

Article

Mathematical Model of Call Center in the Form of Multi-Server Queueing System

Anatoly Nazarov , Alexander Moiseev *  and Svetlana Moiseeva 

Institute of Applied Mathematics and Computer Science, Tomsk State University, 634050 Tomsk, Russia; nazarov.tsu@gmail.com (A.N.); smoiseeva@mail.ru (S.M.)

* Correspondence: moiseev.tsu@gmail.com

Abstract: The paper considers the model of a call center in the form of a multi-server queueing system with Poisson arrivals and an unlimited waiting area. In the model under consideration, incoming calls do not differ in terms of service conditions, requested service, and interarrival periods. It is assumed that an incoming call can use any free server and they are all identical in terms of capabilities and quality. The goal problem is to find the stationary distribution of the number of calls in the system for an arbitrary recurrent service. This will allow us to evaluate the performance measures of such systems and solve various optimization problems for them. Considering models with non-exponential service times provides solutions for a wide class of mathematical models, making the results more adequate for real call centers. The solution is based on the approximation of the given distribution function of the service time by the hyperexponential distribution function. Therefore, first, the problem of studying a system with hyperexponential service is solved using the matrix-geometric method. Further, on the basis of this result, an approximation of the stationary distribution of the number of calls in a multi-server system with an arbitrary distribution function of the service time is constructed. Various issues in the application of this approximation are considered, and its accuracy is analyzed based on comparison with the known analytical result for a particular case, as well as with the results of the simulation.

Keywords: multi-server queue; hyperexponential service; matrix-geometric method; hyperexponential approximation



Citation: Nazarov, A.; Moiseev, A.; Moiseeva, S. Mathematical Model of Call Center in the Form of Multi-Server Queueing System. *Mathematics* **2021**, *9*, 2877. <https://doi.org/10.3390/math9222877>

Academic Editor: Mark Kelbert

Received: 30 September 2021

Accepted: 9 November 2021

Published: 12 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

One of the fastest-growing segments of the telecommunications market is the use of call centers. In this field, most of the costs of maintaining a call center are the salaries of employees who provide various kinds of information services to clients. The number of operators is chosen based on some standard measures, for example, the proportion of customers for which the waiting time for the start of service does not exceed the average time of the request servicing. The solution to the problem depends on the cost relationships between the components of the system. Therefore, the development of mathematical models and methods in this area is an necessary task. Traditional approaches of such studies lay in the field of queueing theory and related topics.

An excellent description of call centers and general issues in applying queueing models to studies were presented in [1]; some other general issues on the topic may be found in [2–5]. In this paper, we consider a call center model in the form of a multi-server queue with an incoming Poisson flow of calls, recurrent service, and an unlimited number of waiting places ($M/G/N$). A number of articles by different authors are devoted to the study of models of this type, for example: [6–10]. Furthermore, there are studies of models with non-Poisson incoming flows, for example: [11–18]. The aforementioned works make it possible to obtain characteristics of the operation of the system mainly using numerical, approximate, or asymptotic methods. Moreover, most of the papers only

evaluate mean values of waiting time or queue length, or the distribution of waiting time, or the distribution of queue length, but only under asymptotic conditions.

Our paper proposes one more approximate approach to solving the problem of studying the $M/G/N$ model, which allows one to obtain an accurate enough approximation of the probability distribution of the number of calls in the system for a wide class of service time distributions. Unlike other publications, our approach allows one to obtain a distribution of the number of calls, and this is performed directly without any asymptotic conditions, as in [8,16]. In addition, the proposed approach is based on substituting an approximation *before* performing the analysis instead of adapting known results for $M/M/N$ queues for the more general cases as was made in [9,10,12,14]. This allows us to obtain the final result in a simpler way with very high precision.

The main idea of the approach is in the approximation of the service time distribution by a two-phase hyperexponential one. The idea is taken from [19], where the problem of waiting time distribution was considered. We apply the idea to find the stationary distribution of the number of calls in this paper. For the model with hyperexponential service time, we use the matrix-geometric method to find the distribution of the number of calls in the system. To estimate the parameters of the hyperexponential distribution approximating a given distribution of the service times, in most cases, we use an approach based on three raw moments, which is in a certain sense similar to the results of [20]. However, for some cases, we have to use an approach based on two moments, and the results show us very high accuracy, even for such an approximation.

The only known exact analytical result is represented by the Pollaczek–Khinchine formula for a system $M/G/1$ with a single server [21,22]. In this paper, we compare the obtained approximation with this formula for the particular case $N = 1$ in the $M/G/N$ model. For the other cases (when the number of servers $N > 1$), we use the simulation results for the comparison. The performed numerical experiments prove the high accuracy of the approach for models $M/G/N$ in a wide class of service times.

The rest of the paper is organized as follows. In Section 2, we propose the mathematical model for the problem domain and formulate the goal of the study. In Section 3, we solve the problem for the partial case of the model with two-phase hyperexponential service times. To obtain the results for this case, we first derive the system of balance equation; after that, we apply the matrix-geometric method to solve it and derive the goal stationary probabilities at the end of the section. In Section 4, we propose a technique to obtain the results for the model $M/G/N$; also, we consider methods of the approximation and various issues of their applications. The results of estimating the accuracy of the proposed approach are presented in Section 5.

2. Mathematical Model

Consider a call center where we have several service devices (operators, servers) (Figure 1). An incoming call is redirected instantly to any free server or should be placed in a waiting queue if all the servers are busy. Calls in the queue are ordered according to the FCFS rule (First Come First Serve). After the completion of the service at any server, the server takes the first call from the queue for its service.

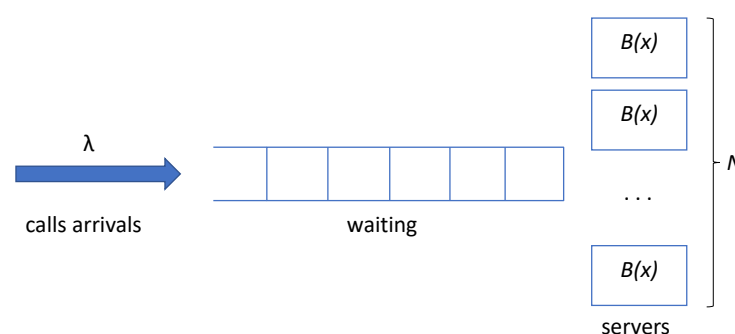


Figure 1. Structure of the model.

