Hindawi
Advances in Operations Research
Volume 2023, Article ID 6620393, 22 pages
https://doi.org/10.1155/2023/6620393

Review Article

A Review of Birth-Death and Other Markovian Discrete-Time Queues

Muhammad El-Taha

Department of Mathematics and Statistics, University of Southern Maine, 96 Falmouth Street, Portland, ME 04104-9300, USA

Correspondence should be addressed to Muhammad El-Taha; el-taha@maine.edu

Received 9 June 2023; Revised 13 October 2023; Accepted 25 October 2023; Published 5 December 2023

Academic Editor: Arunava Majumder

Copyright © 2023 Muhammad El-Taha. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this review article, we consider discrete-time birth-death processes and their applications to discrete-time queues. To make the analysis simpler to follow, we focus on transform-free methods and consider instances of non-birth-death Markovian discrete-time systems. We present a number of results within one discrete-time framework that parallels the treatment of continuous time models. This approach has two advantages; first, it unifies the treatment of several discrete-time models in one framework, and second, it parallels to the extent possible the treatment of continuous time models. This allows us to draw parallels and contrasts between the discrete and continuous time queues. Specifically, we focus on birth-death applications to the single server discrete-time model with Bernoulli arrivals and geometric service times and provide the reader with a simple rigorous detailed analysis that covers all five scheduling rules considered in the literature, with attention to stationary distributions at slot edges, slot centers, and prearrival epochs. We also cover the waiting time distributions. Moreover, we cover three Markovian models that fit the global balance equations. Our approach provides interesting insights into the behavior of discrete-time queues. The article is intended for those who are familiar with queueing theory basics and would like a simple, yet rigorous introductory treatment to discrete-time queues.

1. Introduction

This review article is intended as an introduction to discrete-time queues by focusing mainly on queueing models that fit the discrete birth-death equations. We cover most scheduling rules in the literature and give the stationary distribution at slot edges, slot centers, and at prearrival instants. We also cover instances of Markovian models that can be solved easily by recursive methods. Specifically, we pick models whose stationary distribution can be solved using global balance equations and cover the multiserver, batch arrival, and finite population Markovian models. Moreover, we focus on transform-free methods to make the analysis simpler to follow and present a number of results within one discrete-time framework. This approach has two advantages; first, it unifies the treatment of several discrete-time models in one framework, and second, it parallels to the extent possible the treatment of continuous time models. Moreover, we address BASTA (Bernoulli Arrivals See Time Averages), sometimes referred to as GASTA (Geometric arrivals See Time Averages), by giving the stationary distribution at prearrival epochs for all scheduling rules.

This article has several key contributions. We give a unified treatment of multiple models within one framework, compare the behavior of these models using multiple scheduling rules, give a direct proof of the distribution function of the waiting times in queue (delay), and assert that the waiting time distribution is the same regardless of the scheduling rule. Moreover, we address the BASTA issue in this simpler framework, note that BASTA in discrete-time queues behaves differently from its continuous time counterpart, and address non-birth-death queueing models with solutions that follow from recursive techniques.

This article's focus is on single station queues. All our models, except for one, deal with single server queues. Meisling [1] appears to be the first to study a queueing

system in discrete time. He used a generating function approach to obtain the system characteristics. Since then, queues in discrete-time have gained popularity due to their wide applicability in computer and communications networks. Hunter [2] gives detailed analysis of single server discrete-time queues using Markovian and generating function methods. Robertazzi [3] covers multiserver models and uses a recursive method to efficiently compute the system stationary distribution. El-Taha et al. [4] use transform-free methods to address and prove the insensitivity of discrete-time queues with processor sharing, loss, and infinite servers. There are a few results in the literature where authors focus on discrete birth-death models. Among these are Daduna [5] and Desert and Daduna [6]. Alfa [7] and Alfa [8] address birth-death processes in both books, but attention is restricted to the late arrival models. Bruneel and Kim [9] consider discrete queueing models that fit with the late arrival model observed at slot edges. Woodward [10] (Chapter 4) studies single server queues using the early arrival scheduling rule and the outside observer's epochs. Halfin [11] discusses when arrivals see time averages for discrete-time queues and shows that the arrivals have to follow a Bernoulli process for GASTA to hold. See also Gravey and Hebuterne [12] where the stationary distribution function at prearrival epochs is given. In discrete-time systems, prearrival probabilities do not always coincide with the time-average probabilities observed at slot edges even with Bernoulli arrivals. See Daduna [5] and Desert and Daduna [6]. There have been recent articles that consider other aspects of discrete-time birth-death processes. Daduna [13] gives a detailed analysis of alternating birth-death processes. Ozawa [14], and Ozawa and Kobayashi [15] , consider discrete-time two-dimensional quasi birth-death processes. Fernández and de la Iglesia [16] study quasi-birth and death multivariate processes. Sasaki [17] gives examples of exactly solvable birth-death processes. See also Lenin and Parthasarathy [18], Daduna and Schassberger [19], Chaudhry [20], Chaudhry et al. [21], Dattatreya and Singh [22], Louvion et al. [23], Schassberger [24], Henderson and Taylor [25], and Neuts [26]. However, our focus is on one-dimensional birth-death processes that are used to describe fifteen instances of discrete-time queues by using five scheduling rules and three reference epochs. This review article highlights include the following:

(1) The article uses a unified approach that combines direct sample path and stochastic techniques and avoids generating-functions methods to provide an accessible summary of all birth-death discrete-time queueing models in one framework.

(2) Provides new insights into these models through a combination of generalizations, new results, new proofs, and comparisons of these models. However, the majority of the results are not new.

(3) Presents in one unified space results for the five scheduling rules in the literature with each model studied using slot edges, slot centers, and prearrival epochs, thus allowing readers to compare these models at fifteen instances of these combinations.

(4) Addresses BASTA and provides formulas for the prearrival probabilities for all five scheduling rules.

(5) Addresses the waiting time distribution functions for all five scheduling rules using a unified approach.

(6) Give three Markovian models that do not fit the birth-death equations contrary to their continuous time counterparts.

The rest of the article is organized as follows. In Section 2, we introduce the generalized birth-death process and give a general solution for the model. In Section 3, we study two special cases where in the first model, a customer that arrives at an idle server can leave within the same slot and in the other model, an arrival to find a server idle cannot leave in the same time slot. It turns out that these two specializations of the birth-death model cover all but one of the situations encountered in all five scheduling rules. In Section 4, we apply these two birth-death models to find the stationary distribution functions for all five scheduling rules at slot centers and at slot edges. In Section 5, we address BASTA issues; specifically, we give formulas for prearrival probabilities for all five scheduling rules. In Section 6, we address the waiting times and show that all rules lead to the same waiting time distribution function. In Section 7, we give instances of Markovian models that can be solved by recursive methods. Specifically, we cover the multiserver, batch arrival, and finite population models. Note that the multiserver and finite population models in continuous time can be represented by birth-death equations. This is not the case for the corresponding discrete-time models. Finally, in Section 8, we give concluding remarks.

2. Generalized Birth-Death Equations

In this section, we start with a sample-path version of the generalized birth-death equations, then introduce the stochastic version, and show how the stochastic birth-death equations fit into our sample path framework. We point out that our focus is exclusively on one-dimensional birth-death models. We use the term "generalized" because we do not make any stochastic assumptions in this section. The results follow by assuming that the relevant limits exist.

To formalize this approach, let $\{Z(\tau), \tau = 1, 2, \cdots\}$ be a discrete-time process with state space $S = I$, where I is the set of non-negative integers. Since we shall be using a sample-path framework, it is helpful to think of $Z = \{Z(\tau), \tau = 1, 2, \cdots\}$ as a deterministic one realization (sample path) of a stochastic process. The process makes a transition from one state to another, possibly itself, at every time instant $\tau = 1, 2, \cdots$. In other words, we allow $Z(\tau)$ to make a transition from a state to itself.

For any state $i \in I$; $j \in I$, let

$$C(i,j;\tau) := \sum_{k=1}^{\tau} \mathbf{1}\{Z(k)=i, Z(k+1)=j\},$$
$$Y(i;\tau) := \sum_{k=1}^{\tau} \mathbf{1}\{Z(k)=i\}. \quad (1)$$

During time $(0,\tau]$, $C(i,j;\tau)$ counts the transitions from state i to state j and $Y(i,\tau)$ is the total time in state i. Now, we define the following limits when they exist:

$$p(i,j) := \lim_{\tau\longrightarrow\infty} \frac{C(i,j;\tau)}{Y(i;\tau)},$$
$$\Lambda(i,j) := \lim_{\tau\longrightarrow\infty} \frac{C(i,j;\tau)}{\tau}, \quad (2)$$
$$\pi(i) := \lim_{\tau\longrightarrow\infty} \frac{Y(i;\tau)}{\tau}.$$

Here, $p(i,j)$ is the conditional long-run fraction of transitions from i to j, $\Lambda(i,j)$ is the unconditional long-run transition rate from i to j, and $\pi(i)$ is the long-run fraction of time in state i. In a Markov chain setting, these quantities represent the one-step transition probabilities, the unconditional transition probabilities, and the stationary probabilities, respectively.

A discrete-time generalized birth-death process is a process where from any state the process can make a transition only to a neighboring state or the state itself. We use the term "generalized" because we do not require the process to be Markovian. We give a formal definition as follows.

Definition 1. Let I be the set of non-negative integers. The process $\{Z(\tau), \tau = 1,2,\cdots\}$ is said to be a discrete birth-death process if for each $i \in I$, $b(0) = 0$ and

$$p(i,j) = \begin{cases} a(i), & \text{if } j = i+1, \\ b(i), & \text{if } j = i-1, \\ c(i), & \text{if } j = i, \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

Note that for all $i \in I$, $a(i) + b(i) + c(i) = 1$, and $a(-|i) = b(-|i) = c(-|i) = 0$ for all $i > 0$. Now, we give the generalized birth-death equations.

Lemma 2. *Let $a(k) = p(k,k+1), k = 0,1,\cdots; b(k) = p(k, k-1), k = 1,\cdots$. Then, the generalized birth-death equations are given by*

$$a(k)\pi(k) = b(k+1)\pi(k+1);\ k = 0,1,\cdots. \quad (4)$$

Proof. It follows from the definitions that

$$\Lambda(k,k+1) = \lim_{\tau\longrightarrow\infty} \frac{C(k,k+1;\tau)}{Y(k;\tau)} \frac{Y(k;\tau)}{\tau} = a(k)\pi(k). \quad (5)$$

Similarly,

$$\Lambda(k+1,k) = \lim_{\tau\longrightarrow\infty} \frac{C(k+1,k;\tau)}{Y(k+1,\tau)} \frac{Y(k+1;\tau)}{\tau+1} = a(k+1)\pi(k+1). \quad (6)$$

Now, note that for all k and τ

$$|C(k,k+1;\tau-)C(k+1,k;\tau)| \le 1. \quad (7)$$

In (7), divide by τ and take limits as $\tau \longrightarrow \infty$ to conclude that

$$\Lambda(k,k+1) = \Lambda(k+1,k). \quad (8)$$

This completes the proof of the result.

The equations in (4) are valid without the assumption that the process $\{Z(\tau); \tau = 1,2,\cdots\}$ is a Markov chain. We only need to assume that the relevant limits exist.

In a Markovian stochastic setting, one typically starts with the global balance equations of birth-death process represented by $Z(\tau); \tau = 1,2,\cdots\}$. Then,

$$\pi(k) = a(k-1)\pi(k-1) + b(k+1)\pi(k+1) + c(k)\pi(k);\ k = 0,1,\cdots. \quad (9)$$

Equations (9) are simply the expanded version of the stationary equations encountered in Markov chains (see, for example Alfa [7]). The equations given by (9) can be represented by the flow balance principle, which states that for each state $k \ge 0$: the probability flow out of state k = the probability flow into state k. Equations (9) may be written as

$$a(0)\pi(0) = b(1)\pi(1), \quad (10)$$

$$\pi(1) = a(0)\pi(0) + b(2)\pi(k+1) + c(1)\pi(1), \quad (11)$$

$$\pi(2) = a(1)\pi(1) + b(3)\pi(3) + c(2)\pi(2). \quad (12)$$

Now, add equations (10) and (11) to obtain

$$a(1)\pi(1) = b(2)\pi(2). \quad (13)$$

Add (12) and (13) to obtain

$$a(2)\pi(2) = b(3)\pi(3), \quad (14)$$

and so on. In general, using induction, we obtain

$$a(k-1)\pi(k-1) = b(k)\pi(k);\ k = 1,\cdots. \quad (15)$$

We obtained these same equations in Lemma 2 without the Markovian assumption and using only the assumption that relevant limits exist. The equations in (15) are referred to as the detailed (or local) balance equations. They represent the probability flow between states.

Solution to the generalized birth-death equations.

