mathematics

MDPI

Article

Modeling of Junior Servers Approaching a Senior Server in the Retrial Queuing-Inventory System

Kathirvel Jeganathan [1], Thanushkodi Harikrishnan [2], Kumarasankaralingam Lakshmanan [3], Agassi Melikov [4] and Janos Sztrik [5,*]

[1] Ramanujan Institute for Advanced Study in Mathematics, University of Madras, Chennai 600005, India; kjeganathan@unom.ac.in
[2] Guru Nanak College (Autonomous), University of Madras, Chennai 600042, India; harikrishanan@gurunanakcollege.edu.in
[3] Department of Mathematics, St. Joseph University, Chümoukedima 797115, India; coprime65@gmail.com
[4] Institute of Control Systems, National Academy of Science, Baku AZ 1141, Azerbaijan; agassimelikov@isi.az
[5] Department of Informatics and Networks, Faculty of Informatics, University of Debrecen, 4032 Debrecen, Hungary
* Correspondence: sztrik.janos@inf.unideb.hu

Abstract: This article deals with the queuing-inventory system, composed of c junior servers, a senior server, two finite waiting halls, and an infinite orbit. On occasion, junior servers encounter challenges during customer service. In these instances, they approach the senior server for guidance in resolving the issue. Suppose the senior server is engaged with another junior server. The approaching junior servers await their turn in a finite waiting area with a capacity of c for consultation. Concerning this, we study the performance of junior servers approaching the senior server in the retrial queuing-inventory model with the two finite waiting halls dedicated to the primary customers and the junior servers for consultation. We formulate a level-dependent QBD process and solve its steady-state probability vector using Neuts and Rao's truncation method. The stability condition of the system is derived and the $\mathbb{R}$ matrix is computed. The optimum total cost has been obtained, and the sensitivity analyses, which include the expected total cost, the waiting time of customers in the waiting hall and orbit, the number of busy servers, and a fraction of the successful retrial rate of the model, are computed numerically.

Keywords: multi-server; classical retrial facility; (s, Q) ordering policy

MSC: 60K25

Citation: Jeganathan, K.; Harikrishnan, T.; Lakshmanan, K.; Melikov, A.; Sztrik, J. Modeling of Junior Servers Approaching a Senior Server in the Retrial Queuing-Inventory System. *Mathematics* **2023**, *11*, 4581. https://doi.org/10.3390/math11224581

Academic Editor: Steve Drekic

Received: 17 October 2023
Revised: 3 November 2023
Accepted: 6 November 2023
Published: 8 November 2023

1. Introduction

Over the past two decades, researchers have dedicated significant efforts to developing a range of queuing-inventory models. In this paper, we introduce a distinctive queuing-inventory model characterized by a novel feature: Junior servers can seek guidance from senior servers when facing challenges in customer service. This interactive dynamic injects a fresh perspective into the conventional server–customer relationship. Within this innovative system, junior servers, though proficient in their roles, have the option to consult senior servers when encountering complexities in customer interactions. This collaborative approach, akin to seeking mentorship from a seasoned expert, enhances the efficiency and problem-solving capabilities of the queuing-inventory framework. This research aims to thoroughly investigate this intriguing model, scrutinizing its performance and the intricate interplay between junior and senior servers. Ultimately, it offers invaluable insights into the dynamics of this specialized queuing-inventory system.

1.1. Motivation

The considered queuing-inventory model was developed based on the real-life experience of one of the authors. When the author went to purchase an android from the local

multi-brand mobile showroom located in Chennai, junior sales executives received him. While the junior executive explained the features of the mobile, the author asked a query regarding the credit card offer. Unfortunately, the junior executive could not address the issue. So, the junior executive sought help from a senior executive. Then, both junior and senior executives together addressed the author's query. The incident happened in the author's life and motivated him to develop the retrial queuing-inventory model where less experienced junior servers assist customers by approaching the experienced senior server.

1.2. Literature Review

Positive service time was originally introduced by Melikov and Molchanov [1] as well as Sigman and Simchi-Levi [2] within the context of inventory modeling. They considered a facility that utilizes inventory to cater to customers, with the service time being distributed arbitrarily. Yadavalli et al. [3] conducted an analysis on a perishable inventory model employing continuous review and a multi-server service configuration with MAP. Additionally, they took into account a separate group of negative customers following an independent MAP. In this setup, a negative customer displaces one of the existing customers in the queue. Under the (s, S) ordering policy, Yadavalli et al. [4] examined a multi-server service facility with the inventory system in the finite population in which any arriving customer who finds that all the servers are busy are then sent to the orbit. Nair et al. [5] studied the behavior of a two-dimensional multi-server queuing-inventory system (QIS) with the (s, S) ordering policy. When there are at least $s + 1$ customers in the system, they guarantee a minimal service with rate $(s + 1)\mu$. Krishnamoorthy et al. [6] assumed that the customer may receive an item at the end of the service under Bernoulli's schedule in the QIS with a multi-server facility. For the original two unbounded level challenges, Wang [7] proposed two modeling approximations. They applied enumerative and quasi-Newton search methods in a heuristic manner to fine-tune the quantity of the stock, servers, reorder thresholds, and both the service and retrial facility. Wang et al. [8] studied a priority-type multi-server inventory model with an identical service time distribution. For the case of non-identical service rates, generalized stochastic Petri net (GSPN) models have been used. When service rates are the same, they demonstrate how the two approaches are equivalent.

Hanukov et al. [9] analyzed an inventory model denoted as $M/M/2$, where customer interest is piqued by observing the quantity of available stocks. Jeganathan et al. [10] conducted a comparative study on a Markovian inventory model employing dual servers, examining both homogeneous and heterogeneous server configurations. Their numerical results underscored the superior efficiency of the heterogeneous system. Suganya and Sivakumar [11] delved into QIS featuring retrials, incorporating two distinct servers and incorporating vacation periods. In this scenario, customers arrive following a MAP, and the two concurrent servers offer services in different phases. Jose and Beena [12] investigated a production inventory system employing two servers of differing capabilities, considering retrial customers and allowing for server vacations. One server had the flexibility for multiple breaks, while the other operated continuously without any time off. The assumption was made that either when the inventory's stock level reaches zero or when both conditions are met, the servers would go on vacation. Chakravarthy and Rumyantsev [13] explored batch demands within two distinct models. Both models assumed that the demands followed a Markovian point process. In the first model, if a customer arrived and found the inventory level at zero, the customer would be lost. In the second model, any waiting customers would be lost when the inventory reached zero. Jeganathan and Reiyas [14] analyzed delayed and modified working vacations in the QIS with two distinct servers in which the first server is capable of executing a modified working vacation and the second server can engage in a delayed working vacation.

Chakravarthy et al. [15] investigated a QIS with infinite servers and a queue with a size of infinite. They assumed the self-service facility under exponential distribution and MAP for customer arrival in the system. Hanukov et al. [16] studied the preparatory service in order to decrease the waiting time of customers during the idle time in a multi-server QIS

with stock-dependent arrival. Jeganathan et al. [17] considered two groups of multi-servers in such a way that each group of servers is designed to operate for two types of customers. The first type of customers can purchase the commodity whereas the second type can receive the service only. Rasmi et al. [18] addressed a multi-server queuing-inventory system featuring diverse customer types (K in total) arriving in accordance with a marked Markovian arrival process. Each customer class seeks a distinct type of service and distinct priorities are allocated to each class. Rasmi and Jacob [19] explored a Markovian QIS with c servers and a self-service feature is optionally available. Service for the customers is initiated whenever an item and a server is available. In the event that a customer arrives to find all servers occupied but free inventory is still accessible, they have the choice to either wait in line for a server or opt for self-service. In the multi-server production inventory system, when the on-hand inventory level drops to zero, an immediate replenishment of one item is initiated as an emergency measure to meet customer demands, with no lead time considered by Shajin et al. [20]. Jeganathan et al. [21] studied a queuing-inventory model offering sales of fresh and refurbished items and services alone to the customers managed by the dedicated servers.

Almaqbali et al. [22] analyzed a batch arrival and service pattern for the customers. If there are j customers in a batch, it is called category j and $j = 1, 2, \ldots, k$. A customer in category j can receive service only when j items are in stock. Aghsami et al. [23] proposed a multi-server queuing-inventory model based on the hospital blood bank, considering the request for blood as arrival and blood as inventory. Selvakumar et al. [24] analyzed the home delivery service facility in the QIS considering two distinct servers. The first server was always available in the system in order to maintain the sales of items, and the second server was used for the delivery process of purchased items who is offered the vacation facility. Yue et al. [25] conducted a study on a multi-server QIS with a vacation policy. In this setup, a subset of servers take a collective vacation once the on-hand inventory is exhausted. After each vacation period, if there is still inventory available, these servers return to the system to resume serving customers.

The retrial QIS finds extensive applications across various domains, including supply chain and manufacturing systems. Artalejo et al. [26] pioneered the incorporation of the retry strategy of customers into stochastic inventory systems, examining numerical solutions and optimal decision making. Building upon this work, Ushakumari [27] investigated a QIS with retrials and random lead time, deriving the optimal ordering point. Amirthakodi and Sivakumar [28] examined a QIS with a finite queue and assumed that unsatisfied customers could potentially enter an orbit, where they could attempt service directly if the server is available. The authors investigated the distribution of waiting times for both the queue and retrial queues. Lopez-Herrero and Jesus [29] focused on assessing waiting durations, reorder intervals, and the duration of a pending request within a finite retrial group. Hanukov [30] proposed that customers have the option to move to the orbit during their service, making more efficient use of their time. Sugapriya et al. [31] examined the stock-dependent demand in a retrial QIS. Melikov et al. [32] assumed that in the event of a primary customer arriving when the inventory level is zero, this customer would, according to the Bernoulli scheme, either depart from the system or enter an unbounded buffer to reiterate their request at a later time.

Nithya et al. [33] introduced controlled arrivals in a retrial QIS with an essential interruption and an intermittently available server. Reiyas and Jeganathan [34] conducted a study on a standard retrial QIS that incorporates a two-component demand rate. Jain and Kumar [35] carried out an analysis of the optimization of costs in a QIS with two-level supply modes, retrial demands, and numerous vacations. To achieve this optimization, the researchers employed a genetic algorithm. Jeganatahan et al. [36] studied the classical retrial queuing model with scrap items where the server takes a vacation once the storage becomes full or there are no customers in the queue. Bazizi et al. [37] explored the optimization of an (s, Q) retrial inventory system with partial backlogging demands using a generalized

stochastic Petri net approach. Very recently, Jeganathan et al. [38] studied the asynchronous vacation policy in the retrial QIS with the (s, Q) ordering policy.

1.3. Research Gap

The previously mentioned studies have explored various service configurations within the QIS. However, there is currently no published research in the queuing-inventory domain specifically addressing scenarios where junior servers consult the senior server on behalf of customers. Moreover, there is a scarcity of literature on queuing models involving junior servers approaching the senior server. Chakaravarthy [39] investigated a queuing model where the primary server (senior server) not only serves customers directly but also provides guidance to other servers (junior servers). Junior servers can seek advice only while actively assisting a customer, and the senior server addresses these requests in the order they are received. Priority is given to regular service over consultations, even if it requires interrupting their service to customers. Recently, Hanukov [40] examined a queuing model involving n multi-servers. Initially, a junior server handles the first phase, which may involve gathering information or providing an initial service. The junior server then collaborates with a senior server to complete the service together. However, if the senior server is already assisting another junior server, the latter must join a queue along with other servers awaiting the availability of the senior server. Very recently, Chakravarthy et al. [41] investigated a novel queuing model where the system attempts to recruit secondary servers from the pool of consumers who have already received services and expressed an interest in serving.

To fill this research gap, we present an innovative model featuring a team of c junior servers working alongside a senior server in the QIS, implementing the (s, Q) ordering policy. We conduct a thorough examination of its performance and distinctive attributes.

1.4. Novelty and Contribution of the Model

- **Innovative Service Configuration:** This research introduces a unique stochastic retrial queuing-inventory system where junior servers provide service to the customers and receive consultation from the senior server as per the specific circumstances.
- **Analysis of the System:** The paper contributes by utilizing the Neuts and Rao truncation method to solve the level-dependent QBD and Neuts matrix geometric method to establish the stability condition and calculate the stationary probability vector.
- **Comprehensive Performance Analysis:** Through rigorous numerical analysis, the study delves into various critical aspects including the expected total cost, waiting time, and the workload of junior servers. This detailed investigation provides practical insights into the system's operational efficiency under different modes of operation for junior servers.

Overall, this research advances the understanding of stochastic retrial QIS by proposing an innovative service approach, employing a specialized analytical method, and conducting a thorough performance assessment. It contributes significantly to both theoretical and practical domains within queuing-inventory modeling.

The subsequent sections of this paper are organized as follows: Section 2 outlines the specific assumptions underlying our proposed model. In Section 3, we provide the mathematical formulation and conduct an in-depth analysis of the model. Section 4 calculates the $\mathbb{R}$ matrix and the probability vector in a steady state. We establish a set of metrics to evaluate the system's performance in Section 5. The model's effectiveness is evaluated via numerical analysis in Section 6, and we conclude with a summarizing section in Section 7.

2. Model Description

The study examines an inventory system characterized by a maximum storage capacity of S items, two physical queues with capacities N and c, and an infinite capacity virtual waiting area (Orbit). The primary customers arrive at the waiting area of size N under

the Poisson process at a rate of λ. If the waiting hall is full, any arriving customers compulsorily enter the orbit. The customer in the orbit always tries to enter the waiting hall. The retrial is successful when the number of customers in the waiting hall is less than N. The duration between consecutive attempts of orbital customers, known as the inter-retrial time, is estimated via an exponential distribution with the rate θ and it occurs under the classical retrial policy. The system employs a group of c junior servers to serve customers, with an additional senior server available to assist juniors facing difficulties during service. The service time of each junior server follows an exponential distribution with the parameter μ. Junior servers can operate in two modes: server mode (S-mode) and consultant mode (C-mode).

- When the junior server takes on the duty of offering service to the customers in the waiting hall, he performs the role of a server, known as server mode (S-mode).
- When a junior server seeks assistance from the senior server in solving the encountered problem in S-mode, the junior server receives consultation from the senior server (Consultant), known as consultant mode (C-mode).

If a customer arrives and the inventory is available with at least one free junior server, the service begins immediately. However, if the inventory is empty or all junior servers are occupied, the customer must wait in the waiting area (if the waiting area is not full). If the junior server successfully serves the customer (i.e., no issues arise), having completed the purchase, the customer exits the system with probability p. In cases where the junior server encounters a problem during service, they approach the senior server for consultation. In such a situation, the customer remains in the waiting area and the junior server, acting on the customer's behalf, enters the queue of size c dedicated for the junior servers to consult with the senior server with a probability of q. The junior server informs the issue faced in S-mode to the senior server when it is available. Upon providing a comprehensive explanation of the matter to the senior server, a collaborative effort is undertaken by the senior server and the junior server to successfully provide the service. Upon completion, the customer leaves the system after making a purchase of an item, and the junior server transitions back from C-mode to S-mode. The duration of this service follows an exponential distribution with a rate of α. If both the inventory and customer (excluding the servicing customer) are available, the junior server immediately resumes service in S-mode. Otherwise, they remain idle in S-mode. When the junior server approaches the senior server and finds the senior server is assisting some other junior server, the junior server is required to wait in the waiting area of size c until it is their time to be served.

The system follows a (s, Q) ordering policy, where an order for $Q(=S-s)$ items is placed from an external supplier as soon as the inventory level reaches the designated threshold, s. The lead time for orders also follows an exponential distribution with the rate β. The relationship between the number of junior servers and the reorder point is defined as $c < s$. The flow chart of the considered model is given in Figure 1.