Consider a model of the system in the following form. Let calls arrive at the system according to the stationary Poisson process with intensity λ , let there be an unlimited number of places for waiting calls, and also, let there be N operators (servers) working independently from each other with service times distributed with distribution function $B(x)$. For the goals of the study, we suppose that at least two finite raw moments exist for this distribution (the reasons may be found in Section 4).

Denote the stationary probabilities that we have exactly i calls in the system (both under the service and in the queue) by $\pi(i)$, $i \in \{0, 1, 2, \dots\}$. The goal of the study is to find this probability distribution. The condition of existence of the stationary regime is obvious and standard for such a class of models:

$$\lambda < N\mu,$$

where μ is a rate of service in one server (the value that is inverse to the mean of a random variable with distribution function $B(x)$).

The approach that we use is based on the approximation of the distribution $B(x)$ by the hyperexponential distribution with two phases. So, we first try to obtain the solution for the model with hyperexponential service times. After that, we propose methods for the approximation of an arbitrary probability distribution by the hyperexponential one. At the end of the paper, we estimate errors of the solution of the original problem on the basis of comparison with known analytical results or simulation results.

The choice of hyperexponential distribution for the goals of the study is made based on the fact that this distribution allows studies to be carried out on queueing systems where it is used, and it can also approximate many other types of random variables distributions with a quite good accuracy.

3. Solution of the Problem for the Case of Hyperexponential Service Time

Let the service times be hyperexponentially distributed with the CDF:

$$B(x) = q_1(1 - e^{-\mu_1 x}) + q_2(1 - e^{-\mu_2 x}), \quad (1)$$

where $q_1 = q$, $q_2 = 1 - q$, $0 < q < 1$. Actually, this hyperexponential distribution is a mixture of two exponential distributions with parameters μ_1 and μ_2 chosen with probabilities q_1 and q_2 , respectively. So, actually, a server performs the service of a single call according to the exponential distribution with a chosen parameter. We name the servers that are currently working with parameter μ_1 as working at the first phase, and the servers that are currently working with parameter μ_2 as working at the second phase of the hyperexponential distribution.

Denote the number of servers that are working at the first phase in time moment t by $n_1(t)$ and servers working at the second phase by $n_2(t)$. At any moment, the number of free servers equals $N - n_1(t) - n_2(t)$. Furthermore, denote the number of calls in the queue in moment t by $i(t)$. Notice that it may be greater than zero only if $n_1(t) + n_2(t) = N$.

Let us introduce the following notations for the stationary probabilities:

- $P(n_1, n_2) = \Pr\{n_1(t) = n_1, n_2(t) = n_2\}$ for $0 \leq n_1 + n_2 \leq N - 1$ (when not all servers are busy),
- $P_n(i) = \Pr\{n_1(t) = n, n_2(t) = N - n, i(t) = i\}$ when all servers are busy. Here, $0 \leq n \leq N$ and $i = 0, 1, 2, \dots$

So, the goal of the study for this model is to find these probabilities.

3.1. Balance Equations

We can construct the following system of the balance equations for the stationary regime. When the queue is empty, we have

$$-P(0, 0)\lambda + P(0, 1)\mu_2 + P(1, 0)\mu_1 = 0, \quad (2)$$

$$-P(n_1, n_2)(\lambda + n_1\mu_1 + n_2\mu_2) + P(n_1 - 1, n_2)\lambda q_1 + P(n_1, n_2 - 1)\lambda q_2 + P(n_1, n_2 + 1)(n_2 + 1)\mu_2 + P(n_1 + 1, n_2)(n_1 + 1)\mu_1 = 0 \quad (3)$$

for $1 \leq n_1 + n_2 \leq N - 1$; for the non-empty queue, we can write

$$-P_n(0)(\lambda + n\mu_1 + (N - n)\mu_2) + P(n - 1, N - n)\lambda q_1 + P(n, N - n - 1)\lambda q_2 + P_{n-1}(1)(N - n + 1)\mu_2 q_1 + P_n(1)(n\mu_1 q_1 + (N - n)\mu_2 q_2) + P_{n+1}(1)(n + 1)\mu_1 q_2 = 0, \quad (4)$$

$$-P_n(i)(\lambda + n\mu_1 + (N - n)\mu_2) + P_n(i - 1)\lambda + P_{n-1}(i + 1)(N - n + 1)\mu_2 q_1 + P_n(i + 1)(n\mu_1 q_1 + (N - n)\mu_2 q_2) + P_{n+1}(i + 1)(n + 1)\mu_1 q_2 = 0 \quad (5)$$

for $0 \leq n \leq N$ and $2 \leq i \leq \infty$.

We propose an algorithm for solving this system of equations with two stages. The first stage is based on the matrix-geometric method [23,24]. In this case, solution $P_n(i)$ of part (5) of the system for $i > 0$ will be expressed using term $P_n(0)$. This fact allows us to rewrite the rest of the system (2)–(4) in a closed form of the system of $(N + 1)(N + 2)/2$ equations with the same number of variables. The second stage of the proposed algorithm will solve this system.

3.2. The First Stage: Matrix-Geometric Method

Consider part (5) of system (2)–(5) for $i > 0$.

Denote the row vector

$$\mathbf{P}(i) = [P_0(i) \ P_1(i) \ \dots \ P_N(i)]$$

and vector $\mathbf{P} = \mathbf{P}(0)$ to separate it from other components of the solution. Then, we can rewrite (5) in the matrix form:

$$\begin{aligned} -\mathbf{P}(1)\mathbf{D} + \mathbf{P}\lambda + \mathbf{P}(2)\mathbf{A} &= \mathbf{0}, \\ -\mathbf{P}(i)\mathbf{D} + \mathbf{P}(i - 1)\lambda + \mathbf{P}(i + 1)\mathbf{A} &= \mathbf{0}, \text{ for } i \geq 2, \end{aligned} \quad (6)$$

where matrices \mathbf{D} and \mathbf{A} have the form

$$\mathbf{D} = \begin{bmatrix} \lambda + N\mu_2 & 0 & 0 & \dots & 0 \\ 0 & \lambda + \mu_1 + (N - 1)\mu_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda + 2\mu_1 + (N - 2)\mu_2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda + N\mu_1 \end{bmatrix},$$

$$\mathbf{A} = \begin{bmatrix} N\mu_2 q_2 & N\mu_2 q_1 & 0 & \dots & 0 & 0 \\ \mu_1 q_2 & \mu_1 q_1 + (N - 1)\mu_2 q_2 & (N - 1)\mu_2 q_1 & \dots & 0 & 0 \\ 0 & 2\mu_1 q_2 & 2\mu_1 q_1 + (N - 2)\mu_2 q_2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & N\mu_1 q_2 & N\mu_1 q_1 \end{bmatrix}.$$

