Multiserver bulk service discrete-time queue with finite buffer and renewal input

V. Goswami \(^a,\ast\), G.B. Mund \(^b\)

\(^a\) School of Computer Application, KIIT University, Bhubaneswar-751024, India
\(^b\) Department of Computer Science and Engineering, KIIT University, Bhubaneswar-751024, India

ARTICLE INFO

Article history:
Received 24 July 2008
Received in revised form 11 December 2008
Accepted 9 January 2009

Keywords:
Bulk-service
Finite-buffer
Discrete-time
Multiserver
Waiting-time

ABSTRACT

This paper analyzes a discrete-time finite-buffer multi-server bulk-service queueing system in which the interarrival- and service-times are, respectively, arbitrarily and geometrically distributed. Using the supplementary variable and the imbedded Markov-chain techniques, the queue is analyzed for the early arrival system. We obtain state probabilities at prearrival, arbitrary and outside observer’s observation epochs. Some performance measures, waiting-time distribution in the queue along with some numerical results, and special cases of the model have also been discussed. Finally, it is shown that in the limiting case the results obtained in this paper tend to the continuous-time counterpart.

© 2009 Elsevier Ltd. All rights reserved.

doi:10.1016/j.camwa.2009.01.033

1. Introduction

Discrete-time queueing systems have gained importance because of their ample applications in various fields, such as computer and communication systems, telecommunication networks, production management, etc. They can also be used to model various mechanisms in Broadband Integrated Services Digital Network (BISDN), which provides high flexibility of network access, dynamic bandwidth allocation on demand and flexible capacity allocation, since in such an environment, information is digitized and segmented into small packets. A detailed discussion on applications of discrete-time queues is available in [1–3].

In discrete-time queueing systems, the arrivals and departures can occur simultaneously at a boundary epoch of a slot. In the case of simultaneity their order may be taken care of by either arrival-first (AF) or departure-first (DF) management policies, which are also known as ‘late arrival systems with delayed access’ (LAS-DA) and ‘early arrival systems’ (EAS), respectively. More details on this topic have been reported in [4,5]. The earliest work on discrete-time multi-server queues was due to Chan and Maa [6] wherein they discussed the GI/Geo/m queue with EAS and obtained the distribution of the number of customers in the system at the prearrival epoch. Furthermore, Chaudhry and Gupta [7], and Chaudhry et al. [8] have carried out a detailed analysis of the same queueing model and obtained the state probabilities at prearrival, arbitrary and outside observer’s observation epochs. Similar work has been reported in [9,10]. The behavior of multiserver buffers with geometric service times and bursty input traffic has been analyzed in [11]. The performance analysis and optimal control of Geo/Geo/c queue under LAS-DA has been discussed by Artalejo et al. [12]. The transient behavior of the Geo/Geo/m/m queue has been studied by Chaudhry and Gupta [13]. For the GI/Geo/m/m queue, Chaudhry and Gupta [14] have obtained the distribution of number of busy channels at various epochs. The discrete-time finite-buffer GI/Geo/m/N queue with early arrival system has been analyzed by Chaudhry et al. [15]. Further, Gupta et al. [16] have discussed the same queueing model...
for EAS and LAS-DA, and developed a recursive procedure to obtain system length distributions at prearrival, arbitrary and outside observer’s observation epochs.

All the above studies on a multi-server queue have been carried out under the assumption that a server serves the customer one at a time. However, there are many instances where the services are carried out in batches to increase the service rate. Bulk-service queues with a single server have been investigated by Gupta and Goswami [17], Chaudhry and Chang [18], Alfa and He [19] and Yi et al. [20]. Recently, Goswami et al. [21] have analyzed a discrete-time infinite buffer bulk-service Geo/Geo^b/m queue. Some analytic computational results for discrete-time bulk-service queues have been reported in [22].

The general uncorrelated arrival process appears to be more appropriate and reasonable than the geometrical distribution, as the memoryless property of the arrival process does not always meet the need of the applications and also it can include the special cases of geometrical and deterministic etc. Nowadays due to complicated and irregular service mechanisms in telecommunication networks, there has been an increased interest in discrete-time models with non-deterministic service times in which customers are served in batches. Such a type of queues have potential applications in many areas, for example, in the loading and unloading of cargos at a seaport; in traffic signal systems; in communication systems (each message consisting of different packets); in transportation systems (buses, guided tours, medical evacuation systems, air and maritime traffic, etc.); in manufacturing systems (electrical, electronic and mechanical industries making products such as cars, computers, machines, etc., where each job is considered as the accumulation of different tasks); in a computer network where jobs are processed in batches.

The study of finite-buffer queueing models rather than infinite buffer queues is important because in real applications we usually observe the limited waiting space. However, from a theoretical viewpoint both (infinite- and finite-buffer) queueing models have importance. The results for the infinite-buffer queues can be obtained from those of the corresponding finite-buffer counterparts by taking the finite-buffer parameter to be sufficiently large. It may be mentioned here that the modeling and analysis of a discrete-time finite-buffer multi-server queue with bulk-service is more involved and quite different than the corresponding continuous-time counterpart. The advantage of analyzing a discrete-time queue is that one can obtain the continuous-time result from it as a limiting case but the converse is not true.

In this paper, we consider a discrete-time finite-buffer multi-server bulk-service queueing system with renewal arrivals in which the interarrival- and service-times are, respectively, arbitrarily and geometrically distributed. We analyze the early arrival system and obtain the state probabilities at prearrival, arbitrary and outside observer’s observation epochs using the supplementary variable method with remaining interarrival time as the supplementary variable and the imbedded Markov chain approach. One may note that the state probabilities at different epochs are often used to evaluate various system performance measures such as the average number of customers in the queue, the average waiting-time in the queue, the blocking probability, etc. The analysis of actual waiting-time distributions measured in slots, and some special cases of our model have also been discussed. Finally, numerical results have been presented in the form of tables and graphs.

