

Article

Optimizing Priority Queuing Systems with Server Reservation and Temporal Blocking for Cognitive Radio Networks

Jnana Ranjan Behera ¹, Agbotiname Lucky Imoize ^{2,*} , Sudhansu Sekhar Singh ¹, Subhranshu Sekhar Tripathy ³  and Sujit Bebertta ⁴ 

¹ School of Electronics, KIIT Deemed to be University, Bhubaneswar 751024, India; jnanaranjan07@gmail.com (J.R.B.); ssinghket@kiit.ac.in (S.S.S.)

² Department of Electrical and Electronics Engineering, Faculty of Engineering, University of Lagos, Akoka, Lagos 100213, Nigeria

³ School of Computer Engineering, KIIT Deemed to be University, Bhubaneswar 751024, India; subhranshu.008@gmail.com

⁴ Department of Computer Science, Ravenshaw University, Cuttack 753003, India; sujitlebertta1@gmail.com

* Correspondence: aimoize@unilag.edu.ng

Abstract: In the domain of cognitive radio (CR), unlicensed users have the opportunity to efficiently use available spectrum bands without interfering with licensed primary users (PUs). Our study addresses the challenge of secondary user (SU) spectrum shortage due to high arrival rates of licensed users. We propose two models aimed at improving the average total waiting time for SUs in such scenarios. These models incorporate non-acquired and preemptive priority mechanisms within the M/D/1 model of a PU delay system. Through quantitative evaluations and Monte Carlo simulations, we evaluate the performance of these models. Our findings show significant improvements in average waiting time for both PUs and SUs, especially under the priority scheme. Furthermore, we explore these models in the context of real-time systems, considering the limited buffer capacity for both user types. This further improves the average waiting time for PUs and SUs in both priority schemes. Our contribution lies in providing effective solutions to mitigate SU shortages in CR networks, providing insight into priority-based approaches and real-time system considerations.

Keywords: cognitive radio networks; admission control; spectrum allocation; user authorization; priority queuing



Citation: Behera, J.R.; Imoize, A.L.; Singh, S.S.; Tripathy, S.S.; Bebertta, S. Optimizing Priority Queuing Systems with Server Reservation and Temporal Blocking for Cognitive Radio Networks. *Telecom* **2024**, *5*, 416–432. <https://doi.org/10.3390/telecom5020021>

Academic Editor: Sotirios K. Goudos

Received: 17 March 2024

Revised: 18 May 2024

Accepted: 28 May 2024

Published: 31 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The widespread adoption of smartphones, tablets, and other types of digital devices has led to an enormous surge in traffic on the internet in the past decade. Consequently, there is a need to effectively use the wireless spectrum due to a lack of more wireless resources and resources on the radio spectrum. Cognitive radio is a promising solution for addressing the issue of bandwidth scarcity. Unauthorized customers may utilize previously reserved frequencies for authorized individuals thanks to cognitive wireless technology. Cognitive radio technologies are classified into two types: frequency-sharing and heterogeneity. Numerous conceivable variations of the transfer of data in cognitive radio networks (CRNs) necessitate taking into account various queueing systems as their descriptors; for an example, refer to [1]. Here, we apply the resulting assumptions regarding the CRN's operational circumstances. It is assumed that the bandwidth of each network cell is split into identical sub-bands or channels, each available for usage by high-priority users (HPUs) and low-priority users (LPUs). There might be more than one channel. The single-server example widely discussed in the literature barely applies to CRNs. There is no waiting buffer in the unlikely event that all available bandwidth is used [2]. Spectrum measurement is perfect because LPUs can precisely measure or predict how much of a channel is occupied by HPUS or other LPUs. The so-called overlay mode is used by the

cell. Under these conditions, if an HPU arrives while every channel is occupied, one LPU's service must be stopped. It involves a spectrum handoff wherein an LPU can quickly settle an additional idle channel if it becomes available after an HPU reclaims its track. If the inactive channel is unavailable, the LPU may attempt to join the service later, at random.

1.1. Motivations

It is essential to emphasize the critical role of efficient spectrum utilization in CRNs, where unlicensed users seek to leverage unused spectrum bands without causing interference to licensed PUs. As the arrival rate of licensed users increases, the potential for SU deprivation grows, necessitating robust mechanisms to mitigate waiting-time disparities. The proposed models, integrating M/D/1 queuing frameworks with both non-preemptive and preemptive priority mechanisms, aim to address this challenge by optimizing the average total waiting time for both PUs and SUs. By contrasting these models with an M/D/1 system incorporating PU delay, the study seeks to quantify the performance improvements achieved through priority-based scheduling strategies. Through quantitative evaluations and Monte Carlo simulations, the efficacy of the proposed models is rigorously assessed, demonstrating significant enhancements in waiting-time metrics under both non-preemptive and preemptive priority schemes. Moreover, by extending the analysis to real-time systems with finite buffer capacities, the study underscores the practical relevance of the proposed approaches to scenarios wherein resource constraints must be considered. Overall, the study not only contributes to advancing the theoretical foundations of queuing theory in cognitive radio networks, but also offers practical insights for optimizing system performance in real-world deployment scenarios, thereby facilitating more efficient and equitable spectrum utilization for all users.

Apart from the previously mentioned hypotheses, we propose that there exists a possibility in this cell to disseminate the bandwidth condition to every user. There are two outcomes for this status: either the bandwidth is open for LPUs to sense, or it is temporarily closed off (blocked) to LPUs. This proposal is unrestrictive for radio access networks.

1.2. Contributions

The contributions of the present article are as follows:

- The study introduces two priority models, non-pre-emptive and pre-emptive, within the M/D/1 framework to address SU starvation in CRNs with a high PU arrival rate.
- The research compares the proposed non-pre-emptive and pre-emptive priority models and an M/D/1 model with PU delay to determine which improves the average total waiting time for PUs and SUs.
- Numerical analysis and Monte Carlo simulations are used to analyze the suggested models' impact on waiting times in different scenarios, helping to elucidate system dynamics.
- To model real-time systems, the research extends the proposed models to include finite buffer capacity for both PUs and SUs, leading to significant improvements in average total waiting time with non-pre-emptive and pre-emptive priority.
- The suggested models improve average total waiting time for PUs and SUs, providing useful insights for priority-scheme design and optimization in CR networks with high PU arrival rates.

The structure of the paper is as follows. Section 2 provides a comparison of related studies. The dynamics of the system are developed in Section 3. Section 4 presents the performance analysis of the proposed algorithm in terms of the convergence speed and time complexity. Section 5 deals with the numerical results achieved for the study. In Section 6, a comparative analysis of major key performance metrics obtained for the proposed system and the existing studies is presented in a comprehensive manner. Section 7 discusses the challenges and limitations of the present study. Finally, a summary of the study is found in Section 8, along with potential avenues for the proposed model's generalization.

2. Related Work

The retrial phenomena described in the retrial-queue literature bear similarities to the sensing process. If a client entering a retry queue cannot locate an available server, they will be barred and will need to try again later. Clients who are blocked enter the orbit, a line. The M/G/1/1 retrial queue was analyzed by Keilson et al. [3] to determine the system's efficiency metrics. Falin [4] established a limiting hypothesis for the waiting period experienced by consumers during heavy traffic, whereas Falin [5] developed the waiting-time distribution of customers in orbit. Falin [6] calculated the average, variance, and probability distribution based on the total number of trials. The models we use in this study feature two classes of users, PUs and SUs, but there is only one class of users in these models. While accessing a server, SUs must perceive, and when a PU arrives, an SU may be interrupted in transmission. Correlated retrial queuing architectures that feature two-way interaction with two customer classes without disruptions were showcased by Artalejo and Phung-Duc [7]. A queuing framework for cognitive radio networks with reactive-decision spectrum handoff was put forward by Salameh et al. [8].

