

Effect of Parameters on $\text{Geo}^a/\text{Geo}^b/1$ Queues: Theoretical Analysis and Simulation Results

A. Lorente^{1*}, M. S. Sánchez²

¹Departamento de Matemáticas y Computación, Universidad de Burgos, Escuela Politécnica Superior, Avda. Cantabria s/n, 09006 Burgos, Spain

²Departamento de Matemáticas y Computación, Universidad de Burgos, Facultad de Ciencias, Pza. Misael Bañuelos s/n, 09001 Burgos, Spain

Email: *alorente@ubu.es, ssanchez@ubu.es

How to cite this paper: Lorente, A. and Sánchez, M.S. (2018) Effect of Parameters on $\text{Geo}^a/\text{Geo}^b/1$ Queues: Theoretical Analysis and Simulation Results. *Applied Mathematics*, 9, 153-170.

<https://doi.org/10.4236/am.2018.92011>

Received: December 14, 2017

Accepted: February 25, 2018

Published: February 28, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This paper analyzes a discrete-time $\text{Geo}^a/\text{Geo}^b/1$ queuing system with batch arrivals of fixed size a , and batch services of fixed size b . Both arrivals and services occur randomly following a geometric distribution. The steady-state queue length distribution is obtained as the solution of a system of difference equations. Necessary and sufficient conditions are given for the system to be stationary. Besides, the uniqueness of the root of the characteristic polynomial in the interval $(0, 1)$ is proven which is the only root needed for the computation of the theoretical solution with the proposed procedure. The theoretical results are compared with the ones observed in some simulations of the queuing system under different sets of parameters. The agreement of the results encourages the use of simulation for more complex systems. Finally, we explore the effect of parameters on the mean length of the queue as well as on the mean waiting time.

Keywords

Discrete-Time Queuing System, Batch Arrivals, Batch Services, Stationary Systems

1. Introduction

Queuing systems are being studied from the beginning of the XX century, trying to model fluctuations arising in a queue, such as number of clients, waiting time, service time, etc. In fact, its origin can be found in the study of congestion problems on telephone networks. The first work on these problems was published in 1909 by Agner Krarup Erlang [1] who demonstrated that telephone calls were

well modeled by a Poisson distribution.

Besides the queues in the telephone exchanges that were the origin of the theory of queues, nowadays, batch queues are common in transports, distribution and telecommunications. Consequently, the words “customers” or “clients” are used in general terms, whether it is about people or, say, printer jobs waiting their turn. Chaudhry and Templeton [2] offer a complete analysis of queuing systems.

In particular, models in discrete-time have received much attention in the last years, mostly focusing on the calculus of the length queue distribution and waiting time. In pursuing these goals, different approaches are used and with different models that arise when considering varying characteristics such as the service policy, or distribution of arrivals or of inter-arrivals time.

The present paper focuses on discrete time queuing systems for batch arrivals and services, both following a geometric distribution and both with fixed size, though different. Following the notation introduced by Kendall [3], this is denoted as $\text{Geo}^a/\text{Geo}^b/1$.

A general discrete time $\text{GI}^X/\text{G}^Y/1$ model was studied by Alfa and He [4] by the Markov chain associated to the number of customers in the queue. Chaudhry and Gupta [5] compute the length queue distribution and the waiting time through the PGF (Probability Generating Function) of the distributions involved for the $\text{GI}^X/\text{Geo}/1$ model, and more recently, [6], they focus on the waiting time for the $\text{GI}^X/\text{G}/1$ model. In [7], Chaudhry and Kim present a multi-server model with deterministic service times.

In [8], Borkakaty *et al.* focus on the busy period of the system $\text{GI}^a/\text{G}^b/1$, with fixed sizes for arrival and service batches by lattice path approach. The method is first applied for models in discrete time, handling then the model in continuous time as the limiting case.

For arrivals of variable size following a Poisson distribution, Claeys *et al.* [9] analyze the model $\text{M}^X/\text{GI}^c/1$, where the server remains idle until there is at least I customers waiting and can serve at most c customers simultaneously. With geometric distribution for arrivals and services, in [10], they present the model $\text{Geo}^X/\text{Geo}^c/1$ under two different policies: in the first, the server starts a new service although the number of clients does not reach c ; in the second, the server waits until the number of clients reaches c .

A related model, the $\text{GI}/\text{Geo}/1$ queue under N-policy, is analyzed by Lim *et al.* [11]. In this model, customers arrive and are served one by one, but the server waits until the number of customers reaches N before starting services.

Bruneel *et al.* [12] examine the model with fixed service capacity, while in [13] they consider systems with variable service capacity. Focusing in the interval between arrivals (keeping it constant) and constant time of service, Zhao and Campbell [14] use the roots, both inside and outside the unit circle, of the denominator of the generating function for the system $\text{D}^X/\text{D}^m/1$. Related to this methodology, Janssen and Leeuwaarden [15] discuss about the method of finding the roots of $A(z) - z^s$ (where $A(z)$ is the PGF of the number of arriving customers in each slot time), both in order to clarify the probabilistic interpreta-

tion of the roots and to find them numerically. They present a solution for the stationary queue length distribution that does not depend on roots.

On the other hand, Yao *et al.* [16] consider computer simulation for discrete models as a method to avoid the difficulties of theoretical analysis.

The present paper shows how the specific characteristics of the model being studied make it possible to obtain the theoretical solutions without the need of finding all the roots of $A(z) - z^s$. In particular, we prove the existence of a unique root of the characteristic polynomial in $(0, 1)$, which suffices to obtain the solution, provided the system is stationary. Conditions for the system to be stationary are also given.

2. Analytic Solution

2.1. Model Description

The model under study is a discrete-time model denoted as $\text{Geo}^a/\text{Geo}^b/1$: customers arrive in batches of fixed size a , and they are served in batches of fixed size b . If there is not at least b customers at the end of a service, the server does not start a new service until the number of waiting customers reaches b .

It is assumed that arrivals are independent, service times are independent, and services are independent of arrivals unless the number of waiting customers were less than b . Batch sizes a and b are supposed to be relative prime integers; otherwise, we would take their greatest common divisor as unit.

At each time slot, an arrival of a customers can happen with probability α , or not, with probability $1 - \alpha$. Furthermore, if there is a service running, it can finish with probability β , or not, with probability $1 - \beta$. Besides, inter-arrival times are geometrically distributed with parameter α , and service times are geometrically distributed with parameter β . Then, traffic density is $\rho = a\alpha/(b\beta)$.

