

DISTRIBUTION OF THE NUMBER OF CUSTOMERS SERVED IN AN M/G/1 RETRIAL QUEUE

M. J. LOPEZ-HERRERO,* *Universidad Complutense de Madrid*

Abstract

We present a recursive method of computation for the probability that at most k customers were served during the busy period of an M/G/1 retrial queue.

Keywords: Retrial queue; busy period; number of customers served

AMS 2000 Subject Classification: Primary 60K25

1. Introduction

In retrial systems an arriving customer who finds the server busy is obliged to leave the service area and join a pool of unsatisfied customers called the ‘orbit’. From the orbit, each customer reapplies for service after a random amount of time. Retrial queues have been used to model problems in telephone, computer and communication systems. A review of the main results and literature and a discussion of practical situations can be found in the monograph [9]. In addition, a complete bibliography is given in [1] and [2].

As occurs in classical models (see for instance [5], [7], [10], [11], and [13]), the system performance can be described in terms of its main characteristics, such as limit distribution, waiting time, busy period, number of customers served, etc. In this paper, an M/G/1 queue with exponential retrial times is considered. In particular, we focus on I , the number of customers served during the busy period, L ; the system busy period is defined to be the period of time between an epoch when an arriving customer finds an empty system and the first departure epoch at which the system is empty again.

The analysis of I and L on retrial systems has been the subject of several papers. The joint Laplace transform was obtained by Falin in [8], the structure of the busy period was studied by Artalejo and Falin in [3], and in [4] Artalejo and Lopez-Herrero determined a closed expression for $E[L^2]$. When exponential service times are considered, Choo and Conolly [6] developed a method for the calculation of the moments of L . Related to the number of customers served, the theoretical solution in terms of its Laplace transform presents important practical limitations. However, explicit formulae for the probabilities $P\{I = k\}$ when $k \leq 4$, for $E[I^2]$, and an estimation of maximum entropy for the distribution of I were obtained in [12].

This paper extends the previous work by reducing the distribution of I to the solution of a set of linear equations. Some numerical examples are also given.

2. Model description

We consider a single server queue to which customers arrive according to a Poisson process with rate λ . Any arriving customer who finds the server busy joins the orbit. Each individual

Received 18 February 2002; revision received 17 April 2002.

* Postal address: Escuela de Estadística, Universidad Complutense de Madrid, 28040 Madrid, Spain.

Email address: lherrero@estad.ucm.es

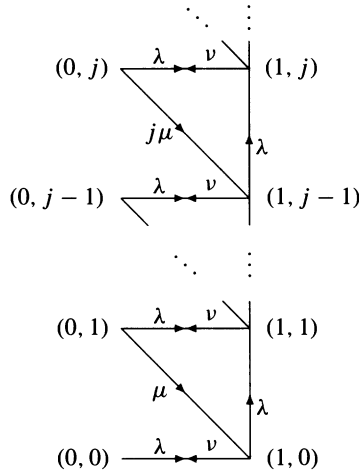


FIGURE 1: State space and transitions.

customer in orbit generates a Poisson stream of rate $\mu > 0$ of repeated requests for service. The service times are independent with common probability distribution function $B(x)$ (where $B(0) = 0$), Laplace–Stieltjes transform $\beta(\theta)$, and mean β_1 . The flow of primary arrivals, the intervals between successive repeated attempts, and service times are assumed to be mutually independent.

At any arbitrary time t , the system is described in terms of the bidimensional process $Y(t) = (C(t), N(t))$, where $C(t)$ is 0 or 1, according to whether the server is free or busy, and $N(t)$ is the number of customers in orbit at time t . As is well known, the stability condition $\rho = \lambda\beta_1$ guarantees that the limiting distribution of the process $Y(t)$ exists and is positive. Figure 1 illustrates the transitions among the states in the process $Y(t)$ for the pure Markovian case (that is, $B(x) = 1 - \exp\{-\nu x\}$, $\nu = 1/\beta_1$).

3. Number of customers served: recurrent relation

In what follows, we assume that $\rho < 1$. We consider the Markov chain embedded at departures epochs. Let us assume that at time $t = 0$ the system is empty, i.e. $C(0) = N(0) = 0$, and a single primary customer arrives at time $t = 0$. Then note that the busy period corresponds to the first return time of the state 0 in that chain.