The blocking probability, the average number of sensing SUs, and the transmission time are among the outcome metrics that the authors calculated by restricting the number of SUs that can feel simultaneously (i.e., by considering a finite orbit size). According to Salameh et al. [8], an SU keeps detecting servers one by one until it locates one that is idle. By removing assumptions and opening up protocols and interfaces among various elements, this multi-vendor, interoperable technology enables the incorporation of intelligence into RAN to support a variety of implementation settings [9,10]. In keeping with the current trend, this study introduces a federation layer to an O-RAN architecture to facilitate effective admission management, dynamic traffic forecasting, and service monitoring. A weighted-sum-based multi-objective optimization framework for resource allocation in cloud systems was suggested by Shrimali et al. [11]. This framework utilized the fuzzy-logic method to provide the coefficients for the specified objectives. The adaptive algorithm uses the coefficients above to produce Pareto-optimal alternatives. A combination of techniques that employs a multi-point algorithm in heterogeneous networks to determine the best RAT selection was put forward by Goudarzi et al. [12]. The scheme was implemented by applying biogeography-based optimization (BBO) to RAT selection probabilities generated by a Markov decision process (MDP). The study [13] proposed an M/D/1 model of a PU delay system with both non-preemptive and preemptive priority. The average waiting time for PU was 18% and that for SU was 32% under the non-preemptive scheme in their proposed approach. Further, the waiting times for PUs and SUs were observed to be 22% and 7%, respectively, with the preemptive approach. A comprehensive summary showing the comparison between the above studies is discussed in Table 1 below.

Table 1. Comparative of relevant studies surveyed.

Authors	Models Used	Contributions	Limitations
Keilson et al. [3]	M/G/1/1 retrial queue	Analysis of system efficiency metrics	Focus on single-class user models
Falin [4]	M/G/1/1 retrial queue	Establishes limiting hypothesis for waiting period during heavy traffic	Single-class user models
Falin [5]	M/G/1/1 retrial queue	Develops waiting-time distribution of customers in orbit	Single-class user models
Falin [6]	M/G/1/1 retrial queue	Calculates average, variance, and probability distribution based on total number of trials	Single-class user models

Table 1. Cont.

Authors	Models Used	Contributions	Limitations
Artalejo and Phung-Duc [7]	Correlated retry queueing architectures	Showcases two-way interaction with two customer classes without disruptions	N/A
Salameh et al. [8]	Queueing framework for cognitive radio networks with reactive-decision spectrum handoff	Introduces blocking probability, average number of sensing SUs, and transmission time metrics	Finite orbit size assumption
Shrimali et al. [11]	Multi-objective optimization framework for resource allocation in cloud systems	Proposes weighted-sum-based multi-objective optimization framework utilizing fuzzy logic	-
Goudarzi et al. [12]	Multi-point algorithm in heterogeneous networks for RAT selection	Introduces combination of techniques for best RAT selection using biogeography-based optimization	Implementation scheme focused on RAT selection probabilities from MDP
Proposed framework	Priority-based M/D/1 queue (with preemption and non-preemption)	Effective in mitigating SU spectrum outage in CRNs. Provides insight into priority-based admission of requests. Consideration of real-time system constraints.	Scalability with increasing users and implementation complexity.

The studies focus on CRNs that aim to improve PU and SU management via admission control or queueing models. This motivated us to devise an effective admission-control mechanism to assure PU area coverage, while the SUs utilize spectrum resources fairly and efficiently. The M/D/1 queue strategy proposed in this study is promising for prioritizing and admission control. The proposed strategy may increase spectrum consumption by PUs with priority or with secondary priorities. In CRNs, it will reduce latency and boost performance.

3. Proposed Framework

The queueing theory is a useful analytical tool that can be utilized in the process of designing and optimizing admission-control systems in cognitive radio networks (CRNs) [13]. In the context of CRNs, where dynamic spectrum access and effective resource usage are of the utmost importance, queueing theory makes it possible to conduct a systematic examination of the queueing dynamics, service rates, and user arrivals in the network [14]. It is possible to conduct an analysis of the performance indicators that are essential for admission control by modeling the network as a queueing system [15]. These metrics include queue lengths, waiting times, and levels of system utilization. This technique makes it possible to gain a thorough understanding of how admission decisions affect network performance. This, in turn, makes it possible to build effective strategies that optimize spectrum usage, decrease interference, and assure efficient exploitation of available resources in cognitive radio environments. Overall, the utilization of queueing theory in admission control for cognitive radio networks makes a contribution to the improvement of overall network efficiency and the fulfillment of quality-of-service standards [16].

In the workflow of a cognitive radio network, requests from both PUs and SUs are managed by the server to optimize spectrum utilization and minimize interference, as presented in Figure 1. Upon arrival, requests are queued and prioritized, with PUs given precedence over SUs to access spectrum bands for communication [17]. Once PU requests have been serviced, the server addresses SUs, allowing them to exploit the unused spectrum opportunistically. Responses indicating request status are provided to both PUs and SUs, facilitating efficient

resource allocation and communication. Continuous monitoring and adaptive adjustments ensure responsive handling of requests, optimizing system performance and user satisfaction. This workflow underscores the dynamic nature of spectrum management in cognitive radio networks, which balances the needs of PUs and SUs while maintaining efficient spectrum utilization and minimizing interference. Table 2 provides a comprehensive summary of the important mathematical notations used in the manuscript.

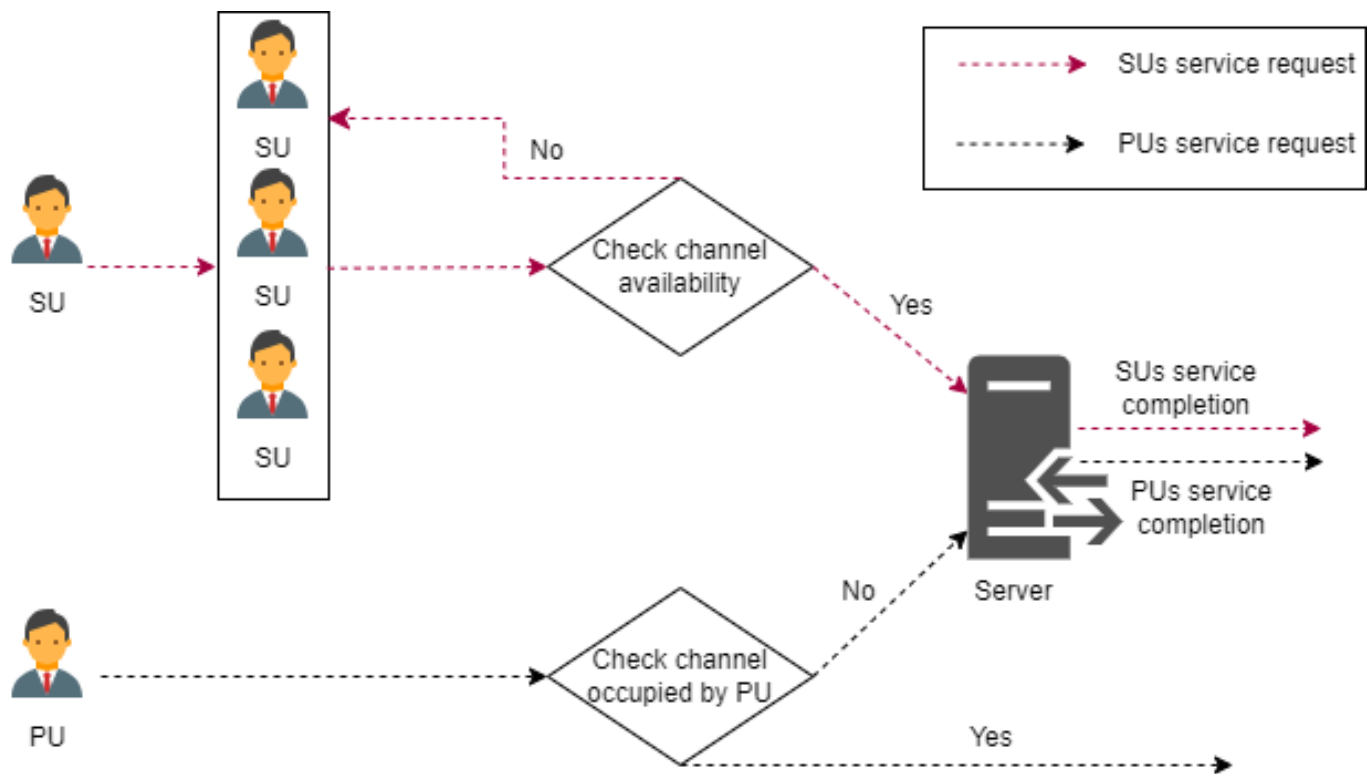


Figure 1. Workflow in the proposed CRN architecture among PUs and SUs.

Table 2. Description of mathematical notations.