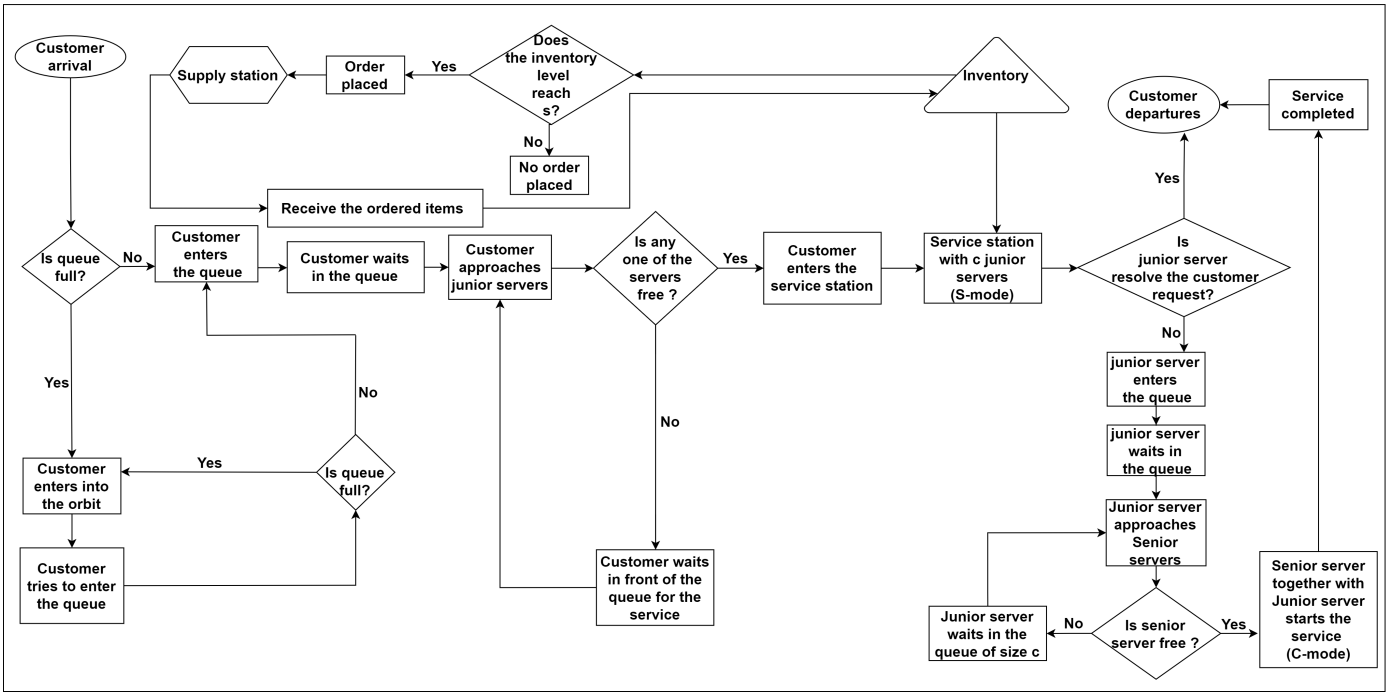

Figure 1. Model flow chart.

3. Mathematical Analysis

Consider the following random variables $A_1(t)$, $A_2(t)$, $A_3(t)$, and $A_4(t)$, representing the number of customers in the orbit at time t, the number of customers in the waiting hall at time t, the number of junior servers in C-mode at time t, and the number of items in the inventory at time t, respectively. The collection $A(t) = \{(A_1(t), A_2(t), A_3(t), A_4(t)) : t \geq 0\}$ constitutes a four-dimensional stochastic process with the state space $P = \bigcup_{i=1}^{2} P_i$ where

$P_1 = \{(a_1, a_2, a_3, a_4) \mid a_1 \in \overline{0, \infty}, a_2 \in \overline{0, c}, a_3 \in \overline{0, a_2}, a_4 \in \overline{a_3, S}\}$
$P_2 = \{(a_1, a_2, a_3, a_4) \mid a_1 \in \overline{0, \infty}, a_2 \in \overline{c+1, N}, a_3 \in \overline{0, c}, a_4 \in \overline{a_3, S}\}$

The stochastic process $A(t)$, which is characterized by discrete state space and continuous time, exhibits the Markov property. Furthermore, it can be shown that any state in P is accessible from any other state. Therefore, we may classify $A(t)$ as a continuous time irreducible Markov chain (CTIMC). The CTIMC is described by the infinitesimal generator matrix given below.

$$T = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ 3 \\ \vdots \end{array} \begin{array}{c} \begin{array}{ccccccc} 0 & 1 & 2 & 3 & 4 & \dots \end{array} \\ \begin{pmatrix} T_{0,0} & T_{0,1} & & & & \\ T_{1,0} & T_{1,1} & T_{0,1} & & & \\ & T_{2,0} & T_{2,2} & T_{0,1} & & \\ & & T_{3,0} & T_{3,3} & T_{0,1} & \\ & & & \ddots & \ddots & \ddots \end{pmatrix} \end{array}, \tag{1}$$

where the $T_{0,1}$ matrix symbolizes the transitions of the customers joining the orbit. This happens when a primary customer arrives and there is no place for customers in the waiting hall.

$$T_{0,1} = \begin{cases} \lambda I_{(c+1)S+1-\frac{c(c-1)}{2}}, & a_2 = N;\ a_2' = a_2; \\ 0, & \text{Otherwise.} \end{cases}$$

The matrices $T_{a_1,0}$ for $a_1 = \overline{1,\infty}$ represent the transitions of retrial customers entering the waiting hall.

$$T_{a_1,0} = \begin{cases} a_1\theta I_{S+1-a_3}, & a_2 = \overline{0,c-1};\ a_2' = a_2+1; \\ & a_3 = \overline{0,a_2};\ a_3' = a_3; \\ a_1\theta I_{S+1-a_3}, & a_2 = \overline{c,N-1};\ a_2' = a_2+1; \\ & a_3 = \overline{0,c};\ a_3' = a_3; \\ 0, & \text{Otherwise.} \end{cases}$$

The matrices T_{a_1,a_1} for $a_1 = \overline{0,\infty}$ represent the transitions of all other remaining parameters, and the entries along the diagonal are populated by the total sum of elements in their respective rows, with an opposite sign to ensure that the sum of all row entries equals zero.

$$T_{a_1,a_1} = \begin{cases} \mathbb{D}_{a_2}, & a_2 = \overline{0,c};\ a_2' = a_2; \\ \mathbb{D}_c, & a_2 = \overline{c+1,N-1};\ a_2' = a_2; \\ \mathbb{D}_N, & a_2 = N;\ a_2' = a_2; \\ \mathbb{F}_{a_2}, & a_2 = \overline{0,c-1};\ a_2' = a_2+1; \\ \mathbb{F}_c, & a_2 = \overline{c,N-1};\ a_2' = a_2+1; \\ \mathbb{E}_{a_2}, & a_2 = \overline{1,c};\ a_2' = a_2-1; \\ \mathbb{E}_{c+1}, & a_2 = \overline{c+1,N};\ a_2' = a_2-1; \\ 0, & \text{Otherwise.} \end{cases}$$

For $a_2 = \overline{0,c}$

$$\mathbb{D}_{a_2} = \begin{cases} \beta, & a_3 = \overline{0,a_2};\ a_3' = a_3; \\ & a_4 = \overline{a_3,s};\ a_4' = a_4+Q; \\ \bar{\delta}_{a_20}(a_4-a_3)q\mu, & a_3 = \overline{0,a_2-1};\ a_3' = a_3+1; \\ & a_4 = \overline{a_3+1,a_2};\ a_4' = a_4; \\ \bar{\delta}_{a_20}(a_2-a_3)q\mu, & a_3 = \overline{0,a_2-1};\ a_3' = a_3+1; \\ & a_4 = \overline{a_2+1,S};\ a_4' = a_4; \\ -[\lambda+\beta+a_1\theta+ & a_3 = \overline{0,a_2};\ a_3' = a_3; \\ \bar{\delta}_{a_20}\bar{\delta}_{a_2a_3}(a_4-a_3)\mu+\bar{\delta}_{a_20}\bar{\delta}_{a_30}\alpha], & a_4 = \overline{a_3,a_2};\ a_4' = a_4; \\ -[\lambda+H(s-a_4)\beta+a_1\theta+ & a_3 = \overline{0,a_2};\ a_3' = a_3; \\ \bar{\delta}_{a_20}(a_2-a_3)\mu+\bar{\delta}_{a_20}\bar{\delta}_{a_30}\alpha], & a_4 = \overline{a_2+1,S};\ a_4' = a_4; \\ 0, & \text{Otherwise.} \end{cases}$$

$$\mathbb{D}_N = \begin{cases} \beta, & a_3 = \overline{0,c};\ a_3' = a_3; \\ & a_4 = \overline{a_3,s};\ a_4' = a_4+Q; \\ (a_4-a_3)q\mu, & a_3 = \overline{0,c-1};\ a_3' = a_3+1; \\ & a_4 = \overline{a_3+1,c};\ a_4' = a_4; \\ (c-a_3)q\mu, & a_3 = \overline{0,c-1};\ a_3' = a_3+1; \\ & a_4 = \overline{c+1,S};\ a_4' = a_4; \\ -[\lambda+\beta+ & a_3 = \overline{0,c};\ a_3' = a_3; \\ \bar{\delta}_{a_3c}(a_4-a_3)\mu+\bar{\delta}_{a_30}\alpha], & a_4 = \overline{a_3,c};\ a_4' = a_4; \\ -[\lambda+H(s-a_4)\beta+ & a_3 = \overline{0,c};\ a_3' = a_3; \\ (c-a_3)\mu+\bar{\delta}_{a_30}\alpha], & a_4 = \overline{c+1,S};\ a_4' = a_4; \\ 0, & \text{Otherwise.} \end{cases}$$

For $a_2 = \overline{0, c}$

$$\mathbb{F}_{a_2} = \begin{cases} \lambda, & a_3 = \overline{0, a_2};\ a_3' = a_3; \\ & a_4 = \overline{a_3, S};\ a_4' = a_4; \\ 0, & \text{Otherwise.} \end{cases}$$

For $a_2 = \overline{1, c}$

$$\mathbb{E}_{a_2} = \begin{cases} (a_4 - a_3)p\mu, & a_3 = \overline{0, a_2 - 1};\ a_3' = a_3; \\ & a_4 = \overline{a_3 + 1, a_2};\ a_4' = a_4 - 1; \\ (a_2 - a_3)p\mu, & a_3 = \overline{0, a_2 - 1};\ a_3' = a_3; \\ & a_4 = \overline{a_2 + 1, S};\ a_4' = a_4 - 1; \\ \alpha, & a_3 = \overline{1, a_2};\ a_3' = a_3 - 1; \\ & a_4 = \overline{a_3, S};\ a_4' = a_4 - 1; \\ 0, & \text{Otherwise.} \end{cases}$$

$$\mathbb{E}_{c+1} = \begin{cases} (a_4 - a_3)p\mu, & a_3 = \overline{0, c - 1};\ a_3' = a_3; \\ & a_4 = \overline{a_3 + 1, c};\ a_4' = a_4 - 1; \\ (c - a_3)p\mu, & a_3 = \overline{0, c - 1};\ a_3' = a_3; \\ & a_4 = \overline{c + 1, S};\ a_4' = a_4 - 1; \\ \alpha, & a_3 = \overline{1, c};\ a_3' = a_3 - 1; \\ & a_4 = \overline{a_3, S};\ a_4' = a_4 - 1; \\ 0, & \text{Otherwise.} \end{cases}$$

3.1. Neuts and Rao Matrix Geometric Approximation

The structure of Equation (1) indicates that the assumed Markov chain, denoted as $\{A(t),\ t \geq 0\}$, conforms to the level-dependent QBD process. To solve this system, we employ Neuts and Rao's truncation method [42], which involves capping the orbit level at a specified point, denoted as M. This truncation shifts the system from being level-dependent to level-independent. Essentially, the equilibrium of the system is determined by setting $T_{a_1,0} = T_{M,0}$ and $T_{a_1,a_1} = T_{M,M}$ for all $a_1 \geq M$. In this scenario, the modified generator matrix for $\{A(t),\ t \geq 0\}$ is given by

$$\hat{T} = \begin{pmatrix} T_{01} & T_{10} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ T_{10} & T_{11} & T_{10} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & T_{20} & T_{22} & T_{10} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & T_{M0} & T_{MM} & T_{10} & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & T_{M0} & T_{MM} & T_{10} & \mathbf{0} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

In order to obtain the necessary stationary probability vector, we use the rate matrix T^*, which is generated using this equation $T^* = T_{M0} + T_{MM} + T_{10}$ and is given by

$$T^* = \begin{cases} \hat{\mathbb{D}}_{a_2}, & a_2 = \overline{0, c};\ a_2' = a_2; \\ \hat{\mathbb{D}}_c, & a_2 = \overline{c + 1, N - 1};\ a_2' = a_2; \\ \hat{\mathbb{D}}_N, & a_2 = N;\ a_2' = a_2; \\ \hat{\mathbb{F}}_{a_2}, & a_2 = \overline{0, c - 1};\ a_2' = a_2 + 1; \\ \hat{\mathbb{F}}_c, & a_2 = \overline{c, N - 1};\ a_2' = a_2 + 1; \\ \mathbb{E}_{a_2}, & a_2 = \overline{1, c};\ a_2' = a_2 - 1; \\ \mathbb{E}_{c+1}, & a_2 = \overline{c + 1, N};\ a_2' = a_2 - 1; \\ 0, & \text{Otherwise.} \end{cases}$$

For $a_2 = \overline{0,c}$

$$\hat{\mathbb{D}}_{a_2} = \begin{cases} \beta, & a_3 = \overline{0,a_2};\ a_3' = a_3; \\ & a_4 = \overline{a_3,s};\ a_4' = a_4 + Q; \\ \bar{\delta}_{a_20}(a_4 - a_3)q\mu, & a_3 = \overline{0,a_2-1};\ a_3' = a_3 + 1; \\ & a_4 = \overline{a_3+1,a_2};\ a_4' = a_4; \\ \bar{\delta}_{a_20}(a_2 - a_3)q\mu, & a_3 = \overline{0,a_2-1};\ a_3' = a_3 + 1; \\ & a_4 = \overline{a_2+1,S};\ a_4' = a_4; \\ -[\lambda + \beta + M\theta + & a_3 = \overline{0,a_2};\ a_3' = a_3; \\ \bar{\delta}_{a_20}\bar{\delta}_{a_2a_3}(a_4 - a_3)\mu + \bar{\delta}_{a_20}\bar{\delta}_{a_30}\alpha], & a_4 = \overline{a_3,a_2};\ a_4' = a_4; \\ -[\lambda + H(s - a_4)\beta + M\theta + & a_3 = \overline{0,a_2};\ a_3' = a_3; \\ \bar{\delta}_{a_20}(a_2 - a_3)\mu + \bar{\delta}_{a_20}\bar{\delta}_{a_30}\alpha], & a_4 = \overline{a_2+1,S};\ a_4' = a_4; \\ 0, & \text{Otherwise.} \end{cases}$$

$$\hat{\mathbb{D}}_N = \begin{cases} \beta, & a_3 = \overline{0,c};\ a_3' = a_3; \\ & a_4 = \overline{a_3,s};\ a_4' = a_4 + Q; \\ (a_4 - a_3)q\mu, & a_3 = \overline{0,c-1};\ a_3' = a_3 + 1; \\ & a_4 = \overline{a_3+1,c};\ a_4' = a_4; \\ (c - a_3)q\mu, & a_3 = \overline{0,c-1};\ a_3' = a_3 + 1; \\ & a_4 = \overline{c+1,S};\ a_4' = a_4; \\ -[\lambda + \beta + & a_3 = \overline{0,c};\ a_3' = a_3; \\ \bar{\delta}_{a_3c}(a_4 - a_3)\mu + \bar{\delta}_{a_30}\alpha], & a_4 = \overline{a_3,c};\ a_4' = a_4; \\ -[\lambda + H(s - a_4)\beta + & a_3 = \overline{0,c};\ a_3' = a_3; \\ (c - a_3)\mu + \bar{\delta}_{a_30}\alpha], & a_4 = \overline{c+1,S};\ a_4' = a_4; \\ 0, & \text{Otherwise.} \end{cases}$$

