symmetry

MDPI

Article

Transient Analysis of a Continuous-Service Markovian Queueing Model with Offline and Online Customers

Ramupillai Sudhesh [1], Paulsamy Balakrishnan [2,*] and Ratchaga Dass Sebasthi Priya [1]

[1] Department of Mathematics, UCE—BIT Campus, Anna University, Tiruchirappalli 620 024, Tamilnadu, India; sudhesh@aubit.edu.in (R.S.); sebasthipriya@aubit.edu.in (R.D.S.P.)
[2] Department of Mathematics, Adhiparasakthi Engineering College, Melmaruvathur 603 319, Tamilnadu, India
* Correspondence: balakrishnan@apec.edu.in

Abstract

This study examines a single-server Markovian queueing system featuring continuous service and an infinite number of customers at both ends—namely, offline and online clients. Offline customers are conventional clients who arrive at the system following a Poisson process, while online customers are assumed to be endlessly present in the system. All service times are exponentially and identically distributed and independent. Utilizing generating functions and Laplace transform techniques, this study derives exact analytical expressions for the system size probabilities in both transient and steady states. Furthermore, it evaluates key performance measures for each state and provides graphical representations to illustrate the system's dynamics, thereby enriching the understanding of its operational behavior. This work contributes to the advancement of priority-based queueing models and proposes a novel framework applicable to hybrid service architectures in contemporary digital ecosystems.

Keywords: online and offline customers; time-dependent and time-independent system size probabilities; generating functions; Laplace transform; Modified Bessel functions

MSC: 60K25

check for updates

Academic Editors: George Karabatsos and Quanxin Zhu

Received: 9 April 2025
Revised: 6 June 2025
Accepted: 2 July 2025
Published: 9 July 2025

Citation: Sudhesh, R.; Balakrishnan, P.; Sebasthi Priya, R.D. Transient Analysis of a Continuous-Service Markovian Queueing Model with Offline and Online Customers. *Symmetry* **2025**, *17*, 1097. https://doi.org/10.3390/sym17071097

Copyright: © 2025 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

Traditional queueing models, such as the $M/M/1$ and $M/M/c$ systems, assume that servers alternate between active and idle states based on customer arrivals. However, in many contemporary applications—particularly those involving real-time processing, automated systems, or high-demand services—servers must operate continuously without interruption. For instance, in cloud computing environments, virtual machines are required to handle both real-time user requests (offline customers) and background tasks (online customers) without downtime. Similarly, in organ transplant networks, the availability of donors (offline arrivals) must be matched with an ever-present waiting list (online queue) to ensure efficient allocation. To address such scenarios, we examine a continuous-service, single-server Markovian queue that incorporates two distinct types of customers. Thay are

1. Offline Customers: Arrive according to a Poisson process and join a queue.
2. Online Customers: Represent an infinite preexisting queue, ensuring the server always has tasks to process.

A key feature of this model is that the server never idles. Offline arrivals do not interrupt ongoing service but are given priority after each service completion. If offline

customers are present, the server attends to them first; otherwise, it continues processing online tasks. This mechanism ensures uninterrupted operation, making the model particularly suitable for systems where service delays or idle time are undesirable or costly.

Analyzing such a queueing system requires advanced mathematical techniques to capture both transient and steady-state behaviors. Since the service times for both offline and online customers follow an exponential distribution, the system lends itself to solutions using generating functions and Modified Bessel functions. These tools enable the derivation of transient-state probabilities and allow us to explore how the system evolves over time before reaching equilibrium.

The Modified Bessel function, in particular, provides a robust framework for analyzing systems characterized by infinite queues and priority-based service disciplines. Its utility in solving differential-difference equations arising in queueing theory makes it an ideal tool for the analysis of the proposed model. By employing these mathematical techniques, we derive key performance metrics, including the expected number of tasks in the system, the average waiting time for offline and online customers, and the probability of system saturation under varying arrival rates.

There are some gaps in Existing Queueing Models, such as

1. Assumption of Server Idleness: Classical queueing models assume that servers remain idle when no customers are present. However, in practical applications such as cloud computing, automated manufacturing, and healthcare systems, servers (or service resources) must operate continuously to avoid costly downtime or delays.
2. Single Customer Source: Most queueing models consider only a single type of customer arrival stream, thereby overlooking scenarios where two distinct customer populations (e.g., real-time requests versus background tasks) simultaneously compete for service.
3. Lack of Priority-Based Continuous Service: While priority queues have been extensively studied, they often involve preemptive mechanisms, where higher-priority tasks interrupt ongoing service. In contrast, many real-world systems—such as organ transplant networks and automated production lines—require non-preemptive priority handling to ensure operational stability and fairness.

To address the above limitations, this study introduces a novel single-server Markovian queueing model characterized by the following features:

1. Continuous Service Operation: The server never idles, ensuring uninterrupted task processing. This feature is essential for applications such as real-time cloud computing, automated logistics, and emergency healthcare systems.
2. Dual Customer Sources:
    - Offline Customers: Arrive according to a Poisson process (rate λ) and receive non-preemptive priority upon each service completion.
    - Online Customers: Represent an infinite preexisting queue, guaranteeing that the server always has tasks to process, even during periods of low offline demand.
3. Priority-Based Non-Preemptive Scheduling: Offline customers do not interrupt ongoing service but are served before online tasks once the current service is completed. This mechanism maintains fairness while ensuring system efficiency and operational continuity.
4. Advanced Analytical Framework:
    - Transient and Steady-State Analysis: Employs generating functions, Laplace transforms, and Modified Bessel functions to derive exact closed-form solutions for system size probabilities.

- Performance Metrics: Computes average waiting times, queue lengths, and saturation probabilities for both customer types under varying arrival conditions.

Real-World Motivations and Applications:

This model is rigorously grounded in the challenges of the following contemporary systems:

- **Cloud Computing and Digital Platforms:**
  - Balancing real-time user requests (offline) with background batch jobs (online) in data centers.
  - Achieving zero downtime in microservice-based architectures where idle servers incur financial penalties.
- **Healthcare Systems:**
  - Managing emergency cases (offline) alongside scheduled procedures (online) in hospitals.
  - Enhancing organ transplant networks, where offline donor arrivals must be matched with an ever-present online waiting list.
- **Smart Manufacturing and Automation:**
  - Handling rush orders (offline) while maintaining uninterrupted production lines (online).
  - Preventing operational bottlenecks in Industry 4.0 systems where machinery must remain continuously active.

Theoretical and Practical Significance

1. **Bridging a Critical Gap in Queueing Theory:**
   - Extends classical models by incorporating both dual customer sources and continuous service mechanisms.
   - Provides exact analytical solutions for both transient and steady-state behaviors, offering a more rigorous alternative to simulation-based methods.
2. **Actionable Insights for System Designers:**
   - Informs the development of service policies (e.g., optimal prioritization strategies).
   - Enables cost–benefit analyses for resource allocation in high-demand environments.
3. **Validation Through Numerical Experiments:**
   - Demonstrates the model's scalability and real-world applicability through graphical representations of system dynamics.

The remainder of this paper is structured as follows:

- Section 2 presents a literature review, summarizing prior research on priority queues, continuous-service models, and their applications in modern industries.
- Section 3 discusses practical implementations of the proposed model, with a focus on use cases such as the lost and found service desk in airports.
- Section 4 provides the detailed mathematical formulation and modeling of the proposed queueing system.
- Sections 5 and 6 develop the analytical solutions for transient and steady-state system size probabilities, respectively.
- Section 7 evaluates key performance metrics, including the mean queue length and waiting times.
- Section 8 presents a cost analysis aimed at identifying optimal service policies.
- Section 9 offers numerical results that illustrate transient behaviors and validate the theoretical findings.

- Section 10 concludes the paper with a summary of its key contributions and outlines directions for future research.

2. Literature Review

Queueing theory has been widely applied to optimize healthcare systems and improve patient access. The researchers Green [1] and Worthington [2] employed queueing models to analyze hospital waiting lists and healthcare service delivery. A notable application in organ transplantation is explored by Boxma et al. [3]. Zenios et al. [4] developed an innovative kidney allocation approach using dynamic index policies. Their methodology employed dynamic programming techniques to simultaneously achieve two key objectives: improving quality-adjusted life expectancy for recipients and decreasing average waiting times for transplants.

Waiting line problems have also been extensively studied in the context of manufacturing systems, with specific methodologies detailed in [5,6]. Cheng [7], along with the authors in [8,9], investigated the role of queueing theory in the sharing economy.

Several studies have analyzed single-server systems that serve two classes of customers. Gelenbe [10] introduced innovative models, including queueing networks with both positive and negative customers, which effectively represent dynamic behaviors observed in real-world systems. Parthasarathy [11] examined a single-server queue serving two customer types (Type I and Type II) under a gated service discipline. In another study, Choudhury and Deka [12] explored a single-server queue with two service phases, incorporating server interruptions and probabilistic vacation periods.

Previous research on priority queueing systems, including the work by Jaiswal [13], has extensively examined both single-server and multi-server configurations with multiple priority classes. These studies have investigated various service disciplines, encompassing both preemptive and non-preemptive priority schemes. A common characteristic of these models is their treatment of service time distributions, where all customer classes are typically assumed to follow the same probability distribution, although potentially with varying service rates. Klimenok et al. [14] studied a single-server queue with limited capacity, receiving two customer types through a batch-based Markov arrival process. Kim and Kim [15] analyzed a retrial queueing system involving two distinct customer categories: Type-2 customers entered a retrial orbit during busy periods, while Type-1 customers formed an unbounded queue.