Our main goal is to obtain a recursive scheme to compute the probability distribution of I . Firstly, we recall that the probability density $\{c_k\}$ of the number of arrivals during a service time is given by

$$c_k = \int_0^\infty e^{-\lambda t} \frac{(\lambda t)^k}{k!} dB(t), \quad k \geq 0.$$

At any departure epoch we observe the number of customers in orbit and define $x_i^{(k)}$, $1 \leq i \leq k$, to be the conditional probability that, starting in the state i , the embedded chain reaches state 0 and the server gives service to at most k more customers. In other words, $x_i^{(k)}$ is the probability that, starting at a departure which leaves i customers behind, in the rest of the busy period up to k customers were served.

The probability distribution of I has the following expression in terms of $\{x_i^{(k)}\}$:

$$P\{I = 1\} = c_0, \tag{1}$$

$$P\{I \leq k\} = c_0 + \sum_{j=1}^{k-1} c_j x_j^{(k-1)}, \quad k > 1. \tag{2}$$

The equation (1) is obtained by noting that no other customer arrives during the service time of the unique customer served. When $k > 1$, the event $\{I \leq k\}$ occurs when at least $j \leq k - 1$ customers arrive during the first service time and subsequently the system empties out by giving service to up to $k - 1$ more customers.

The probabilities $\{x_i^{(k)}\}$ are obtained from a recursive scheme which is derived as follows. For $k = 1$, the preceding departure left one customer in orbit. To serve at most one customer during the rest of the busy period it is necessary that the customer in orbit was the next entering service and, moreover, during that customer's service no other arrival occurs. Consequently, $x_1^{(1)}$ satisfies the equation

$$x_1^{(1)} = \frac{\mu}{\lambda + \mu} c_0. \tag{3}$$

For $k \geq 2$ a similar argument is used. We distinguish between the alternatives where the next customer served is a primary arrival or comes from the orbit. In either case, we partition the subsequent event on the number of arrivals that may occur during that service. The following linear equations are obtained in this way for the quantities $\{x_i^{(k)}\}$, $1 \leq i \leq k$.

$$x_i^{(k)} = \begin{cases} \frac{\mu}{\lambda + \mu} c_0 + \sum_{j=1}^{k-1} \frac{\lambda c_{j-1} + \mu c_j}{\lambda + \mu} x_j^{(k-1)} & \text{for } i = 1, \\ \frac{i\mu}{\lambda + i\mu} c_0 x_{i-1}^{(k-1)} + \sum_{j=i}^{k-1} \frac{\lambda c_{j-i} + i\mu c_{j-i+1}}{\lambda + i\mu} x_j^{(k-1)} & \text{for } 2 \leq i \leq k - 1, \\ \frac{k\mu}{\lambda + k\mu} c_0 x_{k-1}^{(k-1)} & \text{for } i = k. \end{cases} \tag{4}$$

For any fixed level $k \geq 1$, the equations (3) and (4) give a recursive scheme which can be written in matrix form as

$$\mathbf{x}^k = \mathbf{T} \tilde{\mathbf{x}}^{k-1},$$

where \mathbf{x}^k is the vector with components $\{x_i^{(k)}, 1 \leq i \leq k\}$, $\tilde{\mathbf{x}}^{k-1}$ represents the column vector $(1, \mathbf{x}^{k-1})^\top$ and \mathbf{T} is the $k \times k$ matrix

$$\mathbf{T} = \begin{pmatrix} a_1 & b_{11} & b_{12} & \cdots & b_{1,k-1} \\ 0 & a_2 & b_{22} & \cdots & b_{2,k-1} \\ 0 & 0 & a_3 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & b_{k-1,k-1} \\ 0 & 0 & \cdots & 0 & a_k \end{pmatrix}$$

with

$$a_i = \frac{i\mu c_0}{\lambda + i\mu} \quad \text{for } 1 \leq i \leq k,$$

$$b_{ij} = \frac{\lambda c_{j-i} + i\mu c_{j-i+1}}{\lambda + i\mu} \quad \text{for } 1 \leq i \leq j \leq k - 1.$$

In order to check the validity of the recursive scheme, we can observe that (1) and (2) when $k \leq 4$ lead, after some algebra, to the explicit expressions (3.1)–(3.4) in [12].

On the other hand, under high retrial intensity, the M/G/1 retrial queue approaches the classical M/G/1 queue (without retrials) operating under random order discipline [9]. Consequently, when $\mu \rightarrow \infty$, the random variable I should have the same probabilistic behaviour as the number of customers served during the busy period of the classic M/G/1 queue. In this sense, numerical results in the next section show that (3) and (4) give probabilities which are asymptotically consistent with formula (5.157) in [11].