For $a_2 = \overline{0,c}$

$$\hat{\mathbb{F}}_{a_2} = \begin{cases} \lambda + M\theta, & a_3 = \overline{0,c};\ a_3' = a_3; \\ & a_4 = \overline{a_3,S};\ a_4' = a_4; \\ 0, & \text{Otherwise.} \end{cases}$$

Theorem 1. *The steady state probability vector ψ to the matrix T^* is given by*

$$\psi^{(a_2)} = \psi^{(0)}\chi_{a_2},\ \forall\ a_2 \in \overline{0,N} \tag{2}$$

where

$$\chi_0 = I;\ \chi_i = \prod_{j=1}^{i}\Gamma_j$$

$$\Gamma_j = \begin{cases} -\hat{\mathbb{F}}_{j-1}[\hat{\mathbb{D}}_j + \Gamma_{j+1}E_{j+1}]^{-1}, & j \in \overline{1,c}, \\ -\hat{\mathbb{F}}_c[\hat{\mathbb{D}}_c + \Gamma_{j+1}E_{c+1}]^{-1}, & j \in \overline{c+1,N-1}, \\ -\hat{\mathbb{F}}_c\hat{\mathbb{D}}_N^{-1}, & j = N. \end{cases}$$

and $\psi^{(0)}$ is obtained by solving the equations

$$\psi^{(0)}(\hat{\mathbb{D}}_0 + \Gamma_1\mathbb{E}_1) = \mathbf{0}, \tag{3}$$

$$\sum_{a_2=0}^{N}\psi^{(0)}\mathbf{e} = 1. \tag{4}$$

Proof. The probability vector ψ in steady state satisfies the below equations:

$$\psi T^* = 0, \tag{5}$$

$$\psi \mathbf{e} = 1. \tag{6}$$

In Equation (5), after explicitly expressing ψ and T^* and simplifying it, we arrive at the $N+1$ set of equations as follows:

$$\psi^{(a_2)}\hat{\mathbb{D}}_{a_2} + \psi^{(a_2+1)}\mathbb{E}_{a_2+1} = 0,\ a_2 = 0, \tag{7}$$

$$\psi^{(a_2-1)}\hat{\mathbb{F}}_{a_2-1} + \psi^{(a_2)}\hat{\mathbb{D}}_{a_2} + \psi^{(a_2+1)}\mathbb{E}_{a_2+1} = 0,\ a_2 \in \overline{0,c}, \tag{8}$$

$$\psi^{(a_2-1)}\hat{\mathbb{F}}_{c} + \psi^{(a_2)}\hat{\mathbb{D}}_{c} + \psi^{(a_2+1)}\mathbb{E}_{c+1} = 0,\ a_2 \in \overline{c+1, N-2}, \tag{9}$$

$$\psi^{(a_2-1)}\hat{\mathbb{F}}_{c} + \psi^{(a_2)}\hat{\mathbb{D}}_{a_2} = 0,\ a_2 = N-1. \tag{10}$$

The steady-state probability vectors $\psi^{(a_2)}$ for all a_2 in the set $\overline{0,N}$ are obtained by solving the system of Equations (7) and (10) iteratively. This allows us to express these vectors in terms of the steady-state probability vector $\psi^{(0)}$ as shown in Equation (2). Solving Equations (3) and (4), we obtain $\psi^{(0)}$ and use the vector $\psi^{(0)}$ to obtain ψ. □

Theorem 2. *The inequality*

$$\sum_{a_3=0}^{c}\sum_{a_4=a_3}^{S}\psi^{(M,N,a_3,a_4)}\lambda < \left\{\sum_{a_2=0}^{c}\sum_{a_3=0}^{a_2}\sum_{a_4=a_3}^{S} + \sum_{a_2=c+1}^{N-1}\sum_{a_3=0}^{c}\sum_{a_4=a_3}^{S}\right\}\psi^{(M,a_2,a_3,a_4)}M\theta \tag{11}$$

gives the stability condition for the infinitesimal generator matrix, $\hat{T}$.

Proof. Based on the result given by Neuts [43], the existence of a steady-state probability vector ψ for the modified infinitesimal generator matrix $\hat{T}$ is dependent upon the fulfillment of the following condition.

$$\psi T_{0,1}\mathbf{e} < \psi T_{M,0}\mathbf{e}. \tag{12}$$

Through the above Theorem (1) in inequality (12) and writing elaborately on all ψ, $T_{0,1}$, $T_{M,0}$ and $\mathbf{e}$ and simplifying it, the required stability condition in (11) is obtained. □

3.2. Limiting Probability Distribution

The generating matrix T yields the steady-state probability vector $\varphi = (\varphi^{(0)}, \varphi^{(1)}, \varphi^{(2)}, \cdots)$, which meets the stability requirement. Consequently, the Markov process denoted as

$$\{(A_1(t),\ A_2(t),\ A_3(t),\ A_4(t)),\ t \geq 0\},$$

with a state space denoted as P is classified as regular. Hence, the ultimate probability distribution is denoted as

$$\varphi^{(a_1,a_2,a_3,a_4)} = \lim_{t\to\infty} Pr[A_1(t) = a_1, A_2(t) = a_2, A_3(t) = a_3, A_4(t) = a_4 \mid A_1(0) = 0, A_2(0) = 0, A_3(0) = 0, A_4(0) = 0].$$

It exists and is observed to be independent of the initial state.

4. Computation of $\mathbb{R}$ Matrix

To obtain the steady-state probability vector $\varphi = (\varphi^{(0)}, \varphi^{(1)}, \varphi^{(2)}, \dots)$, it is essential to calculate the rate matrix $\mathbb{R}$.

Theorem 3. *The rate matrix* $\mathbb{R}$ *is the minimal non-negative solution of the matrix quadratic equation*

$$T_{M,0}\mathbb{R}^2 + T_{M,M}\mathbb{R} + T_{0,1} = \mathbf{0}, \tag{13}$$

and the $\mathbb{R}$ *matrix structure is*

$$\mathbb{R} = \begin{array}{c} \\ 0 \\ 1 \\ \vdots \\ N-1 \\ N \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & \cdots & S-1 & S \end{array} \\ \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ \mathfrak{A}^{(0)} & \mathfrak{A}^{(1)} & \cdots & \mathfrak{A}^{(N-1)} & \mathfrak{A}^{(N)} \end{pmatrix} \end{array} \tag{14}$$

For $a_2 \in \overline{0, N}$

$$\mathfrak{A}^{(a_2)} = \begin{array}{c} \\ 0 \\ 1 \\ \vdots \\ c-1 \\ c \end{array} \begin{array}{c} \begin{array}{ccccc} 0 & 1 & \cdots & c-1 & c \end{array} \\ \begin{pmatrix} \Re_{a_2}^{0,0} & \Re_{a_2}^{0,1} & \cdots & \Re_{a_2}^{0,c-1} & \Re_{a_2}^{0,c} \\ \Re_{a_2}^{1,0} & \Re_{a_2}^{1,1} & \cdots & \Re_{a_2}^{1,c-1} & \Re_{a_2}^{1,c} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \Re_{a_2}^{c-1,0} & \Re_{a_2}^{c-1,1} & \cdots & \Re_{a_2}^{c-1,c-1} & \Re_{a_2}^{c-1,c} \\ \Re_{a_2}^{c,0} & \Re_{a_2}^{c,1} & \cdots & \Re_{a_2}^{c,c-1} & \Re_{a_2}^{c,c} \end{pmatrix} \end{array} \tag{15}$$

For $u, v \in \overline{0, c}$

$$\Re_{a_2}^{u,v} = \begin{array}{c} \\ u \\ u+1 \\ \vdots \\ S-1 \\ S \end{array} \begin{array}{c} \begin{array}{ccccc} v & v+1 & \cdots & S-1 & S \end{array} \\ \begin{pmatrix} \vartheta_{a_2,u,v}^{u,v} & \vartheta_{a_2,u,v}^{u,v+1} & \cdots & \vartheta_{a_2,u,v}^{u,S-1} & \vartheta_{a_2,u,v}^{u,S} \\ \vartheta_{a_2,u,v}^{u+1,v} & \vartheta_{a_2,u,v}^{u+1,v+1} & \cdots & \vartheta_{a_2,u,v}^{u+1,S-1} & \vartheta_{a_2,u,v}^{u+1,S} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \vartheta_{a_2,u,v}^{S-1,v} & \vartheta_{a_2,u,v}^{S-1,v+1} & \cdots & \vartheta_{a_2,u,v}^{S-1,S-1} & \vartheta_{a_2,u,v}^{S-1,S} \\ \vartheta_{a_2,u,v}^{S,v} & \vartheta_{a_2,u,v}^{S,v+1} & \cdots & \vartheta_{a_2,u,v}^{S,S-1} & \vartheta_{a_2,u,v}^{S,S} \end{pmatrix} \end{array} \tag{16}$$

Proof. The matrix quadratic Equation (13) is fulfilled by the rate matrix $\mathbb{R}$, which is constructed based on the block tridiagonal structure of the infinitesimal modified generating matrix $\hat{T}$. At first, it is presumed that (14) contains the unknown $\mathbb{R}$-matrix. In fact, the $\mathbb{R}$-matrix structure may be determined simply by counting the number of non-zero rows in the $T_{0,1}$ matrix. Since each row in the final block of the $T_{0,1}$ matrix contains at least one non-zero element, all of the rows in the final block of the $\mathbb{R}$-matrix should be regarded as non-zero rows. These presumptions lead to the structure of the unknown $\mathbb{R}$-matrix being as shown in (14). The following set of non-linear homogeneous equations is derived by using each block matrix in Equation (13).
If $a_2 = N; a_2' \in \overline{0, N}; a_3 \in \overline{0, c-1}; a_3' \in \overline{0, c}; a_4 \in \overline{a_3, c}; a_4' \in \overline{a_3', S};$

$$\left[\lambda + \beta + \delta_{0a_4} a_4 \mu + \bar{\delta}_{a_3 0}\alpha\right] \vartheta_{a_2,a_3,a_3'}^{a_4,a_4'} + \beta \vartheta_{a_2,a_3,a_3'}^{Q+a_4,a_4'} + \delta_{a_4 0} q(a_4 - a_3)\mu \vartheta_{a_2,a_3,a_3'}^{a_4,a_4'} + \lambda \delta_{a_2' N} = 0. \tag{17}$$

If $a_2 = N; a_2' \in \overline{0, N}; a_3 \in \overline{0, c-1}; a_3' \in \overline{0, c}; a_4 \in \overline{c+1, S}; a_4' \in \overline{a_3', S};$

$$\left[\lambda + H(s - a_4)\beta + c\mu + \bar{\delta}_{a_3 0}\alpha\right] \vartheta_{a_2,a_3,a_3'}^{a_4,a_4'} + H(s - a_4)\beta \vartheta_{a_2,a_3,a_3'}^{Q+a_4,a_4'} + q(c - a_3)\mu \vartheta_{a_2,a_3,a_3'}^{a_4,a_4'} + \lambda \delta_{a_2' N} = 0. \tag{18}$$

If $a_2 = N; a_2' \in \overline{0, N}; a_3 = c; a_3' \in \overline{0, c}; a_4 \in \overline{a_3, c}; a_4' \in \overline{a_3', S}$;

$$\left[\lambda + \beta + \alpha\right] \vartheta^{a_4, a_4'}_{a_2, a_3, a_3'} + \beta \vartheta^{Q+a_4, a_4'}_{a_2, a_3, a_3'} + \lambda \delta_{a_2' N} = 0. \tag{19}$$

If $a_2 = N; a_2' \in \overline{0, N}; a_3 = c; a_3' \in \overline{0, c}; a_4 \in \overline{c+1, S}; a_4' \in \overline{a_3', S}$;

$$\left[\lambda + H(s - a_4)\beta + \alpha\right] \vartheta^{a_4, a_4'}_{a_2, a_3, a_3'} + H(s - a_4)\beta \vartheta^{Q+a_4, a_4'}_{a_2, a_3, a_3'} + \lambda \delta_{a_2' N} = 0. \tag{20}$$

Solving the above set of Equations (17)–(20) via Gauss–Seidel Method, we obtain the exact entries of the rate matrix $\mathbb{R}$. □

Theorem 4. *The probability vector* $\varphi^{(a_1)}$, $\forall\, a_1 = 1, 2, 3, \ldots$ *of the Markov chain can be derived by* $\varphi^{(a_1)} = \begin{cases} \varphi^{(0)} \Lambda_{a_1}, & \forall\, a_1 = \overline{0, M}, \\ \varphi^{(0)} \Lambda_M \mathbb{R}^{a_1 - M}, & \forall\, a_1 > M. \end{cases}$
where $\mathbb{R}$ *is the solution of the matrix quadratic Equation* (13) *and*

$$\Lambda_{a_1} = \begin{cases} I, & \text{if } a_1 = 0, \\ \prod_{i=0}^{a_1} T_{0,1} A_i, & \text{if } a_1 = \overline{1, M}. \end{cases}$$

$$X_j = \begin{cases} -(T_{0,1} X_{j+1} T_{j+2,0} + T_{(j+1),(j+1)})^{-1}, & \forall\, j = \overline{0, M-2}, \\ [-(T_{M,M} + \mathbb{R} T_{M,0})]^{-1}, & \forall\, j = M - 1. \end{cases}$$

and

$$\varphi^{(0)} = \left[I + \sum_{a_1=1}^{M-1} \prod_{j=0}^{i-1} T_{0,1} X_j + \prod_{j=0}^{M-1} T_{0,1} X_j (I - \mathbb{R})^{-1} \right]^{-1}$$

Proof. Let $\varphi = (\varphi^{(0)}, \varphi^{(1)}, \varphi^{(2)}, \ldots)$ be a probability vector which satisfies

$$\varphi \hat{T} = \mathbf{0} \text{ and } \varphi \mathbf{e} = 1. \tag{21}$$

Let us use the matrix geometric method to solve (21). Let us consider a rate matrix $\mathbb{R}$ which is the solution of the matrix quadratic equation $\mathbb{R}^2 T_{M,0} + \mathbb{R} T_{M,M} + T_{0,1} = \mathbf{0}$. Let us assume that

$$\varphi^{(a_1)} = \varphi^{(M)} \mathbb{R}^{(a_1 - M)} \quad \forall\, a_1 = \overline{M, \infty}. \tag{22}$$

By solving $\varphi T^* = \mathbf{0}$, we obtain the following system of equations

$$\varphi^{(0)} T_{0,0} + \varphi^{(1)} T_{1,0} = \mathbf{0} \tag{23}$$

$$\varphi^{(a_1 - 1)} T_{0,1} + \varphi^{(a_1)} T_{a_1, a_1} + \varphi^{(a_1 + 1)} T_{a_1 + 1 0} = \mathbf{0}\ \forall\, a_1 = \overline{1, n-1} \tag{24}$$

$$\varphi^{(M-1)} T_{0,1} + \varphi^{(M)} (T_{M,M} + \mathbb{R} T_{M,0}) = \mathbf{0} \tag{25}$$

$$\text{and } \left[\sum_{a_1=0}^{M-1} \varphi^{(a_1)} + \varphi^{(M)} (I - \mathbb{R})^{-1} \right] \mathbf{e} = 1. \tag{26}$$

