Queueing Models and
Service Management
Vol. 7, No. 1, page 43–65, 2024

QMSM

Waiting Times in the Queueing System GI/M^{a,b}/c

Jing Gai* and Mohan Chaudhry

Department of Mathematics and Computer Science
Royal Military College of Canada
Kingston, ON K7K 7B4, Canada

(*Received December 2023; accepted February 2024*)

Abstract: In this paper, we present analytical techniques for determining the waiting times for a complex bulk-service, multi-server queueing system GI/M^{a,b}/c, where inter-arrival times follow an arbitrary distribution. The introduction of quorum $''a''$ increases the complexity of the model, and as of now, it remains unaddressed for this specific system. We derive a closed-form formula for computing the mean waiting-time. Numerical results are presented for the queueing systems with inter-arrival time distributions of Erlang, deterministic, and uniform. Additionally, We verify Little's Formula for these systems.

Keywords: Bulk Service; Multi-Server; Queues; Quorum; Waiting-time Distribution.

1. Introduction

Bulk-service queues are widely applied in numerous areas. Examples of their applications can be seen in shuttle-bus services, freight trains, express elevators, tour operators, and batch servicing in manufacturing processes. Queueing systems with multiple servers constitute a vital category of queueing processes, providing extensive practical applications across various fields. These systems are characterized by the ability to serve a group (or batch) of customers simultaneously by multiple servers. One example of this process is performance analysis of blood testing procedure for detecting viruses like HIV, HBV, HCV, where the expected outcomes are either positive or negative [1]. In this procedure, blood samples from multiple individuals would be mixed and tested as one. If the test comes back negative, everyone in the pool is clear, otherwise, each member in the pool is then tested individually. COVID-19 has affected the whole world for the past 3 years. It is extremely important to quickly, efficiently and economically isolate infected persons. It has been proved that group testing is the fastest, cheapest and most efficient technique. Multi-server bulk- service GI/M^{a,b}/c queue, with bounded batch sizes of $''a''$ and $''b''$, would be an appropriate model for such pandemic situations.

Compared to well-developed non-bulk queueing systems, bulk-service systems have an extensive mathematical theory. They are more complex and harder to deal with. This topic,

* Corresponding author
Email: Jing.Gai@rmc.ca

due to its perceived applicability, has attracted the attention of many researchers over several decades. At an early stage, some simple bulk-service models, such as single-server systems GI/M^b/1 and M/M^a/1 were studied by Shyu [13] and Gross et al. [7], respectively. Neuts [9] first introduced a quorum bulk service rule to create more complex models necessary to describe certain realistic situations. He considered a queueing system with Poisson arrivals and a general service-time distribution M/G$^{a,\,b}$/1, where a is the quorum and b is the capacity of the server. Easton and Chaudhry [5] extended these results to the case where the inter-arrival times were Erlangian with the η-stage, E$_\eta$/M$^{a,\,b}$/1. Later, Chaudhry and Madill [3] also, Samanta and Bank [11] solved for a more general queueing system GI/M$^{a,\,b}$/1. An alternate method was given in Neuts' book [10], where he describes the application of his matrix geometric approach to the GI/PH$^{a,\,b}$/1 system, which has a phase-type service-time distribution. However, these systems are single-server queues. For many other variations of bulk-service queues, such as bulk service queues with vacations or bulk-service queues of the type M/G/1, one may view the survey paper written by Sasikala and Indhira [12]. In this survey, which had over 100 publications, most of the models considered were single server queues.

Multi-server queueing systems constitute a significant category of queueing processes with wide-ranging practical applications. Nevertheless, these systems pose greater complexity and are more challenging to deal with compared to single-server queueing systems. This is especially true when the inter-arrival time distribution is arbitrary. Medhi [8] investigated a queue with Poisson arrivals M/M$^{a,\,b}$/c, but his method was not analytically tractable for $c > 2$. Related work was conducted by Sim [16] using algorithmic methods. Sim [15] also solved the η-phase Erlangian arrivals E$_\eta$/M$^{a,\,b}$/c system. A formula for passenger waiting time distribution in the queue was derived. However, this formula contained $c \times (a-1)+1$ unknowns and these variables were obtained through a recursive scheme. Additionally, the mean waiting-time remains unprovided, and Little's formula has not been verified.

The most relevant model to the model GI/M$^{a,\,b}$/c is GI/M^b/c, where the quorum was set to 1. Goswami et al. [6] solved the finite-buffer GI/M^b/c model by the supplementary variable technique. Shyu [14], as well as Chaudhry and Templeton [4], dealt with the distribution of the number of customers in the system without considering the server being busy or idle. Therefore, there is no information regarding server utilization. Moreover, the numerical results for the system GI/M^b/c are not available. The quorum $''a''$ refers to the minimum number of customers that are required in the waiting line before service starts. For example, a grouping blood test will not start until the quorum is met. Similarly, in transportation problems, a bus may not start until we have the quorum. This is an important policy desired by the service providers to reduce the business cost and maximize server utilization. The addition of the quorum policy makes the model closer to the real situation, but it also makes the model more complex to study. In consideration of this, a two-dimensional Markov chain is involved, where the first dimension corresponds to the state of the servers (busy or idle), and the second dimension corresponds to the number of customers in the queue.

To ensure the practical applicability of our model, we investigate both analytic and computational aspects to assess the performance of the queueing system GI/M$^{a,\,b}$/c system in the steady state. It appears that this particular system has not yet been addressed in the existing queueing literature. The model GI/M$^{a,b}/c$ that we examine integrates most of the models [3, 4, 5, 7, 8, 11, 13, 14, 15, 16, 19] mentioned earlier as special cases. In a previous work [2], we presented closed-form solutions for the queue-length distributions at three different epochs: pre-arrival epoch (p.a.e.), random epoch (r.e.), and post-departure epoch (p.d.e.). Our analysis covered both the busy and idle states of the system. In this paper, we specifically explore waiting times in the queueing system GI/M$^{a,b}/c$. We confirm our model's validity by generating numerical results, demonstrating the desired level of accuracy at trivial computational costs.

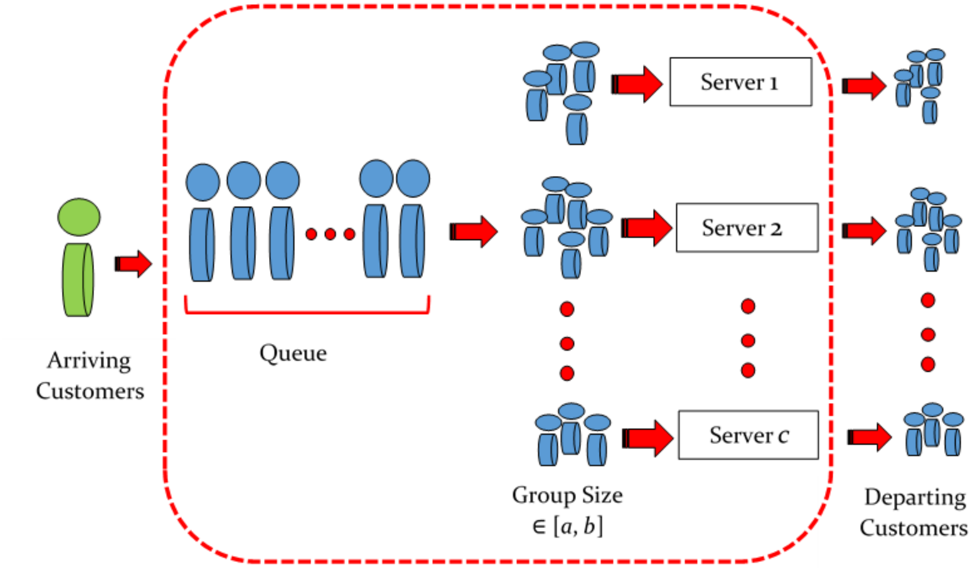

Figure 1. GI/M$^{a,b}/c$ Queueing Model

2. Model Description

In this continuous-time queueing system GI/M$^{a,b}/c$ (see Figure 1), there are c independent servers, each serving at the rate of μ. Customers arrive at a rate of λ according to a renewal process with an arbitrary inter-arrival time distribution $A(t)$. One of the idle c servers starts the service as soon as the number of customers (including the new arriving customer) in the queue reaches quorum $''a''$. Each of the c servers is capable of serving up to b customers simultaneously. This means that if the server completes a service and finds less than the quorum $''a''$ in the queue, it will remain idle until $''a''$ is reached. The service times of each server are independently, identically exponentially distributed random variables (i.i.e.d.r.v.$'$s). We consider the system to be in a steady state with the traffic intensity $\rho = \dfrac{\lambda}{bc\mu} < 1$. The queue discipline follows first-come first-serve (FCFS) by batches.