Therefore, if there are n customers in the system, Equations (1) to (4) describe the different probabilities.

$$P(\bar{A}\bar{E} | n) = P(\text{no arrival and no exit} | n) = \begin{cases} 1 - \alpha & \text{if } n < b \\ (1 - \alpha)(1 - \beta) & \text{if } n \geq b \end{cases} \quad (1)$$

$$P(A\bar{E} | n) = P(\text{arrival and no exit} | n) = \begin{cases} \alpha & \text{if } n < b \\ \alpha(1 - \beta) & \text{if } n \geq b \end{cases} \quad (2)$$

$$P(\bar{A}E | n) = P(\text{no arrival and exit} | n) = \begin{cases} 0 & \text{if } n < b \\ (1 - \alpha)\beta & \text{if } n \geq b \end{cases} \quad (3)$$

$$P(AE | n) = P(\text{arrival and exit} | n) = \begin{cases} 0 & \text{if } n < b \\ \alpha\beta & \text{if } n \geq b \end{cases} \quad (4)$$

If the system reaches an equilibrium, and q_n denotes the probability of “ n customers in the system”, then the following Equation (5) holds,

$$q_n = P(\bar{A}\bar{E} | n)q_n + P(A\bar{E} | n - a)q_{n-a} + P(\bar{A}E | n + b)q_{n+b} + P(AE | n - a + b)q_{n-a+b} \quad (5)$$

where all q_m with $m < 0$ are 0.

Note that, according to Equations (1)-(4), the probabilities of *exit* or *no exit* are different depending on whether n is greater than b or not. Consequently, all equations with any subscript less than b are *particular equations*.

Since the lowest subscript in Equation (5) is $n - a$, then there are $a + b$ particular equations, from $n = 0$ to $n = a + b - 1$, the ones detailed in Equations (6)-(11), that also should take into account the relation between a and b .

If $a < b$ the particular equations are:

$$q_n = (1 - \alpha)q_n + (1 - \alpha)\beta q_{n+b} \quad \text{if } 0 \leq n < a \quad (6)$$

$$q_n = (1 - \alpha)q_n + \alpha q_{n-a} + (1 - \alpha)\beta q_{n+b} + \alpha\beta q_{n-a+b} \quad \text{if } a \leq n < b \quad (7)$$

$$q_n = (1 - \alpha)(1 - \beta)q_n + \alpha q_{n-a} + (1 - \alpha)\beta q_{n+b} + \alpha\beta q_{n-a+b} \quad \text{if } b \leq n < a + b \quad (8)$$

While, if $a > b$, the equations become:

$$q_n = (1 - \alpha)q_n + (1 - \alpha)\beta q_{n+b} \quad \text{if } 0 \leq n < b \quad (9)$$

$$q_n = (1 - \alpha)(1 - \beta)q_n + \alpha q_{n-a} + (1 - \alpha)\beta q_{n+b} \quad \text{if } b \leq n < a \quad (10)$$

$$q_n = (1 - \alpha)(1 - \beta)q_n + \alpha q_{n-a} + (1 - \alpha)\beta q_{n+b} + \alpha\beta q_{n-a+b} \quad \text{if } a \leq n < a + b \quad (11)$$

Remember that a and b are supposed to be relative prime integers. In any case, if $a = b$, the problem would reduce to $a = b = 1$, since each block can be treated as a unit.

The general equation for $n \geq a + b$ in Equation (12) is obtained by replacing Equations (1)-(4) into Equation (5).

$$q_n = (1 - \alpha)(1 - \beta)q_n + \alpha(1 - \beta)q_{n-a} + (1 - \alpha)\beta q_{n+b} + \alpha\beta q_{n-a+b} \quad (12)$$

2.2. Steady-State Equations

Rearranging Equation (12), for $n \geq a + b$ we have

$$\alpha(1 - \beta)q_{n-a} + [(1 - \alpha)(1 - \beta) - 1]q_n + \alpha\beta q_{n-a+b} + (1 - \alpha)\beta q_{n+b} = 0 \quad (13)$$

whose characteristic polynomial is

$$P(x) = \alpha(1 - \beta) + [(1 - \alpha)(1 - \beta) - 1]x^a + \alpha\beta x^b + (1 - \alpha)\beta x^{a+b} \quad (14)$$

The steady-state distribution, if it exists, must be a particular solution of Equation (13). The general solution of Equation (13) is

$$q_n = \sum_k a_k x_k^n \quad (n \geq a + b) \quad (15)$$

where x_k are the roots of the characteristic polynomial in Equation (14).

Then, with the aim of studying the roots of Equation (14), we distinguish two cases:

Case $a < b$:

If $a < b$, exponents in Equation (14) are already in ascending order. As a polynomial, $P(x)$ is continuous and differentiable in \mathbb{R} . Its derivative:

$$\begin{aligned}
P'(x) &= a[(1-\alpha)(1-\beta)-1]x^{a-1} + b\alpha\beta x^{b-1} + (a+b)(1-\alpha)\beta x^{a+b-1} \\
&= x^{a-1}(-a[1-(1-\alpha)(1-\beta)] + b\alpha\beta x^{b-a} + (a+b)(1-\alpha)\beta x^b) \quad (16) \\
&= x^{a-1}Q(x)
\end{aligned}$$

Since a, b are positive integers and α, β lay in $(0,1)$, $Q(x)$ is an increasing polynomial and $Q(0) < 0$. Consequently, $Q(x)$ has a unique root in $(0, \infty)$, and so has $P'(x)$. Therefore $P(x)$ has at most two roots in $(0, \infty)$.

Case $a > b$:

If $a > b$, Equation (17) shows $P(x)$ with exponents in ascending order

$$P(x) = \alpha(1-\beta) + \alpha\beta x^b + [(1-\alpha)(1-\beta)-1]x^a + (1-\alpha)\beta x^{a+b} \quad (17)$$

Its derivative:

$$\begin{aligned}
P'(x) &= b\alpha\beta x^{b-1} + a[(1-\alpha)(1-\beta)-1]x^{a-1} + (a+b)(1-\alpha)\beta x^{a+b-1} \\
&= x^{b-1}(b\alpha\beta - a[1-(1-\alpha)(1-\beta)]x^{a-b} + (a+b)(1-\alpha)\beta x^a) \quad (18) \\
&= x^{b-1}S(x)
\end{aligned}$$

By differentiating again the polynomial $S(x)$ in Equation (18) is

$$\begin{aligned}
S'(x) &= -a(a-b)[1-(1-\alpha)(1-\beta)]x^{a-b-1} + a(a+b)(1-\alpha)\beta x^{a-1} \\
&= x^{a-b-1}(-a(a-b)[1-(1-\alpha)(1-\beta)] + a(a+b)(1-\alpha)\beta x^b) \quad (19)
\end{aligned}$$

Like in the previous case, the second factor in Equation (19) is an increasing polynomial in the interval $(0, \infty)$ and negative at 0. Then $S'(x)$ has a unique root in $(0, \infty)$. This implies that $S(x)$ has at most two roots in $(0, \infty)$, and so has $P'(x)$.