Solving (4) (equivalently (15)) recursively, one obtains

$$\pi(k) = \Pi_{j=1}^{k} \frac{a(j-1)}{b(j)} \pi(0), \quad k \ge 1. \quad (16)$$

Now, assuming that $\{\pi(k); k \in I\}$ exist, $\sum_{k=1}^{\infty} \pi(k) = 1$, and normalizing, we obtain

$$\pi(0) = \left[1 + \sum_{m=1}^{\infty} \Pi_{j=1}^{m} \frac{a(j-1)}{b(j)}\right]^{-1} = \left[\sum_{m=0}^{\infty} \Pi_{j=1}^{m} \frac{a(j-1)}{b(j)}\right]^{-1}, \tag{17}$$

where a $\prod$ over an empty set is 1. The second form is more compact but requires us to assume $a(-1) = 1$. Note that $\pi(0) > 0$ if and only if $\sum_{m=1}^{\infty} \Pi_{j=1}^{m} a(j-1)/b(j) < \infty$. Therefore,

$$\pi(k) = \frac{\Pi_{j=1}^{k} a(j-1)/b(j)}{\left[\sum_{m=0}^{\infty} \Pi_{j=1}^{m} a(j-1)/b(j)\right]}, k \geq 1. \tag{18}$$

This solution is valid without any stochastic assumptions. We only assumed the existence of limits. We shall investigate Markovian queueing models using detailed balance equations. Specifically, we focus on models that can be represented as generalized birth-death equations. In Section 7, we discuss a number of models using global balance equations that can be solved by iterative methods. Since most applications involve the geometric distribution function, we give a quick review of this distribution. A random variable X is said to have a geometric pmf with parameter p, $0 < p < 1$, if

$$P(X = n) = q^{n-1} p \ (n = 1, 2, \cdots; p > 0, q = 1 - p). \tag{19}$$

The geometric distribution has properties similar to the exponential distribution. The first and second moments are given by $E[X] = 1/p$ and $E[X^2] = 2/p^2 - 1/p$, and the variance is also given by $(X) = q/p^2$. The complement of the cumulative distribution functions (CDF) is $P(X \geq k) = q^{k-1}$ Moreover, it has the memoryless property; i.e., $P(X = n + k \mid X > n) = P(X = k)$; $k = 1, 2, \cdots$. In particular, if $k = 1$, we obtain $P(X = n + 1 \mid X > n) = p$.

2.1. A State-Dependent Generalized Birth-Death Process. Consider a state-dependent birth-death process that represents a Bernoulli queue where an arrival will occur with probability $\alpha(j) \in (0, 1)$ if the system state is $j \geq 0$, i.e., there are j customers in the system. Similarly, a service completion will occur with probability $\beta(j) \in (0, 1)$ when the system state is $j \geq 1$. When $j = 0$, we assume $\beta(0) = 0$ or $\beta(0) = \beta$, where β is a constant such that $0 < \beta < 1$. Now, we use the birth-death model in the previous section with

$$\begin{aligned} a(j) &= \alpha(j)(1 - \beta(j)); \ j \geq 0; \\ b(j) &= \beta(j)(1 - \alpha(j)); \ j \geq 1. \end{aligned} \tag{20}$$

Substitute in (18) to get

$$\pi(n) = \Pi_{j=1}^{n} \frac{\alpha(j-1)(1 - \beta(j-1))}{\beta(j)(1 - \alpha(j))} \pi(0), n \geq 1; \tag{21}$$

where

$$\pi(0) = \left[\sum_{k=0}^{\infty} \Pi_{j=1}^{k} \frac{\alpha(j-1)(1 - \beta(j-1))}{\beta(j)(1 - \alpha(j))}\right]^{-1}. \tag{22}$$

The birth-death equations are not particularly useful at this level of generality. In the next section, we consider two special instances that are not only useful but cover a good number of situations related to Markovian single server queues in discrete time.

We point out that the results in this section are not new. Daduna [5] and Desert and Daduna [6] give a formula for a birth-death process with state-dependent arrival and service completion probabilities, for the case where $\beta(0) > 0$. Alfa [7] considers a birth-death process with $\beta(0) = 0$ and uses matrix geometric and generating functions methods in his analysis. Our birth-death process does not have these restrictions; thus, we are able to study a wider class of systems with a larger selection of observation epochs by switching between $\beta(0) = 0$ and $\beta(0) = \beta$ where $0 < \beta < 1$ is a constant. This leads to a unified and simplified treatment of all scheduling rules at multiple reference points, as we shall see in Section 3.

3. Two Special Birth-Death Processes

In this section, we are motivated by single server queues with Bernoulli arrivals and geometric service times. That is, Markovian single server queues that can be modeled by a birth-death process. At this point, we do not consider specific scheduling rules or specific observation epochs. In general, there are five scheduling rules and three observation epochs that are of interest, giving us a large number of situations that will be covered in Section 4. For the vast majority of situations, as we shall see later, the stationary distribution will be given by one of the two cases we cover in the two subsections as follows.

3.1. The Birth-Death Process with $\beta(0) = 0$. Consider a Markovian single server discrete-time queue with infinite waiting room. Arrivals follow a Bernoulli process such that the probability of arrival at any given time instant is α. Equivalently, the time between arrivals follows a geometric distribution with mean $1/\alpha$. Service times are i.i.d. such that

$$P(S = k) = (1 - \beta)^{k-1} \beta; \quad 0 < \beta < 1, k = 1, 2, \cdots, \tag{23}$$

that is, the service times follow a geometric distribution function with mean $1/\beta$. This means that $\alpha(k) = \alpha$; $k = 0, 1, \cdots$, and $\beta(k) = \beta$; $k = 1, 2, \cdots$. In addition, we assume $\beta(0) = 0$. See Figure 1 for a flow balance diagram. Substitute in (21) to obtain

$$\pi(n) = \frac{1}{1 - \beta} \left[\frac{\alpha(1 - \beta)}{\beta(1 - \alpha)}\right]^{n} \pi(0), \tag{24}$$

where $\sum_{n=0}^{\infty} \pi(n) = 1$ implies

$$\begin{aligned} \pi(0) &= \left[1 + \sum_{i=1}^{\infty} \frac{1}{1 - \beta} \left[\frac{\alpha(1 - \beta)}{\beta(1 - \alpha)}\right]^{i}\right]^{-1}, \\ &= \left[1 + \frac{\alpha}{\beta(1 - \alpha)} \sum_{j=0}^{\infty} \left[\frac{\alpha(1 - \beta)}{\beta(1 - \alpha)}\right]^{j}\right]^{-1}. \end{aligned} \tag{25}$$

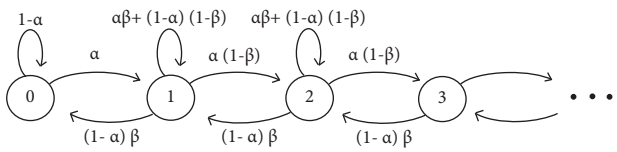

Figure 1: State diagram for a birth-death model with $\beta(0) = 0$.

The system is stable, i.e., $\pi(0) > 0$ if and only if $\alpha(1-\beta)/\beta(1-\alpha) < 1$, which implies

$$\pi(0 = \left[1 + \frac{\alpha}{\beta(1-\alpha)} \frac{1}{1-\alpha(1-\beta)/\beta(1-\alpha)}\right]^{-1}. \tag{26}$$

The stability condition $\alpha(1-\beta)/\beta(1-\alpha) < 1$ is implied by $\alpha/\beta < 1$. This is so because $\alpha < \beta$ implies that $1-\alpha > 1-\beta$ which implies that $(1-\beta)/(1-\alpha)$, thus $\alpha(1-\beta)/\beta(1-\alpha) < 1$. Simplify to obtain

$$\begin{aligned}\pi(0) &= \left[1 + \frac{\alpha}{\beta(1-\alpha)} \frac{\beta(1-\alpha)}{\beta(1-\alpha) - \alpha(1-\beta)}\right]^{-1}, \\ &= \left[1 + \frac{\alpha}{\beta-\alpha}\right]^{-1}, \\ &= 1 - \frac{\alpha}{\beta}.\end{aligned} \tag{27}$$

Therefore,

$$\pi(n) = \left[\frac{\beta-\alpha}{\beta(1-\beta)}\right]\left[\frac{\alpha(1-\beta)}{\beta(1-\alpha)}\right]^n, n = 0, 1, \cdots. \tag{28}$$

Let $\rho = \alpha/\beta$ and $\gamma = \alpha(1-\beta)/\beta(1-\alpha)$. Then, we have the following result.

Theorem 3. *Consider the state-independent birth-death model with $\beta(0) = 0$. Then, the stationary distribution function is given by*

$$\pi(n) = \begin{cases} \rho(1-\gamma)\gamma^{n-1}, & n = 1, 2, \cdots, \\ 1-\rho, & n = 0. \end{cases} \tag{29}$$

This is the same formula given by El-Taha et al. [4] for the $M/G/1$ round Robin model. The round Robin model is known to have the insensitivity property; that is, its stationary distribution does not depend on the shape of the service time distribution function but only on its mean. See also Hunter [2]. Now, we give performance measures and show that their proofs are similar to those of the $M/M/1$ case.

Theorem 4. *The mean number of customers in the system, L, and the mean number of customers in the queue, L_q, are given by*

$$\begin{aligned} L &= \frac{\alpha(1-\alpha)}{\beta-\alpha}, \\ Lq &= \frac{\alpha^2}{\beta} \frac{1-\beta}{\beta-\alpha}. \end{aligned} \tag{30}$$

Proof. By definition $L = \sum_{n=0}^{\infty} \pi(n)$, therefore,

$$\begin{aligned} L &= \frac{\beta-\alpha}{\beta(1-\beta)} \sum_{n=0}^{\infty} n\left[\frac{\alpha(1-\beta)}{\beta(1-\alpha)}\right]^n, \\ &= \frac{\beta-\alpha}{\beta(1-\beta)} \frac{\alpha(1-\beta)}{\beta(1-\alpha)} \sum_{n=1}^{\infty} n\left(\frac{\alpha(1-\beta)}{\beta(1-\alpha)}\right)^{n-1}, \\ &= \frac{\alpha}{\beta^2} \frac{\beta-\alpha}{1-\alpha} \frac{1}{[1-\alpha(1-\beta)/\beta(1-\alpha)]^2}, \\ &= \frac{\alpha}{\beta^2} \frac{\beta-\alpha}{1-\alpha} \frac{1}{[(\beta(1-\alpha)-\alpha(1-\beta))/\beta(1-\alpha)]^2}, \\ &= \frac{\alpha}{\beta^2} \frac{\beta-\alpha}{1-\alpha} \frac{\beta^2(1-\alpha)^2}{(\beta-\alpha)^2}, \\ &= \alpha \frac{1-\alpha}{\beta-\alpha}, \end{aligned} \tag{31}$$

where we have used the relation $\sum_{n=1}^{\infty} n\rho^{n-1} = 1/(1-\rho)^2$. See Gross et al. [27] for details. Now,

$$\begin{aligned} Lq &= L - (1 - \pi(0)), \\ &= \frac{\alpha(1-\alpha)}{\beta-\alpha} - \frac{\alpha}{\beta}, \\ &= \frac{\alpha\beta(1-\alpha) - \alpha(\beta-\alpha)}{\beta(\beta-\alpha)}, \\ &= \frac{\alpha\beta - \alpha^2\beta - \alpha\beta + \alpha^2}{\beta(\beta-\alpha)}, \\ &= \frac{\alpha^2 - \alpha^2\beta}{\beta(\beta-\alpha)}, \\ &= \frac{\alpha^2}{\beta} \frac{1-\beta}{\beta-\alpha}. \end{aligned} \tag{32}$$

□

3.1.1. Mean Delay of the B/G/1 Model. We can use Little's law to evaluate the waiting time in the system and the queue W and W_q, respectively. Instead, we will use an intuitive approach similar to the one used to compute W_q in the continuous $M/G/1$ case.

Here, we still assume that we have a Bernoulli arrival process with parameter α, but general discrete service times with mean $E[S] = 1/\beta$ and second moment $E[S^2]$. We assume that the system is stable in the sense that $\rho = \alpha E[S] < 1$. Moreover, we assume a FIFO discipline. We use an intuitive argument to give the following closed form expression for the mean delay in the system.

Theorem 5. *The mean waiting time in the queue (excluding service times) is given by*

$$Wq = \frac{\alpha\left(E\left[S^2\right] - E[S]\right)}{2(1-\rho)}. \tag{33}$$

Proof. We show that (33) holds using an intuitive argument similar to the $M/G/1$ case. First, note that the log-run fraction of time the server is busy (i.e., probability that the server is busy) is equal to $\rho = \alpha ES$. Moreover, the remaining limiting time-average service time for the customer in service at arrival instants is given by $E[S_r] = (E[S^2] - E[S])/2ES$. Let V be the (virtual) waiting time for a randomly arriving customer. On average, this customer finds L_q customers ahead of him/her in addition to the one in service. Therefore, $V = L_q E[S] + R$, where $R = \rho \times E[S_r] = \alpha(E[S^2] - E[S])/2$ is the residual time of the customer in service. Now, BASTA (Bernoulli arrivals see time averages), El-Taha and Stidham ([28], Chapter 3), and FIFO imply that $V = W_q$, and Little's law implies that $L_q = \alpha W_q$. Therefore,

$$W_q = \alpha W_q E[S] + \alpha \frac{E\left[S^2\right] - E[S]}{2}. \tag{34}$$

Simplify to obtain (33).

