

Article

# Retrial Queues with Unreliable Servers and Delayed Feedback

Agassi Melikov <sup>1</sup>, Sevinj Aliyeva <sup>2</sup> and Janos Sztrik <sup>3,\*</sup>

<sup>1</sup> Institute of Control Systems, National Academy of Science, Baku AZ 1141, Azerbaijan; agassi.melikov@gmail.com

<sup>2</sup> Faculty of Applied Mathematics and Cybernetics, Baku State University, Baku AZ 1148, Azerbaijan; s@aliyeva.info

<sup>3</sup> Department of Informatics and Networks, Faculty of Informatics, University of Debrecen, 4032 Debrecen, Hungary

\* Correspondence: sztrik.janos@inf.unideb.hu

**Abstract:** In this paper, models of unreliable multi-server retrial queues with delayed feedback are examined. The Bernoulli retrial is allowed upon the arrival of both primary (from outside) and feedback customers (from orbit), as well as the Bernoulli feedback that may occur after each service in this system. Servers can break down both during the service of customers and when they are idle. If a server breaks down during the service of a customer, then the interrupted customer, in accordance with the Bernoulli scheme, decides either to leave the system or join a common orbit of retrial and feedback customers. An approximate method, based on the space merging approach of three-dimensional Markov chains, is proposed for the calculation of the steady-state probabilities, as well as performance measures of the system. The results of the numerical experiments are demonstrated.

**Keywords:** multi-server queue; unreliable servers; retrial customers; delayed feedback; finite and infinite orbits; balking from orbit; calculation method



**Citation:** Melikov, A.; Aliyeva, S.; Sztrik, J. Retrial Queues with Unreliable Servers and Delayed Feedback. *Mathematics* **2021**, *9*, 2415. <https://doi.org/10.3390/math9192415>

Academic Editor: Juan Eloy Ruiz-Castro

Received: 19 August 2021  
Accepted: 23 September 2021  
Published: 28 September 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

The theory of retrial queue (RQ) is widely implemented for modeling the functioning of real-world systems. State of the art theory of RQ might be found in the recent review [1]. The general property of RQ is the following: if upon arrival of a primary customer, all servers of the system and/or waiting rooms are full, then a primary customer can join the group of customers that are waiting for the service outside of the system. This virtual waiting room is called orbit for retrial customers. As a rule, in the classical theory of RQ, it is assumed that the system has no information about the number of retrial customers in the orbit, and retrial customers generate their requests independently of each other.

In classical RQ, both primary and retrial customers leave the system after getting the required service. However, in many cases, some customers require an arbitrary number of optional services aside from the essential services. In such systems, some customers provide feedback to the system for optional service. These customers are referred to as feedback customers. Note that a customer can provide feedback either instantaneously after leaving the server or after some delay. Retrial queues with delayed feedback (RQwDFB) are very useful tools for modeling stochastic processes arising in communication networks, call centers, queueing-inventory systems, etc. This paper analyzes the steady-state behavior of an unreliable multi-server retrial queueing system, first with essential services and then with an arbitrary number of optional services.

## 2. Related Work

To see the importance of RQwDFB, the interested reader is referred to papers [2–9], and their references lists.

In the above-mentioned papers, the models of RQwDFB with reliable servers are investigated. However, in practice, the servers of the systems are subject to random

breakdowns; therefore, it is necessary to study the model of RQwDFB with unreliable servers. Models of RQwDFB with unreliable servers have been investigated in a small number of papers in the last few years, see [10–17]. In the indicated papers, single-server models are considered.

Note that the models of unreliable multi-server RQ without feedback phenomena were investigated using the Space Merging Method (SMM) in [18–20]. Similar models have been investigated in [21] using an asymptotic approach. In indicated papers, it is assumed that the parameters of the Bernoulli trial that determine decisions for primary customers to join the orbit or leave the system over the absence of free servers are state-independent. Further, parameters of the Bernoulli trial that determine decisions for customers to enter the orbit that are preempted due to a server failure or abandoning the service are state-independent. Moreover, in known models, as a rule, it is assumed that both active and idle breakdowns occur at the same rate. These assumptions essentially restrict the application areas of the obtained results.

To our best knowledge, models of unreliable multi-server RQwDFB have not been studied. In this paper, models of unreliable multi-server RQwDFB are investigated under more general assumptions using SMM. The method is the same; however, the system is different, and the parameters of the Bernoulli trial that determine appropriate decisions in the system are state-dependent for the model with a finite orbit size. The last assumption leads to more complex equations and solutions. For historical justice, it is necessary to note that the SMM was initially used to calculate the steady-state probabilities of the two-dimensional Markov chain (2D MC) in [22]. In the last decade, [23–25] have developed similar algorithms to study 2D MC in various areas of application. In all papers, the authors note the high efficiency of SMM (both in the sense of accuracy and complexity) in comparison with other numerical methods.

This paper is structured as follows: Basic mathematical models of the investigated multi-server unreliable RQwDFB are described in Section 3. The generator matrix of the appropriate 2D MC that represents the mathematical model of the system with a finite orbit size is created in Section 4. Here, the balance equations method of calculating its steady-state probabilities is indicated, and explicit formulas for the performance measures of the investigated system are developed, as well. In Section 5, the approximate SMM for calculating steady-state probabilities and performance measures of the system with an infinite orbit size is proposed. The results of numerical experiments that show the behavior of performance measures versus system parameters are demonstrated in Section 6. Conclusions are given in Section 7.

### 3. Model Description

A pictorial representation of the investigated RQwDFB with unreliable servers is shown in Figure 1.

First, consider the model with a finite capacity of orbit for retrial customers, i.e., assume that the maximal size of orbit is equal to  $R$ ,  $R < \infty$ . The system contains  $K > 1$  identical and unreliable servers. It is assumed that servers can independently break down in both busy and idle modes. The failure rate of servers is different in their various statuses, i.e., failures of servers in both busy and idle statuses occur via two independent Poisson processes with rates of  $\theta_1$  and  $\theta_2$ , respectively. Each server also has its own repairman and the repair times for servers are exponentially distributed with the common rate  $\zeta$ . If the failure of a server occurs during the busy status (active breakdown), then a customer that is preempted due to a server failure either leaves the system (i.e., is lost) with a probability (w.p.)  $\gamma_r$  or enters the retrial orbit w.p.  $1 - \gamma_r$ , provided that there are  $r$  customers in the orbit,  $r = 0, 1, \dots, R - 1$ ;  $\gamma_R = 1$ .

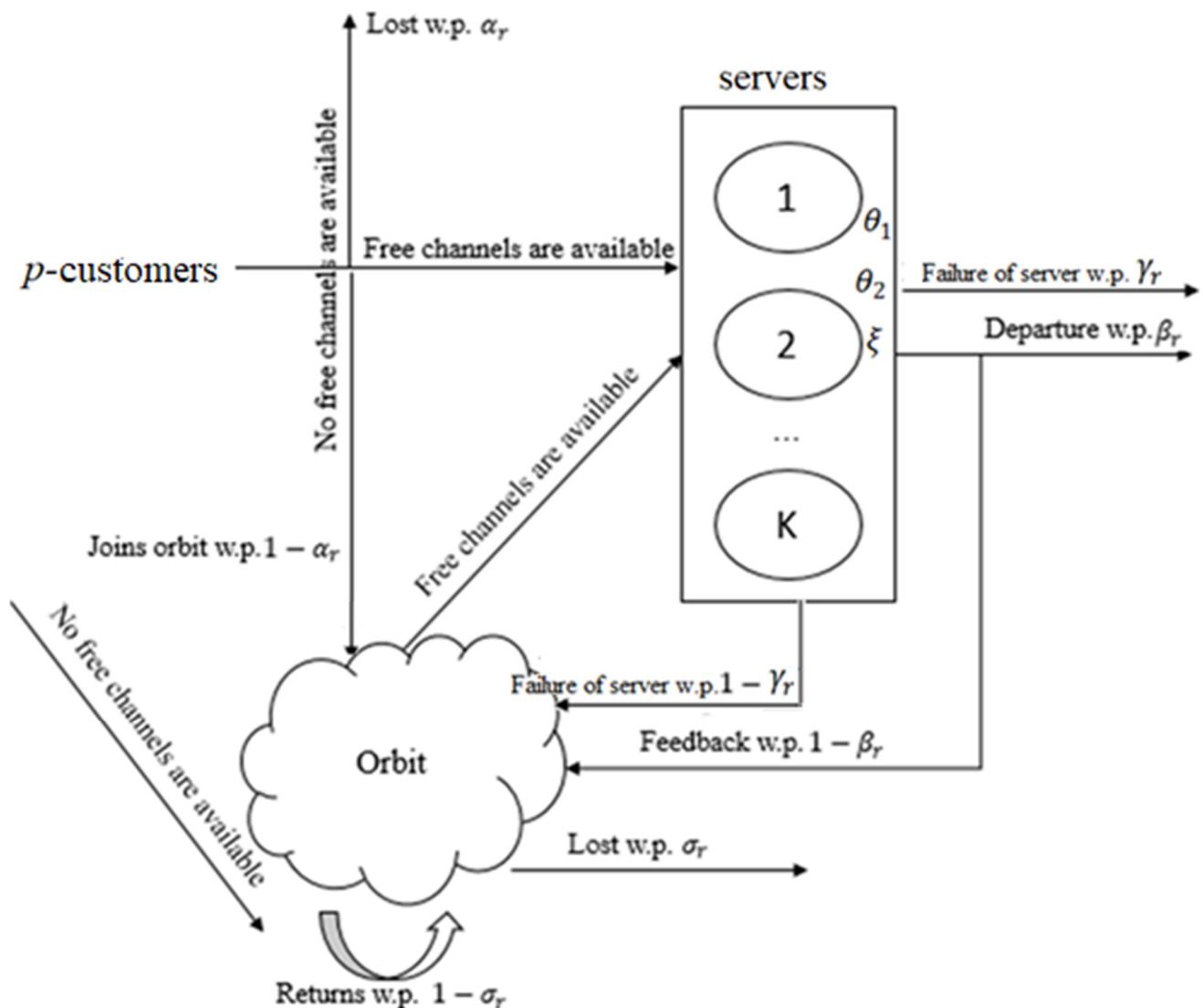


Figure 1. A graphical representation of the unreliable multi-server RQwDFB under investigation.

Primary customers ( $p$ -customers) form Poisson flow with rate  $\lambda$ . An incoming  $p$ -customer is accepted for service if at this moment the total number of busy and failed servers is less than  $K$ ; otherwise, they either leave the system w.p.  $\alpha_r$  or go to orbit w.p.  $1 - \alpha_r$  provided that there are  $r$  customers in the orbit,  $r = 0, 1, \dots, R - 1$ ;  $\alpha_R = 1$ .

Sojourn times of retrial customers ( $r$ -customers) in the orbit are exponentially distributed with a mean  $\eta^{-1}$  and here a classical retrial scheme is considered, i.e., if there are  $r$  customers in the orbit, then the arrival intensity of  $r$ -customers is  $r\eta$  (linear retrial rate). If upon the arrival of  $r$ -customer there is no free and serviceable server, then they either leave the orbit (i.e., are lost) w.p.  $\sigma_r$  or return to the orbit w.p.  $1 - \sigma_r$  provided that there are  $r$  customers in the orbit,  $r = 1, \dots, R$ .

After the completion of the service process, the customer either leaves the system forever w.p.  $\beta_r$  or goes to the orbit for repeated service w.p.  $1 - \beta_r$  provided that there are  $r$  customers in the orbit,  $r = 0, 1, \dots, R - 1$ ;  $\beta_R = 1$ . Customers who require re-service are called feedback customers ( $f$ -customers). For simplicity, in the orbit, both kinds of customers (i.e.,  $r$ -customers and  $f$ -customers) are treated as the same and they are referred to as  $r$ -customers. Here, multiple re-servicing processes (feedbacks) are allowed.

For the model with an infinite orbit capacity for retrial and feedback customers, i.e., when  $R = \infty$ , we assume that  $\alpha_r < 1$ ,  $\beta_r < 1$ ,  $\gamma_r < 1$  for any value of  $r$ .

The service times for all kinds of customers ( $p$ -,  $r$ -, and  $f$ -customers) are assumed to be exponentially distributed with a common mean  $\mu^{-1}$ .