We write solution $\mathbf{P}(i)$ of system (6) in the matrix-geometric form:

$$\mathbf{P}(i) = \mathbf{P} \cdot \mathbf{R}^i. \quad (7)$$

Substituting it into (6), we derive a quadratic equation for the matrix \mathbf{R} :

$$\mathbf{R}^2 \mathbf{A} - \mathbf{R} \mathbf{D} + \lambda \mathbf{I} = \mathbf{O}. \quad (8)$$

This equation can be solved by the iterative method starting from zero matrix \mathbf{O} :

$$\mathbf{R}_0 = \mathbf{O}, \quad \mathbf{R}_{n+1} = (\mathbf{R}_n^2 \mathbf{A} + \lambda \mathbf{I}) \mathbf{D}^{-1}$$

until absolute values of the entries of matrix $\mathbf{R}_n^2 \mathbf{A} - \mathbf{R}_n \mathbf{D} + \lambda \mathbf{I}$ become small enough.

To find vector \mathbf{P} , which is important for (7), we consider vector $\mathbf{P}(1) = \mathbf{P} \cdot \mathbf{R}$, whose entries can be written in the form

$$P_n(1) = \sum_{k=0}^N P_k(0) R_{kn}, \text{ for } n = 0, 1, \dots, N.$$

Substituting these expressions into system (2)–(5), we obtain

$$\begin{aligned} & -P(0,0)\lambda + P(0,1)\mu_2 + P(1,0)\mu_1 = 0, \\ & -P(n_1, n_2)(\lambda + n_1\mu_1 + n_2\mu_2) + P(n_1 - 1, n_2)\lambda q_1 + P(n_1, n_2 - 1)\lambda q_2 \\ & + P(n_1, n_2 + 1)(n_2 + 1)\mu_2 + P(n_1 + 1, n_2)(n_1 + 1)\mu_1 = 0, \text{ for } 1 \leq n_1 + n_2 \leq N - 1, \\ & -P_n(0)[\lambda + n_1\mu_1 + (N - n)\mu_2] + P(n - 1, N - n)\lambda q_1 + P(n, N - n - 1)\lambda q_2 \\ & + \sum_{k=0}^N P_k(0)\{R_{k,n-1}(N - n + 1)\mu_2 q_1 + R_{kn}[n\mu_1 q_1 + (N - n)\mu_2 q_2] + R_{k,n+1}(n + 1)\mu_1 q_2\} = 0. \end{aligned}$$

Denoting

$$S_{kn} = R_{kn} - 1(N - n + 1)\mu_2 q_1 + R_{kn}[n\mu_1 q_1 + (N - n)\mu_2 q_2] + R_{k,n+1}(n + 1)\mu_1 q_2, \quad (9)$$

we rewrite the system in the following form:

$$-P(0,0)\lambda + P(0,1)\mu_2 + P(1,0)\mu_1 = 0, \quad (10)$$

$$\begin{aligned} & -P(n_1, n_2)(\lambda + n_1\mu_1 + n_2\mu_2) + P(n_1 - 1, n_2)\lambda q_1 + P(n_1, n_2 - 1)\lambda q_2 \\ & + P(n_1, n_2 + 1)(n_2 + 1)\mu_2 + P(n_1 + 1, n_2)(n_1 + 1)\mu_1 = 0, \quad (11) \\ & \text{for } 1 \leq n_1 + n_2 \leq N - 1, \end{aligned}$$

$$\begin{aligned} & -P_n(0)[\lambda + n_1\mu_1 + (N - n)\mu_2] + P(n - 1, N - n)\lambda q_1 \\ & + P(n, N - n - 1)\lambda q_2 + \sum_{k=0}^N P_k(0)S_{kn} = 0. \quad (12) \end{aligned}$$

This is a system of $(N + 1)(N + 2)/2$ linear equations with the same number of variables $P(0,0)$, $P(n_1, n_2)$ for $1 \leq n_1 + n_2 \leq N - 1$, and $P_n(0)$. Due to the fact that this system is a homogeneous, its non-trivial solution can be determined in a form with a constant multiplier which can be obtained from the normalization condition. In the next section, we propose an algorithm that reduces system (10)–(12) consisting of $(N + 1)(N + 2)/2$ equations to the homogeneous system with $N + 1$ equations with variables $P_n(0)$, $n = 0, 1, \dots, N$ that are entries of vector \mathbf{P} , which we wish to derive.

3.3. The Second Stage: Deriving Entries of Vector \mathbf{P}

To solve system (10)–(12), let us use the following notation for vectors with growing dimensions:

$$\begin{aligned} \mathbf{P}_0 &= [P(0,0)], \mathbf{P}_1 = [P(0,1) \ P(1,0)], \mathbf{P}_2 = [P(0,2) \ P(1,1) \ P(2,0)], \dots, \\ \mathbf{P}_n &= [P(0,n) \ P(1,n-1) \ \dots \ P(n-1,1) \ P(n,0)], \ 0 \leq n \leq N - 1. \end{aligned}$$

Using these notations, we rewrite Equation (10) in the matrix form:

$$-\mathbf{P}_0 \mathbf{D}_0 + \mathbf{P}_1 \mathbf{A}_1 = 0, \quad (13)$$

where

$$\mathbf{D}_0 = \lambda, \mathbf{A}_1 = \begin{bmatrix} \mu_2 \\ \mu_1 \end{bmatrix}.$$