The states occurring at the instants immediately before the arrivals form an embedded Markov chain (I.M.C.). The state observed by an arriving customer can be described by

(S_n, n), where $n \geq 0$ is the queue-length and S_n is a supplementary flag defined as

$$S_n = \begin{cases} I(k), & \text{if } k \text{ servers are idle}, \quad 1 \leq k \leq c, \quad 0 \leq n \leq a-1, \\ B, & \text{if all servers are busy}, \quad n \geq 0. \end{cases} \tag{1}$$

We define the system as busy if all the servers are busy ($S_n = B$), and idle if at least one server is idle ($S_n = I(k), k$ is the number of idle servers). The queue-length n can be written as $n = qb + n_0, 0 \leq n_0 \leq b-1$, where q is the nearest lower non-negative integer of the fraction n/b, denoting the available number of full size batches (the batch size is b) in the queue waiting for service.

We define the following probabilities:

1. $[l|m;t]$ and $[l|m]$, $0 \leq l \leq m \leq c$. In this situation, $q = 0$, and there are less than a customers waiting in the queue at the beginning of the period. Here,

$$[l|m;t] = \binom{m}{l}(1-e^{-\mu t})^l (e^{-\mu t})^{m-l}$$

is the conditional probability that l of m servers complete services during an inter-arrival period of duration t, given that m servers are busy ($c-m$ servers are idle) at the beginning of the period. Moreover, $[l|m]$ is defined as

$$[l|m] = \int_0^\infty [l|m;t]dA(t), \qquad 0 \leq l \leq m \leq c. \tag{2}$$

2. $\{l|c;q\}$, $0 \leq l \leq c$, is the conditional probability that l of c servers become idle during an inter-arrival period, given that all c servers are busy at the beginning of the period, and q ($q \geq 1$) batches of customers are waiting for the services. Assume that a time V has elapsed when the last batch of q batches enters service. In this case, the c servers have been processed at a rate of $c\mu$ until time V has elapsed. When all c servers are busy, the number of departed batches follows a Poisson process with a rate $c\mu$. The time V is Erlang-distributed, so it is the sum of q exponential random variables with a rate $c\mu$, implying that the probability density function (p.d.f.) of V is given by

$$p(v) = \frac{(c\mu)(c\mu v)^{q-1}e^{-c\mu v}}{(q-1)!}, \qquad v > 0.$$

After all the waiting q batches leave the queue, there is time $t-V$ remaining to have l batches processed. The probability that these l batches complete the service during period $t-V$ is $[l|c;t-V]$. Therefore

$$\{l|c;q\} = \int_0^\infty \int_0^t \binom{c}{l}(1-e^{-(t-v)\mu})^l (e^{-(t-v)\mu})^{c-l} \frac{(c\mu)(c\mu v)^{q-1}e^{-c\mu v}}{(q-1)!} dv dA(t). \tag{3}$$

3. $(l|c)$, $0 \leq l \leq c$, is the conditional probability that l batches complete service during an inter-arrival period of duration t, given that all the c servers are busy at the beginning of the period and still busy at the end of the period. When all the servers are busy

during an inter-arrival time period, for the queueing model GI/M$^{a,b}/c$, the service times for batches are independent identically distributed random variables (i.i.d.r.v.), having exponential distributions. Thus, the number of batches that complete service during an arbitrary inter-arrival time will have a Poisson distribution, which implies that the probability of l service completions during an inter-arrival time A is

$$\int_0^\infty \frac{e^{-c\mu t}(c\mu t)^l}{l!} dA(t), \qquad 0 \le l \le c, \tag{4}$$

and the probability generating function (p.g.f.) is

$$D(z) = \bar{a}(c\mu(1-z)), \tag{5}$$

where $\bar{a}(\alpha)$ is the Laplace–Stieltjes transform (L.-S.T.) of $A(t)$, i.e., $\bar{a}(\alpha) = \int_0^\infty \exp(-\alpha t) dA(t)$ and

$$K_0 = \bar{a}(c\mu) = \int_0^\infty \exp(-c\mu t) dA(t). \tag{6}$$

3. Queue-Length Distributions at Pre-Arrival Epoch

For ease of reference, we provide the results established in Chaudhry and Gai [2], which are necessary for finding waiting times for the queueing model GI/M$^{a,b}/c$ in the subsequent sections.

Let J_r be the system state on the arrival of the r^{th} customer who sees n customers in the queue. The entry of the one-step transition probability matrix (t.p.m.) T from state (S_i, i) to state (S_j, j) is

$$[T_{(S_i,i),(S_j,j)}] = P(J_{r+1} = (S_j, j) | J_r = (S_i, i)), \qquad i \ge 0, j \ge 0,$$

implying that the $(r+1)^{th}$ arriving customer sees j customers waiting in the queue with the server state S_j, given that the previous r^{th} arriving customer saw i customers waiting in the queue with the server state S_i.

The Markov chain for this system is two-dimensional rather than the usual one-dimensional. The t.p.m. can be formed as four sub-matrices as

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_{Idle \to Idle} & \mathbf{T}_{Idle \to Busy} \\ \mathbf{T}_{Busy \to Idle} & \mathbf{T}_{Busy \to Busy} \end{bmatrix}. \tag{7}$$

Since the Markov chain under consideration is irreducible, positive recurrent and aperiodic, it has a limiting distribution if and only if $\rho = \frac{\lambda}{bc\mu} < 1$. In view of this, $\lim_{r \to \infty} P(J_r = (S_n, n)) = X(S_n, n)$ exists. In this case, the limiting distribution is given by $\mathbf{X} = \mathbf{XT}$ where $\mathbf{T}$ is t.p.m. defined in equation 7, and the vector $\mathbf{X}$ has the form

$$\begin{aligned} \mathbf{X} =& [X(I(c), 0), \cdots, X(I(c), a-1), \cdots, X(I(1), 0), \cdots, \\ & X(I(1), a-1), X(B, 0), \cdots X(B, 1), \cdots], \end{aligned} \tag{8}$$

where $X(I(k),n), 0 \leq n < a$ and $X(B,n), n \geq 0$, respectively, denote the p.a.e. (pre-arrival epoch) unnormalized probabilities that an arriving customer sees n customers in queue, k of c servers idle, and n customers in queue, with all servers busy. If such a vector **X** exists, it will represent the vector of steady state probabilities at p.a.e. up to a certain normalizing constant.

The queue-length distributions at p.a.e. can be evaluated by using the following theorems outlined in Chaudhry and Gai [2].

Theorem 1. *For the queueing system GI/M^{a,b}/c, in the steady state case, the busy-server probabilities of queue-length at pre-arrival epoch are given by* $P^-(B,n) = X(B,n)/C_N = w^n/C_N, n \geq 0$, *where* w *is a real root inside the unit circle of equation* $D(z^b) = z = \bar{a}(c\mu(1-z^b))$ *and* C_N *is a normalizing constant given by* $C_N = \sum_{j=1}^{c}\sum_{i=0}^{a-1} X(I(j),i) + \frac{1}{1-w.}$

Theorem 2. *For the queueing system GI/M^{a,b}/c, in the steady state case, the idle server probabilities of queue-length at the pre-arrival epoch are given by* $P^-(I(k),n) = X(I(k),n)/C_N, 0 \leq n < a-1, 1 \leq k \leq c$, *where* C_N *is a normalizing constant given by* $C_N = \sum_{j=1}^{c}\sum_{i=0}^{a-1} X(I(j),i) + \frac{1}{1-w}$ *and* $X(I(k),n)$ *satisfy the following equations:*

$$X(I(1),a-1) = \frac{1}{(1-w)K_0}(1 - w^{a-b} + K_0 w^{a-b-1} - K_0), \tag{9}$$

$$X(I(k),j) = \sum_{m=1}^{k} X(I(m),j-1)[(k-m)|(c-m)] + w^{j-1}([k|c] + J(k)), 1 < j < a-1, \tag{10}$$

$$X(I(k),0) = \sum_{m=1}^{k+1} X(I(m),a-1)[(k-m+1)|(c-m+1)] + \frac{w^{a-b-1}-1}{1-w}J(k), \tag{11}$$

where $J(k) = \sum_{i=1}^{\infty} w^{ib}\{k|c;i\}$, K_0 *is defined in equation 6, and*

$$J(k) = c\mu w^b \int_0^{\infty}\int_0^{t} \binom{c}{k}(1 - e^{-(t-v)\mu})^k (e^{-(t-v)\mu})^{c-k} e^{-c\mu v(1-w^b)} dv dA(t). \tag{12}$$

