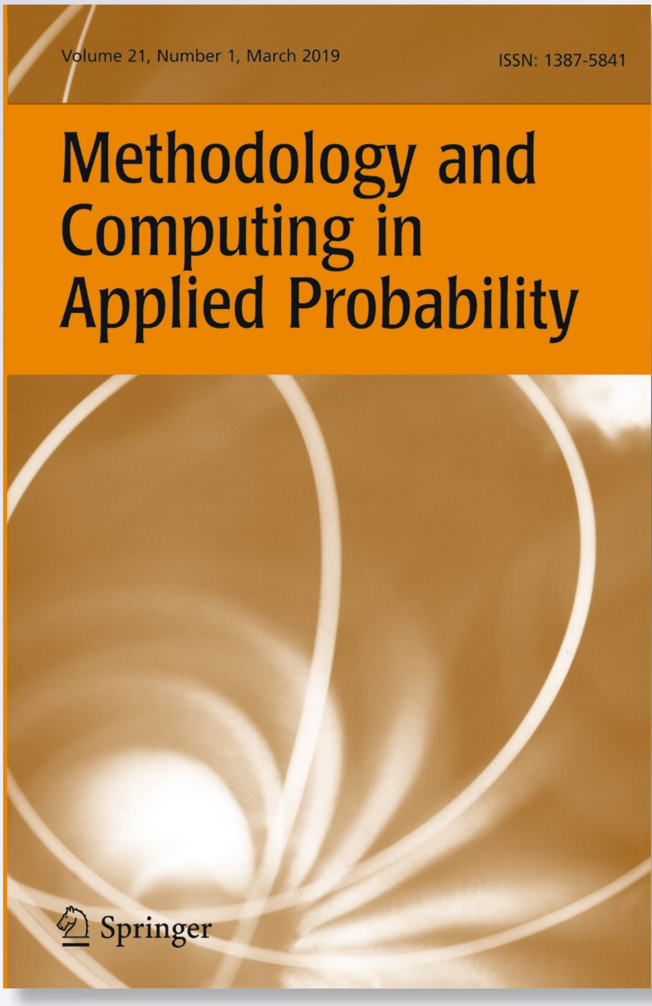
The Queue G e o/G/1/N +1 Revisited

M. L. Chaudhry & Veena Goswami

Methodology and Computing in Applied Probability

ISSN 1387-5841
Volume 21
Number 1

Methodol Comput Appl Probab (2019)
21:155-168
DOI 10.1007/s11009-018-9645-0



Methodol Comput Appl Probab (2019) 21:155–168
https://doi.org/10.1007/s11009-018-9645-0

The Queue $Geo/G/1/N+1$ Revisited

M. L. Chaudhry[1] · **Veena Goswami**[2]

Received: 19 September 2017 /
Accepted: 28 May 2018 / Published online: 4 June 2018

Abstract This paper presents an alternative steady-state solution to the discrete-time $Geo/G/1/N+1$ queueing system using roots. The analysis has been carried out for a late-arrival system using the imbedded Markov chain method, and the solutions for the early arrival system have been obtained from those of the late-arrival system. Using roots of the associated characteristic equation, the distributions of the numbers in the system at various epochs are determined. We find a unified approach for solving both finite- and infinite-buffer systems. We investigate the measures of effectiveness and provide numerical illustrations. We establish that, in the limiting case, the results thus obtained converge to the results of the continuous-time counterparts. The applications of discrete-time queues in modeling slotted digital computer and communication systems make the contributions of this paper relevant.

Keywords Discrete-time · Finite buffer · Roots · Queue

Mathematics Subject Classification 2010 60K25 · 68M20 · 90B22

✉ Veena Goswami
veena_goswami@yahoo.com

M. L. Chaudhry
chaudhry-ml@rmc.ca

1 Department of Mathematics and Computer Science, Royal Military College of Canada, P.O. Box 17000, STN Forces, Kingston, ON K7K 7B4, Canada

2 School of Computer Applications, Kalinga Institute of Industrial Technology, Bhubaneswar, 751024, India

1 Introduction

The last two decades have seen a rise in the analysis and applications of discrete-time queueing systems in various fields. With the advent of new technologies, the popularity of discrete-time queueing systems can be well understood. However, when compared to continuous-time queueing systems, there is less literature on discrete-time queueing systems due to the perceived difficulty in its analysis. A key complication during formulation of discrete-time queueing systems is that the systems allow for simultaneous arrivals and departures. Despite these complications, discrete-time systems are adopted in applications that are inherently discrete, like when modeling computer and communication systems. Extensive analysis of discrete-time queueing systems and their applications have been reported in Alfa (2016), Bruneel and Kim (2012), Takagi (1993a), and Woodward (1994).

In real-life, most queues have finite buffer space to store incoming packets. An important design problem in such queues is determining a buffer size such that the loss probability is below a pre-specified value. The results for queues with finite buffer-space hold true for any value of the traffic intensity ρ. However, when we consider queues with infinite buffer space, the results are good only for $\rho < 1$. From the server's point of view, the queueing systems with finite buffer space operate most efficiently when ρ is near unity. In such situations, one of the main concerns of the system designers is the estimation of blocking probability, which, in general, is kept small to avoid loss of packets. Further, one may obtain the results of infinite buffer space by taking a large value of the finite buffer space. Gravey and Hébuterne (1992) investigated simultaneity in discrete-time Bernoulli inputs single server queues with finite- and infinite- buffers. Lee et al. (1999) computed the state probabilities of the $Geo/G/1/K$ queue directly from the state equations. However, they did not obtain the system-length distributions at a pre-arrival epoch or at a random epoch. The analytic study related to the impact of the size of the waiting room on a queue with a random input process has been carried out in Finch (1958). Computational analysis of continuous-time $M/G/1/N+1$ and $GI/M/1/N+1$ queues using roots has been presented in Chaudhry et al. (1991). Historically, when MAPLE and Mathematica could not find a large number of roots (they do now), a software package called QROOT developed by Chaudhry (1992) was used by him and his collaborators to find a large number of roots and use them in solving several queueing models. The algorithm for finding such roots is available in Chaudhry et al. (1990). It may be remarked here that MAPLE can now not only find roots that are close to each other (a concern expressed by several researchers including (Akar 2006) and Neuts (1981)), but even the repeated roots.