All times involved in the model, as well as the arrival of  $p$ -customers from the outside, the arrival of  $r$ -customers from the orbit, and the service of customers, are mutually independent of each other.

Our goal is to determine the stationary distribution of the given system, as well as to calculate its performance measures: loss probabilities of both  $p$ -customers and  $r$ -customers, the average number of both failed and busy servers, and the average number of  $r$ -customers in the orbit.

#### 4. Balance Equations Method for the Model with a Finite Orbit Size

The state of the system is defined by a three-dimensional vector  $(b, f, r)$ , where  $b$  is the number of busy servers,  $f$  is the number of failure servers, and  $r$  is the number of  $r$ -customers in the orbit. In other words, the mathematical model of the given system is 3D MC with the following state space (SS):

$$S = \bigcup_{r=0}^R S_r, S_r \cap S_{r'} = \emptyset, r \neq r', \tag{1}$$

where  $S_r = \{(b, f, r) : 0 \leq b + f \leq K\}, r = 0, 1, \dots, R$ .

It is clear that in the state  $(b, f, r) \in S$ , the number of idle and serviceable servers is equal to  $K - b - f$ . From (1), we conclude that the SS of the investigated 3D MC geometrically represents a prism with height  $R$ . Consider the problem of constructing the generator of the indicated 3D MC.

Let us fix the value of  $r$  and first consider determining transition intensities between the states within a subset (level),  $S_r$ . In order to be short, a transition from state  $(b, f, r) \in S_r$  to state  $(b', f', r') \in S_{r'}$  with intensity  $a$  is denoted by  $(b, f, r) \xrightarrow{a} (b', f', r')$ . Let the initial state be  $(b, f, r) \in S$ . As a result of analyzing possible transitions, we have the following relations:

- $(b, f, r) \xrightarrow{\lambda} (b + 1, f, r)$  is carried out upon receipt of the  $p$ -customer if  $b + f < K$ ;
- $(b, f, r) \xrightarrow{b\mu_r} (b - 1, f, r)$  is carried out upon departure of the customer from the system or their loss due to a server's active breakdown, where  $\mu_r = \mu\beta_r + \theta_1\gamma_r$ ;
- $(b, f, r) \xrightarrow{(K-b-f)\theta_2} (b, f + 1, r)$  is carried out when the failure of a server occurs in its idle status;
- $(b, f, r) \xrightarrow{f\xi} (b, f - 1, r)$  is carried out when the repair of a failed server is completed;

Now consider determining the transition intensities between the states of the various subsets (levels). Note that only transitions between states of neighboring levels are allowed. For such kinds of transitions, we have the following relations:

- $(b, f, r) \xrightarrow{\lambda(1-\alpha_r)} (b, f, r + 1)$  is carried out upon the joining of the arrived  $p$ -customer to orbit when  $b + f = K$  and  $r < R$ ;
- $(b, f, r) \xrightarrow{b\mu(1-\beta_r)} (b - 1, f, r + 1)$  is carried out when the serviced customer feedbacks to the orbit when  $r < R$ ;
- $(b, f, r) \xrightarrow{f\theta_1(1-\gamma_r)} (b, f + 1, r + 1)$  is carried out when the customer goes to orbit due to a server's active breakdown when  $r < R$ ;
- $(b, f, r) \xrightarrow{r\eta} (b + 1, f, r - 1)$  is carried out upon receipt of the  $r$ -customer if  $b + f < K$  when  $r > 0$ ;
- $(b, f, r) \xrightarrow{r\eta\sigma_r} (b, f, r - 1)$  is carried out when an arrived  $r$ -customer is lost, if  $b + f = K$ , when  $r > 0$ .

The transition intensity from state  $(b, f, r)$  to state  $(b', f', r')$  is denoted by  $q((b, f, r), (b', f', r'))$ . From the above relations, we conclude that the positive elements of the generator of the investigated 3D MCs are determined as follows:

$$q((b, f, r), (b', f', r')) = \begin{cases} \lambda & \text{if } b + f < K, (b', f', r') = (b + 1, f, r) \\ b\mu_r & \text{if } (b', f', r') = (b - 1, f, r) \\ (K - b - f)\theta_2 & \text{if } (b', f', r') = (b, f + 1, r) \\ f\zeta & \text{if } (b', f', r') = (b, f - 1, r) \\ \lambda(1 - \alpha_r) & \text{if } r < R, b + f < K, (b', f', r') = (b, f, r + 1) \\ b\mu(1 - \beta_r) & \text{if } r < R, (b', f', r') = (b - 1, f, r + 1) \\ f\theta_1(1 - \gamma_r) & \text{if } r < R, (b', f', r') = (b, f + 1, r + 1) \\ r\eta & \text{if } b + f < K, r > 0, (b', f', r') = (b + 1, f, r - 1) \\ r\eta\sigma_r & \text{if } b + f = K, r > 0, (b', f', r') = (b, f, r - 1) . \end{cases} \tag{2}$$

Hereinafter, the equality of vectors means that their corresponding components are equal to one another.

From the relations (2), we conclude that states of constructed finite 3D MCs are communicated with each other, i.e., a stationary mode exists in this system. Let  $p(b, f, r)$  denote the stationary probability of the state  $(b, f, r) \in S$ . These probabilities can be found using the following balance equations:

$$\sum_{(b', f', r') \in S^+_{(b, f, r)}} q((b, f, r), (b', f', r')) p(b, f, r) = \sum_{(b', f', r') \in S^-_{(b, f, r)}} q((b', f', r'), (b, f, r)) p(b', f', r') ; \tag{3}$$

$$\sum_{(b, f, r) \in S} p(b, f, r) = 1 , \tag{4}$$

where  $S^+_{(b, f, r)}$  is the set of those states of  $S$ , which can be reached from the state  $(b, f, r)$  in one step, and  $S^-_{(b, f, r)}$  is the set of those states of  $S$ , from which one can get to the state  $(b, f, r)$  in one step. Note that the sets  $S^+_{(b, f, r)}$  and  $S^-_{(b, f, r)}$  are determined from the relations (2).

The dimension of the balance Equations (3) and (4) is equal to  $(K + 1)(K + 2)(R + 1)/2$ . Modern software packages allow the linear equations of arbitrary (finite) dimensions to be solved.

The desired performance measures are calculated via steady-state probabilities. The average number of busy ( $NBS_{av}$ ) and failed servers ( $NFS_{av}$ ), as well as the average number of  $r$ -customers in the orbit ( $NCO_{av}$ ), are calculated as the mathematical expectations of the corresponding random variables, i.e.,

$$NBS_{av} = \sum_{b=1}^K b \sum_{(b, f, r) \in S} p(b, f, r) ; \tag{5}$$

$$NFS_{av} = \sum_{f=1}^K f \sum_{(b, f, r) \in S} p(b, f, r) ; \tag{6}$$

$$NCO_{av} = \sum_{r=1}^R r \sum_{(b, f, r) \in S} p(b, f, r) . \tag{7}$$

The loss probabilities of  $p$ -customers ( $P_p$ ) and  $r$ -customers ( $P_r$ ) are calculated as

$$P_p = \sum_{r=0}^R \alpha_r \sum_{(b, f, r) \in S} p(b, f, r) \delta(b + f, K) ; \tag{8}$$

$$P_r = \sum_{r=1}^R \sigma_r \sum_{(b, f, r) \in S} p(b, f, r) \delta(b + f, K) . \tag{9}$$

Hereinafter  $\delta(i, j)$  are Kronecker's delta.

### 5. The Space Merging Method for the Model with an Infinite Orbit Size

Methodological difficulties occur in solving the investigated problem for the model with an infinite orbit size, i.e., when  $R = \infty$ . Now consider this problem for the case where the indicated above probabilities,  $\alpha_r, \beta_r, \gamma_r, \sigma_r$ , are state-independent, i.e., assume that  $\alpha_r = \alpha, \beta_r = \beta, \gamma_r = \gamma, \sigma_r = \sigma$ .

The mathematical model of the given system is 3D MC with the following state space (SS):

$$E = \bigcup_{r=0}^{\infty} E_r, E_r \cap E_{r'} = \emptyset, r \neq r', \tag{10}$$

where  $E_r = \{(b, f, r) : 0 \leq b + f \leq K\}, r = 0, 1, \dots$

From (10), we conclude that the SS of the investigated 3D MC geometrically represents an infinite prism. The generator of the presently examined infinite 3D MC is determined similarly to (2), wherein the right side condition,  $r < R$ , should be omitted.

For simplicity, let  $p(b, f, r)$  again denote the stationary probability of the state  $(b, f, r) \in E$ . Below, we show that for the case of the linear retrial rate, the stationary mode exists for any positive values of input parameters; however, in the case of the constant retrial rate, the fulfillment of some stability conditions is required.

From the form of the SS (8), we conclude that the constructed 3D MC does not belong to the class of quasi-birth and death (QBD) processes (see [26]). Therefore, unfortunately, to calculate the steady-state probabilities of the investigated 3D MC, we cannot apply well-worked matrix-geometric methods. For this reason, we propose an alternate approach below to solve the indicated problem. The proposed approach is based on the space merging principles of a multi-dimensional MC and it allows for the calculation of steady-state probabilities and performance metrics via explicit formulas. Moreover, the obtained formulas contain quantities that are tabulated, i.e., Erlang’s B-formula as well as the formula for calculating the average number of busy servers in Erlang’s classical model.

It is known that satisfying the following condition is required for the correct application of the SMM: the state space of the system should be decomposed into classes in such a way that the rates of transitions between states within classes are much larger than rates of transitions between states from different classes. The fulfillment of this condition can be ensured at certain ratios between the initial parameters of the system under study. So, suppose that the arrival rate of  $p$ -customers is much larger than the arrival rate of  $r$ -customers, i.e.,  $\eta \ll \lambda$ .

When this assumption is fulfilled, the transitions between the states within classes  $E_r$  (see (10)) are much larger than transition classes between the states of different classes. By using this fact, the following merge function in the SS (10) is defined:

$$U_1((b, f, r)) = \langle r \rangle, (b, f, r) \in E_r, \tag{11}$$

where  $\langle r \rangle$  is the merged state, which includes all micro-states from the class  $E_r, r = 0, 1, \dots$ . The set of merged states  $\langle r \rangle$  is denoted by  $\Omega_1 = \{\langle r \rangle : r = 0, 1, \dots\}$ .

According to hierarchical SMM (see [7,27]), the approximate values of the steady-state probabilities of the initial 3D MC, denoted by  $\tilde{p}(b, f, r)$ , are calculated as follows:

$$\tilde{p}(b, f, r) = \rho_r(b, f) \pi_1(\langle r \rangle), \tag{12}$$

where  $\rho_r(b, f)$  is the probability of the state,  $(b, f, r)$ , within a splitting model with SS  $E_r$ , and  $\pi_1(\langle r \rangle)$  is the probability of the merged state  $\langle r \rangle \in \Omega_1$ .

From (12), conclude that to calculate the steady-state probabilities of the initial 3D MC, it is necessary to find the steady-state probabilities,  $\rho_r(b, f)$ , of the infinite number of 2D MCs with a finite SS as well as the steady-state probabilities,  $\pi_1(\langle r \rangle)$ , of one 1D MC with the SS  $\Omega_1$ .

First, consider the calculation of the steady-state probabilities,  $\rho_r(b, f)$ , of the 2D MCs with the SS  $E_r$ . The transitions diagram for the splitting model with the SS  $E_r$  is shown in Figure 2.

