

Article

Asymptotic Diffusion Analysis of Multi-Server Retrial Queue with Hyper-Exponential Service

Alexander Moiseev * and Anatoly Nazarov  and Svetlana Paul 

Institute of Applied Mathematics and Computer Science, Tomsk State University, Tomsk 634050, Russia; nazarov.tsu@gmail.com (A.N.); paulsv82@mail.ru (S.P.)

* Correspondence: moiseev.tsu@gmail.com

Received: 15 March 2020; Accepted: 1 April 2020; Published: 3 April 2020



Abstract: A multi-server retrial queue with a hyper-exponential service time is considered in this paper. The study is performed by the method of asymptotic diffusion analysis under the condition of long delay in orbit. On the basis of the constructed diffusion process, we obtain approximations of stationary probability distributions of the number of customers in orbit and the number of busy servers. Using simulations and numerical analysis, we estimate the accuracy and applicability area of the obtained approximations.

Keywords: multi-server retrial queue; hyper-exponential service time; asymptotic analysis; diffusion process

1. Introduction

Nowadays, retrial queues (RQ) are very popular mathematical models of various real systems. These may be call centers [1,2] where a calling customer, if he or she find all operators busy, tries to make a new call again after some (random) time. In the case of modeling of data transmission via TCP [3] if a data packet is lost or corrupted, the packet should be retransmitted after certain time. In cellular mobile networks, more complex models of retrial queues are used (see [4,5]). Furthermore, we see great opportunities in using multi-server retrial queues for modeling and design data processing systems with high-intensity data flows [6,7]. These are not only fields where retrial queues may be applied (e.g., survey [8]).

A retrial queue is a such queueing system with limited number of servers where customers that arrived and do not find a free server are not lost, but go to special ‘waiting’ state, or are moved to special buffer (*orbit*) for waiting. In the orbit, the customers are not organizing in a queue (e.g., first come, first served (FCFS)), but wait for a random time and then try to obtain the service again. There are a great number of RQs modifications considered by various authors—From ordinary single-server or multi-server systems to retrial queues with several orbits, priorities, collisions and so on. Detail information about retrial queue models, their analysis and applications you may find in surveys [8–12] and others.

Most of the authors consider single-line retrial queues and or retrial queues with exponentially distributed service time. In this paper, we consider a multi-line retrial queue with hyper-exponential service time, but we hold in mind a such model with arbitrary distributed service time as the aim for future work.

Unfortunately, exact analytical solutions for retrial queues may be derived in very rare cases. Due to this, the most authors use numerical methods, approximations or even simulation methods to study more complex retrial queues [13–17]. This also applies to multi-server models [16] including systems with non-exponential service time [13].

Traditionally, we study retrial queues under various asymptotic conditions such as heavy load or long delay [18–21]. In these studies we and our colleagues obtained approximations for stationary

probability distributions of the number of customers in the orbit that are applicable in appropriate conditions. Unlike these previous works, in this paper, we apply a new approach—The asymptotic diffusion method [22]—To perform more detailed and accurate analysis of the model. This will give more precise approximations and wider areas of their applicability.

The rest of the paper is organized as follows. In Section 2, we describe in detail a mathematical model of the considered retrial queue and derive balance equations. In Sections 3 and 4, we perform asymptotic analysis of obtained equations under condition of long delay in the orbit. In Section 5, we introduce the method of asymptotic diffusion analysis and construct a diffusion process which is used in Section 6 to build approximations for steady-state probability distributions of the number of busy servers and the number of customers in the orbit. Using simulations and numerical experiments, we determine the applicability area of suggested approximations in Section 7.

2. Mathematical Model

Consider a retrial queue with N servers and Poisson arrivals with intensity λ . Service times are independent and identically distributed as hyper-exponential variables with cumulative distribution function

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

where $\mu_1, \mu_2 > 0, 0 < q_1 < 1$, and $q_1 + q_2 = 1$.

When a customer arrives, if there is a free server the customer moves to it for service, otherwise the customer goes to the orbit and stays there during random period exponentially distributed with parameter σ . After this period the customer tries to get a service again according to the described algorithm.

Let us denote the number of busy servers at the moment t by $n(t) \in \{0, 1, \dots, N\}$, and the number of customers in the orbit by $i(t) \in \{0, 1, 2, \dots\}$. The problem is to find steady-state probability distribution of the two-dimensional process $\{n(t), i(t)\}$ that we denote by

$$P(n, i) = P\{n(t) = n, i(t) = i\}$$

for all $n \in \{0, 1, \dots, N\}$ and $i \in \{0, 1, 2, \dots\}$ independently of t .

To solve the problem, let us consider the stochastic process $\{n_1(t), n_2(t), i(t)\}$, where $n_1(t)$ and $n_2(t)$ are the numbers of busy servers that realizes the service time of the first phase or the second phase respectively. The process is Markovian, hence we can write balance equations for its probability distribution

$$P(n_1, n_2, i, t) = P\{n_1(t) = n_1, n_2(t) = n_2, i(t) = i\},$$

where $0 \leq n_1 + n_2 \leq N, i \in \{0, 1, 2, \dots\}$. So, we can write differential equations

$$\begin{aligned} \frac{\partial P(n_1, n_2, i, t)}{\partial t} = & -(\lambda + n_1\mu_1 + n_2\mu_2 + i\sigma)P(n_1, n_2, i, t) + q_1\lambda P(n_1 - 1, n_2, i, t) + q_2\lambda P(n_1, n_2 - 1, i, t) \\ & + q_1(i + 1)\sigma P(n_1 - 1, n_2, i + 1, t) + q_2(i + 1)\sigma P(n_1, n_2 - 1, i + 1, t) \\ & + (n_1 + 1)\mu_1 P(n_1 + 1, n_2, i, t) + (n_2 + 1)\mu_2 P(n_1, n_2 + 1, i, t) \end{aligned} \tag{2}$$

for the case $n_1 + n_2 \leq N - 1$, and

$$\begin{aligned} \frac{\partial P(n_1, n_2, i, t)}{\partial t} = & -(\lambda + n_1\mu_1 + n_2\mu_2)P(n_1, n_2, i, t) \\ & + q_1\lambda P(n_1 - 1, n_2, i, t) + q_2\lambda P(n_1, n_2 - 1, i, t) + \lambda P(n_1, n_2, i - 1, t) \\ & + q_1(i + 1)\sigma P(n_1 - 1, n_2, i + 1, t) + q_2(i + 1)\sigma P(n_1, n_2 - 1, i + 1, t) \end{aligned} \tag{3}$$

for the case $n_1 + n_2 = N$.

Let us introduce partial characteristic functions

$$H(n_1, n_2, u, t) = \sum_{i=0}^{\infty} e^{ju_i} P(n_1, n_2, i, t),$$

where $j = \sqrt{-1}$. Then we can rewrite the system of balance Equations (2)–(3) as follows:

$$\begin{aligned} \frac{\partial H(n_1, n_2, u, t)}{\partial t} = & -(\lambda + n_1\mu_1 + n_2\mu_2)H(n_1, n_2, u, t) + q_1\lambda H(n_1 - 1, n_2, u, t) + q_2\lambda H(n_1, n_2 - 1, u, t) \\ & + (n_1 + 1)\mu_1 H(n_1 + 1, n_2, u, t) + (n_2 + 1)\mu_2 H(n_1, n_2 + 1, u, t) \\ & - j\sigma q_1 e^{-ju} \frac{\partial H(n_1 - 1, n_2, u, t)}{\partial u} - j\sigma q_2 e^{-ju} \frac{\partial H(n_1, n_2 - 1, u, t)}{\partial u} + j\sigma \frac{\partial H(n_1, n_2, u, t)}{\partial u} \end{aligned} \tag{4}$$

for $n_1 + n_2 \leq N - 1$, and

$$\begin{aligned} \frac{\partial H(n_1, n_2, u, t)}{\partial t} = & -(\lambda + n_1\mu_1 + n_2\mu_2)H(n_1, n_2, u, t) \\ & + q_1\lambda H(n_1 - 1, n_2, u, t) + q_2\lambda H(n_1, n_2 - 1, u, t) + \lambda e^{ju} H(n_1, n_2, u, t) \\ & - j\sigma q_1 e^{-ju} \frac{\partial H(n_1 - 1, n_2, u, t)}{\partial u} - j\sigma q_2 e^{-ju} \frac{\partial H(n_1, n_2 - 1, u, t)}{\partial u} \end{aligned} \tag{5}$$

for $n_1 + n_2 = N$.