In discrete-time queues, time is assumed to be divided into intervals of the same length called slots. We solve finite waiting space $Geo/G/1/N+1$ for both late and early arrival systems using the roots method which computes numerical results efficiently. Limited work has been reported on the computational aspect of the $Geo/G/1/N+1$ system. The roots method has never been applied to discrete-time finite-buffer space queues. Another characteristic of applying the roots method is that the number of roots to be found is independent of the number of waiting spaces. Also, it shows that both the systems $Geo/G/1/N+1$ and $Geo/G/1$ can be solved using the roots method leading to a unified approach for solving both the systems, see Chaudhry (2000) for $Geo/G/1$. Further, the efficiency of the roots method relative to matrix-analytic has been shown in several papers, see, e.g., a recent paper (and references therein) by Singh et al. (2016). In this paper, we find closed-form expressions for the limiting distribution of the number in the system in terms of roots of the characteristic equation. The outcomes for $Geo/G/1/N+1$ can be computed for any service-time distribution having a rational probability generating function. We obtain steady-state system-length distributions at various epochs using roots of the associated characteristic equation. We also examine the measures of effectiveness for both the systems, and provide numerical examples for the same.

The organization of this paper is as follows. Section 2 presents the analysis of $Geo/G/1/N+1$ system for both LAS-DA and EAS policies. We obtain steady-state system-length distributions at post-departure, random, pre-arrival and outside observer's epochs. Section 3 discusses some performance measures of the system. Section 4 presents numerical results for both LAS-DA and EAS. Section 5 shows that in the limiting case our analytic results tend to their continuous-time counterparts. Section 6 draws the conclusions.

2 The $Geo/G/1/N+1$ System

A discrete-time queueing system is specified by time-slotted service, where the events (arrival of packets and their onward transmissions) may occur simultaneously around slot boundaries. In the case of simultaneity, their order may be taken care of by either departure-first (DF) or arrival-first (AF) management policies, which are also known as early arrival system (EAS) and late arrival system with delayed access (LAS-DA), respectively. According to AF policy, arrivals take precedence over departures, while under DF policy the opposite is true. The inter-arrival times are independent and geometrically distributed as $a_n = \lambda\bar{\lambda}^{n-1},\ 0 < \lambda < 1,\ n \geq 1$ with mean inter-arrival time $a = 1/\lambda$. We denote $\bar{x} = 1-x$ for any real number $x \in [0, 1]$. The service times $\{B_n,\ n \geq 1\}$ are independent and identically distributed random variables with probability mass function (pmf) $b_i = P(B_n = i)$, $i \geq 1$, corresponding probability generating function (pgf) $B(z) = \sum_{i=1}^{\infty} b_i z^i$ and mean service time $E[S] = B^{(1)}(1) = b = 1/\mu$, where $B^{(n)}(k)$ is the n-th derivative of $B(z)$ with respect to z at $z = k$. The traffic intensity ρ is equal to λ/μ. The system has a finite buffer of capacity $N+1$ to retain incoming arrivals. It is assumed that inter-arrival times and service times are mutually independent and the service discipline follows first-come, first-served (FCFS).

2.1 The LAS-DA System

In this subsection, we study a discrete-time $Geo/G/1/N+1$ queue under the late arrival system with delayed access. In LAS-DA, potential arrivals take place in the interval $(t-, t)$ and potential departures take place in the interval $(t, t+)$. Various time epochs at which events occur in LAS-DA are shown in Fig. 1.

Let L_n^+ be the system length left by the departing nth customer and let A_{n+1} be the number of customers arriving during the service time of the $(n+1)$th customer. The discrete

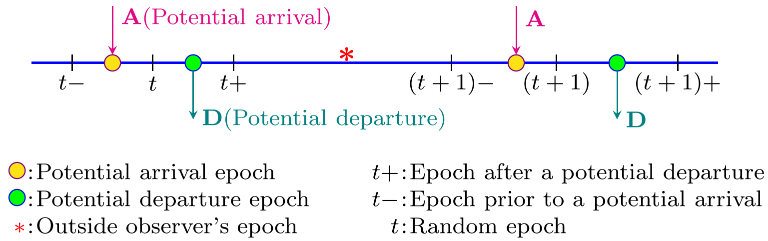


Fig. 1 Various time epochs in LAS-DA

time process N_n^+ forms a Markov chain, L_{n+1}^+ may be expressed in terms of L_n^+ and of a random variable A_{n+1} which is independent of L_n^+. Then

$$L_{n+1}^+ = \begin{cases} min(L_n^+ - 1 + A_{n+1}, N), & \text{if } L_n^+ \geq 1, \\ min(A_{n+1}, N), & \text{if } L_n^+ = 0. \end{cases} \tag{1}$$

Let k_j be the probability that j arrivals occur during a service time. For all $n \geq 1$,

$$k_j = \lim_{n\to\infty} P(A_{n+1} = j) = \sum_{k=j}^{\infty} b_k \binom{k}{j} \lambda^j \bar{\lambda}^{k-j}, \;\; j \geq 0.$$

Define the pgf of the sequence $\{k_j, \;\; j = 0, 1, \ldots\}$ by $K(z) = \sum_{j=0}^{\infty} k_j z^j$. Thus

$$\begin{aligned} K(z) &= \sum_{j=0}^{\infty}\sum_{k=j}^{\infty} b_k \binom{k}{j} \lambda^j \bar{\lambda}^{k-j} z^j = \sum_{k=0}^{\infty} b_k \sum_{j=0}^{k} \binom{k}{j} (\lambda z)^j \bar{\lambda}^{k-j} \\ &= \sum_{k=0}^{\infty} b_k (\bar{\lambda} + \lambda z)^k = B(\bar{\lambda} + \lambda z). \end{aligned} \tag{2}$$