Other performance measures can now be obtained immediately. For instance, $W = W_q + ES$, and using Little's law, $L = \alpha W$, and $L_q = \alpha W_q$. A more rigorous proof of (33) can be obtained using a discrete version of $H = \lambda G$ (see El-Taha and Stidham [28] and El-Taha [29]). Now, let the service times be geometric with parameter β so that $E[S] = 1/\beta, E[S^2] = 2/\beta^2 - 1/\beta$, and $E[S^2] - E[S] = 2/\beta^2 - 2/\beta$. Using (33) and noting that $\rho = \alpha/\beta$; we obtain

$$\begin{aligned} Wq &= \frac{\alpha\left(2/\beta^2 - 2/\beta\right)}{2(1-\alpha/\beta)} = \frac{\alpha}{\beta} \times \frac{(1/\beta - 1)}{1-\alpha/\beta} = \frac{\alpha}{\beta}\,\frac{1-\beta}{\beta-\alpha}, \\ W &= Wq + \frac{1}{\beta} = \frac{\beta - \alpha\beta}{\beta(\beta-\alpha)} = \frac{1-\alpha}{\beta-\alpha}. \end{aligned} \tag{35}$$

We treated the $B/Geo/1$ model as a special case of the $B/G/1$ model. The results here are consistent with the birth-death model as can be verified using Little's law. □

3.2. The Birth-Death Process with $\beta(0) = \beta$. Here, we consider the same model as in Subsection 3.1 except that in this case $\beta(0) = \beta$. Referring to the generalized birth-death model, we have

$$\begin{aligned} a(j) &= \alpha(1-\beta); \quad j \geq 0, \\ b(j) &= \beta(1-\alpha); \quad j \geq 1. \end{aligned} \tag{36}$$

A state diagram for this model is given in Figure 2.

Substitute in (21) and (22) to obtain

$$\pi(n)) = \left[\frac{\alpha(1-\beta))}{\beta(1-\alpha)}\right]^n \pi(0), \quad n \geq 1, \tag{37}$$

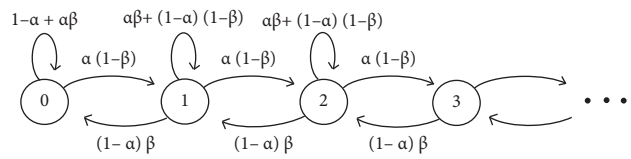

Figure 2: State diagram for a birth-death model with $\beta(0) = \beta$.

where

$$\pi(0) = 1 - \frac{\alpha(1-\beta))}{\beta(1-\alpha)} = \frac{1}{1-\alpha}\left(1 - \frac{\alpha}{\beta}\right). \tag{38}$$

We see from (38) that $\pi(0) > 0$ if $\rho = \alpha/\beta < 1$. Then,

Theorem 6. *Consider the state-independent birth-death model with $\beta(0) = \beta$. Then, the stationary distribution function is given by*

$$\pi(n) = \gamma^n(1-\gamma), \quad n \geq 0. \tag{39}$$

Note that the mean of the distribution given by (39) is given by

$$L = \frac{\gamma}{1-\gamma} = \frac{\alpha(1-\beta)}{\beta-\alpha}, \tag{40}$$

which is not the same as L given by the distribution function corresponding to the case $\beta(0) = 0$. Using Little's law will result in the wrong expression for the mean waiting time W as pointed out by Desert and Daduna [6]. We believe the reason for this is the fact that when $\beta(0) = \beta > 0$, there will be customers that will enter and leave the system in state 0 so that these customers are not counted when the system is observed at the corresponding observation epochs.

4. Applications of Birth-Death Processes to Discrete-Time Queues

In this section, we consider various scheduling rules at various observation epochs. We start by identifying five scheduling rules and various observation epochs of interest.

4.1. Scheduling Rules for Discrete-Time Queues. In this subsection, we discuss five scheduling rules. These rules are the early arrival system (EAS), the late arrival system with immediate access (LAS-IA), the late arrival system with delayed access (LAS-DA), the late arrivals with arrivals first system (LA-AF), and the late arrivals with departures first rule (LA-DF).

In discrete-time queues, time is divided into slots of equal length of one unit. Slot edges are numbered by τ, where $\tau = 1, \cdots$. It is assumed that arrivals and departures occur only on slot boundaries. Contrary to continuous time queues, here, we need to keep track of the order of arrivals and departures in each slot. Depending on the behavior of the actual system, the order of potential arrivals and departures at any given slot varies significantly. In the literature, one finds the early arrival system (EAS) where an

arrival occurs at the beginning (before a potential departure), and the late arrival system (LAS) where an arrival occurs at the end (after a potential departure) of a time slot. Late arrival systems are further refined into two subsystems. For the late arrival system with immediate access (LAS-IA), an arrival can start service immediately and possibly leave at the start of the next time slot if the arrival finds an idle server. For the late arrival system with delayed access (LAS-DA), the arrival waits until the next time slot to start service. For details about different scheduling regimes, one may consult Hunter [2] and Chaudhry [20]. Others schedule potential arrivals and departures at the end of a time slot. In one such rule, at the end of any time slot, potential departures occur first, then potential arrivals, and then the system state is observed. This is the convention used by El-Taha et al. [4] and Daduna [5] and Desert and Daduna [6].

The convention where an arriving customer can enter and leave an empty queue in the same time slot (immediate access) is considered in Chapter 6 of Robertazzi [3] for discrete models that use "virtual cut-through" routing where a packet starts transmission before it is completely received at its current node. In this article, we follow the notation setup as in Hunter [2], Chaudhry et al. [21], El-Taha et al. [4], and Desert and Daduna [6]. We assume work conserving queueing discipline, i.e., the server is not idle when there is work in the system (note that in LAS-DA if an arrival finds an idle server, the delayed access is not counted because the server becomes available to serve only at slot boundaries). Next, we describe each of the scheduling rules.

4.1.1. EAS Scheduling Rule. In the early arrival system (EAS), potential arrivals are scheduled to occur before potential departures. More specifically, a potential arrival in slot $(\tau, \tau+1]$ occurs in $(\tau, \tau+)$, and a potential departure in slot $(\tau-1, \tau]$ occurs in $(\tau-, \tau)$. Moreover, if an arrival finds an idle server, it goes into service immediately and can potentially depart in the same time slot.

The EAS is also referred to as the departure first (DA) system by other authors. In this situation, if one focuses on the time instance, say, τ, then we have $\tau- < D < \tau < A < \tau+$, where D and A refer to potential departures and arrivals, respectively. See Gravey and Hebuterne [12] for a reference on this.

4.1.2. LAS-IA and LAS-DA Scheduling Rules. In this late arrival system (LAS), the order of potential arrivals and departures is reversed so that a potential departure occurs early in a time slot and potential arrivals occur at the end of the slot so that $\tau- < A < \tau < D < \tau+$. More specifically, a potential departure in slot $(\tau, \tau+1]$ occurs in $(\tau, \tau+)$, and a potential arrival in slot $(\tau-1, \tau]$ occurs in $(\tau-, \tau)$. Moreover, if an arrival finds an idle server and goes into service immediately and can potentially depart at the start of the following time slot, the system is called immediate access (IA), if the arrival waits until the start of the next slot to start service, then the system is called delayed access (DA).

4.1.3. LA-DF (Late Arrivals with Arrivals First). In this late arrivals with departures first system, both potential arrivals and departures occur late in the slot so that $\tau-- < D < \tau- < A < \tau$. An arrival that finds an idle server starts service at τ.

4.1.4. LA-AF (Late Arrivals with Departures First). In this late arrivals with departures first system, both potential arrivals and departures occur late in the slot so that $\tau-- < A < \tau- < D < \tau$. An arrival that finds an idle server starts service at τ. See Figure 3 for a depiction of the above-stated scheduling rules.

In the following subsections, for each of the scheduling rules, we deal with single server queues with Bernoulli arrivals and geometric service times. More specifically, let the random variables A and S represent the interarrival and service times, respectively. Assume that interarrival times and service times are i.i.d. and independent of each other. Let the mean interarrival times $E(A) = 1/\alpha$, and mean service times $ES = 1/\beta$, where $0 < \alpha, \beta < 1$, and let the traffic intensity $\rho = \alpha/\beta < 1$. We refer to the observation epochs at slot edges as the random observer epochs. This is consistent with the notion of the random observer in continuous time queues. We also follow the literature by referring to the slot centers as the outside observer epochs.

Whether the system state is observed at slot edges or slot centers, all these five models, except for LAS-DA, fit the birth-death equations covered in Sections 2 and 3. In these cases, we have

$$\begin{aligned} a(j) &= \alpha(1-\beta), \quad j = 1, \cdots, \\ b(i) &= \beta(1-\alpha), \quad j = 1, \cdots, \\ a(0) &= \alpha(1-\beta(0)). \end{aligned} \tag{41}$$

In each case, we need only to determine if $\beta(0) = 0$ or $\beta(0) = \beta$. When the system state is observed at random observer epochs, we check if an arrival can depart in the same slot of its arrival. This can happen if in an EAS rule, an arrival with one unit of service arrives to find an idle server. In this case, $\beta(0) = \beta$; otherwise, $\beta(0) = 0$. It would be instructive if the reader creates a probability flow diagram like Figure 1. For the outside observer, we have a similar situation except we think of a modified slot $(\tau - 1/2, \tau + 1/2]$. We note that LAS-DA model follows the birth-death process when observed at slot centers. At slot edges, the model follows the birth-death process when $j \geq 2$. So, we need to pay special attention to this case as we shall see in Subsection 4.2.

4.2. The Random Observer Stationary Distribution. Here, our interest is in the process $\{Z(\tau), \tau = 1, 2, \cdots\}$; that is, the state of the system is observed at slot edges τ, $\tau = 1, 2, \cdots$. Specifically, we are interested in the stationary distribution function of the process given by $\{\pi(.)\}$ for various scheduling rules.

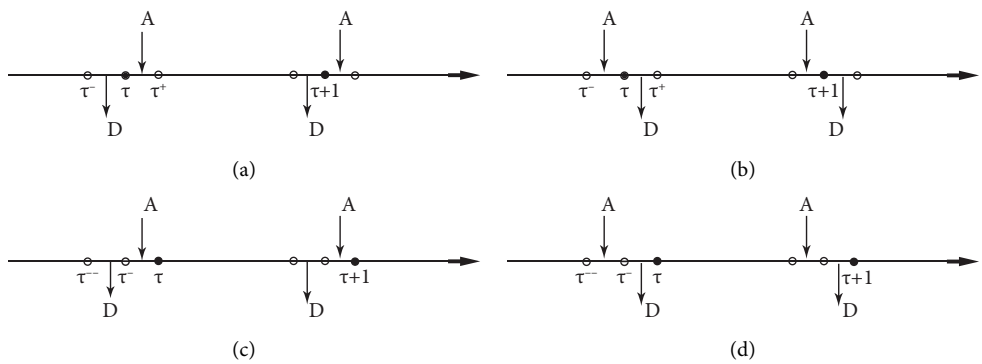

Figure 3: Scheduling rules with A and D represent potential arrivals and departures. (a) EAS rule. (b) LAS rule. (c) LA-DF rule. (d) LA-AF rule.

4.2.1. Queues with EAS Rule. In this model, a potential arrival occurs before potential departures in a time slot, and when a customer arrives to find a server idle, the customer will enter service immediately and therefore may leave in the same time slot.

Here, we are interested in the stationary distribution function at the slot edges, and then we have a birth-death model with $\beta(0) = \beta$. Therefore, (16) gives the stationary distribution function.

4.2.2. Queues with LAS-IA Rule. Here, we are interested in the distribution function at the slot edges, and then we have a birth-death model with $\beta(0) = 0$. Therefore, (12) gives stationary distribution function.

4.2.3. Queues with LA-AF and LA-DF Rules. Here, we are interested in the distribution function at slot edges, and then we have a birth-death model with $\beta(0) = 0$. Therefore, (12) gives the stationary distribution function.

4.2.4. Queues with LAS-DA Rule. This model observed at the slot edges does not fit a birth-death process because of the transition to and from state one. However, for states $n \geq 2$, the process of the number of customers in the system behaves like a birth-death process. See Figure 4 for details. Therefore, $\pi(n) = \gamma^{n-2}\pi(2), n \geq 2$.

A customer that arrives to find an idle server cannot leave in the same time slot and starts service at the beginning of the next time slot. We call this state $1a$. Note that a service completion cannot occur in state $1a$. From state $1a$, the process can transition to state $1b$ if no arrivals occur in the next time slot. State $1b$ behaves like state 1 in the LAS-IA model. The following are the balance equations for states $0, 1a$, and $1b$.

$$\begin{aligned}
\pi(0) &= \beta(1-\alpha)\pi(1b) + (1-\alpha)\pi(0),\\
\pi(1a) &= \alpha\pi(0) + \alpha\beta\pi(1b),\\
\pi(1b) &= (1-\alpha)\pi(1a) + \beta(1-\alpha)\pi(2) + (1-\alpha)(1-\beta)\pi(1b).
\end{aligned} \tag{42}$$

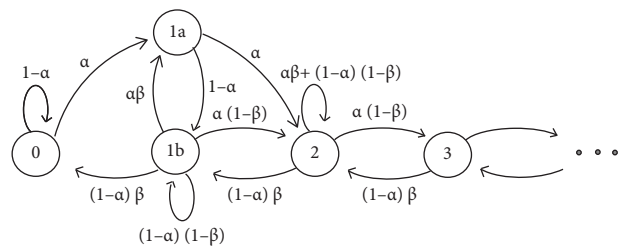

Figure 4: State diagram for the LAS-DA rule.

One can replace the balance equation for state $1b$ with the balance equation across the cut $S = \{0, 1a, 1b\}$ and its complement S^c, i.e.,

$$\alpha\pi(1a) + \alpha(1-\beta)\pi(1b) = \beta(1-\alpha)\pi(2). \tag{43}$$

Solve these equations to obtain

$$\begin{aligned}
\pi(1a) &= \frac{\alpha}{1-\alpha}\pi(0),\\
\pi(1b) &= \frac{\alpha}{\beta(1-\alpha)}\pi(0),\\
\pi(1) &= \frac{\alpha(1+\beta)}{\beta(1-\alpha)}\pi(0),\\
\pi(2) &= \frac{\alpha^2}{\beta^2(1-\alpha)^2}\pi(0).
\end{aligned} \tag{44}$$

Now, $\sum_{n=2}^{\infty}\pi(n) = \sum_{n=2}^{\infty}\gamma^{n-2}\pi(2) = \pi(2)/(1-\gamma)$ so that $\sum_{n=2}^{\infty}\pi(n) = \alpha^2\pi(0)/(\beta-\alpha)\beta(1-\alpha)$. Simplify to obtain the following.