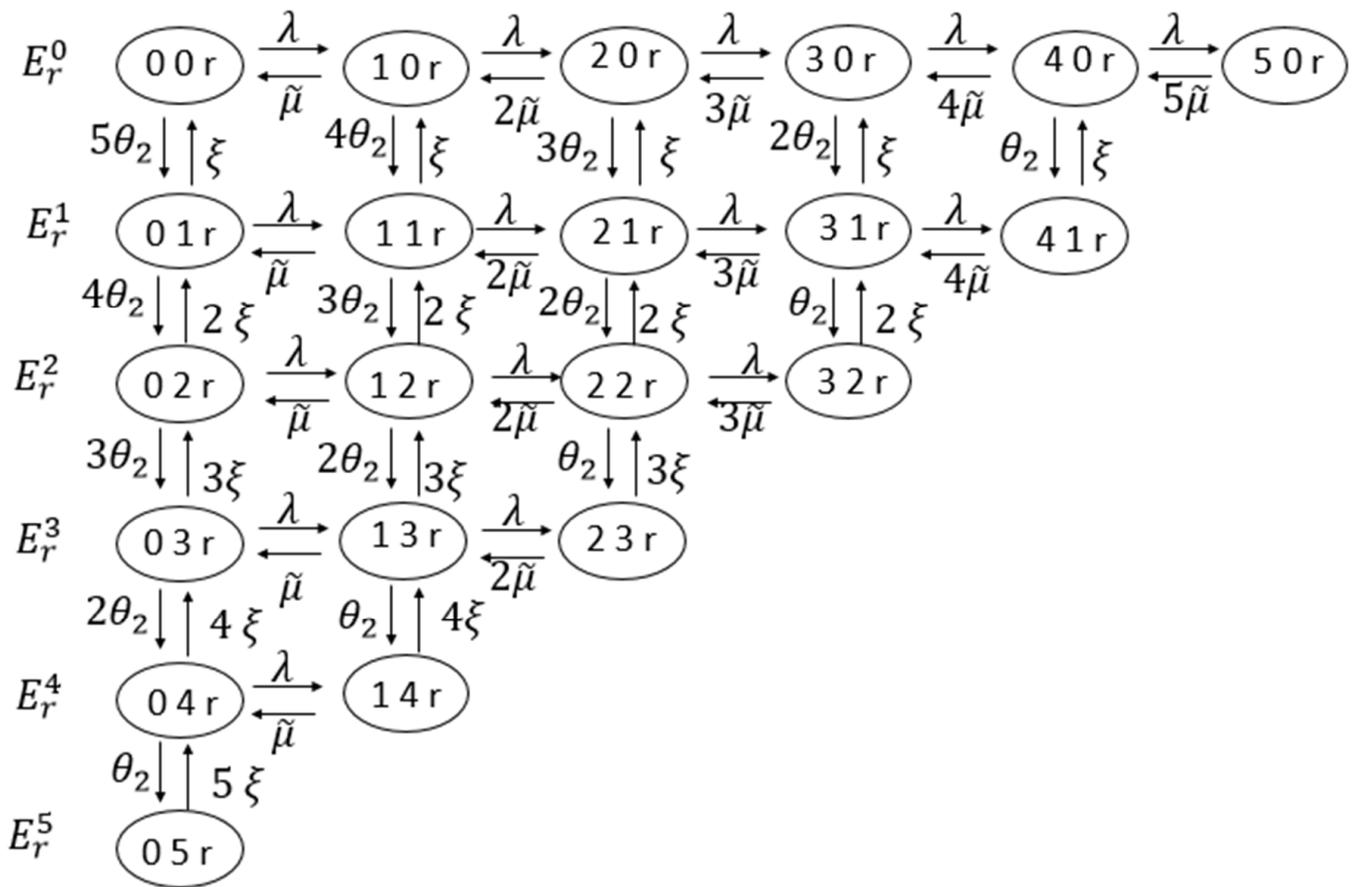


Figure 2. State diagram for the split model with SS  $E_r$ ,  $K = 5$ .

The transition intensity from state  $(b, f, r) \in E_r$  to state  $(b', f', r) \in E_r$  is denoted by  $q_r((b, f, r), (b', f', r))$ . From (2), we conclude that these quantities are determined as follows:

$$q_r((b, f, r), (b', f', r)) = \begin{cases} \lambda & \text{if } b + f < K, (b', f', r) = (b + 1, f, r) \\ b\tilde{\mu} & \text{if } (b', f', r) = (b - 1, f, r) \\ (K - b - f)\theta_2 & \text{if } (b', f', r) = (b, f + 1, r) \\ f\xi & \text{if } (b', f', r) = (b, f - 1, r), \end{cases} \quad (13)$$

where  $\tilde{\mu} = \mu\beta + \theta_1\gamma$ .

From (13), we conclude that  $q_r((b, f, r), (b', f', r))$  are identical for any value of  $r$ , i.e., the steady-state probabilities,  $\rho_r(b, f)$ , do not depend on  $r$  because, below, we fix the value of  $r$  and omit the subscript  $r$  in the notation of these quantities.

The calculation of the steady-state probabilities,  $\rho(b, f)$ , for a moderately sized SS,  $S_r$ , might be performed via solving balance equations. However, it is possible to develop explicit formulas for their approximate calculations. For this purpose, the merging procedure (the second level of the hierarchy) is applied to these 2D MCs.

For the correct application of the SMM in the second level, it is assumed that  $\theta_1 \gg \theta_2$ . Note that this assumption is a realistic one since, in real situations, the failure intensity of servers in a busy mode is much higher than for those in an idle mode. Under this assumption, consider the following splitting of  $S_r$ :

$$E_r = \bigcup_{f=0}^K E_r^f, E_r^f \cap E_r^{f'} = \emptyset, f \neq f', \quad (14)$$

where  $E_r^f = \{(b, f) \in E_r : b = 0, 1, \dots, K; b + f \leq K\}$ ,  $f = 0, 1, \dots, K$ .

Similar to (11), based on splitting (14) in the SS,  $E_r$ , the following merge function is defined:

$$U_2((b, f)) = \langle\langle f \rangle\rangle, (b, f) \in E_r^f,$$

where  $\langle\langle f \rangle\rangle$  is a merge state which includes all micro-states from the subclass  $E_r^f$ . The set of merged states  $\langle\langle f \rangle\rangle$  is denoted by  $\Omega_2 = \{\langle\langle f \rangle\rangle : f = 0, 1, \dots, K\}$ .

In accordance with [7,27], we have:

$$\rho(b, f) \approx \rho_f(b)\pi_2(\langle\langle f \rangle\rangle), \tag{15}$$

where  $\rho_f(b)$  is the probability of the state  $(b, f)$  within the splitting model with space  $E_r^f$ , and  $\pi_2(\langle\langle f \rangle\rangle)$  is the probability of the merged state  $\langle\langle f \rangle\rangle \in \Omega_2$ .

In a subclass,  $E_r^f, f = 0, 1, \dots, K$ , the second component of the state vector is a constant and equals  $f$ . Therefore, in a splitting model with space  $E_r^f$ , each state,  $(b, f)$ , can be represented simply by the first component, i.e., by  $b, b = 0, 1, \dots, K - f$ .

The transition intensity from state  $(b, f) \in E_r^f$  to state  $(b', f) \in E_r^f$  is denoted by  $q(b, b')$ . So, from relations (2) we conclude that these quantities are independent of the parameter  $r$ , and are determined as follows:

$$q(b, b') = \begin{cases} \lambda, & \text{if } b' = b + 1, \\ b(\mu\beta + \theta_1\gamma), & \text{if } b' = b - 1. \end{cases} \tag{16}$$

From (16), we conclude that the state probabilities,  $\rho_f(b)$ , of the splitting models with the SS  $E_r^f, f = 0, 1, \dots, K - 1$ , coincides with the state probabilities of Erlang's model,  $M/M/K-f/K-f$ , with load  $\nu = \lambda/(\mu\beta + \theta_1\gamma)$ , i.e., the desired probabilities are calculated by well-known Erlang's formulas:

$$\rho_f(b) = \frac{\nu^b}{b!} / \sum_{i=0}^{K-f} \frac{\nu^i}{i!}, b = 0, 1, \dots, K - f. \tag{17}$$

The splitting model with the SS  $E_r^K$  contains only one state  $(0, K, r)$ , i.e., for this model, we set  $\rho_K(0) = 1$ .

The transition intensity from merged state  $\langle\langle f \rangle\rangle \in \Omega_2$  to merged state  $\langle\langle f' \rangle\rangle \in \Omega_2$  is denoted by  $q(\langle\langle f \rangle\rangle, \langle\langle f' \rangle\rangle)$ . By using (2) and (17), after certain mathematical transformations, we conclude that the indicated quantities are determined as follows:

$$q(\langle\langle f \rangle\rangle, \langle\langle f' \rangle\rangle) = \begin{cases} \Lambda_f, & \text{if } f' = f + 1, \\ f\zeta, & \text{if } f' = f - 1, \end{cases} \tag{18}$$

where  $\Lambda_f = \theta_2 \sum_{i=0}^{K-f-1} \rho_f(i)(K - f - i)$ .

Therefore, from relations (18), we conclude that the probabilities of merged states are calculated as follows:

$$\pi_2(\langle\langle f \rangle\rangle) = \frac{1}{f! \zeta^f} \prod_{i=0}^{f-1} \Lambda_i \pi_2(\langle\langle 0 \rangle\rangle), f = 1, \dots, K, \tag{19}$$

where  $\pi_2(\langle\langle 0 \rangle\rangle)$  is found from the normalizing condition, i.e.,  $\sum_{f=0}^K \pi_2(\langle\langle f \rangle\rangle) = 1$ .

Now consider the problem of finding the probabilities of merged states  $\pi_1(\langle r \rangle), \langle r \rangle \in \Omega_1$ . The transition intensity from state  $\langle r \rangle \in \Omega_1$  to state  $\langle r' \rangle \in \Omega_1$  is denoted

by  $q(\langle r \rangle, \langle r' \rangle)$ . By using (2) (for this model), (17), and (19), we conclude that these quantities are calculated as follows:

$$q(\langle r \rangle, \langle r' \rangle) = \begin{cases} \omega, & \text{if } r' = r + 1, \\ r\phi, & \text{if } r' = r - 1, \end{cases} \tag{20}$$

where

$$\begin{aligned} \omega &= (\mu(1 - \beta) + \theta_1(1 - \gamma)) \underbrace{\sum_{j=0}^{K-1} \sum_{i=1}^{K-j} i \rho_j(i) \pi_2(\langle\langle j \rangle\rangle)}_{B_{av}(K-j, \nu)} + \lambda(1 - \alpha) \underbrace{\sum_{i=0}^K \rho_i(K-i) \pi_2(\langle\langle i \rangle\rangle)}_{E_B(K-i, \nu)} = \\ &= (\mu(1 - \beta) + \theta_1(1 - \gamma)) \sum_{j=0}^{K-1} B_{av}(K-j, \nu) \pi_2(\langle\langle j \rangle\rangle) + \lambda(1 - \alpha) \sum_{i=0}^K E_B(K-i, \nu) \pi_2(\langle\langle i \rangle\rangle); \\ \phi &= \eta \left( \underbrace{\sum_{i=0}^{K-1} \rho_0(i) \pi_2(\langle\langle 0 \rangle\rangle)}_{1-E_B(K, \nu)} + \underbrace{\sum_{i=0}^{K-2} \rho_1(i) \pi_2(\langle\langle 1 \rangle\rangle)}_{1-E_B(K-1, \nu)} + \dots + \underbrace{\rho_{K-1}(0) \pi_2(\langle\langle K-1 \rangle\rangle)}_{1-E_B(1, \nu)} \right) + \\ &+ \eta \sigma \left( \underbrace{\rho_0(K) \pi_2(\langle\langle 0 \rangle\rangle)}_{E_B(K, \nu)} + \underbrace{\rho_1(K-1) \pi_2(\langle\langle 1 \rangle\rangle)}_{E_B(K-1, \nu)} + \dots + \underbrace{\rho_K(0) \pi_2(\langle\langle K \rangle\rangle)}_{E_B(0, \nu)} \right) = \\ &= \eta \left( \sum_{i=1}^K (1 - E_B(i, \nu)) \pi_2(\langle\langle K-i \rangle\rangle) + \sigma \sum_{i=0}^K E_B(i, \nu) \pi_2(\langle\langle K-i \rangle\rangle) \right). \end{aligned}$$

Hereinafter,  $E_B(m, \nu)$  and  $B_{av}(m, \nu)$  denote Erlang’s B-formula and the average number of busy servers for the  $M/M/m/m$  system with a load,  $\nu$ , respectively, i.e.,

$$E_B(m, \nu) = \frac{\nu^m}{m!} \bigg/ \sum_{i=0}^m \frac{\nu^i}{i!}; E_B(0, \nu) = 1; B_{av}(m, \nu) = \nu(1 - E_B(m, \nu)).$$

From relations (20). we conclude that the probabilities of merged states coincide with the probabilities of the states of the model  $M/M/\infty$  with the load  $\psi = \omega/\phi$ , i.e.,

$$\pi_1(\langle r \rangle) = \frac{\psi^r}{r!} \exp(-\psi), \quad r = 0, 1, 2, \dots \tag{21}$$

By using (12), (15), (17), and (21), the approximate values of the steady-state of the initial 3D MC are found. So, for any positive values of the loading parameters in this system, there exists stationary mode, and after some standard mathematical transformations, we obtain the following approximate formulas for calculating the performance measures:

$$P_p \approx \alpha \sum_{i=0}^K E_B(i, \nu) \pi_2(\langle\langle K-i \rangle\rangle); \tag{22}$$

$$P_r \approx \sigma(1 - e^{-\psi}) \sum_{i=0}^K E_B(i, \nu) \pi_2(\langle\langle K-i \rangle\rangle); \tag{23}$$

$$NBS_{av} \approx \sum_{i=1}^K i \sum_{j=0}^{K-i} \rho_j(i) \pi_2(\langle\langle j \rangle\rangle); \tag{24}$$

$$NFS_{av} \approx \sum_{i=1}^K i \pi_2(\langle\langle i \rangle\rangle); \tag{25}$$

$$NCO_{av} \approx \psi \tag{26}$$

Note. The average number of busy ( $NBS_{av}$ ) and failed servers ( $NFS_{av}$ ) (see Formulas (24) and (25)) does not depend on the parameters  $\eta$  and  $\sigma$ . These facts are explained by the accepted assumption that  $\eta \ll \lambda$ , i.e., the impact of the indicated parameters on two performance measures is not essential. Other performance measures depend on all system parameters.