The one-step transition probabilities for the underlying Markov chain are obtained as

$$p_{ij} = Pr\{L_{n+1}^+ = j | L_n^+ = i\} = \begin{cases} k_j, & 0 \leq j \leq N-1, i = 0 \\ k_{j-i+1}, & i-1 \leq j \leq N-1, 1 \leq i \leq N \\ \sum_{l=N}^{\infty} k_l, & j = N, i = 0 \\ \sum_{l=N-i+1}^{\infty} k_l, & j = N, 1 \leq i \leq N. \end{cases} \tag{3}$$

leading to the transition probability matrix P for the Markov chain as

$$P = (p_{ij}) = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ N \end{array} \begin{array}{c} \begin{array}{ccccccc} 0 & 1 & 2 & 3 & \ldots & N-1 & N \end{array} \\ \begin{pmatrix} k_0 & k_1 & k_2 & k_3 & \ldots & k_{N-1} & 1 - \sum_{i=0}^{N-1} k_i \\ k_0 & k_1 & k_2 & k_3 & \ldots & k_{N-1} & 1 - \sum_{i=0}^{N-1} k_i \\ 0 & k_0 & k_1 & k_2 & \ldots & k_{N-2} & 1 - \sum_{i=0}^{N-2} k_i \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & k_0 & 1 - k_0 \end{pmatrix} \end{array}.$$

Let $\mathbf{p}^+ = (p_0^+, p_1^+, \ldots, p_N^+)$ be the limiting distribution of the imbedded chain. Using the transition probabilities of Eq. 3 in the system of equations $\mathbf{p}^+ P = \mathbf{p}^+$, the steady-state probabilities at departure instants are as follows

$$p_n^+ = p_0^+ k_n + \sum_{j=1}^{n+1} p_j^+ k_{n-j+1}, \;\; 0 \leq n \leq N-1, \tag{4}$$

$$p_N^+ = p_0^+ \sum_{n=N}^{\infty} k_n + \sum_{j=1}^{N} p_j^+ \sum_{n=N-j+1}^{\infty} k_n, \tag{5}$$

with the normalization condition $\sum_{j=0}^{N} p_j^+ = 1$. It may be noted that Eq. 5 is expressed by the first N equations of Eq. 4, thus, Eq. 5 is redundant and it will not be considered hereafter. Let $P^+(z)$ be the pgf of the sequence p_n^+. Multiplying Eq. 4 by z^n and summing over $n = 0, 1, \ldots, N$, we get pgf as

$$P^+(z) = \sum_{n=0}^{N} p_n^+ z^n = p_0^+ \sum_{n=0}^{N-1} k_n z^n + \sum_{n=0}^{N-1} z^n \sum_{j=1}^{n+1} p_j^+ k_{n-j+1} + p_N^+ z^N$$

$$= p_0^+ \left(K(z) - \sum_{n=N}^{\infty} k_n z^n \right) + \sum_{j=1}^{N} p_j^+ z^{j-1} \left(K(z) - \sum_{l=N-j+1}^{\infty} k_l z^l \right) + p_N^+ z^N, \quad (6)$$

where the power series $K(z)$ is absolutely convergent. Simplifying Eq. 6 and taking $\hat{p}_{N+j} = (p_0^+ + p_1^+)k_{N+j} + p_2^+ k_{N+j-1} + \ldots + p_{N-1}^+ k_{j+2} + p_N^+ k_{j+1}$, we obtain

$$P^+(z) = \frac{p_0^+(1-z)K(z)}{K(z) - z} + \frac{z^{N+1}\left(\sum_{j=0}^{\infty} \hat{p}_{N+j} z^j - p_N^+ \right)}{K(z) - z}. \quad (7)$$

The left-hand side of Eq. 7 is a pgf for the number of customers in the system after a departure for $0 \le n \le N$. It is evident that p_n^+ is the coefficient of z^n, $n = 0, 1, \ldots, N$ in the first term and the second term contains higher powers of z^n, $n \ge N + 1$. The second term of right-hand side of Eq. 7 may be omitted, as it makes no contribution to the left-side part. In addition, since $K(z)$ is a probability generating function, the statement is true for any $K(z)$. Thus, p_n^+, $n = 0, 1, \ldots, N$ can be evaluated from the following

$$P^+(z) = \frac{p_0^+(1-z)K(z)}{K(z) - z}, \quad (8)$$

which is the same as pgf of an infinite buffer case (see Chaudhry (2000) or Takagi (1993a)). It may be noted that Eq. 8 represents the pgf of infinite buffer $Geo/G/1$ queue when $\rho < 1$. In the case of $Geo/G/1/N + 1$ queue, the probabilities for all values of ρ need to be computed. It may be mentioned here that if at a random epoch the number of states is $N + 1$, it will be N at a departure epoch. To derive the probabilities $\{p_n^+\}_0^N$, one of the ways is a power series expansion of the right-hand-side of Eq. 8 for a suitable region of convergence, and computation of $\{p_n^+\}$ in terms of $\{p_0^+\}$ which in turn can be computed using normalization condition. Also, without using the probability generating function one can get the solution recursively, for details see Bhat (2015). The system $Geo/G/1$ has been solved using the roots method, see Chaudhry (2000) or Kobayashi and Mark (2009). We obtain $\{p_n^+\}_0^N$ using roots and partial-fraction expansion of $P^+(z)$ in order to give, as stated earlier, a unified approach to solving the system $Geo/G/1/N+1$ as well as $Geo/G/1$ (see Chaudhry (2000)). Further, since the pgf of many distributions can be approximated by a rational function, we assume that $B(z)$ is a rational function of z given by

$$B(z) = \frac{U(z)}{V(z)}, \quad (9)$$

where $V(z)$ and $U(z)$ are polynomials of degree k and less than k, respectively. Using (9) in Eq. 8, we get

$$P^+(z) = \frac{p_0^+(z-1)f(z)}{zg(z) - f(z)}, \quad (10)$$

where

$$K(z)=\frac{U(\bar{\lambda}+\lambda z)}{V(\bar{\lambda}+\lambda z)}=\frac{f(z)}{g(z)}.$$

The denominator of Eq. 10 is a polynomial of degree $(k+1)$ with $z=1$ being a zero of both the numerator and the denominator. The term

$$zg(z)-f(z)=0, \tag{11}$$

has $(k+1)$ roots, one of the roots being 1, and the other k roots (real or complex), assumed to be distinct, denoted as $\alpha_i;\ i=1,2,\ldots,k$.