As $S'(x)$ must be negative until its unique root in $(0, \infty)$ and positive after the root, $S(x)$ is at first decreasing, it reaches a relative minimum and turns to be an increasing function. Because $S(0) > 0$, either $S(x)$ remains above the horizontal axis in $(0, \infty)$ (maybe reaching it at the relative minimum), or $S(x)$ crosses the horizontal axis at two different points. However, the former case would imply, Equation (18), that $P'(x)$ is never negative and therefore $P(x)$ is never decreasing, which is impossible because $P(0) > 0$ and $P(1) = 0$.

We conclude that $S(x)$, and thus $P'(x)$, crosses the horizontal axis at two different points s_1 and s_2 ($s_1 < s_2$) in $(0, \infty)$ with $P'(x) > 0$ if $x \in (0, s_1)$, $P'(x) < 0$ if $x \in (s_1, s_2)$ and again $P'(x) > 0$ if $x > s_2$. Accordingly, $P(x)$ is increasing in $(0, s_1) \cup (s_2, \infty)$ and decreasing in (s_1, s_2) . In the first interval $(0, s_1)$ there are no roots of $P(x)$, since $P(0) > 0$. Consequently, as the only remaining root of $P'(x)$ is in (s_1, ∞) , in this interval there is at most two roots of $P(x)$.

This proves that, in both cases ($a < b$ and $a > b$), the characteristic polynomial has at most two roots in the interval $(0, \infty)$. One of these roots is 1. Let's suppose that there is another root, r , in $(0,1)$. Then $P(x)$ must be decreasing up to r and then increasing until 1. In particular, $P'(1)$ must be positive but

$$\begin{aligned}
P'(1) &= a[(1-\alpha)(1-\beta)-1] + b\alpha\beta + (a+b)(1-\alpha)\beta \\
&= a - a\alpha - a\beta + a\alpha\beta - a + b\alpha\beta + a\beta - a\alpha\beta + b\beta - b\alpha\beta \\
&= -a\alpha + b\beta
\end{aligned} \tag{20}$$

Thus, $P'(1) > 0 \Leftrightarrow a\alpha < b\beta$.

This implies that if $a\alpha \geq b\beta$ there are no roots of $P(x)$ in $(0,1)$, then we cannot construct any particular solution as defined in Equation (15), which should generate a probability distribution.

On the other hand, if $a\alpha < b\beta$, the root of $P(x)$ in $(0,1)$ is unique, let x_0 be this root. Therefore, for any $u > 0$, the series in Equation (21)

$$\sum_{n=a+b}^{\infty} ux_0^n \tag{21}$$

is a convergent series with positive terms.

We can choose $u > 0$ such that

$$\sum_{n=0}^{\infty} q_n = \sum_{n=0}^{a+b-1} q_n + \sum_{n=a+b}^{\infty} ux_0^n = 1 \tag{22}$$

that is,

$$u = \left(1 - \sum_{n=0}^{a+b-1} q_n\right) \frac{1-x_0}{x_0^{a+b}} \tag{23}$$

Inserting Equation (23) into Equation (21), the probability of n customers in the system, q_n for $n \geq a+b$, is written as

$$q_n = \left(1 - \sum_{n=0}^{a+b-1} q_n\right) \frac{1-x_0}{x_0^{a+b}} x_0^n \tag{24}$$

depending on the first $a+b$ probabilities of the succession. We determine these initial probabilities by replacing q_n ($n \geq a+b$) in the $a+b$ particular equations and solving the resulting linear system.

This provides the distribution of the total number of customers. Equation (25) gives the rule to obtain the number of customers waiting before service

$$P(n \text{ customers waiting}) = \begin{cases} q_n + q_{n+b} & \text{if } n < b \\ q_{n+b} & \text{if } n \geq b \end{cases} \tag{25}$$

Also from these equations, the average number of customers in the system is calculated by summing up the series in the following Equation (26)

$$\sum_{n=0}^{a+b-1} nq_n + \sum_{n=a+b}^{\infty} n \left(1 - \sum_{n=0}^{a+b-1} q_n\right) \frac{1-x_0}{x_0^{a+b}} x_0^n \tag{26}$$

which gives

$$\sum_{n=0}^{a+b-1} nq_n + \left(1 - \sum_{n=0}^{a+b-1} q_n\right) \left(\frac{x_0}{1-x_0} + a + b\right) \tag{27}$$

and the average number of customers waiting is computed as:

$$\sum_{n=0}^{b-1} n(q_n + q_{n+b}) + \sum_{n=b}^{\infty} nq_n = \sum_{n=0}^{b-1} nq_n + \sum_{n=0}^{a+b-1} nq_{n+b} + \left(1 - \sum_{n=0}^{a+b-1} q_n\right) \left(\frac{x_0}{1-x_0} + a\right) \tag{28}$$

3. Numerical Results and Discussion

3.1. Theoretical Probabilities versus Relative Frequencies

The equations in the previous section can be easily programmed so they provide a straightforward procedure to compute the probabilities. To illustrate the procedure as well as comparing the probabilities against relative frequencies computed with simulated results (for the cases where the probabilities were difficult to compute), we simulated the system for some given values and for around 10,000 cycles. For the sake of simplicity, the case studies we chose were with small numbers for a and b .

For the first simulation study, fixed values of $a = 2$, $b = 3$ and $\beta = 0.5$ were chosen. By varying α , the traffic density ρ is also varied, so that, for instance, $\alpha = 0.3$ gives $\rho = 0.4$, whereas for $\alpha = 0.6$, a heavier traffic density, $\rho = 0.8$, is obtained.

In both cases, the first five probabilities $\{q_0, q_1, q_2, q_3, q_4\}$ come from the linear system obtained by substituting and rearranging Equations (6)-(8), and (12) (the details can be found in [17]), namely

$$\begin{pmatrix} -\alpha & 0 & \alpha & \frac{1-\alpha}{2} & 0 \\ 0 & -\alpha & 0 & 0 & \frac{1-\alpha}{2} \\ \alpha & 0 & -\alpha & \frac{\alpha}{2} & 0 \\ 0 & \alpha & 0 & -\frac{1+\alpha}{2} & \frac{\alpha}{2} \\ 0 & 0 & \alpha & 0 & -\frac{1+\alpha}{2} \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \\ q_4 \end{pmatrix} - \frac{1-\alpha}{2}(1-x_0) \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ x_0 & x_0 & x_0 & x_0 & x_0 \\ x_0^2 & x_0^2 & x_0^2 & x_0^2 & x_0^2 \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \\ q_4 \end{pmatrix} \\ - \frac{\alpha}{2}(1-x_0) \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \\ q_4 \end{pmatrix} = -\frac{1-\alpha}{2}(1-x_0) \begin{pmatrix} 0 \\ 0 \\ 1 \\ x_0 \\ x_0^2 \end{pmatrix} - \frac{\alpha}{2}(1-x_0) \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

where x_0 is the root of the characteristic polynomial in $(0,1)$. The remaining probabilities obey to the geometric formula in Equation (24) with $a+b=5$.