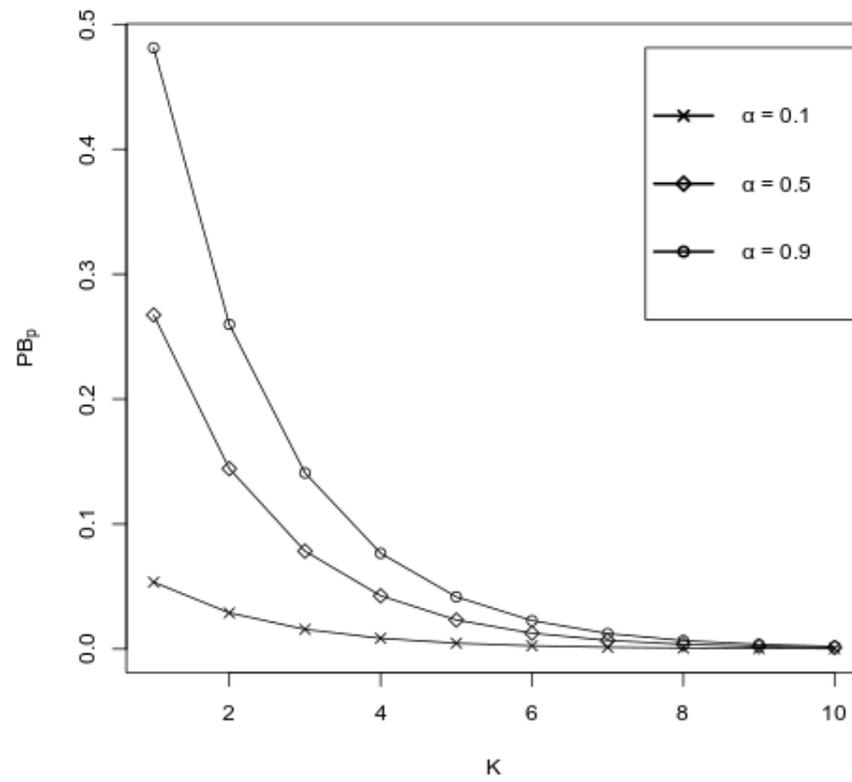
## 6. Numerical Results

In this section, the qualitative behavior of the performance measures is explored through a few numerical experiments. Due to the size limitations for the present paper, only the results for the model with an infinite orbit size have been considered here.

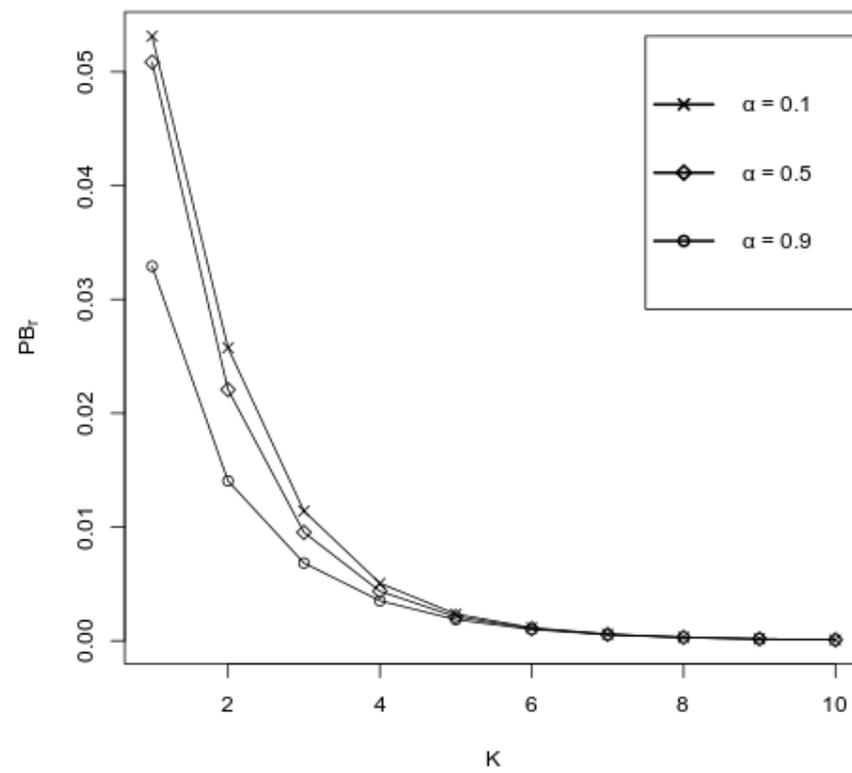
It can be noticed that the proposed algorithms allowed for a simple investigation of the behavior of the performance measures over any parameter of the model without any computational difficulties. For the sake of brevity, we have omitted the results of numerical experiments that demonstrate the high accuracy of the algorithm developed for the approximate calculation of steady-state probabilities of the system under study (appropriate results might be found in [9]). Moreover, for brevity, we only demonstrate experiments that present the behavior of performance measures with respect to the number of servers over various values of the probabilities  $\alpha, \beta, \gamma$  and  $\sigma$ .

For the numerical calculations, the values of the model parameters are selected as follows:  $\lambda = 5, \mu = 3, \theta_1 = 5, \theta_2 = 1, \xi = 1, \eta = 1$ . Note that since the performance measures depend on many parameters, the analysis of the numerical experiments indicated below concerns only the selected values of the initial data. However, in some cases, the conclusions are general.

Figure 3a–e represent the behavior of performance measures as functions of  $K$  over various values of  $\alpha$ , where values of other probabilities are fixed as follows:  $\beta = 0.4, \gamma = 0.7, \sigma = 0.1$ . It can be observed from Figure 3a,b that both  $PB_p$  and  $PB_r$  decrease with respect to  $K$ , where an increase of  $\alpha$  is favorable for  $PB_r$ , whereas  $PB_p$  increases with respect to  $\alpha$ ; for large values of  $K$ , i.e., when  $K \geq 7$ , the impact of the values of  $\alpha$  is negligible. It is seen from Figure 3c,d that both the  $NBS_{av}$  and the  $NFS_{av}$  increase, and their values are almost independent of  $\alpha$ . Figure 3e exhibits the impact of  $K$  and  $\alpha$  over the  $NCO_{av}$ . It is seen from the graph that the values of the decrease with respect to both  $K$  and  $\alpha$ . As in Figure 3a,b, here, when  $K \geq 7$ , the impact of the values of  $\alpha$  on the values of the  $NCO_{av}$  is negligible.

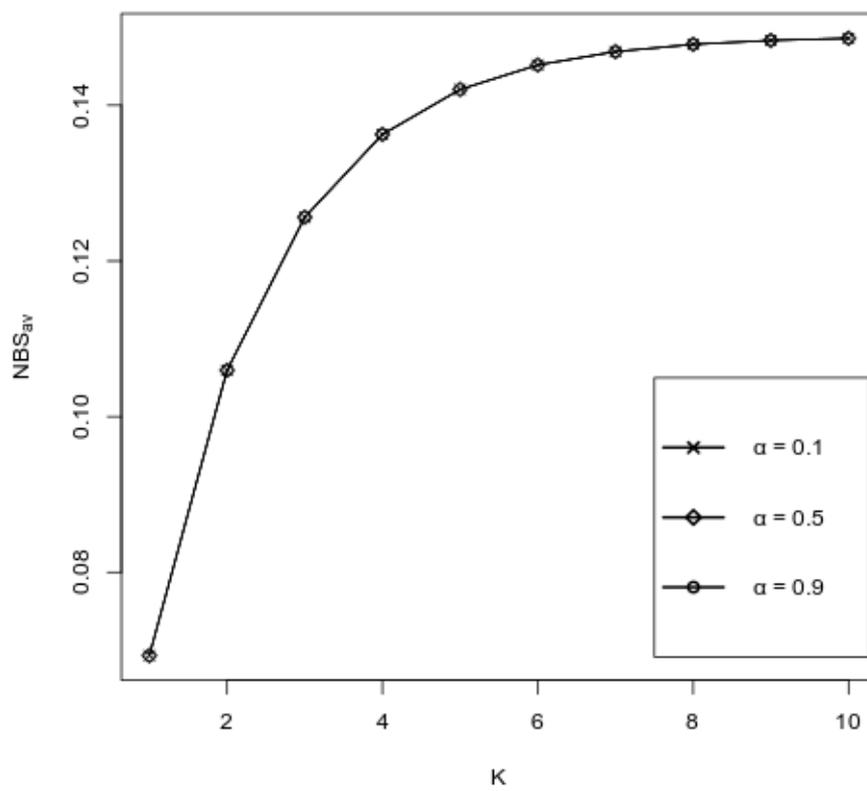


(a)

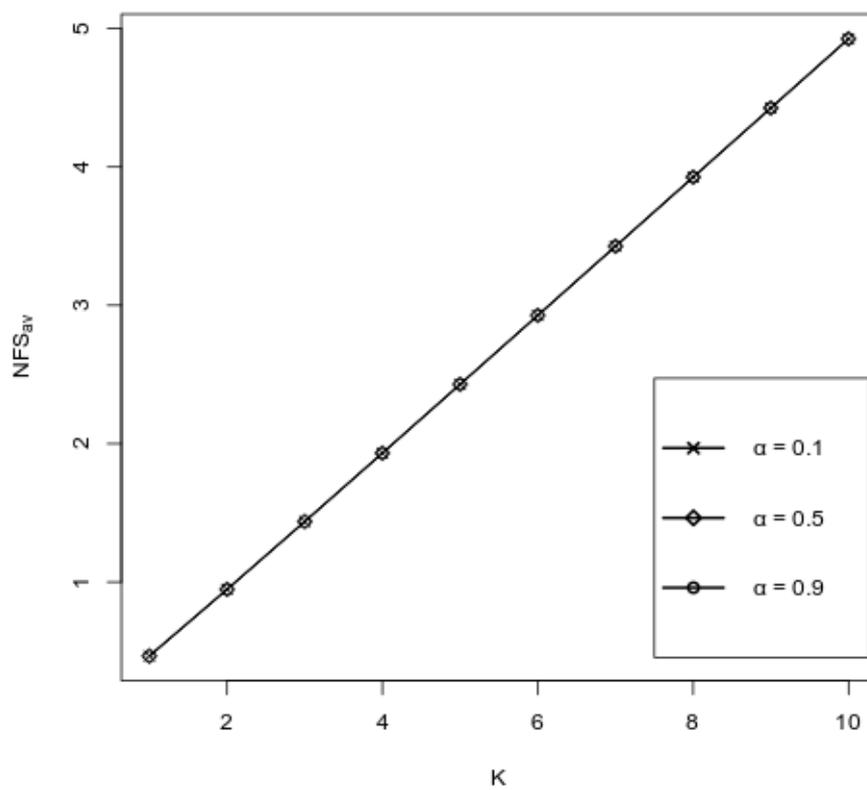


(b)

Figure 3. Cont.