Symbols	Description
P	Primary users
S	Secondary users
k	Users in the system
q	The system queue of CRN server
λ_P	Arrival rate of PUs
λ_S	Arrival rate of SUs
μ_P	Service rate of PUs
μ_S	Service rate of SUs
N_P	Number of PUs in the system
N_S	Number of SUs in the system
N	Total number of slots in the system
W_k	Waiting time of packets in system
W_P^q	Average waiting time in queue for PU delay
$W_P^{q(non-premp)}$	Average waiting time in queue for non-preemptive PU delay
W_S^q	Average waiting time in queue for SU delay
$W_S^{q(non-premp)}$	Average waiting time in queue for non-preemptive SU delay
ρ_P	Utilization factor for PU delay
ρ_S	Utilization factor for SU delay

The performance of the various proposed methods is assessed by examining the average total waiting times for both the PU and the SU under different scenarios. Initially, performance was assessed for a delay model based on PUs that has already been developed. The study in [4] focuses on analyzing the priority queueing models to improve the Quality of Service (QoS) for the SU. These methods, which have an infinite buffer size, are suggested for the purpose of performance assessment. Subsequently, identical designs are proposed, albeit with a limited buffer capacity [18,19]. The M/D/1 model is considered for both the PU queue and the SU queue in all schemes. It should be noted that the average overall waiting time for the packet in the system, denoted as W_k , is equal to the sum of the average service time of the user, denoted as X_k , and the average waiting time for the user in the queue, denoted as W_k^q , as follows:

$$W_k = \overline{X_k} + W_k^q \quad (1)$$

Delay by PUs

The study in [4] utilizes the M/D/1 priority queueing model and incorporates a PU delay, a term that refers to the number of time slots that are postponed in order to transmit PU packets. This delay enhances the SU waiting time and prevents famine. The duration of the delay will fluctuate based on the frequency of delay requests and the specific purpose for which this system is employed [20,21]. This approach incorporates a time delay, T_D , prior to the commencement of a time slot [4]. The study in [4] utilizes the M/D/1 priority queueing model and incorporates a PU delay, a term that refers to the number of time slots that are postponed for the transmission of the PU's packets. This delay improves the SU waiting time and prevents hunger. The duration of the delay will be contingent upon the frequency of delay requests and the specific purpose for which this scheme is employed. This approach incorporates a time delay, T_D , prior to the commencement of a time slot [4].

(i). Scheme 1: M/D/1 with PU delay and non-pre-emptive priority model

The model described in Section 3.1 of Scheme 1, which incorporates PU delay [4], is further improved by incorporating non-preemptive priority [22]. The PU packet experiences a delay equal to a specific number of time slots (PU delay, $d = 1$). Consequently, the PU packet is unable to disrupt the transmission in the presence of a served SU packet. We assume an unlimited buffer and zero delay time. Despite the absence of a SU in the queue, the PU packet experiences a delay and is subsequently transmitted in the following time slot. T_D , the delay time till the start of a time slot, is not accounted for in this approach.

In Scheme 1, the average total waiting time of the tagged PU packet encompasses the average service time as well as the service times of the current PU and SU (if there is one) [23]. The system that does not prioritize or interrupt ongoing tasks. The formula for calculating the average total waiting time, given that the PU delay is i (an integer) timeslots, was developed as follows:

$$W_P^q(d) = \frac{1}{2} \left(\frac{\lambda_P}{\mu_P^2} + \frac{\lambda_S}{\mu_S^2} \right) + \frac{N_P^q}{\mu_P} + \frac{i}{\mu_S}, \forall d = i \quad (2)$$

Next, employing Little's hypothesis, N^q is equal to λW^q , as follows:

$$W_P^{q(non-premp)}(d) = \frac{1}{2} \left(\frac{\frac{\lambda_P}{\mu_P^2} + \frac{\lambda_S}{\mu_S^2} + \frac{i}{\mu_S}}{1 - \rho_P} \right) \quad (3)$$

The equation representing the average total waiting time of the processing unit in Scheme 1 is displayed below:

$$W_P^{q(non-premp)}(d) = X_P + \frac{1}{2} \left(\frac{\frac{\lambda_P}{\mu_P^2} + \frac{\lambda_S}{\mu_S^2} + \frac{i}{\mu_S}}{1 - \rho_P} \right) \quad (4)$$

The average cumulative waiting time of the labeled SU packet is thereafter indicated. The calculation includes the average service time, as well as the service time of the identified SU packet and the combined service times of the PU and SU. The future arrivals at the PU line have no influence on it.

$$W_S^q(d) = \frac{1}{2} \left(\frac{\lambda_P}{\mu_P^2} + \frac{\lambda_S}{\mu_S^2} \right) + \frac{N_P^q}{\mu_P} + \frac{N_S^q}{\mu_S}, \forall d = i \quad (5)$$

By applying Little's theorem, this can be deduced as demonstrated below.

$$W_S^{q(non-premp)}(d) = \frac{1}{2} \left(\frac{\frac{\lambda_P}{\mu_P^2} + \frac{\lambda_S}{\mu_S^2}}{(1 - \rho_P)(1 - \rho_S)} \right) \quad (6)$$

Ultimately, when the average service time of SU, $\overline{X_S}$, is included, the resulting statement is as follows:

$$W_S^{(non-premp)}(d) = \overline{X_S} + W_S^{q(non-premp)}(d) \quad (7)$$

In this work, the following two equations are derived for simulation purposes, assuming that the service time, denoted as X_S , is equal to one time slot for both the service and processing times.

$$W_P^{(non-premp)}(d) = 1 + \frac{1}{2} \left(\frac{\frac{\lambda_P}{\mu_P^2} + \frac{\lambda_S}{\mu_S^2} + \frac{i}{\mu_S}}{1 - \rho_P} \right) \quad (8)$$

$$W_S^{(non-premp)}(d) = 1 + \frac{1}{2} \left(\frac{\frac{\lambda_P}{\mu_P^2} + \frac{\lambda_S}{\mu_S^2}}{(1 - \rho_P)(1 - \rho_S)} \right) \quad (9)$$

(ii). *Scheme 2: M/D/1 with PU delay and pre-emptive priority model*

Scheme 2 aims to enhance the average total waiting time of the PU by sacrificing the QoS of the SU [24,25]. This is because the performance of the SU is influenced by the arrival rate of the PU.

This model proposes the combination of the pre-emptive priority system with the approach described in reference [4]. The PU packet will have a delay for a designated number of time intervals. However, following the PU delay, if there is an SU packet being processed, the PU packet has the ability to interrupt the transmission in order to receive its own service. However, the timing of the packets' arrival is a determining factor [26]. The PU packet experiences a delay of one unit of time. Infinite buffer capacity is assumed, as for the non-pre-emptive priority with the PU delay scheme. In addition, this scheme does not incorporate the time delay, TD, prior to the start of a time slot, which differentiates it from the scheme in [4].

The average total waiting time of the tagged PU packet is determined by the sum of the mean service time and the combined service times of both PU and SU packets in the system. Furthermore, it experiences a delay of i occurrences prior to transmission, resulting in the following equation:

$$W_P^{q(premp)}(d) = \frac{1}{2} \left(\frac{\lambda_P}{\mu_P^2} \right) + \frac{N_P^q}{\mu_P} + \frac{i}{\mu_S}, \forall d = i \quad (10)$$

By applying Little's theorem, we derive the following equation:

$$W_P^{q(prem)}(d) = \frac{\frac{1}{2} \left(\frac{\lambda_P}{\mu_P^2} \right) + \frac{N_P^q}{\mu_P} + \frac{i}{\mu_S}}{1 - \rho_P} \quad (11)$$

Finally, an expression describing the average total waiting time for each PU packet is derived.

$$W_P^{(premp)}(d) = c_p + W_P^{q(prem)}(d) \quad (12)$$

where c_p represents the average time it takes to complete a task.

The mean waiting time in the system for the tagged SU packet is determined by considering the mean service time, the average service times of all PUs and SUs present, and the service time of the marked packet. The outcome of this leads to the following mathematical expression:

$$W_S^{q(prem)}(d) = \frac{1}{2} \left(\frac{\lambda_P}{\mu_P^2} + \frac{\lambda_S}{\mu_S^2} \right) + \frac{N_P^q}{\mu_P} + \frac{N_S^q}{\mu_S}, \forall d = i \quad (13)$$

When Little's theorem is utilized, it is expressed as follows:

$$W_S^{q(prem)}(d) = \frac{1}{2} \left(\frac{\frac{\lambda_P}{\mu_P^2} + \frac{\lambda_S}{\mu_S^2}}{(1 - \rho_P)(1 - \rho_S)} \right) \quad (14)$$

Finally, the derivation concludes with the expression for the average total waiting time of the SU.

$$W_S^{(premp)}(d) = c_S + W_S^{q(prem)}(d) \quad (15)$$

The mean completion time, denoted as c_S , is represented by [13].