From (25) $\varphi^{(M)} = \varphi^{(M-1)} T_{0,1} X_{M-1}$, where $X_{M-1} = [-(T_{M,M} + \mathbb{R} T_{M,0})]^{-1}$.
From (24) $\varphi^{(M-1)} = \varphi^{(M-2)} T_{0,1} X_{M-2}$, where $X_{M-2} = [-(T_{(M-1)\times(M-1)} + T_{0,1} G_{M-1} T_{M,0})]^{-1}$

Using (24) again, similarly we obtain $\varphi^{(M-2)} = \varphi^{(M-3)} T_{0,1} X_{M-3}$
where $X_{M-3} = [-(T_{(M-2),(M-2)} + T_{0,1} X_{M-2} T_{M,0})]^{-1}$

In general

$$\varphi^{(a_1)} = \varphi^{(a_1 - 1)} T_{0,1} X_{a_1 - 1}\ \forall\, a_1 = \overline{1, M} \tag{27}$$

where $X_{a_1} = \begin{cases} [-(T_{M,M} + \mathbb{R}T_{M,0})]^{-1} & a_1 = M-1 \\ [-(T_{a_1,a_1} + T_{0,0}A_{a_1}T_{a_1+1,1})]^{-1} & a_1 = \overline{1, M-2} \end{cases}$

From (27)

$$\varphi^{(a_1)} = \varphi^{(0)}\Lambda_{a_1} \; \forall \; a_1 = \overline{0, M} \tag{28}$$

$$\aleph_{a_1} = \begin{cases} I & \text{if } a_1 = 0 \\ \prod_{i=0}^{a_1} T_{0,1}A_i & \text{if } a_1 = \overline{1, M} \end{cases} \tag{29}$$

By substituting (28) in (25), we obtain

$$\varphi^{(0)}\left[I + \sum_{a_1=0}^{M-1} \aleph_{a_1} + \aleph_M(I - \mathbb{R})^{-1}\right]\mathbf{e} = 1 \tag{30}$$

By substituting (29) in (30), we obtain

$$\varphi^{(0)} = \left[I + \sum_{a_1=1}^{M-1}\prod_{i=0}^{a_1-1} T_{0,1}X_i + \prod_{i=0}^{M-1} T_{0,1}X_i(I - \mathbb{R})^{-1}\right]^{-1}. \tag{31}$$

Using (31) in (28), we can obtain each $\varphi^{(i)}$. □

5. System Performance Measures

In this section, various performance measures of the system are defined to construct the total cost function and to illustrate numerical examples.

1. **Mean Inventory level:**

$$Z_1 = \sum_{a_1=0}^{\infty}\sum_{a_2=0}^{c}\sum_{a_3=0}^{a_2}\sum_{a_4=a_3}^{S} a_4\varphi^{(a_1,a_2,a_3,a_4)} + \sum_{a_1=0}^{\infty}\sum_{a_2=c+1}^{N}\sum_{a_3=0}^{c}\sum_{a_4=a_3}^{S} a_4\varphi^{(a_1,a_2,a_3,a_4)}$$

2. **Mean Reorder Rate:**

$$Z_2 = \sum_{a_1=0}^{\infty}\sum_{a_2=1}^{c}\sum_{a_3=0}^{a_2-1}(a_2 - a_3)p\mu\varphi^{(a_1,a_2,a_3,s+1)} + \sum_{a_1=0}^{\infty}\sum_{a_2=c+1}^{N}\sum_{a_3=0}^{c-1}(c - a_3)p\mu\varphi^{(a_1,a_2,a_3,s+1)}$$
$$+ \sum_{a_1=0}^{\infty}\sum_{a_2=1}^{c}\sum_{a_3=1}^{a_2}\alpha\varphi^{(a_1,a_2,a_3,s+1)} + \sum_{a_1=0}^{\infty}\sum_{a_2=c+1}^{N}\sum_{a_3=1}^{c}\alpha\varphi^{(a_1,a_2,a_3,s+1)}$$

3. **Mean Number of Customers in Waiting Hall:**

$$Z_3 = \sum_{a_1=1}^{\infty}\sum_{a_2=0}^{c}\sum_{a_3=0}^{a_2}\sum_{a_4=a_3}^{S} a_2\varphi^{(a_1,a_2,a_3,a_4)} + \sum_{a_1=0}^{\infty}\sum_{a_2=c+1}^{N}\sum_{a_3=0}^{c}\sum_{a_4=a_3}^{S} a_2\varphi^{(a_1,a_2,a_3,a_4)}$$

4. **Mean Number of Customers Enter into the Waiting Hall:**

$$Z_4 = \sum_{a_1=0}^{\infty}\sum_{a_2=0}^{c}\sum_{a_3=0}^{a_2}\sum_{a_4=a_3}^{S} \lambda\varphi^{(a_1,a_2,a_3,a_4)} + \sum_{a_1=0}^{\infty}\sum_{a_2=c+1}^{N-1}\sum_{a_3=0}^{c}\sum_{a_4=a_3}^{S} \lambda\varphi^{(a_1,a_2,a_3,a_4)}$$

5. **Mean Waiting Time of Customers in Waiting Hall:**

$$Z_5 = \frac{Z_3}{Z_4}$$

6. **Mean Number of Junior Servers in C-mode:**

$$Z_6 = \sum_{a_1=1}^{\infty}\sum_{a_2=1}^{c}\sum_{a_3=1}^{a_2}\sum_{a_4=a_3}^{S} a_3\varphi^{(a_1,a_2,a_3,a_4)} + \sum_{a_1=0}^{\infty}\sum_{a_2=c+1}^{N}\sum_{a_3=1}^{c}\sum_{a_4=a_3}^{S} a_3\varphi^{(a_1,a_2,a_3,a_4)}$$

7. **Mean Number of Junior Servers Enter into C-mode:**

$$Z_7 = q\mu\left[\sum_{a_1=0}^{\infty}\sum_{a_2=1}^{c}\sum_{a_3=0}^{a_2-1}\sum_{a_4=a_3+1}^{a_2}(a_4 - a_3)\varphi^{(a_1,a_2,a_3,a_4)} + \right.$$

$$\sum_{a_1=0}^{\infty}\sum_{a_2=1}^{c}\sum_{a_3=0}^{a_2-1}\sum_{a_4=a_2+1}^{S}(a_2-a_3)\varphi^{(a_1,a_2,a_3,a_4)}+\sum_{a_1=0}^{\infty}\sum_{a_2=c+1}^{N}\sum_{a_3=0}^{c-1}\sum_{a_4=a_3+1}^{c}(a_4-a_3)\varphi^{(a_1,a_2,a_3,a_4)}+$$
$$\sum_{a_1=0}^{\infty}\sum_{a_2=c+1}^{N}\sum_{a_3=0}^{c-1}\sum_{a_4=c+1}^{S}(c-a_3)\varphi^{(a_1,a_2,a_3,a_4)}\Bigg]$$

8. **Mean Waiting Time of Junior Servers in C-mode:**
$$Z_8=\frac{Z_6}{Z_7}$$
9. **Mean Number of Busy Junior Servers in S-mode:**
$$Z_9=\sum_{a_1=0}^{\infty}\sum_{a_2=1}^{c}\sum_{a_3=0}^{a_2-1}\sum_{a_4=a_3+1}^{a_2}(a_4-a_3)\varphi^{(a_1,a_2,a_3,a_4)}+\sum_{a_1=0}^{\infty}\sum_{a_2=1}^{c}\sum_{a_3=0}^{a_2-1}\sum_{a_4=a_2+1}^{S}(a_2-a_3)\varphi^{(a_1,a_2,a_3,a_4)}$$
$$+\sum_{a_1=0}^{\infty}\sum_{a_2=c+1}^{N}\sum_{a_3=0}^{c-1}\sum_{a_4=a_3+1}^{c}(a_4-a_3)\varphi^{(a_1,a_2,a_3,a_4)}+\sum_{a_1=0}^{\infty}\sum_{a_2=c+1}^{N}\sum_{a_3=0}^{c-1}\sum_{a_4=c+1}^{S}(c-a_3)\varphi^{(a_1,a_2,a_3,a_4)}$$
10. **Mean Number of Idle Junior Servers in S-mode:**
$$Z_{10}=c-(Z_6+Z_9)$$
11. **Mean Number of Customers in the Orbit:**
$$Z_{11}=\sum_{a_1=1}^{M-1}\varphi^{(a_1)}\mathbf{e}+(M\varphi^{(M)}(I-\mathbb{R})^{-1}+\varphi^{(M)}\mathbb{R}(I-\mathbb{R})^{-2})\mathbf{e}$$
12. **Mean Number of Customers Enter into Orbit:**
$$Z_{12}=\sum_{a_1=0}^{\infty}\sum_{a_3=0}^{a_2}\sum_{a_4=a_3}^{S}\lambda\varphi^{(a_1,N,a_3,a_4)}$$
13. **Mean Waiting Time of Customers in Orbit:**
$$Z_{13}=\frac{Z_{11}}{Z_{12}}$$
14. **Successful Rate of Retrial:**
$$Z_{14}=\sum_{a_1=1}^{M-1}\sum_{a_2=0}^{c}\sum_{a_3=0}^{a_2}\sum_{a_4=a_3}^{S}a_1\theta\varphi^{(a_1,a_2,a_3,a_4)}+\sum_{a_1=1}^{M-1}\sum_{a_2=c+1}^{N-1}\sum_{a_3=0}^{c}\sum_{a_4=a_3}^{S}a_1\theta\varphi^{(a_1,a_2,a_3,a_4)}$$
$$+(M\theta\varphi^{(M)}(I-\mathbb{R})^{-1}+\varphi^{(M)}\theta\mathbb{R}(I-\mathbb{R})^{-2})\mathbf{e}$$
15. **Overall Rate of Retrial:**
$$Z_{15}=\theta Z_{11}$$
16. **Fraction of Successful Rate of Retrial:**
$$Z_{16}=\frac{Z_{14}}{Z_{15}}$$

The Expected Total Cost

The expected total expenditure of the considered queuing-inventory system is defined as

$$E_{TC}=c_1Z_1+c_2Z_2+c_3Z_3+c_4Z_6+c_5Z_{11}+c_6Z_9+c_7Z_{10} \tag{32}$$

where

c_1 refers to the expense associated with holding each unit of an item.
c_2 signifies the cost incurred for setting up each unit of an item.
c_3 represents the cost accrued for each customer in the waiting hall per unit of time.
c_4 denotes the cost incurred for each junior server in C-mode per unit of time.
c_5 pertains to the cost accrued for each customer in orbit per unit of time.
c_6 represents the expenditure linked to each engaged junior server.
c_7 signifies the cost associated with each idle junior server.

6. Cost Analysis and Numerical Illustration

The study will employ the system's cost and parameter values to analyze the four-dimensional stochastic multi-server QIS under examination. The investigation will involve the study of the total cost, active servers, proportion of the successful retrial rate, and waiting times for customers in the orbit, waiting area, and servers in both S-mode and C-mode, by varying the parameters. Based on the results obtained from the stability

conditions and normalizing property, the following parameters and costs of the Markov process are assumed: $S = 28; s = 10; c_1 = 0.008; c_2 = 1; c_3 = 2; c_4 = 2; c_5 = 0.1; c_6 = 2; c_7 = 1; Q = S - s; N = 6; c = 3; M = 100; p = 0.5; q = 1 - p; \theta = 0.9; \lambda = 1; \mu = 3; \alpha = 8;$ and $\beta = 2$; for the analysis of the numerical discussions.

6.1. Analysis on the Expected Total Cost

In this section, the expected total cost is analyzed under the parameters. The obtained convex in Figure 2 ensures the model's efficiency. The impacts of each parameter on E_{TC} are given in Tables 1–3. The characteristics of the parameters are listed below:

- As the arrival rate λ rises, there is a notable upswing in E_{TC} due to the increased presence of customers inside the waiting hall.
- Elevating the probability value p hinders junior servers from approaching their senior counterparts. As a result, customers inside the waiting hall tend to depart promptly at the end of service in S-mode, resulting in a decrease in E_{TC}.
- With an increase in the retry rate θ, the anticipated orbit level diminishes as an increasing number of customers enter the waiting area. Since the expected orbit level correlates with the total cost, this leads to a reduction in E_{TC}.
- Higher service rates μ and α lead to a drop in each customer's mean service duration. Consequently, Z_3 decreases, contributing to a decrease in E_{TC}.
- Raising the rate β results in a decrease in the average lead time per order. Consequently, the expected total cost decreases.
- Altering the number of junior servers in the system incurs additional expenses. Hence, an increase in c leads to an increase in E_{TC}.
- An expansion in the waiting hall size corresponds to a rise in Z_3. As a result, the overall cost exhibits a positive correlation with the parameter N.

Expected total cost is a pivotal measure utilized to assess and enhance the effectiveness of a system. The efficiency of the model is significantly influenced by factors like the service rate, reorder rate, and the quantity of servers.

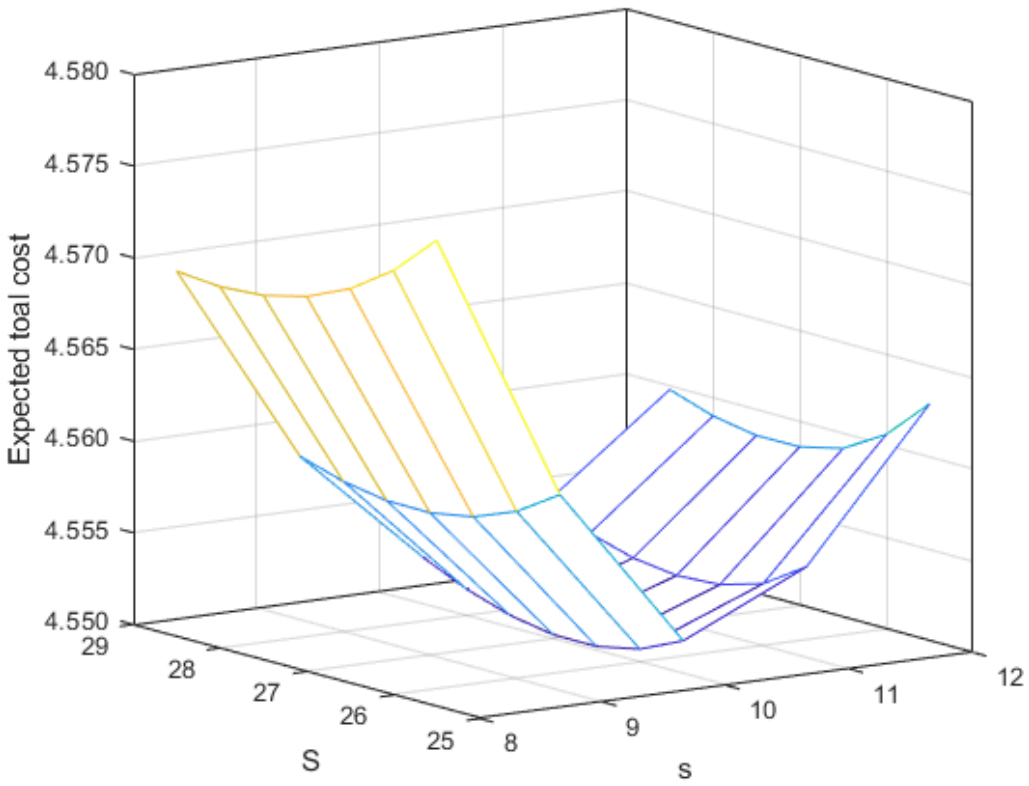

Figure 2. Expected total cost for S vs. s.