We can write two equations from (11), where $n_1 + n_2 = 1$ in the matrix form:

$$\mathbf{P}_0 \mathbf{B}_0 - \mathbf{P}_1 \mathbf{D}_1 + \mathbf{P}_2 \mathbf{A}_2 = \mathbf{0}, \quad (14)$$

where

$$\mathbf{B}_0 = [\lambda q_2 \ \lambda q_1], \ \mathbf{D}_1 = \begin{bmatrix} \lambda + \mu_2 & 0 \\ 0 & \lambda + \mu_1 \end{bmatrix}, \ \mathbf{A}_2 = \begin{bmatrix} 2\mu_2 & 0 \\ \mu_1 & \mu_2 \\ 0 & 2\mu_1 \end{bmatrix}.$$

Further, for the case $n_1 + n_2 = n$, we can write corresponding equations from (11) in the form

$$\mathbf{P}_{n-1} \mathbf{B}_{n-1} - \mathbf{P}_n \mathbf{D}_n + \mathbf{P}_{n+1} \mathbf{A}_{n+1} = \mathbf{0} \text{ for } 1 \leq n \leq N-1, \quad (15)$$

where matrices \mathbf{B}_{n-1} with size $n \times (n+1)$, \mathbf{D}_n with size $(n+1) \times (n+1)$ and \mathbf{A}_{n+1} with size $(n+2) \times (n+1)$ have the following structures:

$$\mathbf{B}_{n-1} = \begin{bmatrix} \lambda q_2 & \lambda q_1 & 0 & \dots & 0 & 0 \\ 0 & \lambda q_2 & \lambda q_1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda q_2 & \lambda q_1 \end{bmatrix},$$

$$\mathbf{D}_n = \begin{bmatrix} \lambda + n\mu_2 & 0 & 0 & \dots & 0 & 0 \\ 0 & \lambda + \mu_1 + (n-1)\mu_2 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda + (n-1)\mu_1 + \mu_2 & 0 \\ 0 & 0 & 0 & \dots & 0 & \lambda + n\mu_1 \end{bmatrix},$$

$$\mathbf{A}_{n+1} = \begin{bmatrix} (n+1)\mu_2 & 0 & 0 & \dots & 0 & 0 \\ \mu_1 & n\mu_2 & 0 & \dots & 0 & 0 \\ 0 & 2\mu_1 & (n-1)\mu_2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 2\mu_2 & 0 \\ 0 & 0 & 0 & \dots & n\mu_1 & \mu_2 \\ 0 & 0 & 0 & \dots & 0 & (n+1)\mu_2 \end{bmatrix}.$$

Notation $\mathbf{0}$ in (14) and (15) means vector of zeros with corresponding sizes.

Finally, we write the last $N+1$ equations of system (10)–(12) in the form

$$\mathbf{P}_{N-1} \mathbf{B}_{N-1} + \mathbf{P} \mathbf{S} - \mathbf{P} \mathbf{D}_N = \mathbf{0}, \quad (16)$$

where \mathbf{S} is a matrix with entries S_{kn} from (9).

So, gathering all matrix Equations (13)–(16), we can rewrite system (10)–(12) in the matrix form:

$$\begin{cases} -\mathbf{P}_0 \mathbf{D}_0 + \mathbf{P}_1 \mathbf{A}_1 = 0, \\ \mathbf{P}_0 \mathbf{B}_0 - \mathbf{P}_1 \mathbf{D}_1 + \mathbf{P}_2 \mathbf{A}_2 = \mathbf{0}, \\ \mathbf{P}_{n-1} \mathbf{B}_{n-1} - \mathbf{P}_n \mathbf{D}_n + \mathbf{P}_{n+1} \mathbf{A}_{n+1} = \mathbf{0} \text{ for } 1 \leq n \leq N-1, \\ \mathbf{P}_{N-1} \mathbf{B}_{N-1} + \mathbf{P} \mathbf{S} - \mathbf{P} \mathbf{D}_N = \mathbf{0}. \end{cases} \quad (17)$$

From the first equation of the system, we can express:

$$\mathbf{P}_0 = \mathbf{P}_1 \mathbf{A}_1 \mathbf{D}_0^{-1} = \mathbf{P}_1 \mathbf{G}_1,$$

where $\mathbf{G}_1 = \mathbf{A}_1 \mathbf{D}_0^{-1}$.

From the second equation, we express:

$$\begin{aligned}\mathbf{P}_1 \mathbf{D}_1 - \mathbf{P}_0 \mathbf{B}_0 &= \mathbf{P}_2 \mathbf{A}_2, \\ \mathbf{P}_1 \mathbf{D}_1 - \mathbf{P}_1 \mathbf{G}_1 \mathbf{B}_0 &= \mathbf{P}_2 \mathbf{A}_2, \\ \mathbf{P}_1 &= \mathbf{P}_2 \mathbf{A}_2 (\mathbf{D}_1 - \mathbf{G}_1 \mathbf{B}_0)^{-1} = \mathbf{P}_2 \mathbf{G}_2,\end{aligned}$$

where $\mathbf{G}_2 = \mathbf{A}_2 (\mathbf{D}_1 - \mathbf{G}_1 \mathbf{B}_0)^{-1}$.

From the next equations of system (17) for $n \leq N - 1$, we obtain

$$\begin{aligned}\mathbf{P}_n \mathbf{D}_n - \mathbf{P}_{n-1} \mathbf{B}_{n-1} &= \mathbf{P}_{n+1} \mathbf{A}_{n+1}, \\ \mathbf{P}_n \mathbf{D}_n - \mathbf{P}_n \mathbf{G}_n \mathbf{B}_{n-1} &= \mathbf{P}_{n+1} \mathbf{A}_{n+1}, \\ \mathbf{P}_n &= \mathbf{P}_{n+1} \mathbf{A}_{n+1} (\mathbf{D}_n - \mathbf{G}_n \mathbf{B}_{n-1})^{-1} = \mathbf{P}_{n+1} \mathbf{G}_{n+1},\end{aligned}\quad (18)$$

where $\mathbf{G}_{n+1} = \mathbf{A}_{n+1} (\mathbf{D}_n - \mathbf{G}_n \mathbf{B}_{n-1})^{-1}$.