This paper is organized as follows: Model description and analysis of the queueing model is given in Section 2. In Section 3, we discuss the outside observer’s distribution. The waiting-time analysis is carried out in Section 4. Section 5 presents some special cases of our model. The numerical results in the form of tables and graphs are presented in Section 6. Section 7 concludes the paper. In the Appendix, it has been shown that our results tend to the continuous-time counterparts in the limiting case.

2. Model description and analysis

We consider a discrete-time GI/Geo^b/m/N queue wherein interarrival times are independently identically distributed (i.i.d.) random variables with common probability mass function (p.m.f.) \( a_n = P(A = n), n \geq 1 \), probability generating function (p.g.f.) \( A(z) = \sum_{n=0}^{\infty} a_n z^n \) \( (a_0 = 0) \) and mean interarrival time \( a = A^{(1)}(1), \) where \( A^{(1)}(1) \) is the first derivative of \( A(z) \) with respect to \( z \) at \( z = 1. \) There are \( m \) servers and the service times \( S \) (independent of batch size) of batches are independent and geometrically distributed with p.m.f. \( P(S = n) = (1 – \mu)^{n-1} \mu, \) \( 0 < \mu < 1, n \geq 1 \) and mean service time \( 1/\mu. \) The customers are served in batches with a minimum of one and a maximum of \( b \) customers per server. The system has finite-buffer space of size \( N, \) that is, at any time the system can accommodate at most \( N + mb \) customers. The traffic intensity is given by \( \rho = 1/(amb \mu). \) Furthermore, the probability that \( j \) servers complete their service in the next interval given that there are \( i \) busy servers is given by

\[
c(j|i) = \begin{cases} \binom{i}{j} \mu^j (1-\mu)^{i-j}, & 0 \leq j \leq \min(i, m), \\ 0, & \text{otherwise}. \end{cases}
\]

Let us assume that the time axis is slotted into intervals of unit length and the time axis be marked by \( 0, 1, 2, \ldots, t, \ldots. \) We assume that a potential arrival occurs in \((t, t+)\) and a potential departure takes place in \((t-, t).\) The various time epochs at which events occur are depicted in Fig. 1.

The state of the system just before a potential arrival at time \( t \) is described by the following random variables:

- \( N_0(t) \): number of customers in the queue excluding the batches which are in services,
- \( U(t) \): remaining interarrival time for the next arrival.
In the steady-state, let us define

\[ Q_{0,k}(u) = \lim_{t \to \infty} P[N_q(t) = 0, U(t) = u, k \text{ servers are busy}], \quad u \geq 0, 0 \leq k \leq m - 1, \]
\[ Q_{n,m}(u) = \lim_{t \to \infty} P[N_q(t) = n, U(t) = u, \text{all servers are busy}], \quad u \geq 0, 0 \leq n \leq N. \]

We introduce the following probability generating functions:

\[ Q_{0,k}^*(z) = \sum_{u=0}^{\infty} Q_{0,k}(u)z^u, \quad 0 \leq k \leq m - 1, \quad Q_{n,m}^*(z) = \sum_{u=0}^{\infty} Q_{n,m}(u)z^u, \quad 0 \leq n \leq N. \]

Let \( Q_{0,k} \equiv Q_{0,k}^*(1) \) be the probability that \( k (0 \leq k \leq m - 1) \) servers are busy with no customers waiting in the queue at an arbitrary epoch, and \( Q_{n,m} \equiv Q_{n,m}^*(1) \) be the probability that all servers are busy with \( n (0 \leq n \leq N) \) customers waiting in the queue at an arbitrary epoch.

Relating the states of the system at two consecutive epochs \( t \) and \( (t + 1) \), and using the definitions and probabilities defined above, we obtain the following equations, in the steady-state, for \( u \geq 1, \)

\[ Q_{0,0}(u-1) = \sum_{j=0}^{m} c(j)Q_{0,j}(u) + a_u \sum_{j=0}^{m-1} c(j+1+j+1)Q_{0,j}(0), \quad \text{(2)} \]
\[ Q_{0,k}(u-1) = \sum_{j=k}^{m} c(j-k+j)Q_{0,j}(u) + \sum_{j=m+1}^{m+k} c(j-k|m) \sum_{i=1}^{b} Q_{i+(j-m-1)b,m}(u) \]
\[ + a_u \sum_{j=k}^{m} c(j-k+j)Q_{0,j-1}(0) + a_u \sum_{j=m+1}^{m+k} c(j-k|m) \sum_{i=1}^{b} Q_{i+(j-m-1)b-1,m}(0), \quad 1 \leq k \leq m, \quad \text{(3)} \]
\[ Q_{n,m}(u-1) = \sum_{j=0}^{m} c(j|m)Q_{a+jb,m}(u) + a_u \sum_{j=0}^{m} c(j|m)Q_{a+jb-1,m}(0), \quad 1 \leq n \leq N - mb - 1, \quad \text{(4)} \]
\[ Q_{n,m}(u-1) = \sum_{j=0}^{[\frac{N-n}{b}]} c(j|m)Q_{n+jb,m}(u) + a_u \sum_{j=0}^{[\frac{N-n}{b}]} c(j|m)Q_{n+jb-1,m}(0) \]
\[ + a_u \gamma_n Q_{n,m}(0), \quad N - mb \leq n \leq N, \quad \text{(5)} \]

where \([x]\) is the greatest integer contained in \( x \), and

\[ \gamma_n = \begin{cases} c(i|m), & \text{if } i = \frac{N-n}{b}\text{ is an integer}, \\ 0, & \text{otherwise}. \end{cases} \]

Multiplying (2) to (5) by \( z^u \) and summing over \( u \) from 1 to \( \infty \), we obtain the following

\[ zQ_{0,0}^*(z) = \sum_{j=0}^{m} c(j)Q_{0,j}^*(z) + A(z) \sum_{j=0}^{m-1} c(j+1+j+1)Q_{0,j}(0) - \sum_{j=0}^{m} c(j)Q_{0,j}(0). \quad \text{(6)} \]
of busy servers at prearrivalepoch of the mutually independent identically distributed random variables with common distribution function \(B\). Evaluation of \(Q\) to determine the prearrivalepoch probabilities.