Remark If the Eq. 11 has either repeated roots or the roots that are very close to each other, we can find them using modern mathematical software packages. The MAPLE script below illustrates the computation of repeated roots for the equation

$$h(x)=(x-1)(x-2)^3(x-3)^2(x-5).$$

restart : Digits :=10: with(RootFinding):
$h:=(x-1)*(x-2)^3*(x-3)^2*(x-5);$
Analytic(h, x, re = -1 .. 10, im = -2 .. 10);
2.00000000000000, 2.00000000000000, 2.00000000000000, 3.00000000000000,
3.00000000000000, 1.00000000000000, 5.00000000000000

From Eq. 10, we obtain

$$P^{+}(z)=\frac{Tf(z)}{\prod_{i=1}^{k}(z-\alpha_i)}, \tag{12}$$

where T is a normalizing constant. If all the roots are not distinct, a little change in the partial-fraction method is required (Kobayashi et al. (2011), pp. 221). As we are discussing finite buffer queueing system, there are three cases $\rho<1$, $\rho=1$ and $\rho>1$.

- When $\rho<1$ all the roots $\alpha_i, i=1,2,\ldots,k$ remain outside the unit circle $\mid z\mid=1$.
- When $\rho=1$, one root is equal to one, say $\alpha_1=1$ and the other roots $\alpha_i, i=2,3,\ldots,k$ are outside the unit circle $\mid z\mid=1$.
- When $\rho>1$, one root is inside the interval $(0,1)$, say α_1 and the other roots $\alpha_i, i=2,3,\ldots,k$ are outside the unit circle $\mid z\mid=1$.

It is observed that when $\rho>0$ increases, one positive real root approaches the origin from right to left. The other α_i roots stay outside the unit circle. Applying partial-fraction method, we get from Eq. 12 as

$$P^{+}(z)=T\sum_{i=1}^{k}\frac{C_i}{z-\alpha_i}, \tag{13}$$

where

$$C_i=\frac{f(\alpha_i)}{\prod_{j=1,j\neq i}^{k}(\alpha_i-\alpha_j)}.$$

Using (13), we get

$$p_n^{+}=T\sum_{i=1}^{k}\frac{-C_i}{\alpha_i^{n+1}},\quad n=0,1,\ldots,N. \tag{14}$$

Applying the normalization condition $\sum_{n=0}^{N} p_n^+ = 1$, we can find the unknown T as

$$T = \begin{cases} -\left[\sum_{i=1}^{k} \frac{C_i}{\alpha_i} \times \frac{1-\alpha_i^{-(N+1)}}{1-\alpha_i^{-1}}\right]^{-1}, & \text{if } \rho \neq 1, \\ -\left[C_1(N+1) + \sum_{i=2}^{k} \frac{C_i}{\alpha_i} \times \frac{1-\alpha_i^{-(N+1)}}{1-\alpha_i^{-1}}\right]^{-1}, & \text{if } \rho = 1. \end{cases}$$

It is seen that, once all the roots are computed, we can find the distribution of the number in the system.

Remark When the service time follows geometric distribution with mean $1/\mu$, we have $K(z) = \frac{\mu(\bar{\lambda}+\lambda z)}{1-\bar{\mu}(\bar{\lambda}+\lambda z)}$. Using (8), we find $P(z) = p_0\left[1 + \frac{\lambda z}{\bar{\lambda}\mu - \lambda\bar{\mu}z}\right] = p_0\left[1 + \frac{\lambda z}{\bar{\lambda}\mu}\left\{1 - \frac{\lambda\bar{\mu}z}{\bar{\lambda}\mu}\right\}^{-1}\right]$. Expanding and collecting the coefficient of z^n, we get p_n in terms of p_0 and using normalization condition the unknown p_0 is computed.

2.1.1 System-Length Distribution at a Random Epoch

Suppose $\{p_n,\ n = 0, 1, \ldots, N+1\}$ is the probability that at a random epoch there are n customers in the system. Let $\{\tilde{p}_n,\ n = 0, 1, \ldots, N+1\}$ be the probability that an arrival at a slot boundary finds n customers in the system whether it joins the queue or not. Since the property that Bernoulli-arrivals-see-time-averages holds, we get

$$\tilde{p}_n = p_n, \quad n = 0, 1, \ldots, N+1. \tag{15}$$

As arrivals and departures take place one by one, the probability distribution $\{\tilde{p}_n\}_0^N$ for the number in the system immediately before an arrival is the same as the probability distribution $\{p_n^+\}_0^N$ for the number in the system immediately after a service-completion, excluding those that find the system blocked. In other words, $\{\tilde{p}_n\}_0^N$ is the conditional probability that an arriving customer finds the system is not blocked. Since the probability that the system is not blocked is $1 - p_{N+1}$, the effective rate of arrival is $\lambda(1 - p_{N+1})$. The mean number of busy servers (the carried load a') is $a' = \sum_{n=1}^{N+1} p_n = 1 - p_0$. The carried load a' is also defined as the fraction of the offered load ($\rho = \lambda b$) that is accepted by the system. It is also the probability that the server is busy immediately at an arbitrary slot boundary. Using the law that rate-in must equal rate-out, we get

$$\lambda(1 - p_{N+1}) = \mu(1 - p_0) = \mu\, a'. \tag{16}$$

Further, using the level crossing law which expresses that the distribution seen by an arriving customer whether he joins the queue or not is the same as that seen by departing customer, we have

$$\lambda p_n = \mu(1 - p_0)p_n^+. \tag{17}$$

Considering Eqs. 16 and 17, we get

$$p_n = \frac{(1 - p_0)p_n^+}{\rho} = \frac{a'}{\rho}p_n^+, \quad 0 \le n \le N \text{ and } p_{N+1} = 1 - \frac{a'}{\rho}. \tag{18}$$