Consider equality (18) for the case $n = N - 1$. We can write:

$$\mathbf{P}_{N-1} = \mathbf{P} \mathbf{G}_N \quad (19)$$

because $\mathbf{P}_N = \mathbf{P}$. Let us substitute expression (19) into the last equation of system (17). Then we obtain the following equation for vector \mathbf{P} :

$$\mathbf{P} (\mathbf{G}_N \mathbf{B}_{N-1} + \mathbf{S} - \mathbf{D}_N) = \mathbf{0}, \quad (20)$$

which is the homogeneous linear system for entries $P_n(0)$ of vector \mathbf{P} . This system can be solved without any problems. So, we obtain vector \mathbf{P} .

After that, we can substitute vector \mathbf{P} as vector \mathbf{P}_N into the expressions

$$\mathbf{P}_n = \mathbf{P}_{n+1} \mathbf{G}_{n+1}$$

for $n = N - 1, N - 2, \dots, 1, 0$, and find vectors \mathbf{P}_n iteratively for all $n \in \{0, 1, \dots, N - 1\}$.

Finally, applying expressions (7), we find vectors $\mathbf{P}(i)$ for all $i \geq 0$.

3.4. Stationary Probability Distribution of the Number of Calls in the System

As it is mentioned in Section 3.2, we obtained all vectors \mathbf{P}_n in the previous section, just in a form with a constant multiplier. To find true values of the vectors, we should apply the normalization condition

$$\sum_{n=0}^{N-1} \mathbf{P}_n \mathbf{e}_{n+1} + \mathbf{P} (\mathbf{I} - \mathbf{R})^{-1} \mathbf{e}_{N+1} = \mathbf{1}, \quad (21)$$

where \mathbf{e}_n are column vectors consisting of ones.

Denote stationary probabilities:

$$\begin{aligned}r(n) &= \Pr\{n(t) = n, i(t) = 0\} \text{ for } 0 \leq n \leq N - 1, \\ p(i) &= \Pr\{n(t) = N, i(t) = i\} \text{ for } i \geq 0.\end{aligned}\quad (22)$$

Here, $r(n)$ means the probability that n servers are busy, and $p(i)$ means that all N servers are busy and there are exactly i calls waiting in the queue.

Based on (21), we can write the following:

$$r(n) = \mathbf{P}_n \mathbf{e}_{n+1}, \text{ for } n = 0, 1, \dots, N - 1, \quad (23)$$

$$p(i) = \mathbf{P} \mathbf{R}^i \mathbf{e}_{N+1}, \text{ for } i \geq 0. \quad (24)$$

Considering the stationary probabilities $\pi(i)$, $i \in \{0, 1, 2, \dots\}$ that we have exactly i calls in the system and using notations (22), we can write the following:

$$\pi(i) = \begin{cases} r(i) & \text{for } 0 \leq i \leq N-1, \\ p(i-N) & \text{for } i \geq N, \end{cases} \quad (25)$$

and, so, we can evaluate probabilities $\pi(i)$ by using expressions (23) and (24).

4. Approximation for the Case of Arbitrary Distributed Service Times

Consider the system shown in Figure 1 with i.i.d. service times with arbitrary distribution function $B(x)$. To approximate probabilities $\pi(i)$ from (25), we use expressions (23)–(24), supposing that distribution $B(x)$ can be approximated by the hyperexponential distribution in form (1). So, the problem is reduced to establishing parameters μ_1 , μ_2 , q_1 , and $q_2 = 1 - q_1$ that provide enough accuracy for the approximation of distribution $B(x)$ by the hyperexponential distribution (1).

It is not hard to obtain the expressions for the parameters of the two-phase hyperexponential approximation if we have two or three finite raw moments of the original distribution $B(x)$. These expressions are presented in the following subsections.

4.1. Estimating Parameters of Hyperexponential Approximation Basing on Two Moments

Suppose distribution $B(x)$ has two finite raw moments that we denote as b_1 and b_2 , respectively. We may estimate parameters μ_1 , μ_2 , q_1 , and q_2 from the following equalities for the moments:

$$\begin{aligned} \frac{q_1}{\mu_1} + \frac{q_2}{\mu_2} &= b_1, \\ 2\frac{q_1}{\mu_1^2} + 2\frac{q_2}{\mu_2^2} &= b_2. \end{aligned} \quad (26)$$

The following expressions may be derived by an easy method:

$$\begin{aligned} c^2 &= \frac{b_2 - b_1^2}{b_1^2}, \quad q_1 = \frac{1}{2} \left(1 - \sqrt{\frac{c^2 - 1}{c^2 + 1}} \right), \quad q_2 = 1 - q_1, \\ \mu_1 &= 2\frac{q_1}{b_1}, \quad \mu_2 = 2\frac{q_2}{b_1}. \end{aligned} \quad (27)$$

4.2. Estimating Parameters of Hyperexponential Approximation Basing on Three Moments

If we suppose that distribution $B(x)$ has three finite raw moments that we denote as b_1 , b_2 , and b_3 , respectively, then the parameters of the hyperexponential distribution can be derived as the solution of the following system of equations:

$$\begin{aligned} \frac{q_1}{\mu_1} + \frac{q_2}{\mu_2} &= b_1, \\ 2\frac{q_1}{\mu_1^2} + 2\frac{q_2}{\mu_2^2} &= b_2, \\ 6\frac{q_1}{\mu_1^3} + 6\frac{q_2}{\mu_2^3} &= b_3. \end{aligned} \quad (28)$$

Solving system (28) is not hard. If we denote

$$u = \frac{b_3 - 3b_1b_2}{3(b_2 - 2b_1^2)}, \quad v = \frac{2b_1b_3 - 3b_2^2}{6(b_2 - 2b_1^2)}, \quad (29)$$

we obtain

$$\begin{aligned} q_1 &= \frac{1}{2} \left(1 - \frac{u - 2b_1}{\sqrt{u^2 - 4v}} \right), \quad q_2 = 1 - q_1, \\ \mu_1 &= \frac{u - \sqrt{u^2 - 4v}}{2v}, \quad \mu_2 = \frac{u + \sqrt{u^2 - 4v}}{2v}. \end{aligned} \quad (30)$$

It is clear that the approximation based on three moments is more accurate than one based on two moments. Therefore, it should be preferable if it is possible.