4. Waiting - Time Distributions

Define

- Random variable $V_Q(S_n,n)$ as the actual waiting time in queue for a customer who finds the system in state (S_n,n) on his arrival, with cumulative distribution function (c.d.f.) $F_{V_Q(S_n,n)}(t)$;

- Random variable $A_w(a-j-1)$ as waiting time for an arriving customer who must wait for $a-j-1$ arrivals to complete the quorum, with c.d.f. $F_{A_w(a-j-1)}(t)$. Where $j \in [0, a-1]$ is the number of customers waiting in the queue who are in front of him;
- Random variable $B_w(q+1)$ as waiting time for an arriving customer who must wait for $q+1$ service completions, with c.d.f. $F_{B_w(q+1)}(t)$, where $q \geq 0$ represents the number of full-sized (batch size $= b$) batches. There are total $qb+ n_0$ ($n_0 \in [a-1, b-1]$) customers waiting in queue who are in front of him with $s = 1$ if the system is busy; 0, otherwise.

Then $V_Q(S_n, n) = \text{Max}[A_w(a-j-1), B_w(q+s)]$, where $V_Q(S_n, n)$ is a random variable of the maximum of two independent random variables $A_w(a-j-1)$ and $B_w(q+s)$. Thus its p.d.f will be

$$dF_{V_Q(S_n,n)}(t) = F_{A_w(a-j-1)}(t) \cdot dF_{B_w(q+s)}(t) + dF_{A_w(a-j-1)}(t) \cdot F_{B_w(q+s)}(t), \tag{13}$$

and the unconditional waiting-time

$$dF_{V_Q}(t) = \sum_s \sum_n dF_{V_Q(S_n,n)}(t). \tag{14}$$

For an arriving customer, he will find the system in one of the following four classes of states:

1. $V_Q(S_n, n) = 0$, $dF_{V(S_n,n)}(t) = \delta(t)dt$, where $\delta(t)$ is the Dirac delta function.
   In this case, the system is idle (at least one server is idle) and the number of customers waiting in the queue is $a-1$ (the system state: $S_n = idle, n = a-1$). An arriving customer does not need to wait, and he immediately gets into service whenever he enters the system.
2. $V_Q(S_n, n) = A_w(a-j-1)$, $dF_{V_Q(S_n,n)}(t) = dF_{A_w(a-j-1)}(t)$.
   In this case, the system is idle, and the number of customers waiting in queue is less than $a-1$ (the system state: $S_n = idle, n = j \in [1, a-1]$). An arriving customer has to wait for the arrival of $a-j-1$ customers to reach the quorum $''a''$.
3. $V_Q(S_n, n) = B_w(q+1)$, $dF_{V_Q(S_n,n)}(t) = dF_{B_w(q+1)}(t)$.
   In this case, the system is busy and there are q full-sized batches and n_0 ($n_0 \in [a-1, b-1]$) of customers waiting in the queue (the system state: $S_n = Busy, n = qb + n_0$). An arriving customer will join the $(q+1)$ th batch and he has to wait for the completion of services of $q+1$ batches.
4. $V_Q(S_n, n) = Max[A_w(a-j-1), B_w(q+1)]$.
   In this case, the system is busy and $n = qb+j, j < a-1$ customers wait in queue (the system state: $S_n = Busy, n = qb+j, j < a-1$). An arriving customer has to wait until both the services of $q+1$ batches are completed and $a-j-1$ customers arrive.

Combining the above four cases, the probability distribution function (p.d.f.) for the unconditional waiting time $dF_V(t)$ in equation 14 can be written as

$$dF_{V_Q}(t) = \sum_{k=1}^{c} P^-(I(k), a-1)\delta(t)dt$$

$$+\sum_{k=1}^{c}\sum_{j=0}^{a-2}P^{-}(I(k),j)dF_{A_w(a-j-1)}(t)$$
$$+\sum_{q=0}^{\infty}\sum_{n_0=a-1}^{b-1}P^{-}(B,qb+n_0)dF_{B_w(q+1)}(t)$$
$$+\sum_{q=0}^{\infty}\sum_{j=0}^{a-2}P^{-}(B,qb+j)dF_{V_Q(S_n,n)}(t), \tag{15}$$

where $P^{-}(I(k),n)$ and $P^{-}(B,n)$ can be evaluated by using theorems 1 and 2. However, the result in equation 15 is too complex to be useful if the inter-arrival time distributions are not closed (as we assumed it is arbitrary).

5. Mean Waiting-time

Although the waiting-time distribution cannot be developed in general beyond equation 15, its expected value is still of interest in order to obtain a quantitative assessment of the waiting-time costs.

Define W_Q as the mean waiting time for a customer who just arrived the system, using equation 15,

$$W_Q = \int_0^{\infty} t\,dF_V(t) = \underbrace{\sum_{k=1}^{c}\sum_{j=0}^{a-2}P^{-}(I(k),j)E[A_w(a-j-1)]}_{T(16)-1}$$
$$+\underbrace{\sum_{q=0}^{\infty}\sum_{n_0=a-1}^{b-1}P^{-}(B,qb+n_0)E[B_w(q+1)]}_{T(16)-2}$$
$$+\underbrace{\sum_{q=0}^{\infty}\sum_{j=0}^{a-2}P^{-}(B,qb+j)\int_0^{\infty} t\,dF_{V_Q(S_n,n)}(t)}_{T(16)-3}. \tag{16}$$

Equation 16 poses mathematical challenges for usability as the last two terms, $T(16)-2$ and $T(16)-3$, involve the infinite series, and also the third term $T(16)-3$ incorporates an indefinite integral. To mitigate the approximation issues arising from truncating infinite sums and enhance computational efficiency, we proceed to derive the closed form of the mean waiting-time.

First, for convenience, let us define two expressions which will become apparent in the following derivations.

$$T_w = \frac{1}{C_N\left(1-w^b\right)}\left[aw^{a-1}+\frac{w^a-w^b}{1-w}\right], \tag{17}$$

$$T_1 = 1 - T_w + \frac{1}{C_N}\frac{aw^{a-1}}{1-w^b}, \tag{18}$$

where w is the root of equation 5.

A. Simplifying T(16)-1:

Using the results from theorems 1 and 2, in addition to the following two relations (see the proof in Appendix A),

$$\begin{aligned}\sum_{k=1}^{c} P^{-}(I(k), j-1) &= \sum_{k=1}^{c} P^{-}(I(k), a-1) + \frac{1}{C_N}\frac{w^{a-1} - w^{j-1}}{1-w^b},\\ \sum_{k=1}^{c} (P^{-}(I(k), a-1) &= \frac{1}{a}(1-T_w),\end{aligned} \tag{19}$$

and the fact that $E[A_w(a-j-1)] = (a-j-1)/\lambda$, the term T(16)-1 can be simplified as

$$\begin{aligned}&\underbrace{\sum_{k=1}^{c}\sum_{j=1}^{a-2} P^{-}(I(k), j)E[A_w(a-j-1)]}_{T(16)-1}\\ &= \frac{1}{\lambda}\sum_{j=1}^{a-1}(a-j)\left[\sum_{k=1}^{c}(P^{-}(I(k), a-1) + \frac{1}{C_N}\frac{w^{a-1}-w^{j-1}}{1-w^b}\right]\\ &= \frac{a(a-1)}{2\lambda}\sum_{k=1}^{c}(P^{-}(I(k), a-1) + \frac{1}{C_N\lambda}\frac{a(a-1)}{2}\frac{w^{a-1}}{1-w^b} - \frac{1}{C_N\lambda}\sum_{j=1}^{a-1}(a-j)\frac{w^{j-1}}{1-w^b}\\ &= \frac{a-1}{2\lambda}(1-T_w) + \frac{1}{C_N\lambda}\frac{a(a-1)}{2}\frac{w^{a-1}}{1-w^b} - \frac{1}{C_N\lambda}\sum_{j=1}^{a-1}(a-j)\frac{w^{j-1}}{1-w^b}.\end{aligned} \tag{20}$$