Setting $n = 0$ in Eq. 18 and using (16), we obtain

$$a' = \frac{\rho}{p_0^+ + \rho}. \tag{19}$$

Substituting Eqs. 19 into 18, we get

$$p_n = \frac{p_n^+}{p_0^+ + \rho}, \quad n = 0, 1, \ldots, N, \tag{20}$$

$$p_{N+1} = 1 - \sum_{n=0}^{N} p_n = 1 - \frac{1}{p_0^+ + \rho}. \tag{21}$$

Thus, we can compute all the probabilities at random epoch $\{p_n,\ n = 0, 1, \ldots, N + 1\}$.

2.1.2 System-Length Distribution at a Pre-arrival Epoch

Let p_n^- $(0 \le n \le N)$ be the probability that there are n customers in the system at a slot boundary.

$$\begin{aligned} p_n^- &= P\{n \text{ in the system at } t- \mid \text{arrival occurs before the end of the slot}\} \\ &= \frac{\lambda p_n}{\sum_{i=0}^{N} \lambda p_i} = \frac{\lambda p_n}{\lambda(1 - p_{N+1})} = \frac{p_n}{1 - p_{N+1}} = p_n^+, \quad n = 0, 1, \ldots, N. \end{aligned}$$

Remark 2 When $N \to \infty$, the model reduces to $Geo/G/1/\infty$ queueing system with $p_n^- = p_n = p_n^+$.

2.1.3 System-Length Distribution at an Outside Observer's Epoch

Let p_n^o $(n = 0, 1, \ldots, N + 1)$ be the probability that there are n customers in the system at an outside observer's observation epoch. From Fig. 1, one may notice that the outside observer's observation epoch in LAS-DA system falls somewhere in the time interval $(t+, (t + 1)-)$. That is, an outside observer's observation epoch falls in a time interval that commences just after a potential departure and instantly before a potential arrival. Therefore, random epoch and outside observer's distributions are the same, that is, $p_n^o = p_n,\ 0 \le n \le N + 1$.

2.2 The Early Arrival System

In an early arrival system, a potential arrival occurs in $(t, t+)$ and a potential departure takes place in $(t-, t)$. We obtain the results for $Geo/G/1/N + 1$ queue with EAS from $Geo/G/1/N + 1$ queue with LAS-DA. Since we have already obtained the distributions of the numbers in system at various epochs in LAS-DA, by using those results we obtain similar distributions for EAS. Various-time epochs at which events occur in EAS are shown in Fig. 2.

Further, let q_n^+ $(0 \le n \le N)$ be the post-departure epoch stationary probabilities in the case of EAS. Observing Figs. 1 and 2, we can relate the departing customer's distributions

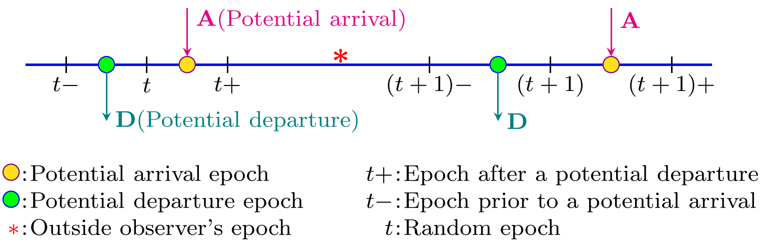


Fig. 2 Various time epochs in early-arrival system (EAS)

of LAS-DA and EAS. Noting that in the finite buffer-space the post-departure probabilities will go up to N, they are given by

$$p_0^+ = \bar{\lambda} q_0^+ \tag{22}$$
$$p_n^+ = \bar{\lambda} q_n^+ + \lambda q_{n-1}^+, \qquad n = 1, 2, \ldots, N-1, \tag{23}$$
$$p_N^+ = q_N^+ + \lambda q_{N-1}^+. \tag{24}$$

As the probabilities p_n^+ are known, we can obtain the probabilities $\{q_n^+\}_0^N$, recursively, from Eqs. 22–24.

2.2.1 System-Length Distribution at a Random Epoch

Let q_n $(0 \le n \le N+1)$ be the probability that at a random epoch there are n customers in the EAS system. Now, we develop relations among q_n and q_n^+. In this case, the input rate is $\lambda(1 - q_{N+1})$ and the output rate is the inverse of $E[S] + \frac{\bar{\lambda} q_0^+}{\lambda}$ which leads to

$$\lambda(1 - q_{N+1}) = \frac{\lambda}{\rho + \bar{\lambda} q_0^+}.$$

This gives

$$q_{N+1} = 1 - \frac{1}{\rho + \bar{\lambda} q_0^+}. \tag{25}$$

Following the argument relating p_n to p_n^+, we obtain

$$q_n = (1 - q_{N+1}) q_n^+ = \frac{q_n^+}{\rho + \bar{\lambda} q_0^+}, \quad n = 0, 1, \ldots, N. \tag{26}$$

2.2.2 System-Length Distribution at a Pre-arrival Epoch

Let q_n^- $(0 \le n \le N)$ be the probability that there are n customers in the system at a pre-arrival epoch. Then

$$\begin{aligned} q_n^- &= P\{n \text{ in the system at } t- \mid \text{arrival occurs before the end of the slot}\} \\ &= \frac{\lambda q_n}{\sum_{i=0}^N \lambda q_i} = \frac{\lambda q_n}{\lambda(1 - q_{N+1})} = \frac{q_n}{1 - q_{N+1}} = q_n^+, \quad n = 0, 1, \ldots, N. \end{aligned}$$