Let us use linear finite difference operators $\mathbf{A}, \mathbf{B}, \mathbf{I}_0, \mathbf{I}_1$ to rewrite the system Equations (4) and (5) in a compact form. Then using notation $\mathbf{H}(u, t)$ for $(N + 1) \times (N + 1)$ matrix function with indexes $n_1, n_2 = 0, \dots, N$ and entries equal to $H(n_1, n_2, u, t)$ for $n_1 + n_2 \leq N$ or zero otherwise, we can rewrite Equations (4) and (5) as follows:

$$\frac{\partial \mathbf{H}(u, t)}{\partial t} = \mathbf{H}(u, t) (\mathbf{A} + e^{ju} \mathbf{B}) + j\sigma \frac{\partial \mathbf{H}(u, t)}{\partial u} (\mathbf{I}_0 - e^{-ju} \mathbf{I}_1). \tag{6}$$

Here

$$\begin{aligned} \mathbf{H}(u, t)\mathbf{A} = & -(\lambda + n_1\mu_1 + n_2\mu_2)H(n_1, n_2, u, t) + q_1\lambda H(n_1 - 1, n_2, u, t) + q_2\lambda H(n_1, n_2 - 1, u, t) \\ & + \psi(n_1 + n_2, N) [(n_1 + 1)\mu_1 H(n_1 + 1, n_2, u, t) + (n_2 + 1)\mu_2 H(n_1, n_2 + 1, u, t)], \\ \mathbf{H}(u, t)\mathbf{B} = & [1 - \psi(n_1 + n_2, N)] \lambda H(n_1, n_2, u, t), \\ \mathbf{H}(u, t)\mathbf{I}_0 = & \psi(n_1 + n_2, N) H(n_1, n_2, u, t), \\ \mathbf{H}(u, t)\mathbf{I}_1 = & q_1 H(n_1 - 1, n_2, u, t) + q_2 H(n_1, n_2 - 1, u, t), \end{aligned}$$

where indicator $\psi(n_1 + n_2, N)$ is defined as follows:

$$\psi(n_1 + n_2, N) = \begin{cases} 1, & \text{if } 0 \leq n_1 + n_2 < N, \\ 0, & \text{otherwise: } n_1 + n_2 = N. \end{cases}$$

We write these operators on the right hand of matrix function affected by them similar to notations of matrix operations.

Let us denote summing operator by \mathbf{E} . Applying this operator on Equation (6) and taking into account that

$$(\mathbf{A} + \mathbf{B})\mathbf{E} = \mathbf{0} \tag{7}$$

and

$$(\mathbf{I}_0 - \mathbf{I}_1)\mathbf{E} = \mathbf{0}, \tag{8}$$

we obtain the following scalar equation:

$$\frac{\partial \mathbf{H}(u, t)}{\partial t} \mathbf{E} = \left(e^{ju} - 1 \right) \left\{ \mathbf{H}(u, t) \mathbf{B} + j\sigma e^{-ju} \frac{\partial \mathbf{H}(u, t)}{\partial u} \mathbf{I}_0 \right\} \mathbf{E}. \tag{9}$$

It will be an additional equation to the system Equation (6).

3. First Stage of the Asymptotic Analysis

System of Equations (6) and (9) can not be solved in a direct way, so, we will solve it under asymptotic condition of long delay in the orbit: $\sigma \rightarrow 0$.

Denoting $\sigma = \varepsilon$, we make the following substitutions in system Equations (6) and (9):

$$\tau = t\varepsilon, \quad u = \varepsilon w, \quad \mathbf{H}(u, t) = \mathbf{F}(w, \tau, \varepsilon). \tag{10}$$

We obtain

$$\varepsilon \frac{\partial \mathbf{F}(w, \tau, \varepsilon)}{\partial \tau} = \mathbf{F}(w, \tau, \varepsilon) \left(\mathbf{A} + e^{j\varepsilon w} \mathbf{B} \right) + j \frac{\partial \mathbf{F}(w, \tau, \varepsilon)}{\partial w} \left(\mathbf{I}_0 - e^{-j\varepsilon w} \mathbf{I}_1 \right), \tag{11}$$

$$\varepsilon \frac{\partial \mathbf{F}(w, \tau, \varepsilon)}{\partial \tau} \mathbf{E} = \left(e^{j\varepsilon w} - 1 \right) \left\{ \mathbf{F}(w, \tau, \varepsilon) \mathbf{B} \mathbf{E} + j e^{-j\varepsilon w} \frac{\partial \mathbf{F}(w, \tau, \varepsilon)}{\partial w} \mathbf{I}_0 \mathbf{E} \right\}. \tag{12}$$

We will solve system Equations (11) and (12) under asymptotic condition $\varepsilon \rightarrow 0$.

Let $x = x(\tau) = \lim_{\varepsilon \rightarrow 0} E\{\varepsilon i(\tau)\}$ be a scalar function which determines the average number of customers in the orbit normalized by $\varepsilon = \sigma$ asymptotically for $\varepsilon \rightarrow 0$. Denote probabilities of the number of busy servers working at the first phase (n_1) or second phase (n_2) of given hyper-exponential distribution of service time by $R(n_1, n_2, x)$ for $n_1, n_2 = 0, \dots, N$ and $n_1 + n_2 \leq N$. Let $\mathbf{R}(x)$ be $(N + 1) \times (N + 1)$ left-top triangle matrix with entries equal to $R(n_1, n_2, x)$ for $n_1 + n_2 \leq N$ and others equal to zero.

Theorem 1. Under asymptotic condition $\varepsilon \rightarrow 0$, probabilities $R(n_1, n_2, x)$ are determined by expressions

$$\begin{aligned} R(n_1, n_2, x) &= \frac{\rho_1(x)^{n_1}}{n_1!} \cdot \frac{\rho_2(x)^{n_2}}{n_2!} R(0, 0, x), \\ R(0, 0, x) &= 1 / \sum_{n_1+n_2 \leq N} \frac{\rho_1(x)^{n_1}}{n_1!} \cdot \frac{\rho_2(x)^{n_2}}{n_2!}, \end{aligned} \tag{13}$$

where

$$\rho_1(x) = q_1 \frac{(\lambda + x)}{\mu_1}, \quad \rho_2(x) = q_2 \frac{(\lambda + x)}{\mu_2}, \tag{14}$$

and $x = x(\tau)$ is a solution of the ordinary differential equation

$$x'(\tau) = \mathbf{R}(x) (\mathbf{B} - x(\tau) \mathbf{I}_0) \mathbf{E}. \tag{15}$$

Proof. Making asymptotic transition $\varepsilon \rightarrow 0$ in Equations (11) and (12) and denoting $\lim_{\varepsilon \rightarrow 0} \mathbf{F}(w, \tau, \varepsilon) = \mathbf{F}(w, \tau)$, we derive

$$\begin{aligned} \mathbf{F}(w, \tau) (\mathbf{A} + \mathbf{B}) + j \frac{\partial \mathbf{F}(w, \tau)}{\partial w} (\mathbf{I}_0 - \mathbf{I}_1) &= 0, \\ \frac{\partial \mathbf{F}(w, \tau)}{\partial \tau} \mathbf{E} &= jw \left\{ \mathbf{F}(w, \tau) \mathbf{B} \mathbf{E} + j \frac{\partial \mathbf{F}(w, \tau)}{\partial w} \mathbf{I}_0 \mathbf{E} \right\}. \end{aligned} \tag{16}$$

Let us find a solution of system Equation (16) in the form

$$\mathbf{F}(w, \tau) = \mathbf{R}(x) e^{jw x(\tau)}. \tag{17}$$

Substituting Equation (17) into Equation (16), we derive

$$\mathbf{R}(x) \{ \mathbf{A} + \mathbf{B} - x(\tau) (\mathbf{I}_0 - \mathbf{I}_1) \} = 0, \tag{18}$$

$$x'(\tau) = \mathbf{R}(x) (\mathbf{B} - x(\tau)\mathbf{I}_0) \mathbf{E}. \tag{19}$$

So, Equation (18) is a homogeneous system of linear equations for $R(n_1, n_2, x)$ which solution depends on $x(\tau)$. Equation (19) (that coincides with Equation (15)) is an ordinary differential equation of the first order for function $x(\tau)$. So, we can find a solution of Equation (18) and substitute it into Equation (19).