The mean cumulative waiting time for SU is provided as follows:

$$W_S^{(premp)}(d) = \frac{\overline{X}_S}{(1 - \rho_P)} + \frac{1}{2} \left(\frac{\frac{\lambda_P}{\mu_P^2} + \frac{\lambda_S}{\mu_S^2}}{(1 - \rho_P)(1 - \rho_S)} \right) \quad (16)$$

Next, the assumption that there is only one slot for service time is utilized, and the equations for calculating the average total waiting time for both PU and SU are produced.

$$W_P^{(premp)}(d) = 1 + \frac{1}{2} \left(\frac{\frac{\lambda_P}{\mu_P^2} + \frac{i}{\mu_S}}{1 - \rho_P} \right) \quad (17)$$

$$W_S^{(premp)}(d) = \frac{1}{(1 - \rho_P)} + \left(\frac{\frac{1}{2} \left(\frac{\lambda_P}{\mu_P^2} + \frac{\lambda_S}{\mu_S^2} \right)}{(1 - \rho_P)(1 - \rho_S)} \right) \quad (18)$$

The pseudocode presented below in Algorithm 1 describes the mechanism of user prioritization in CRN, where the capacity and service rate are the major factors that are considered. It can manage arrivals and departures, along with the capacity limitations at various levels so that the whole system is leveraged to the maximum degree while maintaining fairness between the PU and SUs.

Algorithm 1: Algorithmic workflow of the proposed framework.

```

1. Initialize system parameters:  $\lambda_P, \lambda_S, \mu_P, \mu_S, K$ 
2. Initialize system state:
3.  $W_S^q \rightarrow$  Set waiting time of SUs to 0
4.  $W_P^q \rightarrow$  Set waiting time of PUs to 0
5. For  $i = 1 : 10000$  do
6.     Generate random number of arrivals for PUs according to Poisson distribution with rate  $\lambda_P$ 
7.     Generate random number of arrivals for SUs according to Poisson distribution with rate  $\lambda_S$ 
8.     Update the current number of PUs and SUs in the system:
9.     If  $K = FULL$ 
10.        Discard incoming requests
11.     End If
12.     If  $K \neq FULL$ 
13.        Dequeue first user from buffer  $K$ 
14.        Update  $W_S^q$  and  $W_P^q$  as per user type
15.        Update remaining service time of users in  $K$ 
16.     End If
17.     Repeat steps 12 to 16 until all users get service
18.     End For
19. Calculate total average waiting time  $W_S$  and  $W_P$  for SUs and PUs
20. Exit

```

4. Performance Analysis of Proposed Algorithm

To estimate the time complexity and convergence speed of the given Algorithm 1 to simulate the proposed models in a CRN, we need to analyze each step of the pseudocode and derive the complexity in terms of the number of users N and the buffer capacity K .

4.1. Time Complexity

The initialization phase is a constant-time operation and therefore has complexity of $O(1)$. Arrival times for PUs and SUs are generated using a Poisson distribution $O(N)$, where N is the total number of users (both PUs and SUs). To determine the type of arrival and enqueue the user $O(1)$ per user, the following conditions are considered. If the buffer is full, the process of discarding the user also has a constant complexity of $O(1)$ per user. Dequeueing the first user and updating the total waiting times have a constant complexity of $O(1)$. Updating the remaining service time for users in the buffer is considered as $O(K)$, since each user in the buffer might need an update. The process of repeating until all users are processed has complexity $O(N)$. Calculating average waiting times involves simple arithmetic operations and has a constant complexity of $O(1)$. Therefore, the overall complexity associated with generating arrival times is $O(N)$, while that associated with enqueueing and processing users is $O(N.B)$ and that associated with calculating averages is $O(1)$. Hence, combining these steps, the overall time complexity of the algorithm is found to be $O(N) + O(N.B) + O(1) = O(N.B)$.

4.2. Convergence Speed

The convergence speed of the algorithm is related to how quickly it reaches a stable state wherein the performance metrics (e.g., average waiting times) do not change significantly with additional iterations. For a number of users N , an increase in the number of users leads to better estimates of average waiting times but increases the computation time. The larger the buffer capacity K , the more users can be queued, affecting the total waiting times and their stabilization. For practical precision, the number of iterations required for convergence depends on the variability in arrival and service times. Typically, the system needs enough iterations to capture the stochastic behavior accurately. Empirically, convergence is often achieved after a sufficient number of arrivals and services have been simulated, and this number is typically $O(N)$. The algorithm processes each user in $O(K)$ time, so the convergence speed can be approximated by the time it takes to process all

users sufficiently. If each user needs to be processed multiple times for stabilization, the convergence iterations can be estimated as a factor of the number of users N .

Combining these factors, the convergence speed can be approximated by the product of the number of users N and the buffer capacity K , i.e., $O(N.K)$. Hence, the time complexity of the algorithm is $O(N.K)$, and the convergence speed is typically reached after $O(N)$ iterations, with each iteration taking $O(K)$ time, resulting in an overall complexity of $O(N.K)$.

5. Results

This section presents the numerical and simulation findings of the four strategies, as stated in part 3. The study conducted in [4], as discussed in Section 2, was employed to evaluate and contrast our findings. The result figures incorporate plots from [4] for the purpose of comparison. Monte Carlo simulation (MCS) was conducted for all results, and they have been incorporated into all the figure plots. The legends of the graphs designate the plot for the scheme proposed in [4] as ‘Scheme ref [4]’. The Monte Carlo simulation, abbreviated as ‘MCS’, is denoted solely by markers. The numerical graph displays the frequency of PU delays ($d = 0, 1, 2$), as represented by lines. Table 3 provides an overview of the simulation parameters used in the study to evaluate the proposed M/D/1 queueing-based CRN.

Table 3. Simulation parameters used in the study.

Parameter	Value
Service Rate (μ)	[11.00, 11.20, 11.40, 11.60] requests/sec
Arrival Rate (λ_P)	[19.00, 19.50, 20.00, 20.50] requests/sec
Arrival Rate (λ_S)	[5:10] requests/sec
Buffer Capacity	50 requests
Priority Scheme	PU > SU
Processing Time (d)	[0, 1, 2] sec/request
System Stability	Stable
Packet Accumulation	20 packets
Detection Time	0.1 s
Accuracy	95%
Interference Mitigation	Frequency-hopping
Performance Metrics	Avg. waiting time: 2 s

Figure 2 presents the impact of the utilization factor on the mean number of packets. It measures how aggressively a resource is used and affects the system packet count. This diagram shows how utilization-factor variations affect the mean number of packets, revealing the system’s behavior as the utilization factor fluctuates. In Figure 2, the analysis examines how changes in the utilization factor, which reflects resource-usage intensity in a system, affect waiting time. It shows how well the system handles demand and allocates resources based on wait times. Figure 3 describe the impact of utilization factor ρ_P on the average number of packets in the network.

In Figure 4, we present a comprehensive comparison of the average number of packets across a spectrum of service rates, delineated by different ranges of λ_S , reflecting the arrival rates of SUs. This visualization allows for a nuanced understanding of how changes in service rates impact the accumulation of packets within the network, offering insights into the system’s performance under varying levels of SU activity. Additionally, Figure 5 illustrates a detailed comparison of the average number of packets within the network with regard to the service rate for primary users (PUs), with λ_P varying across different scenarios. By comparing these metrics against service rates, this figure provides valuable insights into the behavior of packet accumulation for PUs under different arrival-rate conditions, facilitating a comprehensive analysis of network dynamics and resource utilization in cognitive radio environments.