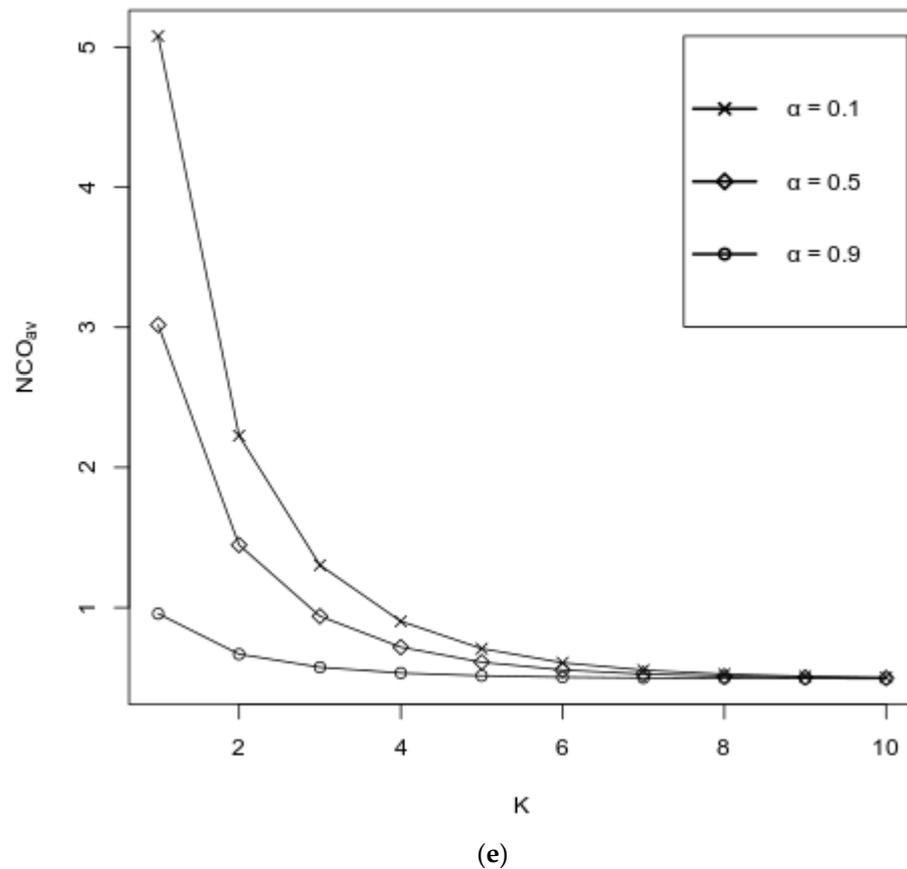


(c)



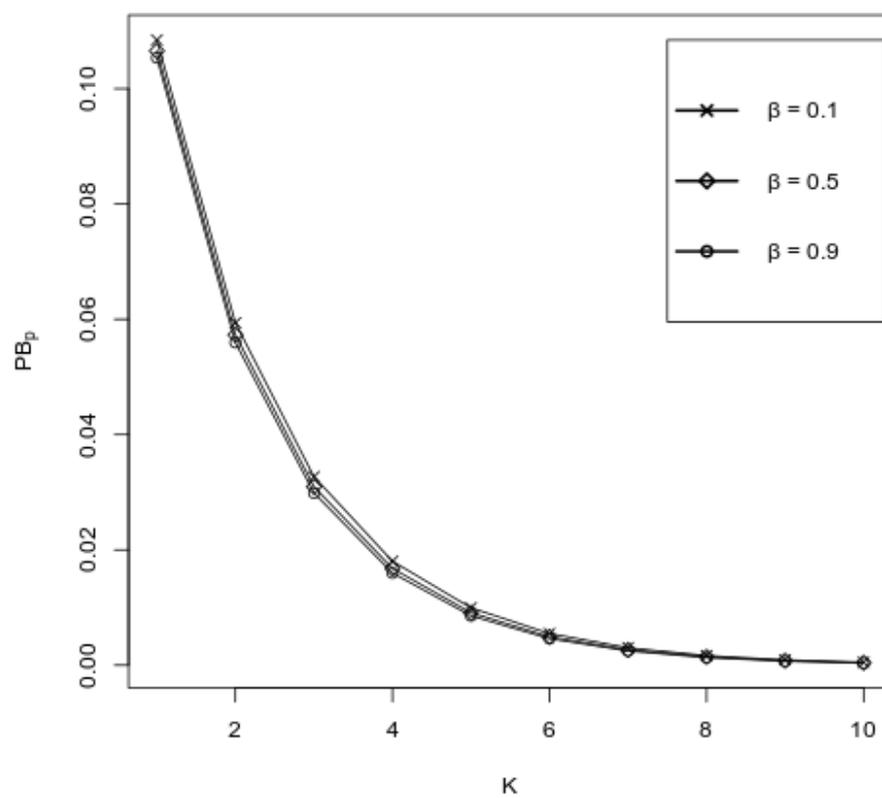
(d)

Figure 3. Cont.

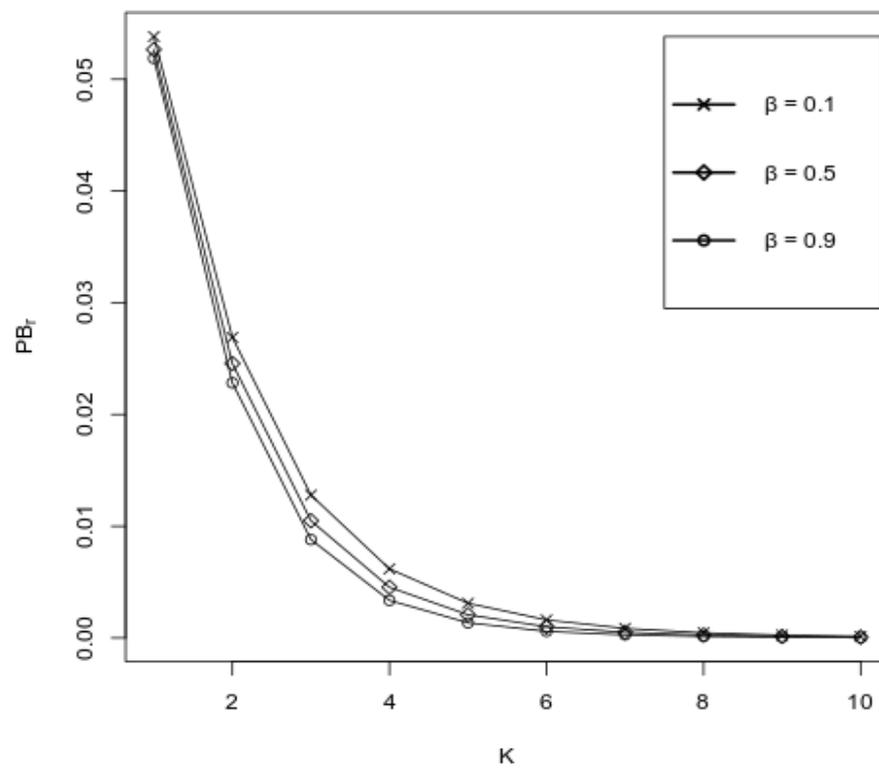


**Figure 3.** Loss probabilities of  $p$ -customers (a) and  $r$ -customers (b), the average number of busy servers (c) and failed servers (d) and the average number of  $r$ -customers in the orbit (e) versus the number of servers under various values of  $\alpha$ .

Figure 4a–e describe the behavior of performance measures as functions of  $K$  over various values of  $\beta$ , where values of other probabilities are fixed as follows:  $\alpha = 0.2$ ,  $\gamma = 0.7$ ,  $\sigma = 0.1$ . It is seen from Figure 4a,b that both  $PB_p$  and  $PB_r$  decrease, and their values are almost independent on  $\beta$ ; some dependence on  $\beta$  is observed for  $PB_r$  under small values of  $K$ , i.e., under small values of  $K$ , the performance measure  $PB_r$  is decreasing by one with respect to  $\beta$ . As was expected, the  $NBS_{av}$  increases with respect to  $K$  and decreases when  $\beta$  increases (see Figure 4c). Figure 4d shows that the  $NFS_{av}$  increases with respect to  $K$ , and its values are almost independent of  $\beta$ . However, it is seen from Figure 4e that the values of the  $NCO_{av}$  increase with respect to  $K$  and decrease when  $\beta$  increases, and that the impact of the values of  $\beta$  on the values of the  $NCO_{av}$  is essential for a large value of  $K$ .

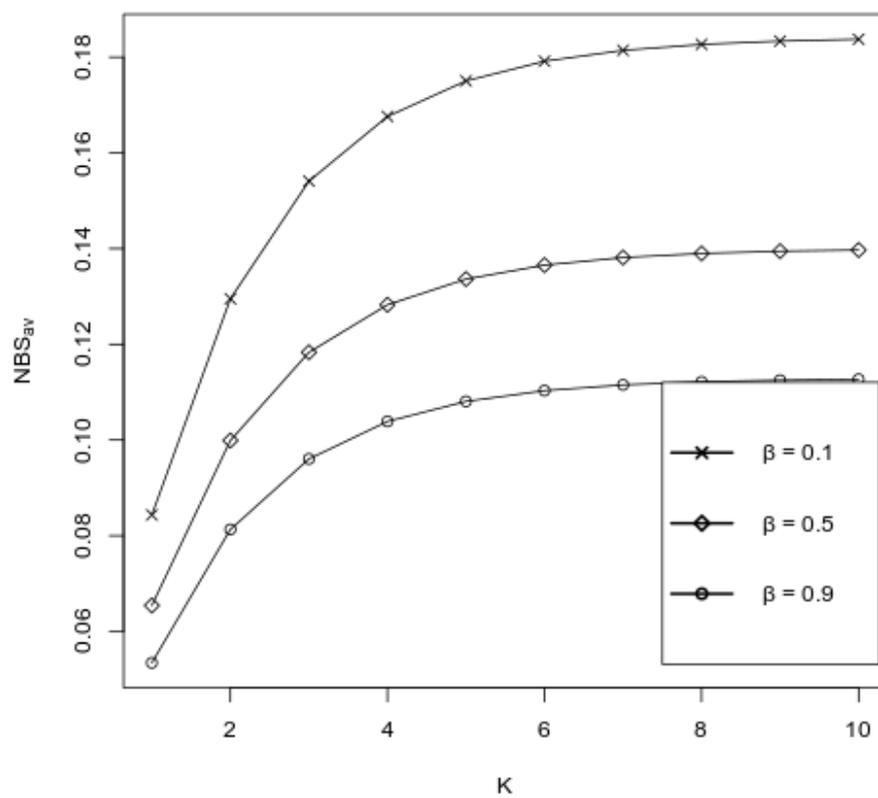


(a)

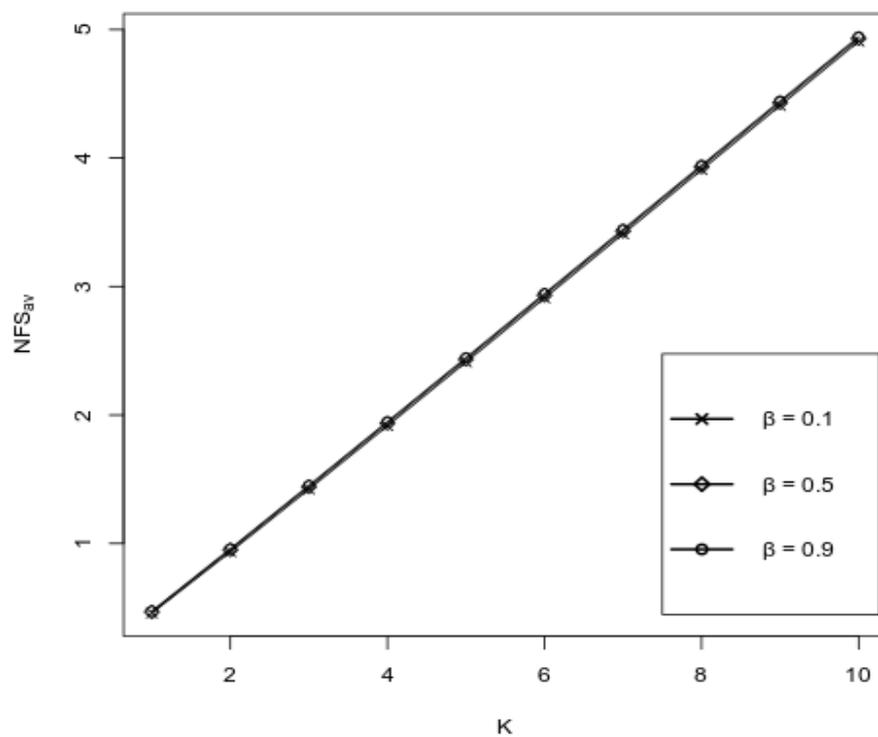


(b)

Figure 4. Cont.