Let us solve system Equation (18). To do this, we rewrite matrix Equation (18) in the form of the following homogeneous system of linear equations:

$$q_1(\lambda + x)R(n_1 - 1, n_2, x) + q_2(\lambda + x)R(n_1, n_2 - 1, x) - (\lambda + x + n_1\mu_1 + n_2\mu_2)R(n_1, n_2, x) + (n_1 + 1)\mu_1R(n_1 + 1, n_2, x) + (n_2 + 1)\mu_2R(n_1, n_2 + 1, x) = 0$$

for $n_1 + n_2 \leq N - 1$, and

$$q_1(\lambda + x)R(n_1 - 1, n_2, x) + q_2(\lambda + x)R(n_1, n_2 - 1, x) - (n_1\mu_1 + n_2\mu_2)R(n_1, n_2, x) = 0$$

for $n_1 + n_2 = N$. Substituting here expressions Equations (13) and (14), we obtain true identities. So, the theorem is proved. \square

Let us denote

$$a(x) = \mathbf{R}(x) (\mathbf{B} - x(\tau)\mathbf{I}_0) \mathbf{E}. \tag{20}$$

Corollary 1. *Let*

$$r(x) = \sum_{n_1+n_2=N} R(n_1, n_2, x), \tag{21}$$

then function $a(x)$ may be written in the form

$$a(x) = (\lambda + x(\tau))r(x) - x(\tau) \tag{22}$$

Proof. From Equation (20) we derive

$$\begin{aligned} a(x) &= \mathbf{R}(x) (\mathbf{B} - x(\tau)\mathbf{I}_0) \mathbf{E} = a(\mathbf{R}(x)\mathbf{B}\mathbf{E} - x(\tau)\mathbf{R}(x)\mathbf{I}_0\mathbf{E}) \\ &= \lambda \sum_{n_1+n_2=N} R(n_1, n_2, x) - x \sum_{n_1+n_2 \leq N-1} R(n_1, n_2, x) \\ &= \lambda \sum_{n_1+n_2=N} R(n_1, n_2, x) - x \left(\sum_{n_1+n_2 \leq N} R(n_1, n_2, x) - \sum_{n_1+n_2=N} R(n_1, n_2, x) \right) \\ &= (\lambda + x) \sum_{n_1+n_2=N} R(n_1, n_2, x) - x \sum_{n_1+n_2 \leq N} R(n_1, n_2, x). \end{aligned}$$

Taking into account that $\sum_{n_1+n_2 \leq N} R(n_1, n_2, x) = 1$ and using notation Equation (21), we obtain Equation (22). \square

4. Second Stage of the Asymptotic Analysis

Taking into account that characteristic function of process $\frac{1}{\sigma}x(\sigma t)$ of the normalized number of customers in the orbit has form $\exp\{j\omega x(\tau)\}$ (see Equation (17)), let us make the following substitution in system Equation (9):

$$\mathbf{H}(u, t) = e^{j\frac{u}{\sigma}x(\sigma t)} \mathbf{H}^{(1)}(u, t).$$

(here $\mathbf{H}^{(1)}(u, t)\mathbf{E}$ is characteristic function of centered process $i(t) - \frac{1}{\sigma}x(\sigma t)$). Then we obtain the system

$$\begin{aligned} & \frac{\partial \mathbf{H}^{(1)}(u, t)}{\partial t} + jux'(\sigma t)\mathbf{H}^{(1)}(u, t) = \mathbf{H}^{(1)}(u, t) \left(\mathbf{A} + e^{ju}\mathbf{B} \right) \\ & + j\sigma \left[\frac{j}{\sigma}x(\sigma t)\mathbf{H}^{(1)}(u, t) + \frac{\partial \mathbf{H}^{(1)}(u, t)}{\partial u} \right] \left(\mathbf{I}_0 - e^{-ju}\mathbf{I}_1 \right), \\ & \left[\frac{\partial \mathbf{H}^{(1)}(u, t)}{\partial u} + jux'(\sigma t)\mathbf{H}^{(1)}(u, t) \right] \mathbf{E} \\ & = \left(e^{ju} - 1 \right) \left\{ \mathbf{H}^{(1)}(u, t)\mathbf{B}\mathbf{E} + j\sigma e^{-ju} \left[\frac{j}{\sigma}x(\sigma t)\mathbf{H}^{(1)}(u, t) + \frac{\partial \mathbf{H}^{(1)}(u, t)}{\partial u} \right] \mathbf{I}_0\mathbf{E} \right\}. \end{aligned} \tag{23}$$

Taking into account notation Equation (20), we can rewrite this system in the form

$$\begin{aligned} & \frac{\partial \mathbf{H}^{(1)}(u, t)}{\partial t} + jua(x)\mathbf{H}^{(1)}(u, t) = \mathbf{H}^{(1)}(u, t) \left[\mathbf{A} + e^{ju}\mathbf{B} - x \left(\mathbf{I}_0 - e^{-ju}\mathbf{I}_1 \right) \right] \\ & + j\sigma \frac{\partial \mathbf{H}^{(1)}(u, t)}{\partial u} \left(\mathbf{I}_0 - e^{-ju}\mathbf{I}_1 \right), \\ & \frac{\partial \mathbf{H}^{(1)}(u, t)}{\partial u} \mathbf{E} + jua(x)\mathbf{H}^{(1)}(u, t)\mathbf{E} \\ & = \left(e^{ju} - 1 \right) \left[\mathbf{H}^{(1)}(u, t) \left(\mathbf{B} - e^{-ju}x\mathbf{I}_0 \right) + e^{-ju}j\sigma \frac{\partial \mathbf{H}^{(1)}(u, t)}{\partial u} \mathbf{I}_0 \right] \mathbf{E}. \end{aligned} \tag{24}$$

Denoting $\sigma = \varepsilon^2$ and making the following substitutions in Equation (24)

$$\tau = t\varepsilon^2, \quad u = \varepsilon w, \quad \mathbf{H}^{(1)}(u, t) = \mathbf{F}^{(1)}(w, \tau, \varepsilon), \tag{25}$$

we obtain system of differential equations

$$\begin{aligned} & \varepsilon^2 \frac{\partial \mathbf{F}^{(1)}(w, \tau, \varepsilon)}{\partial \tau} + j\varepsilon wa(x)\mathbf{F}^{(1)}(w, \tau, \varepsilon) = \mathbf{F}^{(1)}(w, \tau, \varepsilon) \left[\mathbf{A} + e^{j\varepsilon w}\mathbf{B} - x \left(\mathbf{I}_0 - e^{-j\varepsilon w}\mathbf{I}_1 \right) \right] \\ & + j\varepsilon \frac{\partial \mathbf{F}^{(1)}(w, \tau, \varepsilon)}{\partial w} \left(\mathbf{I}_0 - e^{-j\varepsilon w}\mathbf{I}_1 \right), \\ & \varepsilon^2 \frac{\partial \mathbf{F}^{(1)}(w, \tau, \varepsilon)}{\partial \tau} \mathbf{E} + j\varepsilon wa(x)\mathbf{F}^{(1)}(w, \tau, \varepsilon)\mathbf{E} \\ & = \left(e^{j\varepsilon w} - 1 \right) \left[\mathbf{F}^{(1)}(w, \tau, \varepsilon) \left(\mathbf{B} - e^{-j\varepsilon w}x\mathbf{I}_0 \right) + e^{-j\varepsilon w}j\varepsilon \frac{\partial \mathbf{F}^{(1)}(w, \tau, \varepsilon)}{\partial w} \mathbf{I}_0 \right] \mathbf{E}. \end{aligned} \tag{26}$$