In [16], Opher Baron et al. studied queue redirection strategies in a single-server network serving two customer types with varying service time requirements. Tikhonenko [17] analyzed an $M/G/1/\infty$ queue with two customer classes—external and internal. Hanukov [18] proposed a combined queueing-inventory model serving two categories: skeptical and trusting customers. For the latter group, the service process was divided into two phases: an initial opening service followed by a complementary one.

Linhong Li et al. [19] explored optimization techniques for systems serving two customer categories, applying service decomposition under constrained inventory conditions. Yitong Zhang et al. [20] investigated individual equilibrium decisions and socially optimal strategies in a fluid queueing system with two parallel customer types, factoring in alternating normal and partial failure states in the service buffer.

Queueing theory also finds applications in multimedia and telecommunication systems. Parthasarathy et al. [21] examined a specialized queueing model for multimedia synchronization involving two media streams. The model processed packets in pairs (one from each stream) and used continued fraction techniques to derive performance metrics. In [22], Sudhesh et al. studied Denial-of-Service (DoS) scenarios in IEEE 802.16 networks, incorporating system disasters into their transient probability analysis.

Latouche [23] analyzed bilateral matching queues with paired inputs using matrix geometric methods. Transient analysis in queueing theory is mathematically more complex than steady-state analysis. In Markovian systems, studying transient behavior often involves addressing added complexities such as queue discipline variations, server repair processes, and other dynamic factors. For example, Sudhesh et al. [24] performed a transient analysis of a Geo/Geo/1 queueing system under catastrophic event scenarios.

3. Motivating Example

This study investigates a continuous-service Markovian queueing system featuring a single server, where both inter-arrival times and service durations follow exponential probability distributions. Under these conditions, the server operates continuously without interruption. The system accommodates two distinct types of customers: offline and online. It is assumed that an infinite number of online customers already exist in the system, while offline customers arrive according to a Poisson process.

Offline customers must wait in the queue, as they cannot bypass those (either offline or online) already waiting for service. However, priority is given to offline customers over online customers. Consequently, the server always prioritizes offline customers when at least one is present in the queue. After completing the service for one offline customer, the server immediately begins servicing the next offline customer, if any are waiting. Online customers are served only when no offline customers are in the queue. Due to the infinite backlog of online customers and the server's uninterrupted operation, online service occurs only in the absence of offline arrivals.

This single-server Markovian queueing model with dual customer sources can be effectively mapped to a real-world scenario—specifically, the lost and found (LF) service desk at an airport:

1. **Server and Service Mechanism:** The single server represents the LF desk agent. Service times (exponentially distributed with rate μ) correspond to the time taken to process a lost item claim or inquiry. The service discipline ensures that walk-in passengers (offline customers) are always prioritized over online inquiries (online customers).
2. **Customer Arrivals:**
    - **Offline (Regular) Customers:** Passengers who physically arrive at the LF desk, modeled as a Poisson arrival process with rate λ.
    - **Online Customers:** Emails or digital queries submitted via the airport's lost property portal, assumed to have an infinite backlog.
3. **Priority-Based, Non-Preemptive Service Policy:** The desk always serves walk-in passengers first, one at a time. Online queries are addressed only when there are no walk-in customers. A new walk-in arrival does not interrupt an ongoing service for an online query, ensuring a non-preemptive service structure.
4. **System States (X(t)) and their Meaning in the LF Context:**
    - **Positive States (k > 0):**
      k = 1: One walk-in passenger is being served, with none waiting.
      k = 2: One walk-in passenger is being served, with one more waiting.
    - **Negative States (k < 0):**
      k = −1: One walk-in passenger is waiting while an online query is being processed.
      k = −2: Two walk-in passengers are in the queue during the service of an online query.
      k = 0: No walk-in customers; the desk is currently handling an online query.

Key Implications for Airport Lost and Found Operations

- **Prioritization Efficiency:** Ensures in-person passengers receive immediate attention, improving customer satisfaction.
- **Online Query Backlog:** With an infinite number of online submissions, the desk alternates between walk-in and online modes without experiencing idle time.
- **Non-Preemption:** Prevents interruptions in ongoing online query processing upon the arrival of walk-in customers, promoting fairness and operational consistency.
- **State-Dependent Dynamics:** Queue lengths and service transitions reflect real-world behavior at LF desks, offering practical insights for optimizing service operations.

This Markovian model effectively captures the dual-channel service dynamics of an airport lost and found desk, where walk-in claimants are given priority over digital requests, while ensuring continuous utilization of the service agent. The state transitions provide a robust framework for analyzing queue lengths, waiting times, and optimal resource allocation in such hybrid service environments.

Assuming exponentially distributed service times for both offline and online customers, we utilize generating functions and Modified Bessel functions to compute transient queue-length probabilities in this single-server, dual-source system. Furthermore, the Laplace transform method is employed to derive steady-state queue-length probabilities. These analytical results are validated against established studies, notably the work of Song Chew [25]. Computational illustrations further demonstrate the evolution of queue-length probabilities in both transient and steady-state regimes for both customer types. Additionally, key performance metrics are explored to offer deeper insights into the system's operational dynamics.

4. Model Description

This study examines a single-server Markovian queueing system characterized by continuous service and an unbounded customer population originating from two independent sources: offline and online arrivals. Offline (regular) customers arrive according to a Poisson process with the rate λ, and the system assumes an infinite number of online customers are always present. All service times are exponentially distributed with the rate μ.

A key characteristic of the model is that the server always gives precedence to offline customers, serving them in the order of arrival. Online customer service is initiated only when no offline customers remain in the system. Once the server begins serving an online customer, the service continues uninterrupted, even if new offline customers arrive. This non-preemptive service discipline ensures that ongoing services are not disrupted.

Let $\{X(t), t \geq 0\}$ represent the number of offline customers (clients) in the system at any time t.

Define $P_n(t)$, $(n = \cdots -3, -2, -1, 0, 1, 2, 3 \ldots)$ as the probability of having n offline customers at time t. The system is in a positive state when offline customers are being served, and in a non-positive state when the server is attending to online customers. Specifically, let k represent the total number of offline clients currently in the system.

- $k = 0$ indicates that no offline customers are present, and an online customer is being served.
- $k = 1$ indicates that there is one offline customer in the system (being served, with no queue).
- $k = 2$ denotes two offline customers (one in service, one waiting).
- In general, $k = j$ means there are j offline customers, with one being served and $(j - 1)$ in the queue.

For non-positive states:

- $k = -1$ indicates that one offline customer is waiting while an online customer is being served.
- $k = -2$ indicates that two offline customers are waiting in the queue during the service of an online customer.
- More generally, $k = -j$ indicates j offline customers are queuing while the server is engaged with an online customer.

This formulation allows the model to account for all possible system states while maintaining continuous operation and respecting the offline priority policy. Figure 1 illustrates the state transition diagram of the proposed queueing model, highlighting the system's evolution through various positive and non-positive states.

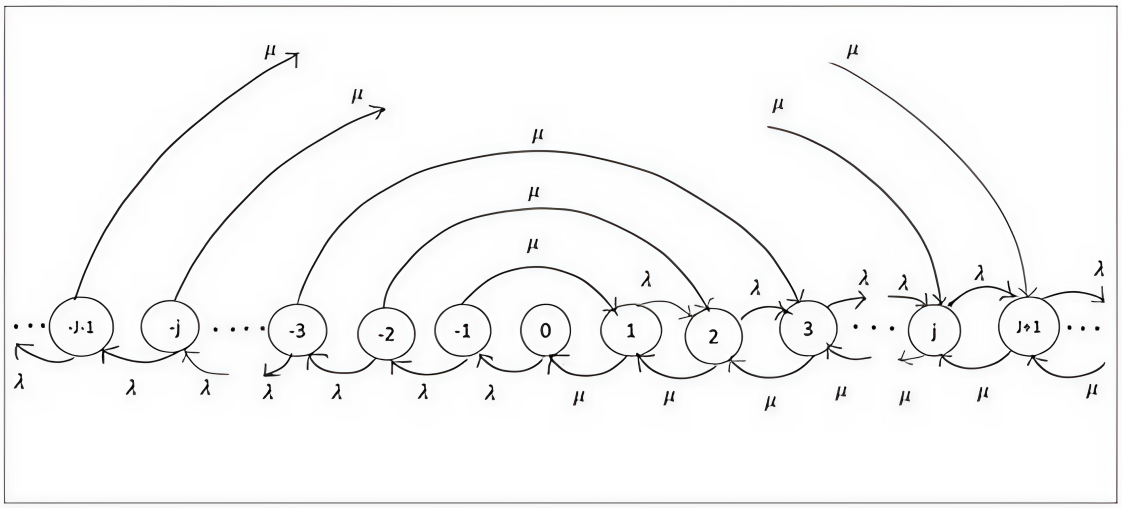

Figure 1. Transition diagram.

