View metadata

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

EFMD EQUIS ACCREDITED

ePub^{WU} Institutional Repository

Walter Böhm

A transient Analysis of M/G/1 Queues with N-policy

Working Paper

Original Citation:
Böhm, Walter (1991) A transient Analysis of M/G/1 Queues with N-policy. *Forschungsberichte / Institut für Statistik*, 20. Department of Statistics and Mathematics, WU Vienna University of Economics and Business, Vienna.

This version is available at: http://epub.wu.ac.at/738/
Available in ePub^{WU}: July 2006

ePub^{WU}, the institutional repository of the WU Vienna University of Economics and Business, is provided by the University Library and the IT-Services. The aim is to enable open access to the scholarly output of the WU.

http://epub.wu.ac.at/

A transient Analysis of M/G/1 Queues with N-policy

Walter Böhm

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 20
1991

http://statmath.wu-wien.ac.at/

A transient analysis of M/G/1 queues with N-policy

Walter Böhm

We consider an M/G/1 queueing model with N-policy operating. This means, that the server will start up only if a queue of a prescribed length has built up. For this model the time dependent distribution of the queue length is given by simple renewal arguments without resorting to integral transform techniques.

1 Introduction

In this paper we discuss M/G/1 queueing systems with N-policy operating. The underlying idea, inspired by cost optimization issues (Heyman (1968)), is to turn off the server, whenever the queue becomes empty and resume service only if a new queue of a prescribed length has built up. Models of this type have received considerable attention over the last years because they have important applications in data transmission and computer science.

We will put our main emphasis on the transient distribution of the queue length because it is one of the major characteristics of a queueing system. Furthermore, once the distribution of the queue length is known, other characteristics, like the distribution of the virtual waiting time may be derived, at least in principle, in a straight forward way. The results to be presented below are derived by rather simple renewal arguments and avoid the use of integral transforms. As a nice byproduct our results give rise to an apparently new representation of the queue length distribution in the ordinary M/G/1 model without N-policy, which is simpler than similar formulas reported by other authors.

The queueing system we will consider consists of a single server and satisfies the following assumptions:

- Customers arrive according to a Poisson process with constant rate $\lambda > 0$. Thus the probability function of $A(t)$, the number of arrivals in $(0, t]$, is given by

$$P(A(t) = k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}, \qquad k \geq 0.$$

- Customers are served one at a time, the service times being independently and identically distributed with distribution function $B(t)$, which is assumed to have finite expectation.

- N-policy operating: whenever the queue becomes empty, the server is turned off and remains idle until a new queue of a prescribed length, say $N \geq 1$, has built up. Then service resumes and continues until the queue is empty again.

Let $Q(t)$ denote the number of customers in the system at time t, including the one being in service, and define an indicator variable $\xi(t)$ by

$$\xi(t) = \begin{cases} 1 & \text{if the server is busy at time } t \\ 0 & \text{otherwise} \end{cases}$$

2 The joint distribution of $Q(t)$ and $\xi(t)$

Let v_n denote the service time of the $n-$th customer after system start-up and define $V_n = \sum_{i=1}^{n} v_i$. The process V_n is a renewal process by independence of the service times and $D(t) = \max\{n : V_n \leq t\}$ equals the number of service completions during $(0, t]$. We have

$$\begin{aligned} P(D(t) = n) &= B_n(t) - B_{n+1}(t) \\ &= C_n(t), \end{aligned}$$

where $B_n(t)$ denotes the $n-$fold convolution of $B(t)$ with itself.

Our task of deriving the joint distribution of $Q(t)$ and $\xi(t)$ will be facilitated by introducing the non-markovian process

$$X_m(t) = m + A(t) - D(t), \tag{1}$$

m being any integer. The importance of $X_m(t)$, which has been recognized as early as 1956 by Champernowne, is due to the following fact: assume that at time zero there are

$m > 0$ customers waiting and a new service starts. Then the transitions of $X_m(t)$ coincide with the transitions of $Q(t)$ as long as $Q(t) > 0$. Thus if T_m denotes the length of a busy period initiated by m customers, i.e. $T_m = \inf\{t : Q(t) = 0, Q(0) = m\}$, then we have equivalently

$$T_m = \inf\{t : m + A(t) - D(t) = 0\}$$

and the distribution of the queue length during a busy period may be expressed as

$$P(Q(t) = k, T_m > t | Q(0) = m) = P(X_m(t) = k, T_m > t). \tag{2}$$

Because of this close relationship between the processes $Q(t)$ and $X_m(t)$ it seems to be appropriate to formulate some basic results about $X_m(t)$ first.