B. Simplifying T(16)-2:

Using $E[B_w(q+1)] = (q+1)/c\mu$ and $P^{-}(B, qb+n_0) = \frac{w^{qb+n_0}}{C_N}$ (theorem 1),

$$\begin{aligned}&\underbrace{\sum_{q=0}^{\infty}\sum_{n_0=a-1}^{b-1} P^{-}(B, qb+n_0)E[B_w(q+1)]}_{T(16)-2}\\ =&\frac{1}{C_N c\mu}\sum_{q=0}^{\infty}(q+1)w^{qb}\sum_{n_0=a-1}^{b-1} w^{n_0}\\ =&\frac{1}{C_N c\mu}\frac{1}{(1-w^b)^2}\frac{w^{a-1}-w^b}{1-w}.\end{aligned} \tag{21}$$

C. Simplifying T(16)-3:

The term T(16)-3 of equation 16 requires more analysis than the first two. For GI/M^{a,b}/c, the service time to complete $q+1$ batches has a gamma distribution with parameters $c\mu$ and $q+1$, then c.d.f. is

$$F_{B_w(q+1)}(t) = 1 - \sum_{i=0}^{q} (c\mu t)^i \exp(-c\mu t)/i! \tag{22}$$

and p.d.f. is

$$dF_{B_w(q+1)}(t) = \frac{(c\mu)^{q+1} t^q \exp(-c\mu t)}{q!} dt. \tag{23}$$

Using equations 13, 22 and 23, and the result of

$$\int_0^\infty t dF_{A_w(a-j-1)}(t) = E[A_w(a-j-1)] = (a-j-1)/\lambda.$$

T(16)-3 can be written as

$$\begin{aligned}
&\underbrace{\sum_{q=0}^{\infty}\sum_{j=0}^{a-2} P^-(B, qb+j)\int_0^\infty t dF_{V_Q(S_n,n)}(t)}_{T(16)-3} \\
=&\sum_{q=0}^{\infty}\sum_{j=0}^{a-2} P^-(B, qb+j)\int_0^\infty t F_{A_w(a-j-1)}(t)\frac{(c\mu)^{q+1} t^q \exp(-c\mu t)}{h!} dt \\
&+\sum_{q=0}^{\infty}\sum_{j=0}^{a-2} P^-(B, qb+j)\int_0^\infty t(1-\sum_{i=0}^{q}(c\mu t)^i \exp(-c\mu t)/i!) dF_{A_w(a-j-1)}(t) \\
=&\sum_{q=0}^{\infty}\sum_{j=0}^{a-2} P^-(B, qb+j)\frac{a-j-1}{\lambda} \\
&+\sum_{q=0}^{\infty}\sum_{j=0}^{a-2} P^-(B, qb+j)\frac{(c\mu)^{q+1}}{h!}\int_0^\infty t^{q+1}\exp(-c\mu t) F_{A_w(a-j-1)}(t) dt \\
&-\sum_{q=0}^{\infty}\sum_{j=0}^{a-2} P^-(B, qb+j)\frac{1}{c\mu}\sum_{i=0}^{q}\int_0^\infty \frac{(c\mu t)^{i+1}\exp(-c\mu t)}{i!} dF_{A_w(a-j-1)}(t).
\end{aligned} \tag{24}$$

The first term of equation 24 can be further simplified as

$$\sum_{q=0}^{\infty}\sum_{j=0}^{a-2} P^-(B, qb+j)\frac{a-j-1}{\lambda} = \frac{1}{\lambda C_N(1-w^b)}\sum_{j=1}^{a-1} w^{j-1}(a-j), \tag{25}$$

which will cancel with the last term of equation 20. The last two terms in equation 24 can be considered as Laplace-Stieltjes transforms, and it can be shown (see Appendix B) that

$$\begin{aligned}&\sum_{q=0}^{\infty}\sum_{j=0}^{a-2}P^{-}(B,qb+j)\frac{(c\mu)^{q+1}}{q!}\int_{0}^{\infty}t^{q+1}\exp(-c\mu t)F_{A_w(a-j-1)}(t)dt\\&-\sum_{q=0}^{\infty}\sum_{j=0}^{a-2}P^{-}(B,qb+j)\frac{1}{c\mu}\sum_{i=0}^{q}\int_{0}^{\infty}\frac{(c\mu t)^{i+1}\exp(-c\mu t)}{i!}dF_{A_w(a-j-1)}(t)\\=&\frac{(a-1)w^{a-1}}{C_N c\mu(1-w^b)^2}.\end{aligned} \tag{26}$$

Plugging equations 25 and 26 into equation 24, we have

$$\begin{aligned}&\underbrace{\sum_{q=0}^{\infty}\sum_{j=0}^{a-2}P^{-}(B,qb+j)\int_{0}^{\infty}tdF_{V(B,n)}(t)}_{T(43)-3}\\&=\frac{1}{\lambda C_N(1-w^b)}\sum_{j=1}^{a-1}w^{j-1}(a-j)+\frac{(a-1)w^{a-1}}{C_N c\mu(1-w^b)^2}.\end{aligned} \tag{27}$$

Finally, we can obtain W_Q by adding the results in equations 20, 21 and 27,

$$\begin{aligned}W_Q=&\frac{a-1}{2\lambda}(1-T_w)+\frac{1}{C_N\lambda}\frac{a(a-1)}{2}\frac{w^{a-1}}{1-w^b}-\frac{1}{C_N\lambda}\sum_{j=1}^{a-1}(a-j)\frac{w^{j-1}}{1-w^b}\\&+\frac{1}{C_N c\mu}\frac{1}{(1-w^b)^2}\frac{w^{a-1}-w^b}{1-w}\\&+\frac{1}{C_N\lambda(1-w^b)}\sum_{j=1}^{a-1}w^{j-1}(a-j)+\frac{(a-1)w^{a-1}}{C_N c\mu(1-w^b)^2}\\=&\frac{a-1}{2\lambda}\underbrace{\left[1-T_w+\frac{1}{C_N}\frac{aw^{a-1}}{1-w^b}\right]}_{T_1}+\frac{T_w}{c\mu(1-w^b)}\\=&\frac{a-1}{2\lambda}T_1+\frac{T_w}{c\mu(1-w^b)}.\end{aligned} \tag{28}$$

The queueing model GI/M^{a,b}/c is highly intricate. However, our expression for the mean waiting-time is presented in a concise and closed form, in terms of system parameters (λ,μ,a,b,c), root w and normalizing constant C_N. This alleviates computational challenges, especially for heavy-traffic or large batch size queueing systems.

6. Special cases

1. $E_\eta/M^{a,b}/c$ [15] and $M/M^{a,b}/c$ [8], [16].
If we assume that the inter-arrival time is Erlangian with phase η, then the time for waiting for $a-j-1$ arrivals has a gamma distribution with parameter $\eta\lambda$ and $\eta(a-j-1)$. Thus, using equations 22 and 23, the waiting-time p.d.f. $dF_V(t)$ in equation 15 can be further simplified to

$$\begin{aligned} dF_{V_Q}(t) = &\left[1 - \sum_{i=0}^{a-j-2} (\eta\lambda t)^i \exp(-\eta\lambda t)/i!\right] \frac{(c\mu)^{q+1} t^q \exp(-c\mu t)}{q!} dt \\ &+ \left[1 - \sum_{i=0}^{q} (c\mu t)^i \exp(-c\mu t)/i!\right] \frac{(\eta\lambda)^{a-j-1} t^{a-j-2} \exp(-\eta\lambda t)}{(a-j-2)!} dt. \end{aligned} \tag{29}$$

The proof of equation 29 is shown in Appendix C. Sim [15] addressed $E_\eta/M^{a,b}/c$ system by using a different approach. The formula he proposed for passenger waiting-time distribution contained $c \times (a-1)+1$ unknowns and these variables are resolved through a recursive scheme. However, his work did not extend beyond this point, and no supplementary results were accessible. Our p.d.f. plots (see Figure 3) generated using equation 29 match those obtained using Sim's method. As [15] did not provide the mean waiting-time, a direct comparison with their results is not possible. Nevertheless, since we have achieved an identical queue-length distribution at r.e. and confirmed Little's Formula (refer to section 7-1), it indirectly validates the accuracy of our formula for the mean waiting-time for this model.

$M/M^{a,b}/c$ is a special case of the $E_\eta/M^{a,b}/c$ by setting $\eta = 1$. Obviously, the p.d.f. of waiting-time from two different approaches should be identical. A mean waiting-time formula was presented in [16]. We numerically confirmed that the results obtained using equation 28 and the formula in [16] are consistent.