2.1. Queue-length distribution at prearrivalepoch

Let \(Q_{0,k}(z) = \sum_{j=k}^{m} c(j-k|j)Q_{0,j}(z) + \sum_{j=m+1}^{m+k} c(j-k|m) \sum_{b=j-1}^{b=m} Q_{i+(j-m-1)b,m}(z) + \sum_{j=m+1}^{m+k} c(j-k|m) \sum_{b=1}^{b=m} (A(z)Q_{i+(j-m-1)b-1,m}(0) - Q_{i+(j-m-1)b,m}(0)) + \sum_{j=k}^{m} c(j-k|j) (A(z)Q_{0,j-1}(0) - Q_{0,j}(0)), \quad 1 \leq k \leq m. \tag{7}\)

\[ zQ_{n,m}(z) = \sum_{j=0}^{N-n} c(j|m)Q_{n+jb,m}(z) + \sum_{j=0}^{N-n} c(j|m) (A(z)Q_{n+jb-1,m}(0) - Q_{n+jb,m}(0)), \quad 1 \leq n \leq N - mb - 1, \tag{8}\]

\[ zQ_{n,m}(z) = \sum_{j=0}^{N-n} c(j|m)Q_{n+jb,m}(z) + \sum_{j=0}^{N-n} c(j|m) (A(z)Q_{n+jb-1,m}(0) - Q_{n+jb,m}(0)) + A(z)T_nQ_{N,m}(0), \quad N - mb \leq n \leq N. \tag{9}\]

Adding (6) to (9), we obtain

\[
\sum_{k=0}^{m-1} Q_{0,k}(z) + \sum_{n=0}^{N} Q_{n,m}(z) = \frac{A(z) - 1}{z - 1} \left\{ \sum_{k=0}^{m-1} Q_{0,k}(0) + \sum_{n=0}^{N} Q_{n,m}(0) \right\}. 
\]

Taking the limit as \(z \to 1\) in the above and using the normalization condition \(\sum_{k=0}^{m-1} Q_{0,k} + \sum_{n=0}^{N} Q_{n,m} = 1\), we obtain

\[
\sum_{k=0}^{m-1} Q_{0,k}(0) + \sum_{n=0}^{N} Q_{n,m}(0) = \frac{1}{\lambda} = \lambda \quad \text{(say)}. \tag{10}\]

The above Eq. (10) represents the fact that the probability that an arrival is about to occur is equal to the arrival rate of customers. Further, it may be noted that \(\sum_{k=0}^{m-1} Q_{0,k}(z) + \sum_{n=0}^{N} Q_{n,m}(z) = \frac{A(z) - 1}{z - 1}\) represents the probability generating function of the residual interarrival time, which is a known result for renewal processes.

2.1. Queue-length distribution at prearrivalepoch

Let \(Q_{0,k}\) be the probability that \(k (0 \leq k \leq m - 1)\) servers are busy with no customers waiting in the queue at the prearrivalepoch, and \(Q_{n,m}\) be the probability that all servers are busy with \(n (0 \leq n \leq N)\) customers waiting in the queue at prearrivalepoch. Applying Bayes' theorem, we have

\[
Q_{0,k} = P(\text{arrival is about to occur and there are} \ n \ \text{customers in the queue with} \ k \ \text{busy servers}) \\
= \frac{Q_{n,k}(0)}{\sum_{j=0}^{m-1} Q_{0,j}(0) + \sum_{l=0}^{N} Q_{l,m}(0)}, \quad n = 0, 0 \leq k \leq m - 1; 0 \leq n \leq N, k = m.
\]

Further, using (10) in the above expression, we obtain

\[
Q_{0,k} = \frac{1}{\lambda} Q_{0,k}(0), \quad n = 0, 0 \leq k \leq m - 1; 0 \leq n \leq N, k = m. \tag{11}\]

To determine the prearrivalepoch probabilities \(Q_{0,k}\) and \(Q_{n,m}\), we first need to evaluate \(Q_{0,k}(0)\) and \(Q_{n,m}(0)\). However, the evaluation of \(Q_{0,k}(0)\) and \(Q_{n,m}(0)\) directly from Eq. (6) to (9) seems to be difficult. Therefore, we first compute prearrivalepoch probabilities using the imbedded Markov chain technique.

Let customers arrive at time epochs \(T_1, T_2, \ldots, \) and assume that the interarrivaltimes \(T_{n+1} - T_n (n = 0, 1, \ldots; T_0 = 0)\) are mutually independent identically distributed random variables with common distribution function \(B(u) = P(T_{n+1} - T_n \leq u), n = 0, 1, \ldots\). Let \(N^n = N^n(T_n - 0)\) and \(J^n\), respectively, denote the number of customers in the queue and the number of busy servers at prearrivalepoch of the \(n\)th customer. Then \(\{N^n, J^n\}, n \geq 0\) is an imbedded two-dimensional Markov chain with the state space \(\Omega = \{(0, k) \cup (n, m) : 0 \leq k \leq m - 1, 0 \leq n \leq N\}\). Let \(Q = (q_{ij})\) be the transition probability matrix, and \(q^- = [Q_{0,0}, Q_{0,1}, \ldots, Q_{0,m}, Q_{1,m}, \ldots, Q_{n,m}]\) be a row vector of prearrivalepoch probabilities. Then \(q^-\) can be
obtained by solving the system of equations $q^{-1}Q = q^{-1}$. The transition probabilities $q_i$'s are given by

\[
q_{ij} = \begin{cases} 
V_{i+1,j} & 0 \leq i \leq m - 1, 1 \leq j \leq \min(i+1, m-1), \\
V_{i-mk+,j+m,j} & m \leq i \leq N + m - 1, 1 \leq j \leq m - 1, \\
\beta_{i-mk+,j+m,j} & m - 1 \leq i \leq N + m - 1, j = m, \\
\beta_{k} & m + kb \leq i \leq N + m - 1, j = i - kb + 1, \text{ where } k \geq 0 \text{ is an integer,} \\
q_{N+m-1,j} & i = N + m, 1 \leq j \leq N + m, \\
1 - \sum_{n=1}^{\infty} q_{i,n} & 0 \leq i \leq N + m, j = 0, \\
0 & \text{otherwise}.
\end{cases}
\]

