A Sample Path Analysis of the M/M/1 Queue

Francois Baccelli; William A. Massey

Journal of Applied Probability, Vol. 26, No. 2 (Jun., 1989), 418-422.

Stable URL:
http://links.jstor.org/sici?sici=0021-9002%28198906%2926%3A2%3C418%3AASPAOT%3E2.0.CO%3B2-W

Journal of Applied Probability is currently published by Applied Probability Trust.

J. Appl. Prob. **26**, 418–422 (1989)
Printed in Israel

A SAMPLE PATH ANALYSIS OF THE $M/M/1$ QUEUE

FRANÇOIS BACCELLI,* *INRIA*
WILLIAM A. MASSEY,** *AT&T Bell Laboratories*

Abstract

The exact solution for the transient distribution of the queue length and busy period of the $M/M/1$ queue in terms of modified Bessel functions has been proved in a variety of ways. Methods of the past range from spectral analysis (Lederman and Reuter (1954)), combinatorial arguments (Champernowne (1956)), to generating functions coupled with Laplace transforms (Clarke (1956)). In this paper, we present a novel approach that ties the computation of these transient distributions directly to the random sample path behavior of the $M/M/1$ queue. The use of Laplace transforms is minimized, and the use of generating functions is eliminated completely. This is a method that could prove to be useful in developing a similar transient analysis for queueing networks.

TRANSIENT ANALYSIS; BUSY PERIOD; BESSEL FUNCTIONS; RANDOM WALKS; REFLECTION PRINCIPLE

The exact solution for the transient distribution of the $M/M/1$ queue in terms of modified Bessel functions is well known. During the past four decades there has been no shortage of techniques for obtaining its solution. Approaches of the past have ranged from spectral methods (Lederman and Reuter (1954)), combinatorial arguments (Champernowne (1956)), to generating functions coupled with Laplace transforms (Clarke (1956)). As recently as 1987 the transient distribution for the $M/M/1$ queue remained a topic of interest, as the many papers of Abate and Whitt ((1987), for an example) as well as the paper by Parthasarathy (1987) both indicate. This paper grew out of the authors' current interest in developing techniques to derive explicit solutions for the transient distribution of some queueing networks (see Baccelli and Massey (1988)). We want to view the analysis of the $M/M/1$ queue as a stepping stone to some higher-dimensional analogue. Jackson networks, for example, can be viewed as multidimensional versions of the $M/M/1$ queue. In this paper, we present a derivation of the transient distribution for the queue length and the busy period of the $M/M/1$ queue that follows purely from the *sample path behavior* of the process. We eschew the analytical approaches of relying heavily on Laplace transform techniques and eliminate the use of generating functions altogether. The methods that we use instead are an amalgam of techniques like the

Received 22 April 1988.
* Postal address: INRIA, Centre de Sophia Antipolis, 2004 route des Lucioles, 06565 Valbonne Cedex, France.
** Postal address: AT&T Bell Laboratories, Murray Hill, NJ 07974, USA.

reflection principle for random walks and stopping time arguments used by Takács (1962). Despite the popularity of these methods, it appears that we have applied them to the $M/M/1$ queue in a novel way. Moreover, we feel that we have a derivation that is more organically related to the underlying probabilistic structure of the $M/M/1$ queue. To begin, we let $Q(t)$ be our $M/M/1$ queue length process with Poisson arrival rate λ, and exponential server rate μ. We associate with this process $Z(t) = Z(0) + N_\lambda(t) - N_\mu(t)$, where $N_\lambda(t)$ and $N_\mu(t)$ are two independent Poisson processes with rates λ and μ respectively, and $Z(0) = Q(0)$. The transitions of $Z(t)$ are identical to those of $Q(t)$ except at the zero state, where $Z(t)$ is allowed to become negative. One key feature of $Z(t)$ is that it is a nearest-neighbor random walk on the integers. So for example, if $Z(0) = m$ and $Z(t) = n$ with $m < n$, then each corresponding sample path must visit every state between m and n. We will construct the transient distributions for the queue length and busy period of $Q(t)$ in terms of the distribution for this intermediate process $Z(t)$. This will give the solution of the former quantities in terms of modified Bessel functions by the theorem below.