Let us denote

$$\lim_{\varepsilon \rightarrow 0} \mathbf{F}^{(1)}(w, \tau, \varepsilon) = \mathbf{F}^{(1)}(w, \tau), \quad \lim_{\varepsilon \rightarrow 0} \frac{\partial \mathbf{F}^{(1)}(w, \tau, \varepsilon)}{\partial \tau} = \frac{\partial \mathbf{F}^{(1)}(w, \tau)}{\partial \tau}, \quad \lim_{\varepsilon \rightarrow 0} \frac{\partial \mathbf{F}^{(1)}(w, \tau, \varepsilon)}{\partial w} = \frac{\partial \mathbf{F}^{(1)}(w, \tau)}{\partial w}.$$

We can prove the following theorem.

Theorem 2. Function $\mathbf{F}^{(1)}(w, \tau)$ has the following form:

$$\mathbf{F}^{(1)}(w, \tau) = \Phi(w, \tau)\mathbf{R}(x), \tag{27}$$

where non-zero entries $R(n_1, n_2, x)$ of matrix $\mathbf{R}(x)$ are determined by Theorem 1, and scalar function $\Phi(w, \tau)$ is a solution of differential equation

$$\frac{\partial \Phi(w, \tau)}{\partial \tau} = a'(x)w \frac{\partial \Phi(w, \tau)}{\partial w} + b(x) \frac{(jw)^2}{2} \Phi(w, \tau). \tag{28}$$

Here function $a(x)$ is determined by Equation (20), $b(x)$ has the form

$$b(x) = a(x) + 2\mathbf{g}(x) (\mathbf{B} - x\mathbf{I}_0) \mathbf{E} + 2x\mathbf{R}(x)\mathbf{I}_0\mathbf{E}, \tag{29}$$

and top-left triangle matrix $\mathbf{g}(x)$ is determined by the following system of equations:

$$\begin{aligned} \mathbf{g}(x) [\mathbf{A} + \mathbf{B} + x (\mathbf{I}_1 - \mathbf{I}_0)] &= a(x)\mathbf{R}(x) + \mathbf{R}(x) (x\mathbf{I}_1 - \mathbf{B}), \\ \mathbf{g}(x)\mathbf{E} &= 0. \end{aligned} \tag{30}$$

Proof. Let us write only the members of the first equation of system Equation (26) that contains only zero or the first order of ε :

$$\begin{aligned} j\varepsilon w a(x)\mathbf{F}^{(1)}(w, \tau, \varepsilon) &= \mathbf{F}^{(1)}(w, \tau, \varepsilon) [\mathbf{A} + \mathbf{B} + j\varepsilon w\mathbf{B} - x (\mathbf{I}_0 - \mathbf{I}_1 - j\varepsilon w\mathbf{I}_1)] \\ &+ j\varepsilon \frac{\partial \mathbf{F}^{(1)}(w, \tau, \varepsilon)}{\partial w} (\mathbf{I}_0 - \mathbf{I}_1) + O(\varepsilon^2). \end{aligned} \tag{31}$$

Solution of this equation we write in the form

$$\mathbf{F}^{(1)}(w, \tau, \varepsilon) = \Phi(w, \tau) [\mathbf{R}(x) + j\varepsilon w\mathbf{f}(x)] + O(\varepsilon^2), \tag{32}$$

where $\Phi(w, \tau)$ and $\mathbf{f}(x)$ are some scalar and matrix functions whose expressions we will obtain later. Substituting this expression into Equation (31), we derive

$$\begin{aligned} j\varepsilon w a(x)\Phi(w, \tau)\mathbf{R}(x) &= \Phi(w, \tau) \{ \mathbf{R}(x) [\mathbf{A} + \mathbf{B} - x (\mathbf{I}_0 - \mathbf{I}_1)] \\ &+ j\varepsilon w [\mathbf{f}(x) (\mathbf{A} + \mathbf{B} - x(\mathbf{I}_0 - \mathbf{I}_1))] \} \\ &+ j\varepsilon \frac{\partial \Phi(w, \tau)}{\partial w} \mathbf{R}(x) (\mathbf{I}_0 - \mathbf{I}_1) + O(\varepsilon^2). \end{aligned} \tag{33}$$

Taking into account Equation (20), dividing Equation (33) by $j\varepsilon w\Phi(w, \tau)$ and tending $\varepsilon \rightarrow 0$, we obtain

$$a(x)\mathbf{R}(x) = \mathbf{f}(x) [\mathbf{A} + \mathbf{B} - x (\mathbf{I}_0 - \mathbf{I}_1)] + \mathbf{R}(x) (\mathbf{B} - x\mathbf{I}_1) + \frac{\partial \Phi(w, \tau) / \partial w}{w\Phi(w, \tau)} \mathbf{R}(x) (\mathbf{I}_0 - \mathbf{I}_1).$$

Let us rewrite this equality in the form of non-homogeneous equation

$$\mathbf{f}(x) [\mathbf{A} + \mathbf{B} + x (\mathbf{I}_1 - \mathbf{I}_0)] = a(x)\mathbf{R}(x) - \mathbf{R}(x) (\mathbf{B} - x\mathbf{I}_1) - \frac{\partial \Phi(w, \tau) / \partial w}{w\Phi(w, \tau)} \mathbf{R}(x) (\mathbf{I}_0 - \mathbf{I}_1). \tag{34}$$

According to superposition principle, we can write a solution of this equation in the form

$$\mathbf{f}(x) = \mathbf{C}\mathbf{R}(x) + \mathbf{g}(x) - \boldsymbol{\varphi}(x) \frac{\partial \Phi(w, \tau) / \partial w}{w\Phi(w, \tau)}. \tag{35}$$

Substituting it into Equation (34), we obtain equations

$$\boldsymbol{\varphi}(x) [\mathbf{A} + \mathbf{B} + x (\mathbf{I}_1 - \mathbf{I}_0)] = \mathbf{R}(x) (\mathbf{I}_0 - \mathbf{I}_1), \tag{36}$$

$$\mathbf{g}(x) [\mathbf{A} + \mathbf{B} + x (\mathbf{I}_1 - \mathbf{I}_0)] = a(x)\mathbf{R}(x) + \mathbf{R}(x) (x\mathbf{I}_1 - \mathbf{B}). \tag{37}$$

Notice that Equation (37) for matrix $\mathbf{g}(x)$ coincides with the first expression of Equation (30), therefore statement Equation (30) is true (the second expression of Equation (30) will be discussed later). Now, consider Equation (18). Let us differentiate it by x , we derive equation

$$\frac{d\mathbf{R}(x)}{dx} [\mathbf{A} + \mathbf{B} - x(\mathbf{I}_0 - \mathbf{I}_1)] - \mathbf{R}(x)(\mathbf{I}_0 - \mathbf{I}_1) = \mathbf{0}.$$

Taking into account Equation (36), we can conclude that

$$\boldsymbol{\varphi}(x) = \frac{d\mathbf{R}(x)}{dx}, \tag{38}$$

where $\boldsymbol{\varphi}(x)\mathbf{E} = \frac{d\mathbf{R}(x)}{dx}\mathbf{E} = \frac{d(\mathbf{R}(x)\mathbf{E})}{dx} = \mathbf{0}$.