4.3. How to Apply the Approximations

Actually, estimation of the parameters in such a way may lead us to the following several types of results, including some unusual cases:

1. $\mu_1 > 0, \mu_2 > 0, 0 < q_1 < 1, 0 < q_2 < 1$.
2. $\mu_1 > 0, \mu_2 > 0, q_1 > 1, q_2 < 0$.
3. Pairs μ_1, μ_2 and q_1, q_2 are complex numbers with their conjugates where all real parts are positive.
4. One of parameters μ_1 or μ_2 has a negative value of its real part.

Obviously, in the cases 2 and 3, such values cannot be values of the parameters of hyperexponential distribution (1), but in the next section, we show that these values may be applied for approximation (25) and the result will have enough good accuracy. Moreover, when we deal with the third case with complex values of the parameters, corresponding function (1) is not a distribution function, but approximation (25) still works.

If you meet case 4, it is better to use estimation scheme (27) using two moments, b_1 and b_2 . This excludes the possibility of obtaining negative values for parameters μ_1 and μ_2 .

5. Accuracy of the Obtained Approximation

The analytical result for the single-server queue $M/G/1$ is well known and it is expressed by the Pollaczek–Khinchine formula. Unfortunately, we cannot compare obtained approximation (25) with the Pollaczek–Khinchine formula analytically, but we can compare the results numerically by performing evaluations on the basis of analytical expressions. For the case of a multi-server queue, there are no exact analytical results in the literature, so we compare obtained approximation (25) with the results of the simulation.

Consider single-server queueing system $M/G/1$. According to the Pollaczek–Khinchine formula, the characteristic function of the number of calls in the system has the form

$$H(u) = (1 - \lambda b) \frac{\beta(\lambda - e^{ju}\lambda)(1 - e^{ju})}{\beta(\lambda - e^{ju}\lambda) - e^{ju}}. \quad (31)$$

Here,

$$\beta(u) = \int_0^\infty e^{-ux} dB(x)$$

is the Laplace–Stieltjes transform of distribution function $B(x)$ of service time. Applying the inverse Fourier transform

$$\Pi(i) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-jui} H(u) du, \quad (32)$$

we obtain probability distribution $\Pi(i), i = 0, 1, 2, \dots$ of the number of calls in the system.

We determine the accuracy of approximation (25) by using the Kolmogorov distance for discrete distributions

$$\Delta = \max_{0 \leq i < \infty} \left| \sum_{k=0}^i (\pi(k) - \Pi(k)) \right|. \quad (33)$$

For the numerical experiment, we take the following three types of distribution functions $B(x)$ with mean $b_1 = 1$:

1. Gamma distribution with shape parameter α and inverse scale parameter β . The raw moments of the distribution are evaluated as follows:

$$b_1 = \frac{\alpha}{\beta} = 1, \quad b_2 = \frac{\alpha(\alpha + 1)}{\beta^2}, \quad b_3 = \frac{\alpha(\alpha + 1)(\alpha + 2)}{\beta^3}.$$

2. Weibull distribution with parameters γ and $\beta = \frac{1}{\Gamma\left(1 + \frac{1}{\gamma}\right)}$, and with the raw moments

$$b_1 = \beta \Gamma\left(1 + \frac{1}{\gamma}\right) = 1, \quad b_2 = \beta^2 \Gamma\left(1 + \frac{2}{\gamma}\right), \quad b_3 = \beta^3 \Gamma\left(1 + \frac{3}{\gamma}\right).$$

3. Lognormal distribution with parameters σ^2 and $a = -\frac{\sigma^2}{2}$. Its raw moments are the following:

$$b_1 = \exp\left\{a + \frac{\sigma^2}{2}\right\} = 1, \quad b_2 = \exp\{2a + 2\sigma^2\}, \quad b_3 = \exp\left\{3a + \frac{9}{2}\sigma^2\right\}.$$

Consider single-server systems with the Poisson arrival process with intensity $\lambda = 0.8$ and service times with distribution functions $B(x)$ from the list above. The results of estimation parameters of hyperexponential approximation (1) using expressions (29) and (30) for these distribution functions and corresponding values of Kolmogorov distance (33) (line Δ_1) for the resulting approximation (25) are presented in Tables 1–3. Here, we compare distribution (25) with analytical results obtained using the Pollaczek–Khinchine formula (31) and (32). For the lognormal distribution in interval $0.4 < \sigma^2 < 0.7$, applying formulas (29) and (30) leads to the case $\mu_2 < 0$; therefore, we use estimation expressions (28) for this interval. The corresponding results are presented in Table 4. Furthermore, we cannot obtain the result in the case of gamma distribution with shape parameter $\alpha = 2$. We think that this case can be resolved directly, but it requires additional study.

Table 1. Values of parameters μ_1 , μ_2 , and q of the hyperexponential approximation obtained for gamma distribution function for various values of its shape parameter α , and corresponding Kolmogorov distances Δ for the main result.

α	0.02	0.1	0.5	1	1.5	1.9	2.1	5	100
μ_1	0.030	0.141	0.586	1	1.368	1.737	$2 - 0.026j$	$2 - j$	$2 - 1.39j$
μ_2	3.370	3.859	3.414	3	2.632	2.263	$2 + 0.026j$	$2 + j$	$2 + 1.39j$
q	0.022	0.109	0.500	1	1.765	4.177	$0.5 - 38.75j$	$0.5 - 1.5j$	$0.5 - 1.41j$
Δ_1	0.032	0.016	0.002	0	1.5×10^{-4}	3.7×10^{-7}	3.7×10^{-7}	8.8×10^{-4}	0.002
Δ_5	0.057	0.019	0.004	0.002	0.002	0.004	0.002	0.002	0.004

Table 2. Values of parameters μ_1 , μ_2 , and q of the hyperexponential approximation obtained for Weibull distribution function for various values of its parameter γ , and corresponding Kolmogorov distances Δ for the main result.

γ	0.7	0.9	1	1.05	1.12	1.18	1.2	10	100
μ_1	0.471	0.711	1	1.104	1.275	1.515	$1.61 - 0.44j$	$2 - 1.38j$	$2 - 1.41j$
μ_2	2.026	2.049	2	1.958	1.866	1.686	$1.61 + 0.44j$	$2 + 1.38j$	$2 + 1.41j$
q	0.311	0.687	1	1.239	1.868	6.101	$0.5 - 3.63j$	$0.5 - 1.41j$	$0.5 - 1.41j$
Δ_1	0.004	0.0005	0	0.0001	0.0002	0.0001	0.0001	0.002	0.002
Δ_5	0.005	0.006	0.002	0.001	0.002	0.001	0.002	0.005	0.006

Table 3. Values of parameters μ_1 , μ_2 , and q of the hyperexponential approximation obtained for lognormal distribution function for various values of its parameter σ^2 (except values $0.4 < \sigma^2 < 0.7$), and corresponding Kolmogorov distances Δ for the main result.