2. $GI/M^b/c$ [4].
In this case, since $a-j-1$ is zero, the equation 15 can be simplified to

$$\begin{aligned} dF_{V_Q}(t) = &\sum_{k=1}^{c} P^-(I(k),0)\delta(t)dt + \sum_{q=0}^{\infty}\sum_{n_0=0}^{b-1} P^-(B, qb+n_0) dF_{B_{w(q+1)}}(t) \\ = &\frac{1}{C_N}\left\{\frac{w}{1-w}\delta(t)dt + c\mu \sum_{n_0=0}^{b-1} w^{n_0}\left[\sum_{q=0}^{\infty} \frac{(c\mu t w^b)^q}{q!}\right] \exp(-c\mu t)dt\right\} \\ = &\frac{1}{C_N}\left\{\frac{w}{1-w}\delta(t)dt + c\mu \sum_{n_0=0}^{b-1} w^{n_0} \exp(-c\mu t(1-w^b)dt\right\} \end{aligned}$$

by using equation 23, $P^-(B, qb+n_0) = w^{qb+n_0}/C_N$, and $\sum_{k=1}^{c} P^-(I(k),0) = 1 - P_B^- = \frac{w}{C_N(1-w)}$. This matches the result in [4] for the queue $GI/M^b/c$.

3. GI/M^{a,b}/1 [3].
   The system GI/M^{a,b}/1 is a special case of GI/M$^{a,b}/c$ when $c = 1$. If $c = 1$, the root equation (5) is simplified to $D(z) = z = \bar{a}(\mu(1-z))$, which agrees with the root equation in [3]; consequently, the same results of $X(B,0), \cdots, X(B,1), \cdots, X(B,M)$ can be obtained. Since $k = m = c = 1, [0|0] = 1, [1|1] = 1 - [0|1] = 1 - K_0$, and $\sum_{i=1}^{\infty} w^{ib}\{1|1;i\} = \frac{1}{(1-w^b)}(w^b - w + (1-w^b)K_0)$. Equation 10 can be simplified to

$$\begin{aligned} X(I(1),j) &= X(I(1),j-1) + w^{j-1}(1 - K_0 + \frac{1}{(1-w^b)}(w^b - w + (1-w^b)K_0)) \\ &= X(I(1),j-1) + w^{j-1}\frac{1-w}{1-w^b}. \end{aligned}$$

   Then, we can further simplify C_N (defined in Theorem 1) as

$$C_N = \sum_{j=0}^{a-1} X(I(1),j) + \frac{1}{1-w} = \frac{aC_1 + K_0(w^a - w^b)}{K_0(1-w^b)(1-w)},$$

   where, $C_1 = (1-K_0)(1 + w^a - w^b) - (w - K_0)w^{a-b-1}$.

   Plug C_N and T_w into equation 28, set $c = 1$, we have

$$\begin{aligned} W_Q &= \frac{a-1}{2\lambda} \underbrace{\left[1 - T_w + \frac{1}{C_N}\frac{aw^{a-1}}{1-w^b}\right]}_{T_1} + \frac{T_w}{\mu(1-w^b)} \\ &= \frac{a-1}{2\lambda}\left[1 - \frac{w^a - w^b}{C_N(1-w)(1-w^b)}\right] + \frac{1}{C_N\mu(1-w^b)^2}\left[aw^{a-1} + \frac{w^a - w^b}{1-w}\right] \\ &= \frac{C_1 a(a-1)}{2\lambda(aC_1 + K_0(w^a - w^b))} + \frac{1}{C_N\mu(1-w^b)^2}\left[aw^{a-1} + \frac{w^a - w^b}{1-w}\right]. \end{aligned} \tag{30}$$

   This agrees with the mean waiting time equation in [3].

4. GI/M/c by [17, 19].
   When $a = b = 1$, $T_w = \frac{1}{C_N}\frac{1}{1-w}$. Plug it into equation 28,

$$W_Q == \frac{T_w}{c\mu(1-w)} = \frac{1}{C_N c\mu(1-w)^2} \tag{31}$$

   agrees the result in queueing system GI/M/c [17, 19].

7. Numerical Results

In this section, we present some numerical results for various inter-arrival time distributions such as η-phase Erlang (E$_\eta$), deterministic (D), and uniform (U). The root equation (see equation (5)), probability density functions (p.d.f.) of inter-arrival time A, and p.d.f. of a random period time R for these three distributions are summarized in Table 1.

Table 1. Root Equations, p.d.f.s of $A(t)$, $R(t)$, and mean value of of $A(t)$ for $E_\eta/M^{a,b}/c$, $D/M^{a,b}/c$ and $U/M^{a,b}/c$.

Inter-arrival time distributions	Root Equations (Equation (14))	p.d.f. of $A(t)$	p.d.f. of $R(t)$	$E(A)$
η-phase Erlang	$\left(\frac{\eta\rho b}{\eta\rho b+1-z^b}\right)^\eta - z = 0$	$\frac{(\lambda\eta)^\eta t^{\eta-1}\exp(-\lambda\eta t)}{(\eta-1)!}$	$\lambda\sum_{n=0}^{\eta-1}\frac{(\lambda\eta t)^n\exp(-\lambda\eta t)}{n!}$	$1/\lambda$
Deterministic	$\exp(-\frac{1-z^b}{\rho b}) - z = 0$	$\delta(t-1/\lambda)$	$\begin{cases}\lambda, & \text{if } t<\frac{1}{\lambda}\\ 0, & \text{if } t\geq\frac{1}{\lambda}\end{cases}$	$1/\lambda$
Uniform	$\frac{\exp(-\frac{1-z^b}{\rho b})}{\varphi c\mu(1-z^b)}\times[\exp(\varphi c\mu(1-z^b)/2)$ – $\exp(-\varphi c\mu(1-z^b))/2]-z=0$ $\varphi = t_2 - t_1$, is the interval width	$1/\varphi$	$\begin{cases}\lambda, & \text{if } t<t_1\\ \frac{1}{\varphi}+\frac{\lambda}{2}-\frac{\lambda t}{\varphi}, & \text{if } t_1\leq t<t_2\\ 0, & \text{if } t\geq t_2\end{cases}$ $t_1=\frac{1}{\lambda}-\frac{\varphi}{2},\quad t_2=\frac{1}{\lambda}+\frac{\varphi}{2}$	$1/\lambda$

7.1. Verification of Little's Formula $L_Q = \lambda W_Q$

The queueing formula $L = \lambda W$, known as Little's Formula, and its generalizations have been studied by many researchers. In the early years, the mathematical proofs were all concerned with single arrival single service systems. Lately, Little's Formula has been verified to the bulk-arrival systems, e.g., $GI^X/M/c$ and $GI^X/M/1$ [19]. It is also known that Little's Formula still holds for the bulk-service with quorum queueing systems $GI/M^{a,b}/1$ [3], $M/M^{a,b}/c$ [16]. Now our current work extends it to the system $GI/M^{a,b}/c$.

We verified the Little's Formula $L_Q = \lambda W_Q$ using various inter-arrival time distributions. Here, we present some examples in Table 2. Clearly, the results presented in Table 2 support Little's Formula. In this verification, we consider three systems: $E_6/M^{a,10}/5$, $D/M^{a,10}/5$ and $U/M^{a,10}/5$ ($[0.875/\lambda, 1.125/\lambda]$, $\phi = 0.25/\lambda$) with varied $a = 1, 4, 7$, and $\rho = 0.1, 0.5, 0.9$. All the examples we considered have the same mean value of the inter-arrival time $E(A) = 1/\lambda$. The mean queue-length at random epoch, L_Q, can be obtained from Chaudhry and Gai [2], and W_Q is calculated using equation [28].

Table 2. Numerical verification of Little's formula, $E_6/M^{a,10}/5$, $D/M^{a,10}/5$ and $U/M^{a,10}/5$ ($[0.875/\lambda, 1.125/\lambda]$, $\phi = 0.25/\lambda$) $a = 1, 4, 7$, and $\rho = 0.1, 0.5, 0.9$.