Matrix $\mathbf{g}(x)$ is a particular solution to non-homogeneous system Equation (37), so it should satisfy some additional condition which we choose in the form $\mathbf{g}(x)\mathbf{E} = \mathbf{0}$ (as was declared in Equation (30)). Then solution $\mathbf{g}(x)$ of system Equation (37) is uniquely determined.

Now, let us consider the second Equation of Equation (26). We substitute expansion Equation (32) in it and write the members of derived equation including up to the second order of ε :

$$\varepsilon^2 \frac{\partial \Phi(w, \tau)}{\partial \tau} + j\varepsilon w a(x) \Phi(w, \tau) [1 + j\varepsilon w \mathbf{f}(x)] \mathbf{E} = \left[j\varepsilon w + \frac{(j\varepsilon w)^2}{2} \right] \left\{ \Phi(w, \tau) [\mathbf{R}(x) + j\varepsilon w \mathbf{f}(x)] (\mathbf{B} - x\mathbf{I}_0 + j\varepsilon w x \mathbf{I}_0) + j\varepsilon \frac{\partial \Phi(w, \tau)}{\partial w} \mathbf{R}(x) \mathbf{I}_0 \right\} \mathbf{E} + O(\varepsilon^3).$$

Then we derive

$$\begin{aligned} & \varepsilon^2 \frac{\partial \Phi(w, \tau)}{\partial \tau} + j\varepsilon w a(x) \Phi(w, \tau) + (j\varepsilon w)^2 a(x) \Phi(w, \tau) \mathbf{f}(x) \mathbf{E} = \\ & \Phi(w, \tau) \left\{ j\varepsilon w \mathbf{R}(x) [\mathbf{B} - x\mathbf{I}_0] + (j\varepsilon w)^2 \mathbf{f}(x) [\mathbf{B} - x\mathbf{I}_0] + (j\varepsilon w)^2 \mathbf{R}(x) x \mathbf{I}_0 \right. \\ & \left. + \frac{(j\varepsilon w)^2}{2} \mathbf{R}(x) [\mathbf{B} - x\mathbf{I}_0] + (j\varepsilon w)^2 \frac{\partial \Phi(w, \tau) / \partial w}{w} \mathbf{R}(x) \mathbf{I}_0 \right\} \mathbf{E} + O(\varepsilon^3). \end{aligned}$$

Applying Equation (20), we obtain

$$\begin{aligned} & \varepsilon^2 \frac{\partial \Phi(w, \tau)}{\partial \tau} + (j\varepsilon w)^2 a(x) \Phi(w, \tau) \mathbf{f}(x) \mathbf{E} = \\ & \Phi(w, \tau) \left\{ (j\varepsilon w)^2 \mathbf{f}(x) [\mathbf{B} - x\mathbf{I}_0] + (j\varepsilon w)^2 \mathbf{R}(x) x \mathbf{I}_0 + \frac{(j\varepsilon w)^2}{2} a(x) \right. \\ & \left. + (j\varepsilon w)^2 \frac{\partial \Phi(w, \tau) / \partial w}{w} \mathbf{R}(x) \mathbf{I}_0 \right\} \mathbf{E} + O(\varepsilon^3). \end{aligned}$$

Let us divide this equation by ε^2 :

$$\begin{aligned} \frac{\partial \Phi(w, \tau)}{\partial \tau} = \Phi(w, \tau) & \left\{ (jw)^2 \mathbf{f}(x) [\mathbf{B} - x\mathbf{I}_0] + (jw)^2 \mathbf{R}(x) x \mathbf{I}_0 + \frac{(jw)^2}{2} a(x) \right. \\ & \left. - (jw)^2 a(x) \mathbf{f}(x) + (jw)^2 \frac{\partial \Phi(w, \tau) / \partial w}{w} \mathbf{R}(x) \mathbf{I}_0 \right\} \mathbf{E}. \end{aligned}$$

So, we derive

$$\frac{\partial\Phi(w, \tau)/\partial\tau}{\Phi(w, \tau)} = \frac{(jw)^2}{2} \{2 [\mathbf{f}(x) (\mathbf{B} - x\mathbf{I}_0) + \mathbf{R}(x)x\mathbf{I}_0 - a(x)\mathbf{f}(x)] \mathbf{E} + a(x)\} - w \frac{\partial\Phi(w, \tau)/\partial w}{\Phi(w, \tau)} \mathbf{R}(x)\mathbf{I}_0\mathbf{E}.$$

Substituting Equation (35) here, we obtain

$$\begin{aligned} \frac{\partial\Phi(w, \tau)/\partial\tau}{\Phi(w, \tau)} &= \frac{(jw)^2}{2} \{2\mathbf{g}(x) (\mathbf{B} - x\mathbf{I}_0) \mathbf{E} + 2\mathbf{R}(x)x\mathbf{I}_0\mathbf{E} + a(x)\} \\ &+ w \frac{\partial\Phi(w, \tau)/\partial w}{\Phi(w, \tau)} \{\boldsymbol{\varphi}(x) (\mathbf{B} - x\mathbf{I}_0) \mathbf{E} - \mathbf{R}(x)\mathbf{I}_0\mathbf{E}\}. \end{aligned} \tag{39}$$

Denoting

$$b(x) = a(x) + 2\mathbf{g}(x) (\mathbf{B} - x\mathbf{I}_0) \mathbf{E} + 2x\mathbf{R}(x)\mathbf{I}_0\mathbf{E}, \tag{40}$$

we can rewrite Equation (39) in the form

$$\frac{\partial\Phi(w, \tau)}{\partial\tau} = w \frac{\partial\Phi(w, \tau)}{\partial w} \{\boldsymbol{\varphi}(x) (\mathbf{B} - x\mathbf{I}_0) \mathbf{E} - \mathbf{R}(x)\mathbf{I}_0\mathbf{E}\} + \frac{(jw)^2}{2} b(x)\Phi(w, \tau). \tag{41}$$

Differentiating Equation (20) by x , we obtain

$$a'(x) = \frac{\partial\mathbf{R}(x)}{\partial x} (\mathbf{B} - x\mathbf{I}_0) \mathbf{E} - \mathbf{R}(x)\mathbf{I}_0\mathbf{E}. \tag{42}$$

Then, taking into account Equation (38) and substituting Equation (42) into Equation (41), we derive the equation

$$\frac{\partial\Phi(w, \tau)}{\partial\tau} = a'(x)w \frac{\partial\Phi(w, \tau)}{\partial w} + b(x) \frac{(jw)^2}{2} \Phi(w, \tau)$$

that coincides with Equation (28). So, the theorem is proved. \square

Corollary 2. Let $g_1(x) = \sum_{n_1+n_2=N} g(n_1, n_2, x)$, $r(x) = \sum_{n_1+n_2=N} R(n_1, n_2, x)$, then function $b(x)$ from Equation (40) may be written in the form

$$b(x) = a(x) + 2[x(1 - r(x)) + (\lambda + x)g_1(x)]. \tag{43}$$

Proof. Let us rewrite Equation (40) in scalar form:

$$\begin{aligned} b(x) &= a(x) + 2 \left[\lambda \sum_{n_1+n_2=N} g(n_1, n_2, x) + x \sum_{n_1+n_2 \leq N-1} g(n_1, n_2, x) - x \sum_{n_1+n_2 \leq N-1} R(n_1, n_2, x) \right] = \\ &a(x) + 2 \left[\lambda g_1(x) + x \left\{ 1 - x \sum_{n_1+n_2=N} R(n_1, n_2, x) \right\} - x \left\{ \sum_{n_1+n_2 \leq N} g(n_1, n_2, x) - g_1(x) \right\} \right]. \end{aligned}$$

Applying here condition $\mathbf{g}(x)\mathbf{E} = 0$, we obtain Equation (43). \square

In the next section, we will use obtained functions $a(x)$ and $b(x)$ as a drift and diffusion coefficients of a diffusion process which we apply to derive approximations for probability distributions under study.