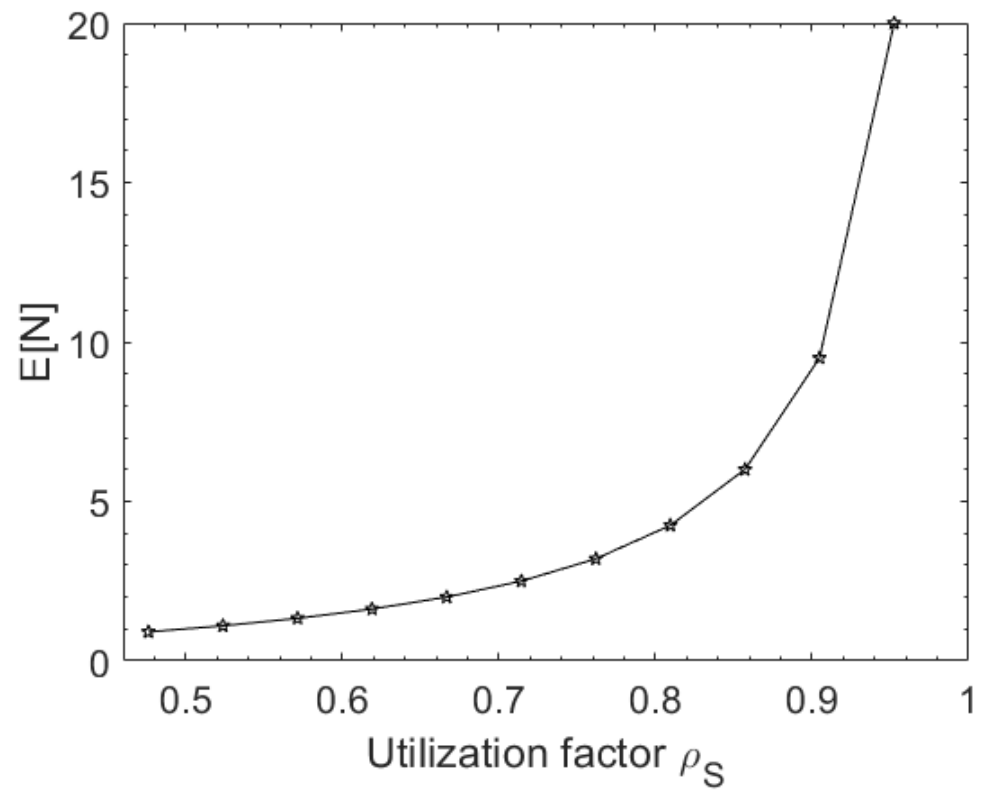


Figure 2. Impact of utilization factor ρ_S on the number of packets in the network.

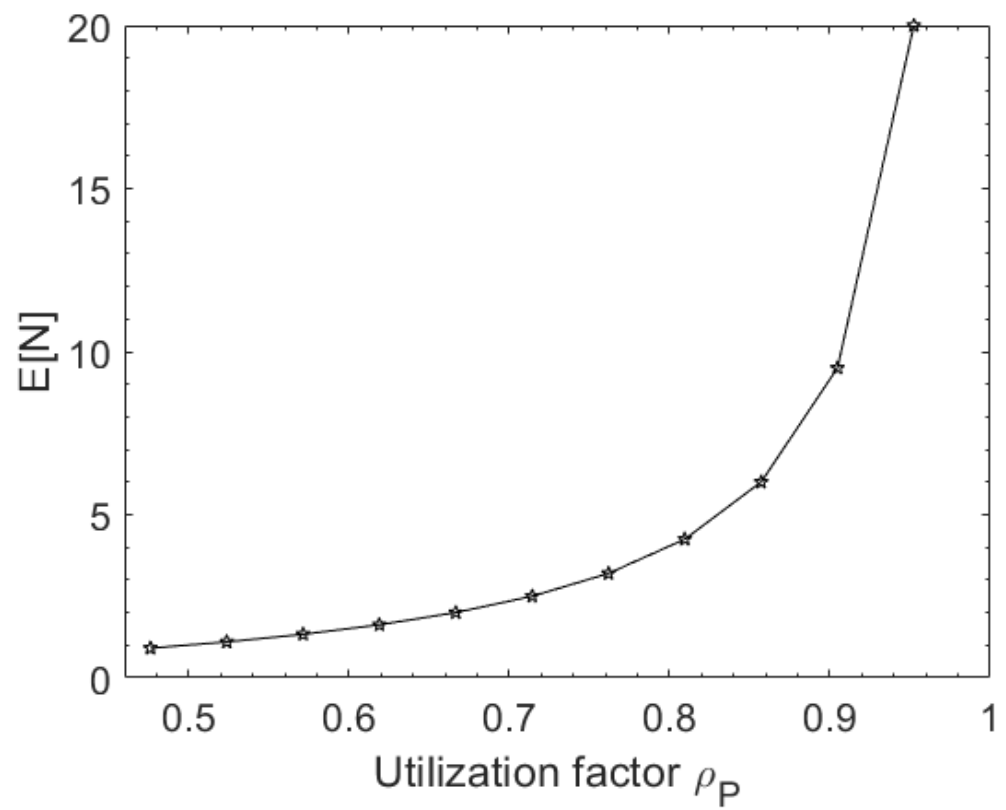


Figure 3. Impact of utilization factor ρ_P on the average number of packets in the network.

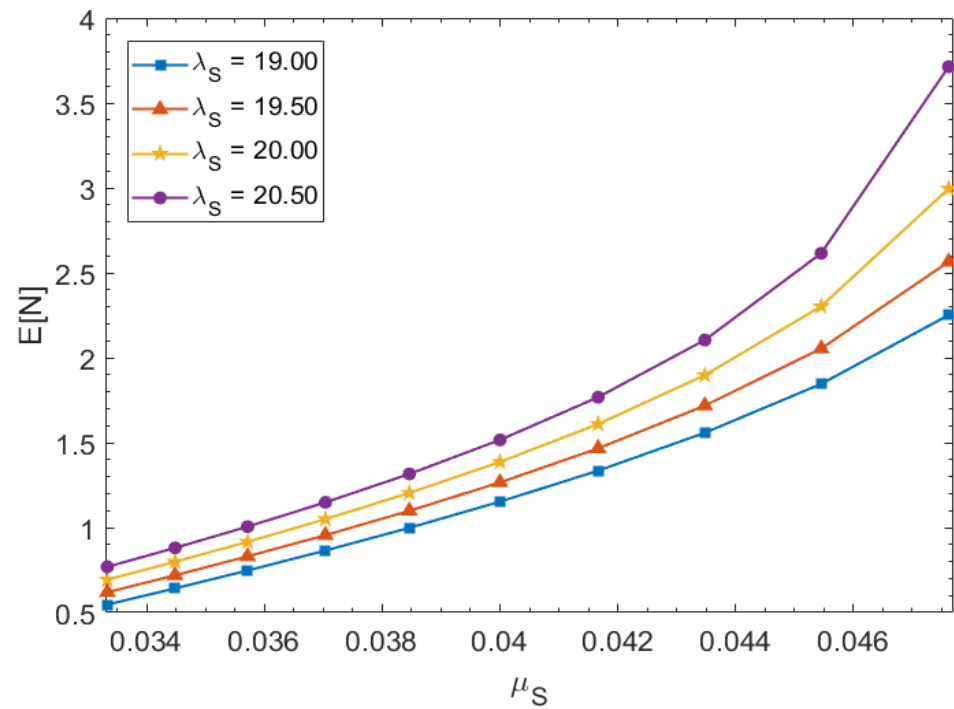


Figure 4. Comparison of average numbers of packets over varying service rates for varying λ_S .