σ^2	0.1	0.2	0.25	0.3	0.4	0.7	0.8	1	1.5
μ_1	$2.04 - 1.19j$	$2.22 - 0.84j$	$2.44 - 0.31j$	1.759	1.348	0.020	0.184	0.216	0.106
μ_2	$2.04 + 1.19j$	$2.22 + 0.84j$	$2.44 + 0.31j$	4.053	62.15	1.000	1.026	1.110	1.173
q	$0.5 - 1.48j$	$0.5 - 2.02j$	$0.5 - 5.79j$	2.342	1.356	2.8×10^{-6}	0.006	0.026	0.017
Δ_1	0.001	0.0008	0.0008	0.001	0.003	0.014	0.007	0.009	0.035
Δ_5	0.004	0.004	0.002	0.003	0.005	0.008	0.005	0.005	0.020

Table 4. Values of parameters μ_1 , μ_2 , and q of the hyperexponential approximation obtained for lognormal distribution function for values of its parameter $0.4 < \sigma^2 < 0.7$, and corresponding Kolmogorov distances Δ for the main result.

σ^2	0.41	0.50	0.55	0.60	0.69
μ_1	$1 - 0.572j$	$1 - 0.462j$	$1 - 0.392j$	$1 - 0.312j$	$1 - 0.056j$
μ_2	$1 + 0.572j$	$1 + 0.462j$	$1 + 0.392j$	$1 + 0.312j$	$1 + 0.056j$
q	$0.5 - 0.286j$	$0.5 - 0.231j$	$0.5 - 0.196j$	$0.5 - 0.156j$	$0.5 - 0.028j$
Δ_1	0.016	0.015	0.015	0.015	0.015
Δ_5	0.010	0.014	0.013	0.013	0.012

In line Δ_5 of the tables, you may find the Kolmogorov distance for distribution (25) evaluated on the basis of hyperexponential approximations of service time distribution for the system with five servers (the intensity of the Poisson arrivals is taken equal to 4). The results are compared with the simulation results (note that the simulation results in the experiments are obtained with an error of 0.002 in terms of the Kolmogorov distance). All numerical evaluations for the examples are made using MathCAD software.

In Figure 2, you may find probability mass functions (PMFs) for some of the considered examples for the multi-server model with five servers. We chose the most interesting and the most “strange” cases of the approximation to use as examples.

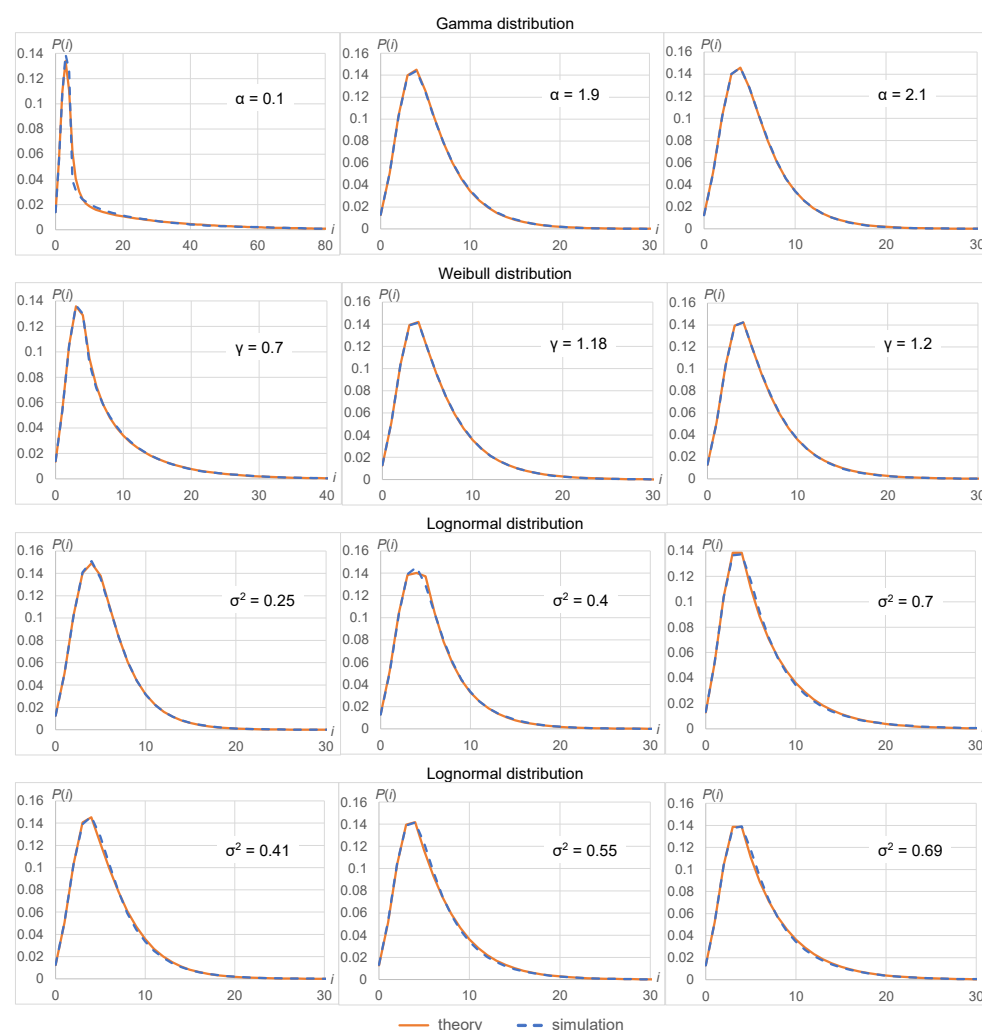


Figure 2. PMFs obtained by the approximation and simulation for multi-server queue with various distributions of service time and their parameters.

