A NOVEL APPROACH FOR CALCULATING BLOCKING PROBABILITIES IN THE M/G/1/N+1 QUEUE.

Matthew Roughan
Department of Applied Mathematics,
University of Adelaide

Abstract:

While the blocking probability in the M/G/1 queue with a finite waiting room is easily calculated, the analytic result for this queue is not commonly given. This is desirable if the blocking probability (or some other performance measure) is to be used in an optimisation procedure. In this paper we present a method of calculating the blocking probability in the M/G/1/N+1 queue using the probability generating function for the equilibrium system size of a modified M/G/1 queue. This method has the further advantage of giving a number of results as a byproduct of the solution of the modified M/G/1 queue.

1 Introduction

The solution to the stationary M/G/1 queue has long been known. For applications the M/G/1/N+1 queueing system is of interest. The stationary distribution for the number of customers in the system, and hence the probability of a customer being blocked, can be calculated through a recurrence relation (Cooper, 1972, page 179) amongst other means. In this paper we use a modified M/G/1 queue to calculate blocking probabilities for such a system. The advantages of this method are that it can be generalised to more complicated systems and that other quantities of interest are calculated as byproducts of the analysis.

The modified M/G/1 queue can be described as follows. The queue starts behaving as a standard M/G/1 queue. If, at the end of a service, there are more than a certain number of customers in the system (say N) the server switches to a different service-time distribution and possibly a different service discipline. It continues with this new behaviour until the system contains N or fewer customers at the end of a service whereupon it switches back to the original server behaviour. This continues for the lifetime of the queue. This modified queue can be used to model the M/G/1/N+1 queue if we take the service times in the second regime to be zero with probability one. We define this system more precisely below and provide a probability generating function for the number of customers in the system in equilibrium derived using an extension of a technique of Baccelli and Makowski (1989). Using Little's law we can find the probability of being in each of the two regimes and this can give the blocking probabilities in the M/G/1/N+1 queue.

2 Definitions

By the M/G/1 queue, we mean the single server system with a potentially infinite queue whose arrival process is an homogeneous Poisson process with rate λ and whose service times are independent, identically distributed random variables with probability distribution function $A(\cdot)$ and mean $1/\mu$. Customers who find the server unoccupied seize it immediately and hold it for their service time. Customers who find the server busy wait in the queue until they receive service. The order of service, or the queue discipline is irrelevant so long as it is noted that it is non-preemptive. In order to obtain the solution we observe the system immediately after services. By PASTA (Wolff, 1989) and the fact that in equilibrium arrivals to the queue see the same distribution that departures leave (Cooper, 1972) the equilibrium distribution of customers in the embedded process is the same as the equilibrium distribution for the queueing process. The probability generating function for the equilibrium behaviour of the system is then

$$g(z) = (1-\rho)\frac{a(z)(1-z)}{a(z)-z},$$

where $\rho = \lambda/\mu$ is the traffic intensity and $a(z)$ is the probability generating function for the number of arrivals during one service time, which given in terms of the Laplace-Stieltjes transform of $A(\cdot)$ is

$$a(z) = A^*(\lambda(1-z)).$$

The M/G/1/N+1 queue is identical to this except that if a customer arrives while there are $N+1$ customers in the system the arriving customer is blocked. By blocked we mean that it leaves the system and does not return. Note that if we consider the process at departure epochs then there are now two types of departures. There are those who depart after receiving service, these may not leave more than N customers in the system as they depart. Secondly there are those customers who leave the system having been blocked. These customers always leave $N+1$ customers in the system upon departure. When we consider the process embedded at departure epochs we include both types of departure.

We also use the solution to a M/G/1 queue modified as follows. The arrivals are again Poisson with rate λ. If there are no more than than N customers in the system immediately before a service, the service time takes distribution $A(\cdot)$ and we say the process is in regime A while if there are more than N customers in the system immediately before the service begins the service time takes distribution $B(\cdot)$ and we say the process is in regime B. Further the service-discipline may be different in the two regimes. We take ρ_a and ρ_b to be the traffic intensities during the respective regimes and note that the queue will be stable if and only if $\rho_b \leq 1$. In this case we can use a martingale technique devised by Baccelli and Makowski (1985,1989) to derive the probability generating function for this process in equilibrium as

$$E\left[z^X\right] = \frac{1}{m}\left[\frac{b(z)(1-z)+\{b(z)-a(z)\}R_N(z)}{b(z)-z}\right],$$

where

$$\begin{aligned} a(z) &= A^*(\lambda(1-z)), \\ b(z) &= B^*(\lambda(1-z)) \end{aligned}$$

and

$$R_N(z) = \left(\mathbf{e}_1 + \left(\frac{h_1}{1-h}\right)\mathbf{e}_N\right)(\mathbf{I}-\mathbf{P}_N)^{-1}\mathbf{z}^t.$$

Here $\mathbf{z} = (z, z^2, \ldots, z^N)$ and

$$\begin{aligned} \mathbf{P}_N &= \begin{pmatrix} a_1 & a_2 & a_3 & \cdots & a_{N-1} & a_N \\ a_0 & a_1 & a_2 & \cdots & a_{N-2} & a_{N-1} \\ 0 & a_0 & a_1 & \cdots & a_{N-3} & a_{N-2} \\ & & & \vdots & & \\ 0 & 0 & 0 & \cdots & a_0 & a_1 \end{pmatrix}, \\ h_1 &= 1 - a_0\mathbf{e}_1(\mathbf{I}-\mathbf{P}_N)^{-1}\mathbf{e}_N, \\ h &= 1 - a_0\mathbf{e}_N(\mathbf{I}-\mathbf{P}_N)^{-1}\mathbf{e}_N. \end{aligned}$$

The mean length of the busy period is m is given by

$$m = \frac{1+(\rho_a-\rho_b)R_N(1)}{1-\rho_b}.$$

This is a generalisation of a result of Roughan (1993).