The $V_{i,j}$'s and $\beta_{j}$'s are given by

\[
V_{ij} = \begin{cases} 
\sum_{n=1}^{\infty} a_n \binom{i}{j} (1 - \mu)^n (1 - (1 - \mu)^n)^{i-j}, & j \leq i, j \geq 0, \\
\sum_{n=1}^{\infty} a_n \left( \sum_{r=1}^{n} \sum_{s=j}^{m} \binom{s}{j} (1 - \mu)^{n-r} (1 - (1 - \mu)^{n-r})^{s-j} \right) \times \sum_{y=m-n+1}^{m} b(i - y - s[m(r - 1)]c(y|m)) & i > m, 0 \leq j \leq m - 1, \\
0 & \text{otherwise},
\end{cases}
\]

\[
\beta_{j} = \sum_{n=1}^{\infty} a_n b(j|mn), \quad j \geq 0,
\]

where

\[
b(j|mk) = \begin{cases} 
\binom{mk}{j} \mu^{j} (1 - \mu)^{mk-j}, & 0 \leq j \leq mk, k \geq 0, \\
0, & \text{otherwise}.
\end{cases}
\]

The $V_{i,j}$ and $\beta_{j}$, respectively, being the probabilities of $(i-j)$ service completion when at least one server remains or becomes idle during an interarrival time, and $j$ service completion when all the $m$ servers are busy during an interarrival time. For a detailed derivation of $V_{i,j}$'s and $\beta_{j}$'s, see [6]. The system of equations have been solved using the algorithm of Grassmann, Taksar and Heyman [23].

### 2.2. Queue-length distribution at arbitrary epoch

To obtain arbitrary epoch probabilities, we develop relations between distributions of number of customers in the queue at prearrival and arbitrary epochs. For this, setting $z = 1$ in (9) down to (7) and using (11), we get the following

\[
Q_{n,m} = \frac{\lambda c(0|m)}{1 - c(0|m)} Q_{n-1,m},
\]

\[
Q_{n,m} = \frac{\lambda c(0|m)}{1 - c(0|m)} \left\{ Q_{n-1,m} - Q_{n,m} \right\}, \quad n = N - 1, N - 2, \ldots, N - b + 1,
\]

\[
Q_{n,m} = \frac{1}{1 - c(0|m)} \left\{ \sum_{j=1}^{N-n} c(j|m)Q_{n+jb,m} + \lambda \left\{ \sum_{j=0}^{N-n} c(j|m) \left( Q_{n+jb-1,m} - Q_{n+jb,m} \right) + \gamma_{n} Q_{n,m} \right\} \right\},
\]

\[
Q_{n,m} = \frac{1}{1 - c(0|m)} \left\{ \sum_{j=1}^{m} c(j|m)Q_{n+jb,m} + \lambda \left\{ \sum_{j=0}^{m} c(j|m) \left( Q_{n+jb-1,m} - Q_{n+jb,m} \right) \right\} \right\}, \quad n = N - mb - 1, N - mb - 2, \ldots, 1,
\]

\[
Q_{0,m} = \frac{1}{1 - c(0|m)} \left[ \sum_{j=1}^{m} c(j|m) \sum_{i=1}^{b} Q_{i+(j-1)b,m} + \gamma_{0} \left( Q_{0,m-1} - Q_{0,m} \right) \right] + \sum_{j=1}^{m} c(j|m) \sum_{i=1}^{b} \left( Q_{i+(j-1)b-1,m} - Q_{i+(j-1)b,m} \right),
\]
\[
Q_{0,k} = \frac{1}{1 - c(0|k)} \left[ \sum_{j=k+1}^{m} c(j-k|j)Q_{0,j} + \sum_{j=m+1}^{m+k} c(j-k|m) \sum_{i=1}^{b} Q_{i+(j-m-1)b,m} \right.
\]
\[
+ \lambda \left\{ \sum_{j=k}^{m} c(j-k|j) \left( Q_{0,j-1} - Q_{0,j} \right) + \sum_{j=m+1}^{m+k} c(j-k|m) \times \sum_{i=1}^{b} \left( Q_{i+(j-m-1)b-1,m} - Q_{i+(j-m-1)b,m} \right) \right\},
\]
\[
k = m - 1, m - 2, \ldots, 2, 1.
\]

Using the normalization condition,
\[
Q_{0,0} = 1 - \sum_{k=1}^{m-1} Q_{0,k} - \sum_{n=0}^{N} Q_{n,m}.
\]

It can be seen from the above set of expressions that once we know prearrival epoch probabilities, the arbitrary epoch probabilities can be easily computed.