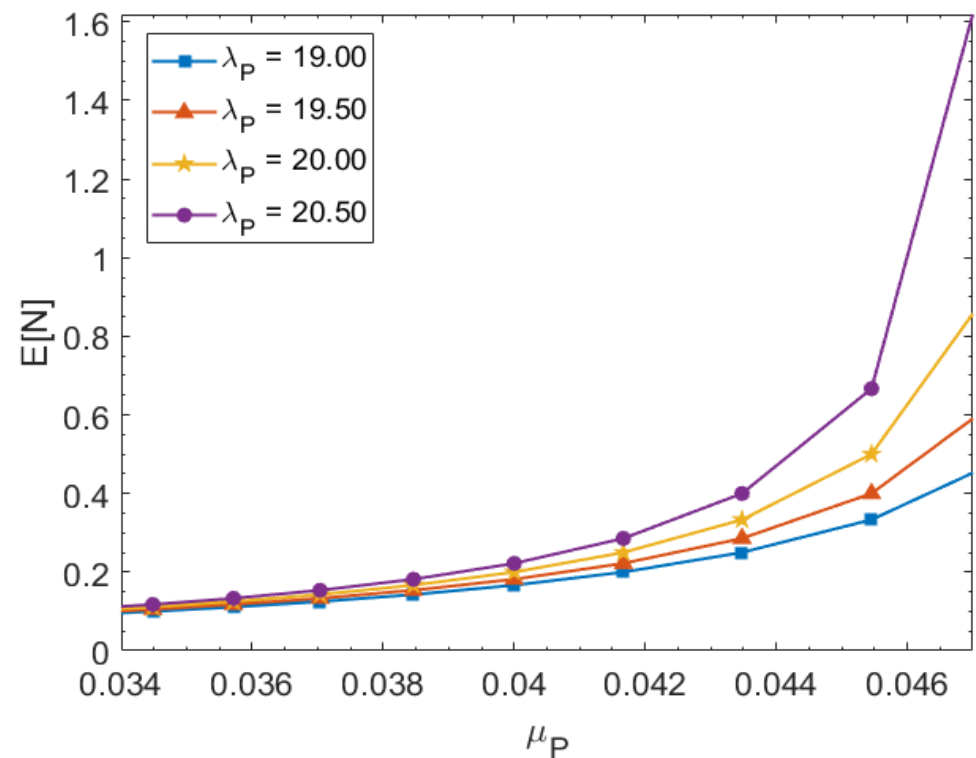


Figure 5. Comparison of average packets in network with respect to service rates for PUs with varying values of λ_P .

Figure 6 shows a comparison of network-induced SU delays with respect to service rate for different values of λ_S for SUs. It represents the delays experienced by SUs at intervals within the network as the level of service provided to SUs varies while the arrival rates for SUs, λ_S , also vary. Each curve represents a different value of λ_S , allowing a comparative analysis of how changes in λ_S affect delays caused by different SUs at different

service levels. The x -axis represents the service rate for SUs, while the y -axis shows the delay experienced by SUs. Figure 7 shows the delays experienced by PUs in the network in terms of coverage. It shows how the delay varies as the number of services provided to PUs changes, as the number of arrivals to PUs, λ_P , is also adjusted. For different value of λ_P , a comparative analysis of how changes in λ_P affect delays caused by different PUs at different service levels is depicted for the considered CRN.

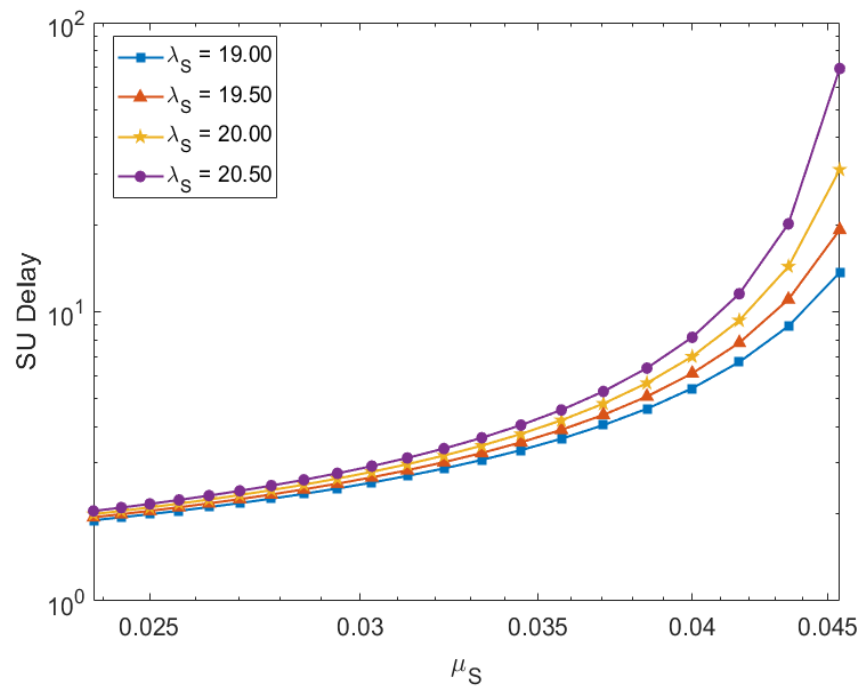


Figure 6. Comparison of SU delay incurred by the network with respect to service rate for SUs with varying λ_S .

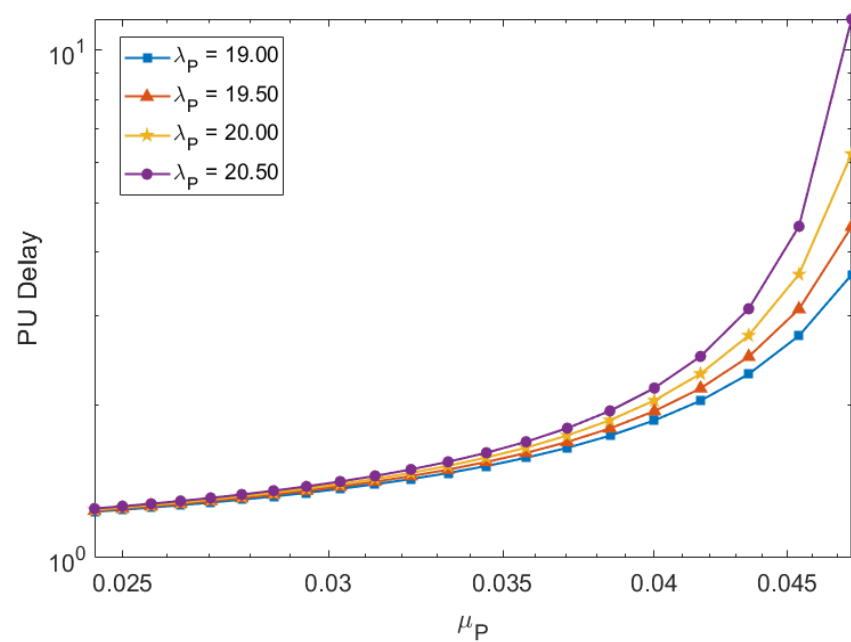


Figure 7. Comparison of PU delay incurred by the network with respect to service rate for PUs with varying λ_P .

For the purpose of obtaining a random sample of size 100 on the system size at service epochs, we used the system simulation. These are then divided into two categories: (a) those requests that ended up leaving behind an empty system; and (b) those that ended up leaving behind a system that was not empty. As a result, the mean and standard deviation of ρ_S is 0.5002 and the 95% confidence interval for ρ_S in the Wilson distribution is (0.5677, 0.7271), as can be observed from Figure 8 below. Additionally, it is worth noting that the mean and standard deviation of ρ_P in Figure 9 are both 0.5000. Additionally, the 95% confidence interval for ρ_P in the Wilson distribution is between 0.5767 and 0.7371.

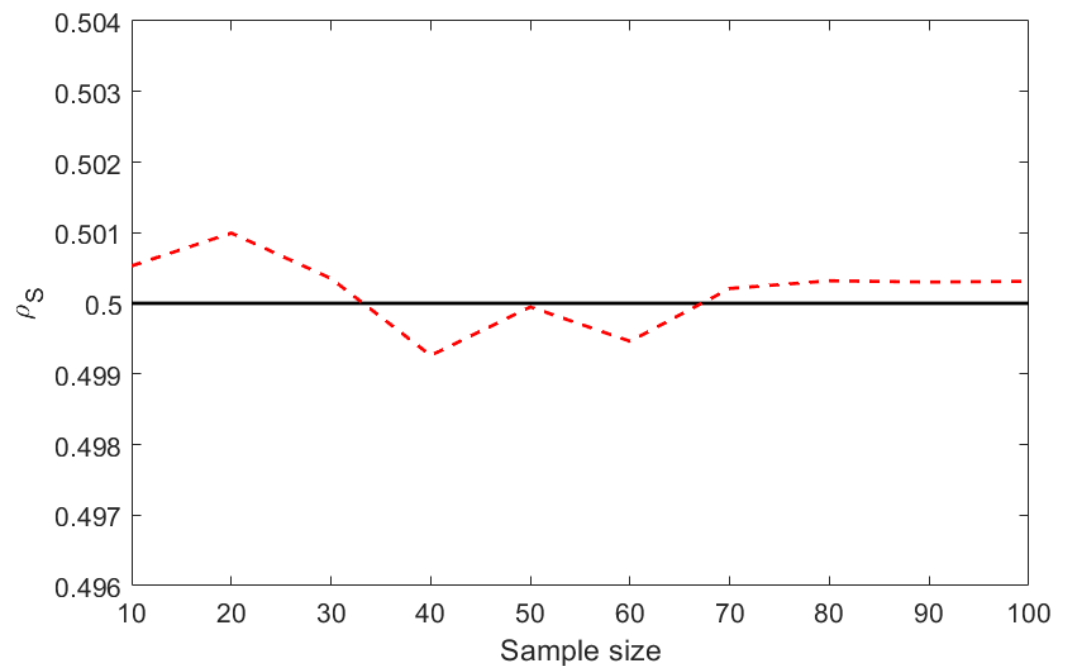


Figure 8. Comparison of varying sample sizes with ρ_S estimated for $\rho_S = 0.5$ under 95% CI.