Table 1. E_{TC} vs. μ, α, and p.

μ	α	p	0.00	0.25	0.50	0.75	1.00
3	7		6.031736	5.065806	4.594173	4.313232	4.126693
	8		5.939920	5.006762	4.550207	4.278155	4.097510
	9		5.870337	4.961519	4.516374	4.251100	4.074969
4	7		5.488890	4.714822	4.332903	4.105023	3.953676
	8		5.396855	4.655079	4.288480	4.069652	3.924297
	9		5.326736	4.609159	4.254219	4.042323	3.901571
5	7		5.171290	4.506030	4.176982	3.980651	3.850284
	8		5.078740	4.445875	4.132323	3.945139	3.820815
	9		5.008045	4.399572	4.097844	3.917677	3.798003

Table 2. E_{TC} vs. λ, θ, β, and p.

λ	θ	β	p	0.00	0.25	0.50	0.75	1.00
1.0	0.1	2		5.941777	5.006924	4.550238	4.278163	4.097513
		3		5.935194	5.002773	4.547378	4.276104	4.095999
		4		5.931683	5.000643	4.545962	4.275120	4.095300
	0.5	2		5.940224	5.006786	4.550212	4.278156	4.097510
		3		5.933647	5.002635	4.547351	4.276097	4.095996
		4		5.930138	5.000505	4.545936	4.275112	4.095297
	0.9	2		5.939920	5.006762	4.550207	4.278155	4.097510
		3		5.933346	5.002612	4.547347	4.276096	4.095995
		4		5.929838	5.000481	4.545932	4.275111	4.095297
1.5	0.1	2		7.539386	5.970785	5.263147	4.848807	4.575142
		3		7.524784	5.960969	5.255680	4.842871	4.570300
		4		7.516844	5.955703	5.251778	4.839841	4.567878
	0.5	2		7.514470	5.968412	5.262689	4.848674	4.575092
		3		7.499993	5.958607	5.255224	4.842738	4.570250
		4		7.492103	5.953344	5.251322	4.839708	4.567828
	0.9	2		7.509099	5.967982	5.262613	4.848653	4.575084
		3		7.494679	5.958184	5.255149	4.842718	4.570243
		4		7.486809	5.952924	5.251248	4.839688	4.567821
2.0	0.1	2		9.652635	7.022696	6.000656	5.429878	5.058704
		3		9.623882	7.005369	5.987205	5.418726	5.049210
		4		9.609102	6.996062	5.980050	5.412893	5.044321
	0.5	2		9.563897	7.005378	5.997239	5.428880	5.058325
		3		9.534545	6.988206	5.983825	5.417744	5.048842
		4		9.519482	6.978943	5.976676	5.411911	5.043953
	0.9	2		9.532591	7.002251	5.996667	5.428720	5.058266
		3		9.503364	6.985142	5.983269	5.417590	5.048786
		4		9.488287	6.975898	5.976124	5.411758	5.043898

Table 3. E_{TC} vs. c, N, and p.

c	N	p	0.00	0.25	0.50	0.75	1.00
3	6		5.939920	5.006762	4.550207	4.278155	4.097510
	7		5.943270	5.007606	4.550499	4.278280	4.097572
	8		5.945190	5.008096	4.550672	4.278356	4.097610
4	6		7.328508	6.329089	5.816626	5.504202	5.293634
	7		7.328949	6.329197	5.816658	5.504214	5.293639
	8		7.329114	6.329223	5.816665	5.504216	5.293640
5	6		8.678503	7.621454	7.063953	6.717419	6.480568
	7		8.678304	7.621446	7.063953	6.717419	6.480568
	8		8.678374	7.621456	7.063955	6.717420	6.480568

6.2. Analysis on Mean Waiting Time of Customers in the Waiting Hall

This section analyzes the customer's waiting time in the waiting hall under the parameter variation shown in Tables 4–6. Each parameter is varied under the probability p.

- As the waiting hall size N is enlarged, there is an observed rise in waiting time attributable to an increase in Z_4.
- Augmenting the service rates for both junior and senior servers reduces their average service period. Hence, customer's waiting time decreases.
- An escalation in the arrival rate λ corresponds to an uptick in primary arrivals, leading to an increase in Z_3. Likewise, as the retry rate θ rises, orbiting customers transition into the waiting hall, influencing Z_5.
- It is a well-established fact that an augmentation in the number of servers generally leads to a decrease in waiting time. Accordingly, the parameter c is reflected in Z_5.
- Increasing the rate β results in a reduction in the average lead time per order and Z_5 diminishes.

By thoroughly analyzing waiting times in a waiting hall, businesses and organizations can implement strategies to enhance customer satisfaction, optimize operations, and improve overall service quality.

Table 4. Z_5 vs. λ, θ, and p.

λ	θ	p 0.00	0.25	0.50	0.75	1.00
1	0.7	0.818034	0.676300	0.576869	0.503096	0.446129
	0.9	0.818077	0.676309	0.576872	0.503096	0.446130
	1.1	0.818122	0.676322	0.576876	0.503098	0.446130
1.5	0.7	0.853010	0.693696	0.587304	0.510087	0.451152
	0.9	0.853583	0.693816	0.587331	0.510093	0.451153
	1.1	0.854016	0.693942	0.587372	0.510108	0.451159
2	0.7	0.944325	0.730811	0.606190	0.521327	0.458600
	0.9	0.949729	0.731874	0.606448	0.521392	0.458614
	1.1	0.952303	0.732582	0.606687	0.521479	0.458648

Table 5. Z_5 vs. c, β, and p.

c	β	p 0.00	0.25	0.50	0.75	1.00
2	1	0.894707	0.718120	0.603122	0.521038	0.459102
	2	0.894216	0.717907	0.602988	0.520932	0.459009
	3	0.894138	0.717889	0.602984	0.520932	0.459010
3	1	0.818223	0.676429	0.576976	0.503191	0.446218
	2	0.818040	0.676300	0.576870	0.503096	0.446129
	3	0.818026	0.676296	0.576868	0.503095	0.446129
4	1	0.810804	0.672472	0.574604	0.501661	0.445179
	2	0.810654	0.672354	0.574502	0.501568	0.445092
	3	0.810646	0.672351	0.574501	0.501567	0.445092

Table 6. Z_5 vs. N, μ, α, and p.

N	μ	α	p 0.00	0.25	0.50	0.75	1.00
5	2	7	1.203648	0.779707	0.579805	0.461953	0.384052
		8	1.178165	0.765030	0.569407	0.453902	0.377485
		9	1.159069	0.753881	0.561460	0.447727	0.372437
	3	7	0.841619	0.551896	0.411114	0.327681	0.272442
		8	0.817536	0.537408	0.400787	0.319667	0.265897
		9	0.799442	0.526400	0.392894	0.313520	0.260864
	4	7	0.669833	0.439553	0.327314	0.260796	0.216771
		8	0.645964	0.425118	0.317012	0.252796	0.210233
		9	0.628029	0.414150	0.309138	0.246659	0.205206

Table 6. *Cont.*

N	μ	α	p 0.00	0.25	0.50	0.75	1.00
6	2	7	1.208293	0.780137	0.579873	0.461968	0.384057
		8	1.182238	0.765400	0.569464	0.453914	0.377489
		9	1.162759	0.754211	0.561510	0.447738	0.372440
	3	7	0.842290	0.551948	0.411122	0.327683	0.272442
		8	0.818072	0.537448	0.400793	0.319668	0.265897
		9	0.799896	0.526432	0.392899	0.313521	0.260864
	4	7	0.670011	0.439567	0.327317	0.260797	0.216771
		8	0.646091	0.425128	0.317013	0.252796	0.210233
		9	0.628128	0.414157	0.309139	0.246659	0.205207
7	2	7	1.223024	0.782166	0.580291	0.462086	0.384098
		8	1.195385	0.767176	0.569825	0.454015	0.377524
		9	1.174808	0.755817	0.561834	0.447827	0.372471
	3	7	0.845365	0.552284	0.411189	0.327701	0.272449
		8	0.820602	0.537717	0.400845	0.319683	0.265902
		9	0.802076	0.526660	0.392943	0.313533	0.260869
	4	7	0.671008	0.439674	0.327338	0.260803	0.216773
		8	0.646849	0.425206	0.317029	0.252800	0.210235
		9	0.628743	0.414219	0.309151	0.246662	0.205208

6.3. Analysis of Mean Waiting Time for Customer in the Orbit

The mean waiting time for the customers in the orbit is analyzed in this section for each parameter under the variation in P shown in Tables 7–9. The impact of the parameters is listed below:

- By increasing N, the waiting hall offers more space, allowing orbital customers to transition into the waiting hall. As a result, their waiting time is reduced.
- Elevating the service rates for both junior and senior servers leads to a reduction in their average service time. Consequently, customers in the waiting hall tend to leave promptly, creating room for orbital customers to enter. This leads to a reduction in the waiting time for orbital customers.
- An increase in λ signifies a rise in primary arrivals, causing Z_3 to also increase. This, in turn, results in an extended Z_{13}.
- Similarly, as the rate θ rises, orbital customers move into the waiting hall, decreasing Z_{13}.
- It is a well-established fact that augmenting the number of servers generally leads to a decrease in waiting time. Therefore, the parameter c contributes to optimizing Z_{13}.
- The rate of parameter β is being increased. It is associated with a reduction in the lead time per order that minimizes Z_{13}.

The examination provides valuable insights into minimizing the waiting time for customers in the orbit. Ultimately, these parameters impact the efficiency of waiting times for customers in the waiting hall and also affect the orbit. A decrease in the waiting time for orbital customers leads to an increase in the business's profitability.

Table 7. Z_{13} vs. N, μ, α, and p.

N	μ	α	p 0.00	0.25	0.50	0.75	1.00
5	2	7	17.974591	16.126918	15.134281	14.557830	14.196432
		8	17.385102	15.680368	14.769172	14.246196	13.923038
		9	16.951712	15.348465	14.494552	14.008187	13.709900
	3	7	15.496179	14.688288	14.214487	13.819129	13.354706
		8	15.039793	14.342994	13.940541	13.577170	13.084691
		9	14.697968	14.073332	13.711969	13.357540	12.824969
	4	7	14.823263	14.195999	13.603035	12.876039	11.961140
		8	14.464449	13.938156	13.340046	12.460582	11.545646
		9	14.175495	13.699796	13.059647	12.004922	11.494981

Table 7. *Cont.*

N	μ	α	p 0.00	0.25	0.50	0.75	1.00
6	2	7	15.398444	13.833228	12.987423	12.502860	12.217659
		8	14.886138	13.443967	12.670814	12.235936	11.988846
		9	14.509286	13.154461	12.432376	12.031705	11.810049
	3	7	13.290115	12.611500	12.260900	12.051000	11.885700
		8	12.899227	12.324800	12.051900	11.904900	11.787000
		9	12.606572	12.101900	11.879100	11.771100	11.680400
	4	7	12.721641	12.279487	11.983881	11.707888	11.390321
		8	12.432918	12.114537	11.906196	11.672482	11.338833
		9	12.205155	11.968338	11.815081	11.598996	11.229460
7	2	7	12.982237	11.655100	10.927200	10.508000	10.264500
		8	12.537662	11.316500	10.653400	10.279400	10.070800
		9	12.210647	11.064900	10.447500	10.104900	9.920400
	3	7	11.171236	10.580468	10.287446	10.142443	10.073247
		8	10.836051	10.338327	10.114754	10.028237	10.010985
		9	10.586480	10.152566	9.976379	9.929820	9.918257
	4	7	10.654105	10.297545	10.115627	10.042687	10.05909
		8	10.410305	10.163202	10.069510	10.034289	10.00849
		9	10.224213	10.053975	9.024084	9.0077514	8.925115

Table 8. Z_{13} vs. λ, θ, and p

λ	θ	p 0.00	0.25	0.50	0.75	1.00
1	0.7	6.168542	4.411438	3.410960	2.348533	1.446406
	0.9	5.039675	3.494077	2.670702	1.839568	1.137603
	1.1	4.338611	2.932733	2.217329	1.524989	0.944842
1.5	0.7	9.004800	7.463500	6.438000	4.978400	3.376200
	0.9	7.149300	5.759400	4.923600	3.820000	2.612900
	1.1	6.016700	4.733400	4.010400	3.113900	2.141300
2	0.7	16.816769	15.988781	15.693862	14.625320	12.438549
	0.9	12.899227	11.969791	11.634857	10.840719	9.288518
	1.1	10.564923	9.601746	9.246775	8.601285	7.401879

Table 9. Z_{13} vs. c, β, and p

c	β	p 0.00	0.25	0.50	0.75	1.00
2	1	35.941785	29.674531	28.507375	28.603095	29.006031
	2	16.676209	13.582353	12.923631	12.734906	12.420229
	3	4.417784	2.637320	1.175358	0.480075	0.234289
3	1	28.543326	27.075633	27.309766	27.759411	28.048548
	2	12.899227	11.969791	11.634857	10.840719	9.288518
	3	2.656402	0.945474	0.300075	0.131414	0.080402
4	1	26.155253	25.882732	26.404486	26.976858	27.352461
	2	11.639618	11.124192	10.458458	8.803575	6.357663
	3	1.770459	0.466486	0.145741	0.071584	0.049899

6.4. Analysis on Mean Waiting Time of Junior Servers in C-Mode

This section analyzes a parameter analysis on the mean waiting time of junior servers in C-mode. The impact of parameters can be seen in Tables 10–12 and is listed below

- As the rate of customer arrival increases, there is a corresponding growth in the quantity of customers present in the waiting area. As a result, there is an increased probability of junior servers operating in C-mode, hence resulting in a corresponding rise in their waiting duration.
- Similarly, an increase in the retrial rate θ allows orbital customers to transition into the waiting hall, resulting in an increase in Z_8

- Increasing the value of c leads to a rise in the average number of busy junior servers. Additionally, the number of junior servers in C-mode also increases with an increase in c.
- An expansion in the capacity of the waiting area corresponds to a proportional increase in the number of busy servers in C-mode, as there is a rise in Z_8.
- The reduction in the number of servers in C-mode is seen when there is an increase in the service rate for the senior server. However, this relationship is reversed for junior servers.
- A reduction in the rate β always leads to a decrease in lead time, consequently resulting in a decrease in Z_8.
- Upon observing the effects of varying the probability p, there is a notable reduction in the waiting time of junior servers in C-mode.

The importance of the number of servers lies in its direct impact on the system's capacity to serve customers. A higher number of servers increases the system's capacity to process customers simultaneously, reducing waiting times and congestion. Conversely, a lower number of servers can lead to longer waiting times and potentially dissatisfied customers.

Table 10. Z_8 vs. N, μ, α, and p.