Table 1 shows the resulting probabilities with $n=1,2,\dots,30$ customers in the system, with different traffic density ρ obtained by varying the probability of incoming customers α . The probabilities have been computed for the total number of customers (first three columns) and also for the number of customers that are waiting (last three columns). The last row of **Table 1** shows the corresponding expectation.

For the total number of customers, relative frequencies obtained when simulating the systems with the same parameters are shown in **Figure 1** and **Figure 2**, where also the theoretical probabilities have been depicted for comparative purposes. We can see in both cases how relative frequencies approach the theoretical values.

Table 1. Theoretical probabilities and expectations for the system with $a = 2$, $b = 3$, $\beta = 0.5$, and different traffic density ρ .

Total number of customers	Probability		Waiting customers	Probability	
	$\alpha = 0.3$	$\alpha = 0.6$		$\alpha = 0.3$	$\alpha = 0.6$
	($\rho = 0.4$)	($\rho = 0.8$)		($\rho = 0.4$)	($\rho = 0.8$)
0	0.1520	0.0362	0	0.2823	0.1449
1	0.1742	0.0476	1	0.3235	0.1904
2	0.2789	0.1220	2	0.3320	0.2161
3	0.1303	0.1087	3	0.0286	0.0778
4	0.1493	0.1428	4	0.0155	0.0643
5	0.0530	0.0941	5	0.0084	0.0531
6	0.0286	0.0778	6	0.0045	0.0439
7	0.0155	0.0643	7	0.0024	0.0363
8	0.0084	0.0531	8	0.0013	0.0300
9	0.0045	0.0439	9	0.0007	0.0248
10	0.0024	0.0363	10	0.0004	0.0205
11	0.0013	0.0300	11	0.0002	0.0170
12	0.0007	0.0248	12	0.0001	0.0140
13	0.0004	0.0205	13	0.0001	0.0116
14	0.0002	0.0170	14	0.0000	0.0096
15	0.0001	0.0140	15	0.0000	0.0079
16	0.0001	0.0116	16	0.0000	0.0065
17	0.0000	0.0096	17	0.0000	0.0054
18	0.0000	0.0079	18	0.0000	0.0045
19	0.0000	0.0065	19	0.0000	0.0037
20	0.0000	0.0054	20	0.0000	0.0031
21	0.0000	0.0045	21	0.0000	0.0025
22	0.0000	0.0037	22	0.0000	0.0021
23	0.0000	0.0031	23	0.0000	0.0017
24	0.0000	0.0025	24	0.0000	0.0014
25	0.0000	0.0021	25	0.0000	0.0012
26	0.0000	0.0017	26	0.0000	0.0010
27	0.0000	0.0014	27	0.0000	0.0008
28	0.0000	0.0012	28	0.0000	0.0007
29	0.0000	0.0010	29	0.0000	0.0006
30	0.0000	0.0008	30	0.0000	0.0005
Expectation:	2.4319	6.4891	Expectation:	1.2473	4.1065

In the case of lighter traffic density, $\rho = 0.4$, the first $a + b = 5$ probabilities, up to 4 customers in the systems, are comparatively large in the system, which is very clear in **Figure 1** (where it is also seen that the probabilities are practically equal to the relative frequencies) and also in the second column of **Table 1** with

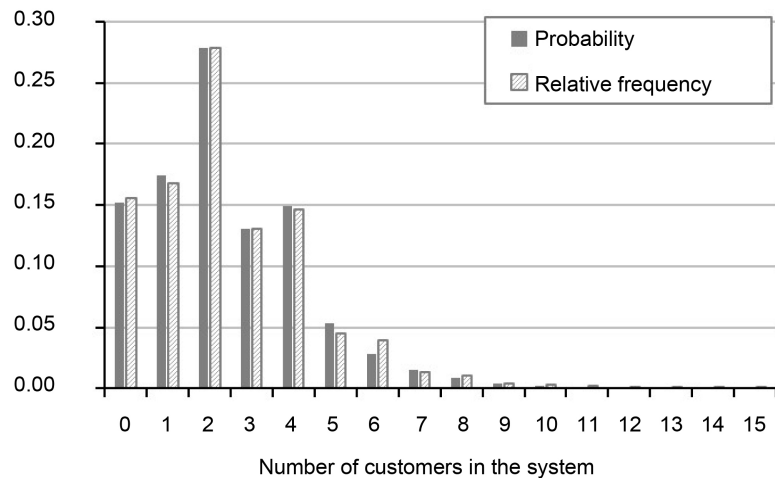


Figure 1. Results for the system with $a = 2$, $b = 3$, $\beta = 0.5$, $\alpha = 0.3$ ($\rho = 0.4$).

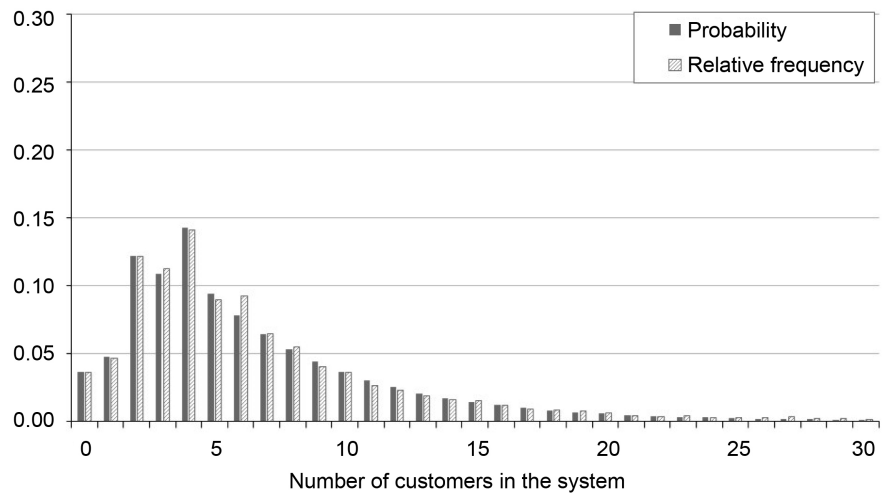


Figure 2. Results for the system for $a = 2$, $b = 3$, $\beta = 0.5$, $\alpha = 0.6$ ($\rho = 0.8$).

probabilities greater than 0.13 for less than 5 customers; while the remaining ones are very small, graphically inappreciable in **Figure 1** from around 11 customers in the system, and smaller than 10^{-4} from 17 customers (**Table 1**).

