{i,k}(u)}] \mathbf{d_y} \, \mathbf{F(x-u+y)}. \quad \textbf{(13)}$$

Taking mathematical expectations on both sides of (11), by (12) and (13), the first part of (9) is obtained by referring to equations (4) and (7).

Now, consider the following function of x:

$$\int_0^{C(i)} I(\{\, v_t \le x, J(t) = j \,/\, J(t, 0-) = i \,\}) dt. \tag{14}$$

This function is an almost everywhere continuous, nondecreasing and piecewise linear function and is differentiable for all $x > 0$. By the definition of $O_{i,j}(x)$, we have

$$\Omega_{i,j}(x) = \frac{d}{dx} \int_0^{C(i)} I(\{\, v_t \le x, J(t) = j \,/\, J(t, 0-) = i \,\}) dt, \quad a.s.. \tag{15}$$

Taking mathematical expectations on both sides and by the theory of regenerative processes, the second equation in (9) is obtained.

Therefore, by Lemma 2, we have

$$\frac{d}{dx} V(x) = diag(\frac{\mathbf{EN(1)}}{\mathbf{EC(1)}}, \ ..., \ \frac{\mathbf{EN(m)}}{\mathbf{EC(m)}}) \int_{\mathbf{y=0}}^{\infty} \int_{\mathbf{u=0}}^{\mathbf{x}} [\mathbf{dW(u)}] \exp\{\mathbf{yQ}\} \mathbf{d_y} \, \mathbf{F(x-u+y)}. \tag{16}$$

Since $?$ is the arrival rate of the *MAP*, we must have $? = EN(i)/EC(i)$, for all $1 \le i \le m$. Taking *LST*s on both sides of (16), we have, for $s > 0$,

$$V^*(s) - V(0) = \lambda \int_{x=0}^{\infty} e^{-sx} \int_{y=0}^{\infty} \int_{u=0}^{x} [dW(u)] \exp\{yQ\} \, d_y \, F(x - u + y) dx$$

$$= \lambda \int_{u=0}^{\infty} dW(u) \int_{y=0}^{\infty} \int_{x=u}^{\infty} e^{-sx} dx [\, d_{yF}(x - u + y)] \exp\{yQ\}$$

$$= \lambda \int_{u=0}^{\infty} dW(u) \int_{y=x-u}^{\infty} \int_{x=u}^{\infty} e^{-sx} dx \, [dF(y)] \exp\{(y + u - x)Q\}$$

(17)

$$= \lambda \int_{u=0}^{\infty} dW(u) \int_{y=0}^{\infty} \int_{x=u}^{x=u+y} e^{-sx} dx \, [dF(y)] \exp\{(y + u - x)Q\}$$

$$= \lambda \int_{u=0}^{\infty} dW(u) \int_{y=0}^{\infty} [\, e^{-su} \exp\{yQ\} - e^{-s(u+y)}](sI + Q)^{-1} dF(y)$$

$$= \lambda \, [\, w^*(s) \int_{0}^{\infty} \exp\{yQ\} dF(y) - w^*(s) \, f^*(s) \,] \, (sI + Q)^{-1}.$$

The matrix $sI+Q$ is invertible since Q is an infinitesimal generator and $s>0$.

For the *MAP*, the probability that there is an arrival in $(t, t+ ?t)$ is approximately $??t$. The probability that there is an arrival in $(t, t+ ?t)$ and finding a waiting time (in LST form) $V^*(s)$ is approximately $V^*(s)D_1 ?t$. Then the waiting time of an arrival is approximately $V^*(s)D_1 ?t/ ? ?t$. Letting $?t$ goes to zero, we obtain (as in Lucantoni, Meier-Hellstern and Neuts (1990) and Neuts (1989b)):

$$W^*(s) = \frac{1}{\lambda} V^*(s) \, D_1.$$

(18)

Finally, by (8), (17) and (18), we are able to obtain the *LST* of the virtual waiting time.