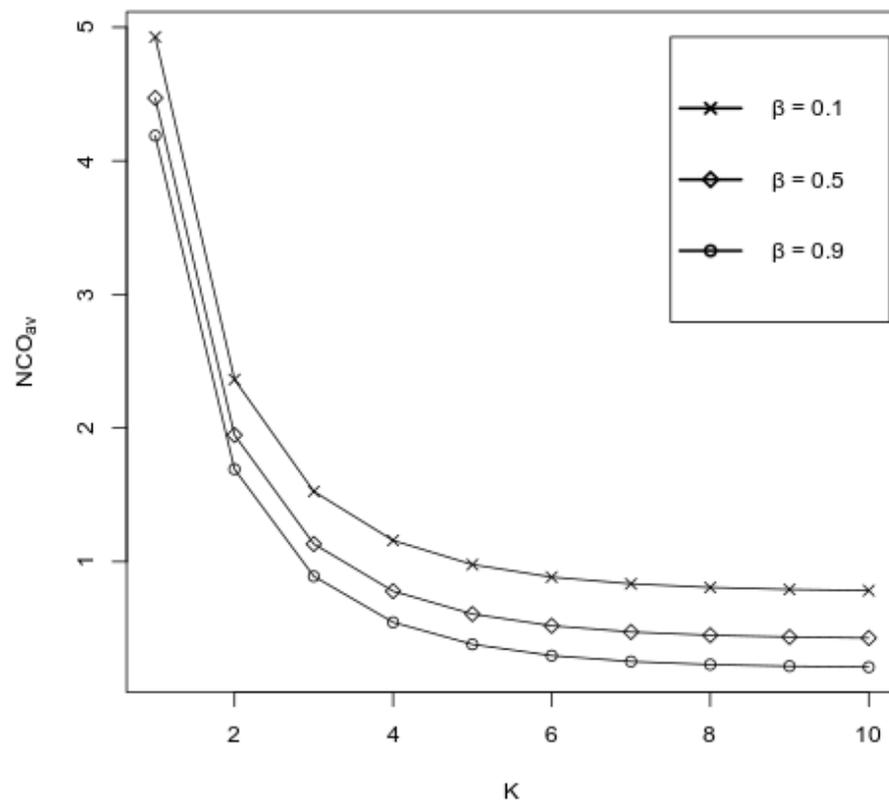


(c)



(d)

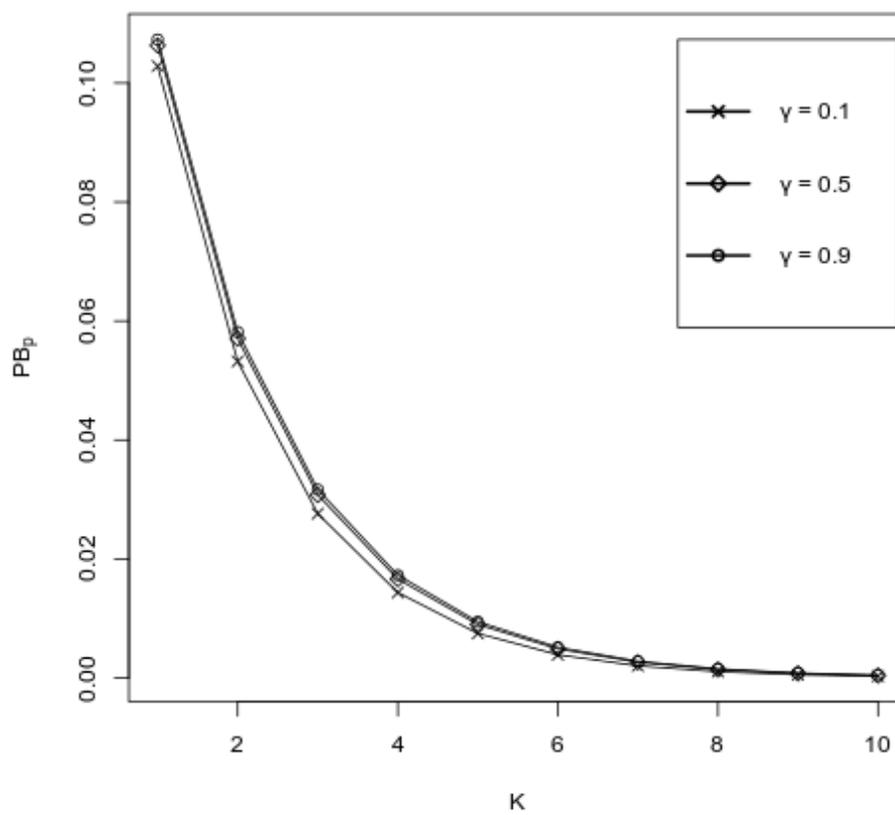
Figure 4. Cont.



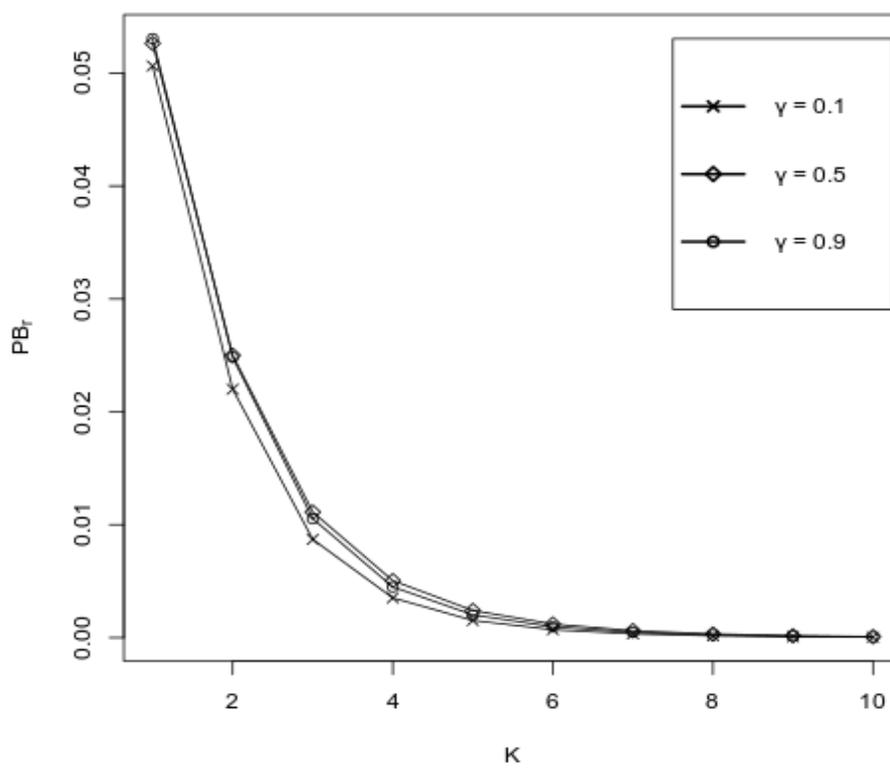
(e)

**Figure 4.** Loss probabilities of  $p$ -customers (a) and  $r$ -customers (b), the average number of busy servers (c) and failed servers (d) and the average number of  $r$ -customers in the orbit (e) versus the number of servers under various values of  $\beta$ .

Figure 5a–e show the behavior of performance measures as functions of  $K$  over various values of  $\gamma$ , where values of other probabilities are fixed as follows:  $\alpha = 0.2, \beta = 0.4, \sigma = 0.1$ . From Figure 5a,b, we see that both  $PB_p$  and  $PB_r$  decrease with respect to  $K$ , and their values are almost independent on  $\gamma$ . Surprisingly, the  $NBS_{av}$  increases when  $\gamma$  increases (see Figure 5c). Such behavior of the  $NBS_{av}$  can be explained as follows: for selected values of the load parameters of the system, an increase in the probability of a customer leaving the system from a broken server leads to an increase in the chance of  $p$ -customers being accepted, i.e., the average number of busy servers increases. Though, for other values of the load parameters, this performance measure might decrease when  $\gamma$  increases. Figure 5d demonstrates that the  $NFS_{av}$  increases with respect to  $K$  and its values are almost independent of  $\gamma$ . The impact of  $\gamma$  on the values of the  $NCO_{av}$  is negligible, and it decreases when  $K$  increases (see Figure 5e).

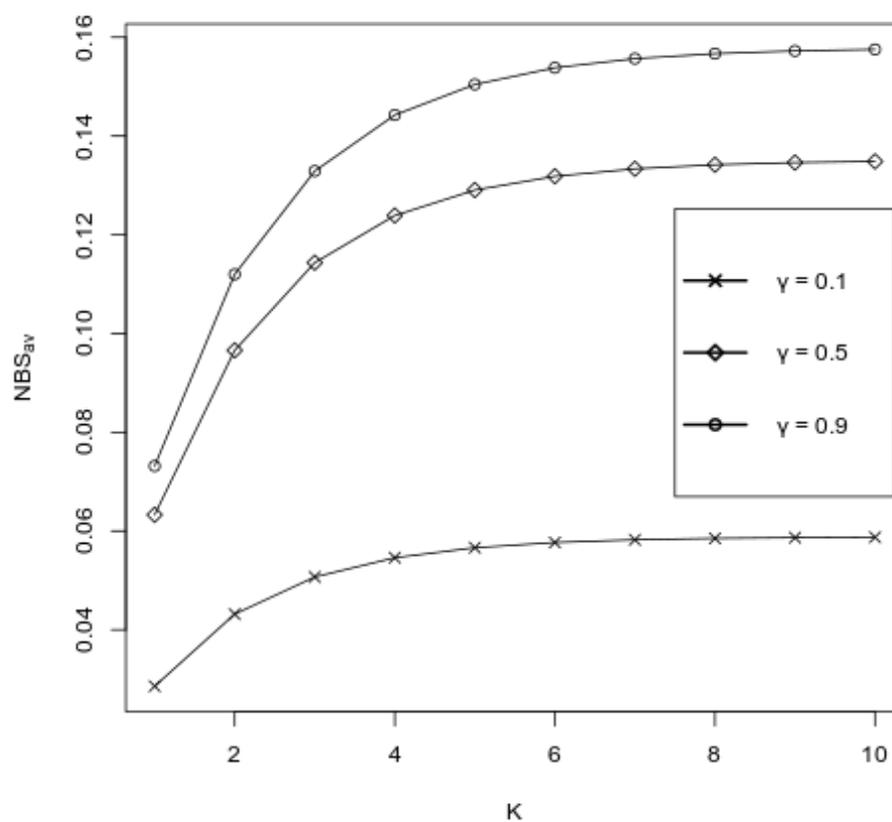


(a)

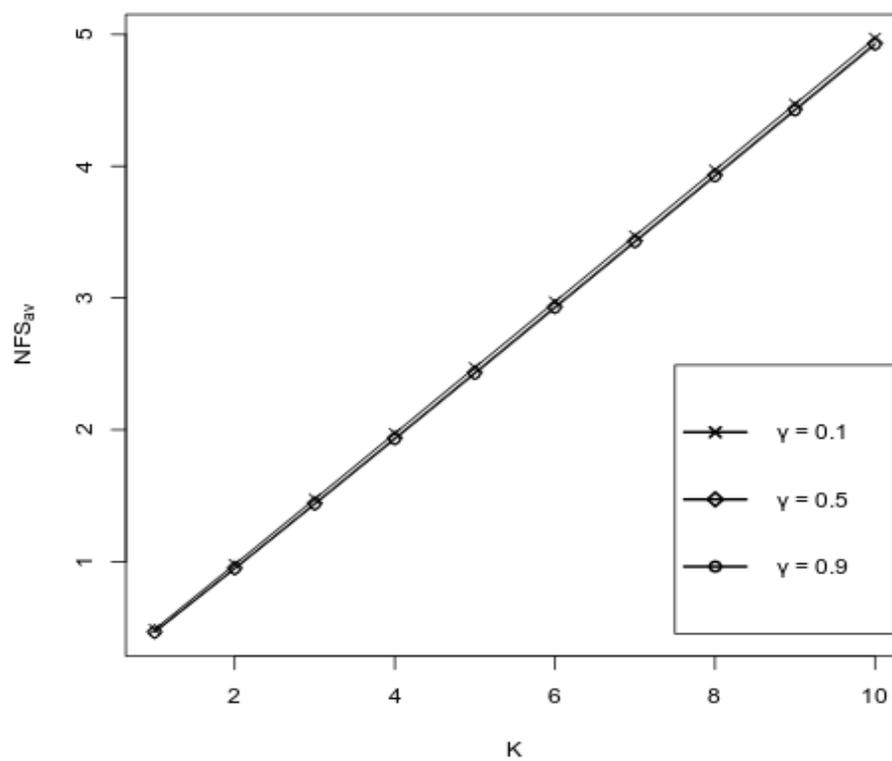


(b)

Figure 5. Cont.