The magnitude of q_2 stands out clearly that reflects the fact that there must be at least 3 customers for the system to start, but they arrive two by two with less probability. Accordingly, the expectation is near 2 customers (2.4319 as can be seen in the last row of **Table 1**) in the system.

This same reason explains the behavior of the probabilities for the number of customers that are waiting their turn, fourth column in **Table 1**, not only the fact that the largest probability is indeed for 2 customers but that the cumulative probability of having up to 2 customers waiting is in fact 0.9378, almost the whole distribution.

Comparatively, the probabilities of the number of customers waiting for their turn are much larger than the probabilities of n customers in the system just up until 2 customers waiting. Then, the probabilities rapidly decrease and are ob-

viously less than their counterparts, and the expectation drops to 1.2473 customers.

The difference between the first five probabilities and the remaining ones is less pronounced in the second system, with $\rho = 0.8$, as we can see in both **Figure 2** and the third column of **Table 1** and also for the probabilities of n customers waiting to go to the server, last column in **Table 1**.

In **Figure 2**, again, there is not difference between the probabilities computed with Equation (24) and the relative frequencies obtained when simulating the system (except may be for 6 customers in the system), the probability reaches its maximum for 4 customers (0.1428 in **Table 1**) but then, contrary to the situation in **Figure 1**, the probabilities softly decrease and remain larger than 8×10^{-4} for up to 30 customers. The expectation increases accordingly up to near 6.5 customers in the system, on average, 4 of them waiting.

Like in the previous case, when looking at the probability of having n customers waiting for their turn, the largest probability is for 2 customers waiting, and again the largest probabilities are indeed for 0, 1 or 2 customers waiting but now the cumulative probability is much smaller, 0.5514, and slowly decreases from 3 customers waiting their turn.

In this case, larger traffic density due to the larger probability of arrival causes larger probabilities for more customers in the system, and thus for the ones waiting because the probability of service is the same.

For illustrative purposes, **Table 2** contains probabilities and relative frequencies computed also with ten thousand cycles when $a > b$, in this case, obtained for $a = 5$, $b = 3$, and probabilities $\alpha = 0.2$, $\beta = 0.4$ (thus, maintaining high traffic density $\rho = 0.83$). The range of values shown now for the number of customers is necessarily larger than in **Table 1** because the root of the characteristic polynomial is 0.9392 implying a very slow decreasing of the probabilities when increasing the number of customers for $n \geq 8 = a + b$, see Equation (24).

As in the previous case, the results for the number of total customers in the system are similar to each other, although in this case, shorter simulation times showed rather high differences between the theoretical probabilities and the observed frequencies.

The largest values are in the first three rows, up to 2 customers in the system, especially for exactly 2 customers. This is due to the needed $3 = b$ customers to start the service. Then, from 3 to $7 = a + b - 1$ customers, both the probabilities and observed frequencies increase and, at that point, they start to slowly decrease from 8 customers in the system, although the high traffic density diffuses the differences among these three block of probabilities (or frequencies). The slow decrease is apparent in the fact that there is still 0.8 in a thousand rate of having 70 customers in the system, with almost 16 customers on average, 15.72 according to **Table 2**.

3.2. Effect of the Parameters on the Queue Length and Waiting Time

With the aim of studying the effect of the parameters on the waiting time and on

Table 2. Relative frequencies observed in a simulation of 10,000 cycles and theoretical probabilities computed for $a = 5$, $b = 3$, $\alpha = 0.2$, $\beta = 0.4$.

Total number of customers	Relative frequency	Probability	Total number of customers	Relative frequency	Probability
0	0.0446	0.0470	36	0.0038	0.0064
1	0.0472	0.0544	37	0.0051	0.0061
2	0.0781	0.0729	38	0.0038	0.0057
3	0.0296	0.0294	39	0.0049	0.0053
4	0.0308	0.0340	40	0.0062	0.0050
5	0.0476	0.0455	41	0.0047	0.0047
6	0.0476	0.0477	42	0.0055	0.0044
7	0.0593	0.0553	43	0.0061	0.0042
8	0.0378	0.0373	44	0.0054	0.0039
9	0.0390	0.0350	45	0.0039	0.0037
10	0.0333	0.0329	46	0.0040	0.0034
11	0.0343	0.0309	47	0.0066	0.0032
12	0.0316	0.0290	48	0.0038	0.0030
13	0.0310	0.0273	49	0.0043	0.0029
14	0.0234	0.0256	50	0.0028	0.0027
15	0.0265	0.0240	51	0.0031	0.0025
16	0.0258	0.0226	52	0.0020	0.0024
17	0.0228	0.0212	53	0.0021	0.0022
18	0.0233	0.0199	54	0.0022	0.0021
19	0.0213	0.0187	55	0.0005	0.0020
20	0.0157	0.0176	56	0.0017	0.0018
21	0.0195	0.0165	57	0.0013	0.0017
22	0.0174	0.0155	58	0.0004	0.0016
23	0.0159	0.0146	59	0.0011	0.0015
24	0.0139	0.0137	60	0.0004	0.0014
25	0.0144	0.0128	61	0.0004	0.0013
26	0.0121	0.0121	62	0.0002	0.0013
27	0.0093	0.0113	63	0.0002	0.0012
28	0.0126	0.0106	64	0.0005	0.0011
29	0.0076	0.0100	65	0.0001	0.0010
30	0.0095	0.0094	66	0.0003	0.0010
31	0.0066	0.0088	67	0.0003	0.0009
32	0.0055	0.0083	68	0.0001	0.0009
33	0.0078	0.0078	69	0.0002	0.0008
34	0.0046	0.0073	70	0.0004	0.0008
35	0.0048	0.0069	Expectation	14.4072	15.72

the length of the queue, in the two possible situations $a < b$ and $b < a$, we maintain the size of the blocks, both entering and exiting the system (the same as in section 3.1), and modify the rate service from high ($\beta = 0.2$), medium ($\beta = 0.5$), to low ($\beta = 0.7$). Systematic simulations were conducted by varying the arrival rate α with the purpose of obtaining traffic density ρ from 0.1 to 0.9.