As we see from the presented results, distribution (25), using hyperexponential approximation of the service time, has very high accuracy in almost all cases (the worst results are only obtained for gamma distribution with a small value of the shape parameter). You may find various values of the parameters of the hyperexponential approximation in Tables 1–4, including strange ones (complex values, $q > 1$), but the final result (25) works properly for these cases too.

6. Discussion

The study presented in the paper shows that a two-phase hyperexponential approximation can be applied for service time distribution in the $M/G/N$ queue. This approximation provides highly accurate results for the goal probability distribution of the number of calls in the system even in cases where the approximation cannot be used as a distribution function. A comparison of the results with analytical results for the case of a single-server queue and simulation results for the case of a multi-server queue show the high accuracy of the approach.

Furthermore, in the paper, the analytical result for the system with hyperexponential service time is derived on the basis of the matrix-geometric approach. The result can be used both directly for the analysis of the $M/H_2/N$ queues or together with the hyperexponential approximation of the service time distribution for the analysis of the $M/G/N$ queues.

We believe that hyperexponential approximation can be applied in studies that investigate other types of queueing models where it can allow one to obtain a solution of a problem or, at least, obtain it more easily than by other methods. For example, we see good possibilities in analyzing systems of $G/G/1$ and $G/G/N$ types using the approximation and approach proposed in the paper.

Author Contributions: Conceptualization, A.N.; methodology, A.N.; software, A.M.; validation, A.M. and S.M.; formal analysis, S.M.; investigation, A.N.; writing—original draft preparation, A.N. and S.M.; writing—review and editing, A.M.; visualization, A.M.; supervision, A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gans, N.; Koole, G.; Mandelbaum, A. Telephone Call-centers: Tutorial, Review and Research Prospects. *Manuf. Serv. Manag.* **2003**, *5*, 79–141. [\[CrossRef\]](#)
2. Koole, G.; Mandelbaum, A. Queueing Models of Call Centers: An Introduction. *Ann. Oper. Res.* **2002**, *113*, 41–59. [\[CrossRef\]](#)
3. Borst, S.; Mandelbaum, A.; Reiman, M.I. Dimensioning large call centers. *Oper. Res.* **2004**, *52*, 17–34. [\[CrossRef\]](#)
4. Stollatz, R.; Helber, S. Performance analysis of an inbound call center with skills-based routing. *OR Spectr.* **2004**, *26*, 331–352. [\[CrossRef\]](#)
5. Stepanov, S.N.; Stepanov, M.S. Construction and analysis of a generalized contact center model. *Autom. Remote Control* **2014**, *75*, 1936–1947. [\[CrossRef\]](#)
6. Nozaki, S.A.; Ross, S.M. Approximations in finite-capacity multi-server queues with Poisson arrivals. *J. Appl. Prob.* **1978**, *15*, 826–834. [\[CrossRef\]](#)
7. Boxma, O.J.; Cohen, J.W.; Hufels, N. Approximations of the mean waiting time in an $M/G/s$ queueing system. *Oper. Res.* **1979**, *27*, 1115–1127. [\[CrossRef\]](#)
8. Whitt, W. Comparison conjectures about the $M/G/s$ queue. *OR Lett.* **1983**, *2*, 203–209. [\[CrossRef\]](#)
9. Kimura, T. Diffusion approximation for an $M/G/m$ queue. *Oper. Res.* **1983**, *31*, 304–321. [\[CrossRef\]](#)
10. Lisovskaya, E.Y.; Moiseeva, S.P. Study of the Queueing Systems $M/GI/N/\infty$. *Commun. Comput. Inf. Sci.* **2015**, *564*, 175–185.
11. Whitt, W. The effect of variability in the $GI/G/s$ queue. *J. Appl. Prob.* **1980**, *17*, 1062–1071. [\[CrossRef\]](#)
12. Shore, H. Simple Approximations for the $GI/G/c$ Queue-I: The Steady-State Probabilities. *J. Oper. Res. Soc.* **1988**, *39*, 279–284.
13. Whitt, W. Approximations for the $GI/G/m$ queue. *Prod. Oper. Manag.* **1993**, *2*, 114–161. [\[CrossRef\]](#)
14. Kimura, T. Approximations for multi-server queues: System interpolations. *Queueing Syst.* **1994**, *17*, 347–382. [\[CrossRef\]](#)

15. Escobar, M.; Odoni, A.; Roth, E. Approximate solution for multi-server queueing systems with Erlangian service times. *Comput. Oper. Res.* **2002**, *29*, 1353–1374. [[CrossRef](#)]
16. Whitt, W. A diffusion approximation for the G/GI/n/m queue. *Oper. Res.* **2004**, *52*, 922–941. [[CrossRef](#)]
17. Sun, B.; Dudin, A.N. The MAP/PH/N multi-server queueing system with broadcasting service discipline and server heating. *Aut. Control Comp. Sci.* **2013**, *47*, 173–182. [[CrossRef](#)]
18. Brandwajn, A.; Begin, T. Preliminary Results on a Simple Approach to G/G/c-Like Queues. In Proceedings of the International Conference on Analytical and Stochastic Modeling Techniques and Applications, Madrid, Spain, 9–12 June 2009; Al-Begain, K., Fiems, D., Horváth, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; pp. 159–173.
19. Ryzhikov, Y.I. Three methods for computing time characteristics of open queueing systems. *Autom. Remote Control* **1993**, *54*, 283–289.
20. Gupta, V.; Harchol-Balter, M.; Dai, J.G.; Zwart, B. On the inapproximability of M/G/K: Why two moments of job size distribution are not enough. *Queueing Syst.* **2010**, *64*, 5–48. [[CrossRef](#)]
21. Pollaczek, F. Über eine Aufgabe der Wahrscheinlichkeitstheorie. *Math. Z.* **1930**, *32*, 64–100. [[CrossRef](#)]
22. Khintchine, A.Y. Mathematical theory of a stationary queue. *Mat. Sb.* **1932**, *39*, 73–84.
23. Neuts, M.F. *Matrix-Geometric Solutions in Stochastic Models*; Johns Hopkins University Press: Baltimore, MD, USA, 1981; p. 352.
24. Neuts, M.F. Matrix-analytic methods in queueing theory. *Eur. J. Oper. Res.* **1984**, *15*, 2–12. [[CrossRef](#)]