ρ	λ	a	$E_6/M^{a,10}/5$			$D/M^{a,10}/5$			$U/M^{a,10}/5$		
			W_Q	λW_Q	L_Q	W_Q	λW_Q	L_Q	W_Q	λW_Q	L_Q
0.1	5	1	0.0816	0.4082	0.4082	0.0800	0.3998	0.3998	0.0800	0.4001	0.4001
		4	0.3000	1.5000	1.5000	0.3000	1.5000	1.5000	0.3000	1.5000	1.5000
		7	0.6000	3.0000	3.0000	0.6000	3.0000	3.0000	0.6000	3.0000	3.0000
0.5	25	1	0.2481	6.2025	6.2025	0.2464	6.1610	6.1610	0.2465	6.1623	6.1623
		4	0.1892	4.7317	4.7317	0.1876	4.6909	4.6909	0.1877	4.6921	4.6921
		7	0.1527	3.8177	3.8177	0.1514	3.7841	3.7841	0.1514	3.7852	3.7852
0.9	45	1	1.0502	47.259	47.259	1.0344	46.547	46.547	1.0349	46.569	46.569
		4	1.0229	46.031	46.031	1.0078	45.350	45.350	1.0083	45.372	45.372
		7	0.9395	42.276	42.276	0.9250	41.624	41.624	0.9254	41.644	41.644

7.2. Examples

In this section, we provide numerical results for the queueing model GI/M^{a,b}/c to illustrate the behavior of system performance measures.

In our first example, We assumed that a blood testing lab receives 30 blood samples per time unit, with a service rate of 10 per time unit. The inter-arrival time follows an Erlang distribution with the phase η set to 6. The minimum batch size a is fixed at 5. Figure 2 provides the the mean waiting-time for varied batch capacities b from 8 to 20, and number of servers c ranging from 6 to 15.

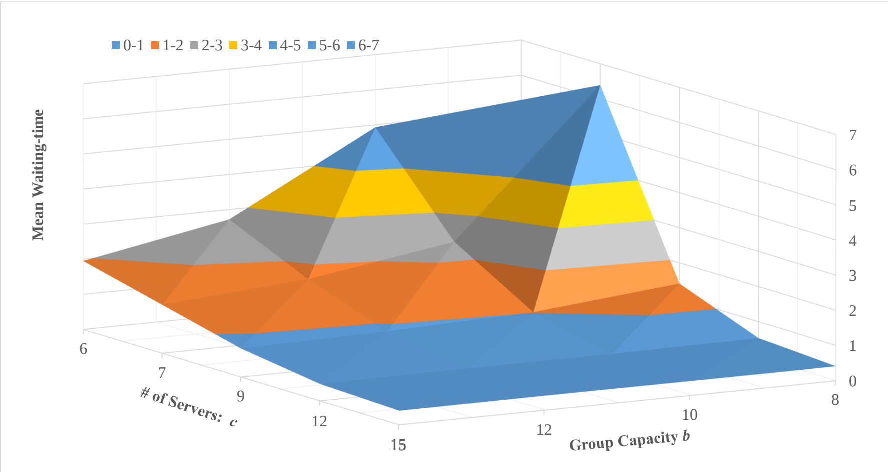

Figure 2. Mean Waiting-time for E$_6$/M^{5,b}/c.
$\lambda = 30, \mu = 1/10; b = 8, 10, 12, 15; c = 6, 7, 9, 12, 15.$

Numerous researchers have applied various queueing models to analyze blood testing procedures for virus detection. Bar-Lev et al. [1]) offered a comprehensive review in this field. Mathematical models serve as simplified representations of real-world phenomena, their simplifications, while essential for manageability and solvability, come with inherent limitations. Bar-Lev et al. [1] addressed the issue using M /M^{a,b}/c queueing model with impatient customers, but their group testing size "b" is restricted to 2. Tamrakar and Banerjee [18] explored M ^{X}/G^{a,y}/1 queue with optional service and queue length-dependent single (multiple) vacation, yet it is confined to a single-server system. Notably, these models may not be suitable for extensive scenarios, such as a large influenza pandemic like COVID-19.

In our study, the GI/M^{a,b}/c model demonstrates versatility by accommodating an arbitrary inter-arrival time distribution. Moreover, presenting all performance measures in closed forms affirms the validity of our model, ensuring accurate numerical results with minimal computational costs. This makes our model particularly well-suited for handling large parameters such as the arrival rate (λ), group testing sizes (b), and the number of servers (c), etc. However, if the value of the waiting-time is large, it becomes imperative to address another crucial aspect in our model: the aging or expiration date (as some blood samples may have expired). This will be a focus of one of our future works to enhance the model's suitability for real-world scenarios.

Table 3. Effect of Quorum for D/$M^{a,50}$/10, $\rho = 0.5$, $\lambda = 250$; $a = 1, 4, 7$, and $\mu = 1$.

Quorum *a*	Prob. Of Empty Waiting Line	Prob. of Idle System	Mean of Idle Servers	Mean Waiting-time at p.a.e.	Mean Queue-Length at p.a.e.	Mean Queue-Length at p.d.e.
1	0.032005	0.000656	0.000664	0.125418	30.857436	6.370870
5	0.032080	0.018147	0.019568	0.123391	30.365213	6.586150
10	0.032286	0.081815	0.098511	0.116949	28.802141	7.361795
15	0.032527	0.203639	0.285200	0.106389	26.240370	8.658705
20	0.032515	0.384797	0.653063	0.093165	23.034688	10.400283
25	0.031772	0.597199	1.265352	0.080853	20.059318	12.472310
30	0.029974	0.784359	2.089419	0.073993	18.423583	14.745989
35	0.027388	0.905056	2.976585	0.074409	18.572761	17.124420
40	0.024626	0.964050	3.783753	0.080213	20.043488	19.558312
45	0.022111	0.987648	4.458506	0.088696	22.171064	22.023560
50	0.019968	0.995983	5.009977	0.098208	24.551219	24.504550

Table 3 illustrates the effect of the quorum value (a) on various system performance measures, including the probability of an empty waiting line (column 2), the probability of the system being idle (column 3), the mean number of idle servers (column 4), the mean waiting-time at p.a.e. (column 5), the mean queue-length at p.a.e. (column 6), and at p.d.e. (column 7). Certain outcomes presented in Table 3 are derived from findings in another one of our research papers [2]. In this example, we choose a higher arrival rate of $\lambda = 250$ and a higher group testing size of $b = 50$. We set the service rate of $\mu = 1$, and the number of servers $c = 10$.

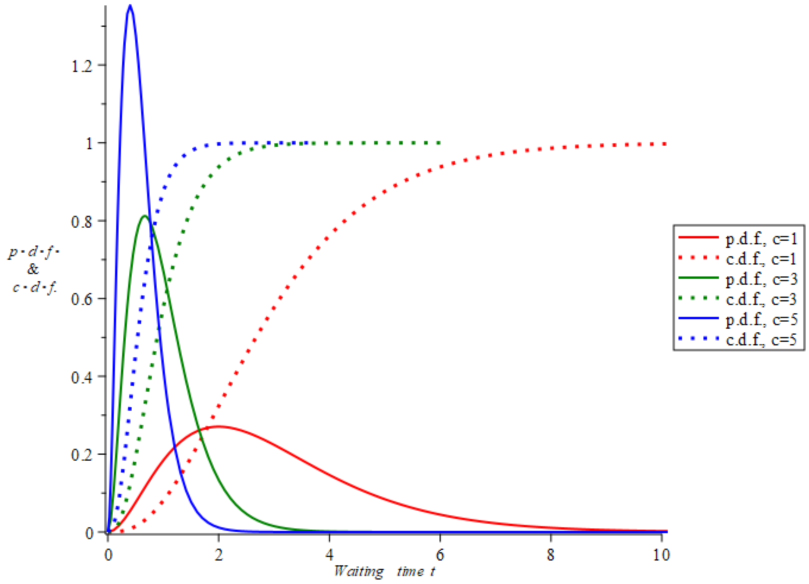

Figure 3. p.d.f. & c.d.f. of Waiting-time for $E_2/M^{5,10}/c$. $\rho = 0.9$, $\mu = 1$; c = 1, 3, 5

Figure 2 and Table 3 demonstrate the capability of our model in providing crucial system performance measures necessary for optimizing the problem and achieving the desired outcome. Noteworthy progress has already been made in this area, and the findings from this research will be presented in our forthcoming paper.

Figure 3 presents the waiting-time densities for $E_2/M^{5,\,10}/c$, with a traffic intensity set to 0.9, and a service rate of $\mu = 1$. In such a high- traffic system, the probabilities of the system being idle are 2.80%, 4.16% and 4.69%, accompanied by the mean queue-length of 47.36, 46.72 and 46.48. These values correspond to the number of servers $c = 1$, 3, 5, respectively.