N	μ	α	p 0.00	0.25	0.50	0.75	1.00
5	2	7	0.214168	0.173785	0.153179	0.140672	0.132265
		8	0.205232	0.165357	0.144930	0.132515	0.124167
		9	0.197841	0.158327	0.138030	0.125685	0.117382
	3	7	0.311016	0.243614	0.208336	0.186755	0.172203
		8	0.300123	0.233510	0.198503	0.177056	0.162586
		9	0.291243	0.225200	0.190392	0.169047	0.154639
	4	7	0.548656	0.420817	0.348025	0.302173	0.270858
		8	0.534660	0.408228	0.335948	0.290339	0.259165
		9	0.523450	0.398053	0.326153	0.280728	0.249662
6	2	7	0.214574	0.173874	0.153211	0.140687	0.132273
		8	0.205650	0.165451	0.144964	0.132532	0.124176
		9	0.198264	0.158424	0.138066	0.125702	0.117391
	3	7	0.312723	0.243974	0.208462	0.186813	0.172234
		8	0.301814	0.233874	0.198631	0.177115	0.162617
		9	0.292917	0.225563	0.190521	0.169106	0.154671
	4	7	0.560953	0.423295	0.348857	0.302539	0.271048
		8	0.546575	0.410671	0.336772	0.290703	0.259353
		9	0.535072	0.400464	0.326969	0.281088	0.249849
7	2	7	0.215571	0.174074	0.153276	0.140714	0.132287
		8	0.206497	0.165628	0.145023	0.132557	0.124189
		9	0.199010	0.158583	0.138120	0.125726	0.117403
	3	7	0.315726	0.244621	0.208679	0.186906	0.172281
		8	0.304383	0.234445	0.198826	0.177200	0.162661
		9	0.295190	0.226081	0.190700	0.169185	0.154712
	4	7	0.571828	0.426441	0.349984	0.303037	0.271302
		8	0.555640	0.413467	0.337792	0.291157	0.259587
		9	0.542875	0.403013	0.327913	0.281511	0.250067

Table 11. Z_8 vs. λ, θ, and p.

λ	θ	p 0.00	0.25	0.50	0.75	1.00
1	0.7	0.298673	0.230565	0.195358	0.173921	0.159507
	0.9	0.301814	0.233874	0.198631	0.177115	0.162617
	1.1	0.302669	0.234794	0.199552	0.178019	0.163501
1.5	0.7	0.391362	0.295437	0.244605	0.213324	0.192213
	0.9	0.395540	0.300339	0.249639	0.218302	0.197071
	1.1	0.396814	0.301838	0.251218	0.219896	0.198652
2	0.7	0.484258	0.357994	0.293037	0.252562	0.225074
	0.9	0.487276	0.363284	0.298753	0.258276	0.230623
	1.1	0.488741	0.365234	0.300915	0.260522	0.232891

Table 12. Z_8 vs. c, β, and p.

c	β	p 0.00	0.25	0.50	0.75	1.00
2	1	0.200141	0.149298	0.123846	0.108641	0.098556
	2	0.200125	0.149298	0.123846	0.108641	0.098556
	3	0.200117	0.149297	0.123845	0.108640	0.098555
3	1	0.301820	0.233874	0.198631	0.177115	0.162617
	2	0.301810	0.233873	0.198630	0.177114	0.162616
	3	0.301818	0.233872	0.198629	0.177113	0.162615
4	1	0.315977	0.249250	0.212910	0.189968	0.174199
	2	0.315622	0.249228	0.212908	0.189967	0.174199
	3	0.315402	0.249212	0.212905	0.189966	0.174198

6.5. Analysis of Busy Junior Servers in S-Mode

This section analyzes the impact of parameters on the busy junior servers in server mode, which are displayed in Figures 3–9. The observation of the parameters are given below:

- As arrivals increase, Z_3 rises. This, in turn, raises the probability of junior servers being in S-mode.
- Similarly, when the retrial rate θ is raised, orbital customers have a higher chance of entering the waiting hall. Consequently, there is an uptick in the count of junior servers in S-mode.
- Increasing the value of c results in a higher average of occupied junior servers. Simultaneously, there is a rise in Z_9 with an increase in c.
- Expanding the capacity of the waiting hall leads to a corresponding increase in the number of occupied servers in S-mode due to the upswing in the average number of servers in S-mode.
- With an increase in the service rates for both senior and junior servers, Z_9 decreases due to the reduction in average service time per customer.
- An increase in the rate β consistently leads to a reduction in lead time. Consequently, there is a decrease that can be seen in Z_9.
- An examination of the effect of varying the probability p in combination with each parameter consistently reveals a decrease in the number of junior servers in S-mode.

In service mode, the number of servers can vary based on factors such as demand patterns, time of day, or operational decisions. Having the flexibility to adjust the number of servers allows for responsiveness to the changing levels of demand. Increasing the number of servers during high-demand periods can help manage waiting times and improve customer satisfaction. Conversely, reducing the number of servers during low-demand periods can help optimize resource utilization.

6.6. Analysis of Fraction of Successful Retrial Rate

The fraction of the successful retrial rate is analyzed in this section under the parameter variation, shown in Figures 10–16. The observations are listed below:

- There is a gradual increase in Z_{16} while varying the parameter β. However, after some point, the increase is not notably significant.
- An immediate response of increasing c can be observed in Z_{16} as it increases.
- The fraction of the successful retrial rate decreases when the arrival rate of the customer increases. After some values, it moves faster towards the value zero.
- An increase in the service rate directly decreases the service time per customer. It is suitable for the service rates of both junior and senior servers. As a result, they increase the rate of successful retrials.
- Increasing the capacity of the waiting hall increases the chance of a retrial's success. Hence Z_{16} increases when we increase N.
- As similar to lambda, while increasing the rate θ, the waiting hall soon becomes full, which decreases the chance of a successful retrial.

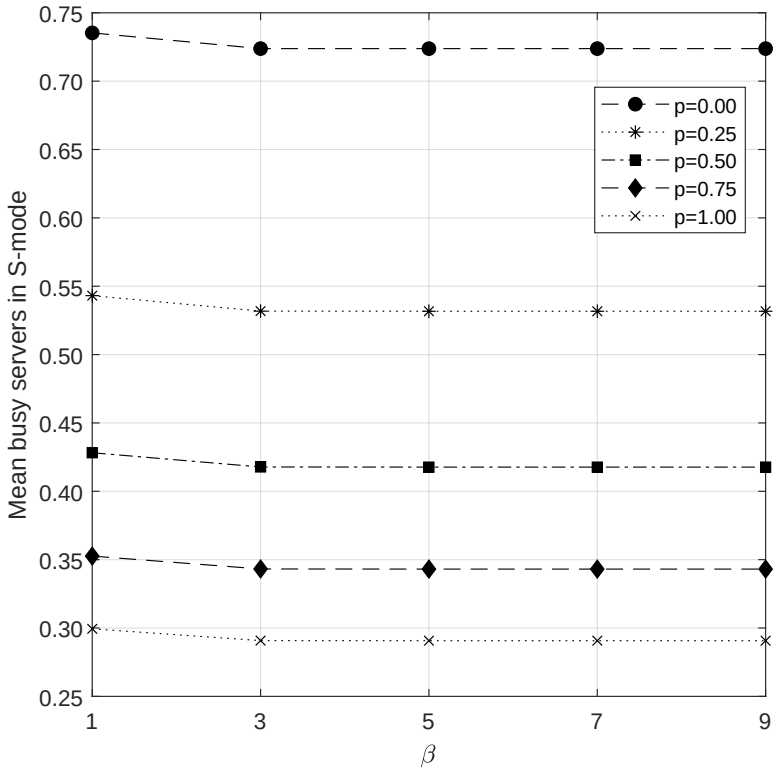

Figure 3. Mean busy servers in S-mode vs. p and β.

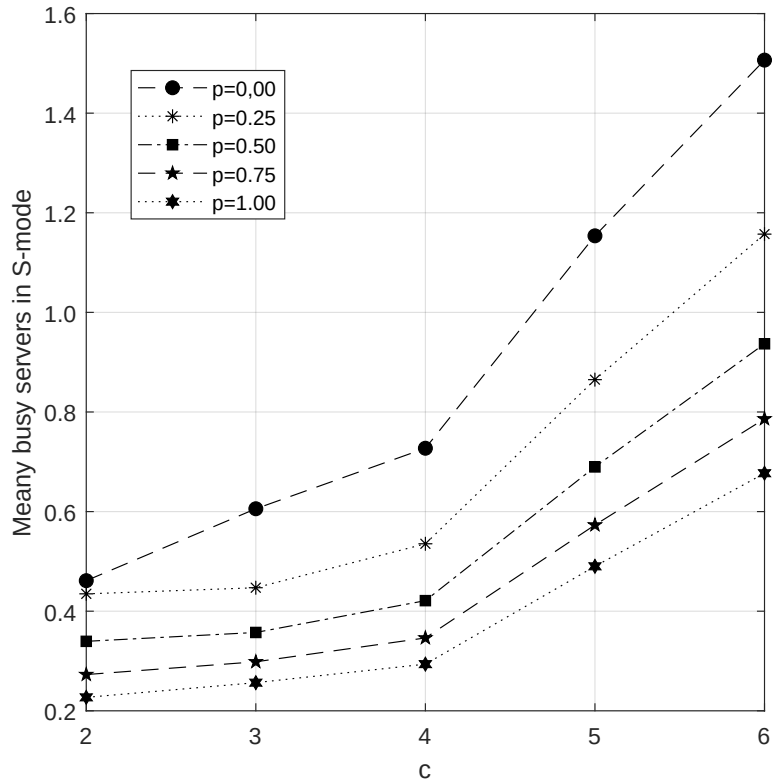

Figure 4. Mean busy servers in S-mode vs. p and c.

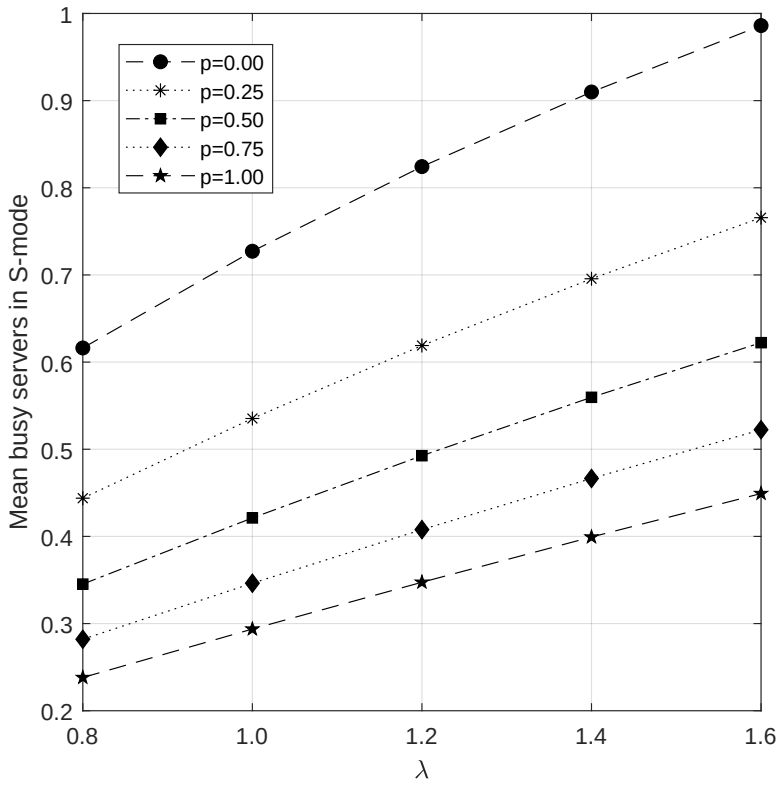

Figure 5. Mean busy servers in S-mode vs. p and λ.

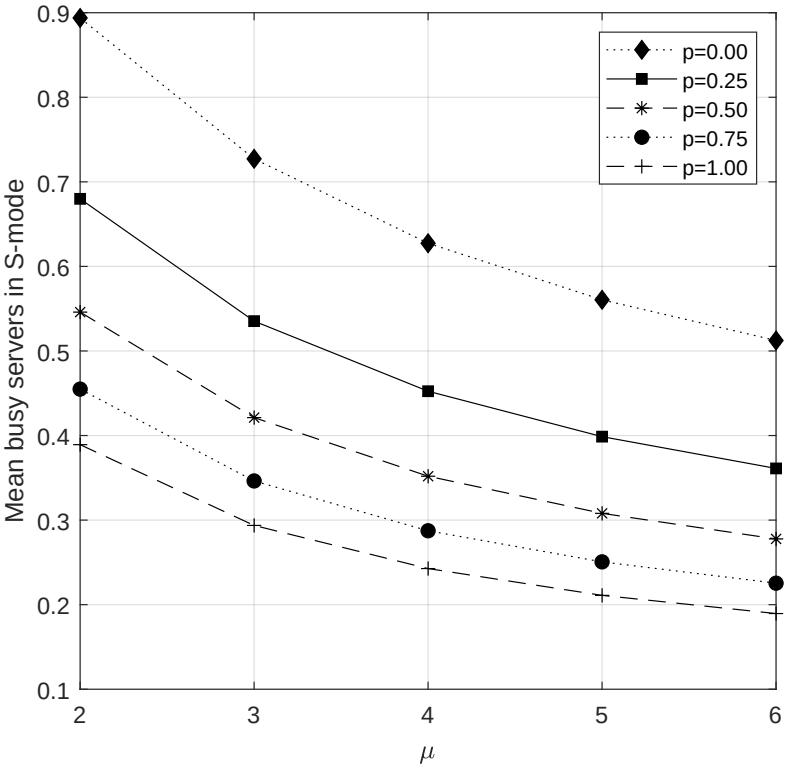

Figure 6. Mean busy servers in S-mode vs. p and μ.

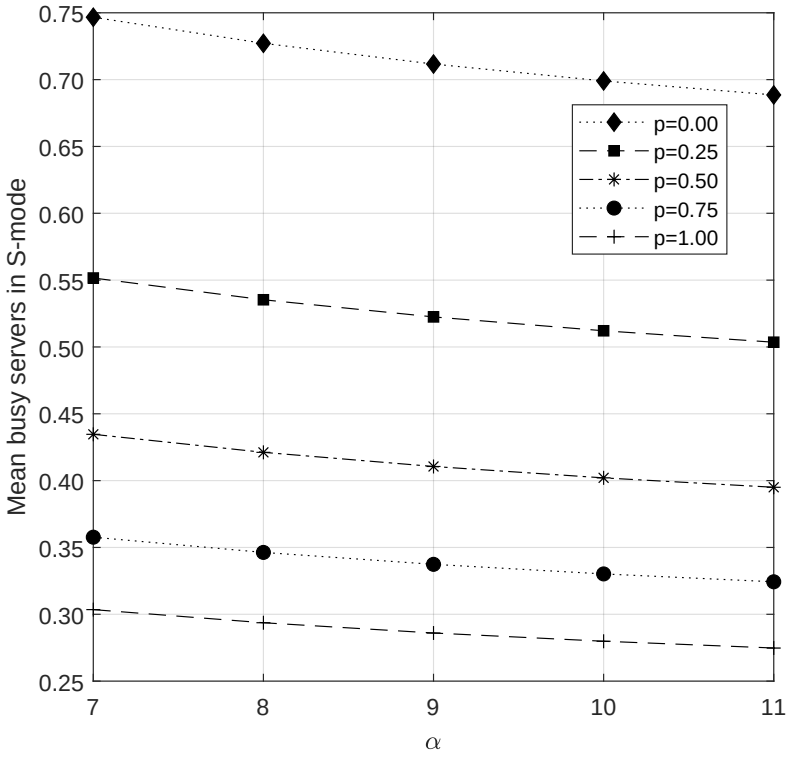

Figure 7. Mean busy servers in S-mode vs. p and α.