Let us start with the transition probabilities of $X_m(t)$. In particular define $p_{m,k}(t) = P(X_m(t) = k)$. Then

$$\begin{aligned} p_{m,k}(t) &= \sum_{j\geq 0} P(m + A(t) - D(t) = k, D(t) = j) \\ &= \sum_{j\geq 0} C_j(t) e^{-\lambda t} \frac{(\lambda t)^{k-m+j}}{(k-m+j)!}. \end{aligned} \tag{3}$$

The density of the stopping time T_m is well known and given by

$$dG_m(t) = \sum_{j\geq m} \frac{m}{j} e^{-\lambda t} \frac{(\lambda t)^{j-m}}{(j-m)!} dB_j(t). \tag{4}$$

This formula has been derived first by Prabhu (1960) by analytical methods and by Takács (1961) by purely combinatorial arguments. It is important to observe that the density (4) has the property

$$dG_a(t) * dG_b(t) = dG_{a+b}(t) \tag{5}$$

for positive integers a and b, because $T_a + T_b = T_{a+b}$.

The transition probabilities (2) during busy periods can now be expressed in terms of $p_{m,k}(t)$ and $dG_m(t)$. To see this we note that for $m > 0$

$$p_{m,k}(t) = P(X_m(t) = k, T_m > t) + P(X_m(t) = k, T_m \leq t).$$

However, we have ("$*$" denoting convolution):

$$P(X_m(t) = k, T_m \leq t) = dG_m(t) * p_{0,k}(t),$$

since the event $\{X_m(t) = k, T_m \leq t\}$ implies that at some time $0 < s < t$ the process $X_m(s)$ must have reached the zero state for the first time.

Hence we find that

$$\begin{aligned} s_{m,k}(t) &= P(X_m(t)=k, T_m > t) \qquad (6)\\ &= p_{m,k}(t) - dG_m(t) * p_{0,k}(t), \end{aligned}$$

which may be stated more explicitely as

$$\begin{aligned} s_{m,k}(t) &= \sum_{j\geq 0} C_j(t) e^{-\lambda t} \frac{(\lambda t)^{k-m+j}}{(k-m+j)!} - \\ &- \sum_{j\geq 0} \int_0^t dG_m(s) C_j(t-s) e^{-\lambda(t-s)} \frac{[\lambda(t-s)]^{k+j}}{(k+j)!}. \end{aligned}$$

Consider now the process $Q(t)$ in presence of N-policy. The discussion above suggests (a) to break $Q(t)$ at those time instances where idle and busy periods terminate, these instances being renewal points, and (b) to describe the segments of $Q(t)$ due to busy periods in terms of the auxilliary process $X_m(t)$.

For completed idle periods we note that they always terminate with an arrival. Therefore their lengths have an Erlang density which we denote by

$$q_a(t) = \frac{\lambda^a t^{a-1} e^{-\lambda t}}{(a-1)!}, \qquad a \geq 1.$$

Suppose now that at time zero the server is busy with $m \geq 1$ customers waiting and a service begins. Furthermore let

$$P_i(Q(t)=k, \xi(t)=1 | Q(0)=m, \xi(0)=1)$$

be the probability that at time t there are k customers waiting and the server is busy, provided that there have been $i \geq 0$ completed idle periods and the server was initially busy with m customers in the queue.

If $i=0$, then there has been no idle period at all and

$$\begin{aligned} P_0(Q(t)=k, \xi(t)=1 | Q(0)=m, \xi(0)=1) &= P(X_m(t)=k, T_m > t) \\ &= s_{m,k}(t). \end{aligned}$$

Now let $i \geq 1$. Then we have i completed idle and busy periods, where the first busy period has been initiated by m customers, all others by N customers. Hence we have for $m, k \geq 1$:

$$\begin{aligned} &P_i(Q(t)=k, \xi(t)=1 | Q(0)=m, \xi(0)=1) = \\ &= dG_m(t) * [dG_N(t)]^{(i-1)*} * [q_N(t)]^{i*} * s_{N,k}(t) \\ &= dG_{m+N(i-1)}(t) * q_{Ni}(t) * s_{N,k}(t) \\ &= q_{Ni}(t) * [dG_{m+N(i-1)}(t) * p_{N,k}(t) - dG_{m+Ni}(t) * p_{0,k}(t)], \end{aligned}$$

where f^{i*} denotes i-fold convolution of the function f with itself and we have used (5), (6) and the fact that for Erlang densities and positive integers a, b:

$$q_a(t) * q_b(t) = q_{a+b}(t).$$

Thus upon summation on i and writing the convolutions explicitely as integrals we get for $m, k \geq 1$:

$$P(Q(t) = k, \xi(t) = 1 | Q(0) = m, \xi(0) = 1) = \tag{7}$$

$$= s_{m,k}(t) + \sum_{i \geq 1} \int_0^t q_{Ni}(t-s) \int_0^s [dG_{m+N(i-1)}(u) p_{N,k}(s-u) -$$

$$-dG_{m+Ni}(u) p_{0,k}(s-u)]\, ds.$$

Some remarks concerning this formula are in order: observe first that (7) yields the distribution of $Q(t)$ in the ordinary M/G/1 model without N-policy, if we set $N = 1$. However, the resulting formula is considerably simpler than the classical result given in Prabhu (1965, p. 83), and also simpler than similar results given in Jaiswal (1968, p. 14) and Bhat (1968, p. 25). Second it can be shown, that if the service time distribution is of exponential or Erlangian type, the double integral in (7) reduces to a single integral. And third, if the traffic intensity is close to one, the formula above offers also some computational advantages even if the service time distribution is of general type so that we are left with a double integral, because in this case in the summation on i only a relatively small number of terms will be significant.