Theorem 1. *For all integers m and n, we have*

$$P_m\{Z(t) = n\} = \exp(-(\lambda + \mu)t)\left(\frac{\lambda}{\mu}\right)^{(n-m)/2} I_{n-m}(2t\sqrt{\lambda\mu}).$$

Proof. Using the fact that $N_\lambda(t)$ and $N_\mu(t)$ are independent processes, we get

$$\begin{aligned}
P_m\{Z(t) = n\} &= \Pr\{N_\lambda(t) - N_\mu(t) = n - m\} \\
&= \sum_{k=0}^{\infty} \Pr\{N_\lambda(t) = k\}\Pr\{N_\mu(t) = k - n + m\} \\
&= \sum_{k=0}^{\infty} \frac{\exp(-\lambda t)(\lambda t)^k}{k!} \cdot \frac{\exp(-\mu t)(\mu t)^{k-n+m}}{(k-n+m)!} \\
&= \exp(-(\lambda + \mu)t)\left(\frac{\lambda}{\mu}\right)^{(n-m)/2} \cdot \sum_{k=0}^{\infty} \frac{(t\sqrt{\lambda\mu})^{2k-n+m}}{k!(k-n+m)!} \\
&= \exp(-(\lambda + \mu)t)\left(\frac{\lambda}{\mu}\right)^{(n-m)/2} I_{n-m}(2t\sqrt{\lambda\mu}),
\end{aligned}$$

where $I_n(\cdot)$ is the nth modified Bessel function.

For all integers n, we define a random stopping time $T_n = \inf\{t \mid Z(t) = n\}$. We will let $P_m\{T_n = t\}$ denote the density of T_n, given that $Z(0) = m$.

Lemma 2. *For all non-negative integers, we have*

$$P_0\{T_n = t\} = \left(\frac{\lambda}{\mu}\right)^n P_0\{T_{-n} = t\}.$$

Proof. We prove this for the Laplace transform version of this lemma, namely

$$E_0(\exp(-sT_n)) = \left(\frac{\lambda}{\mu}\right)^n E_0(\exp(-sT_{-n})).$$

By sample path arguments including the strong Markov property (see Takács (1962), pp. 32–38), we have $E_0(\exp(-sT_n)) = E_0(\exp(-sT_1))^n$ and $E_0(\exp(-sT_{-n})) = E_0(\exp(-sT_{-1}))^n$, so it is sufficient to prove that the lemma holds for the case $n = 1$. Using sample path arguments again, we have

$$E_0(\exp(-sT_1)) = \frac{\lambda}{\lambda+\mu+s} + \frac{\mu}{\lambda+\mu+s} E_0(\exp(-sT_1))^2.$$

This formula is more transparent when we note that $E_{-1}(\exp(-sT_1)) = E_0(\exp(-sT_2)) = E_0(\exp(-sT_1))^2$. Similarly, we have for $E_0(\exp(-sT_{-1}))$,

$$E_0(\exp(-sT_{-1})) = \frac{\mu}{\lambda+\mu+s} + \frac{\lambda}{\lambda+\mu+s} E_0(\exp(-sT_{-1}))^2.$$

Dividing the last equation by $E_0(\exp(-sT_{-1}))^2$, we see that $E_0(\exp(-sT_1))$ and $E_0(\exp(-sT_{-1}))^{-1}$ both solve the same quadratic equation. However, $|E_0(\exp(-sT_1))| < 1$, and for similar reasons $|E_0(\exp(-sT_{-1}))^{-1}| > 1$. Therefore $E_0(\exp(-sT_1))$ and $E_0(\exp(-sT_{-1}))^{-1}$ are two *distinct* roots of the same quadratic equation. From this, it follows that

$$\frac{E_0(\exp(-sT_1))}{E_0(\exp(-sT_{-1}))} = \frac{\lambda}{\mu},$$

and this finishes the proof.

Now we can solve for the busy period distribution of $Q(t)$. We let $*$ denote the convolution operation for probability densities.

Theorem 3. *For all positive integers m and n, we have*

$$P_m\{Q(t)=n,\ T_0>t\} = P_m\{Z(t)=n\} - \left(\frac{\lambda}{\mu}\right)^n P_m\{Z(t)=-n\}.$$

Proof. The derivation goes as follows:

$$\begin{aligned} P_m\{Z(t)=n\} &= P_m\{Z(t)=n,\ T_0>t\} + P_m\{Z(t)=n,\ T_0 \leqq t\} \\ &= P_m\{Q(t)=n,\ T_0>t\} + P_m\{T_0=t\} * P_0\{T_n=t\} * P_n\{Z(t)=n\} \\ &= P_m\{Q(t)=n,\ T_0>t\} \\ &\quad + \left(\frac{\lambda}{\mu}\right)^n P_m\{T_0=t\} * P_0\{T_{-n}=t\} * P_{-n}\{Z(t)=-n\} \\ &= P_m\{Q(t)=n,\ T_0>t\} + \left(\frac{\lambda}{\mu}\right)^n P_m\{Z(t)=-n\}. \end{aligned}$$

Consider the term $P_m\{Z(t)=n,\ T_0 \leqq t\}$. This probability can be viewed as the measure of all paths that start at m, terminate at n, and touch 0 along the way. All such paths are equivalently expressed as paths originating at m that hit 0, then hit n, and ultimately terminate at n. The probability of this sequence of events is $P_m\{T_0=t\} * P_0\{T_n=t\} * P_n\{Z(t)=n\}$. The remaining sequence of steps follows from making similar arguments, appealing to Lemma 2, and using the translation invariance of the process $Z(t)$, namely that $P_m\{Z(t)=n\}=P_{m+k}\{Z(t)=n+k\}$ for all k, and similarly $P_m\{T_n=t\}=P_{m+k}\{T_{n+k}=t\}$. Finally, rearranging terms completes the proof.