5. Method of Asymptotic Diffusion Analysis

Taking into account notation $\sigma = \varepsilon^2$ and substitutions Equation (25) from the previous section, we consider stochastic process

$$\sqrt{\sigma} \left[i(\sigma t) - \frac{1}{\sigma} x(\sigma t) \right] = \sqrt{\sigma} \left[i(\tau) - \frac{1}{\sigma} x(\tau) \right]$$

under asymptotic condition $\sigma \rightarrow 0$. Here $i(\tau)$ is the number of customers in the orbit and function $x(\tau)$ is determined in the previous section as a solution of differential equation $x'(\tau) = a(x)$. We can prove the following statement.

Lemma 1. *Asymptotic stochastic process*

$$y(\tau) = \lim_{\sigma \rightarrow 0} \sqrt{\sigma} \left[i(\tau) - \frac{1}{\sigma} x(\tau) \right]$$

is a solution of stochastic differential equation

$$dy(\tau) = a'(x)y d\tau + \sqrt{b(x)} dw(\tau) \tag{44}$$

that depends on continuous parameter x .

Proof. Consider Equation (28) with $a(x)$ and $b(x)$ determined by Equations (22) and (40). Solution of this equation determines asymptotic characteristic function for centered and normalized stochastic process $\sqrt{\sigma} \left\{ i(\tau) - \frac{1}{\sigma} x(\tau) \right\}$ of the number $i(t)$ of customers in the orbit under condition $\sigma \rightarrow 0$.

Let us make in this equation an inverse Fourier transform on argument w . Then for probability density function (p.d.f.) $P(y, \tau) = \frac{\partial P \{y(\tau) < y\}}{\partial y}$ of process $y(\tau)$, we obtain the equation

$$\frac{\partial P(y, \tau)}{\partial \tau} = -\frac{\partial}{\partial y} \{a'(x)yP(y, \tau)\} + \frac{1}{2} \frac{\partial^2}{\partial y^2} \{b(x)P(y, \tau)\},$$

which is the Fokker–Planck equation for p.d.f. $P(y, \tau)$. Hence, stochastic process $y(\tau)$ is a diffusion process with drift coefficient $a'(x)y$ and diffusion coefficient $b(x)$. Therefore, process $y(\tau)$ is a solution of stochastic differential Equation (44). □

Let us consider the following stochastic process:

$$z(\tau) = x(\tau) + \varepsilon y(\tau) \tag{45}$$

(here $\varepsilon = \sqrt{\sigma}$ as before.)

Lemma 2. *Stochastic process $z(\tau)$ is a solution to stochastic differential equation*

$$dz(\tau) = a(z)d\tau + \sqrt{\sigma b(z)} dw(\tau) \tag{46}$$

with a precision up to an infinitesimal of order ε^2 .

Proof. Due to $x(\tau)$ is a solution of differential equation $dx(\tau) = a(x)d\tau$ and process $y(\tau)$ satisfies Equation (44), the following equality is true:

$$dz(\tau) = d[x(\tau) + \varepsilon y(\tau)] = [a(x) + \varepsilon y a'(x)] d\tau + \varepsilon \sqrt{b(x)} dw(\tau). \tag{47}$$

We can represent its coefficients in the form

$$a(x) + \varepsilon ya'(x) + O(\varepsilon^2) = a(x + \varepsilon y) = a(z),$$

$$\varepsilon \sqrt{b(x)} + O(\varepsilon^2) = \varepsilon \sqrt{b(x) + O(\varepsilon)} = \varepsilon \sqrt{b(x + \varepsilon y)} = \varepsilon \sqrt{b(z)},$$

and, so, we can rewrite Equation (47) as follows:

$$dz(\tau) = a(z)d\tau + \sqrt{\sigma b(z)}dw(\tau) + O(\varepsilon^2).$$

It coincides with Equation (46) with a precision up to infinitesimal $O(\varepsilon^2)$. \square

Suppose that the system is in steady-state regime. Consider stationary p.d.f. $s(z)$ for process $z(\tau)$:

$$s(z, \tau) = s(z) = \frac{\partial P \{z(\tau) < z\}}{\partial z}.$$

Let us prove the following statement.

Theorem 3. Stationary p.d.f. $s(z)$ of stochastic process $z(\tau)$ has the form

$$s(z) = \frac{C}{b(z)} \exp \left\{ \frac{2}{\sigma} \int_0^z \frac{a(x)}{b(x)} dx \right\}, \tag{48}$$

where C is a normalizing constant.

Proof. Due to $z(\tau)$ is a solution of stochastic differential Equation (46), the process is diffusion with drift coefficient $a(z)$ and diffusion coefficient $b(z)$. Therefore, its p.d.f. $s(z)$ is a solution of the Fokker–Planck equation

$$-\frac{\partial}{\partial z} \{a(z)s(z)\} + \frac{1}{2} \frac{\partial^2}{\partial z^2} \{\sigma b(z)s(z)\} = 0.$$

This equation is an ordinary differential equation of the second order. Solving it and taking into account normalization condition and boundary condition $s(\infty) = 0$, we obtain Equation (48). The theorem is proved. \square

6. Approximations of the Stationary Distributions

Consider matrix characteristic function $\mathbf{F}^{(1)}(w, \tau)$ (Theorem 2) of the joint distribution of the number of customers in the orbit and the number of servers working in the first and at the second phases. Due to Equation (27), it equals to production of characteristic function $\Phi(w, \tau)$ of the number of customers in the orbit and matrix $\mathbf{R}(x)$ that represents two-dimensional probability distribution of the number of servers working in the first and at the second phases. Therefore, we can represent stationary probability distribution under study

$$P\{n_1(t) = n_1, n_2(t) = n_2, i(t) = i\} = P(n_1, n_2, i, t)$$

in the form of production

$$P(n_1, n_2, i, t) = R(n_1, n_2)P(i),$$

where $R(n_1, n_2)$ is the joint two-dimensional stationary distribution of the number of servers working in the first and in the second phases of hyper-exponential service time Equation (1), $P(i)$ is the stationary distribution of the number of customers in the orbit.

To construct an approximation $\hat{P}(i)$ for discrete probability distribution $P(i)$, consider p.d.f. $s(z)$ of stochastic process $z(\tau)$. In the non-asymptotic case we can represent $z(\tau)$ as follows:

$$z(\tau) = x(\tau) + \varepsilon y(\tau) = x(\tau) + \sqrt{\sigma} \left\{ \sqrt{\sigma} \left[i(\tau) - \frac{1}{\sigma} x(\tau) \right] \right\} = x(\tau) + \sigma i(\tau) - x(\tau) = \sigma i(\tau). \quad (49)$$

There are various ways to construct probability distribution $\hat{P}(i)$ of discrete process $\sigma i(\tau)$ on the base of p.d.f. $s(z)$ of continuous stochastic process $z(\tau)$. We will use the following one.

Using Equations (48) and (49), we write non-negative function $G(i)$ of discrete argument $i \in \{0, 1, \dots\}$ in the form

$$G(i) = \frac{C}{b(\sigma i)} \exp \left\{ \frac{2}{\sigma} \int_0^{\sigma i} \frac{a(x)}{b(x)} dx \right\}. \quad (50)$$

Taking into account normalization condition, we can write

$$\hat{P}(i) = \frac{G(i)}{\sum_{i=0}^{\infty} G(i)}. \quad (51)$$

This probability distribution we will use as an approximation for probability distribution $P(i) = P\{i(t) = i\}$ of the number of customers in the orbit of considered retrial queue in steady-state regime.