Consider now the case $\xi(t) = 0$, i.e. the server is idle at time t. Again let i denote the number of completed idle periods. Then we have in the case $i = 0$ for $0 \leq k < N$ and $m > 0$:

$$P_0(Q(t) = k, \xi(t) = 0 | Q(0) = m, \xi(0) = 1) = dG_m(t) * h_k(t),$$

where

$$h_k(t) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}, \qquad k \geq 0.$$

And if $i \geq 1$:

$$\begin{aligned} P_i(Q(t) = k, \xi(t) = 0 | Q(0) = m, \xi(0) = 1) &= dG_m(t) * [dG_N(t)]^{i*} * [q_N(t)]^{i*} * h_k(t) \\ &= dG_{m+Ni}(t) * h_{k+Ni}(t), \end{aligned}$$

since for positive integers a, b:

$$q_a(t) * h_b(t) = h_{a+b}(t).$$

Thus we get for $0 \le k < N, m > 0$:

$$P(Q(t)=k,\xi(t)=0|Q(0)=m,\xi(0)=1)=\sum_{i\ge 0}\int_0^t h_{k+Ni}(t-s)dG_{m+Ni}(s). \qquad (8)$$

In a similar way the joint distribution of $Q(t)$ and $\xi(t)$ may be derived in the case $\xi(0)=0$. To deal with the case $\xi(0)=0$ and $\xi(t)=1$, we observe that the initial left censored idle period (we assume that $0 \le m < N$ customers are waiting at time zero) requires $N-m$ arrivals to be completed. Thus

$$\begin{aligned} & P(Q(t)=k,\xi(t)=1|Q(0)=m,\xi(0)=0)= \\ &= q_{N-m}(t) * P(Q(t)=k,\xi(t)=1|Q(0)=N,\xi(0)=1) \\ &= \int_0^t q_{N-m}(u)s_{N,k}(t-u)\,du + \qquad (9) \\ &+ \sum_{i\ge 1}\int_0^t q_{N(i+1)-m}(t-s)\int_0^s [dG_{Ni}(u)p_{N,k}(s-u)- \\ & -dG_{N(i+1)}(u)p_{0,k}(s-u)]\,ds. \end{aligned}$$

If $\xi(0)=0$ and $\xi(t)=0$, then we have for $0 \le k, m < N$, $k \ge m$:

$$P_0(Q(t)=k,\xi(t)=0|Q(0)=m,\xi(0)=0)=h_{k-m}(t),$$

and for $i \ge 1$:

$$\begin{aligned} & P_i(Q(t)=k,\xi(t)=0|Q(0)=m,\xi(0)=0)= \\ &= q_{N-m}(t) * P_{i-1}(Q(t)=k,\xi(t)=0|Q(0)=N,\xi(0)=1) \\ &= q_{N-m}(t) * dG_{Ni}(t) * h_{k+N(i-1)}(t) \\ &= dG_{Ni}(t) * h_{k-m+Ni}(t). \end{aligned}$$

And therefore

$$\begin{aligned} & P(Q(t)=k,\xi(t)=0|Q(0)=m,\xi(0)=0)= \qquad (10) \\ &= h_{k-m}(t)+\sum_{i\ge 0}\int_0^t h_{k-m+Ni}(t-s)dG_{Ni}(s). \end{aligned}$$

3 References

[1] Bhat U. N. (1968) A Study of the Queueing Systems M/G/1 and GI/M/1. *Lecture Notes in Operations Research and Mathematical Economics*, Springer, Berlin.

[2] Champernowne D.G. (1956) An elementary method of the solution of the queueing problem with a single server and constant parameters. *J. Roy. Stat. Soc. Ser. B.*, **18**, 125-128.

[3] Heymann D.P. (1968) Optimal operating policies for M/G/1 queueing systems. *Opns. Res.*, **16**, 362-382.

[4] Jaiswal N.K. (1968) *Priority Queues.* Academic Press, New York.

[5] Prabhu N.U. (1960) Some results for the queue with Poisson arrivals. *J. Roy. Stat. Soc. Ser. B.*, **22**, 104-107.

[6] Prabhu N.U. (1965) *Queues and Inventories.* John Wiley and Sons, New York.

[7] Takács L. (1961) The probability law of the busy period for two types of queueing processes. *Opns. Res.* **9**, 402-407.