The matrix q_{ij} corresponds to the transition probabilities of the Markov chain. The (i, j)-component q_{ij} is defined as

$$q_{ij} = \begin{cases} \lambda & \text{if} \quad j = i+1, \quad i \geq 1 \\ \lambda & \text{if} \quad j = i-1, \quad i \leq 0 \\ \mu & \text{if} \quad j = i-1, \quad i \geq 1 \\ \mu & \text{if} \quad j = -i, \quad i \leq -1 \end{cases}$$

Then, $P_n(t)$ satisfies the forward Chapman Kolmogorov equations for OFFLINE customers,

$$P_0'(t) = -\lambda P_0(t) + \mu P_1(t), \tag{1}$$

$$P_1'(t) = \mu[P_{-1}(t) + P_2(t)] - (\lambda + \mu)P_1(t) \tag{2}$$

$$P_2'(t) = \mu[P_{-2}(t) + P_3(t)] - (\lambda + \mu)P_2(t) + \lambda P_1(t) \tag{3}$$

$$P_3'(t) = \mu[P_{-3}(t) + P_4(t)] - (\lambda + \mu)P_3(t) + \lambda P_2(t) \tag{4}$$

$$P_n'(t) = \mu[P_{-n}(t) + P_{n+1}(t)] - (\lambda + \mu)P_n(t) + \lambda P_{n-1}(t), n = 2, 3, 4, \ldots \tag{5}$$

and for ONLINE customers, we have

$$P_{-1}'(t) = \lambda P_0(t) - (\lambda + \mu)P_{-1}(t), \tag{6}$$

$$P_{-n}'(t) = \lambda P_{-(n-1)}(t) - (\lambda + \mu)P_{-n}(t), n = 1, 2, 3, \ldots \tag{7}$$

Probabilities for transient system size in the continuous-service queue appear in the following section.

5. Time-Dependent Probabilities

This section employs generating functions and Laplace transforms to obtain closed-form solutions for the time-dependent (transient state) system size probabilities of both online and offline customers.

Let us assume that initially the system is idle, i.e., $P_0(0) = 1$.

From Equation (5), by applying Laplace Transforms, we have

$$\hat{f}_{-n}(s) = \left(\frac{\lambda}{s+\lambda+\mu}\right)^n \hat{f}_0(s) = \chi_n(s)\hat{f}_0(s), \tag{8}$$

where $\chi_n(s) = \left(\frac{\lambda}{s+\lambda+\mu}\right)^n$.

On inversion, we obtain, for $n = 1, 2, 3, \ldots$

$$\begin{aligned} P_{-n}(t) &= \chi_n(t) * P_0(t) \\ &= \lambda^n e^{-(\lambda+\mu)t} \frac{t^{n-1}}{(n-1)!} * P_0(t). \end{aligned} \tag{9}$$

The Laplace transform of (1) yields

$$\hat{f}_0(s) = \frac{1}{s+\lambda} + \frac{\mu}{s+\lambda}\hat{f}_1(s).$$

On inversion, we obtain

$$P_0(t) = e^{-\lambda t} + \mu e^{-\lambda t} * P_1(t). \tag{10}$$

Let $G(z,t) = \sum_{n=0}^{\infty} P_n(t)z^n$ and $G_1(z,t) = \sum_{n=1}^{\infty} P_{-n}(t)z^n$ be the generating functions under the initial condition $G(z,0) = 1$.

From Equations (5) and (7),

$$G'(z,t) + \left[\lambda + \mu - \lambda z - \frac{\mu}{z}\right] G(z,t) = \mu G_1(z,t) + \left[\mu - \lambda z - \frac{\mu}{z}\right].$$

By integrating the equation above, we obtain

$$\begin{aligned} G(z,t) = \sum_{n=0}^{\infty} P_n(t)z^n &= \int_0^t \left[\mu G_1(z,y) + P_0(y)\left(\frac{-\lambda z^2 + \mu z - \mu}{z}\right)\right] e^{-(\lambda+\mu)(t-y)} e^{\left(\lambda z + \frac{\mu}{z}\right)(t-y)} dy \\ &\quad + e^{-(\lambda+\mu)(t)} e^{\left(\lambda z + \frac{\mu}{z}\right)(t)}. \end{aligned} \tag{11}$$

One can assume that if $\alpha = 2\sqrt{\lambda\mu}$ and $\beta = \sqrt{\frac{\lambda}{\mu}}$, then

$$e^{\left(\lambda z + \frac{\mu}{z}\right)(t)} = \sum_{n=-\infty}^{\infty} (\beta z)^n I_n(\alpha(t)), \tag{12}$$

where $I_n(\alpha t)$ stands for the Modified Bessel function of the first kind, a widely established mathematical function.

Evaluation of $P_n(t)$, $n \geq 1$

Equating the coefficients of z^n on both sides of Equation (11), we obtain, for $n = 1, 2, 3, \ldots$

$$\begin{aligned} P_n(t) &= \mu \int_0^t e^{-(\lambda+\mu)(t-y)} (P_{-1}(y)\beta^{n-1} I_{n-1}(.) + P_{-2}(y)\beta^{n-2} I_{n-2}(.) + \ldots \\ &+ P_{-n}(y) I_0(.) + P_{-(n+1)}(y)\beta^{-1} I_1(.) + P_{-(n+2)}(y)\beta^{-2} I_2(.) + \ldots \\ &+ P_{-(n+n)}(y)\beta^{-n} I_n(.) + P_{-(2n+1)}(y)\beta^{-(n+1)} I_{n+1}(.) + \ldots) dy \\ &+ \int_0^t P_0(y) e^{-(\lambda+\mu)(t-y)} \left[-\lambda\beta^{n-1} I_{n-1}(.) + \mu\beta^n I_n(.) - \mu\beta^{n+1} I_{n+1}(.) \right] dy \\ &+ e^{-(\lambda+\mu)t} \beta^n I_n(\alpha t). \end{aligned} \tag{13}$$

The above equation remains valid for $n = -1, -2, -3 \ldots$ when the left side is set to zero, and using the property $I_{-n}(.) = I_n(.)$, for $n = 1, 2, 3, \ldots$

$$\begin{aligned} 0 &= \mu \int_0^t e^{-(\lambda+\mu)(t-y)} (P_{-1}(y)\beta^{n-1} I_{n+1}(.) + P_{-2}(y)\beta^{n-2} I_{n+2}(.) + \ldots \\ &+ P_{-n}(y) I_{2n}(.)] + P_{-(n+1)}(y)\beta^{-1} I_{2n+1}(.) + P_{-(n+2)}(y)\beta^{-2} I_{2n+2}(.) + \ldots \\ &+ P_{-(n+n)}(y)\beta^{-n} I_{2n+n}(.) + \ldots) dy \\ &+ \int_0^t P_0(y) e^{-(\lambda+\mu)(t-y)} \left[-\lambda\beta^{n-1} I_{n+1}(.) + \mu\beta^n I_n(.) - \mu\beta^{n+1} I_{n-1}(.) \right] dy \\ &+ e^{-(\lambda+\mu)t} \beta^n I_n(\alpha t). \end{aligned} \tag{14}$$

From (13) and (14), for $n = 1, 2, 3, \ldots$

$$\begin{aligned} P_n(t) &= \mu \int_0^t e^{-(\lambda+\mu)(t-y)} \left[\sum_{k=1}^{n} P_{-k}(y)\beta^{n-k} [I_{n-k}(.) - I_{n+k}(.)] \right] dy \\ &+ \mu \int_0^t e^{-(\lambda+\mu)(t-y)} \left[\sum_{k=1}^{\infty} P_{-(k+n)}(y)\beta^{-k} [I_k(.) - I_{2n+k}(.)] \right] dy \\ &+ \int_0^t e^{-(\lambda+\mu)(t-y)} P_0(y) \left[-\lambda\beta^{n-1} + \mu\beta^{n+1} [I_{n-1}(.) - I_{n+1}(.)] \right] dy, \end{aligned} \tag{15}$$

where $I_n(.) = I_n(\alpha(t-y))$. Now, let us consider, from (15),

$$P_n(t)_A = \int_0^t e^{-(\lambda+\mu)(t-y)} P_0(y) \left[-\lambda\beta^{n-1} + \mu\beta^{n+1} [I_{n-1}(.) - I_{n+1}(.)] \right] dy. \tag{16}$$

For n = 1,

$$P_1(t)_A = \int_0^t e^{-(\lambda+\mu)(t-y)} P_0(y) \left[-\lambda + \mu\beta^2 [I_0(.) - I_2(.)] \right] dy = 0, \tag{17}$$

where $\beta^2 = \frac{\lambda}{\mu}$. Similarly, $P_1(t)_A = P_2(t)_A = P_3(t)_A = \ldots P_n(t)_A = 0$.