3. Outside observer's distribution

The outside observer's observation epoch falls in a time interval after a potential arrival and before a potential departure. The probability \(Q_{0,k}^o\) (\(0 \leq k \leq m - 1\)) that the outside observer sees \(k\) servers busy with no customers in the queue and the probability \(Q_{n,m}^o\) (\(0 \leq n \leq N\)) that \(m\) servers busy with \(n\) customers in the queue, can be obtained by observing arbitrary and outside observer's observation epochs presented in Fig. 1.

\[
Q_{0,0} = \sum_{j=0}^{m} c(j|j)Q_{0,j}^o,
\]
\[
Q_{0,k} = \sum_{j=k}^{m} c(j-k|j)Q_{0,j}^o + \sum_{j=m+1}^{m+k} c(j-k|m) \sum_{i=1}^{b} Q_{i+(j-m-1)b,m}^o, \quad 1 \leq k \leq m,
\]
\[
Q_{n,m} = \sum_{j=0}^{\min\{m,\frac{N-n}{b}\}} c(j|m)Q_{n+jb,m}^o, \quad 1 \leq n \leq N - b,
\]
\[
Q_{n,m} = c(0|m)Q_{n,m}^o, \quad N - b + 1 \leq n \leq N.
\]

Now, solving for \(Q_{0,k}^o\) and \(Q_{n,m}^o\), we obtain

\[
Q_{n,m}^o = \frac{Q_{n,m}}{c(0|m)}, \quad n = N, N - 1, \ldots, N - b + 1,
\]
\[
Q_{n,m}^o = \frac{1}{c(0|m)} \left[ Q_{n,m} - \sum_{j=1}^{\min\{m,\frac{N-n}{b}\}} c(j|m)Q_{n+jb,m}^o \right], \quad n = N - b, N - b - 1, \ldots, 2, 1,
\]
\[
Q_{0,k}^o = \frac{1}{c(0|k)} \left[ Q_{0,k} - \sum_{j=k+1}^{m} c(j-k|j)Q_{0,j}^o - \sum_{j=m+1}^{m+k} c(j-k|m) \sum_{i=1}^{b} Q_{i+(j-m-1)b,m}^o \right], \quad n = m, m - 1, \ldots, 2, 1,
\]
\[
Q_{0,0}^o = Q_{0,0} - \sum_{j=1}^{m} Q_{0,j}^o.
\]

4. Waiting-time in the queue

In this section, we obtain an actual waiting-time (in queue) distribution (measured in slots) of a customer under the FCFS discipline. Let us define the random variable \(T_q\) as “time spent waiting in the queue” of an arrival with the corresponding p.m.f. \(w_k = P(T_q = k), k \geq 0\). An arriving customer may observe the system in any one of the following two cases.

Case 1. \(w_0 = P(T_q = 0)\).

That is, there are no customers in the queue and ‘\(I\)’ (\(0 \leq i \leq m - 1\)) servers are busy. The probability that the customer does not wait is given by

\[
w_0 = P(T_q = 0) = \frac{\sum_{i=0}^{m-1} Q_{0,i}^o}{1 - Q_{N,m}^o}.
\]
**Case 2.** \( w_k = P(T_q = k), k \geq 1. \)

This occurs, (i) if prior to an arrival, there are \( nb + s (0 \leq n \leq v - 1, 0 \leq s \leq b - 1, \) where \( v = \left[ \frac{N}{b} \right] \) customers waiting in the queue and all servers are busy, then the arriving customer joins the queue and waits till the service completion of \((n + 1)\) batches of customers.

(ii) if prior to an arrival, there are \( vb + s (0 \leq s \leq N - 1 - vb) \) customers waiting in the queue and all servers are busy, then the arriving customer joins the queue and waits in the queue till the service completion of \((v + 1)\) batches of customers.

Therefore, the probability that the customer waits for greater than \( k \) slots is

\[
P(T_q > k) = \frac{1}{1 - PBL} \left[ \sum_{j=0}^{v-1} \sum_{n=0}^{b-1} Q_{b+n,m} \sum_{i=0}^{j} b(i|m)k + \sum_{n=0}^{N-vb} Q_{b+n,m} \sum_{i=0}^{v} b(i|m)k \right], \quad k \geq 1,
\]

where \( PBL = Q_{N,m}^- \) is the blocking probability of an arriving customer. Using (23), we get

\[
P(T_q > 0) = 1 - w_0.
\]

Thus, the probability that an arriving customer will have to wait exactly \( k \) slots is

\[
w_k = P(T_q > k - 1) - P(T_q > k), \quad k \geq 1.
\] (24)

The average waiting-time in the queue (\( W_q = \sum_{k=1}^{\infty} kw_k \)) is given by

\[
W_q = \frac{1}{1 - Q_{N,m}^-} \sum_{k=1}^{\infty} \left[ \sum_{j=0}^{v-1} \sum_{n=0}^{b-1} Q_{b+n,m} \sum_{i=0}^{j} \{ b(i|m(k-1)) - b(i|m) \} \right.
\]

\[
+ \sum_{n=0}^{N-vb} Q_{b+n,m} \sum_{i=0}^{v} \{ b(i|m(k-1)) - b(i|m) \} \right].
\]

One may note here that using Little’s rule, we can also obtain the average waiting-time in the queue (\( W_q \)) using \( W_q = L_q^0/\lambda' \), where \( L_q^0 = \sum_{n=1}^{N} Q_{n,m}^0 \) is the average queue-length at an outside observer’s observation epoch and \( \lambda' = \lambda (1 - Q_{N,m}^-) \) is the effective arrival rate.

5. **Special cases**

In this section, results pertaining to some models have been deduced from our model by taking specific values for the parameters \( m, b, \) and \( N. \)

**Case 1: \( b = 1, \) that is, the batch size is one.** The model reduces to \( GI/Geo/m/N \) queue. To obtain relations between prearrival and arbitrary epoch probabilities, let us define \( Q_n = Q_{0,n}, \) \( 0 \leq n \leq m, \) and \( Q_n = Q_{n-m,m}, m + 1 \leq n \leq N + m, \) where \( Q_n = Pr\{ \text{number of customers in system is } n \}. \)