Theorem 7. *The random observer stationary distribution function of the LAS-DA queue is given by*

$$\pi(n) = \begin{cases} \rho^2(1-\gamma)\gamma^{n-2}; & n = 2, 3, \cdots,\\ (\alpha+\rho)(1-\rho); & n = 1,\\ (1-\alpha)(1-\rho); & n = 0. \end{cases} \tag{45}$$

An alternative approach to obtain this distribution function is to use the relationship between the stationary

distribution functions at observation epochs τ and potential prearrival epochs as given by Hunter [2], page 204. But his approach requires that we know a priori, the distribution function at potential prearrival instants. Our approach is simpler and based on solving birth-death equations for $n \geq 2$ and global balance equations for $n \leq 2$.

4.3. The Outside Observer Stationary Distribution. Here, our interest is in the process $\{Z(\tau - .5), \tau = 1, 2, \cdots\}$; that is, the state of the system is observed at the outside observer epochs, i.e., at the slot centers $\tau - .5$, $\tau = 1, 2, \cdots$. Specifically, for various scheduling rules, we are interested in the stationary distribution function of the process given by $\{\pi^O(.)\}$, defined as

$$\pi^O(i) := \lim_{\tau \to \infty} \frac{Y(i; \tau - .5)}{(\tau - .5)}. \tag{46}$$

Here, $\pi^O(i)$ is the stationary probability that the process $\{Z(\tau - .5)\}$ is in state i, where state i is observed at the slot centers. Let $u = \tau - .5$, $\tau = 1, 2, \cdots$ and replicate the analysis in Sections 2 and 3, using $\{Z(u)\}$, we obtain similar results, where $\pi^O(.)$ replaces $\pi(.)$ for the results in (29) and (39). Specifically, we have the following theorem.

Theorem 8. *Consider the birth-death process of Section 3, but now, the state is observed at slot centers.*

(1) Let $\beta(0) = 0$, then the stationary distribution function is given by

$$\pi^O(n) = \begin{cases} \rho(1-\gamma)\gamma^{n-1}, & n = 1, 2, \cdots, \\ 1 - \rho, & n = 0. \end{cases} \tag{47}$$

(2) Let $\beta(0) = \beta$, then the stationary distribution function is given by

$$\pi^O(n) = \gamma^n (1 - \gamma), n \geq 0. \tag{48}$$

Note how $\pi^{O,\beta(0)=0}(n) = \pi^{\beta(0)=\beta}(n)$ and $\pi^{O,\beta(0)=\beta}(n) = \pi^{\beta(0)=0}(n)$ for all $n = 0, 1. \cdots$. Next, we give the stationary distribution function at the slot centers for each of the five scheduling rules.

4.3.1. Queues with EAS Rule. Here, we are interested in the distribution function at slot centers, then we have a birth-death model with $\beta(0) = 0$. Therefore, the stationary distribution function, $\{\pi^O(.)\}$, is given by (47). Note that the stationary distribution at the random observer epochs is associated with $\beta(0) = \beta$.

4.3.2. Queues with LAS-IA Rule. Here, we are interested in the distribution function at slot centers, and then we have a birth-death model with $\beta(0) = \beta$. Therefore, the stationary distribution function $\{\pi^O(.)\}$ is given by (48). In contrast, the stationary distribution at the random observer epochs is associated with $\beta(0) = 0$.

4.3.3. Queues with LA-AF and LA-DF Rules. Here, we are interested in the distribution function at slot centers, then we have a birth-death model with $\beta(0) = 0$. Therefore, the stationary distribution function, $\{\pi^O(.)\}$, is given by (47). This is the same as the stationary distribution at random observer epochs.

4.3.4. Queues with LAS-DA Rule. Here, we are interested in the distribution function at slot centers, and then we have a birth-death model with $\beta(0) = 0$. Therefore, the stationary distribution function $\{\pi^O(.)\}$ is given by (47).

Hunter [2] gives formulas for the EAS rule at the random observer epoch, and the LAS-DA and LAS-IA rules at the outside observer's epoch using generating function methods. Desert and Daduna [6] give formulas for the EAS, LA-DF, and LA-AF at the random observer epochs. See also Alfa [7], Alfa and Kim [8], Woodward [10], and Chaudhry et al. [21], among others. In this section, we presented a unified treatment of all these scheduling rules at slot edges and slot centers.

5. Bernoulli Arrivals See Time Averages

We investigate when Bernoulli Arrivals See Time Averages (BASTA), also referred to as GASTA (Geometric Arrivals See Time Averages). Interestingly, and contrary to the continuous time case, Bernoulli arrivals do not necessarily see time averages in the same sense as it happens in the continuous time case. We assume a single-server discrete-time queueing model with at most one potential arrival and/or departure in a time slot. How Bernoulli arrivals see time averages depends on the scheduling of the order of arrivals and departures in a time slot.

5.1. Characterization of BASTA. We start by giving a general characterization of BASTA for a discrete-time process with an imbedded arrival process without consideration of any scheduling rules. Let $\{T_n, n = 1, 2, \cdots\}$ be an imbedded point process associated with process $\{Z(\tau), \tau = 1, \cdots\}$, such that T_n is the arrival instant of the nth arrival. Let $N = \{N(\tau), \tau = 1, 2, \cdots\}$ be an associated counting process such that $N(\tau)$ counts the points of $\{T_n, n = 1, 2, \cdots\}$ in $[0, \tau]$, that is, process N counts the number of arrivals in $[0, \tau]$. Note that we assume one possible arrival at any given time instance.

For any state $k \in S$, $N(k; \tau) := \sum_{u=1}^{\tau} \mathbf{1}\{Z(T_u^-) = k\}$ counts the state k arrivals during $(0, \tau]$. Now, define the following limits when they exist.

$$\alpha := \lim_{\tau \longrightarrow \infty} \frac{N(\tau)}{\tau},$$
$$\alpha(k) := \lim_{\tau \longrightarrow \infty} \frac{N(k; \tau)}{Y(k; \tau)}, \tag{49}$$
$$\pi^A(k) := \lim_{\tau \longrightarrow \infty} \frac{N(k; \tau)}{N(\tau)}.$$

With $\{T_n\}$ a simple imbedded arrival process, we interpret α as the long-run arrival frequency, $\alpha(k)$ as the state-k long-run arrival frequency, and $\pi^A(k)$ as the long-run frequency of arrivals that find the system in state k. In a stochastic system, w.p.1, α is the probability of arrival at any given time instant, $\alpha(k)$ is the state k arrival probability, and $\pi^A(k)$ is the prearrival probability of finding k customers in the system upon arrival. The following sample path result states that prearrival probabilities equal time-average probabilities if and only arrival probabilities are state independent.

Theorem 9. *For any state $k \in S$, assume that all quantities are well defined. Then,*

$$\alpha(k)\pi(k) = \alpha\pi^A(k); \text{ for all } k = 0, 1, \cdots. \tag{50}$$

Moreover, if $\alpha(k) = \alpha$ for all $k \in \mathcal{S}$, then

$$\pi^A(k) = \pi(k); \text{ for all } k = 0, 1, \cdots. \tag{51}$$

Proof. It follows from the definitions that

$$\begin{aligned}\alpha(k)\pi(k) &= \lim_{\tau\to\infty}\left(\frac{N(k;\tau)}{Y(k;\tau)}\right)\times\frac{Y(k;\tau)}{\tau} = \lim_{\tau\to\infty}\frac{N(k;\tau)}{\tau},\\ \alpha\pi^A(k) &= \lim_{\tau\to\infty}\left(\frac{N(\tau)}{\tau}\right)\times\frac{N(k;\tau)}{N(\tau)} = \lim_{\tau\to\infty}\frac{N(k;\tau)}{\tau}.\end{aligned} \tag{52}$$

This proves the first part of the theorem. The second part is straightforward.

A continuous analog of the above sample-path version is proved in El-Taha and Stidham [28]. Note that the condition $\alpha(k) = \alpha$ for all $k \in S$ is the equivalent of the Lack-of-Bias (LBA) condition given by Makowski et al. [30].

Now, assume we are working with a stochastic process with an imbedded point process. Specifically, consider $\{Z(\tau, A(\tau)\}$, where $\{Z(\tau); \tau \geq 1\}$ is a discrete-time stochastic process and for each τ, $\{A(\tau) = 1\}$ if an arrival occurs and 0 otherwise. Typically, one makes the following Lack of Anticipation assumption (LAA) which is sufficient for BASTA to hold.

5.1.1. Lack of Anticipation Assumption (LAA). Assume that

$$\begin{aligned}&P(A(\tau) = k \mid Z(m-) = z_m; 1 \leq m \leq \tau-) = p(k),\\ &\sum_{k=0}^{1} p(k) = 1,\end{aligned} \tag{53}$$

for all τ, $z_1, \cdots, z_\tau$; and $\{p(k)\}$ is Bernoulli p.m.f. with parameter α.

The LAA assumption says that future arrivals are independent of the history of the process Z. It turns out that this is a too strong condition for ASTA. On the other hand, the weaker condition LBA says the future arrivals and the present state of the system are uncorrelated. This weak condition that is not easy to verify in practice. A stronger condition that says future arrivals are independent of the current state of the process seems to avoid these issues and work well for our purposes. Therefore, we will use the following lack of dependence assumption (LDA).

5.1.2. Lack of Dependence Assumption (LDA). Assume that for all $n \geq 0$

$$P(A(\tau) = k \mid Z(\tau-) = n) = p(k); \sum_{k=0}^{1} p(k) = 1, \tag{54}$$

for all τ, all n, and $\{p(k)\}$ is Bernoulli p.m.f. with parameter α.

Remark 10. The important assumption here is that LDA holds. The Bernoulli arrivals assumption by itself is not sufficient for BASTA to hold. We need the LDA independence assumption. To see why LDA is important, consider a system with Bernoulli arrivals that occur at times 2τ with probability α (geometric interarrival times) and let the service of kth arrival be one-half the next interarrival time. Then, all arrivals will see the system in state 0, but the system will spend half the time in state 0, that is $\pi^A(0) = 1$, but $\pi(0) = 0.5$ and $\pi(1) = 0.5$.

Since $\{A(\tau)\}$ is a Bernoulli process, one can see that under LDA the conditional state k-arrival probabilities

$$\alpha(k) = P(A(\tau) = 1 \mid Z(\tau-) = k)\, w.p.1, \tag{55}$$

and the prearrival probabilities

$$\pi^A(k) = P(Z(\tau-) = k \mid A(\tau) = 1)\, w.p.1, \tag{56}$$

for all $k \geq 0$.

Note that if the LDA holds, then $\alpha(k) = \alpha$ for all $k \in \mathcal{S}$. When $\{A(\tau)\}$ is a Bernoulli process, Theorem 9 is referred to as BASTA and sometimes GASTA for geometric arrivals see time averages. Several authors including Halfin [11], Makowski et al. [30], El-Taha and Stidham [28], and El-Taha and Stidham [31] have addressed the discrete-time BASTA, in the framework of a stochastic discrete-time process with an imbedded point process, and its variants. The BASTA issue in discrete-time queueing models with specific scheduling rules will be addressed next.

5.2. BASTA for Queues with Scheduling Rules. In a queueing system, one can think of the time instants $\tau, \tau = 1, 2, \cdots$ as the epochs where the system state is observed. Potential arrivals (departures) can be scheduled right before (after) or right after (before) a time instant τ. In such situations $\tau-$can be taught as a potential prearrival instant if a potential arrival is scheduled to occur in $(\tau-, \tau)$. Moreover, if a potential arrival is scheduled to occur in $(\tau, \tau+)$, then τ will be considered a potential prearrival instant.

BASTA suggests that, similar to the continuous time case, when arrivals follow a Bernoulli process, the prearrival probabilities will equal the corresponding random observer probabilities. This is true for a discrete-time process with an imbedded Bernoulli point process. However, it turns out that when we invoke scheduling rules, BASTA in the sense discussed above does not necessarily hold. For each of the five scheduling rules, we will give the expression for the prearrival probabilities (which may or may not equal the random observer probabilities). We assume arrivals to be i.i.d. and work with an equivalent LDA assumption that does not require the entire history of the process. In each case, we also refine the LDA assumption to fit the specified scheduling rule.

5.2.1. The EAS Rule. Here, we give a BASTA related relationship and related results using the generalized birth-death model. Because the arrivals are i.i.d. and taking into account that in EAS rule arrivals occur right after slot edges, the LDA assumption takes the following form.