Hence, for $n = 1, 2, 3, \ldots$, Equation (15), reduces as

$$\begin{aligned} P_n(t) &= \mu \int_0^t \left(\sum_{k=1}^{n} P_{-k}(y)\beta^{n-k} [I_{n-k}(.) - I_{n+k}(.)] \right. \\ &\left. + \sum_{k=1}^{\infty} P_{-(k+n)}(y)\beta^{-k} [I_k(.) - I_{2n+k}(.)] \right) e^{-(\lambda+\mu)(t-y)} dy. \end{aligned}$$

Through Laplace Transforms on Equation (15), we obtain

$$\begin{aligned}
\hat{f}_n(s) &= \frac{\mu\hat{f}_0(s)}{\sqrt{p^2-\alpha^2}}\sum_{k=1}^{n}\hat{f}_{-k}(s)\beta^{n-k}\left[\left(\frac{p-\sqrt{p^2-\alpha^2}}{\alpha}\right)^{n-k}-\left(\frac{p-\sqrt{p^2-\alpha^2}}{\alpha}\right)^{n+k}\right] \\
&+\frac{\mu\hat{f}_0(s)}{\sqrt{p^2-\alpha^2}}\sum_{k=1}^{\infty}\hat{f}_{-(k+1)}(s)\beta^{-k}\left[\left(\frac{p-\sqrt{p^2-\alpha^2}}{\alpha}\right)^{k}-\left(\frac{p-\sqrt{p^2-\alpha^2}}{\alpha}\right)^{2n+k}\right] \\
&= \frac{\mu\hat{f}_0(s)}{\sqrt{p^2-\alpha^2}}\sum_{k=1}^{n}\left(\frac{\lambda}{p}\right)^{k}\beta^{n-k}\left[\left(\frac{p-\sqrt{p^2-\alpha^2}}{\alpha}\right)^{n-k}-\left(\frac{p-\sqrt{p^2-\alpha^2}}{\alpha}\right)^{n+k}\right] \\
&+\frac{\mu\hat{f}_0(s)}{\sqrt{p^2-\alpha^2}}\sum_{k=1}^{\infty}\left(\frac{\lambda}{p}\right)^{k+n}\beta^{-k}\left[\left(\frac{p-\sqrt{p^2-\alpha^2}}{\alpha}\right)^{k}-\left(\frac{p-\sqrt{p^2-\alpha^2}}{\alpha}\right)^{2n+k}\right]. \qquad (18)
\end{aligned}$$

After some calculations and inversion, we obtain $P_n(t)$ in a compact form for $n = 1, 2, 3\ldots$ as follows:

$$\begin{aligned}
P_n(t) &= \mu\sum_{k=1}^{n}\lambda^k\beta^{n-k} * e^{-(\lambda+\mu)(t)}\frac{t^{k-1}}{(k-1)!}[I_{n-k}(\alpha t)-I_{n+k}(\alpha t)] * P_0(t) \\
&+\mu\sum_{k=1}^{\infty}\lambda^{k+n}(t)\beta^{-k} * e^{-(\lambda+\mu)(t)}\frac{t^{k+n-1}}{(k+n-1)!}[I_k(\alpha t)-I_{2n+k}(\alpha t)] * P_0(t) \qquad (19)
\end{aligned}$$

For $n = 1$, Equation (19) becomes

$$P_1(t) = \sqrt{\lambda\mu}e^{-(\lambda+\mu)t}\sum_{k=1}^{\infty}(\lambda\mu)^{\frac{k}{2}}\frac{t^{k-1}}{(k-1)!}\frac{2(k+1)}{\alpha t}I_{k+1}(\alpha t) * P_0(t) \qquad (20)$$

The Laplace transform of the above equation yields

$$\hat{f}_1(s) = \sum_{k=1}^{\infty}\left[\frac{p-\sqrt{p^2-\alpha^2}}{2p}\right]^k\hat{f}_0(s).$$

After further calculation, we obtain

$$\hat{f}_1(s) = \left[\frac{p-\sqrt{p^2-\alpha^2}}{p+\sqrt{p^2-\alpha^2}}\right]\hat{f}_0(s), \qquad (21)$$

where $p = s + \lambda + \mu$.

On inversion, we obtain

$$P_1(t) = \frac{2I_2(\alpha t)}{t} * P_0(t). \qquad (22)$$

Hence, all transient probabilities $P_n(t)$, $P_{-n}(t)$ are computed in terms of $P_0(t)$, and according to the normalization principle, $\sum_{n=0}^{\infty}P_n(t) + \sum_{n=1}^{\infty}P_n(t) = 1$, in principle, one can obtain the probability $P_0(t)$, thereby enabling the explicit determination of all system size probabilities.

6. Steady-State Probabilities

This section derives explicit expressions for the steady-state system size probabilities of the proposed model through Laplace transform techniques.

Applications of the Tauberian theorem for Laplace transforms in Equation (18) yield

$$\begin{aligned} P_n &= \lim_{s\to 0} s\hat{f}_n(s) \\ &= \left(\frac{\mu}{\mu-\lambda}\right)\lim_{s\to 0} s\hat{f}_0(s)\left(\frac{\lambda}{\mu}\right)^n\left[\sum_{k=1}^{n}\left(\frac{\mu^k-\lambda^k}{(\lambda+\mu)^k}\right)+\left(\frac{\mu^n-\lambda^n}{\mu^n}\right)\sum_{k=1}^{\infty}\left(\frac{\lambda}{\lambda+\mu}\right)^{k+n}\right] \\ &= P_0\left(\frac{\rho^n}{1-\rho}\right)\left[\frac{\mu-\lambda}{\mu+\lambda}+\frac{\mu^2-\lambda^2}{(\lambda+\mu)^2}+\cdots+\frac{\mu^n-\lambda^n}{(\lambda+\mu)^n}+\frac{\mu^n-\lambda^n}{(\lambda+\mu)^n}\rho\right]. \end{aligned}$$

This is valid for $n = 1, 2, 3, \ldots$.
For $n = 1$, the equation reduces to

$$P_1 = \lim_{s\to 0} s\hat{f}_1(s) = \left(\frac{\rho}{1-\rho}\right)\lim_{s\to 0} s\hat{f}_0(s)\left[\frac{\mu-\lambda}{\mu+\lambda}+\frac{\mu-\lambda}{\mu+\lambda}\rho\right] = \rho P_0.$$

When $n = 2$, we obtain

$$\begin{aligned} P_2 = \lim_{s\to 0} s\hat{f}_2(s) &= \left(\frac{\rho^2}{1-\rho}\right)\lim_{s\to 0} s\hat{f}_0(s)\left[\frac{\mu-\lambda}{\mu+\lambda}+\frac{\mu^2-\lambda^2}{(\mu+\lambda)^2}+\frac{\mu^2-\lambda^2}{(\mu+\lambda)^2}\rho\right] \\ &= \rho\left(\frac{(1+\rho)^2-1}{1+\rho}\right)P_0. \end{aligned}$$

Similarly, when $n = 3$, we obtain

$$P_3 = \lim_{s\to 0} s\hat{f}_3(s) = \rho^2\left(\frac{(1+\rho)^3-1}{(1+\rho)^2}\right)P_0.$$

In general, for $n = 1, 2, 3, \ldots$

$$P_n = \lim_{s\to 0} s\hat{f}_n(s) = \rho^{n-1}\left(\frac{(1+\rho)^n-1}{(1+\rho)^{n-1}}\right)P_0. \tag{23}$$

Using the normalization principle, we obtain

$$\sum_{n=1}^{\infty}\hat{f}_{-n}(s)+\hat{f}_0(s)+\sum_{n=1}^{\infty}\hat{f}_n(s) = \frac{1}{s}.$$

Using Equations (8) and (23), and after some calculations and as a limiting case, we obtain

$$P_0 = \frac{1-\rho}{1+\rho}. \tag{24}$$

By substituting the value of P_0 in (8) and (23), we obtain the steady-state system size probabilities for $n = 1, 2, 3, \ldots$ as follows:

$$\lim_{s\to 0} s\hat{f}_n(s) = P_n = \rho^{n-1}(1-\rho)\left(1-\frac{1}{(1+\rho)^n}\right) \tag{25}$$

and

$$\lim_{s\to 0} s\hat{f}_{-n}(s) = P_{-n} = \frac{\rho^n(1-\rho)}{(1+\rho)^{n+1}} \tag{26}$$

Thus, we find that all the system size probabilities in steady state for online and offline customers are as follows:

$$\begin{aligned} P_0 &= \frac{1-\rho}{1+\rho} \\ P_n &= \rho^{n-1}(1-\rho)\left(1-\frac{1}{(1+\rho)^n}\right) \\ P_{-n} &= \frac{\rho^n(1-\rho)}{(1+\rho)^{n+1}}. \end{aligned}$$

Another Method to Find Steady-State Probabilities

The left-hand sides of Equations (1)–(7) can also be adjusted to zero in order to obtain the probability of a time-independent system size for the provided model.

$$\begin{aligned} 0 &= -\lambda P_0 + \mu P_1 \\ 0 &= -(\lambda+\mu)P_1 + \mu P_2 + \mu P_{-1} \\ 0 &= -(\lambda+\mu)P_2 + \mu(P_{-2}+P_3) + \lambda P_1 \end{aligned}$$

In general, for $n = 0, 1, 2, 3, 4, \ldots$

$$0 = -(\lambda+\mu)P_n + \mu(P_{-n}+P_{n+1}) + \lambda P_{n-1}$$

From the above equations, we obtain, for $n = 0, 1, 2, \ldots$

$$\mu P_{n+1} = (\lambda+\mu)P_n - \lambda P_{n-1} - \mu P_{-n}. \tag{27}$$

Similarly, from Equations (6) and (7), we obtain, for $n = 1, 2 \ldots$

$$0 = \lambda P_{-n} - (\lambda+\mu)P_{-(n-1)}.$$

After some simplifications, the above equations provide the following result for $n = 1, 2, 3, \ldots$:

$$\mu P_{-n} = \lambda P_{-(n-1)} - \lambda P_{-n}. \tag{28}$$

Again, employing some mathematical calculations, we obtain, for $n = 1, 2, 3, \ldots$

$$\mu P_n = \lambda P_{n-1} + \lambda P_{-(n-1)} \tag{29}$$

By adding all the above equations, we obtain

$$P_1 + P_2 + P_3 \ldots = \rho \tag{30}$$

From Equation (24), we obtain

$$\lambda P_0 = \left(\frac{\lambda+\mu}{\lambda}\right) P_{-1}$$

Thus, we have

$$P_{-n} = \left(\frac{\lambda}{\lambda+\mu}\right)^n P_0 \tag{31}$$

Using the normalization principle, we obtain

$$P_0 = \frac{1-\rho}{1+\rho} \tag{32}$$

By substituting the value of P_0 into Equations (25) and (27), we obtain the time-independent system size probabilities for $n = 1, 2, 3, \ldots$ as

$$\begin{aligned} P_n &= \rho^{n-1}(1-\rho)\left(1 - \frac{1}{(1+\rho)^n}\right) \\ P_{-n} &= \frac{\rho^n(1-\rho)}{(1+\rho)^{n+1}} \end{aligned} \tag{33}$$

These results coincide with those derived from transient system size probabilities using the Laplace transform, as well as with the established findings in the literature.