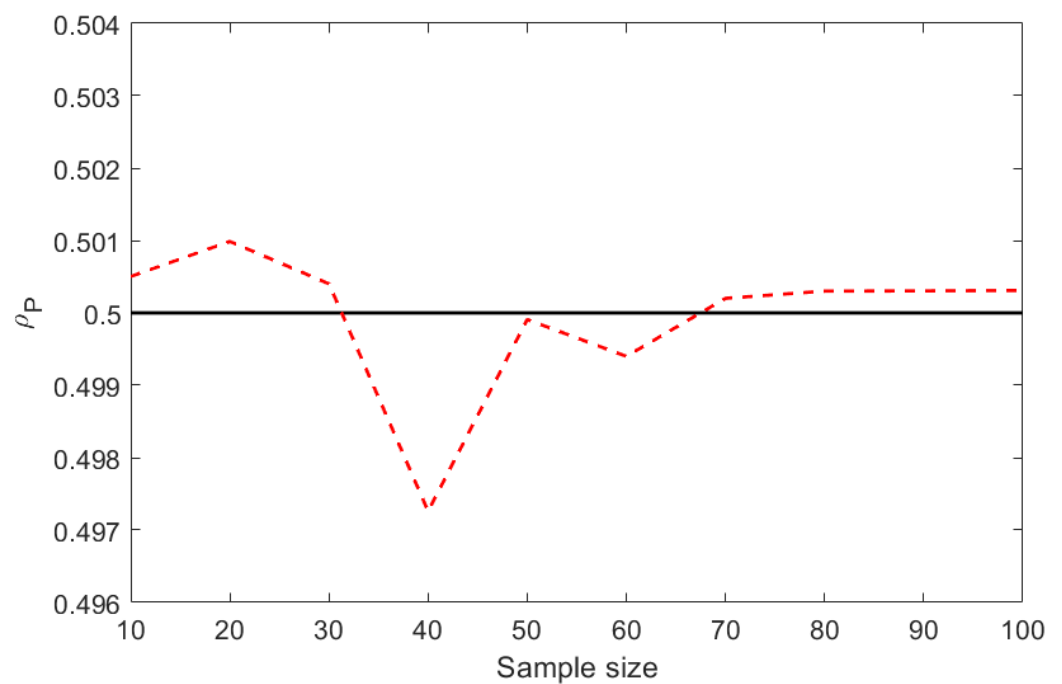


Figure 9. Comparison of varying sample sizes with ρ_P estimated for $\rho_P = 0.5$ under 95% CI.

6. Comparative Analysis

In this section, we provide a comparison of our proposed study with some existing state-of-the-art studies deployed commonly for managing the QoS of both PUs and SUs in CRNs. For the sake of analysis, we considered the studies that were most similar to our work to observe the specific improvements in QoS. The metrics used to evaluate the model are buffer size, delay between PUs and SUs, and network congestions. Table 4 provides the comparison of different studies with the proposed work.

Table 4. Comparative analysis of state-of-the-art literature with the proposed framework.

Related Works	Buffer Size	Delay	Network Congestion	Throughput
Perveen et al. [27]	✗	✓	✓	✗
Phung-Duc et al. [28]	✓	✗	✓	✗
Mahmood et al. [29]	✓	✗	✓	✓
Villota-Jacome et al. [30]	✗	✓	✗	✓
Luu et al. [31]	✗	✓	✓	✓
Proposed Framework	✓	✓	✓	✓

7. Challenges and Limitations

The challenges encountered by CRNs with PU and SU delays may be multi-faceted and depend on the characteristics of the communication protocol. This challenge in CRNs is mostly encountered due to dependency over arbitrary thresholds towards allocation of servers, which may in turn lead to suboptimal performance under dynamic network conditions [32]. Furthermore, CRNs can face challenges due to limited capabilities with limited sensitivity for accurately identifying the available spectrum, contributing to efficient resource utilization and causing congestion problems in practical applications, especially in different environments with different models and data causing delay and packet loss. Besides the optimization of resource allocation between active PU-SU links, the selection of optimal requirements for admission-control policies poses computational and heuristic challenges, and for understanding network dynamics [32].

For CRNs, the mentioned constraints provide valuable insights for improving system performance and QoS [32,33]. By addressing the reliance on arbitrary thresholds defining hysteresis paths, researchers can optimize responsive control mechanisms for server reservation, enabling more efficient spectrum management by increasing sensing capacity through accurate identification of available spectra to optimize resource usage and maximize spectrum utilization. In addition, the development of congestion control techniques helps to reduce collisions and ensure services with optimal QoS by prioritizing network resources and prioritizing critical services. The integration of deep learning techniques can further enhance CRNs through dynamic scheduling and resource allocation based on real-time rate of congestion in the CRN, ultimately providing improved network efficiency and experience [33–36].

8. Conclusions and Future Scope

The radio spectrum, which is used for wireless communication, is a limited resource. This study explores different OSA methods to enhance the QoS for both PUs and SUs. This study introduces innovative priority queueing models that incorporate PU delay and finite buffer size. These models have demonstrated a substantial enhancement in the QoS for both PU and SU. As a result of utilizing solely the model described in reference [4], the SU encounters no challenges in receiving prompt service without excessive waiting time. Nevertheless, the processing unit experiences a small increase in its waiting time. Therefore, the authors of this work conducted more research by incorporating priority models with the PU delay. Previous research has shown that implementing the non-preemptive priority

with PU delay leads to a further decrease in the average overall waiting time for the SU. Similarly, the pre-emptive priority with the PU delay model significantly enhanced the performance of the average total PU waiting time. The combination of non-preemptive priority and PU delay undeniably yields a favorable effect on the system compared to the PU delay model.

Author Contributions: Conceptualization, J.R.B., A.L.I. and S.S.S.; methodology, J.R.B.; software, S.S.T. and S.B.; validation, J.R.B., A.L.I., S.S.S., S.S.T. and S.B.; formal analysis, J.R.B.; investigation, J.R.B., A.L.I., S.S.S., S.S.T. and S.B.; resources, J.R.B., A.L.I., S.S.S., S.S.T. and S.B.; data curation, J.R.B.; writing—original draft preparation, J.R.B.; writing—review and editing, A.L.I., S.S.S., S.S.T. and S.B.; visualization, J.R.B., A.L.I., S.S.S., S.S.T. and S.B.; supervision, A.L.I. and S.S.S.; project administration, A.L.I., S.S.S., S.S.T. and S.B.; funding acquisition, A.L.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data will be made available on request from the authors.