For all $n \geq 0$

$$p(A(\tau+) = 1 \mid Z(\tau) = n) = p(A(\tau+) = 1). \tag{57}$$

Theorem 12. *Consider a single server queueing system with EAS Rule. Then,*

$$\pi^A(n) = \frac{\alpha(n)\pi(n)}{\sum_{k=0}^{\infty}\alpha(k)\pi(k)}. \tag{58}$$

In particular, if LDA holds, then

$$\pi^A(n) = \pi(n); n \geq 0. \tag{59}$$

Proof. Using the law of total probability, we obtain

$$\begin{aligned}\lim_{\tau\longrightarrow\infty} p(A(\tau+) = 1) &= \lim_{\tau\longrightarrow\infty}\sum_{k\in S} p(A(\tau+) = 1 \mid Z(\tau) = k)\\ &\quad\cdot p(Z(\tau) = k),\\ &= \sum_{k\in S}\alpha(k)\pi(k).\end{aligned} \tag{60}$$

Now, it follows that

$$\begin{aligned}\pi^A(n) &= \lim_{\tau\longrightarrow\infty} p(Z(\tau) = n \mid A(\tau+) = 1)\\ &= \lim_{\tau\longrightarrow\infty}\frac{p(Z(\tau) = n, A(\tau+) = 1)}{p(A(\tau+) = 1)}\\ &= \lim_{\tau\longrightarrow\infty}\frac{p(A(\tau+) = 1 \mid Z(\tau) = n)p(Z(\tau) = n)}{p(A(\tau) = 1)}\\ &= \frac{\alpha(n)\pi(n)}{\sum_{k=0}^{\infty}\alpha(k)\pi(k)}.\end{aligned} \tag{61}$$

This proves the first part of the theorem. The second part follows by noting that the LDA assumption implies $\alpha(n) = \alpha$ for all $n \geq 0$.

The relation in the first part of Theorem 12 is the discrete-time counterpart of a similar one that has been used in continuous-time stochastic models as the basis for a proof of PASTA (Cooper [32]). □

5.2.2. The LAS-DA and LAS-IA Rules. Here, we derive relationships between prearrival probabilities and time average probabilities using the generalized birth-death model for both the LAS-IA and LAS-DA rules. We assume late arrival, that is, a potential arrival occurs at $(\tau-, \tau)$. We also assume that the prearrival observed instance falls after the occurrence of a potential departure. Because potential departures occur before potential arrivals in any time slot, the number of customers an arrival sees depends on whether an actual departure occurs in $(\tau - 1, (\tau - 1)+)$.

Lemma 14. *For any LAS generalized birth-death model with state n arrival and service probabilities $\alpha(n)$ and $\beta(n)$, respectively, we have*

$$\pi^A(n) = \frac{1}{\alpha}[\alpha(n)(1-\beta(n)\pi(n) + \alpha(n+1)\beta(n+1)\pi(n+1)], \tag{62}$$

where $\alpha = \sum_{k=0}^{\infty}\alpha(k)\pi(k)$.

In particular, if $\alpha(n) = \alpha$ for all $n \geq 0$ (state independent), then

$$\pi^A(n) = (1 - \beta(n))\pi(n) + \beta(n+1)\pi(n+1). \tag{63}$$

Proof. Now, if the state at a potential prearrival instant $\tau-$ is n, i.e., an arrival sees n customers in the system, then the state at τ is $n + 1$. So,

$$\begin{aligned}
\pi^A(n) &= \lim_{\tau\longrightarrow\infty} p(Z(\tau)=n+1\mid A(\tau-)=1),\\
&= \lim_{\tau\longrightarrow\infty} \frac{p(Z(\tau)=n+1, A(\tau-)=1)}{p(A(\tau-)=1)},\\
&= \lim_{\tau\longrightarrow\infty} \sum_{r\in S} \frac{p(Z(\tau)=n+1, Z(\tau-1)=r, A(\tau-)=1)}{p(A(\tau-)=1)},\\
&= \lim_{\tau\longrightarrow\infty} \sum_{r\in\{n,n+1\}} p(Z(\tau)=n+1|Z(\tau-1)=r, A(\tau-)=|1)p(A(\tau-)=1|Z(\tau-1)=r),\\
&\quad\times \frac{p(Z(\tau-1)=r)}{p(A(\tau-)=1)},\\
&= \frac{[\alpha(n)(1-\beta(n)\pi(n)+\alpha(n+1)\beta(n+1)\pi(n+1)]}{\sum_{k=0}^{\infty}\alpha(k)\pi(k)}.
\end{aligned} \tag{64}$$

The lemma follows by noting that $\alpha=\sum_{k=0}^{\infty}\alpha(k)\pi(k)$.

Daduna [5] uses this type of argument. Using (63), we see that BASTA does not hold here in the sense that prearrival probabilities equal time-average probabilities where the average is taken with respect to time instants at the slot edges, τ. We can use this relationship to derive an expression for the prearrival probabilities. Instead, we will use a different approach where we relate prearrival probabilities to the outside observer probabilities.

5.2.3. LDA for LAS Rules. For the LAS rules, the lack of anticipation assumption takes the form.

$$p(A(\tau-)=1\mid Z(\tau-.5)=n)=p(A(\tau-)=1). \tag{65}$$

Lemma 15. *For any LAS generalized birth-death model with state n arrival and service probabilities $\alpha(n)$ and $\beta(n)$, respectively, we have*

$$\alpha\pi^A(n)=\alpha(n)\pi^O(n) \text{ for all } n\geq 0. \tag{66}$$

In particular, if LDA holds, then

$$\pi^A(n)=\pi^O(n); n=0,1,\cdots, \tag{67}$$

Proof. For all $n\geq 0$,

$$\begin{aligned}
\pi^A(n) &= \lim_{\tau\to\infty} p(Z(\tau-.5)=n\mid A(\tau-)=1),\\
&= \lim_{\tau\to\infty} \frac{p(Z(\tau-.5)=n, A(\tau-)=1)}{p(A(\tau-)=1)},\\
&= \lim_{\tau\to\infty} \frac{p(A(\tau-)=1\mid Z(\tau-.5)=n)P(Z(\tau-.5)=n)}{p(A(\tau-)=1)},\\
&= \frac{\alpha(n)\pi^O(n)}{\alpha}.
\end{aligned} \tag{68}$$

Note that the LDA assumption implies that $\alpha(n)=\alpha$ for all $n\geq 0$. The lemma follows by noting that $\alpha=p(A(\tau-)=1)$.

Instead of $\tau-0.5$, one can use any $u\in((\tau-1)+,\tau-)$ which is referred to as the outside observer interval.

5.2.4. The LA-AF Rule. Here, we consider the LA-AF rule. In this rule, both potential arrivals and departures occur at the end of a time slot with arrivals occurring before departures.

For the LA-AF models, the lack of dependence assumption takes the form.

$$p(A(\tau--)=1\mid Z(\tau-1)=n)=p(A(\tau--)=1). \tag{69}$$

Note that Z can be observed for any u in the interval $[\tau-1,\tau--)$ since for any such u, the system state does not change.

Lemma 17. *For any LA-AF generalized birth-death model with state n arrival and service probabilities $\alpha(n)$ and $\beta(n)$, respectively, we have*

$$\alpha\pi^A(n)=\alpha(n)\pi(n) \text{ for all } n\geq 0. \tag{70}$$

In particular, if LDA holds, then

$$\pi^A(n)=\pi(n)\, n=0,1,\cdots, \tag{71}$$

Proof. For all $n\geq 0$

$$\begin{aligned}
\pi^A(n) &= \lim_{\tau\to\infty} p(Z(\tau-1)=n\mid A(\tau--)=1)\\
&= \lim_{\tau\to\infty} \frac{p(Z(\tau-1)=n, A(\tau--)=1)}{p(A(\tau--)=1)}\\
&= \lim_{\tau\to\infty} \frac{p(A(\tau--)=1\mid Z(\tau-1)=n)P(Z(\tau-1)=n)}{p(A(\tau--)=1)}\\
&= \frac{\alpha(n)\pi(n)}{\alpha}.
\end{aligned} \tag{72}$$

Note that the LDA assumption implies that $\alpha(n) = \alpha$ for all $n \geq 0$. The lemma follows by noting that $\alpha = p(A(\tau - -) = 1)$.

Remark 18. Note that in LDA, if we selected $u = \tau - 0.5$ instead of $\tau - 1$, then we would have ended up with the distribution $\pi^O(.)$, instead $\pi(.)$. So, the choice of $\tau - 0.5$ as an observation epoch allows another independent pathway to prove BASTA implies that for the LA-AF rule $\pi^A(n) = \pi^O(n)$; $n = 0, 1 \cdots$.

5.2.5. The LA-DF Rule. We assume LA-DF, that is, a potential arrival occurs at $(\tau-, \tau)$. In this model, a prearrival observed instance falls after the occurrence of a potential departure. Because departures occur before arrivals, the number of customers an arrival sees depends on whether an actual departure occurs in $(\tau - 1, \tau-)$.

For the LA-DF model, the lack of dependence assumption takes the form.

$$p(A(\tau) = 1 \mid Z(u) = n) = p(A(\tau) = 1);\ u \in (\tau - -, \tau-). \quad (73)$$

Lemma 20. *For any LA-DF generalized birth-death model with state n arrival and service probabilities $\alpha(n)$ and $\beta(n)$, respectively, we have*

$$\pi^A(n) = \frac{1}{\alpha}[\alpha(n)(1 - \beta(n)\pi(n) + \alpha(n+1)\beta(n+1)\pi(n+1)], \quad (74)$$

where $\alpha = \sum_{k=0}^{\infty}\alpha(k)\pi(k)$.

In particular, if LDA holds, then $\alpha(n) = \alpha$ for all $n \geq 0$ (state independent), and

$$\pi^A(n) = (1 - \beta(n))\pi(n) + \beta(n+1)\pi(n+1). \quad (75)$$

The statement and proof of this result are similar to those of Lemma 17.

Again, using (75), we see that BASTA does not hold here in the sense that prearrival probabilities equal random observer probabilities. The following corollary gives an expression for the prearrival probabilities.

Corollary 21. *Assume LDA holds, and let the service time be geometric with mean $1/\beta$. Then,*

$$\pi^A(n) = (1 - \gamma)\gamma^n, \quad n = 0, 1, \cdots. \quad (76)$$

Proof. Note that LDA implies that $\alpha(n) = \alpha$ for all $n \geq 0$, so that (75) holds. For the LA-DF rule, $\beta(0) = 0$, so it follows from (75) that $\pi^A(0) = (1 - \rho) + \beta\rho(1 - \gamma = 1 - \gamma$. Moreover, $\pi^A(n) = (1 - \beta)\pi(n) + \beta\gamma\pi(n)$, $n \geq 1$. Simplify to obtain the desired result.

Note that in this LA-DF rule, the prearrival stationary distribution does not equal the stationary distribution at slot edges (random observer) and not slot centers (outside observer). However, it is equal to one of the two birth-death forms discussed in Sections 2 and 3.

Gravey and Hebuterne [12] address BASTA related results for the LAS and EAS scheduling rules. They show that, using the results of Halfin [11], for LAS rules, the distribution function at arrival instants equals the distribution function at the outside observer epochs. They also conclude that the same result does not hold for the EAS rule (in the sense that arrivals see time averages at the outside observer instants). Daduna [5] studies LA-DF and LA-AF. He shows that BASTA holds for the LA-AF in the sense that arrivals see the same as random observers. Referring to LA-DF, he states that BASTA does not function the same as the continuous time case. Desert and Daduna [6] obtain formulas for the distribution functions at prearrival epochs for the three scheduling rules EAS, LA-DF, and LA-AF. Our results complement their conclusions in the sense that we give the distribution function at prearrival instants for all five scheduling rules. Moreover, instead of using the LAA assumption, we state a specific weaker LDA assumption for each scheduling rule. This, we believe, leads to a better understanding of BASTA in regard with applying it in discrete-time queues with specific scheduling rules. □

5.3. Summary. In Table 1, we provide a summary of the results for stationary distribution functions for five scheduling rules at three observation epochs.

The results in this summary are not new, for the most part, neither the use of birth-death processes. However, the use of one birth-death equation to generate all the results for the random and outside observers is novel. Several results appear in Hunter [2] who deals with EAS, LAS-IA, and LAS-DA scheduling rules. Specifically, for the EAS rule, Hunter [2] pages 197 and 199, respectively, give the outside and random observer's results. Moreover, for the EAS and LAS-IA rules, the arrival-times distribution function can be seen to follow from Example 9.4.1 of Hunter [2], page 248. Similarly, the LAS-DA arrival-times distribution function follows from Example 9.4.2 of Hunter [2], page 252. For LAS-DA outside observer, and LAS-IA outside and random observers, the distribution functions can be obtained by utilizing relationships between scheduling rules at various embedding epochs, as given by Hunter [2] pages 9 and 253. The results obtained by Hunter, however, utilize generating function techniques. By contrast, results for the random and outside observer's epochs in the table follow from using one birth-death model. For the LAS-DA random observer, one can manipulate the relationships given by Hunter [2], page 204, to obtain the distribution function for this case. Our approach given by (45) is simpler. Additionally, the LA-DF random observer distribution function can be obtained from El-Taha et al. [4] who give the result for the $B/G/1$-LA-DF round Robin model. The results are the same due to the insensitivity of the round Robin model to service times distribution function. The LA-DF and LA-AF outside observer's distribution functions are given by Daduna [5], and the

Table 1: Scheduling rules limiting distributions at different epochs.