8. Conclusions

Waiting time distribution for the queue GI/M$^{a,\,b}$/c in the steady state was successfully investigated. We also derived closed-form explicit analytic expressions for the mean waiting-time. They are computationally efficient and stable. We verified Little's Formula for the queueing systems when the inter-arrival time distributions of Erlang, deterministic, and uniform.

GI/M$^{a,\,b}$/c is a more general queueing system with the flexibility to accommodate a wide range of queueing systems. The model's validity was confirmed through MAPLE, producing numerical results with minimal computational costs. It has been demonstrated that, by selecting specific values for parameters a, b, c, and inter-arrival time distributions, the numerical results generated by our model align with those provided in simpler models, as expected.

Acknowledgments

We would like to express our thanks to Dr. Y.Q. Zhao and Dr. F. Rivest for their valuable suggestions and comments. This work was supported by the Royal Military College of Canada Professional Development Allocation.

References

[1] Bar-Lev, S. K., Boxma, O., Kleiner, I., Perry, D. & Stadje, W. (2017). Recycled Incomplete identification procedures for blood screening. *European Journal of Operational Research*, 259(1), 330–343.

[2] Chaudhry, M. & Gai, J. (2022). Analytic and Computational Analysis of GI/M$^{a,\,b}$/c Queueing System . *Mathematics.*, 10(19), 5–26.

[3] Chaudhry, M. & Madill, B. (1987). Probabilities and some measures of efficiency in the queueing system GI/M$^{a,\,b}$/1. *Selecta Statistica Canadiana* , 7, 53–75.

[4] Chaudhry, M. & Templeton, J. (1983). *A First Course in Bulk Queues*; John Wiley & Sons, Inc.: New York, NY, USA,.

[5] Easton, G. & Chaudhry, M. (1982). The queueing system $E_k/M^{a,b}/1$ and its numerical analysis. *Computers & Operations Research*, 9, 197–205.

[6] Goswami, V., Samanta, S., Vijaya Laximi , P. & Gupta, U. (2008). Analyzing a multiserver bulk-service finite-buffer queue. *Applied mathematical modelling*, 32, 1797–1812.

[7] Gross, D., Shortle, J., Thompson, J. & Harris, C. (2008). *Fundamentals of Queueing Theory*; John Wiley and Sons: Hoboken, NJ, USA.

[8] Medhi, J. (1979). Further results on waiting time distribution in Poisson queue under a general bulk wervice rule. *Cahiers du C.E.R.O.*, 21, 183–189.

[9] Neuts, M. (1967). A general class of bulk queues with Poisson input. *The Annals of Mathematical Statistics*, 38, 759–770.

[10] Neuts, M. (1981). *Matrix-Geometric Solution in Stochastic Models—An Algorithmic Approach*; John Hopkins University Press: Baltimore, MD, USA.

[11] Samanta, S. K., & Bank, B. (2022). Extended Analysis and Computationally Efficient Results for the $GI/M^{a,b}/1$. *Communications in Statistics - Theory and Methods* , 51(11), 3739–3760.

[12] Sasikala, S. & Indhira, K. (2016). Bulk service queueing models—A survey. *International Journal of Pure and Applied Mathematics*, 106, 43–56.

[13] Shyu, K. (1960). On the queueing processes in the system GI/M/n with bulk service. *Acta Mathematica Sinica*, 10, 182–189.

[14] Shyu, K. (1964). The waiting time distribution for the queueing processes in the system GI/M/n with bulk service. *Acta Mathematica Sinica*, 14, 796–808.

[15] Sim, S. (1982). On Multi-Vehicle Transportation Systems with Queue-Dependent Dispatching Policies. Ph.D. Thesis, University of Toronto, Toronto, ON, Canada.

[16] Sim, S. & Templeton, J. (1985). Steady state results for the $M/M^{a,b}/c$ batch-service system. *European Journal of Operational Research*, 21, 260–267.

[17] Takacs, L. (1962). *Introduction to the Theory of Queues*; Oxford University Press, New York.

[18] Tamrakar, G. K. & Banerjee, A. (2023). On Steady state analysis of an infinite capacity $M^x/M^{a,y}/1$ queue with optional service and queue length dependent single (multiple) vacation. *Queueing Models and Service Management*, 6(1), 27–61.

[19] Yao, D. & Chaudhry, M. L. (1984). A note on some relations in queue GI/M/c. *Operations Research Letters*, *3*, 53–56.

Appendix

A. Proof of Equation 19

1. $\sum_{k=1}^{c} P^{-}(I(k), j-1) = \sum_{k=1}^{c} (P^{-}(I(k), a-1) + \frac{1}{C_N}\frac{w^{a-1}-w^{j-1}}{1-w^b}, \qquad 1 < j < a-1.$

   Employing the following equation (proved in [2]),

$$\sum_{k=1}^{c} P^{-}(I(k), j) = \sum_{k=1}^{c}\sum_{m=1}^{k} P^{-}(I(m), j-1)\,[k-m|c-m] + P^{-}(B, j-1)\sum_{k=1}^{c}[k|c] + \sum_{i=1}^{\infty} P^{-}(B, ib+j-1)\sum_{k=1}^{c}\{k|c;i\}, \tag{32}$$

then exchanging variables k and m of the summations, using the relations (proved in [2])

$$\sum_{l=1}^{c}\{l|c;q\} + \sum_{i=0}^{q}(i|c) = 1 \qquad for\ q > 0, and$$

$$\sum_{i=m}^{c}[i-m|c-m] = 1, \qquad 0 \le m \le c,$$

the left side of equation 32 can be written as

$$\begin{aligned}\sum_{k=1}^{c} P^{-}(I(k), j) &= \sum_{m=1}^{c} P^{-}(I(m), j-1) + \frac{w^{j-1}}{C_N}\sum_{i=0}^{\infty}(w^b)^i(1-\sum_{l=0}^{i}(l|c)) \\ &= \sum_{m=1}^{c} P^{-}(I(m), j-1) + \frac{w^{j-1}}{C_N}\sum_{i=0}^{\infty}(w^b)^i\left[1-\sum_{l=0}^{i}(l|c)\right] \\ &= \sum_{m=1}^{c} P^{-}(I(m), j-1) + \frac{w^{j-1}}{C_N}\frac{1-w}{1-w^b}\end{aligned} \tag{33}$$

where the following relationship is used

$$\begin{aligned}\sum_{i=0}^{\infty}(w^b)^i\left[1-\sum_{l=1}^{i}(l|c)\right] &= \frac{1}{1-w^b} - \sum_{i=0}^{\infty}(w^b)^i\sum_{l=0}^{i}(l|c) \\ &= \frac{1}{1-w^b} - \sum_{l=0}^{\infty}(l|c)\frac{(w^b)^l}{1-w^b} = \frac{1-w}{1-w^b}.\end{aligned}$$

Since $\sum\limits_{l=0}^{\infty}(l|c)(w^b)^l = w$ by using equation 5. From equation 33, we have

$$\sum_{m=1}^{c} P^-(I(m), j-1) = \sum_{k=1}^{c} P^-(I(k), j) - \frac{w^{j-1}}{C_N}\frac{1-w}{1-w^b}$$
$$= \sum_{k=1}^{c} P^-(I(k), a-1) - \frac{1}{C_N}\frac{1-w}{1-w^b}\sum_{i=j-1}^{a-2} w^i$$
$$= \sum_{k=1}^{c} P^-(I(k), a-1) + \frac{1}{C_N}\frac{w^{a-1}-w^{j-1}}{1-w^b} \tag{34}$$

by recursively replacing the term $\sum\limits_{k=1}^{c} P^-(I(k), j)$ using equation 33.