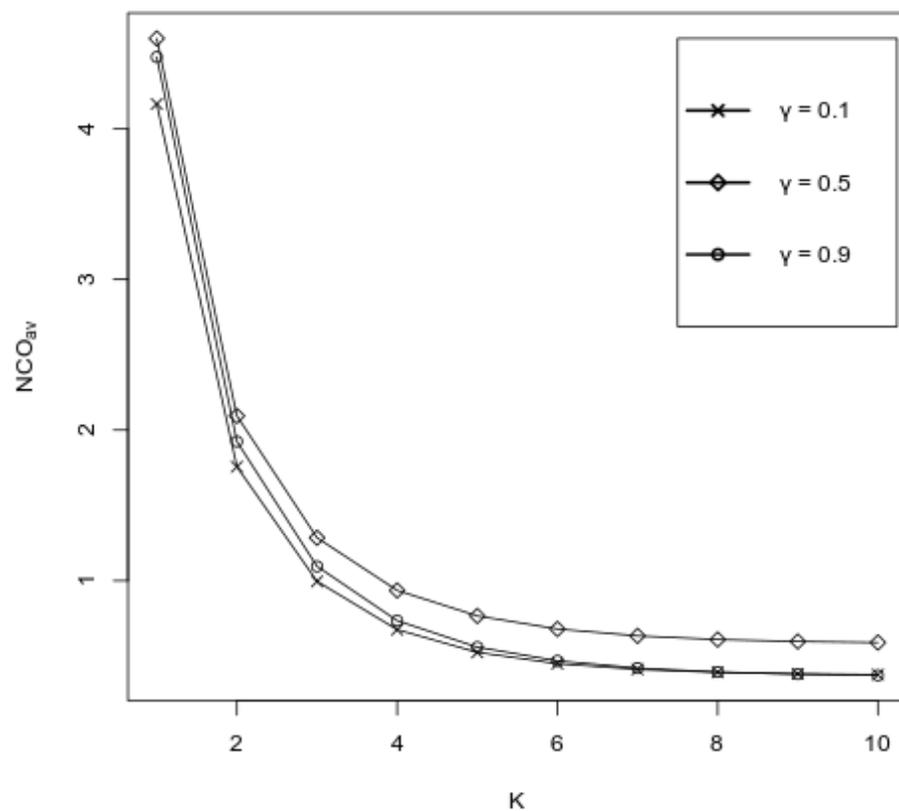


(c)



(d)

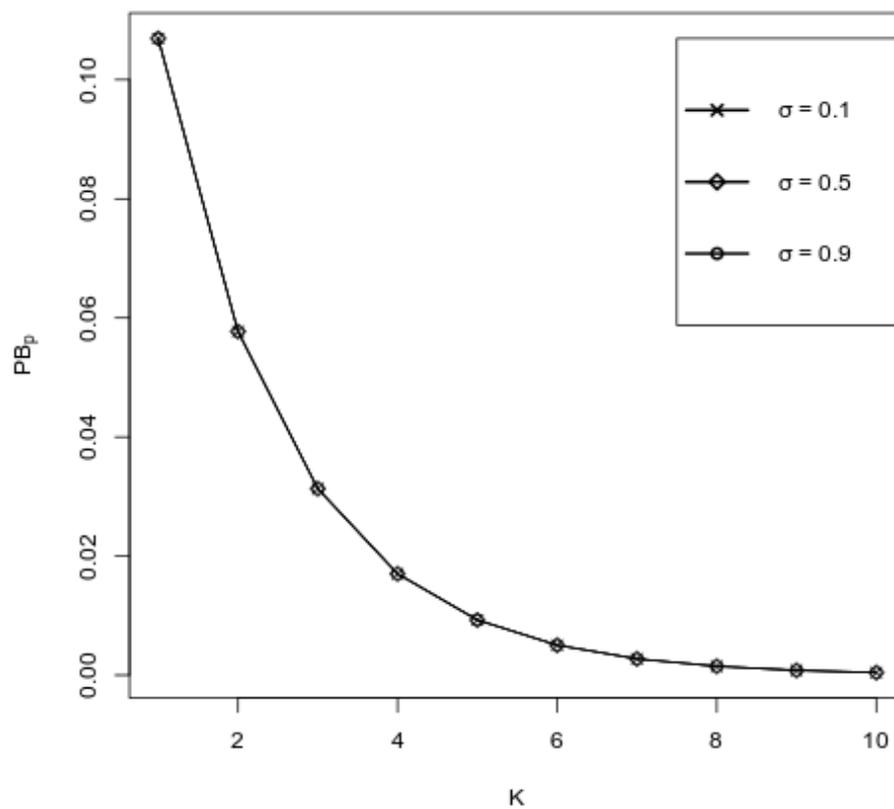
Figure 5. Cont.



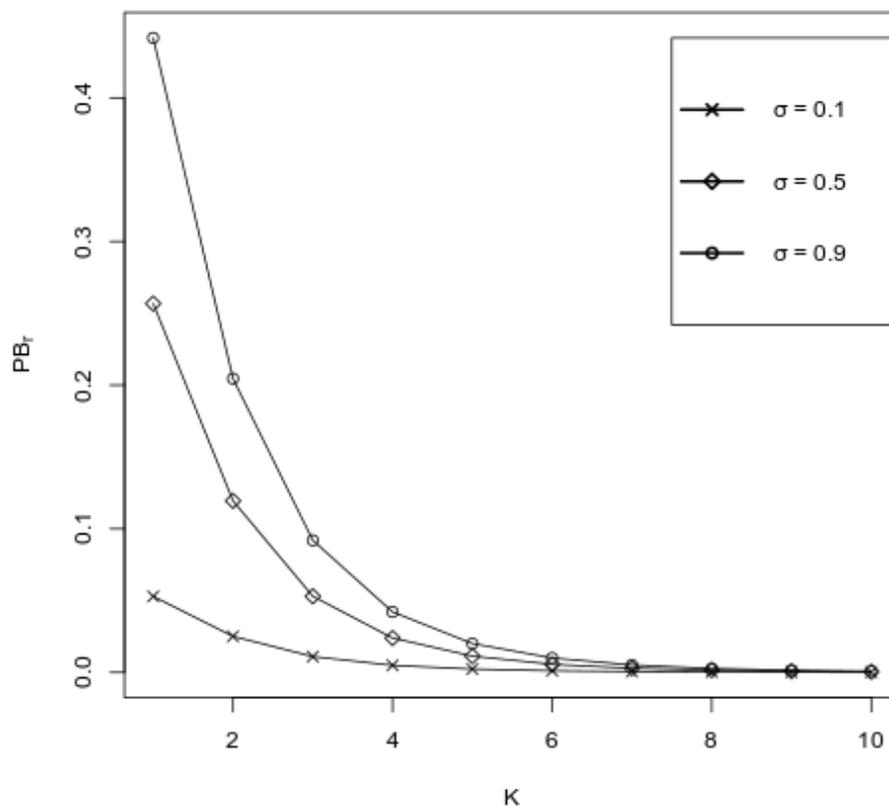
(e)

**Figure 5.** Loss probabilities of  $p$ -customers (a) and  $r$ -customers (b), the average number of busy (c) and failed servers (d) and the average number of  $r$ -customers in the orbit (e) versus the number of servers under various values of  $\gamma$ .

Figure 6a–e show the behavior of performance measures as functions of  $K$  over various values of  $\sigma$ , where the values of other probabilities are fixed as follows:  $\alpha = 0.2, \beta = 0.4, \gamma = 0.7$ . Plot 6(a) exhibits that  $PB_p$  is almost not affected by an increment in  $\sigma$ , and it decreases with the increasing number of servers. Conversely, plot 6(b) exhibits that, for a small value of  $K, PB_r$  is affected by an increment in  $\sigma$ , but when  $K \geq 7$ , the impact of the values of  $\sigma$  is negligible and with increasing  $\sigma, PB_r$  increases. Note that both the  $NBS_{av}$  and the  $NFS_{av}$  increase with respect to  $K$ , and the values are almost independent of  $\sigma$  (see Figure 5c,d). It is seen from Figure 6e that for a small value of  $K$ , the  $NCO_{av}$  is affected by an increment in  $\sigma$ , but when  $K \geq 4$ , the impact of the values of  $\sigma$  is negligible and with increasing  $\sigma$ , the  $NCO_{av}$  decreases.

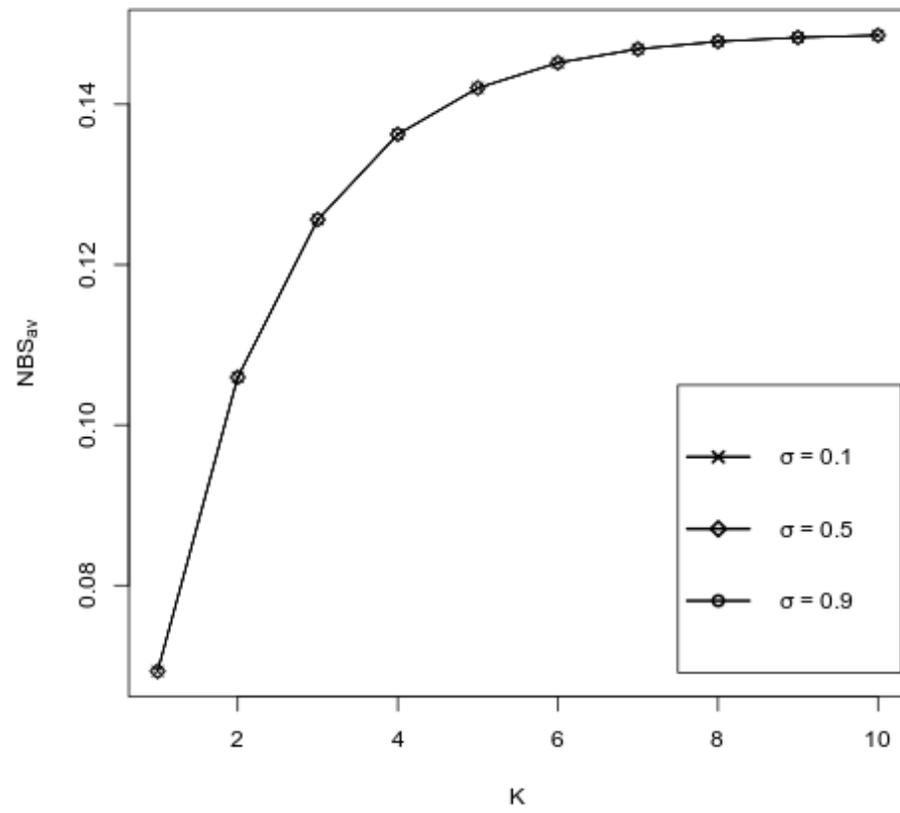


(a)