In all cases, Equations (27) and (28) are used to compute the average number of customers in the system and the average number of customers waiting for service. The average waiting time is obtained from simulations. From it, we also computed the percentage of this waiting time due to the insufficient number of customers in the system, *i.e.* the time spent waiting for the system to reach the minimum b customers to start a service.

The first situation corresponds to the first system studied in Section 3.1, and thus the simulations and computation are made by keeping $a = 2$ and $b = 3$. The resulting waiting time is depicted in **Figure 3** as a function of the traffic density, and for the three values of β taken into account in order to cover slow, medium and fast services, namely $\beta = 0.7$, $\beta = 0.5$ and $\beta = 0.2$, respectively.

It is observed that, for all values of β , the average waiting time is rather high for low traffic density, probably because customers must wait until the number reaches the minimum b . It decreases slowly until medium traffic density and then grows up again for high traffic density due to congestion.

As expected, irrespective of ρ , the time spent waiting decreases with β but not in the same way. The differences among mean times for each ρ are larger when comparing the slow service (open circles) to the medium one (crosses), with smaller differences between the later and the fast one (filled circles). These distances are remarkably different above all for extreme traffic densities, notice the huge increase in the mean for $\rho = 0.9$ when $\beta = 0.7$ as compared to the medium or the fast one.

This behavior is related to the one depicted in **Figure 4**, the percentage of the

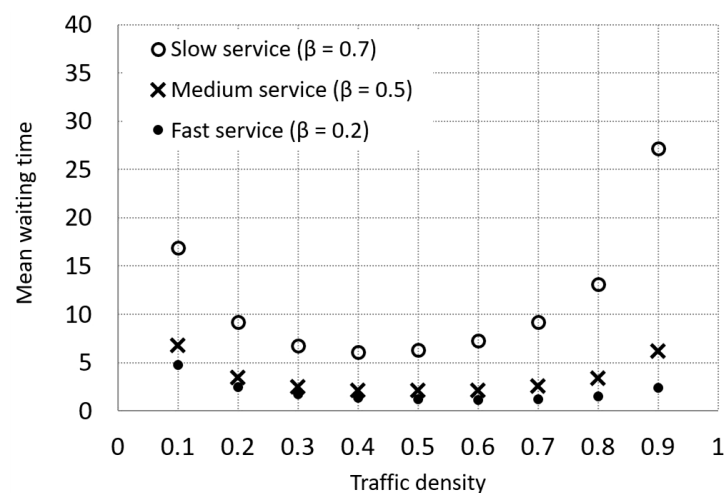


Figure 3. Mean waiting time vs. traffic density for $a = 2$, $b = 3$, and different β .

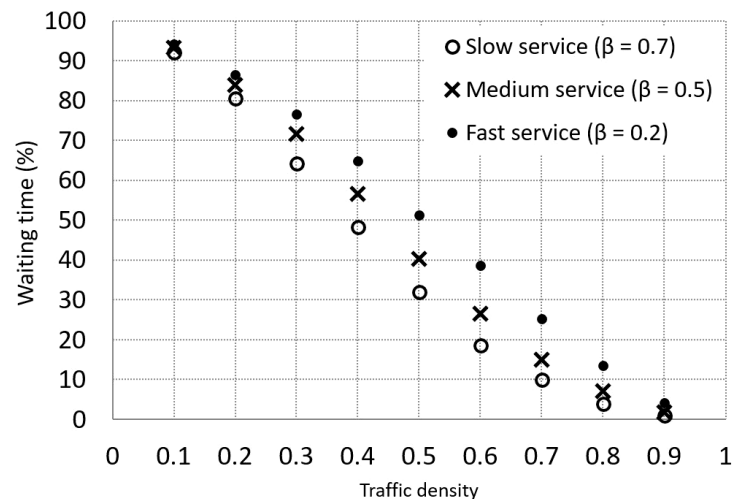


Figure 4. Percentage of waiting time with the server free vs. traffic density for $a = 2$, $b = 3$, and different β .

waiting time that is due to inactivity of the server because there are not enough customers waiting for service, *i.e.*, the server is free but the customers are waiting. At low traffic density, irrespective of the probability of the service, most of the waiting time is due to the insufficient number of customers. For instance, when $\rho = 0.1$, more than 90% of the total waiting time is spent with the server idle (thus, the increasing mean value of the total waiting time already observed in **Figure 3**). As expected, the values coincide again, practically null in this case, for $\rho = 0.9$, where the traffic is so high that more than 90% of the time that the customers spent waiting is because the server is busy (only less than 5% of the waiting time is due to insufficient customers).

Between these two extreme values, **Figure 4** shows a decreasing behavior of the percentage of the time waiting for the minimum number of customers needed to start the service, a decreasing which is almost linear for the fastest service among the cases shown. Globally, for a given traffic density, the faster the service is, the larger the percentage of the time spent waiting that is due to the fact that there are not enough customers.

In relation to the number of customers in the system, **Figure 5** shows the average values for both the total number of customers (black lines) and the number of them waiting for service (in red lines). As expected, in both cases, the mean values are ordered from smallest to largest according to the speed of the service, *i.e.*, the largest mean values are for the slow service (open circles or squares for the total number of customers or for those waiting, respectively), then the ones corresponding to the medium service rate (crosses in **Figure 5**), and the smallest mean values are for the fast service (filled circles or squares).

More interestingly, the rate service β does not seem to have a high influence on the mean values for small traffic densities, regardless of whether we look at the number of total customers or at the number of those that are waiting. In fact, all the mean values are graphically undistinguishable until approximately $\rho =$

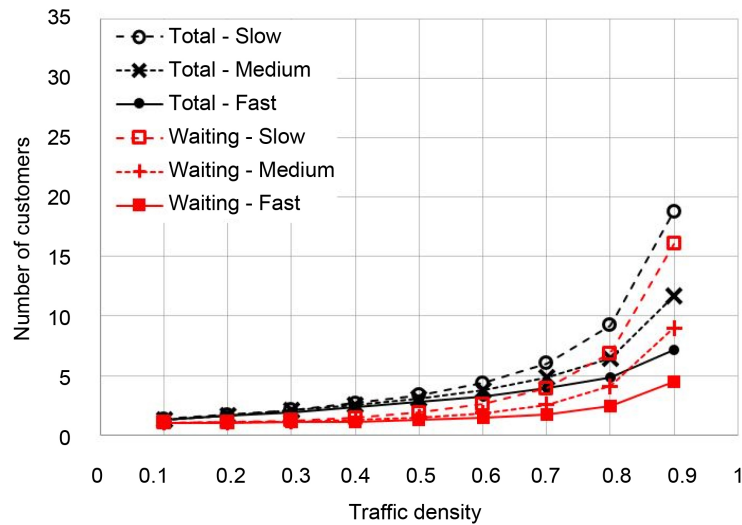


Figure 5. Average number of both total customers (dark lines) and customers waiting for service (red lines) vs. traffic density for $a = 2$, $b = 3$, and different rate services ($\beta = 0.7$ for slow, $\beta = 0.5$ for medium and $\beta = 0.2$ for fast services).