The above can model the M/G/1/N+1 queue as follows. We take the service-times in regime B to be zero with probability one. The service discipline in regime A of our modified process is the same of that of the M/G/1/N+1 queue. In regime B the service discipline becomes last in first out. We do not need to specify this as non-preemptive as the service times are all zero. Thus the system discards customers who arrive at an overful queue. One point to note is that unlike the M/G/1/N+1 queue the departing customers may leave more than $N+1$ customers behind in the system during regime B. This is however not a problem if we note that any customer who leaves more than N customers behind when it leaves must be a discarded or blocked customer. Thus in order to calculate the blocking probability we need merely work out the probability of being in regime B. This we do using Little's law in the next section.

3 Results

Little's law (1961) states

$$L = \lambda W, \tag{1}$$

where L is the mean number of customers in the system, λ is the arrival rate to the system and W is the mean time spent by a customer in the system. If we apply this to the server alone we can see that L is the probability that there is a customer in the system and W is the mean service time. We take the probability of more than N customers being in the system to be p_N. This is also the probability that the system is in the second regime. Hence

$$\begin{aligned} L &= p\{X \neq 0\} = 1 - \frac{1}{m}, \\ W &= \frac{1-p_N}{\mu_a} + \frac{p_N}{\mu_b}, \end{aligned}$$

from which we derive

$$L = \frac{\rho_b + (\rho_a - \rho_b)R_N(1)}{1 + (\rho_a - \rho_b)R_N(1)}, \tag{2}$$

$$\lambda W = p_N(\rho_b - \rho_a) + \rho_b. \tag{3}$$

Substituting (2) and (3) into (1) and rearranging we get an equation

$$p_N = \frac{1 + (\rho_a - 1)R_N(1)}{1 + (\rho_a - \rho_b)R_N(1)}, \tag{4}$$

for p_N (when $\rho_b \neq \rho_a$). Thus we have the probability of being in the second regime of the system. If services during the second regime take zero time with probability one we get $\rho_b = 0$ and hence

$$p_N = \frac{1 + (\rho_a - 1)R_N(1)}{1 + \rho_a R_N(1)}.$$

This then is the blocking probability in the M/G/1/N+1 queue. Further the mean length of the busy period in such a system is given by m and so the mean length of the busy period of this system is

$$m = 1 + \rho_a R_N(1).$$

Interestingly if we look at the case when $\rho_b = \rho_a$ then the solution for the modified M/G/1 queue is the same as that of the standard M/G/1 queue with service-time distribution $A(\cdot)$ regardless of the actual distribution of $B(\cdot)$. If we consider p'_N the probability of having more than N customers in the M/G/1 queue, this can be obtained by taking a limiting set of distributions $B(\cdot)$ such that $\rho_b \to \rho_a$. A variation of the argument above gives

$$p'_N = 1 + (\rho_a - 1) R_N(1)$$

as long as $\rho_a \leq 1$. This probability is clearly not the same as the blocking probability of the M/G/1/N+1 queue. However

$$p_N = \frac{p'_N}{m}.$$

The number of customers consecutively blocked may also be of interest in such systems, and can easily be obtained through this method. In the modified M/G/1 queueing system this will be given by the number of customers who are served consecutively in the second regime which is simply the number in the system at the switch point between regimes one and two since during the second regime we take service times to be zero with probability one. This is intrinsically related to $R_N(z)$ (Roughan, 1993). Thus we can obtain a measure for the number of consecutively blocked customers.

4 Conclusion

The results given are of modest interest in themselves as they have previously been obtained by other means. However, the original results of Baccelli and Makowski have been expanded in Baccelli and Makowski (1991) to systems with Markov modulated Poisson processes as their arrival process. There are a number of other systems which are tractable using this technique (Roughan, 1993). Thus the present method may be used for dealing with these more generalised queues when they are restricted to limited waiting rooms. The additional information gained using this technique make this an attractive approach. Thus the method holds some promise for further work.

5 Acknowledgements

This work has been funded by an APRA and the Teletraffic Research Center at Adelaide University. I would like to thank my supervisor As. Prof. C.E.M. Pearce for many helpful comments and suggestions.

6 References

F.Baccelli and A.M.Makowski, (1985), Direct Martingale Arguments For Stability: the M/GI/1 case, *Systems Control Letters*, **6**, 181-186.

F.Baccelli and A.M.Makowski, (1989), Dynamic, Transient and Stationary Behaviour of the M/GI/1 Queue via Martingales, *Annals of Probability*, **17**, 4, 1691-1699.

F.Baccelli and A.M.Makowski, (1991), Martingale relations for the M/GI/1 queue with Markov modulated Poisson input, *Stochastic Processes and their Applications*, **38**, 99–133.

R.B. Cooper, (1972), *Introduction to Queueing Theory*, The Macmillian Company.

J.D.C. Little, (1961), A proof of the queueing formula: $L = \lambda W$, *Operations Research*, **9**, 3, 383-87.

M. Roughan, (1993) A martingale technique for the solution of a modified M/G/1 queue, *preprint.*

R.W.Wolff, (1989), *Stochastic Modelling and the theory of Queues*, Prentice Hall International.