THEOREM 3. In the *MAP/G/1* queue, we have:

$$V^*(s) = V(0)(sI + D_0 + D_1 G)(sI + D_0 + D_1 f^*(s))^{-1}.$$

(19)

The classical Pollaczek-Khinchin formula is obtained from (19) when $m=1$, i.e., for the *Poisson* input process. The vector generalization of the Pollaczek-Khinchin formula can be obtained from (19) with the following probabilistic interpretation on the matrix $V(0)$.

The (i, j)th element of $V(0)$ is the probability that at an arbitrary time the waiting time is zero and the *MAP* is in phase j, given that the phase just prior to the start of the current busy cycle is i. Let y_{0j} be the probability that at an arbitrary time the queue is empty, $1 \leq j \leq m$. $\mathbf{y}_0 = (y_{01}, y_{02}, ..., y_{0m})$. Let b_j be the probability that a busy cycle starts in phase j. $\mathbf{b} = (b_1, b_2, ..., b_m)$. Then we have

$$\mathbf{b} = \frac{\mathbf{y}_0 \, \mathbf{D}_1}{\mathbf{y}_0 \, \mathbf{D}_1 \, \mathbf{e}} \quad \text{and} \quad \mathbf{y}_0 = \mathbf{b} \, \mathbf{V(0)} .$$

(20)

It has been proved in Lucantoni, Meier-Hellstern and Neuts (1989) that $\mathbf{y}_0 Q=0$. Therefore, when

we multiply both sides of (19) by **b**, the vector form *LST* of the virtual waiting time is obtained, which was obtained first in Ramaswami (1980).

REFERENCES

Asmussen, S. (1989), Ladder heights and the *Markov-Modulated M/G/1* queue, *Stoch. Proc. Appl*, **37**, 313-326.

Asmussen, S. and Perry, D. (1992), On cycle maxima, first passage problems and extreme value theory for queues, *Stochastic Models*, **8**, 421-458.

Brill, P.H. (1975), System point theory in exponential queues, Ph.D. Thesis, University of Toronto, Univ. Microfilms Int. 8901129, Ann Arbor MI.

Brill, P.H. and Posner (1977), Level crossings in point processes applied to queues: single server case, *Operations Research*, **25**, 662-674.

Cohen, J.W. (1976), *On Regenerative Processes in Queueing Theory*, Lecture Notes in Economics and Mathematical Systems **121**, Springer-Verlag, Berlin.

Cohen, J.W. (1977), On up- and downcrossings, *J. Appl. Prob.*, **14**, 405-410.

Doshi, Bharat. (1992), Level-crossing analysis of queues, *Queueing and related Models*, Clarendon Press, 3-33.

Jewkes, E.M. and Buzacott, J.A. (1991), Flow time distributions in a K class *M/G/1* priority feedback queue, *Queueing Systems*, **8**, 183-202.

Lucantoni, D.M., Meier-Hellstern, K.S. and Neuts, M.F. (1990), A single server queue with server vacations and a class of non-renewal arrival processes, *Adv. Appl. Prob.*, **22**, 676-705.

Lucantoni, D.M. (1991), New results on the single server queue with a batch Markovian arrival process, *Stochastic Models*, **7**, 1-46.

Neuts, M.F. (1979), A versatile Markovian point process, *J. Appl. Prob.*, **16**, 764-779.

Neuts, M.F. (1986), Generalizations of the Pollaczek-Khinchin integral equation in the theory of queues, *Adv. Appl. Prob.*, 18, 952-990.

Neuts, M.F. (1989a), The fundamental period of the queue with Markov-Modulated arrivals, *Probability, Statistics and Mathematics*: Papers in Honour of Samuel Karlin, Academic Press, 187-200.

Neuts, M.F. (1989b), *Structured Stochastic Matrices of M/G/1 Type and Their Applications*, New York: Marcel Dekker Inc.

Ramaswami, V. (1980), The *N/G/1* queue and its detailed analysis, *Adv. Appl. Prob.*, **12**, 222-261.

Shanthikumar, J.G. (1980), Some analyses on the control of queues using level crossings of regenerative processes, *J. Appl. Prob.*, **17**, 814-821.