Then from (12)–(18), we obtain

\[
Q_{N+m} = \frac{\lambda c(0|m)}{[1 - c(0|m)]} Q_{N-1+m}^-,\]

\[
Q_{n+m} = \frac{1}{1 - c(0|m)} \left[ \sum_{j=1}^{n} c(j|m)Q_{n-j+m} + \lambda \left[ \sum_{j=0}^{n} c(j|m) \left( Q_{n-j+1+m}^- - Q_{n+j+m}^- \right) \right] \right], \quad n = N - 1, N - 2, \ldots, N - m, N - m + 1, N - m,
\]

\[
Q_{n+m} = \frac{1}{1 - c(0|m)} \left[ \sum_{j=1}^{m} c(j|m)Q_{n+m-j} + \lambda \left[ \sum_{j=0}^{m} c(j|m) \left( Q_{n-j+1+m}^- - Q_{n+j+m}^- \right) \right] \right],
\]

\[
Q_n = \frac{1}{1 - c(0|m)} \left[ \sum_{j=1}^{m+n} c(j-n|m)Q_j + \sum_{j=m+1}^{m+n} c(j-n|m)Q_j + \lambda \left[ \sum_{j=0}^{m+n} c(j-n) \left( Q_{j-1}^--Q_j^- \right) \right] \right], \quad n = m - 1, m - 2, \ldots, 1, 0,
\]

\[
Q_n = \frac{1}{1 - c(0|m)} \left[ \sum_{j=1}^{m+n} c(j-n|m)Q_j + \sum_{j=m+1}^{m+n} c(j-n|m)Q_j + \lambda \left[ \sum_{j=0}^{m+n} c(j-n) \left( Q_{j-1}^--Q_j^- \right) \right] \right], \quad n = m - 1, m - 2, \ldots, 2, 1,
\]
whilst $Q_0$ is found from $Q_0 = 1 - \sum_{n=1}^{N+m} Q_n$. It is to be noted that (13) will not occur in this case. Solving Eqs. (19)-(22) recursively, we obtain outside observer’s distribution as

$$Q_n = \sum_{j=0}^{\min\{m,N+m-n\}} c(j|m)Q_{n+j}^\circ, \quad n = N + m, N + m - 1, \ldots, m + 1,$$

$$Q_n = \sum_{j=n}^{m} c(j-n|j)Q_j^\circ + \sum_{j=m+1}^{m+n} c(j-n|m)Q_j^\circ, \quad n = m, m - 1, \ldots, 0,$$

and our results match analytically with the results presented in Chaudhry et al. [15].

Case 2: $m = b = 1$, that is, the number of servers and the batch size both are one. The model reduces to standard $GI/Geo/1/N$ queue. Let us define $Q_n = Q_{0,n}, 0 \leq n \leq 1, Q_n = Q_{n-1,1}, 2 \leq n \leq N + 1$, where $Q_n = Pr$ (number of customers in system is $n$). Substituting $m = b = 1$ in (12)-(18), we obtain

$$Q_{n+1} = \rho \bar{\mu} Q_n^1,$$

$$Q_n = \rho \mu Q_n^1 + \rho \bar{\mu} Q_n^1, \quad 1 \leq n \leq N,$$

$$Q_0 = 1 - \sum_{n=1}^{N+1} Q_n.$$  

It is to be noted that (13) and (17) will not occur in this case. The outside observer’s distribution after simplification is given as

$$Q_n^\circ = \frac{1}{\rho} \sum_{j=0}^{n+1} \left( \frac{\mu}{\bar{\mu}} \right)^j Q_j, \quad 0 \leq n \leq N + 1,$$

and our results match with the results available in Chaudhry and Gupta [24].

Case 3: $m = 1$, that is, the system has one server. The model reduces to $GI/Geom^b/1/N$ queue. The relations between prearrival and arbitrary epoch probabilities and the outside observer’s observation epochs, respectively, can be derived from (12)-(18), and from (19)-(22). It is to be noted that (17) will not occur in this case.

Case 4: $\rho < 1$ and $N \rightarrow \infty$. That is, when the buffer size is infinite. Then our model reduces to $GI/Geo^b/m/\infty$ queue. In this case, the relations between prearrival and arbitrary epoch probabilities can be derived from (15)-(18). It is to be noted that (12)-(14) will not occur in this case.

6. Numerical results

In this section, we present numerical results in the form of tables and graphs. Table 1 presents the results on queue-length of an interarrival time distribution with two possible values for the interarrival times at prearrival, arbitrary and outside observer’s observation epochs. Various performance measures such as the blocking probabilities, the average queue-length at outside observer’s observation epoch and the average waiting times in the queue using Little’s rule are given at the bottom of Table 1. It can be seen from Table 1 that, for the geometrical interarrival times, the queue length distributions at prearrival and arbitrary epoch probabilities are the same due to the memoryless property of Bernoulli arrivals. The average waiting time distributions for geometrical, deterministic and arbitrary interarrival time distributions are given in Table 2. We have seen that the average waiting-time in the queue using Little’s rule is same as the one obtained using the p.m.f. of the actual waiting-time in the queue of a customer in an accepted batch as it should be. Fig. 2 compares the effect of buffer size ($N$) on the blocking probability (PBL) for various interarrival time distributions with same mean $\lambda = 2, \mu = 0.1$ and $b = m = 3$. The interarrival time distributions are taken as geometric ($\lambda = 0.5$), deterministic ($a_2 = 1$) and arbitrary ($a_2 = 0.2, a_2 = 0.6, a_3 = 0.2$).

The geometrical distribution gives the highest value for the blocking probability and the deterministic distribution gives the smallest value. This is intuitively clear since the blocking probability of the interarrival times is higher for the geometric distribution. This observation can be understood from the fact that the higher the buffer contents, the more customers that get lost for a given amount of buffer space. We further observe that for all distributions considered here, the blocking probability decreases as buffer size $N$ increases and finally reaches to its minimum value zero as it should be. This is due to the fact that the model becomes an infinite-buffer queue. The effect of service rate ($\mu$) on the average waiting-time ($W_q$) is shown in Fig. 3 when interarrival time is deterministic with $a_2 = 1, b = 4$ and $N = 10$ for various values of $m$. It can be seen that for all values of $m$, the average waiting-time monotonically decreases as service rate increases. It can be further observed that $W_q$ decreases as number of servers increases.