2.2.3 System-Length Distribution at an Outside Observer's Epoch

Let q_n^o $(0 \le n \le N+1)$ be the probability that there are n customers in the system at an outside observer's observation epoch. In EAS, the outside observer's observation epoch falls in an interval after a potential arrival and before a potential departure. The outside observer's distribution may be computed using the relation

$$q_0^o = \bar{\lambda} q_0, \tag{27}$$

$$q_n^o = \bar{\lambda} q_n + \lambda q_{n-1}, \qquad n = 1, 2, \ldots N, \tag{28}$$

$$q_{N+1}^o = q_{N+1} + \lambda q_N. \tag{29}$$

It can be easily seen that $p_n^o = p_n = q_n^o$, that is, system-length distribution at outside observer's observation epoch is the same in both LAS-DA and EAS. The queue length at service completion in the EAS model will be less than that of the corresponding LAS-DA model by the number of customers that indeed arrive at that slot boundary. It is to be noted that queue length is always evaluated immediately after a slot boundary, which gives the difference between the two models.

3 Performance Measures

In this section, we discuss various performance measures. It may be mentioned here that Little's law agrees at the outside observer's observation epoch. As we have seen earlier that $p_n = p_n^o = q_n^o$, the average number of customers in the queue (L_q^o) and the average number of customers in the system (L_s^o) will remain the same in both the systems (LAS-DA and EAS). They are given by

$$\begin{aligned} L_q^o &= \sum_{n=1}^{N+1}(n-1)p_n^o = \sum_{n=1}^{N} \frac{np_n^+}{p_0^+ + \rho} + (N+1)\left(1 - \frac{1}{p_0^+ + \rho}\right) - \sum_{n=1}^{N+1} p_n \\ &= \frac{1}{p_0^+ + \rho} \sum_{n=0}^{N} np_n^+ + (N+1) - \frac{N+1}{p_0^+ + \rho} - (1 - p_0) \\ &= \frac{a^{'}}{\rho} \sum_{n=0}^{N} np_n^+ + (N+1) - \frac{(N+1)a^{'}}{\rho} - a^{'}. \end{aligned} \tag{30}$$

and

$$L_s^o = \sum_{n=1}^{N+1} n p_n^o == \frac{a'}{\rho} \sum_{n=0}^{N} np_n^+ + (N+1) - \frac{(N+1)a'}{\rho}. \tag{31}$$

The probability that the server is busy (PB) at a random epoch is given by $1 - p_0 = a^{'}$. The effective arrival rate (λ_e) is given as $\lambda_e = \lambda(1 - p_{N+1}) = \frac{\lambda}{p_0^+ + \rho} = \frac{\lambda a^{'}}{\rho}$. The mean waiting time in the queue ($E[W_q]$) and in the system ($E[W_s]$) can be computed using Little's law as

$$E[W_q] = \frac{L_q^o}{\lambda_e} = \frac{1}{\lambda} \sum_{n=0}^{N} np_n^+ + \frac{N+1}{\lambda}\left(\frac{\rho}{a^{'}} - 1\right) - b, \tag{32}$$

$$E[W_s] = \frac{L_s^o}{\lambda_e} = \frac{1}{\lambda} \sum_{n=0}^{N} np_n^+ + \frac{N+1}{\lambda}\left(\frac{\rho}{a^{'}} - 1\right). \tag{33}$$

Remark 3 If $N \to \infty$, then $a^{'} = \rho$ and

$$L_q^o = \sum_{n=1}^{\infty} n p_n^o - \rho, \quad L_s^o = \sum_{n=1}^{\infty} n p_n^o, \quad E[W_s] = \frac{1}{\lambda}\sum_{n=0}^{\infty} n p_n^o \Rightarrow \lambda E[W_s] = L_s^o,$$

$$E[W_q] = \frac{1}{\lambda}\sum_{n=0}^{\infty} n p_n^o - b \Rightarrow \lambda E[W_q] = L_q^o.$$

The above results match analytically with those of Takagi (1993a).

4 Numerical Results

We present some numerical results in the form of tables to illustrate the analytic results obtained in this paper. Various performance measures such as the mean system length (L_s^o), the mean queue length (L_q^o), the mean waiting-time in the system ($E[W_s]$), the mean waiting-time in the queue ($E[W_q]$), the probability of blocking (PBL) and the probability that the server is busy (PB) are given at the bottom of the tables.

In Table 1, the service-time distribution is taken as Discrete phase-type (DPH), where $b_k = \alpha \mathbf{T^{k\text{-}1}T^0}$, $k = 1, 2, \ldots$, $\mathbf{T^0} = \mathbf{e} - T\mathbf{e}$, $\mathbf{e}$ is the 3×1 column vector with all elements equal to one, pgf is $B(z) = \mathbf{z}\alpha\mathbf{(I\text{-}zT)}^{-1}\mathbf{T^0}$, $|z| \le 1$ and $E[S] = 2.5238095$ with $\alpha = \begin{bmatrix} 0.10 & 0.60 & 0.30 \end{bmatrix}$ and $\mathbf{T} = \begin{bmatrix} 0.30 & 0.30 & 0.40 \\ 0.05 & 0.00 & 0.40 \\ 0.10 & 0.20 & 0.30 \end{bmatrix}$.