	Random observer	Arrival times	Outside observer
EAS	$\pi(n) = (1-\gamma)\gamma^n$ $\pi(0) = 1-\gamma$	$\pi^A(n) = (1-\gamma)\gamma^n$ $\pi^A(0) = 1-\gamma$	$\pi^O(n) = \rho(1-\gamma)\gamma^{n-1}$ $\pi^O(0) = 1-\rho$
LAS-IA	$\pi(n) = \rho(1-\gamma)\gamma^{n-1}$ $\pi(0) = 1-\rho$	$\pi^A(n) = (1-\gamma)\gamma^n$ $\pi^A(0) = 1-\gamma$	$\pi^O(n) = (1-\gamma)\gamma^n$ $\pi^O(0) = 1-\gamma$
LAS-DA	$\pi(n)$ given by (45) $\pi(0) = (1-\alpha)(1-\rho)$	$\pi^A(n) = \rho(1-\gamma)\gamma^{n-1}$ $\pi^A(0) = 1-\rho$	$\pi^O(n) = \rho(1-\gamma)\gamma^{n-1}$ $\pi^O(0) = 1-\rho$
LA-AF	$\pi(n) = \rho(1-\gamma)\gamma^{n-1}$ $\pi(0) = 1-\rho$	$\pi^A(n) = \rho(1-\gamma)\gamma^{n-1}$ $\pi^A(0) = 1-\rho$	$\pi^O(n) = \rho(1-\gamma)\gamma^{n-1}$ $\pi^O(0) = 1-\rho$
LA-DF	$\pi(n) = \rho(1-\gamma)\gamma^{n-1}$ $\pi(0) = 1-\rho$	$\pi^A(n) = (1-\gamma)\gamma^n$ $\pi^A(0) = 1-\gamma$	$\pi^O(n) = \rho(1-\gamma)\gamma^{n-1}$ $\pi^O(0) = 1-\rho$

Note that for all models $n \geq 1$.

LA-DF and LA-AF arrival-times distribution functions are given by Desert and Daduna [6]. We point out that Daduna [5] and Desert and Daduna [6] use birth-death models. Furthermore, the prearrival and outside observer probabilities for the EAS and LAS-DA can also be deduced from the GI/Geom/1 results given by Chaudhry et al. [21] and Takagi [33]. However, their results are obtained using generating functions.

We are able to obtain an expression for the prearrival probability distribution functions for all scheduling rules. This expression is not the same for all the scheduling rules, which led some in the literature to speculate that BASTA does not always hold. We see, at least in the cases covered here, that BASTA holds in its own special way. When one considers discrete-time queues with varied scheduling rules, the arrival-time probabilities will equal the time-average probabilities, but the time-averaging can be at slot boundaries (random observer epochs) or slot centers (outside observer) depending on the applied scheduling rule. Only for EAS and LA-AF rules, the prearrival probabilities equal the random observer probabilities, a result that parallels the continuous time case.

Note that with exception of the random observer LAS-DA model; all other fourteen expressions have one of the two birth-death forms. Moreover, the mean number of customers in the system $L = (\alpha(1-\alpha))/(\beta-\alpha)$ is obtained from the expressions that contain ρ. Applying Little's law, we obtain $W = (1-\alpha)/(\beta-\alpha)$ as we shall see in the next subsection.

An important observation here is that no two stationary distribution functions for any pair of scheduling rules give identical results for the three observation epochs. Note that one can also use other epochs to evaluate time average distribution functions like potential prearrival, potential postarrival, potential predeparture, and potential postdeparture epochs. There is no equivalent for these epochs in the continuous time case, as all these epochs lead to averaging continuously over time.

6. Waiting Times

Now, using the prearrival stationary distribution functions and Little's law, we give the system performance measures and show that their proofs are discrete counterparts of the $M/M/1$ continuous-time case. Now, let T_q be a random variable that represents the time in the queue and let $Wq(j) = p(T_q \leq j), j = 0, 1, \cdots$ be the cumulative distribution function of the time in the queue. Then, we have the flowing preliminary result.

Lemma 22. *For any of the five scheduling rules,*

$$W_q(0) = 1-\gamma. \tag{77}$$

Proof. Note that for any of the EAS, LAS-IA, or LA-DF rules, the customer delay in the queue is 0 if upon arrival to an empty system, the customer starts service immediately. Therefore, $W_q(0) = \pi^A(0)$. Then, the lemma follows from the summary results in Subsection 5.2. Now, assume LAS-DA or LA-AF rule. For these systems, $W_q(0) = \pi^A(0) + \beta\pi^A(1)$, where β is the probability of departure from state 1. This is because in these systems, an arrival that finds one customer with one unit of service does not wait and goes immediately into service. For the LAS-DA, this happens after its one unit delayed access. In this case,

$$W_q(0) = (1-\rho) + \beta\rho(1-\gamma) = 1-\gamma. \tag{78}$$

This completes the proof.

In the following theorem, we give an elementary direct proof for the waiting time distribution. □

Theorem 23. *Let $w_q(j) = p(T_q = j), j = 0, 1, \cdots$ be the probability mass function of the time in queue. Then,*

$$w_q(j) = \begin{cases} \gamma(1-\delta)\delta^{j-1}, & j = 1, 2, \cdots, \\ 1-\gamma, & j = 0. \end{cases} \tag{79}$$

Moreover, the cumulative distribution function of the time in the queue is given by

$$W_q(j) = 1-\gamma\delta^j, j = 0, 1, \cdots, \tag{80}$$

where $\delta = (1-\beta)/(1-\alpha)$.

Proof. We assume any of EAS, LAS-IA, or LA-DF rules. For $j = 1, 2, \cdots$, we have

$$\begin{aligned} w_q(j) &= p\left(T_q = j\right) \\ &= \sum_{n=1}^{j} p\,(n \text{ service completions in } j \text{ time units } \textit{arrival finds } n \textit{ in system}) \cdot \pi^A(n) \\ &= \sum_{n=1}^{j} \binom{j-1}{n-1} \beta^{n-1} (1-\beta)^{j-n} \cdot \beta \cdot \pi^A(n) \\ &= \sum_{n=1}^{j} \binom{j-1}{n-1} \beta^{n} (1-\beta)^{j-n} \times (1-\gamma)\gamma^n. \end{aligned} \tag{81}$$

Note that in the third line, $n-1$ of the departures occur in $j-1$ time units and the last departure occurs at time j. Now, simplify to get

$$\begin{aligned} w_q(j) &= (1-\gamma)(1-\beta)^j \sum_{n=1}^{j} \binom{j-1}{n-1} \left(\frac{\alpha}{1-\alpha}\right)^n, \\ &= (1-\gamma)\,\frac{\alpha}{1-\alpha}(1-\beta)^j \left(1+\frac{\alpha}{1-\alpha}\right)^{j-1}, \\ &= (1-\gamma) \times \frac{\alpha}{1-\alpha}(1-\beta)^j \left(\frac{1}{1-\alpha}\right)^{j-1}, \\ &= \frac{\beta-\alpha}{\beta(1-\beta)} \times \alpha\left(\frac{1-\beta}{1-\alpha}\right)\left(\frac{1-\beta}{1-\alpha}\right)^{j-1}, \\ &= \gamma(1-\delta)\delta^{j-1}, \end{aligned} \tag{82}$$

where we use the fact that $(1-\gamma) = (\beta-\alpha)/(\beta(1-\alpha))$ and $(1-\delta) = (\beta-\alpha)/(1-\alpha)$.

Now, we assume LAS-DA or LA-AF rules. For these systems, for a customer to be delayed j units, there has to be $j+1$ service units and one departure at time $j+1$. Note that because the first service position does not cause any delay, an arrival must see $j+1$ units of work in the system for a j–unit delay. Additionally, when $n = 1$, work in the system (i.e., the remaining service of the customer in service) $j \geq 2$. Therefore,

$$\begin{aligned} w_q(j) &= \sum_{n=1}^{j+1} p\,(n \text{ service completions in } j+1 \text{ time units} \mid \textit{arrival finds } n \textit{ in system}) \cdot \pi^A(n) \\ &= \sum_{n=1}^{j+1} \binom{j}{n-1} \beta^{n-1} (1-\beta)^{j-n+1} \cdot \beta \cdot \pi^A(n) \\ &= \sum_{n=1}^{j+1} \binom{j}{n-1} \beta^{n} (1-\beta)^{j-n+1} \times \rho(1-\gamma)\gamma^{n-1}. \end{aligned} \tag{83}$$

Simplify to obtain

$$\begin{aligned} w_q(j) &= \rho(1-\gamma)(1-\beta)^{j+1} \frac{\beta}{1-\beta} \sum_{n=1}^{j+1} \binom{j}{n-1} \left(\frac{\alpha}{1-\alpha}\right)^{n-1} \\ &= \rho(1-\gamma)\,\frac{\beta}{1-\beta}(1-\beta)^{j+1}\left(1+\frac{\alpha}{1-\alpha}\right)^{j} \\ &= \rho(1-\gamma) \times \beta\left(\frac{1-\beta}{1-\alpha}\right)^{j} \\ &= \frac{\beta-\alpha}{\beta(1-\alpha)} \times \alpha\left(\frac{1-\beta}{1-\alpha}\right)\left(\frac{1-\beta}{1-\alpha}\right)^{j-1} \\ &= \gamma(1-\delta)\delta^{j-1}. \end{aligned} \tag{84}$$

The second part follows by letting $W_q(j) = \sum_{i=0}^{j} w_q(i)$ and simplifying.

Performance measures are obtained using Theorem 23 and Little's law. □

Theorem 24. *The mean waiting time and the number of customers in the queue and in the system are given by*

$$\begin{aligned} W_q &= \frac{\alpha(1-\beta)}{\beta(\beta-\alpha)}, \\ W &= \frac{1-\alpha}{\beta-\alpha}, \\ L &= \frac{\alpha(1-\alpha)}{\beta-\alpha}, \\ Lq &= \frac{\alpha^2(1-\beta)}{\beta(\beta-\alpha)}. \end{aligned} \tag{85}$$

Proof. The mean waiting time in the queue is obtained as

$$\begin{aligned} W_q = E\left[T_q\right] &= \sum_{j=1}^{\infty}(1 - Wq(j)), \\ &= \frac{\alpha(1-\beta)}{\beta(\beta-\alpha)}, \end{aligned} \tag{86}$$

Now, W is obtained using $W = W_q + E[S]$ where $E[S] = 1/\beta$ is the mean service time. Moreover, L and L_q are now obtained using Little's law.

Remark 25. For the EAS and LAS-DA systems, L obtained here is associated with the outside observer distribution. For the LAS-IA, L is associated with the random observer distribution.

Hunter [2] recognizes that the three schedule rules EAS and LAS-DA and LAS-DA have the same waiting time distribution. He gives the waiting time distribution function utilizing generating function techniques. Desert and Daduna [6] obtain waiting time results for EAS, LA-DF, and LA-AF and show that the distribution function is the same for the three scheduling rules using generating function methods. We extend those results to all five scheduling rules and provide a direct unified proof for all cases and for both the density and cumulative distribution functions.

7. Applications Using Global Balance

In this section, we consider three examples of Markovian models, namely, a multiserver, finite source, and batch-arrival single server models that do not fit the birth-death equations, but their stationary distribution functions can still be computed efficiently by recursive methods. It is well known that the multiserver and finite source continuous-time models can be considered as special cases of the continuous birth-death equations. By contrast, the corresponding discrete-time models do not fit the birth-death equations. Here, we use global balance equations to obtain the stationary distribution functions using recursive methods.

7.1. Finite Buffer Multiserver Model. Consider a finite buffer multiserver model denoted by $B/\text{Geo}/c/N$, where B indicates Bernoulli arrivals with parameter α, geometric service times with parameter β, $c \geq 1$ servers, and a finite buffer $N \geq c$, where N represents the number of servers and the waiting space. The loss model $B/\text{Geo}/c/c$ is a special case with $N = c$. The state is the number of customers in the system in steady state. Transitions occur because of arrivals to the system and/or service completions. We assume that in any time slot, departures occur before arrivals. This is consistent with the LAS-IA and LA-DF systems. For $i, k = 0, \cdots N$, the transition probabilities are given as follows. The probability of arrival that takes the system from state i to $i+1$ is given by

$$p(i, i+1) = \alpha(1-\beta)^{\min(i,c)}, \tag{87}$$

the probability that the system moves from state i to $i-k$ is given by

$$\begin{aligned} p(i, i-k) &= (1-\alpha)\binom{i}{k}\beta^k(1-\beta)^{i-k} \\ &\quad + \alpha\binom{i}{k+1}\beta^{k+1}(1-\beta)^{i-k-1}, \\ &\quad i = 0, 1, \cdots, c-1; k = 0, 1, \cdots, i, \\ p(i, i-k) &= (1-\alpha)\binom{c}{k}\beta^k(1-\beta)^{c-k} \\ &\quad + \alpha\binom{c}{k+1}\beta^{k+1}(1-\beta)^{c-k-1}, \\ &\quad i = c, 1, \cdots, N-1; k = 0, 1, \cdots, c, \\ p(N, N) &= (1-\beta)^c + c\alpha\beta(1-\beta)^{c-1}, \end{aligned} \tag{88}$$

and 0, otherwise. Here, $\binom{n}{k} = 0$; if $k < 0$, or $k > n$. We point out that when time is slotted as in communication networks, these transition probabilities do not allow a job to enter and leave state 0, i.e., an empty queue. When that is permitted, then slightly modified transition probabilities can be derived. For details, see Robertazzi ([34], Chapter 6). Now, the global balance equations are given by $(j = 0, \cdots, N)$

$$\pi(j) = \sum_{i=\max(0,j-1)}^{\min(N,j+c)} \pi(i)p(i,j), \sum_{j=0}^{N}\pi(j) = 1. \tag{89}$$

Now, we solve the global balance equations in (89) recursively starting with $\pi(N)$ and working backward. Note that there is no known closed form expression for this multiserver model.