7. Performance Measures

This section focuses on deriving key performance measures of the proposed queueing system, including the mean of time-dependent and time-independent system size probabilities, the energy-saving factor, and system throughput under steady-state conditions.

7.1. Mean System Size in Transient State

In this subsection, we establish a formula to compute the expected number of customers in the system at any given time t. Initially, the system is assumed to be empty. Let $E(X(t))$ denote the expected system size at time t. This measure captures the dynamic behavior of the queue before it reaches equilibrium and provides insights into the system's responsiveness during transient phases.

$$E(X(t)) = \sum_{n=1}^{\infty} nP_{-n}(t) + \sum_{n=0}^{\infty} nP_n(t),$$

and from the Equations (2)–(5) and (7)

$$\begin{aligned} [E(X(t))]' &= \sum_{n=1}^{\infty} nP'_{-n}(t) + \sum_{n=0}^{\infty} nP'_n(t) \\ &= (\lambda - \mu)\sum_{n=1}^{\infty} P_n(t) + \lambda \sum_{n=0}^{\infty} P_{-n}(t) \end{aligned}$$

Here, $[E(X(t))]'$ indicates the instantaneous rate of change of $E(X(t))$ over time t.

After resolving the previous equation and performing more basic algebraic computations, we obtain

$$m(t) = E(X(t)) = (\lambda - \mu)\sum_{n=1}^{\infty} \int_0^t P_n(y)dy + \lambda \sum_{n=1}^{\infty} \int_0^t P_{-n}(y)dy, \tag{34}$$

where $P_n(t)$ and $P_{-n}(t)$ are given in (9) and (19).

By implementing the Laplace Transform, the average number of clients in the system is

$$m(s) = (\lambda - \mu)\sum_{n=1}^{\infty} \frac{\hat{f}_n(s)}{s} + \lambda \sum_{n=1}^{\infty} \frac{\hat{f}_{-n}(s)}{s}$$

7.2. Expected Steady-State System Size

Under steady-state conditions, let L_S denote the average number of customers in the system considering both offline and online modes. Then,

$$\begin{aligned} L_s &= \sum_{n=1}^{\infty} nP_{-n} + \sum_{n=0}^{\infty} nP_n \\ &= \rho\left(\frac{2-\rho}{1-\rho}\right) \end{aligned}$$

From this, we obtain

$$L_q = \frac{\rho}{1-\rho}$$

Applying Little's Law enables the derivation of additional performance measures as

$$\begin{aligned} W_s &= \frac{(2-\rho)\rho}{\lambda(1-\rho)} \\ W_q &= \frac{\rho}{\lambda(1-\rho)} \end{aligned}$$

7.3. Energy-Saving Factor

We can now calculate the energy-saving factor $E(s)$, which is the proportion of idle time spent by the system

$$E(s) = P_0 \times 100 = \left(\frac{1-\rho}{1+\rho}\right) \times 100.$$

7.4. System Throughput in a Transient State

Let $M(t)$ describe the system's throughput or processing rate at time t, with its definition given as

$$M(t) = \mu \times \sum_{n=1}^{\infty} (P_n(t) + P_{-n}(t)),$$

where $P_n(t)$ and $P_{-n}(t)$ are determined using Equations (9) and (19), respectively.

7.5. Throughput of the System in a Steady State

This section evaluates the system throughput U, which represents the rate at which customers are served during periods when the system is not empty. Throughput is calculated based on the service rate μ and reflects the average number of customer requests processed per unit time. Alternatively, it can be interpreted as the mean number of requests successfully handled by the server within a given time interval under steady-state conditions.

$$\begin{aligned} U &= \sum_{n=1}^{\infty} (P_{-n} + P_n) \times \mu \\ &= \sum_{n=1}^{\infty} \left(\frac{\rho^n(1-\rho)}{(1+\rho)^{n+1}} + \rho^{n-1}(1-\rho)\left(1 - \frac{1}{(1+\rho)^n}\right)\right) \times \mu \\ &= \frac{2\lambda\mu}{\lambda+\mu}. \end{aligned}$$

The proportion of customers served is expressed as $N_s = \frac{U}{\lambda}$.

7.6. Remark

The probability that the server is busy $= \sum_{n=1}^{\infty}(P_{-n} + P_n) = \frac{2\rho}{1+\rho}$.

The probability that the server is busy, given that the server is active $= \frac{2\rho}{1-\rho}$.

8. Cost Analysis

This section presents an economic evaluation of the proposed system, defining both Total Expected Revenue (TER) and Total Expected Cost (TEC) as follows:

$$TEC = C_0 P_n + C_1 P_{-n} + C_\mu \mu + C_h L_s$$

where

C_0 : Cost per unit time for the server serving offline clients.
C_1 : Cost per unit time for the server serving online clients.
C_μ : Cost per customer served with rate μ.
C_h : Carrying Cost.

Let R denote the revenue earned by assisting a client (both online and offline modes); then, the Total Expected Revenue (TER) is given by

$TER = R \times L_s$
Case 1: $C_0 = 50, C_1 = 30, C_h = 80, C_\mu = 40, R = 5$
Case 2: $C_0 = 50, C_1 = 30, C_h = 80, C_\mu = 40, R = 5$
Case 3: $C_0 = 50, C_1 = 30, C_h = 80, C_\mu = 40, R = 5$
Case 4: $C_0 = 50, C_1 = 30, C_h = 80, C_\mu = 40, R = 5$
Case 5: $C_0 = 50, C_1 = 30, C_h = 80, C_\mu = 40, R = 5$

Table 1 presents the optimal service rate μ and corresponding Total Expected Cost (TEC). The data shows that both μ and TEC grow with increasing arrival rate λ, independent of cost parameters. This demonstrates a clear relationship between λ and system performance metrics. Tables 2 and 3 reveal how μ and λ influence TEC and Total Expected Revenue (TER). Higher service rates reduce both TEC and TER by decreasing the average system occupancy L_s, while increased arrival rates elevate them due to higher L_s values.

Table 1. Optimal values of μ and the corresponding Total Expected Cost (TEC) for various scenarios and arrival rates.

Case	λ	**0.12**	**0.14**	**0.16**	**0.18**	**0.2**	**0.22**	**0.24**	**0.26**	**0.28**	**0.3**
1	μ	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
	TEC	596.4214	608.6084	619.41	628.6111	636.1066	641.874	645.9531	648.4302	649.4266	649.0901
2	μ	1.7	1.7	1.7	1.7	1.7	1.7	1.7	1.7	1.7	1.7
	TEC	589.9442	596.0133	602.3215	608.7516	615.204	621.5943	627.8516	633.9169	639.7412	645.2848
3	μ	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8
	TEC	621.4159	624.323	627.4774	630.8379	634.3669	638.0303	641.7967	645.6379	649.5279	653.4435
4	μ	3.2	3.2	3.2	3.2	3.2	3.2	3.2	3.2	3.2	3.2
	TEC	635.4247	637.7643	640.323	642.0352	645.9806	649.026	652.1838	655.4322	658.7511	662.1221
5	μ	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5
	TEC	683.9858	685.2936	686.7466	688.3322	690.0383	691.8535	693.7669	695.7684	697.8482	699.9973

Table 2. Total Expected Cost (TEC) and Total Expected Revenue (TER) as functions of the service rate μ.

μ	**Total Expected Cost**	**Total Expected Revenue**
0.2	173.1865	865.93
0.3	132.0253	660.13
0.4	117.2440	586.22

Table 2. *Cont.*

μ	Total Expected Cost	Total Expected Revenue
0.5	110.2410	551.21
0.6	106.9352	534.68
0.7	105.7398	528.7
0.8	105.8710	529.36

Table 3. Total Expected Cost (TEC) and Total Expected Revenue (TER) as functions of arrival rate λ.

λ	Total Expected Cost	Total Expected Revenue
0.12	103.3761	516.8804
0.14	104.7563	523.7816
0.16	106.0832	530.4161
0.18	107.3377	536.6884
0.2	108.5055	542.5275
0.22	109.5766	547.883
0.24	110.5442	552.7211
0.26	111.4044	557.0218
0.28	112.1533	560.7766
0.3	112.7972	563.9858

9. Numerical Illustrations

This section presents a numerical analysis of the proposed queueing model. The parameters are assigned the following values: $\lambda = 0.5$, $\mu = 0.9$. Figures 2 and 3 display the graphs of $P_n(t)$ and $P_{-n}(t)$, respectively. It is assumed that $P_0(0) = 1$; hence, the probability curve for $P_0(0)$ starts at 1, gradually decreases over time, and eventually stabilizes at a steady-state value. The remaining curves for $P_n(t)$ and $P_{-n}(t)$ begin at 0 and asymptotically approach their respective steady-state levels.