Table 1 System length distributions at various epochs

	$Geo/DPH/1/11$ system $\rho = 0.504762$, $\lambda = 0.2$				$\rho = 1.26191$, $\lambda = 0.5$			
n	p_n^+	$p_n = q_n^o$	q_n^+	q_n	p_n^+	$p_n = q_n^o$	q_n^+	q_n
0	0.495275	0.495255	0.619094	0.619069	0.004523	0.003571	0.009045	0.007142
1	0.303638	0.303626	0.224774	0.224765	0.010661	0.008418	0.012276	0.009693
2	0.118656	0.118651	0.092127	0.092123	0.015139	0.011954	0.018002	0.014214
3	0.048651	0.048649	0.037782	0.037780	0.022196	0.017527	0.026391	0.020839
4	0.019959	0.019958	0.015503	0.015503	0.032545	0.025699	0.038699	0.030558
5	0.008189	0.008189	0.006361	0.006361	0.047723	0.037683	0.056746	0.044808
6	0.003360	0.003359	0.002609	0.002609	0.069978	0.055256	0.083210	0.065705
7	0.001378	0.001378	0.001070	0.001070	0.102612	0.081025	0.122014	0.096344
8	0.000564	0.000564	0.000438	0.000438	0.150466	0.118811	0.178918	0.141278
9	0.000230	0.000230	0.000179	0.000179	0.220634	0.174217	0.262350	0.207156
10	0.000094	0.000093	0.000072	0.000072	0.323526	0.255463	0.384702	0.303770
11		0.000048		0.000034		0.210377		0.058492
PB		0.504745		0.380931		0.996429		0.992858
	$L_s^o = 0.845502$, $L_q^o = 0.340755$, $PBL = 0.000048$ $E[W_s] = 4.22772$, $E[W_q] = 1.70386$				$L_s^o = 8.66204$, $L_q^o = 7.66562$, $PBL = 0.210377$ $E[W_s] = 21.9397$, $E[W_q] = 19.4159$			

Numerical results for $Geo/DPH/1/11$ are given in Table 1 for $\rho \in \{0.504762, 1.26191\}$. Table 2 presents results for $Geo/D/1/6$ queueing system for $\rho \in \{0.5, 1.0, 1.25, 1.5\}$. Various computed performance measures are also presented. Further, it is noted that $p_{N+1} \to 1 - 1/\rho$ if $\rho \gg 1$. From Eq. 21, it is seen that $p_{N+1} \to 1 - 1/\rho$ if $p_0^+ = 0$. For large ρ, it is unlikely that the departed customer will see the queue empty and thus $p_0^+ \simeq 0$. When $p_0^+ = 0$, it is observed from Eq. 20 that $p_n = p_n^+/\rho,\ 0 \le n \le N$. When $\rho < 1$, by taking N sufficiently large, the results for $Geo/G/1/\infty$ can be obtained from those of $Geo/G/1/N+1$ system.

All the calculations have been done in double precision and rounded up after the sixth decimal point. Note that the mean number of customers in the queue / system, the mean waiting time in the queue / system, the blocking probability and the probability that the server is busy increase with the increase of traffic intensity in all tables.

5 The Continuous-Time Case

Here we consider the relationship between the discrete-time $Geo/G/1/N+1$ queue and its continuous-time counterpart. Let the time axis be slotted into intervals of equal length Δt, and $\Delta > 0$ is sufficiently small. For the continuous-time $M/G/1/N+1$ queue, we

Table 2 System length distributions at various epochs

n	$Geo/D/1/6$ system $\rho = 0.5, \lambda = 0.1$ p_n^+	$p_n = q_n^o$	q_n^+	q_n	$\rho = 1.0, \lambda = 0.2$ p_n^+	$p_n = q_n^o$	q_n^+	q_n
0	0.500195	0.500095	0.555772	0.555661	0.012008	0.009515	0.096153	0.089285
1	0.346890	0.346821	0.323681	0.323617	0.157827	0.146554	0.173245	0.160871
2	0.116858	0.116835	0.188212	0.174769	0.078299	0.062043	0.022031	0.014674
3	0.028400	0.028395	0.021125	0.021121	0.143228	0.113492	0.192505	0.178755
4	0.006271	0.006270	0.004621	0.004620	0.259107	0.205313	0.192295	0.178560
5	0.001382	0.001382	0.001022	0.001022	0.468767	0.371445	0.192309	0.178572
6		0.000202		0.000010		0.071427		0.035713
PB		0.499905		0.444339		0.990485		0.910715
	$L_s^o = 0.69888$, $L_q^o = 0.19897$, $PBL = 0.000202$				L_s^o =3.06786, $L_q^o = 2.13928$, PBL =0.07143			
	$E[W_s] = 6.99016$, $E[W_q] = 1.99010$				$E[W_s] = 16.5192$, $E[W_q] = 11.5192$			
	$\rho = 1.25, \lambda = 0.25$				$\rho = 1.5, \lambda = 0.3$			
0	0.012008	0.009515	0.016011	0.012687	0.001319	0.000879	0.001884	0.001255
1	0.038595	0.030582	0.046122	0.036547	0.006529	0.004349	0.008520	0.005675
2	0.078299	0.062043	0.089024	0.070542	0.022031	0.014674	0.027821	0.018531
3	0.143228	0.113492	0.161296	0.127809	0.069457	0.046264	0.087301	0.058150
4	0.259107	0.205313	0.291711	0.231148	0.217784	0.145062	0.273706	0.182310
5	0.468767	0.371445	0.527785	0.418211	0.682877	0.454851	0.858236	0.571654
6		0.207610		0.103057		0.333921		0.162425
PB		0.990485		0.987313		0.999121		0.998745
	$L_s^o = 4.41928$, $L_q^o = 3.42880$, $PBL = 0.207610$				$L_s^o = 5.03053$, $L_q^o = 4.03139$, PBL =0.333921			
	$E[W_s] = 22.3086$, $E[W_q] = 17.3086$				$E[W_s] = 25.1748$, $E[W_q] = 20.1747$			

assume that the inter-arrival times are exponentially distributed with mean rate $\hat{\lambda}$. So, we have $\lambda = \hat{\lambda}\Delta t + o(\Delta t)$. If the service times S are measured in multiples of Δt, then in a discrete-time, let $P(S = k\Delta t) = b_k$ with $\sum_{k=1}^{\infty} b_k = 1$, $k\Delta t = x_k$ and $b = n\Delta t$, where the interval $[0, b]$ is divided into intervals of length Δt. The pgf of number of arrivals during an interval $(x_k, x_{k+1}) = 1 - \lambda + \lambda z$. Let

$$\begin{aligned} P(\text{service ends in } (t, t+\Delta t)|\text{ service time} > t) &= f(t)\Delta t + o(\Delta t) \\ \text{and} \quad b_k &= f(x_k)\Delta t + o(\Delta t), \end{aligned}$$

where $f(\cdot)$ is common density function of service times. We establish that in the limiting case, the pgf $B(1-\lambda+\lambda z)$ changes to a Laplace transform in the continuous-time case. Applying the definitions of Riemann's sum, exponential function, and improper integral, we get

$$\begin{aligned} B(1-\lambda+\lambda z) &= \lim_{b\to\infty}\lim_{\Delta\to 0}\sum_{k=1}^{b/\Delta t} f(x_k)\left(1-\hat{\lambda}(1-z)\Delta\right)^{x_k/\Delta t}\Delta t \\ &= \int_0^{\infty} f(x)e^{-\hat{\lambda}(1-z)x}dx = \bar{f}(\hat{\lambda}-\hat{\lambda}z) \\ &= \text{Laplace transform of pdf } f(x) \text{ evaluated at } \hat{\lambda}(1-z). \end{aligned}$$