0.4. From then, the mean values continue their increase with ρ , but the differences due to β become apparent, more importantly for largest traffic densities and also with more dispersion, that is, larger differences among the averages correspond to larger values of ρ , though this dispersion is the same for the total number of customers or when we look only to the customers that are waiting.

Similar conclusions can be drawn with systems with $a = 5$ and $b = 3$ (then $a > b$, as opposed to the preceding case). **Figure 6** shows the effect of traffic density on the mean waiting time computed with the simulations conducted, as before, modifying the rate service from high ($\beta = 0.2$), medium ($\beta = 0.5$), to low ($\beta = 0.7$) and varying the arrival rate a with the purpose of obtaining traffic density ρ from 0.1 to 0.9.

The pattern in **Figure 6** is the same as the one observed in **Figure 3**, at first decreasing with the increase of ρ up to approximately 0.5 and then a faster increase of the waiting time when ρ continues to increase. However, in the present case the mean waiting time is comparatively larger in all cases, especially in the slow service, open circles in **Figure 6**.

It might seem obvious because customers arrive in groups of $a = 5$ instead of $a = 2$ but, in fact, it was not so predictable because the rate of arrival is also varying so that the product remains constant; in other words, there are more clients arriving but they do so more spaced in time to maintain traffic density. When comparing to **Figure 7** that shows the percentage of the time spent waiting because there are not enough customers to start a service, the reason becomes clearer, they are waiting for attention, because for every five arriving, at most three receive service thus there can be up to four customers waiting and still one when the next service starts. This percentage is now smaller, although, as in **Figure 4**, it is comparatively high at lower traffic densities and decreases when

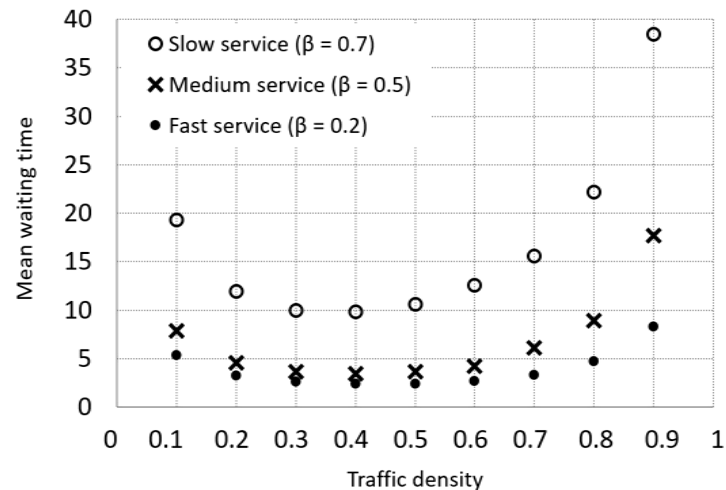


Figure 6. Mean waiting time vs. traffic density for $a = 5$, $b = 3$, and different β .

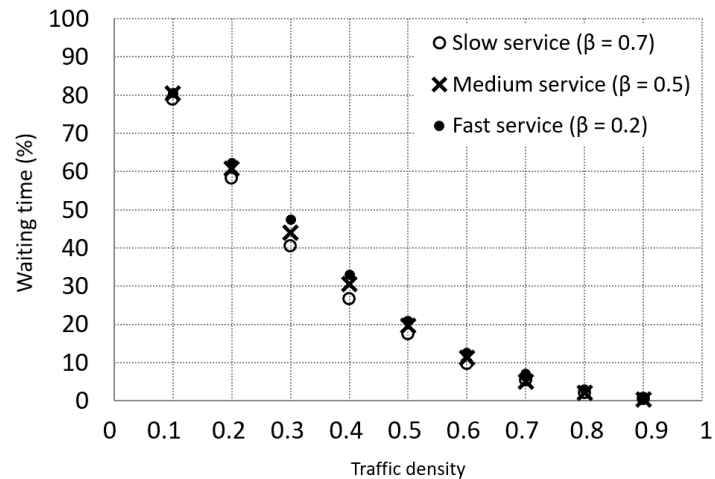


Figure 7. Percentage of the total waiting time with the server free vs. traffic density for $a = 5$, $b = 3$, and varying β .

traffic density increases (up to a point).

Comparing to the case $a < b$, the decreasing pattern for all values of β is now less linear than it appeared to be in **Figure 4** and also the differences among the percentages for the three service rates are smaller than the ones observed in **Figure 4**. This reflects the fact that with $a > b$ the server is less frequently free.

Figure 8 shows the average number of total customers (in black) and the average of the number of customers waiting for service (in red lighter lines), among those in the system. The global increasing pattern of the mean values when increasing the traffic density is the same as the one observed for $a < b$ in **Figure 5**, but now the average number of customers, both total or waiting, is always higher than it was for the analogous case.

Comparing with the preceding system, we see that the number of customers increases with ρ but more rapidly than in **Figure 5** even for low traffic densities

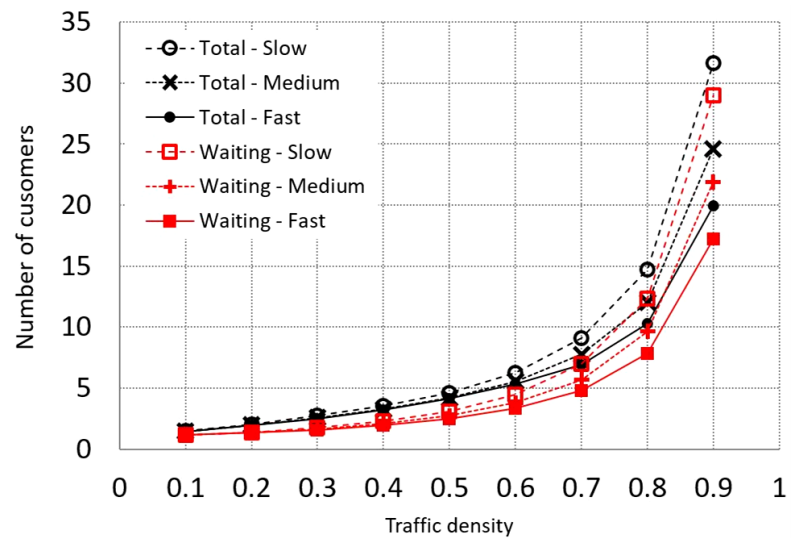


Figure 8. Average number of total customers in the system (black lines) and average number of waiting customers (red lines) vs. traffic density for $a = 5$, $b = 3$. Different service rate: slow for $\beta = 0.7$, medium $\beta = 0.5$, and fast for $\beta = 0.2$.

and, consequently, there are more customers (both total customers and the number of them that are waiting). It is remarkable the sharp increase of the mean number of customers when passing from traffic density 0.8 to 0.9.