Theorem 26. *Let $C(N) = 1$. Then, for $m = N, \cdots, 0$,*

$$\pi(m) = \frac{C(m)}{\sum_{i=0}^{N} C(i)}; m = 0, 1, \cdots, N, \tag{90}$$

where $C(m), m = N-1, \cdots, 0$ are obtained recursively such that

$$C(m) = \frac{\left[C(m+1)(1-p(m+1,m+1)) - \sum_{i=m+2}^{\min(N,m+c+1)} C(i)p(i,m+1)\right]}{\alpha(1-\beta)^{m'}}, \tag{91}$$

where $m' = \min(m, c)$.

Proof. The proof is by induction. From the global balance equations (89)

$$\pi(N) = \pi(N-1)p(N-1,N) + \pi(N)p(N,N), \tag{92}$$

so that

$$\pi(N-1) = \left[\frac{\pi(N)(1-p(N,N))}{p(N-1,N)}\right] = C(N-1)\pi(N). \tag{93}$$

Assume $\pi(k) = C(k)\pi(N)$ for $k = m+1, \cdots, N$ and show that $\pi(m) = C(m)\pi(N)$. Now, for $m = 0, \cdots, N-2$, write the global balance (89) as

$$\begin{aligned}\pi(m+1) &= \sum_{i=\max(0,m)}^{\min(N,m+c+1)} \pi(i)p(i,m+1)\\ &= \pi(m)p(m,m+1) + \pi(m+1)p(m+1,m+1)\\ &\quad + \sum_{i=m+2}^{\min(N,m+c+1)} \pi(i)p(i,m+1),\end{aligned} \tag{94}$$

which can be written as

$$\begin{aligned}\pi(m) &= \frac{\left(\pi(m+1)(1-p(m+1,m+1)) - \sum_{i=m+2}^{\min(N,m+c+1)} \pi(i)p(i,m+1)\right)}{p(m,m+1)}\\ &= C(m)\pi(N).\end{aligned} \tag{95}$$

Noting that $p(m,m+1) = \alpha(1-\beta)^{m'}$, and normalizing, we complete the proof.

Note that

$$p(m,m+1) = \begin{cases} \alpha(1-\beta)^{c} & m = N-1, \cdots, c,\\ \alpha(1-\beta)^{m} & m = c-1, \cdots, 0.\end{cases} \tag{96}$$

The theorem leads to the following Algorithm 1.

This recursive procedure is effective for small N. When N is large, this method can lead to overflow problems. There are several methods in the literature to deal with stability and overflow issues. Gao et al. [34] analyzed a multiserver model with geometric service times and general interarrival times using a generating functions approach.

7.2. Finite Source Discrete-Time Model; B/Geo/1//N. Consider a discrete-time model with N identical machines. Individual machine failures follow a Bernoulli process with parameter α; that is, an up machine fails with probability α at any given time instant. Failed machines join a queue for repair if one is in service; otherwise, a failed machine joins the service immediately. Service times are geometric with parameter β. Repaired machines join up machines immediately upon repair. This queueing system is also known as the finite population model. The state of the system is the number of down machines in steady state taking values in the state space $\{0, \cdots, N\}$. For $i, j = 0, \cdots, N$, the transition probabilities are given by

$$\begin{aligned}p(0,j) &= \binom{N}{j}\alpha^{j}(1-\alpha)^{N-j};\ j = 0, \cdots, N,\\ p(i,j) &= (1-\beta)\binom{N-i}{j-i}\alpha^{j-i}(1-\alpha)^{N-j}\\ &\quad + \beta\binom{N-i}{j-i+1}\alpha^{j-i+1}(1-\alpha)^{N-j-1},\\ i &= 1, \cdots, N;\ j = i-1, \cdots, N,\end{aligned} \tag{97}$$

and 0 otherwise. Note that $p(i, i-1)$ simplifies to

$$p(i,i-1) = \beta(1-\alpha)^{N-i};\ i = 1, \cdots, N. \tag{98}$$

This form of transition probabilities indicate that in state 0, no failed machine can be repaired in the same slot. This again is consistent with the models where departures precede arrivals in a time slot. This is the case for LAS-IA and LA-DF models. Moreover, the global balance equations are given by

$$\pi(j) = \sum_{i=0}^{j+1} \pi(i)p(i,j), \sum_{j=0}^{N} \pi(j) = 1. \tag{99}$$

(1) Let $C(N) = 1$.
(2) For $m = N-1, \cdots, c$
$C(m) = [C(m+1)(1-p(m+1,m+1)) - \sum_{i=m+2}^{\min(N,m+c+1)} C(i)p(i,m+1)]/\alpha(1-\beta)^c$.
(3) For $m = c-1, \cdots, 0$
$C(m) = [C(m+1)(1-p(m+1,m+1)) - \sum_{i=m+2}^{\min(N,m+c+1)} C(i)p(i,m+1)]/\alpha(1-\beta)^m$.
(4) For $m = 0, \cdots, N$, set
$\pi(m) = C(m)/\sum_{i=0}^{N} C(i)$.

Algorithm 1: Stationary probabilities for the multi-server model.

Now, we solve the global balance equations in (99) recursively starting with $\pi(0)$.

Theorem 27. *The stationary distribution* $\{\pi(n), n = 0, 1, \cdots, N\}$ *is given by*

$$\pi(n) = \frac{C(n)}{\sum_{i=0}^{N} C(i)}, \quad n = 0, 1, \cdots, N, \tag{100}$$

where $C(n), n = 0, \cdots, N$ *are obtained recursively such that* $C(0) = 1$, *and*

$$C(n) = \frac{\left(C(n-1)(1-p(n-1,n-1)) - \sum_{m=0}^{n-2} C(m)p(m,n-1)\right)}{p(n,n-1)}, \quad n = 1, \cdots, N. \tag{101}$$

Proof. The proof is by induction. From the global balance equations (99), we have

$$\pi(0)(1-p(0,0)) = \pi(1)p(1,0), \tag{102}$$

so that

$$\begin{aligned}\pi(1) &= \frac{\pi(0)(1-p(0,0))}{p(1,0)} \\ &= \frac{\pi(0)(1-(1-\alpha)^N)}{\beta(1-\alpha)^{N-1}} = C(1)\pi(0).\end{aligned} \tag{103}$$

Assume $\pi(k) = C(k)\pi(N)$ for $k = 1, \cdots, n-1$ and show that $\pi(N) = C(N)\pi(N)$. Now, rewrite the global balance (99) as

$$\pi(n-1) = \sum_{m=0}^{n} \pi(m)p(m,n-1), \tag{104}$$

which gives

$$\begin{aligned}&\pi(n-1)(1-p(n-1,n-1)) \\ &\quad = \sum_{m=0}^{n-2} \pi(m)p(m,n-1) + \pi(n)p(n,n-1),\end{aligned} \tag{105}$$

so that

$$\begin{aligned}\pi(n) = \Bigg[&C(n-1)(1-p(n-1,n-1)) \\ &- \sum_{m=0}^{n-2} C(m)p(m,n-1)\Bigg)\Bigg]\pi(0)/p(n,n-1),\end{aligned} \tag{106}$$

which proves the theorem.

This theorem leads to the following Algorithm 2 to calculate stationary probabilities. □

7.3. Discrete-Time Batch Arrival Model; B^X/Geo/1. Let A be the number of arrivals at any given time instant. Arrivals at different time instants are i.i.d., and $P(A = i) = \alpha_i$, $i = 0, 1, \cdots$, and 0 otherwise; and $\sum_{i=0}^{\infty} \alpha_i = 1$. Service times are i.i.d. They are independent of arrival times. Let the service time S be geometric with parameter $0 < \beta < 1$. For $i, j = 0, \cdots$, the transition probabilities are given by

$$\begin{aligned}p(0,i) &= \alpha_i; i = 0, \cdots, \\ p(i,i-1) &= \alpha_0\beta; i = 1, \cdots, \\ p(i,j) &= \alpha_{j-i}(1-\beta) + \alpha_{j-i+1}\beta; i = 1, \cdots; j = i, i+1, \cdots,\end{aligned} \tag{107}$$

and 0 otherwise. These transition probabilities are consistent with the LAS-IA and LA-DF scheduling rules. Moreover, the global balance equations are given by

$$\pi(j) = \sum_{i=0}^{j+1} \pi(i)p(i,j), \sum_{j=0}^{\infty} \pi(j) = 1. \tag{108}$$

Now, we solve the global balance equations in (108) recursively.

Theorem 28. *The stationary distribution* $\{\pi(n), n = 0, 1, \cdots,\}$ *is given by*

$$\pi(n) = \frac{C(n)}{\sum_{i=0}^{N} C(i)}; n = 0, 1, \cdots, N, \tag{109}$$

where $C(n), n = 0, \cdots, N$ *are obtained recursively such that* $C(0) = 1$, $C(1) = (1-p(0,0))/p(1,0)$, *and for* $n \geq 2$

(1) Let $C(0) = 1$.
(2) For $n = 1, \cdots, N$, use (101) to compute $C(n)$ recursively.
(3) For $n = 0, \cdots, N$, set $\pi(n) = C(n)/\sum_{i=0}^{N} C(i)$.

Algorithm 2: Stationary probabilities for the finite source model.

$$C(n) = \frac{\left[(1 - p(n-1, n-1))C(n-1) - p(0, n-1) - \sum_{i=1}^{n-2} p(i, n-1)C(i)\right]}{p(n, n-1)}. \tag{110}$$

Proof. The proof is by induction. From the global balance equations (108), we have

$$\pi(0) = \pi(0)p(0,0) + \pi(1)p(1,0) \tag{111}$$

so that

$$\pi(1) = \frac{(1 - p(0,0))\pi(0)}{p(1,0)} = C(1)\pi(0). \tag{112}$$

For $n \geq 2$, rewrite the global balance (108) as

$$\begin{aligned}\pi(n-1) &= \sum_{i=0}^{n} \pi(i)p(i, n-1) \\ &= \pi(0)p(0, n-1) + \sum_{i=1}^{n-2} \pi(i)p(i, n-1) + \pi(n-1) \\ &\quad \cdot p(n-1, n-1) + \pi(n)p(n, n-1),\end{aligned} \tag{113}$$

and let $\pi(i) = C(i)\pi(0)$ for $i = 0, \cdots, n-1$. Then,

$$\begin{aligned}\pi(n) &= \frac{\left[(1 - p(n-1, n-1))\pi(n-1) - p(0, n-1)\pi(0) - \sum_{i=1}^{n-2} p(i, n-1)\pi(i)\right]}{p(n, n-1)}, \\ &= C(n)\pi(0).\end{aligned} \tag{114}$$

This completes the proof of the theorem.

Note that we did not use the actual transition probabilities in the theorem. This will be done in the following corollary. We shall need the stationary distribution function of the EAS scheduling rule to use in the analysis of waiting time distribution as follows. The one-step probability transitions for the EAS rule are given by

$$p(0, i) = \begin{cases} \alpha_i + \alpha_{i+1}\beta; & i = 0, \\ \alpha_i(1 - \beta) + \alpha_{i+1}\beta; & i = 1, \cdots. \end{cases} \tag{115}$$

The remaining transition probabilities are unchanged from the previous case. The following corollary gives the details of the solution procedure using the system parameters.

Corollary 29. *The stationary distribution* $\{\pi(n), n = 0, 1, \cdots,\}$ *is given by*

$$\pi(n) = \frac{C(n)}{\sum_{i=0}^{N} C(i)}; \quad n = 0, 1, \cdots, N, \tag{116}$$

where $C(n), n = 0, \cdots, N$ *are obtained recursively such that* $C(0) = 1$,

$$C(1) = \begin{cases} \dfrac{(1 - \alpha_0)}{\alpha_0 \beta}, & \text{for LAS} - \text{IA and LA} - \text{DF}, \\ \dfrac{(1 - \alpha_0 - \alpha_1\beta)}{\alpha_0\beta}, & \text{for EAS}, \end{cases} \tag{117}$$

and for $n \geq 2$

$$C(n) = \frac{\left[\left(1 - \alpha_0(1 - \beta) - \alpha_1\beta\right)C(n-1) - \alpha_{n-1} - \sum_{i=1}^{n-2}\left(\alpha_{n-i-1}(1 - \beta) - \alpha_{n-i}\beta\right)C(i)\right]}{\alpha_0\beta}. \tag{118}$$

This theorem and its corollary lead to the following Algorithm 3 to calculate the stationary probabilities.

Other stopping rules can be used in step 3.

7.3.1. Waiting Time Distribution Function. In the following theorem, we give an elementary direct proof for the waiting time distribution. Let r_j be the probability that a randomly

(1) Choose a value for N, and let $C(0) = 1$ and
(2) Compute $C(1) = (1-\alpha_0)/\alpha_0\beta$ for LAS-IA and LA-DF, or $C(1) = (1-\alpha_0-\alpha_1\beta)/\alpha_0\beta$ for EAS.
(3) For $n = 2, \cdots, N$, compute $C(n)$ using (118).
(4) If $|\sum_{i=0}^{N} C(i) - \sum_{i=0}^{N-1} C(i)| < \epsilon$; stop; otherwise increment N and repeat steps 2 and 3.
(5) For $n = 0, \cdots, N$, set $\pi(n) = C(n)/\sum_{i=0}^{N} C(i)$.