The variation in the queueing delay for different values of the number of servers and the batch size is shown in Fig. 4, when the interarrival time is deterministic with $a_2 = 1, \mu = 0.1$ and $N = 10$. We varied the number of servers $m$ from 2 to 8, while the batch size $b$ is varied from 1 to 5. It is observed that for fixed batch sizes the average waiting-time decreases when the number of servers increases. Furthermore, with a fixed number of servers, the average waiting-time decreases when the
Fig. 2. Effect of $N$ on $PBL$.

Fig. 3. Effect of $\mu$ on $W_q$ with varying $m$.

Fig. 4. The $W_q$ for different values of $m$ and $b$. 
Table 1
Queue-length distributions at prearrival, arbitrary and outside observer’s observation epochs.

<table>
<thead>
<tr>
<th>(n, k)</th>
<th>Geo/Geo^2 /3/25</th>
<th>Gl/Geo^2 /2/20</th>
<th>D/Geo^3 /3/10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\lambda = 0.9, \mu = 0.05, \rho = 0.857143)</td>
<td>(\alpha_2 = 0.7, \alpha_8 = 0.3, \mu = 0.11, \rho = 0.473485)</td>
<td>(d = 4, \mu = 0.1, \rho = 0.277778)</td>
</tr>
<tr>
<td>(Q_{a,k}^{-})</td>
<td>(Q_{a,k})</td>
<td>(Q_{a,k}^{+})</td>
<td>(Q_{a,k}^{-})</td>
</tr>
<tr>
<td>(0, 0)</td>
<td>0.000005</td>
<td>0.000005</td>
<td>0.000007</td>
</tr>
<tr>
<td>(0, 1)</td>
<td>0.000012</td>
<td>0.000012</td>
<td>0.000022</td>
</tr>
<tr>
<td>(0, 2)</td>
<td>0.004817</td>
<td>0.004817</td>
<td>0.006671</td>
</tr>
<tr>
<td>(1, m)</td>
<td>0.063093</td>
<td>0.063093</td>
<td>0.10644</td>
</tr>
<tr>
<td>(2, m)</td>
<td>0.057486</td>
<td>0.057486</td>
<td>0.059958</td>
</tr>
<tr>
<td>(3, m)</td>
<td>0.054831</td>
<td>0.054831</td>
<td>0.057221</td>
</tr>
<tr>
<td>(4, m)</td>
<td>0.052275</td>
<td>0.052275</td>
<td>0.054576</td>
</tr>
<tr>
<td>(5, m)</td>
<td>0.049299</td>
<td>0.049299</td>
<td>0.052040</td>
</tr>
<tr>
<td>(6, m)</td>
<td>0.047747</td>
<td>0.047747</td>
<td>0.049711</td>
</tr>
<tr>
<td>(7, m)</td>
<td>0.045625</td>
<td>0.045625</td>
<td>0.047535</td>
</tr>
<tr>
<td>(8, m)</td>
<td>0.043506</td>
<td>0.043506</td>
<td>0.045413</td>
</tr>
<tr>
<td>(9, m)</td>
<td>0.041360</td>
<td>0.041360</td>
<td>0.043291</td>
</tr>
<tr>
<td>(10, m)</td>
<td>0.039183</td>
<td>0.039183</td>
<td>0.041142</td>
</tr>
<tr>
<td>(11, m)</td>
<td>0.037617</td>
<td>0.037617</td>
<td>0.039026</td>
</tr>
<tr>
<td>(12, m)</td>
<td>0.036581</td>
<td>0.036581</td>
<td>0.037513</td>
</tr>
<tr>
<td>(15, m)</td>
<td>0.030709</td>
<td>0.030709</td>
<td>0.032719</td>
</tr>
<tr>
<td>(18, m)</td>
<td>0.033152</td>
<td>0.033152</td>
<td>0.026728</td>
</tr>
<tr>
<td>(20, m)</td>
<td>0.023615</td>
<td>0.023615</td>
<td>0.027544</td>
</tr>
<tr>
<td>(25, m)</td>
<td>0.064831</td>
<td>0.064831</td>
<td>0.075616</td>
</tr>
<tr>
<td>Sum</td>
<td>1.000000</td>
<td>1.000000</td>
<td>0.999998</td>
</tr>
</tbody>
</table>

Table 2
Waiting-time distribution.

<table>
<thead>
<tr>
<th>(n, k)</th>
<th>Geo/Geo^2 /3/25</th>
<th>Gl/Geo^2 /2/20</th>
<th>D/Geo^3 /3/10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\lambda = 0.9, \mu = 0.05, \rho = 0.857143)</td>
<td>(\alpha_2 = 0.7, \alpha_3 = 0.3, \mu = 0.11, \rho = 0.473485)</td>
<td>(d = 4, \mu = 0.1, \rho = 0.277778)</td>
</tr>
<tr>
<td>(k)</td>
<td>(w_k)</td>
<td>(k)</td>
<td>(w_k)</td>
</tr>
<tr>
<td>0</td>
<td>0.009870</td>
<td>0</td>
<td>0.253001</td>
</tr>
<tr>
<td>1</td>
<td>0.060997</td>
<td>1</td>
<td>0.125991</td>
</tr>
<tr>
<td>2</td>
<td>0.058493</td>
<td>2</td>
<td>0.104747</td>
</tr>
<tr>
<td>3</td>
<td>0.056056</td>
<td>3</td>
<td>0.087077</td>
</tr>
<tr>
<td>4</td>
<td>0.035639</td>
<td>4</td>
<td>0.072391</td>
</tr>
<tr>
<td>5</td>
<td>0.051282</td>
<td>5</td>
<td>0.060181</td>
</tr>
<tr>
<td>6</td>
<td>0.048891</td>
<td>6</td>
<td>0.050311</td>
</tr>
<tr>
<td>7</td>
<td>0.046568</td>
<td>7</td>
<td>0.041592</td>
</tr>
<tr>
<td>8</td>
<td>0.044232</td>
<td>8</td>
<td>0.034577</td>
</tr>
<tr>
<td>9</td>
<td>0.041917</td>
<td>9</td>
<td>0.028745</td>
</tr>
<tr>
<td>10</td>
<td>0.039632</td>
<td>10</td>
<td>0.023897</td>
</tr>
<tr>
<td>20</td>
<td>0.020050</td>
<td>20</td>
<td>0.003768</td>
</tr>
<tr>
<td>30</td>
<td>0.008498</td>
<td>30</td>
<td>0.000594</td>
</tr>
<tr>
<td>126</td>
<td>0.000001</td>
<td>126</td>
<td>0.000001</td>
</tr>
<tr>
<td>(k \geq 127)</td>
<td>0.000000</td>
<td>(k \geq 68)</td>
<td>0.000000</td>
</tr>
<tr>
<td>(\Sigma w_k)</td>
<td>1.000000</td>
<td>(\Sigma w_k)</td>
<td>1.000000</td>
</tr>
<tr>
<td>(W_q)</td>
<td>12.900984</td>
<td>(W_q)</td>
<td>4.428410</td>
</tr>
</tbody>
</table>