4. Conclusions

In this paper, we present a discrete-time queuing system where both arrivals and services occur in batches of fixed size, a for arrivals and b for services, following geometric distributions with probabilities α and β , respectively. In this situation, analytical solutions are obtained by explicitly establishing the steady-state equations, and proving that the system is stationary if and only if $a\alpha < b\beta$.

The specific features of the system allow us avoiding the search of all roots of $A(z) - z^s$ when computing probabilities, since it suffices to take into account the unique real root of the characteristic polynomial in $(0, 1)$, if the stationary requirements hold.

Simulations of the system conducted under different sets of parameters show the similarity between the theoretical probabilities and the relative frequencies obtained in the numerical results. This provides a viable alternative to avoid solving the whole set of equations for each particular case, and then compute the remaining probabilities.

Finally, we explore the effect of parameters on the mean queue length and the mean waiting time. For the cases studied, on average, the waiting time is observed to be high for low traffic density, due probably to the time spent waiting for a group of b customers that starts the server. Then the time decreases as the traffic density increases up to a point where the congestion of the system gives rise again to longer waiting times.

Acknowledgements

Special thanks are due to Dr. M.^a Cruz Valsero, former professor at the University of Valladolid (Spain) for her encouragement, and useful discussions.

References

- [1] Erlang, A.K. (1909) The Theory of Probabilities and Telephone Conversations. *Nyt Tidsskrift for Matematik B*, **20**, 33.
- [2] Chaudhry, M.L. and Templeton, J.G.C. (1983) A First Course in Bulk Queues. Wiley, New York.
- [3] Kendall, D.G. (1953) Stochastic Processes Occurring in the Theory of Queues and Their Analysis by the Method of the Imbedded Markov Chain. *The Annals of Mathematical Statistics*, **24**, 338-354. <https://doi.org/10.1214/aoms/1177728975>
- [4] Alfa, A.S. and He, Q.M. (2008) Algorithmic Analysis of the Discrete Time $GI^X/G^Y/1$ Queueing System. *Performance Evaluation*, **65**, 623-640. <https://doi.org/10.1016/j.peva.2008.02.001>
- [5] Chaudhry, M.L. and Gupta, U.C. (1997) Queue-Length and Waiting-Time Distributions of Discrete-Time $GI^X/Geom/1$ Queueing Systems with Early and Late Arrivals. *Queueing Systems*, **25**, 307-324. <https://doi.org/10.1023/A:1019144116136>
- [6] Chaudhry, M.L. and Gupta, U.C. (2001) Computing Waiting-Time Probabilities in the Discrete-Time Queue: $GI^X/G/1$. *Performance Evaluation*, **43**, 123-131. [https://doi.org/10.1016/S0166-5316\(00\)00038-9](https://doi.org/10.1016/S0166-5316(00)00038-9)
- [7] Chaudhry, M.L. and Kim, N.K. (2003) A Complete and Simple Solution for a Discrete-Time Multi-Server Queue with Bulk Arrivals and Deterministic Service Times. *Operations Research Letters*, **31**, 101-107. [https://doi.org/10.1016/S0167-6377\(02\)00214-6](https://doi.org/10.1016/S0167-6377(02)00214-6)
- [8] Borkakaty, B., Agarwal, M. and Sen, K. (2010) Lattice Path Approach for Busy Period Density of $GI^P/G^b/1$ Queues Using C_2 Coxian Distributions. *Applied Mathematical Modelling*, **34**, 1597-1614. <https://doi.org/10.1016/j.apm.2009.09.005>
- [9] Claeys, D., Walraevens, J., Laevens, K. and Bruneel, H. (2007) A Discrete-Time Queueing Model with a Batch Server Operating under the Minimum Batch Size Rule. *NEW2AN2007: Next Generation Teletraffic and Wired/Wireless Advanced Networking*, **4712**, 248-259. https://doi.org/10.1007/978-3-540-74833-5_21
- [10] Claeys, D., Walraevens, J., Laevens, K. and Bruneel, H. (2010) Delay Analysis of Two Batch-Service Queueing Models with Batch Arrivals: $Geo^X/Geo^c/1$. *4OR*, **8**, 255-269. <https://doi.org/10.1007/s10288-009-0111-2>
- [11] Lim, D.E., Lee, D.H., Yang, W.S. and Chae, K.C. (2013) Analysis of the $GI/Geo/1$ Queue with N -Policy. *Applied Mathematical Modelling*, **37**, 4643-4652. <https://doi.org/10.1016/j.apm.2012.09.037>
- [12] Bruneel, H., Rogiest, W., Walraevens, J. and Wittevrongel, S. (2015) Analysis of a Discrete-Time Queue with General Independent Arrivals, General Service Demands and Fixed Service Capacity. *Mathematical Methods of Operations Research*, **82**, 285-315. <https://doi.org/10.1007/s00186-015-0515-z>
- [13] Bruneel, H., Wittevrongel, S., Claeys, D. and Walraevens, J. (2016) Discrete-Time Queues with Variable Service Capacity: A Basic Model and Its Analysis. *Annals of Operations Research*, **239**, 359-380. <https://doi.org/10.1007/s10479-013-1428-y>
- [14] Zhao, Y.Q. and Campbell, L.L. (1996) Equilibrium Probability Calculations for a Discrete-Time Bulk Queue Model. *Queueing Systems*, **22**, 189-198.

<https://doi.org/10.1007/BF01159401>

- [15] Janssen, A.J.E.M. and van Leeuwaarden, J.S.H. (2005) Analytic Computation Schemes for the Discrete-Time Bulk Service Queue. *Queueing Systems*, **50**, 141-163.
<https://doi.org/10.1007/s11134-005-0402-z>
- [16] Yao, M.H., Chen, X.J. and Zuo, L.C. (2014) Customer Arrival Event Processing on Computer Simulation for Discrete Event System. *Applied Mechanics and Materials*, **513-517**, 2133-2136.
<https://doi.org/10.4028/www.scientific.net/AMM.513-517.2133>
- [17] Lorente, A. (2016) Sistemas de colas en tiempo discreto con entradas y salidas en bloque: Estudio teórico y simulaciones comparativas. Ph.D. Dissertation, Universidad de Burgos, Burgos, Spain.