Figures 4 and 5 present the transient-state mean system size and its variance. The curves illustrate a proportional increase in system size and variability with rising values of λ, assuming μ remains constant.

Figures 6–8 illustrate the system size probabilities in steady-state conditions. Specifically, Figures 6 and 7 highlight the relationship between system size (n) and different arrival rates. A distinct trend is observed: as n increases, the corresponding probability values decrease and ultimately converge toward zero. This pattern suggests that larger system sizes are increasingly less probable, reinforcing the model's ability to manage queue lengths effectively.

Figure 8 offers additional insights by demonstrating the impact of increasing the service rate while keeping the arrival rate constant. The graph reveals an inverse relationship between the service rate and system size. This emphasizes the critical role of service efficiency in maintaining a manageable system size and showcases the delicate balance between arrival and service rates in shaping system dynamics. Understanding this interplay is essential for optimizing system performance.

Figures 9 and 10 depict how the system throughput and Total Energy Cost (TEC) increase as the service rate μ rises across various arrival rates λ. Figures 11 and 12 illustrate the effects of increasing both arrival and service rates on Total Energy Cost (TEC) and Total Expected Revenue (TER). An increase in the arrival rate λ leads to a rise in both the TEC

and TER. In contrast, increasing the service rate μ results in a decrease in both the TEC and TER. These results highlight the complex relationship between input and service dynamics and their implications for cost and performance optimization.

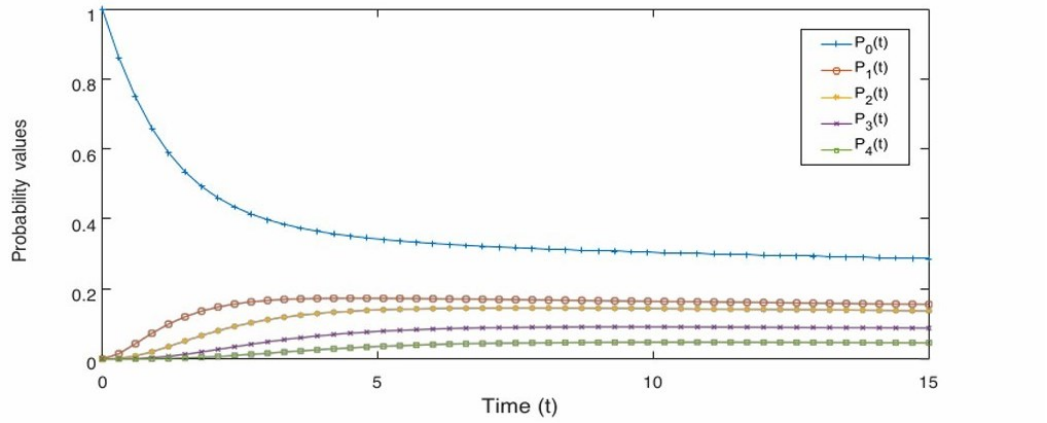

Figure 2. System-size probabilities of the offline customers in the transient state $P_n(t)$.

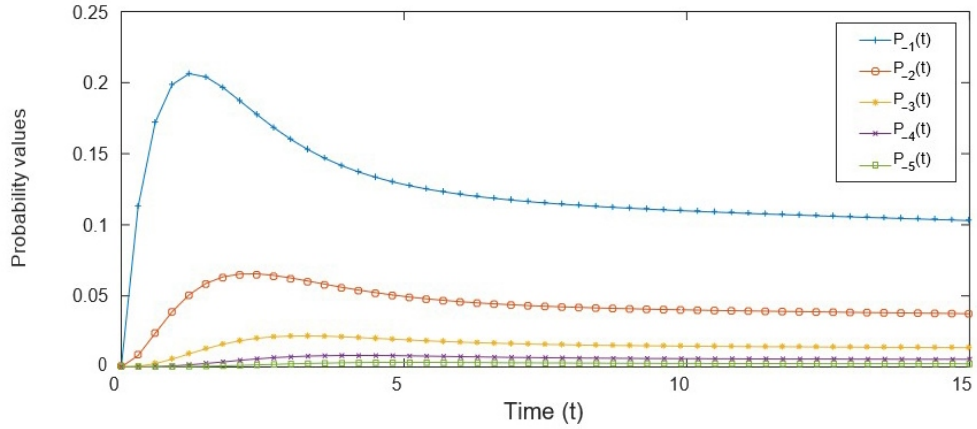

Figure 3. System-size probabilities of the online customers in the transient state $P_{-n}(t)$.

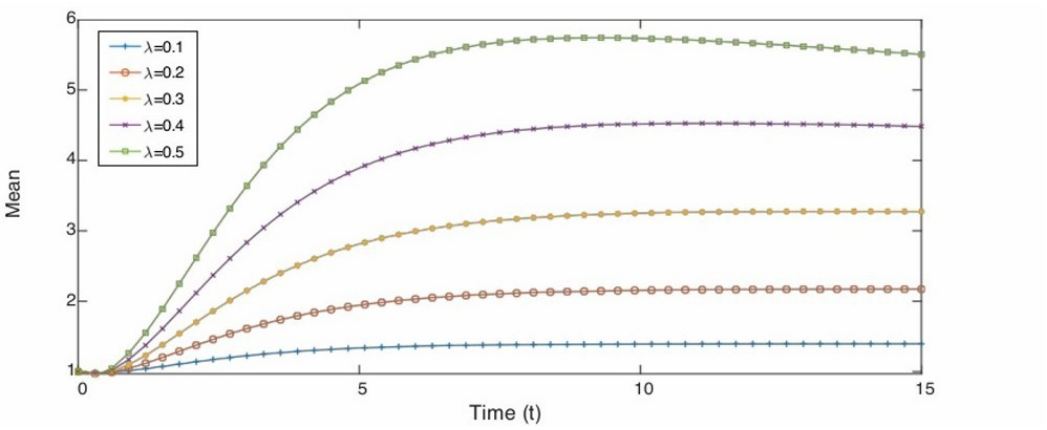

Figure 4. Mean system size in transient state.

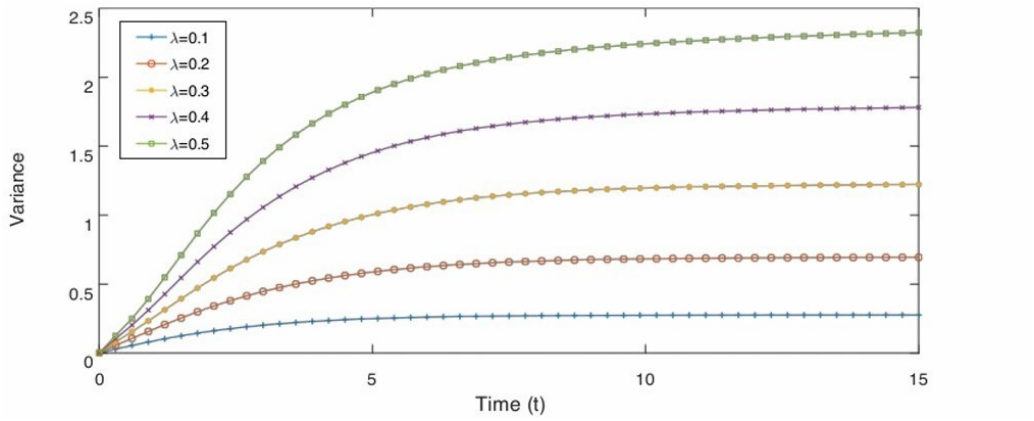

Figure 5. Variance of the system size in transient state.

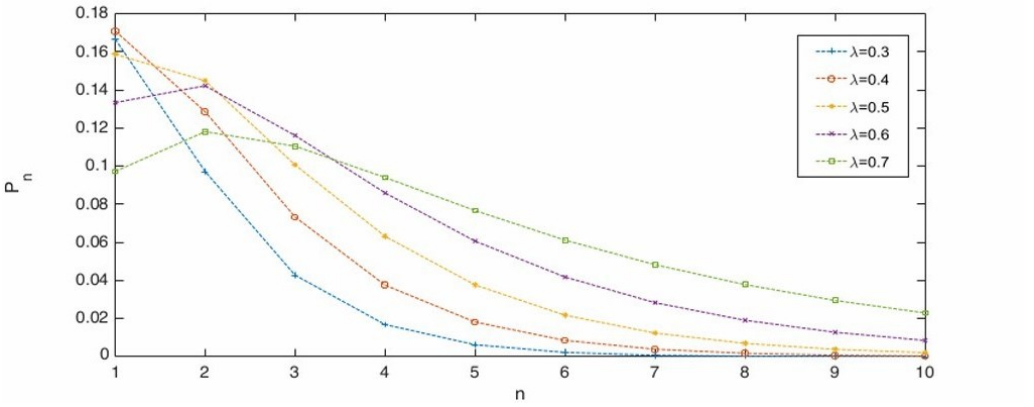

Figure 6. System-size probabilities of the offline customers in the steady state P_n.