Algorithm 3: Stationary probabilities for the batch arrival model.

selected customer in an arriving batch is in position j, $j = 1, \cdots$. Then, see Cox [35], p. 61.,

$$r_j = \frac{\sum_{i=j}^{\infty} \alpha_i}{\sum_{i=1}^{\infty} i\alpha_i}. \tag{119}$$

See also Burke [36] who uses this result to obtain a waiting time distribution function for batch arrival continuous time queues. Since waiting times are independent of the scheduling rule, we select the EAS rule. For the EAS rule, it can be shown, using Theorem 30, that $\pi^A(n) = \pi(n)$ for all $n = 0, \cdots$.

Theorem 30. *Let $w_q(j) = p(T_q = j)$, $j = 0, 1, \cdots$ be the probability mass function of the time in queue. Then,*

$$w_q(j) = \begin{cases} \displaystyle\sum_{n=1}^{j}\sum_{k=0}^{n}\binom{j-1}{n-1}\beta^n(1-\beta)^{j-n}\cdot\pi(k)r_{n+1-k}, & j = 1, 2, \cdots \\ \pi(0)\cdot r_1, & j = 0. \end{cases} \tag{120}$$

Proof. We assume EAS rule and note that the first few steps of proof are similar to the same steps in the proof of Theorem 23. For $j = 1, 2, \cdots$, we have

$$\begin{aligned} w_q(j) &= p\left(T_q = j\right) \\ &= \sum_{n=1}^{j} p\,(n \text{ service completions in } j \text{ time units} \mid \text{arrival is in position } n+1 \text{ to receive service}) \\ &\quad \cdot \sum_{k=0}^{n} \pi^A(k) r_{n+1-k} \\ &= \sum_{n=1}^{j}\binom{j-1}{n-1}\beta^{n-1}(1-\beta)^{j-n}\cdot\beta\cdot\sum_{k=0}^{n}\pi^A(k)r_{n+1-k}. \end{aligned} \tag{121}$$

Note that in the last step, $n-1$ of the departures occur in $j-1$ time units and the last departure occurs at time j. Simplify to obtain the desired result.

The direct nontransform approach in Theorem 30 is similar to the approach used by Chaudhry et al. [37] to compute the waiting time distribution in a batch arrival multiserver system. This approach is later utilized in Chaudhry et al. [38].

7.4. Numerical Results. In this subsection, we give numerical results for the models discussed earlier. The three algorithms described in this section for the multiserver, the finite population and the batch arrivals models are programmed using the Python programming language. The programs were verified carefully to make sure the results are accurate including running the code for special cases with known results. The algorithms and the Python programs provide an easy-to-follow approach to obtain quick numerical results for the stationary distribution functions of the three models discussed in this section.

For the multiserver and finite population models, we use $N = 10$. Moreover, for the multiserver model, we use $c = 4$, $\alpha = 0.9$, and $\beta = 0.3$; and for the finite population model, we use $\alpha = 0.15$ and $\beta = 0.90$. For the batch arrival models, we use $\beta = 0.90$, $\alpha_0 = 0.5$, $\alpha_1 = 0.3$, $\alpha_2 = 0.1$, $\alpha_3 = 0.1$, and 0 otherwise. We iterate on N until the stopping criterion is satisfied with $\epsilon = 10^{-8}$. For the batch arrival LAS-IA and LA-DF cases, we iterated until $N = 106$. For the batch arrival EAS model, we needed $N = 101$. For all models, we report the probabilities $\{p(n), n = 0, \cdots 10\}$. We also give $L = \sum_{n=0}^{N} np(n)$ and $L_q = \sum_{n=c}^{N}(n-c)p(n)$. In batch arrival cases, we, additionally, report $P(X \geq 11)$. The results are given in Table 2.

Table 2: Numerical results for the multiserver, finite population, and batch arrival models.

Model n	Multi-server model $p(n)$	Finite source model $p(n)$	Batch arrival LA models $p(n)$	Batch arrival EAS model $p(n)$
0	0.0018	0.0130	0.1111	0.2222
1	0.0248	0.0501	0.1235	0.1136
2	0.0984	0.1136	0.1125	0.1124
3	0.1776	0.1872	0.1124	0.0901
4	0.1846	0.2299	0.0877	0.0761
5	0.1439	0.2059	0.0748	0.0634
6	0.1121	0.1297	0.0621	0.0530
7	0.0874	0.0545	0.0520	0.0443
8	0.0682	0.0140	0.0434	0.0370
9	0.0523	0.0019	0.0363	0.0309
10	0.0490	0.0001	0.0303	0.0258
$P(X \geq 11)$	—	—	0.1540	0.1312
L	5.0026	4.0781	5.6000	4.8000
Lq	1.4584	3.0911	4.7111	4.0222

In this section, we used a recursive approach to compute numerically the stationary probabilities for three models. One can use these probabilities to compute other performance measures.

8. Concluding remarks

In this article, we use a unified approach that combines direct sample path and stochastic techniques and avoids generating function methods. We start with a general birth-death process that covers all single-server, single-arrival Markovian queueing models covered in this article. We present in one unified space results for all the five most used scheduling rules in the literature, allowing readers to compare these models. We address BASTA issues and note that when considered within the context of a discrete-time process with an imbedded point (arrival) process, ASTA holds at a great level of generality. However, we observe, as others, that BASTA does not hold in the classical sense that prearrival probabilities are equal to the random observer probabilities. Nonetheless, we provide formulas for the prearrival probabilities for all five scheduling rules.

One interesting observation from our summary in Subsection 5.2 is that when considering stationary distributions at random observation epochs, outside observation epochs, and at prearrival epochs, we conclude that no two scheduling rules lead to identical results. Each of the five scheduling rules has merits of its own. One could consider scheduling rules with potential arrivals and departures occurring at the beginning of a time slot, but these do not make sense from the point of view of managing arrivals/departures in a slot, and they have not been considered in the literature to the best of our knowledge.

We also consider waiting time distribution functions. Our results match those of Hunter [2] for three of the scheduling rules covered by Hunter and extend those results to the other two, namely, the LA-DA and LA-AF rules. Specifically, all rules lead to the same waiting time distribution function. This is interesting and worth investigating for systems with general arrival processes and general service times. This also has implications for Little's law and its applicability in discrete-time queues. Finally, we give three Markovian models where the results of two of them (the multiserver and the finite source models) do not fit the birth-death equations contrary to their continuous time counterparts.

The article should be accessible to researchers, engineers, and graduate students interested in learning the basic elements of discrete-time queues. The only requirement is a basic knowledge of probability and stochastic processes. This article should also be useful for people familiar with continuous-time systems and would like an accessible introduction to discrete time-queues, yet the article addresses significant issues in discrete-time queues.

Data Availability

No data were used to support the findings of this study.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] T. Meisling, "Discrete-time queuing theory," *Operations Research*, vol. 6, no. 1, pp. 96–105, 1958.

[2] J. J. Hunter, *Mathematical Techniques of Applied Probability, Volume II: Discrete-Time Models: Tecniques and Applications*, Academic Press, New York, NY, USA, 1983.

[3] T. G. Robertazzi, *Computer Networks and Systems: Queueing Theory and Performance Evaluation*, Springer-Verlag, New York, NY, USA, 3rd edition, 2000.

[4] M. El-Taha, S. Stidham, and R. Anand, "Sample-path insensitivity of symmetric queues in discrete-time," *Nonlinear Analysis: Theory, Methods and Applications*, vol. 30, pp. 1099–1110, 1997.

[5] H. Daduna, *Queueing Networks with Discrete Time Scale*, Springer-Verlag, Berlin, Germany, 2001.

[6] B. Desert and H. Daduna, "Discrete time tandem networks of queues: effects of different regulation schemes for simultaneous events," *Performance Evaluation*, vol. 47, no. 2-3, pp. 73–104, 2002.

[7] A. S. Alfa, *Applied Discrete-Time Queues*, Springer-Verlag, New York, NY, USA, 2nd edition, 2016.

[8] A. S. Alfa, *Queueing Theory for Telecommunications: Discrete Time Modeling of a Single Node System*, Springer-Verlag, Berlin, Germany, 2010.

[9] H. Bruneel and B. G. Kim, *Discrete-time Models for Communications Systems Including ATM*, Kluwer Academic, Boston, MA, USA, 1993.

[10] M. E. Woodward, *Communications and computer networks: modeling with discrete-time queues*, IEEE Computer Society Press, Los Alamitos, CA, USA, 1994.

[11] S. Halfin, "Batch delays versus customer delays," *Bell System Technical Journal*, vol. 62, no. 7, pp. 2011–2015, 1983.

[12] A. Gravey and G. Hebuterne, "Simultaneity in discrete-time single server queues with Bernoulli inputs," *Performance Evaluation*, vol. 14, no. 2, pp. 123–131, 1992.

[13] H. Daduna, "Alternating birth-death processes," 2020, https://arxiv.org/abs/2004.08816.

[14] T. Ozawa, "Exact asymptotics of the stationary tail probabilities in an arbitrary direction in a two-dimensional discrete-time qbd process," 2023, https://arxiv.org/abs/2301.02434.

[15] T. Ozawa and M. Kobayashi, "Exact asymptotic formulae of the stationary distribution of a discrete-time two-dimensional QBD process," *Queueing Systems*, vol. 90, no. 3-4, pp. 351–403, 2018.

[16] L. Fernández and M. D. de la Iglesia, "Quasi-birth-and-death processes and multivariate orthogonal polynomials," *Journal of Mathematical Analysis and Applications*, vol. 499, no. 1, Article ID 125029, 2021.

[17] R. Sasaki, "Exactly solvable discrete time birth and death processes," *Journal of Mathematical Physics*, vol. 63, no. 6, 2022.

[18] R. B. Lenin and P. R. Parthasarathy, "A birth-death process suggested by a chain sequence," *Computers and Mathematics with Applications*, vol. 40, no. 2-3, pp. 239–247, 2000.

[19] H. Daduna and R. Schassberger, "A discrete-time round-robin queue with Bernoulli input and general arithmetic service time distributions," *Acta Informatica*, vol. 15, no. 3, pp. 251–263, 1981.

[20] M. L. Chaudhry, "On numerical computations of some discrete-time queues," in *Computational Probability*, W. K. Grassmann, Ed., pp. 365–408, Springer, Boca Raton, FL, USA, 2000.

[21] M. L. Chaudhry, U. C. Gupta, and J. G. C. Templeton, "On the relations among the distributions at different epochs for discrete-time GI/Geom/1 queues," *Operations Research Letters*, vol. 18, no. 5, pp. 247–255, 1996.

[22] G. R. Dattatreya and L. N. Singh, "Relationships among different models for discrete-time queues," *WSEAS Transactions on Systems*, vol. 4, pp. 1183–1190, 2005.

[23] J. R. Louvion, P. Boyer, and A. Gravey, "A discrete-time single server queue with Bernoulli arrivals and constant service time," *Proceedings of the International Teletraffic Conference (ITC)*, vol. 12, pp. 1–8, 1988.

[24] R. Schassberger, "A definition of discrete product form distributions," *Zeitschrift für Operations Research*, vol. 23, no. 5, pp. 189–195, 1979.

[25] W. Henderson and P. G. Taylor, "Insensitivity in discrete time queues with a moving server," *Queueing Systems*, vol. 11, no. 3, pp. 273–297, 1992.

[26] M. F. Neuts, *Matrix-Geometric Solutions in Stochastic Models-An Algorithmic Approach*, Johns Hopkins University Press, Baltimore, MD, USA, 1981.

[27] D. Gross, J. F. Shortle, J. M. Thompson, and C. Harris, *Fundamentals of Queueing Theory*, John Wiley, New Jersey, NJ, USA, 4th edition, 2008.

[28] M. El-Taha and S. Stidham, *Sample-Path Analysis of Queueing Systems*, Kluwer Academic Publishing, Boston, MA, USA, 1999.

[29] M. El-Taha, "A discrete workload conservation law with applications to discrete-time queueing systems," *Queueing Models and Service Management*, vol. 2, pp. 124–152, 2019.

[30] A. Makowski, B. Melamed, and W. Whitt, "On averages seen by arrivals in discrete-time," in *Proceedings of the IEEE Conference on Decision and Control. Conference Proceedings*, pp. 1084–1086, Tampa, FL, USA, December 1989.

[31] M. El-Taha and S. Stidham, "Filtration of ASTA: a weak convergence approach," *Journal of Statistical Planning and Inference*, vol. 100, no. 2, pp. 171–183, 2002.

[32] R. B. Cooper, *Introduction to Queueing Theory*, North Holland, Amsterdam, Netherlands, 2nd edition, 1981.

[33] H. Takagi, "Queueing analysis, A foundation of performance evaluation," *Discrete-Time Systems*, Vol. 3, North Holland, Amsterdam, Netherlands, 1993.

[34] P. Gao, S. Wittevrongel, and H. Bruneel, "Discrete-time multiserver queues with geometric service times," *Computers and Operations Research*, vol. 31, no. 1, pp. 81–99, 2004.

[35] D. R. Cox, *Renewal Theory*, Chapman and Hall, London, UK, 1962.

[36] P. J. Burke, "Technical note—delays in single-server queues with batch input," *Operations Research*, vol. 23, no. 4, pp. 830–833, 1975.

[37] M. L. Chaudhry, U. C. Gupta, and V. Goswami, "Modeling and analysis of discrete-time multiserver queues with batch arrivals: GI^X/Geom/mX/Geom/m," *INFORMS Journal on Computing*, vol. 13, no. 3, pp. 172–180, 2001.

[38] M. L. Chaudhry, J. J. Kim, and A. D. Banik, "Analytically simple and computationally efficient results for the $GI^X/Geo/c$ queuesX/Geo/cqueues," *Journal of Probability and Statistics*, vol. 2019, Article ID 6480139, 18 pages, 2019.