batch size increases. To ensure minimum average waiting-time in the queue, we can carefully setup the number of servers and the batch size in the system. Fig. 5 illustrates dependence of the blocking probability on the buffer size \(N\) varying from 5 to 30 and the batch size \(b\) varying from 2 to 10. The interarrival time is assumed to be geometrically distributed with \(\lambda = 0.95, \mu = 0.2\) and \(m = 2\). We observe that, for a fixed batch size, the blocking probability decreases as the buffer size increases. Furthermore, with a fixed buffer size, it increases when the batch size decreases. Hence, we can setup an admissible batch size and buffer size in the system in order to have lower blocking probability.

7. Conclusion

In this paper, we have carried out an analysis of finite-buffer discrete-time multiserver bulk-service queues that have potential applications in modeling, telecommunication systems and computer networks, etc. We have obtained the queue-length distributions at prearrival, arbitrary and outside observer’s observation epochs. Various performance measures such as the blocking probability, average queue-length at outside observer’s observation epoch and analysis of actual waiting-time in the queue have been carried out for the EAS model. The results for the LAS-DA model can also be obtained in a similar
manner. The techniques used in this paper can be applied to analyze more complex models such as $\text{DMAP}/\text{Geo}^b/m/N$ and $\text{GI}^X/\text{Geo}^b/m/N$ queues which are left for future investigations.

Acknowledgments

The authors are thankful to the referees for their valuable comments and suggestions which have helped in improving the quality of the presentation of this paper.

Appendix

Here we study the relationship between the discrete-time $\text{GI}/\text{Geo}^b/m/N$ queue and its continuous-time counterpart. For the continuous-time $\text{GI}/\text{M}^b/m/N$ queue, we assume that the interarrival time distribution has probability density function $a(u)$ with mean interarrival time $1/\alpha$. There are $m$ server and service times of batches are assumed to be independent and exponentially distributed with mean service time $1/\beta$. Let the time axis be slotted into intervals of equal length, so that $\lambda = \alpha \Delta$ and $\mu = \beta \Delta$, where $\Delta > 0$ is sufficiently small. Now, using $\mu = \beta \Delta$ in (1), we obtain

$$c(j|i) = \begin{cases} 
1, & i = j = 0; \\
1 - \beta \Delta, & i = 1, j = 0; \\
\beta \Delta, & i = j = 1; \\
1 - \min(i, m) \beta \Delta + o(\Delta), & i = 2, 3, \ldots, m, j = 0; \\
\min(i, m) \beta \Delta + o(\Delta), & i = 2, 3, \ldots, m, j = 1; \\
o(\Delta), & i = 2, 3, \ldots, m, j = 2, 3, \ldots, \min(i, m); \\
0, & \text{otherwise,} 
\end{cases} \quad (25)$$

where $o(\Delta)$ denotes any function of $\Delta$ such that $\lim_{\Delta \to 0} \frac{o(\Delta)}{\Delta} = 0$.

Using $\lambda = \alpha \Delta$, $\mu = \beta \Delta$ and (25) in (12), we obtain

$$Q_{N,m} = (\alpha \Delta) \left\{ \frac{1 - m \beta \Delta + o(\Delta)}{m \beta \Delta - o(\Delta)} \right\} Q_{N-1,m}.$$ 

Dividing both sides by $\Delta$ and taking the limit as $\Delta \to 0$ in the above equation, we obtain

$$Q_{N,m} = \frac{\alpha}{m \beta} Q_{N-1,m} = \rho b Q_{N-1,m}.$$ 

Using (25), putting $\lambda = \alpha \Delta$, $\mu = \beta \Delta$ and taking limit as $\Delta \to 0$ in Eqs. (13)-(18), we get,

$$Q_{n,m} = \rho b \left( Q_{n-1,m} - Q_{n,m}^+ \right), \quad n = N - 1, \ldots, N - b + 1,$$

$$Q_{n,m} = Q_{n+b,m} + \rho b \left( Q_{n-1,m}^+ - Q_{n,m}^+ \right), \quad n = N - b, N - b - 1, \ldots, 1.$$
\[
Q_{0,m} = \sum_{k=1}^{b} Q_{0,k} + \rho b \left( Q_{0,m-1} - Q_{0,m} \right),
\]
\[
Q_{0,k} = \left( \frac{k+1}{k} \right) Q_{0,k+1} + \frac{\rho bm}{k} \left( Q_{0,k-1} - Q_{0,k} \right), \quad k = m - 1, m - 2, \ldots, 1,
\]
\[
Q_{0,0} = 1 - \sum_{k=1}^{m-1} Q_{0,k} - \sum_{n=0}^{N} Q_{n,m},
\]

which match with the relations for the continuous-time GI/\(M^b/m/N\) queue reported in [25]. Further, using (25) and taking limit as \(\Delta \to 0\) in (19)–(22), we obtain \(Q_{0,k}^o = Q_{0,k} (0 \leq k \leq m - 1)\) and \(Q_{n,m}^o = Q_{n,m} (0 \leq n \leq N)\) as it should be.

References