Lemma 4. *For all $t>0$, we have*

$$Q(t)=Z(t)-\inf_{0\leqq s\leqq t} Z(t)\wedge 0.$$

Proof. Let $S_1<S_2<\cdots$ equal the embedded time points where the jumps of $Z(t)$ occur. We then construct $Q(t)$ such that its jumps are coupled with those of $Z(t)$. If $Q_i=Q(S_i)$ and $Z_i=Z(S_i)$, we then have

$$Q_{i+1}=(Q_i+Z_{i+1}-Z_i)^+.$$

Since $a^+=\sup(a,0)$, we have

$$Q_{i+1}-Z_{i+1}=\sup(Q_i-Z_i,-Z_{i+1})=\sup(0,-Z_1,\cdots,-Z_{i+1}).$$

The last step follows by induction and the fact that $Q(0)=Z(0)$. Setting $\sup(-a,-b)=-\inf(a,b)$, gives us

$$Q_{i+1}=Z_{i+1}-\inf_{0\leqq j\leqq i+1} Z_j\wedge 0.$$

Since the sample paths of $Q(t)$ and $Z(t)$ are step functions, the lemma is proved.

This sample path construction of $Q(t)$ now gives us its transient distribution.

Theorem 5. *For all non-negative integers m and n, we have*

$$P_m\{Q(t)<n\}=P_m\{Z(t)<n\}-\left(\frac{\lambda}{\mu}\right)^n P_m\{Z(t)<-n\}.$$

Proof. By Lemma 4, we can construct $Q(t)$ such that of $Z(t)\leqq Q(t)$. This gives us

$$\begin{aligned} P_m\{Z(t)<n\} &= P_m\{Z(t)<n,\ Q(t)<n\}+P_m\{Z(t)<n,\ Q(t)\geqq n\}\\ &= P_m\{Q(t)<n\}+P_m\left\{n+\inf_{0\leqq s\leqq t} Z(s)\wedge 0\leqq Z(t)<n\right\}\\ &= P_m\{Q(t)<n\}+\sum_{k<n} P_m\{T_{k-n}=t\} * P_{k-n}\{T_k=t\} * P_k\{Z(t)=k\}\\ &= P_m\{Q(t)<n\}\\ &\quad+\left(\frac{\lambda}{\mu}\right)^n \sum_{k<n} P_m\{T_{k-n}=t\} * P_{k-n}\{T_{k-2n}=t\} * P_{k-2n}\{Z(t)=k-2n\} \end{aligned}$$

$$= P_m\{Q(t) < n\} + \left(\frac{\lambda}{\mu}\right)^n \sum_{k<n} P_m\{Z(t) = k - 2n\}$$

$$= P_m\{Q(t) < n\} + \left(\frac{\lambda}{\mu}\right)^n P_m\{Z(t) < -n\}.$$

To decompose $P_m\{n + \inf_{0 \leqq s \leqq t} Z(s) \wedge 0 \leqq Z(t) < n\}$, we let k equal $Z(t)$, the terminal value of the path. The value of k will be any integer less than n. The constraining inequalities of this event require that $Z(t) - n$ or $k - n$ exceed some minimal value achieved before time t. This is equivalent to a path that starts at m, falls to hit $k - n$, which is negative, rises to hit k, and ultimately terminates at k. So this probability equals $\Sigma_{k<n} P_m\{T_{k-n} = t\} * P_{k-n}\{T_k = t\} * P_k\{Z(t) = k\}$. The remaining steps are due to arguments similar to those in Theorem 3, and this finishes the proof.

Acknowledgements

This paper came about during the second author's enjoyable two-month stay at INRIA during the fall of 1986.

References

ABATE, J. AND WHITT, W. (1987) Transient behavior of the *M/M/*1 queue: starting at the origin. *Queueing Systems: Theory and Applications* **2**, 41–65.

BACCELLI, F. AND MASSEY, W. A. (1988) A transient analysis of the two-node series Jackson network. *Math. Operat. Res.* To appear.

CHAMPERNOWNE, D. G. (1956) An elementary method of solution of the queueing problem with a single server and constant parameters. *J. R. Statist. Soc.* **B18**, 125–128.

CLARKE, A. B. (1956) A waiting line process of Markov type. *Ann. Math. Statist.* **27**, 452–459.

LEDERMAN, W. AND REUTER, G. E. H. (1954) Spectral theory for the differential equations of simple birth and death processes. *Phil. Trans. R. Soc. London* A**246**, 321–369.

PARTHASARATHY, P. R. (1987) A transient solution to an *M/M/*1 queue: A simple approach. *Adv. Appl. Prob.* **19**, 997–998.

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