2. $\sum\limits_{k=1}^{c} (P^-(I(k), a-1) = \frac{1}{a}(1-T_w)$, where

$$T_w = \frac{1}{C_N(1-w^b)}\left[aw^{a-1} + \frac{w^a - w^b}{1-w}\right].$$

Adding j from 1 to a on the both sides of equation 34, we have

$$\sum_{j=1}^{a}\sum_{k=1}^{c} P^-(I(k), j-1) = a\sum_{k=1}^{c} (P^-(I(k), a-1) + \frac{1}{C_N}\frac{aw^{a-1}}{1-w^b} - \frac{1-w^a}{C_N(1-w^b)(1-w)}$$

The left side of the above equation is the sum of the idle-server p.a.e. probabilities. It is the complement part of the sum of the busy-server p.a.e. probabilities, such as $1 - 1/C_N(1-w)$.
Then $\sum\limits_{k=1}^{c} (P^-(I(k), a-1)$ can be solved as

$$\sum_{k=1}^{c} (P^-(I(k), a-1) = \frac{1}{a}\left(1 - \frac{1}{C_N(1-w^b)}\left(aw^{a-1} + \frac{w^a - w^b}{1-w}\right)\right) = \frac{1}{a}\left(1 - T_w\right).$$

B. Proof of Equation 26

$$\sum_{q=0}^{\infty}\sum_{j=0}^{a-2} P^-(B, qb+j)\frac{(c\mu)^{q+1}}{q\,!}\int_0^{\infty} t^{q+1}\exp(-c\mu t)F_{A\,(a-j-1)}(t)dt$$
$$-\sum_{q=0}^{\infty}\sum_{j=0}^{a-2} P^-(B, qb+j)\frac{1}{c\mu}\sum_{i=0}^{q}\int_0^{\infty}\frac{(c\mu t)^{i+1}\exp(-c\mu t)}{i\,!}dF_{A\,(a-j-1)}(t)$$

$$=\frac{(a-1)w^{a-1}}{C_N c\mu(1-w^b)^2}. \tag{35}$$

By using L.-S.T.'s, the first integral in the left side of above equation can be written as

$$\int_0^\infty t^{q+1}\exp(-c\mu t)F_{A\,(a-j-1)}(t)dt=(-D)^{q+1}\left\{\frac{[\bar{a}(c\mu)]^{a-j-1}}{c\mu}\right\}, \tag{36}$$

and the summation of the second integral as

$$\begin{aligned}&\sum_{i=0}^{q}\int_0^\infty \frac{(c\mu t)^{i+1}\exp(-c\mu t)}{i\,!}dF_{A\,(a-j-1)}(t)\\ =&\sum_{i=0}^{q}\frac{(c\mu)^{i+1}}{i\,!}(-D)^{i+1}\left\{\,[\bar{a}(c\mu)]^{a-j-1}\right\},\end{aligned} \tag{37}$$

where D is the derivative operator defined by $D=d/d\mu$, $D^n=d^n/d\mu^n$, and $\bar{a}(c\mu)$is the L.-S.T. of interarrival-time A and it is defined in equation 6. Since the random variable $A_w(a-j-1)$ is the sum of $a-j-1$ i.i.d. random variables A, then the L.-S.T. of $A_w(a-j-1)$ is $[\bar{a}(c\mu)]^{a-j-1}$.

By using Leibnitz's theorem for the multiple differentiation of a product, we can rewrite equation 36 as

$$\begin{aligned}&(-D)^{q+1}\left\{\,[\bar{a}(c\mu)]^{a-j-1}(c\mu)^{-1}\right\}\\ =&(-1)^{q+1}\sum_{r=0}^{q+1}\binom{q+1}{1}(-1)^r r!\,(c\mu)^{-r-1}D^{q+1-r}\left\{\,[\bar{a}(c\mu)]^{a-j-1}\right\}\,.\end{aligned} \tag{38}$$

Plugging 36, 37 and 38 into the left side of equation 35, simplifying, we have

$$\begin{aligned}&\text{The left side of 35}\\ =&\frac{1}{c\mu}\sum_{q=0}^{\infty}\sum_{j=0}^{a-2}P^-(B,qb+j)\sum_{r=0}^{q+1}\frac{q+1}{(q+1-r)!}(-c\mu D)^{q+1-r}\left\{\,[\bar{a}(c\mu)]^{a-j-1}\right\}\\ &-\frac{1}{c\mu}\sum_{q=0}^{\infty}\sum_{j=0}^{a-2}P^-(B,qb+j)\sum_{i=0}^{q}\frac{1}{i!}(-c\mu D)^{i+1}\left\{\,[\bar{a}(c\mu)]^{a-j-1}\right\}\,.\end{aligned} \tag{39}$$

Exchange the variables $q-r+1$ in the first term of above equation by m, and $i+1$ in the second term by m, then combine the two terms to get

$$\begin{aligned}&\text{The left side of 35}\\ =&\frac{1}{c\mu}\sum_{q=0}^{\infty}\sum_{j=0}^{a-2}P^-(B,qb+j)\sum_{m=0}^{q}\left(\frac{q+1-m}{m\,!}\right)(-c\mu D)^m\left\{\,[\bar{a}(c\mu)]^{a-j-1}\right\}\,.\end{aligned} \tag{40}$$

Next, substitute the results of $P^{-}(B, qb+j) = w^{qb+j}/C_N$ into B6, then exchange variables q and m of the summations to get

$$\begin{aligned}
&\text{The left side of 35}\\
=&\frac{1}{C_N c\mu}\sum_{j=0}^{a-2} w^j \sum_{q=m}^{\infty} w^{qb} \sum_{m=0}^{\infty}\left(\frac{q+1-m}{m\,!}\right)(-c\mu D)^m \left\{\ [\bar{a}(c\mu)]^{a-j-1}\right\}\\
=&\frac{1}{C_N c\mu}\sum_{j=0}^{a-2} w^j \sum_{r=0}^{\infty} (r+1) w^{rb} \sum_{m=0}^{\infty}\left(\frac{(-c\mu w^b D)^m}{m\,!}\right)\left\{\ [\bar{a}(c\mu)]^{a-j-1}\right\}\\
=&\frac{1}{C_N c\mu (1-w^b)^2}\sum_{j=0}^{a-2} w^j \sum_{m=0}^{\infty}\left(\frac{(-c\mu w^b D)^m}{m\,!}\right)\left\{\ [\bar{a}(c\mu)]^{a-j-1}\right\},
\end{aligned} \tag{41}$$

where we replace variables $q-m$ by r, and use $\sum\limits_{r=0}^{\infty}(r+1)w^{rb} = \frac{1}{(1-w^b)^2}$.

Now, we can replace the operator $\sum\limits_{m=0}^{\infty}\left(\frac{(-c\mu w^b D)^m}{m!}\right)$ by $\exp(-c\mu w^b D)$, and the property of $\exp(-c\mu w^b D) f(c\mu) = f(c\mu(1-w^b))$, then 41 will be

$$\begin{aligned}
&\text{The left side of 35}\\
=&\frac{1}{C_N c\mu (1-w^b)^2}\sum_{j=0}^{a-2} w^j \left\{\ \left[\bar{a}(c\mu(1-w^b))\right]^{a-j-1}\right\}\\
=&\frac{1}{C_N c\mu (1-w^b)^2}\sum_{j=0}^{a-2} w^j \left\{\ w^{a-j-1}\right\}\\
=&\frac{(a-1)w^{a-1}}{C_N c\mu (1-w^b)^2}
\end{aligned} \tag{42}$$

by using the root equation (5).

C. Proof of Equation 29

$$\begin{aligned}
dF_{V(B_n,n)}(t) =&(1-\sum_{i=0}^{a-j-2}(\eta\lambda t)^i \exp(-\eta\lambda t)/i!)\frac{(c\mu)^{q+1}t^q \exp(-c\mu t)}{q\,!}dt\\
&+(1-\sum_{i=0}^{q}(c\mu t)^i \exp(-c\mu t)/i\,!)\frac{(\eta\lambda)^{a-j-1}t^{a-j-2}\exp(-\eta\lambda t)}{(a-j-2)!}dt.
\end{aligned}$$

If the random variable, the interarrival-time $A_w(a-j-1)$, has a gamma probability distribution with parameter $\eta\lambda$ and $\eta(a-j-1)$, then

$$F_{A_w(a-j-1)}(t) = 1-\sum_{i=0}^{a-j-2}(\eta\lambda t)^i \exp(-\eta\lambda t)/i!, \tag{43}$$

and

$$dF_{A_w(a-j-1)}(t) = \frac{(\eta\lambda)^{a-j-1} t^{a-j-2} \exp(-\eta\lambda t)}{(a-j-2)!} dt. \quad (44)$$

Plugging above results, together with equations 22 and 23 into equation 15, simplifying, we have

$$\begin{aligned} dF_{V_Q}(t) =& (1 - \sum_{i=0}^{a-j-2} (\eta\lambda t)^i \exp(-\eta\lambda t)/i!) \frac{(c\mu)^{q+1} t^q \exp(-c\mu t)}{q!} dt \\ &+ (1 - \sum_{i=0}^{q} (c\mu t)^i \exp(-c\mu t)/i!) \frac{(\eta\lambda)^{a-j-1} t^{a-j-2} \exp(-\eta\lambda t)}{(a-j-2)!} dt. \end{aligned}$$