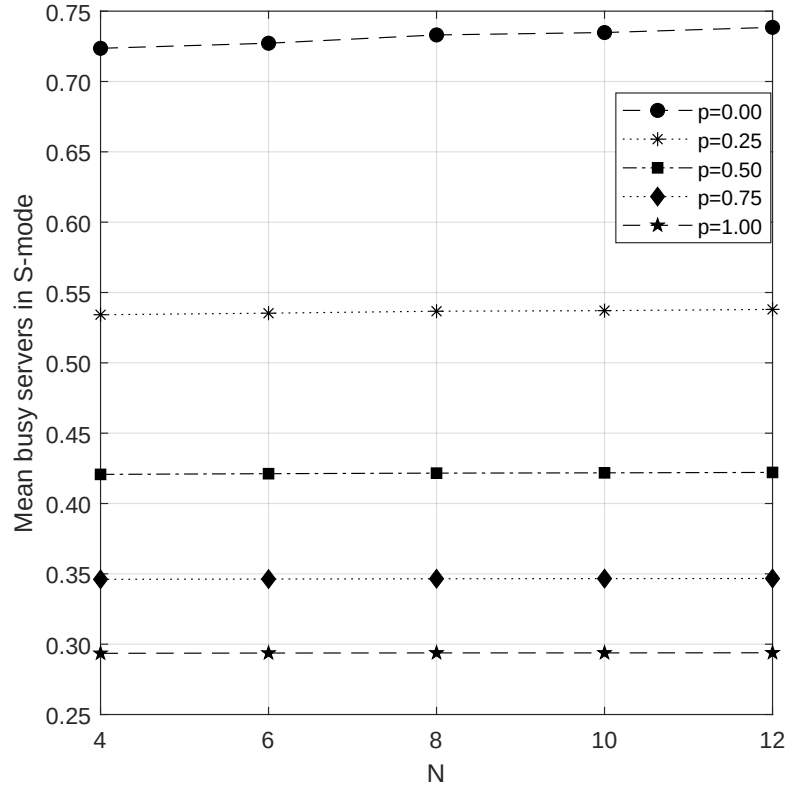

Figure 8. Mean busy servers in S-mode vs. p and N.

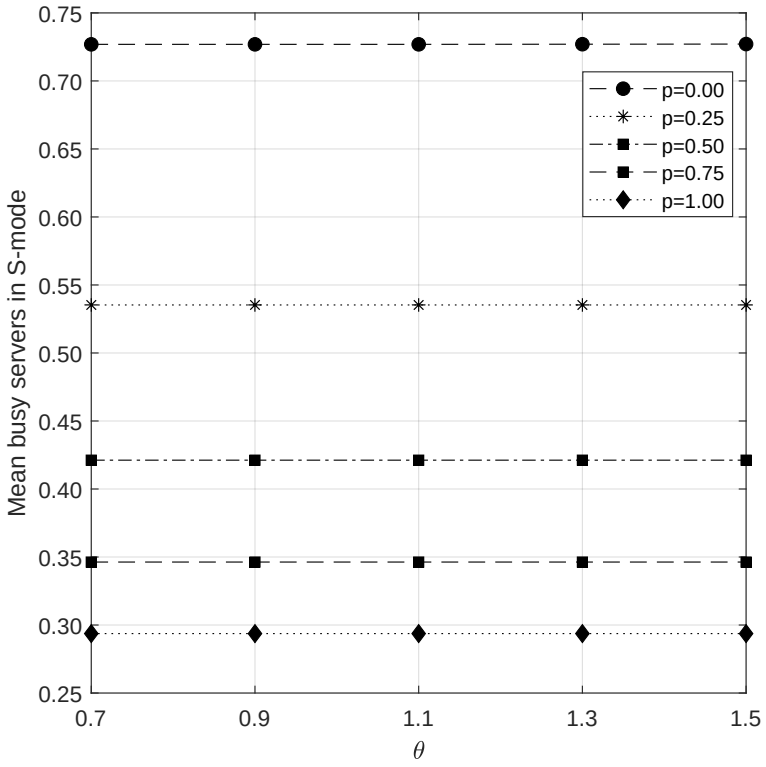

Figure 9. Mean busy servers in S-mode vs. p and θ.

Analyzing the fraction of the successful retrial rate provides valuable insights into the performance and effectiveness of a system in handling retrials. The service rates, number of servers, waiting hall size, and reorder rate directly influence it.

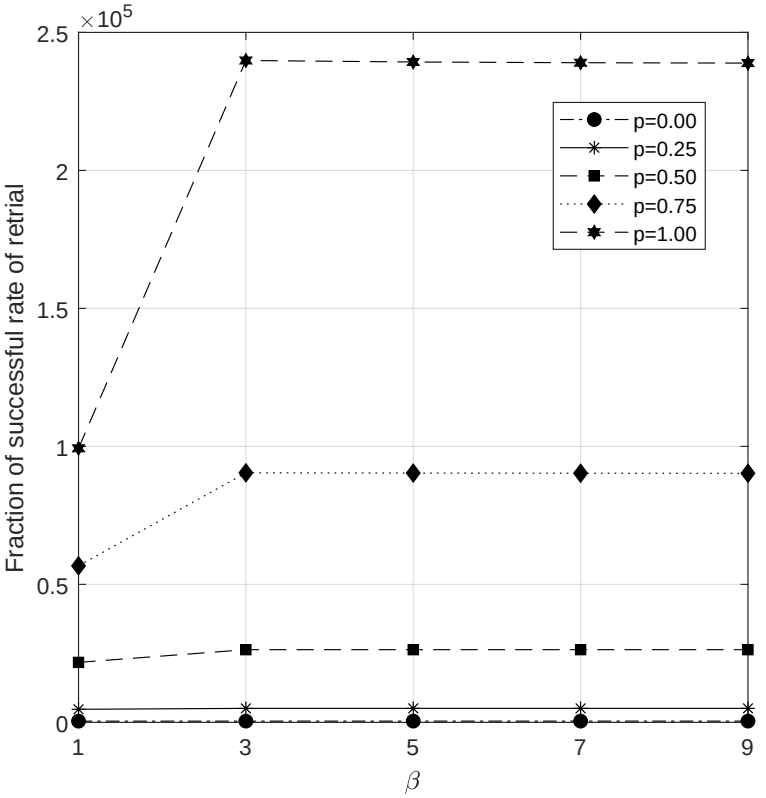

Figure 10. Fraction of successful rate of retrial vs. p and β.

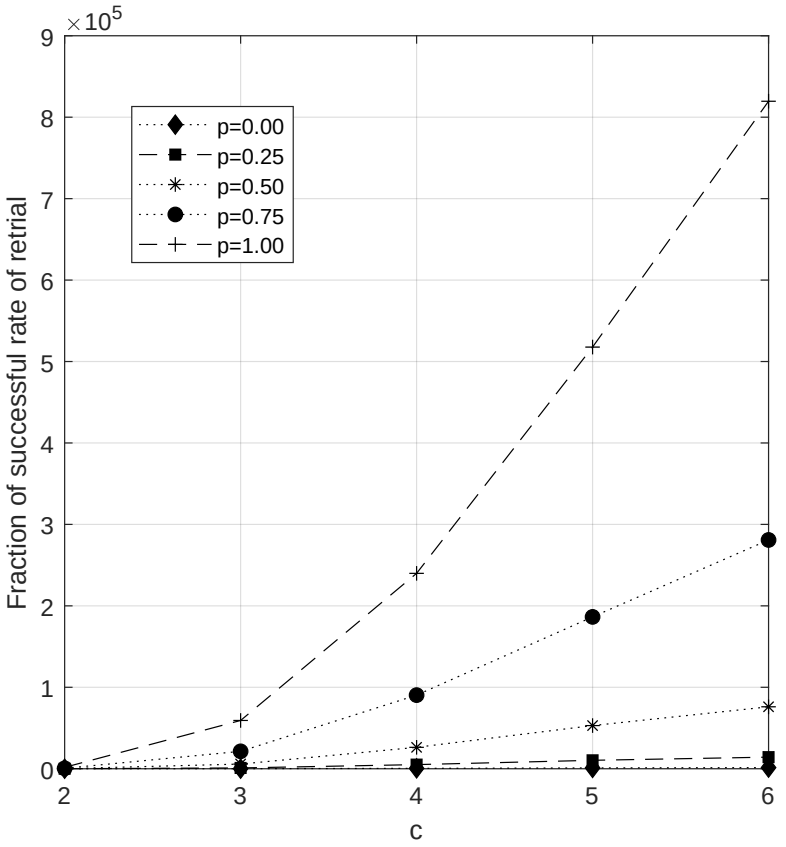

Figure 11. Fraction of successful rate of retrial vs. p and c.

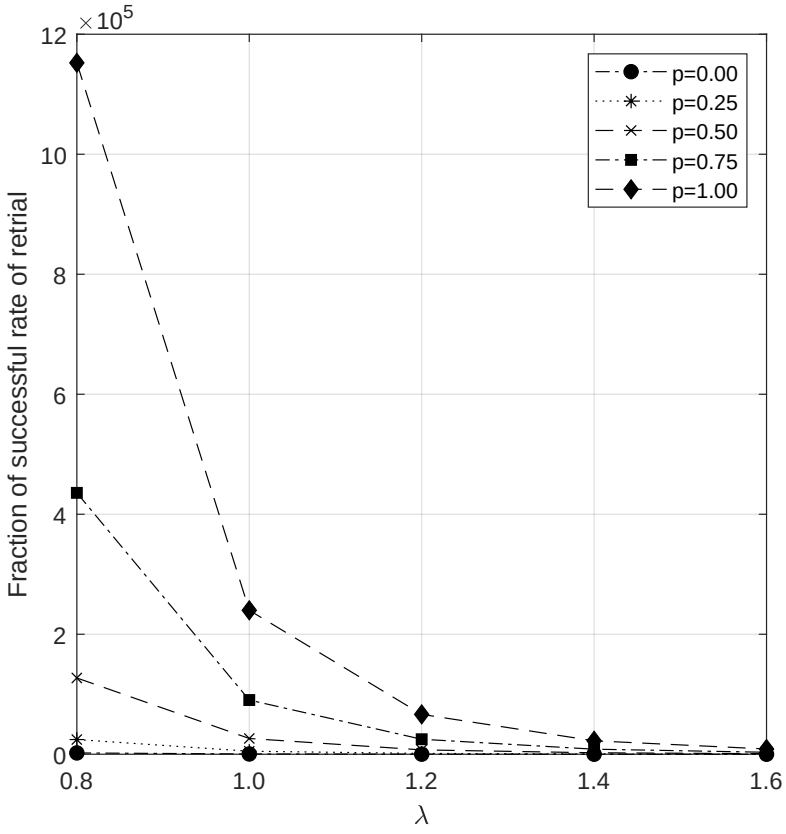

Figure 12. Fraction of successful rate of retrial vs. p and λ.

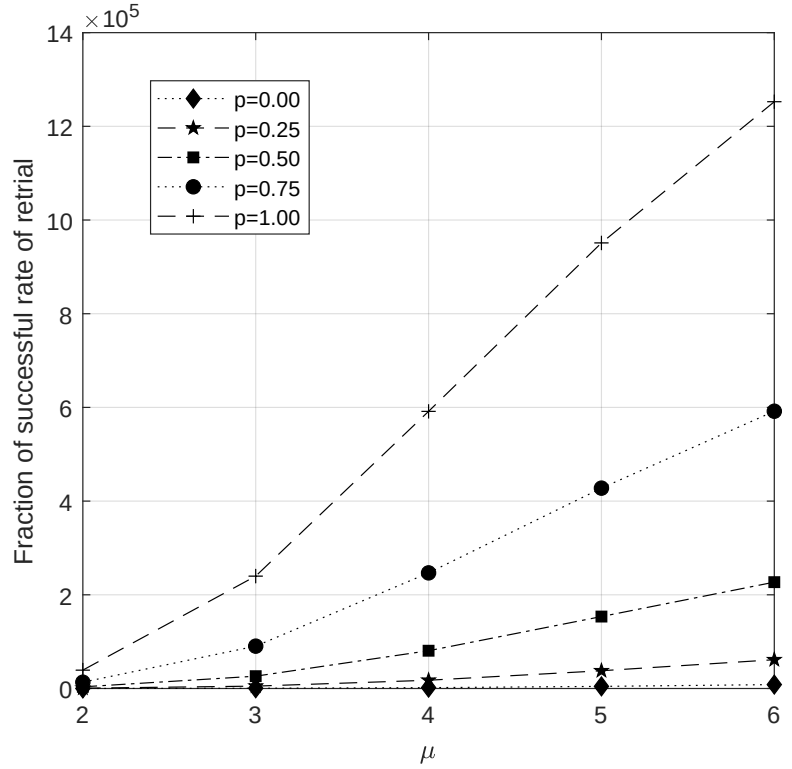

Figure 13. Fraction of successful rate of retrial vs. p and μ.

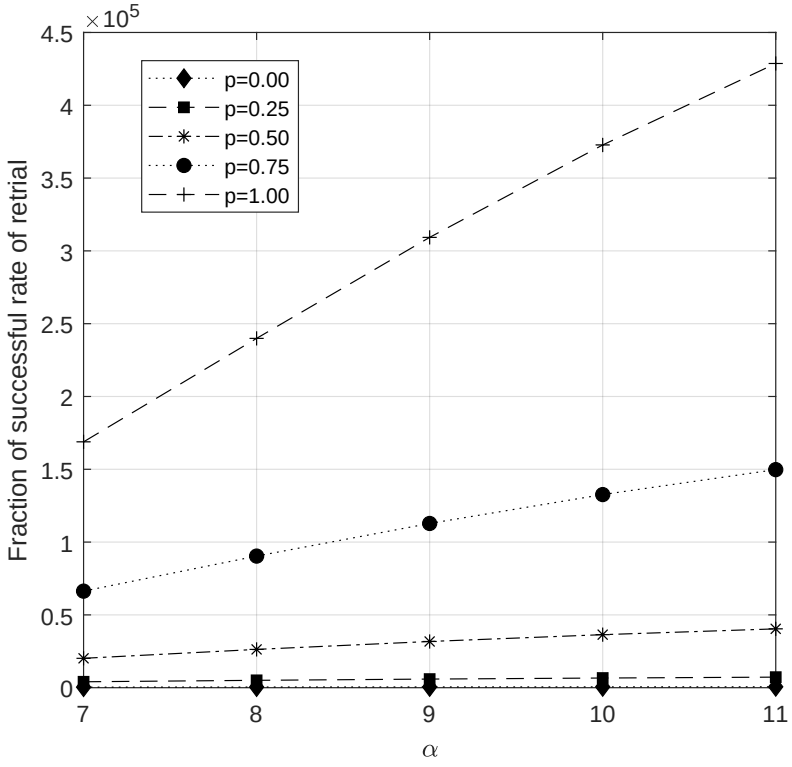

Figure 14. Fraction of successful rate of retrial vs. p and α.

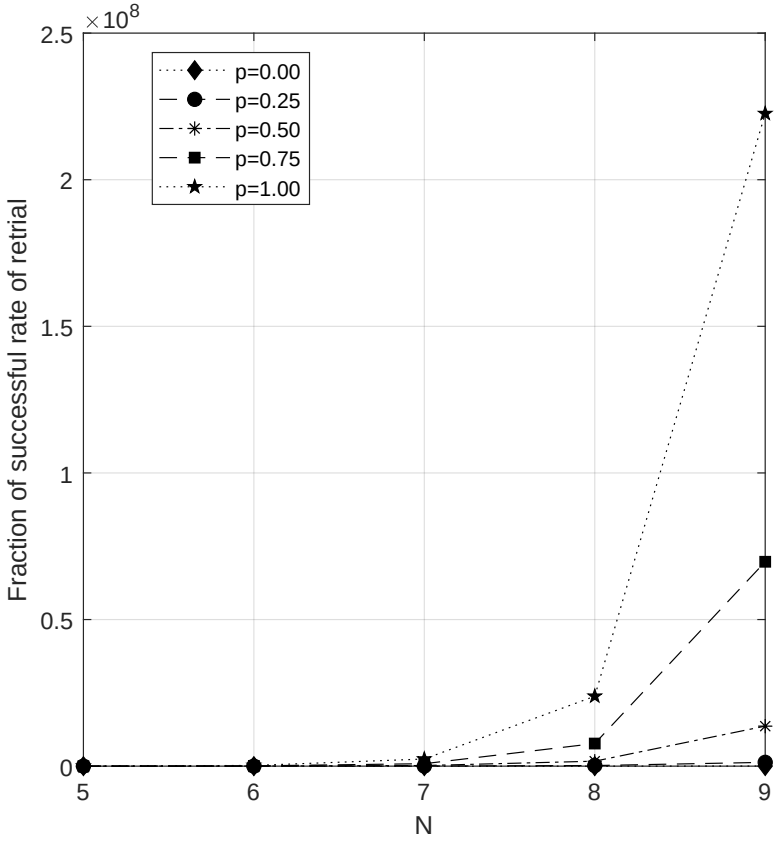

Figure 15. Fraction of successful rate of retrial vs. p and N.