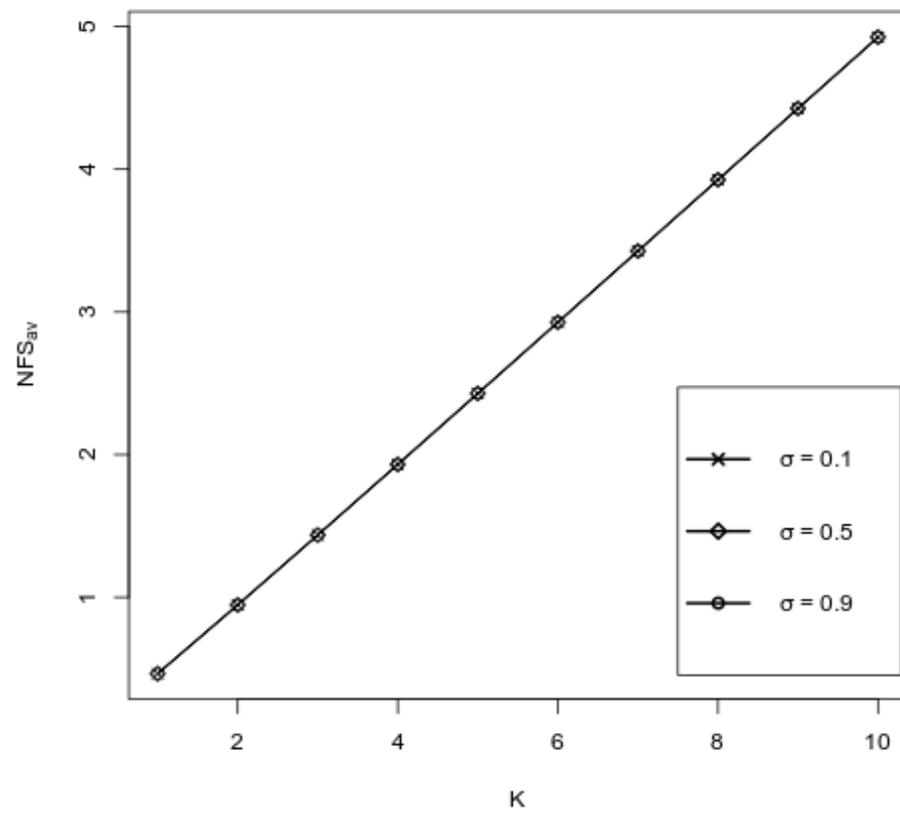


(b)

Figure 6. Cont.

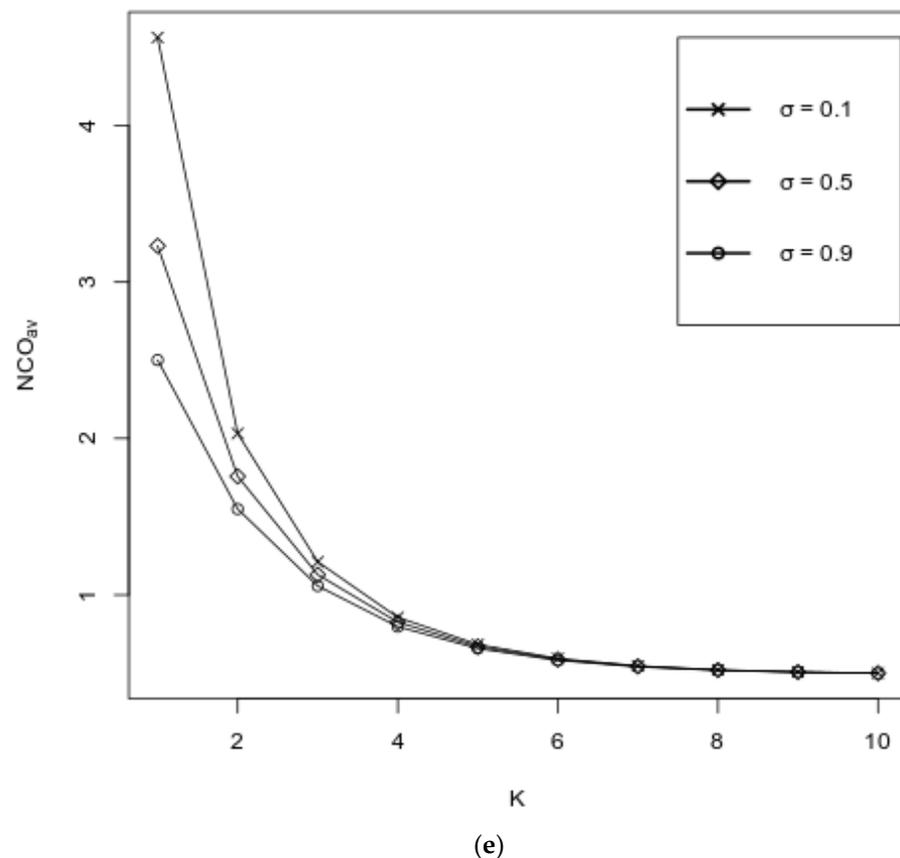


(c)



(d)

Figure 6. Cont.



**Figure 6.** Loss probabilities of  $p$ -customers (a) and  $r$ -customers (b), the average number of busy servers (c) and failed servers (d) and the average number of  $r$ -customers in the orbit (e) versus the number of servers under various values of  $\sigma$ .

## 7. Conclusions

The present paper proposes mathematical models of unreliable multi-server RQwDFB. We studied both models with common finite and infinite orbit capacity for retrial and feedback customers. In a model with a finite orbit capacity, it is assumed that the probabilities of retrial, feedback, and balking from the orbit depend on the current number of customers in the orbit. For such a kind of model, we propose an exact method for calculating its steady-state probabilities as well as its performance measures. For a model with an infinite orbit size, an approximate space merging method is developed for calculating its steady-state probabilities as well as its performance measures.

As for directions for further research, it is proposed that models of RQwDFB with MAP flow and phase-type distributions of service time be studied. Practical interests are represented by the problems of optimizing RQwDFB regarding the selected criteria for the quality of their functioning. These problems are the subjects of future works.

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

**Funding:** The work of János Sztrik is supported by the EFOP-3.6.1-16-2016-00022 project. The project is co-financed by the European Union and the European Social Fund.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors are very grateful to the reviewers for their valuable comments and suggestions, which improved the quality and the presentation of the paper.

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

## References

1. Phung-Duc, T. Retrial Queueing Models: A Survey on Theory and Applications. *arXiv* **2019**, arXiv:1906.09560.
2. Ayyappan, G.; Subramanian, A.M.G.; Sekar, G. M/M/1 retrial queueing system with loss and feedback under non-pre-emptive priority service by matrix-geometric method. *Appl. Math. Sci.* **2010**, *4*, 2379–2389.
3. Ayyappan, G.; Subramanian, A.M.G.; Sekar, G. M/M/1 retrial queueing system with loss and feedback under pre-emptive priority service. *Int. J. Comp. Appl.* **2010**, *2*, 27–34. [[CrossRef](#)]
4. Ayyappan, G.; Devipriya, G.; Subramanian, A.M.G. Analysis of single server queueing system with batch service under multiple vacations with loss and feedback. *Math. Theory Model.* **2014**, *4*, 79–90.
5. Bouchentouf, A.A.; Belarbi, F. Performance evaluation of two Markovian retrial queueing model with balking and feedback. *Acta Univ. Sapientiae Math.* **2013**, *5*, 132–146.
6. Choi, B.D.; Kim, Y.C.; Lee, Y.W. The M/M/c retrial queue with geometric loss and feedback. *Comp. Math. Appl.* **1998**, *36*, 41–52. [[CrossRef](#)]
7. Korolyuk, V.S.; Melikov, A.Z.; Ponomarenko, L.A.; Rustamov, A.M. Methods for analysis of multi-channel queueing models with instantaneous and delayed feedbacks. *Cybern. Syst. Anal.* **2016**, *52*, 58–70. [[CrossRef](#)]
8. Krishnamoorthy, A.; Manjunath, A.S. On Queues with Priority Determined by Feedback. *Calcutta Stat. Assoc. Bull.* **2018**, *70*, 33–56. [[CrossRef](#)]
9. Melikov, A.Z.; Sztrik, J.; Aliyeva, S.H. Analysis of queueing system MMPP/M/K/K with delayed feedback. *Mathematics* **2019**, *7*, 1128. [[CrossRef](#)]
10. Chang, C.J.; Ke, J.C.; Chang, F.M. Unreliable retrial queue with loss and feedback under threshold-based policy. *Int. J. Ind. Syst. Eng.* **2018**, *30*, 1–20. [[CrossRef](#)]
11. Chang, F.M.; Liu, T.H.; Ke, J.C. On an unreliable-server retrial queue with customer feedback and impatience. *Appl. Math. Model.* **2018**, *55*, 171–182. [[CrossRef](#)]
12. Jain, M.; Kaur, S. Bernoulli vacation model for M[X]/G/1 unreliable server retrial queue with Bernoulli feedback, balking and optional service. *RAIRO-Oper. Res.* **2020**, *55*, S2027–S2053. [[CrossRef](#)]
13. Ke, J.C.; Liu, T.H.; Su, S.; Zhang, Z.G. On retrial queue with customer balking and feedback subject to server breakdowns. *Commun. Stat. Theory Methods* **2020**, 1–17. [[CrossRef](#)]
14. Mokaddis, G.S.; Metwally, S.A.; Zaki, B.M. A Feedback Retrial Queueing System with Starting Failures and Single Vacation. *Tamkang J. Sci. Eng.* **2007**, *10*, 183–192.
15. Rajadurai, P.; Sundararaman, M.; Narasimhan, D. Performance analysis of an M/G/1 retrial G-queue with feedback under breakdown services. *Songklanakarin J. Sci. Technol.* **2020**, *42*, 236–247.
16. Rajadurai, P.; Saravananarajan, M.C.; Chandrasekaran, V.M. A study on M/G/1 feedback retrial queue subject to server breakdown and repair under multiple working vacation policy. *Alex. Eng. J.* **2018**, *57*, 947–962. [[CrossRef](#)]
17. Singh, C.J.; Kaur, S.; Jain, M. Waiting time of bulk arrival unreliable queue with balking and Bernoulli feedback using maximum entropy principle. *J. Stat. Theory Pract.* **2017**, *11*, 41–62. [[CrossRef](#)]
18. Brian, P.C. Approximate Analysis of an Unreliable M/M/2 Retrial Queue. Master's Thesis, Air Force Institute of Technology, Wright-Patterson Air Force Base, OH, USA, 2007; p. 84.
19. Elhaddad, M.; Belarbi, F. Approximate analysis of an unreliable M/M/c retrial queue with phase merging algorithm. *New Trends Math. Sci.* **2016**, *4*, 9–21. [[CrossRef](#)]
20. Raiah, L.; Oukid, N. An M/M/2 retrial queue with breakdowns and repairs. *Rom. J. Math. Comp. Sci.* **2017**, *17*, 11–20.
21. Nazarov, A.A.; Sztrik, J.; Kvach, A.; Toth, A. Asymptotic sojourn time analysis of Markov finite-source M/M/1 retrial queueing system with collisions and server subject to breakdowns and repairs. *Ann. Oper. Res.* **2020**, *288*, 417–434. [[CrossRef](#)]
22. Melikov, A.Z.; Fattakhova, M.I. Computational algorithms to optimization of buffer allocation strategies in a packet switching networks. *Appl. Comp. Math.* **2002**, *1*, 51–58.
23. Liang, C.C.; Luh, H. Cost estimation queueing model for large-scale file delivery service. *Int. J. Electr. Commer. Stud.* **2011**, *2*, 19–34.
24. Liang, C.C.; Luh, H. Optimal services for content delivery based on business priority. *J. Chin. Inst. Eng.* **2013**, *36*, 422–440. [[CrossRef](#)]
25. Liang, C.C.; Luh, H. Solving two-dimensional Markov chain model for call center. *Ind. Manag. Data Syst.* **2015**, *115*, 901–922. [[CrossRef](#)]
26. Neuts, M.F. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*; John Hopkins University Press: Baltimore, MD, USA, 1981; 332p.
27. Korolyuk, V.S.; Korolyuk, V.V. *Stochastic Models of Systems*; Kluwer Academic Publishers: Boston, MA, USA, 1999; 185p.