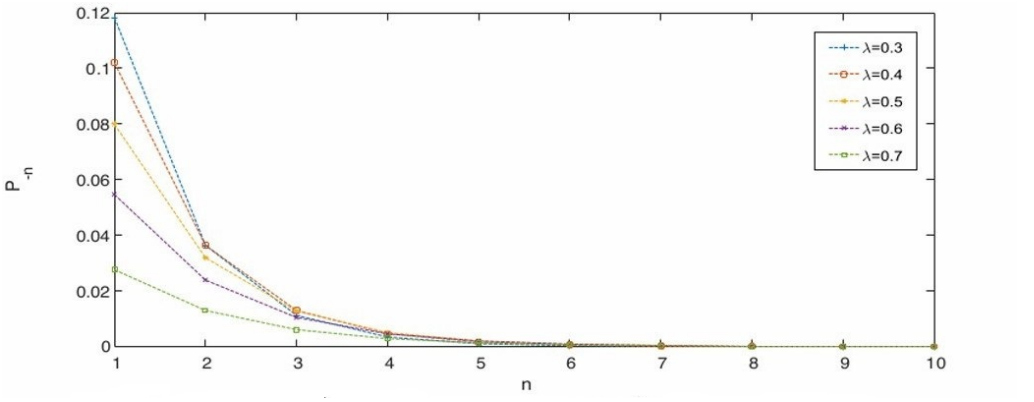

Figure 7. System-size probabilities of the online customers in the steady state P_{-n}.

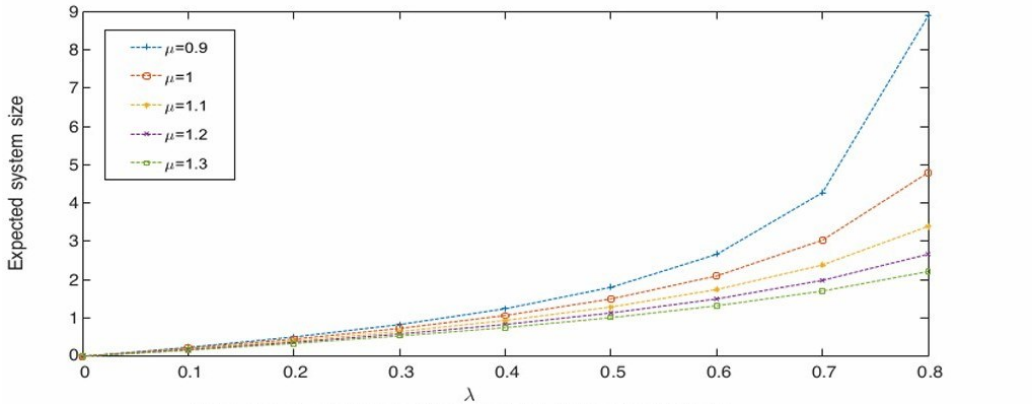

Figure 8. Mean system size in the steady state.

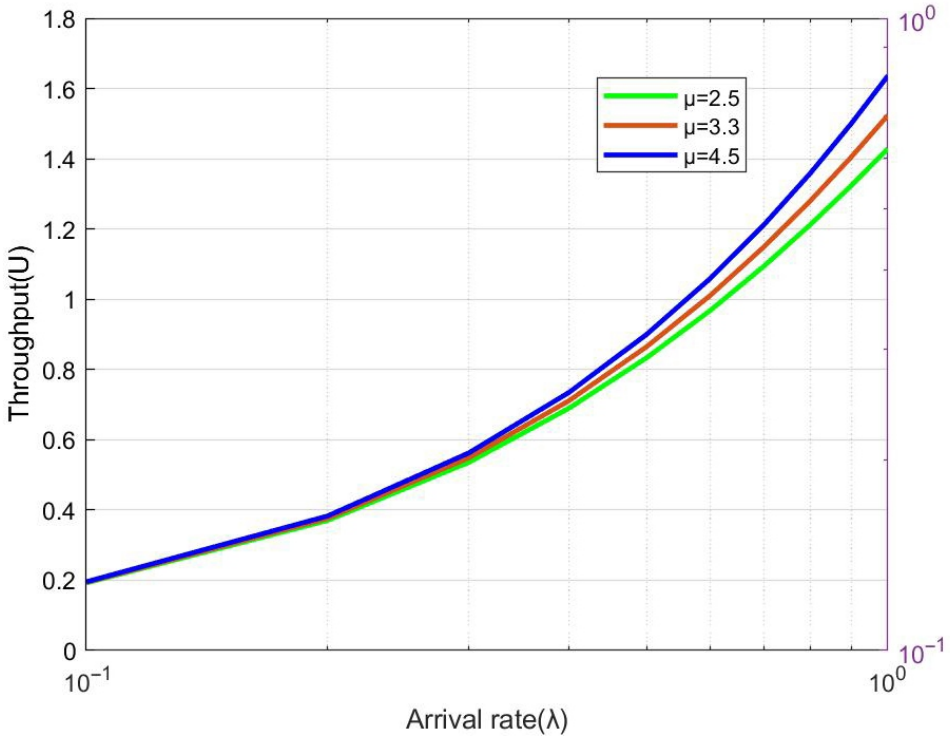

Figure 9. Throughput versus arrival rate (λ).

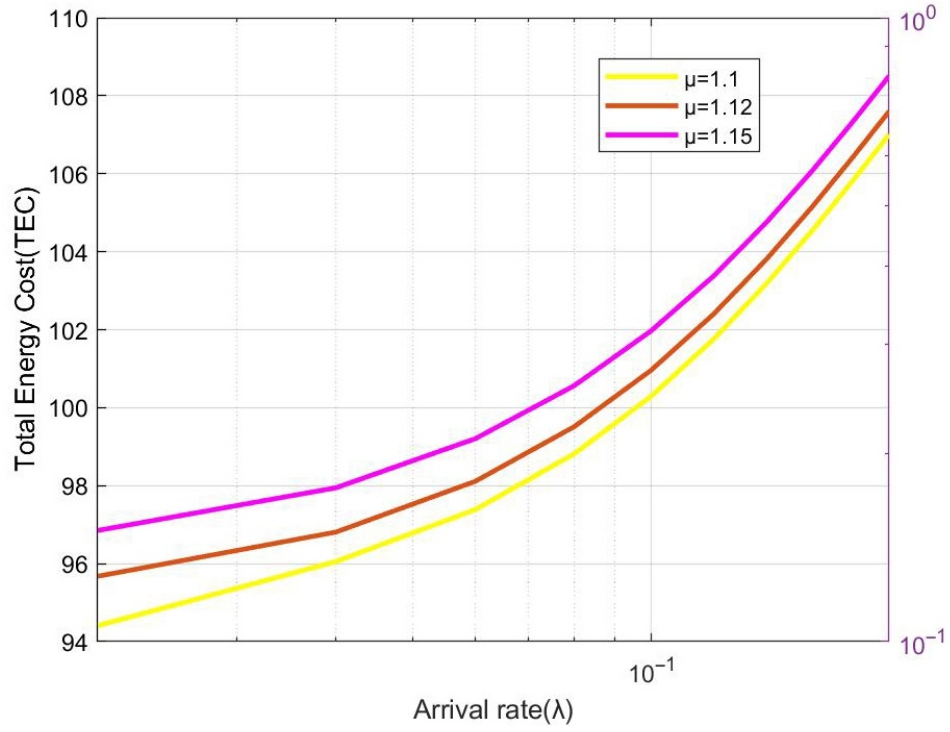

Figure 10. Total Energy Cost versus arrival rate.

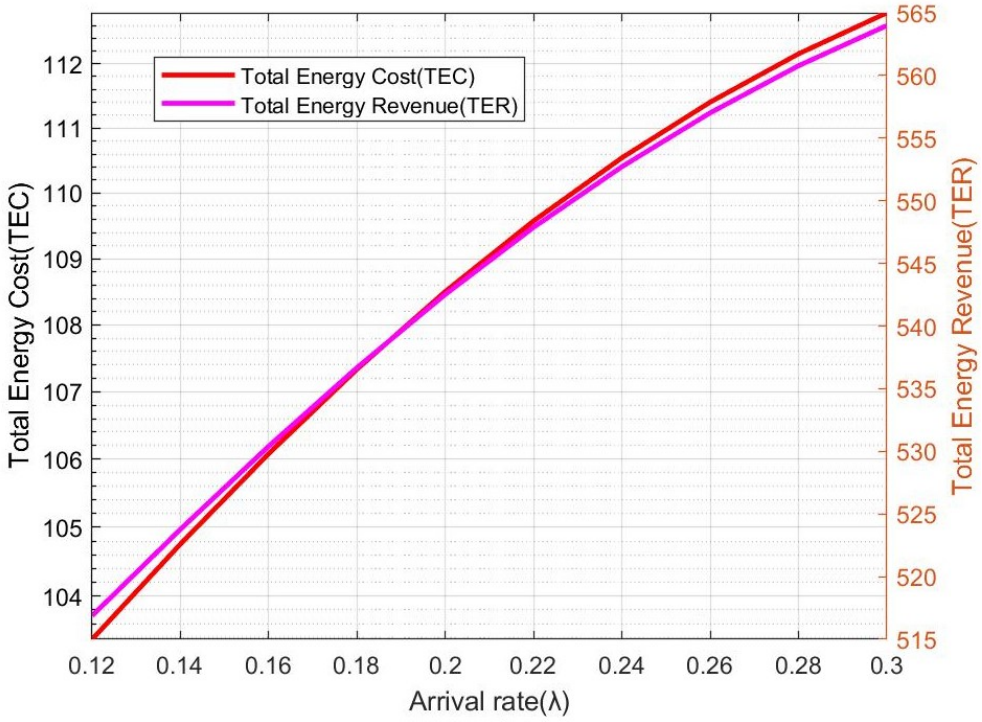

Figure 11. Total Energy Cost and Total Energy Revenue versus arrival rate.