We can construct two different approximations for stationary probabilities $R(n)$ of the number of busy servers. The first one can be build on the base of Formulas (13) and (14). In stationary regime, differential Equation (15) transforms to the equality

$$a(x) = 0,$$

which is an equation for variable x . Let us denote its positive solution by κ . Substituting $x = \kappa$ into Equation (13), we obtain the following expression for calculation of the first-type approximation $\hat{R}_1(n_1, n_2)$ of probabilities $R(n_1, n_2)$:

$$\hat{R}_1(n_1, n_2) = R(n_1, n_2, \kappa). \quad (52)$$

Approximation Equation (51) allows us determine another form of approximation $\hat{R}_2(n_1, n_2)$ for probability distribution $R(n_1, n_2)$ of the number of servers working in the first and in the second phases:

$$\hat{R}_2(n_1, n_2) = \sum_{i=0}^{\infty} R(n_1, n_2, \sigma i) \hat{P}(i), \quad (53)$$

where $R(n_1, n_2, x)$ is calculated using Equation (13).

So, basing on these two types of approximations, we can write two types $\hat{R}_1(n)$ and $\hat{R}_2(n)$ of approximations for the stationary probability distribution $R(n)$ of the number of busy servers as follows:

$$\hat{R}_k(n) = \sum_{\nu=0}^n R_k(\nu, n - \nu), \quad k = 1, 2. \quad (54)$$

The accuracy of the constructed approximations and their applicability area will be discussed in the next section.

7. Applicability Area of Obtained Approximations

To estimate the accuracy of obtained approximations Equations (51)–(54) we perform numerical experiments with usage of simulation modeling. We use the Kolmogorov distance

$$\Delta = \max_{0 \leq i < \infty} \left| \sum_{\nu=0}^i [P_1(\nu) - P_2(\nu)] \right| \tag{55}$$

as an error estimation. Here $P_1(\nu)$ and $P_2(\nu)$ are probability distributions of two discrete variables. For our purposes, one of them will be an approximation from $\hat{P}(i)$, $\hat{R}_1(n)$ or $\hat{R}_2(n)$, and another will be an empiric distribution of the same parameter constructed on the base of simulation results. Simulation model and simulation parameters (e.g., volume of collected statistical data) are chosen in a such way to error of simulation be not have a significant influence on calculation of value Equation (55) for approximations under test.

We consider values of Kolmogorov distance Equation (55) less or equal to 0.05 as enough to an approximation be enough accurate. We reach $\Delta \leq 0.001$ for simulation experiments themselves (when both $P_1(\nu)$ and $P_2(\nu)$ in Equation (55) are taken from simulation results of two experiments with identical parameters).

Special developed software based on ODIS system [23,24] is used to perform simulations. Numerical calculations of probabilities Equations (51)–(54) are made using PTC Mathcad application.

For our purposes of analyzing applicability area of results Equations (51)–(54), let us consider a retrial queue with five servers ($N = 5$) and hyper-exponential service time with parameters

$$\mu_1 = 0.5, \mu_2 = 3, q_1 = 0.4, q_2 = 0.6.$$

Let us denote load of the system by ρ ($0 < \rho < 1$), and value of arrival process intensity λ we determine as $\lambda = \rho N$.

In Table 1, you may find Kolmogorov distances Δ_k for approximations \hat{R}_k from Equation (54) calculated for various values of load parameter ρ and intensity of retrials σ . As we see, while σ is decreasing, the accuracy of the approximations becomes better (Δ_k is decreasing). We highlight values of Kolmogorov distance that satisfy chosen applicability condition $\Delta \leq 0.05$ by boldface font. We can not conclude what approximation from these is better in all cases but the first one seems more preferable in the case of $\rho \leq 0.7$ and the second one is more preferable for $\rho > 0.7$. Furthermore, we may notice that for $\sigma < 1$ both of them are applicable for all cases $\rho \geq 0.6$, and the second approximation is applicable for greater values σ for narrower area ($\rho \geq 0.9$).

Table 1. Values of Kolmogorov distances Δ_k of approximations $\hat{R}_k(n)$ from Equation (54) for various values of σ and ρ .

ρ	σ	2	1	0.5	0.2	0.1
0.6	Δ_1	0.044	0.032	0.020	0.011	0.006
	Δ_2	0.059	0.053	0.042	0.019	0.008
0.7	Δ_1	0.065	0.046	0.030	0.015	0.009
	Δ_2	0.083	0.063	0.038	0.013	0.003
0.8	Δ_1	0.078	0.056	0.037	0.019	0.011
	Δ_2	0.075	0.050	0.025	0.005	0.002
0.9	Δ_1	0.073	0.049	0.030	0.015	0.008
	Δ_2	0.045	0.021	0.008	0.001	0.001

The similar results for approximation of the number of customers in the orbit Equation (51) are presented in Table 2. As in the previous case, this approximation became more accurate while σ is decreasing too. In addition, this approximation became more precise almost always while ρ is growing.

We can assume that it is applicable for all $\sigma < 2$ while $\rho \geq 0.9$ (boldface is used to highlight applicable cases again).

Table 2. Values of Kolmogorov distance of approximation \hat{P} for various values of σ and ρ .

σ	2	1	0.5	0.2	0.1
$\rho = 0.6$	0.085	0.129	0.141	0.082	0.039
$\rho = 0.7$	0.122	0.122	0.081	0.027	0.014
$\rho = 0.8$	0.104	0.070	0.032	0.011	0.014
$\rho = 0.9$	0.052	0.022	0.003	0.012	0.010

8. Call Center Example

Consider an example which illustrates application of obtained results in real system. Some call center has five workstations for operators that service calls from customers. Ordinary call requires time to be serviced equal to $1/m$ in average, where m is some measure of the average productivity of one operator and its workstation. For such calls we use exponentially distributed service time with rate $\mu_2 = m$. Some calls (about 20%) require long time to be serviced. We model them using rate $\mu_1 = 0.2m$ for exponential distribution. So, we can generally model service time in our system as hyper-exponential with parameters $\mu_1 = 0.2m, \mu_2 = m, q_1 = 0.2, q_2 = 0.8$. Let arrival process be Poisson with intensity $\lambda = 2$. If a customer find the system busy (there is no free operator), he or she tries to make a new call after a random time that we suppose exponentially distributed with parameter $\sigma = 0.2$. So, we can consider a model of the call center in the form of multi-server retrial queue with hyper-exponential service time and mentioned parameters.

Let us consider that the call center operation requires some cost and has some penalties. We suppose that we can improve a performance of the call center by increasing an average speed of calls processing m (e.g., involving more qualified personal or deploying better equipment) but we should pay a cost equals to $m \cdot C_s$, where C_s is some constant. Furthermore, we have some losses due to customers that should make repeated calls. We calculate them as $\bar{i} \cdot C_E$ where C_E is some constant and $\bar{i} = \sum_{i=0}^{\infty} i \cdot P_i$ is mathematical expectation of the number of calls in the orbit. In addition, if we have too many "waiting" customers we obtain a penalty which we calculate as $C_L \cdot \sum_{i=w+1}^{\infty} P_i$, where C_L is some constant and w is some upper bound after which calls in the orbit are taken into account for the penalty calculation. Thus we can write the total cost of the call center operation K in the following form:

$$K(m) = m \cdot C_s + C_E \sum_{i=0}^{\infty} i \cdot P_i + C_L \sum_{i=w+1}^{\infty} P_i.$$

Using the following values of constants

$$C_s = 3, \quad C_E = 0.04, \quad C_L = 2,$$

and varying value of m in the interval from 0.75 to 1.2 (with step 0.05) that corresponds to values of the system load ρ from 0.96 to 0.6 respectively, we obtain values of the total cost $K(m)$ for the cases $w = 10, w = 20$ and $w = 50$ (Figure 1). As we see, in all these cases function $K(m)$ has a minimum in the considered range of m . Therefore, when we plan work of the call center we should take into account that increasing of performance of operators and workstations does not always cause the cost be less. Moreover, value of m for the minimum point depends on value of upper bound w .