4. Numerical work

In this section, we present numerical results for the distribution of I . We consider that service times are distributed according to an Erlang-2, an exponential or a hyperexponential random variable with parameters chosen in such a manner that the mean of service times β_1 is 1 in all cases, and the coefficient of variation is 0.8, 1, and 2 respectively.

Figures 2 and 3 show the effect of μ on $E[I]$ and $(\text{var}(I))^{1/2}$ respectively. We present three groups of curves corresponding to $\lambda = \rho = 0.25, 0.5, 0.75$. Continuous curves are associated with Erlang-2 distributed service times, while the curves of circles and rhombuses correspond to hyperexponential service times. Curves associated to the Markovian model are not represented, but their graphs have a similar shape, being between those represented for each level of ρ .

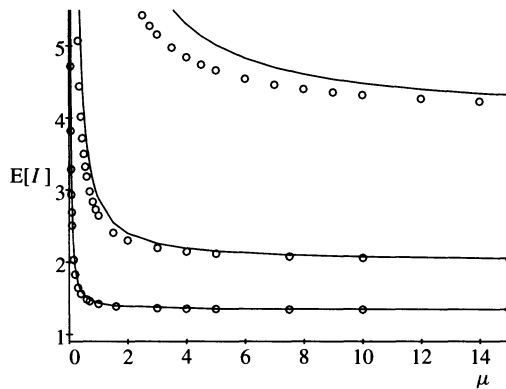


FIGURE 2: $E[I]$ versus μ .

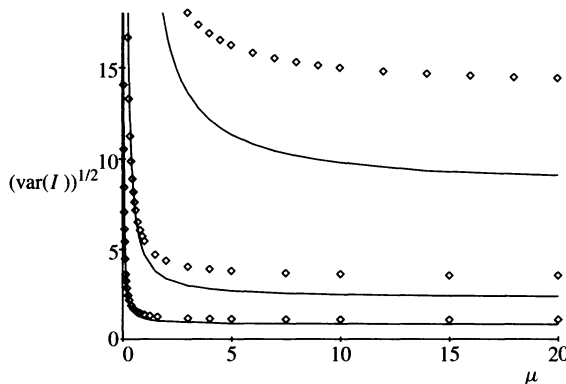


FIGURE 3: $(\text{var}(I))^{1/2}$ versus μ .

TABLE 1: Customers served distribution.

μ	0.01	0.1	0.5	2	5	10	100
$I = 1$	0.647249	0.647249	0.647249	0.647249	0.647249	0.647249	0.647249
	0.666666	0.666666	0.666666	0.666666	0.666666	0.666666	0.666666
	0.727272	0.727272	0.727272	0.727272	0.727272	0.727272	0.727272
$I = 2$	0.003136	0.026663	0.079990	0.127984	0.145436	0.152362	0.159184
	0.002904	0.024691	0.074074	0.118518	0.134680	0.141093	0.147411
	0.002514	0.021370	0.064112	0.102579	0.116567	0.122118	0.127586
$I = 3$	0.002029	0.016986	0.049320	0.064377	0.068635	0.069964	0.071071
	0.001936	0.016199	0.043895	0.060283	0.063874	0.064929	0.065760
	0.001821	0.014841	0.037796	0.048228	0.049327	0.049350	0.049139
$I = 4$	0.001332	0.011847	0.031765	0.039568	0.039843	0.039598	0.039169
	0.001306	0.011505	0.030279	0.037199	0.037328	0.037063	0.036639
	0.001331	0.010968	0.025583	0.028040	0.026850	0.026143	0.025338
$I = 5$	0.000886	0.008835	0.023852	0.026838	0.025655	0.024926	0.024083
	0.000892	0.008682	0.022762	0.025321	0.024236	0.023589	0.022852
	0.000982	0.008516	0.018824	0.018372	0.016849	0.016149	0.015428
$I = 6$	0.000597	0.006927	0.018919	0.019299	0.017584	0.016737	0.015836
	0.000616	0.006862	0.018046	0.018281	0.016769	0.016037	0.015266
	0.000730	0.006874	0.014622	0.013003	0.011590	0.011008	0.010440
$I = 7$	0.000407	0.005640	0.015557	0.014421	0.012565	0.011738	0.010898
	0.000430	0.005620	0.014823	0.013724	0.012107	0.011397	0.010681
	0.000548	0.005722	0.011791	0.009708	0.008487	0.008018	0.007576
$I = 8$	0.000281	0.004728	0.013120	0.011073	0.009251	0.008494	0.007751
	0.000304	0.004734	0.012485	0.010592	0.009011	0.008361	0.007722
	0.000414	0.004881	0.009771	0.007535	0.006496	0.006116	0.005766
$I = 9$	0.000196	0.004057	0.011270	0.008676	0.006965	0.006292	0.005651
	0.000216	0.004077	0.010711	0.008346	0.006862	0.006283	0.005732
	0.000315	0.004248	0.008266	0.006020	0.005134	0.004821	0.004538
$I \geq 10$	0.343882	0.267063	0.111954	0.040511	0.026812	0.022636	0.019105
	0.324724	0.250960	0.106255	0.041065	0.028462	0.024577	0.021262
	0.264068	0.195303	0.081958	0.039238	0.031423	0.028999	0.026911