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

References

1. Paluncic, F.; Alfa, A.S.; Maharaj, B.T.; Tsimba, H.M. Queueing models for cognitive radio networks: A survey. *IEEE Access* **2018**, *6*, 50801–50823. [\[CrossRef\]](#)
2. D’Apice, C.; D’Arienzo, M.P.; Dudin, A.; Manzo, R. Admission control in priority queueing system with servers reservation and temporal blocking admission of low priority users. *IEEE Access* **2023**, *11*, 44425–44443. [\[CrossRef\]](#)
3. Mitrany, I.L.; Avi-Itzhak, B. A many-server queue with service interruptions. *Oper. Res.* **1968**, *16*, 628–638. [\[CrossRef\]](#)
4. Falin, G.I. Waiting time in a single-channel queueing system with repeated calls. *Mosc. Univ. Comput. Math. Cybern.* **1977**, *4*, 83–87.
5. Falin, G.I. An M/G/1 system with repeated calls in heavy traffic. *Mosc. Univ. Comput. Math. Cybern.* **1980**, *6*, 48–50.
6. Falin, G.I. On the waiting-time process in a single-line queue with repeated calls. *J. Appl. Probab.* **1986**, *23*, 185–192. [\[CrossRef\]](#)
7. Artalejo, J.R.; Phung-Duc, T. Markovian retrial queues with two way communication. *J. Ind. Manag. Optim.* **2012**, *8*, 781–806. [\[CrossRef\]](#)
8. Salameh, O.; De Turck, K.; Bruneel, H.; Blondia, C.; Wittevrongel, S. Analysis of secondary user performance in cognitive radio networks with reactive spectrum handoff. *Telecommun. Syst.* **2017**, *65*, 539–550. [\[CrossRef\]](#)
9. Gavrilovska, L.; Rakovic, V.; Denkovski, D. From cloud ran to open ran. *Wirel. Pers. Commun.* **2020**, *113*, 1523–1539. [\[CrossRef\]](#)
10. Niknam, S.; Roy, A.; Dhillon, H.S.; Singh, S.; Banerji, R.; Reed, J.H.; Saxena, N.; Yoon, S. Intelligent O-RAN for beyond 5G and 6G wireless networks. *arXiv* **2020**, arXiv:2005.08374.
11. Shrimali, B.; Bhadka, H.; Patel, H. A fuzzy-based approach to evaluate multi-objective optimization for resource allocation in cloud. *Int. J. Adv. Technol. Eng. Exp.* **2018**, *5*, 140–150. [\[CrossRef\]](#)
12. Goudarzi, S.; Anisi, M.H.; Abdullah, A.H.; Lloret, J.; Soleymani, S.A.; Hassan, W.H. A hybrid intelligent model for network selection in the industrial internet of things. *Appl. Soft. Comput.* **2019**, *74*, 529–546. [\[CrossRef\]](#)
13. Suganthi, N.; Meenakshi, S. Efficient spectrum allocation for secondary users in cognitive radio network using round robin priority algorithm along with reservation channels. *J. Ambient. Intell. Humaniz. Comput.* **2023**, *14*, 16715–16728. [\[CrossRef\]](#)
14. Sun, K.; Liu, Y.; Li, K. Energy harvesting cognitive radio networks with strategic users: A two-class queueing model with retrials. *Comput. Commun.* **2023**, *199*, 98–112. [\[CrossRef\]](#)
15. Krishna Kumar, B.; Navaneetha Krishnan, R.; Sankar, R.; Rukmani, R. Performance analysis of cognitive wireless retrial queueing networks with admission control for secondary users. *Qual. Technol. Quant. Manag.* **2023**, *20*, 633–670. [\[CrossRef\]](#)
16. Zhang, S.; Ni, Z.; Kuang, L.; Jiang, C.; Zhao, X. Traffic Priority-Aware Multi-User Distributed Dynamic Spectrum Access: A Multi-Agent Deep RL Approach. *IEEE Trans. Cogn. Commun. Netw.* **2023**, *9*, 1454–1471. [\[CrossRef\]](#)
17. Winkler, A. Dynamic scheduling of a single-server two-class queue with constant retrial policy. *Ann. Oper. Res.* **2013**, *202*, 197–210. [\[CrossRef\]](#)
18. Cao, Y.; Zhao, Y.; Li, H. A Spectrum Occupancy Model for Primary Users in Cognitive Radio Systems. *J. Comput. Commun.* **2023**, *11*, 78–86. [\[CrossRef\]](#)
19. Dimitriou, I. A queueing model with two classes of retrial customers and paired services. *Ann. Oper. Res.* **2016**, *238*, 123–143. [\[CrossRef\]](#)
20. Kumar, N.; Aryan, P.; Raja, G.L.; Muduli, U.R. Robust Frequency-Shifting Based Control Amid False Data Injection Attacks for Interconnected Power Systems with Communication Delay. *IEEE Trans. Ind. Appl.* **2024**, *60*, 3710–3723. [\[CrossRef\]](#)
21. Zhang, Y.; Wang, J. Effectiveness, fairness and social welfare maximization: Service rules for the interrupted secondary users in cognitive radio networks. *Ann. Oper. Res.* **2023**, *323*, 247–286. [\[CrossRef\]](#)

22. Dey, S.; Saha Misra, I. A novel spectrum handoff switching decision scheme for improved performance of secondary users in cognitive radio network. *Int. J. Commun. Syst.* **2023**, *36*, e5413. [\[CrossRef\]](#)
23. Kulshrestha, R.; Singh, A. Transient analysis of energy-saving strategy for cognitive radio networks using G-queue with heterogeneity. *Comput. Commun.* **2024**, *216*, 295–306. [\[CrossRef\]](#)
24. Li, Y.; Zhai, X.B.; Wang, R.; Zhu, J.; Yao, C. Accumulating Priority Queue for Charging of Unmanned Aerial Vehicles in Cognitive Radio Networks. In Proceedings of the 2023 6th International Symposium on Autonomous Systems (ISAS), Nanjing, China, 23–25 June 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6.
25. Zhang, J.; Liang, C.L.; Gao, H.X.; Yan, H.; Zhu, H.B. Throughput-delay tradeoff for opportunistic spectrum access in cognitive radio networks. *IET Commun.* **2023**, *17*, 2053–2067. [\[CrossRef\]](#)
26. Mohamed, A.G.; Gamal, M. Stability region characterization of a two-user age of information restricted cognitive radio model. *Alex. Eng. J.* **2023**, *73*, 205–215. [\[CrossRef\]](#)
27. Perveen, A.; Abozariba, R.; Patwary, M.; Aneiba, A. Dynamic traffic forecasting and fuzzy-based optimized admission control in federated 5G-open RAN networks. *Neural Comput. Appl.* **2023**, *35*, 23841–23859. [\[CrossRef\]](#)
28. Phung-Duc, T.; Akutsu, K.; Kawanishi, K.I.; Salameh, O.; Wittevrangel, S. Queueing models for cognitive wireless networks with sensing time of secondary users. *Ann. Oper. Res.* **2022**, *310*, 641–660. [\[CrossRef\]](#)
29. Mahmood, A.; Kareem, A.R.; Nasser, A.R. Design of Elastic Call Admission Control in Tidal Load Scenario for Cloud-based Cellular Networks. *Elektron. Elektrotehnika* **2022**, *28*, 85–91. [\[CrossRef\]](#)
30. Villota-Jacome, W.F.; Rendon OM, C.; da Fonseca, N.L. Admission control for 5G core network slicing based on deep reinforcement learning. *IEEE Syst. J.* **2022**, *16*, 4686–4697. [\[CrossRef\]](#)
31. Luu, Q.T.; Kerboeuf, S.; Kieffer, M. Admission control and resource reservation for prioritized slice requests with guaranteed SLA under uncertainties. *IEEE Trans. Netw. Serv. Manag.* **2022**, *19*, 3136–3153. [\[CrossRef\]](#)
32. Kanellopoulos, D.; Sharma, V.K.; Panagiotakopoulos, T.; Kameas, A. Networking architectures and protocols for IoT applications in smart cities: Recent developments and perspectives. *Electronics* **2023**, *12*, 2490. [\[CrossRef\]](#)
33. Dudin, A.; Dudin, S.; Manzo, R.; Rarità, L. Analysis of multi-server priority queueing system with hysteresis strategy of server reservation and retrials. *Mathematics* **2022**, *10*, 3747. [\[CrossRef\]](#)
34. Tam, P.; Kang, S.; Ros, S.; Kim, S. Enhancing QoS with LSTM-Based Prediction for Congestion-Aware Aggregation Scheduling in Edge Federated Learning. *Electronics* **2023**, *12*, 3615. [\[CrossRef\]](#)
35. Adelabu, M.A.; Imoize, A.L.; Ugwu, M.B. Radio Resource Management of WLAN Hotspot Access Points in Next Generation Wireless Networks. *SN Comput. Sci.* **2023**, *4*, 313. [\[CrossRef\]](#)
36. Zhou, X.; Yang, Q.; Zheng, X.; Liang, W.; Wang, K.I.-K.; Ma, J.; Pan, Y.; Jin, Q. Personalized Federated Learning with Model-Contrastive Learning for Multi-Modal User Modeling in Human-Centric Metaverse. *IEEE J. Sel. Areas Commun.* **2024**, *42*, 817–831. [\[CrossRef\]](#)

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.