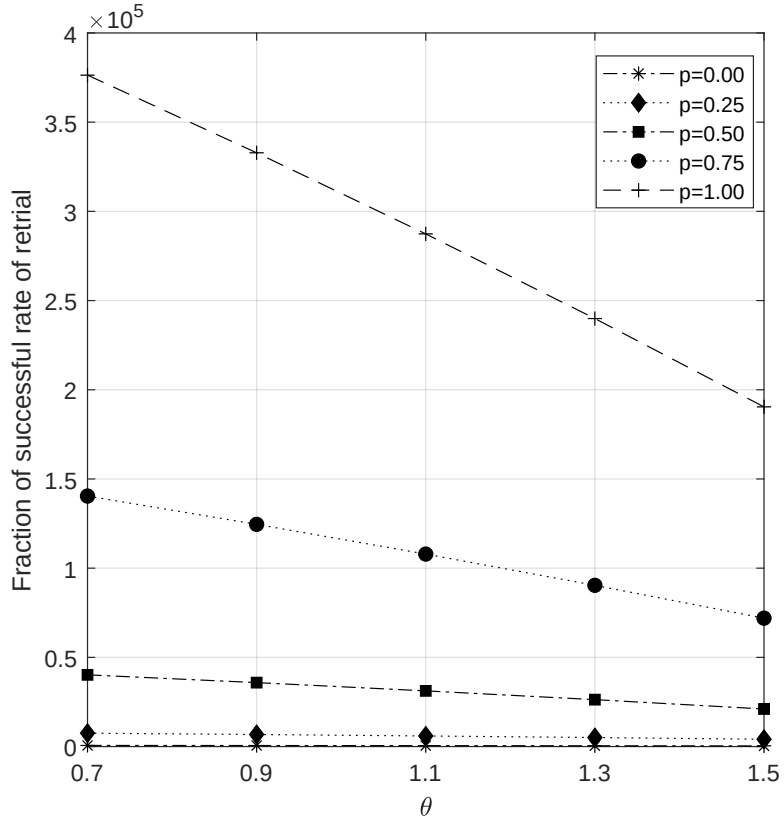

Figure 16. Fraction of successful rate of retrial vs. p and θ.

6.7. Observations

We present the observations that we saw in each numerical analysis.

- The obtained convexity of the total cost helps in identifying efficient configurations that minimize the overall costs.
- The parameters β, μ, and α expose a favorable output for each analysis.
- Though the parameters N and c increase the total cost, they are actually helpful in increasing the fraction of successful retrials.
- When $p = 0$, the model turns out to be the standard QIS without consultation.
- When $p = 1$, the model turns out to be the QIS with compulsory consultation as [40]. Also, with the inventory and retrial facility, the model looks like an extension version of [40].
- The rate θ decreases the waiting time of customers in the orbit when it increases. But, in the case of customers in the waiting hall, it acts inversely.
- The rate λ increases the total cost, waiting time of customers, and number of junior servers in C and S mode.
- An increase in p means that the junior servers are developed as experienced servers.

7. Conclusions

In this article, we have investigated the queuing-inventory model, where junior servers occasionally encounter challenges while serving customers. When faced with such situations, they seek advice from the senior server to resolve the issue. Initially, primary customers meet junior servers to set up service. If the junior server faces no issues while providing service, a customer leaves the system successfully with an item. Suppose that the junior server faces an issue. He enters C-mode for consultation with a senior server. Following this consultation, the senior and junior servers collaborate to finalize the service. When the senior server is engaged with other junior servers, the latter joins a line and waits for their turn. Essentially, in the second phase, junior servers assume the role of customers.

To evaluate this system, a Markovian queuing-inventory model is employed, necessitating the creation of a four-dimensional state space. By applying Neuts and Rao's truncation method, the study calculates the steady-state probability vector of the system and various performance metrics. The investigation also scrutinizes customer waiting times and the average count of active junior servers in both S-mode and C-mode. By varying the probabilities, we gain insight into the impact of the necessity for junior server consultations at the end of service, ranging from compulsory consultation to no consultation. As junior servers accumulate experience from the senior server over time, the need for consultations ultimately decreases. These analyses offer valuable insights into the model's performance and effectiveness. In the future, this model could be extended to incorporate more complex arrival processes (MAP) and phase-type distributions, replacing the current Poisson and exponential distributions.

Author Contributions: Conceptualization, K.J. and T.H.; Software, T.H.; Investigation, K.L., (Sections 1 and 2); A.M., and J.S., (Sections 3–6); Writing—original draft preparation, review, and editing, K.J., T.H., K.L., A.M. and J.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable

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

Notations

The following abbreviations are used in this manuscript:

$\mathbf{0}$	A matrix where all entries are zero
$\mathbf{e}$	A column vector of appropriate dimensions, with each coordinate set to one
I	An identity matrix
δ_{ij}	Kronecker Delta
$\bar{\delta}_{ij}$	$1 - \delta_{ij}$
$H(x)$	Heaviside function
$\overline{a,b}$	$a, a+1, \cdots, b$, where a and b are integers
$\overline{a,\infty}$	$a, a+1, \cdots$ where a is an integer

References

1. Melikov, A.Z.; Molchanov, A.A. Stock optimization in transportation/storage systems. *Cybern. Syst. Anal.* **1992**, *28*, 484–487. [CrossRef]
2. Sigman, K.; Simchi-Levi, D. Light traffic heuristic for an M/G/1 queue with limited inventory. *Ann. Oper. Res.* **1992**, *40*, 371–380. [CrossRef]
3. Yadavalli, V.S.; Sivakumar, B.; Arivarignan, G.; Adetunji, O. A multi-server perishable inventory system with negative customer. *Comput. Ind. Eng.* **2011**, *61*, 254–273. [CrossRef]
4. Yadavalli, V.S.; Sivakumar, B.; Arivarignan, G.; Adetunji, O. A finite source multi-server inventory system with service facility. *Comput. Ind. Eng.* **2012**, *63*, 739–753. [CrossRef]
5. Nair, A.N.; Jacob, M.J.; Krishnamoorthy, A. The multi server M/M/(s, S) queueing inventory system. *Ann. Oper. Res.* **2015**, *233*, 321–333. [CrossRef]
6. Krishnamoorthy, A.; an Manik, R.; Shajin, D. Analysis of a multiserver queueing-inventory system. *Adv. Oper. Res.* **2015**, *2015*, 747328. [CrossRef]
7. Wang, F.F. Approximation and optimization of a multi-server impatient retrial inventory-queueing system with two demand classes. *Qual. Technol. Quant. Manag.* **2015**, *12*, 269–292. [CrossRef]
8. Wang, F.F.; Bhagat, A.; Chang, T.M. Analysis of priority multi-server retrial queueing inventory systems with MAP arrivals and exponential services. *Opsearch* **2017**, *54*, 44–66. [CrossRef]
9. Hanukov, G.; Avinadav, T.; Chernonog, T.; Yechiali, U. A multi-server queueing-inventory system with stock-dependent demand. *IFAC PapersOnLine* **2019**, *52*, 671–676. [CrossRef]
10. Jeganathan, K.; Reiyas, M.A.; Lakshmi, K.P.; Saravanan, S. Two server Markovian inventory systems with server interruptions: Heterogeneous vs. homogeneous servers. *Math. Comput. Simul.* **2019**, *155*, 177–200. [CrossRef]
11. Suganya, C.; Sivakumar, B. MAP/PH (1), PH (2)/2 finite retrial inventory system with service facility, multiple vacations for servers. *Int. J. Math. Oper. Res.* **2019**, *15*, 265–295. [CrossRef]
12. Jose, K.P.; Beena, P. On a retrial production inventory system with vacation and multiple servers. *Int. J. Appl. Comput. Math.* **2020**, *6*, 108. [CrossRef]
13. Chakravarthy, S.R.; Rumyantsev, A. Analytical and simulation studies of queueing-inventory models with MAP demands in batches and positive phase-type services. *Simul. Model. Pract. Theory* **2020**, *103*, 102092. [CrossRef]
14. Jeganathan, K.; Reiyas, M.A. Two parallel heterogeneous servers Markovian inventory system with modified and delayed working vacations. *Math. Comput. Simul.* **2020**, *172*, 273–304. [CrossRef]
15. Chakravarthy, S.R.; Shajin, D.; Krishnamoorthy, A. Infinite server queueing-inventory models. *J. Indian Soc. Probab. Stat.* **2020**, *21*, 43–68. [CrossRef]
16. Hanukov, G.; Avinadav, T.; Chernonog, T.; Yechiali, U. A multi-server system with inventory of preliminary services and stock-dependent demand. *Int. J. Prod. Res.* **2021**, *59*, 4384–4402. [CrossRef]
17. Jeganathan, K.; Harikrishnan, T.; Selvakumar, S.; Anbazhagan, N.; Amutha, S.; Acharya, S.; Rajendra, D.; Joshi, G.P. Analysis of interconnected arrivals on queueing-inventory system with two multi-server service channels and one retrial facility. *Electronics* **2021**, *10*, 576. [CrossRef]
18. Rasmi, K.; Jacob, M.J.; Rumyantsev, A.S.; Krishnamoorthy, A. A multi-server heterogeneous queuing-inventory system with class-dependent inventory access. *Mathematics* **2021**, *9*, 1037. [CrossRef]
19. Rasmi, K.; Jacob, M.J. Analysis of a multiserver queueing inventory model with self-service. *Int. J. Math. Model. Numer. Optim.* **2021**, *11*, 275–291. [CrossRef]
20. Shajin, D.; Krishnamoorthy, A.; Melikov, A.Z.; Sztrik, J. Multi-server queuing production inventory system with emergency replenishment. *Mathematics* **2022**, *10*, 3839. [CrossRef]
21. Jeganathan, K.; Selvakumar, S.; Saravanan, S.; Anbazhagan, N.; Amutha, S.; Cho, W.; Joshi, G.P.; Ryoo, J. Performance of stochastic inventory system with a fresh item, returned item, refurbished item, and multi-class customers. *Mathematics* **2022**, *10*, 1137. [CrossRef]
22. Almaqbali, K.A.; Joshua, V.C.; Krishnamoorthy, A. Multi-class, multi-server queueing inventory system with batch service. *Mathematics* **2023**, *11*, 830. [CrossRef]

23. Aghsami, A.; Samimi, Y.; Aghaie, A. A combined continuous-time Markov chain and queueing-inventory model for a blood transfusion network considering ABO/Rh substitution priority and unreliable screening laboratory. *Expert Syst. Appl.* **2023**, *215*, 119360. [CrossRef]
24. Selvakumar, S.; Jeganathan, K.; Srinivasan, K.; Anbazhagan, N.; Lee, S.; Joshi, G.P.; Doo, I.C. An optimization of home delivery services in a stochastic modeling with self and compulsory vacation interruption. *Mathematics* **2023**, *11*, 2044. [CrossRef]
25. Yue, D.; Ye, Z.; Yue, W. Analysis of a Queueing-Inventory System with Synchronous Vacation of Multiple Servers. *Queueing Models Serv. Manag.* **2023**, *6*, 1–26.
26. Artalejo, J.R.; Krishnamoorthy, A.; Lopez-Herrero, M.J. Numerical analysis of (s, S) inventory systems with repeated attempts. *Ann. Oper. Res.* **2006**, *141*, 67–83. [CrossRef]
27. Ushakumari, P.V. On (s, S) inventory system with random lead time and repeated demands. *J. Appl. Math. Stoch. Anal.* **2006**, *2006*, 81508. [CrossRef]
28. Amirthakodi, M.; Sivakumar, B. An inventory system with service facility and feedback customers. *Int. J. Ind. Syst. Eng.* **2019**, *33*, 374–411. [CrossRef]
29. Lopez-Herrero, M.J. Waiting time and other first-passage time measures in an $(s,\ S)$ inventory system with repeated attempts and finite retrial group. *Comput. Oper. Res.* **2010**, *37*, 1256–1261. [CrossRef]
30. Hanukov, G. A queueing-inventory system in which customers can orbit during the service. *IFAC PapersOnLine* **2022**, *55*, 619–624. [CrossRef]
31. Sugapriya, C.; Nithya, M.; Jeganathan, K.; Anbazhagan, N.; Joshi, G.P.; Yang, E.; Seo, S. Analysis of stock-dependent arrival process in a retrial stochastic inventory system with server vacation. *Processes* **2022**, *10*, 176. [CrossRef]
32. Melikov, A.; Aliyeva, S.; Nair, S.S.; Kumar, B.K. Retrial queuing-inventory systems with delayed feedback and instantaneous damaging of items. *Axioms* **2022**, *11*, 241. [CrossRef]
33. Nithya, N.; Anbazhagan, N.; Amutha, S.; Jeganathan, K.; Park, G.-C.; Joshi, G.P.;Cho, W. Controlled arrivals on the retrial queueing–inventory system with an essential interruption and emergency vacationing server. *Mathematics* **2023**, *11*, 3560 [CrossRef]
34. Reiyas, M.A.; Jeganathan, K. A classical retrial queueing inventory system with two component demand rate. *Int. J. Oper.* **2023**, *47*, 508–533. [CrossRef]
35. Jain, M.; Kumar, I. Cost optimization of a queueing inventory system with two-level supply mode, retrial demands and multiple vacations using a genetic algorithm. *Int. J. Appl. Comput. Math.* **2023**, *9*, 51. [CrossRef]
36. Jeganathan, K.; Koffer, V.A.; Sugapriya, C.; Nagarajan, D. A matrix-analytic method for the steady-state analysis of a Markovian queueing system with scrap items. *Decis. Anal. J.* **2023**, *7*, 100244. [CrossRef]
37. Bazizi, L.; Rahmoune, F.; Lekadir, O.; Labadi, K. Modelling, performance evaluation and optimisation of (s,Q) retrial inventory system with partial backlogging demands: A gspn approach. *Eur. J. Ind. Eng.* **2023**, *17*, 529–569. [CrossRef]
38. Jeganathan, K.; Harikrishnan, T.; Lakshmi, K.P.; Nagarajan, D. A multi-server retrial queueing-inventory system with asynchronous multiple vacations. *Decis. Anal. J.* **2023**, *9*, 100333. [CrossRef]
39. Chakravarthy, S.R. A multi-server queueing model with server consultations. *Eur. J. Oper. Res.* **2014**, *233*, 625–639. [CrossRef]
40. Hanukov, G. A service system where junior servers approach a senior server on behalf of customers. *Int. J. Prod. Econ.* **2022**, *244*, 108351. [CrossRef]
41. Chakravarthy, S.R.; Dudin, A.N.; Dudin, S.A.; Dudina, O.S. Queueing System with Potential for Recruiting Secondary Servers. Mathematics **2023**, *11*, 624. [CrossRef]
42. Neuts, M.F.; Rao, B.M. Numerical investigation of a multiserver retrial model. *Queueing Syst.* **1990**, *7*, 169–189. [CrossRef]
43. Neuts, M.F. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*; Courier Corporation: North Chelmsford, MA, USA, 1994.