This leads to the probability generating function for the number in the system for $M/G/1/N+1$, that is,

$$P^{+}(z) = \frac{p_0^{+}(1-z)\bar{f}(\hat{\lambda}-\hat{\lambda}z)}{\bar{f}(\hat{\lambda}-\hat{\lambda}z)-z},$$

which matches with the results reported in Chaudhry et al. (1991). Further, using $\lambda = \hat{\lambda}\Delta$ and taking the limit as $\Delta \to 0$, in Eqs. 22–24, we have $p_n^{+} = q_n^{+},\ 0 \le n \le N$, as it should. Also, from Eqs. 20, 21 and 25–29, we obtain

$$p_n = q_n = q_n^{o} = \frac{p_n^{+}}{p_0^{+}+\rho},\quad n = 0, 1, \ldots, N,\quad p_{N+1} = q_{N+1} = q_{N+1}^{o} = 1-\sum_{n=0}^{N} p_n = 1-\frac{1}{p_0^{+}+\rho},$$

which match with the results for the $M/G/1/N+1$ system given in Chaudhry et al. (1991). It is noted that taking the limit as $\Delta \to 0$ and $\lambda = \hat{\lambda}\Delta$ in LAS-DA as well as in EAS systems the probabilities tend to be the same, as they should be in case of continuous-time. Further, using (30)–(33) and taking the limit as $\Delta \to 0$, we establish that the mean system length (L_s^{o}), the mean queue length (L_q^{o}), the mean waiting-time in the system ($E[W_s]$) and the mean waiting-time in the queue ($E[W_q]$) match exactly with those of continuous-time $M/G/1/N+1$ system reported in Takagi (1993b).

6 Conclusions

This paper considers the steady-state system-length distributions for the $Geo/G/1/N+1$ queueing system. We obtain an analysis of the system-length distributions at post departure, random, pre-arrival and outside observer's observation epochs for both late and early arrival systems using roots of the associated characteristic equation. We find a unified approach for solving both finite- and infinite- buffer systems by using roots. The performance measures and numerical illustrations for both the systems are carried out. We establish that in the limiting case the results found in this paper tend to its continuous-time counterpart. In general, the analysis of discrete-time finite-buffer single-server queueing system is more involved since the length of time a customer spends in the system may not follow the Markovian property. These queues have applications in systems such as slotted digital computer and communication systems, manufacturing and service facilities, transportation and telecommunications. Further, the method discussed in this paper could apply to more complex discrete-time queueing systems such as finite-space bulk-arrival or bulk-service systems.

Acknowledgements This research was supported (in part) by the Department of National Defense Applied Research Program grant GRC0000B1638.

References

Akar N (2006) Solving the $ME/ME/1$ queue with state-space methods and the matrix sign function. Perform Eval 63(2):131–145

Alfa A (2016) Applied discrete-time queues. Springer, Berlin

Bhat UN (2015) An introduction to queueing theory: modeling and analysis in applications. Birkhäuser, Basel

Bruneel H, Kim BG (2012) Discrete-time models for communication systems including ATM, vol 205. Springer, Berlin

Chaudhry ML (1992) Qroot software package. A & A Publications, Kingston

Chaudhry ML (2000) On numerical computations of some discrete-time queues. In: Grassmann W (ed) Computational probability, vol 10. Kluwer Academic, Boston, pp 365-407

Chaudhry ML, Harris CM, Marchal WG (1990) Robustness of rootfinding in single-server queueing models. ORSA J Comput 2(3):273–286

Chaudhry ML, Gupta UC, Agarwal M (1991) On exact computational analysis of distributions of numbers in systems for $M/G/1/N+1$ and $GI/M/1/N+1$ queues using roots. Comput Oper Res 18(8):679–694

Finch PD (1958) The effect of the size of the waiting room on a simple queue. J Royal Stat Soc Ser B (Methodological):182–186

Gravey A, Hébuterne G (1992) Simultaneity in discrete-time single server queues with bernoulli inputs. Perform Eval 14(2):123–131

Kobayashi H, Mark BL (2009), System modeling and analysis: Foundations of system performance evaluation. Pearson Education India

Kobayashi H, Mark BL, Turin W (2011) Probability, random processes and statistical analysis: applications to communications, Signal processing, queueing theory and mathematical finance. Cambridge University Press, Cambridge

Lee HW, Lee SH, Yoon SH, Ahn B, Park NI (1999) A recursive method for bernoulli arrival queues and its application to partial buffer sharing in atm. Comput Oper Res 26(6):559–581

Neuts MF (1981) Matrix-geometric solutions in stochastic models: an algorithmic approach. Courier Corporation

Singh G, Gupta U, Chaudhry ML (2016) Detailed computational analysis of queueing-time distributions of the $BMAP/G/1$ queue using roots. J Appl Probab 53(4):1078–1097

Takagi H (1993a) Queueing analysis: Discrete-time systems, vol 3. North Holland

Takagi H (1993b) Queueing Analysis: Finite systems, vol 2. North Holland

Woodward ME (1994) Communication and computer networks: modelling with discrete-time queues. Wiley-IEEE Computer Society Press, New York