The behaviour of $E[I]$ and $(\text{var}(I))^{1/2}$ is similar, increasing with increasing traffic intensity and decreasing retrial rate. On the other hand, at any fixed μ we observe that differences between the mean number of customers served and in the variability in both models are larger for high traffic intensities. Moreover, for values of μ , not really small, we have that the mean number of customers served in the Erlangian model is bigger, and with less variability, than in the hyperexponential model; in addition, when $\mu \rightarrow \infty$, both $E[I]$ and $(\text{var}(I))^{1/2}$ are consistent with the characteristics of the corresponding standard model (without retrials).

Our numerical work also shows that the first and second order moments, $E[I]$ and $E[I^2]$, agree with the expressions (3.15) and (3.22) in [12].

The distribution of I is displayed on Table 1. For each μ and each I , the three values given are, from top to bottom, the probabilities for the Erlangian, Markovian, and hyperexponential

models respectively. For this group of results the arrival rate was chosen so that the traffic intensity is $\rho = \frac{1}{2}$ in all three models.

The behaviour of the probability distribution is essentially the same for these models. The most frequent number of customers served during the busy period is 1 and, according to (1), such a probability does not depend on the retrial rate. The high dispersion when μ is small, indicated by Figure 3, is due to the smoothness of those distributions for $I > 1$. The initial idea of the decreasing behaviour of the probability that at least k , $k > 1$, customers were served in the busy period as a function of μ can be visualized in Figures 2 and 3 and confirmed by the numerical results in Table 1.

Acknowledgement

This work was supported by the DGES through project 98-0837.

References

- [1] ARTALEJO, J. R. (1999). Accessible bibliography on retrial queues. *Math. Comput. Modelling* **30**, 1–6.
- [2] ARTALEJO, J. R. (1999). A classified bibliography of research on retrial queues: Progress in 1990–1999. *Top* **7**, 187–211.
- [3] ARTALEJO, J. R. AND FALIN, G. I. (1996). On the orbit characteristics of the M/G/1 retrial queue. *Naval Res. Logistics* **43**, 1147–1161.
- [4] ARTALEJO, J. R. AND LOPEZ-HERRERO, M. J. (2000). On the busy period of the M/G/1 retrial queue. *Naval Res. Logistics* **47**, 115–127.
- [5] ARTALEJO, J. R. AND LOPEZ-HERRERO, M. J. (2001). Analysis of the busy period for the M/M/c queue: An algorithmic approach. *J. Appl. Prob.* **38**, 209–222.
- [6] CHOO, Q. H. AND CONOLLY, B. (1979). New results in the theory of repeated orders queueing systems. *J. Appl. Prob.* **16**, 631–640.
- [7] CONOLLY, B. AND CHAN, J. (1977). Generalised birth and queueing processes: Recent results. *J. Appl. Prob.* **9**, 125–140.
- [8] FALIN, G. I. (1979). A single-line system with secondary orders. *Eng. Cybern.* **17**, 76–83.
- [9] FALIN, G. I. AND TEMPLETON, J. G. C. (1997). *Retrial Queues*. Chapman and Hall, London.
- [10] GHARAMANI, S. (1990). On remaining full busy periods of GI/G/c queues and their relation to stationary point processes. *J. Appl. Prob.* **27**, 232–236.
- [11] KLEINROCK, L. (1975). *Queueing Systems*, Vol. 1, *Theory*. John Wiley, New York.
- [12] LOPEZ-HERRERO, M. J. (2002). On the number of customers served in the M/G/1 retrial queue: First moments and maximum entropy approach. *Comput. Operat. Res.* **29**, 1739–1757.
- [13] NATVIG, B. (1975) On the waiting-time and busy period distributions for a general birth-and-death queueing model. *J. Appl. Prob.* **12**, 524–532.