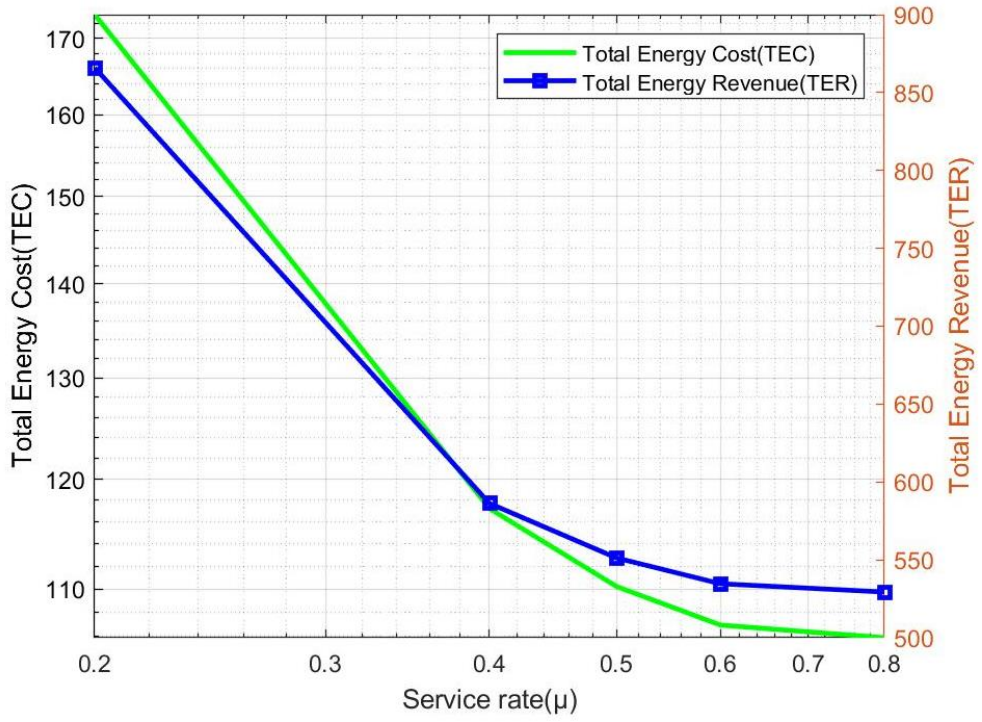

Figure 12. Total Energy Cost and Total Energy Revenue versus service rate

10. Conclusions

This study analyzes a single-server, continuous-service Markovian queueing system with infinite capacity, designed to accommodate two distinct customer classes: offline and online. By employing generating functions, Modified Bessel functions, and Laplace transform techniques, we derive compact analytical expressions for both transient and steady-state system size probabilities. The model effectively captures the system's dynamics

under a non-preemptive priority service discipline, where offline customers are always prioritized over online customers. Comprehensive performance measures, including the mean and variance of system size, throughput, and cost-related metrics, are evaluated in both transient and steady states. Numerical illustrations validate the analytical findings and demonstrate the influence of arrival and service rates on key system parameters such as system size, waiting times, throughput, and energy costs. The proposed model offers a robust framework for analyzing dual-source service systems with continuous operation requirements, making it especially relevant to modern applications in cloud computing, healthcare, and automated service environments. Future work may extend this model to multi-server systems, incorporate customer abandonment or retrial behavior, and explore the effects of non-exponential service time distributions. Such extensions would further enhance the applicability of the model to more complex and realistic service systems.

Author Contributions: Conceptualization, R.S., P.B. and R.D.S.P.; Methodology, R.S., P.B. and R.D.S.P.; Software, P.B. and R.D.S.P.; Validation, R.S., P.B. and R.D.S.P.; Formal analysis, R.D.S.P.; Investigation, R.S. and R.D.S.P.; Resources, R.S., P.B. and R.D.S.P.; Data curation, P.B. and R.D.S.P.; Writing—original draft, P.B. and R.D.S.P.; Writing—review & editing, R.S.; Visualization, R.S.; Supervision, R.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

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

References

1. Green, L.V. Queueing analysis in healthcare. In *Patient Flow: Reducing Delay in Healthcare Delivery*; Hall, R.W., Ed.; Springer: New York, NY, USA, 2006; pp. 281–308.
2. Worthington, D.J. Queueing models for hospital waiting lists. *J. Oper. Res. Soc.* **1987**, *38*, 413–422. [CrossRef] [PubMed]
3. Boxma, O.J.; David, I.; Perry, D.; Stadje, W. A new look at organ transplantation models and double matching queues. *Probab. Eng. Inf. Sci.* **2011**, *25*, 135–155. [CrossRef]
4. Zenios, S.A.; Chertow, G.M.; Wein, L.M. Dynamic Allocation of Kidneys to Candidates on the Transplant Waiting List. *Oper. Res.* **2000**, *48*, 549–569. [CrossRef]
5. Buzacott, J.A.; Shanthikumar, J.G. *Stochastic Models of Manufacturing Systems*; Prentice Hall: Upper Saddle River, NJ, USA, 1993.
6. Gershwin, S.B. *Manufacturing Systems Engineering*; Prentice Hall: Upper Saddle River, NJ, USA, 1994.
7. Cheng, M. Sharing economy: A review and agenda for future research. *Int. J. Hosp. Manag.* **2016**, *57*, 60–70. [CrossRef]
8. Liu, H.L.; Li, Q.L.; Wu, X.; Zhang, C. Two basic queueing models of service platforms in digital sharing economy. *arXiv* **2021**, arXiv:2108.02852.
9. Sutherland, W.; Jarrahi, M.H. The sharing economy and digital platforms: A review and research agenda. *Int. J. Inf. Manag.* **2018**, *43*, 328–341. [CrossRef]
10. Gelenbe, E. Product-form queueing networks with negative and positive customer. *J. Appl. Probab.* **1991**, *28*, 656–663. [CrossRef]
11. Parthasarathy, P.; Vasudevan, K. A single server queue with two types of customers and a gate mechanism. *Am. J. Math. Manag. Sci.* **2010**, *30*, 331–345. [CrossRef]
12. Choudhury, G.; Deka, M. A single server queueing system with two phases of service subject to server breakdown and Bernoulli vacation. *Appl. Math. Model.* **2012**, *36*, 6050–6060. [CrossRef]
13. Jaiswal, N.K. *Priority Queues*; Academic Press: New York, NY, USA, 1968.
14. Klimenok, V.; Dudin, A.; Dudina, O.; Kochetkova, I. Queueing System with Two Types of Customers and Dynamic Change of a Priority. *Mathematics* **2020**, *8*, 824. [CrossRef]
15. Kim, B.; Kim, J. Waiting time distributions in an M/G/1 retrial queue with two classes of customers. *Ann. Oper. Res.* **2017**, *252*, 121–134. [CrossRef]
16. Baron, O.; Berman, O.; Krass, D.; Sherzer, E. Re-direction in queueing networks with two customer types: The inter-departure analysis. *Comput. Oper. Res.* **2025**, *173*, 106867. [CrossRef]
17. Tikhonenko, O.; Ziółkowski, M. Single-server queueing system with external and internal customers. *Bull. Pol. Acad. Sci. Tech. Sci.* **2018**, *66*, 4.

18. Hanukov, G. A queueing-inventory model with skeptical and trusting customers. *Ann. Oper. Res.* **2023**, *331*, 763–786. [CrossRef]
19. Li, L.; Xu, W.; Wang, Z.; Liu, L. Improving efficiency of the queueing system with two types of customers by service decomposition. *AIMS Math.* **2023**, *8*, 25382–25408. [CrossRef]
20. Zhang, Y.; Xu, X.; Zhao, P.; Liu, M. Equilibrium analysis of the fluid model with two types of parallel customers and incomplete fault. *Probab. Eng. Informational Sci.* **2024**, *38*, 539–558. [CrossRef]
21. Parthasarathy, P.R.; Selvaraju, N.; Manimaran, G. A paired queueing system arising in multimedia synchronization. *Math. Comput. Model.* **1999**, *30*, 133–140. [CrossRef]
22. Sudhesh, R.; Shapique, M.; Dharmaraja, S. Analysis of a Multiple Dual-Stage Vacation Queueing System with Disaster and Repairable Server. *Methodol. Comput. Appl. Probab.* **2022**, *24*, 2485–2508.
23. Latouche, G. Queues with paired customers. *J. Appl. Probab.* **1981**, *18*, 684–696. [CrossRef]
24. Sudhesh, R.; Sebasthi Priya, R.; Lenin, R.B. Transient Analysis of a Single Server Discrete-Time Queue with System Disaster. *RAIRO-Oper. Res.* **2017**, *51*, 123–134. [CrossRef]
25. Chew, S. Continuous-Service $M/M/1$ queueing Systems. *Appl. Syst. Innov.* **2019**, *2*, 16. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.