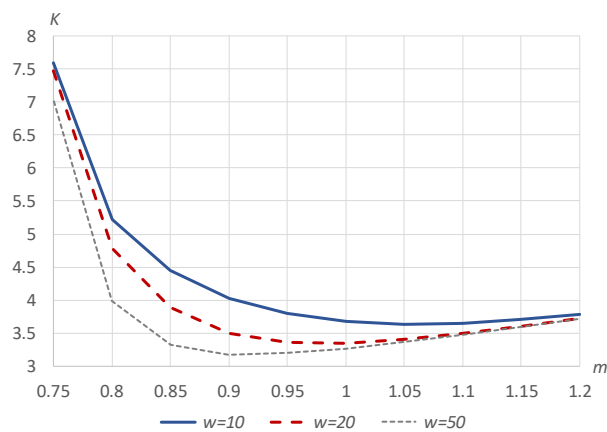


Figure 1. Values of the total cost $K(m)$ for various values of bound w .

9. Discussion

In this paper, we apply a new approach [22] for studying multi-server retrial queue with hyper-exponential service time. Analysis of the model is made under asymptotic condition of long delay in the orbit. Using asymptotic analysis technique [18,19], we obtain parameters of diffusion process Equation (45) which is used to build approximations for stationary probability distributions of the number of customers in the orbit Equation (51) and the number of busy servers Equations (52)–(54).

We use numerical experiments and simulations for estimation of the approximations accuracy. Using Kolmogorov distance Δ Equation (55) as an error of approximation and criteria of approximation applicability in the form $\Delta \leq 0.05$, we obtain applicability areas of proposed approximations. As it shown, these areas are wide enough. At least in the most of cases, they are wider than applicability areas of approximations obtained using basic asymptotic analysis technique as it was made in [18–21].

We suppose to use asymptotic diffusion method and obtained results in future works that will be devoted as to analysis of retrial queues with other configurations both to study of multi-server retrial queue with arbitrary distributed service time. Using hyper-exponential distribution as an approximation for distributions of other types may be the first step in this direction. Further studies may involve supplementary variables technique or using Markov chains based on departure epochs.

We thank the reviewers of this paper whose comments allow us improve the work.

Author Contributions: Conceptualization, A.N.; Formal analysis, A.M.; Investigation, A.M., A.N. and S.P.; Methodology, A.N.; Software, A.M.; Writing—original draft, S.P.; Writing—review & editing, A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References

1. Aguir, S.; Karaesmen, F.; Askin, O.Z.; Chauvet, F. The impact of retrials on call center performance. *OR Spektrum* **2004**, *26*, 353–376. [[CrossRef](#)]
2. Stolletz, R. *Performance Analysis and Optimization of Inbound Call Centers*; Springer: Berlin/Heidelberg, Germany, 2003. [[CrossRef](#)]
3. Avrachenkov, K.; Yechiali, U. Retrial networks with finite buffers and their application to internet data traffic. *Probab. Eng. Inf. Sci.* **2008**, *22*, 519–536.
4. Roszik, J.; Sztrik, J.; Kim, C.S. Retrial queues in the performance modelling of cellular mobile networks using MOSEL. *Int. J. Simul.* **2005**, *6*, 38–47. [[CrossRef](#)]
5. Artalejo, J.R.; Lopez-Herrero, M.J. Cellular mobile networks with repeated calls operating in random environment. *Comput. Oper. Res.* **2010**, *37*, 1158–1166.

6. Ke, X.; Shi, L.; Guo, W.; Chen, D. Multi-Dimensional Traffic Congestion Detection Based on Fusion of Visual Features and Convolutional Neural Network. *IEEE Trans. Intell. Trans. Syst.* **2019**, *20*, 2157–2170. [[CrossRef](#)]
7. Nazarov, A.A.; Moiseev, A.N. Distributed System of Processing of Data of Physical Experiments. *Rus. Phys. J.* **2014**, *57*, 984–990. [[CrossRef](#)]
8. Phung-Duc, T. Retrial Queueing Models: A Survey on Theory and Applications. *arXiv* **2019**, arXiv:1906.09560v1. [[CrossRef](#)]
9. Falin, G.I.; Templeton, J.G.C. *Retrial Queues*; Chapman & Hall: London, UK, 1997.
10. Falin, G.I. A survey of retrial queues. *Queueing Syst.* **1990**, *7*, 127–168.
11. Artalejo, J.R.; Gomez-Corral, A. *Retrial Queueing Systems: A Computational Approach*; Springer: Berlin/Heidelberg, Germany, 2008. [[CrossRef](#)]
12. Shekhar, C.; Raina, A.A.; Kumar, A. A brief review on retrial queue: Progress in 2010–2015. *Int. J. Appl. Sci. Eng. Res.* **2016**, *5*, 324–336.
13. Kim, C.; Mushko, V.; Dudin, A. Computation of the steady state distribution for multi-server retrial queues with phase type service process. *Ann. Oper. Res.* **2012**, *201*, 307–323.
14. Phung-Duc, T.; Masuyama, H.; Kasahara, S.; Takahashi, Y. A matrix continued fraction approach to multiserver retrial queues. *Ann. Oper. Res.* **2013**, *203*, 161–183. [[CrossRef](#)]
15. Neuts, M.; Rao, B. Numerical investigation of a multiserver retrial mode. *Queueing Syst.* **2002**, *7*, 169–189. [[CrossRef](#)]
16. Artalejo, J.; Pozo, M. Numerical calculation of the stationary distribution of the main multiserver retrial queue. *Ann. Oper. Res.* **2002**, *116*, 41–56. [[CrossRef](#)]
17. Ridder, A. Fast simulation of retrial queues. In Proceedings of the 3rd Workshop on Rare Event Simulation and Related Combinatorial Optimization Problems, Pisa, Italy, 5–6 October 2000. [[CrossRef](#)]
18. Moiseeva, E.; Nazarov, A. Asymptotic analysis of RQ-systems M/M/1 on heavy load condition. In Proceedings of the IV International Conference Problems of Cybernetics and Informatics, Baku, Azerbaijan, 12–14 September 2012; pp. 164–166.
19. Fedorova, E. Quasi-geometric and gamma approximation for retrial queueing systems. *Commun. Comput. Inf. Sci.* **2014**, *487*, 123–136.
20. Nazarov, A.; Chernikova, Y. Gaussian approximations of probabilities distribution of states of the retrial queueing system with r-persistent exclusion of alternative customers. *Commun. Comput. Inf. Sci.* **2015**, *564*, 200–209.
21. Lapatin, I.; Nazarov, A. Asymptotic Analysis of the Output Process in Retrial Queue with Markov-Modulated Poisson Input Under Low Rate of Retrials Condition. *Commun. Comput. Inf. Sci.* **2019**, *1141*, 315–324.
22. Nazarov, A.; Phung-Duc, T.; Paul, S.; Lizura, O. Asymptotic-Diffusion Analysis for Retrial Queue with Batch Poisson Input and Multiple Types of Outgoing Calls. *Lect. Notes Comput. Sci.* **2019**, *11965*, 207–222.
23. Moiseev, A.; Demin, A.; Dorofeev, V.; Sorokin, V. Discrete-event approach to simulation of queueing networks. *Key Eng. Mater.* **2016**, *685*, 939–942.
24. Mesheryakov, R.; Moiseev, A.; Demin, A.; Dorofeev, V.; Sorokin, V. Using parallel computing in queueing network simulation. *Key Eng. Mater.* **2016**, *685*, 943–947. [[CrossRef](#)]

