

# Age of Information Minimization Through Reinforcement Learning in IoT Enabled Cognitive Radio Networks

<sup>1</sup>Ragupathi C, <sup>2</sup>Sathish Kumar Ravichandran, <sup>3</sup>Sivasamy A and <sup>4</sup>Deepa R

<sup>1</sup>Department of Electrical and Electronic Engineering, P. T. Lee Chengalvaraya Naicker College of Engineering and Technology, Veliyur, Tamil Nadu, India.

<sup>2</sup>Department of Computer Science and Engineering, School of Advanced Studies, S-VYASA Deemed to be University, Bangalore, Karnataka, India.

<sup>3</sup>Department of Mechanical Engineering, Dr. N.G.P. Institute of Technology, Coimbatore, Tamil Nadu, India.

<sup>4</sup>Department of Computer and Communication Engineering, Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India.

<sup>1</sup>raguptlee@gmail.com, <sup>2</sup>cbsathish@hotmail.com, <sup>3</sup>a.sivasamy@gmail.com, <sup>4</sup>ecedeeepa@gmail.com

Correspondence should be addressed to Ragupathi C : raguptlee@gmail.com

## Article Info

Journal of Machine and Computing (<https://anapub.co.ke/journals/jmc/jmc.html>)

Doi : <https://doi.org/10.53759/7669/jmc202505211>

Received 02 May 2025; Revised from 12 August 2025; Accepted 19 September 2025.

Available online 05 October 2025.

©2025 The Authors. Published by AnaPub Publications.

This is an open access article under the CC BY-NC-ND license. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

**Abstract** – The rapid growth of Internet of Things (IoT) applications has led to an unprecedented demand for efficient wireless communication, particularly in scenarios where devices compete for limited spectrum resources. Cognitive Radio Networks (CRNs) provide a promising solution by enabling dynamic spectrum access, yet maintaining timely information updates remains a challenge. Age of Information (AoI), a metric that measures the freshness of status updates, has become a key performance indicator in such networks. This paper proposes a reinforcement learning (RL)-based scheduling framework designed to minimize AoI in IoT-enabled CRNs. The framework models the dynamic environment of spectrum availability and device activity, and uses an adaptive RL agent to optimize transmission scheduling decisions in real time. A mathematical model of the AoI minimization problem is elaborated and after that, the proposed learning algorithm is designed. Large-scale simulations are done in Python in Google Colab and the performance of these simulations is compared with four state-of-the-art scheduling strategies. Findings demonstrate that the suggested RL-based system is able to generate lower AoI with different network scales, channel characteristics, and traffic dynamics. The results indicate the promise of RL-based solutions to the creation of scalable and reactive IoT communication systems.

**Keywords** – Cognitive Radio Networks, IoT, Reinforcement Learning, Age of Information, Resource Allocation.

## I. INTRODUCTION

The exponential growth of the IoT has turned the communication networks into a very dynamic and data-driven ecosystem. IoT applications have expanded into many areas such as smart cities, healthcare monitoring and automation, transportation, and environmental sensing. These applications are often dependent on periodic updates or transmission of status information from sensors, actuators, and edge devices back to centralized controllers or cloud services in real time. As the number of connected devices continues to increase, networks must handle more complex traffic patterns with a mix of devices and ensure reliable communication and low latency. Traditional approaches of wireless spectrum allocation, typically based on fixed frequency allocation, cannot satisfy the need for large-scale connectivity for IoT scenarios. This provides the motivation for flexible spectrum management strategies which are able to dynamically allocate resources and adapt to traffic fluctuations [1].

The CRNs have become a promising solution to the spectrum scarcity issue by allowing devices to access portions of frequency bands that have been licensed to others on an opportunistic basis. In CRNs, secondary users (SUs), e.g., IoT nodes, are able to transmit information over channels temporally vacated from PUs without causing harmful interference [2]. This is a paradigm that improves the overall spectrum utilization and increases the capacity of the network. However, in addition to resolving the problem of spectrum availability, CRNs bring with them further challenges associated with maintaining timely and efficient communication. IoT applications, especially those that include mission-critical control

systems or real-time monitoring, require low-latency communications with an emphasis on the freshness of information. Traditional metrics such as throughput, packet delivery ratio and delay are not sufficient to capture the timeliness of delivered information in scenarios where outdated data can degrade system performance [3].

#### *Motivation of This Work*

To overcome this limitation, the concept of AoI has been introduced. AoI is a performance measure of the time since the last successfully received update was generated at the source. Unlike classical delay metrics, AoI measures the freshness of the data at the receiver directly making it very applicable for IoT applications such as autonomous driving, industrial automation and remote health monitoring. A low AoI implies the information at the decision making node is up-to-date and this is vital to accurate control and timely response [4, 5]. However, the fact is that minimizing AoI in CRNs is nontrivial because of the unpredictable channel availability as well as interference constraints, heterogeneous traffic patterns, and energy limitations of IoT devices.

There have been previous research attempts in wireless networks to enhance the AoI performance using different methods [6]. Scheduling policies have been proposed to round-robin, random access and prioritize transmission opportunities based on age, among others. While these approaches can provide some improvements in AoI, they are often based on simplified models of network behavior and don't fully account for the stochastic nature of CRNs. Moreover, optimization based techniques, such as convex optimization and linear programming, have been used for AoI minimization problems. Although these approaches may provide optimal solutions in some assumptions, the main drawback is the high computational complexity and low scalability in large IoT deployments. In practice, optimization problems need to be solved in real time, but this becomes infeasible as the size of the network, the variability of traffic and uncertainty of channels increases [7].

#### *Challenges in AoI, IoT and CRNs*

In the past years, machine learning (ML) methods have demonstrated an enormous potential to solve complex resource allocation problems in wireless networks. Reinforcement learning (RL) in particular has become a powerful decision-making paradigm applied to dynamic environments. RL allows agents to acquire optimal policies from interaction with the environment, without having to explicitly model the system's dynamics. This makes it suitable for CRNs, where channel availability, interference and traffic patterns can change unpredictably. RL-based approaches have already been used for problems like spectrum sensing, power allocation and channel assignment. However, research on the use of RL for AoI minimization in IoT-enabled CRNs is still at an early stage. Most existing works are focused on maximizing throughput or minimizing delay, without taking data freshness directly into consideration.

Another issue with CRNs is finding a balance between AoI performance and energy efficiency and fairness among devices. Many of the nodes in IoT are battery-powered devices, and constant transmission to keep AoI to the minimum can be draining resources quickly. Designing a scheduling policy that minimizes AoI while conserving energy poses a problem that requires careful consideration of trade-offs. Additionally, the question of ensuring fairness is critical in IoT scenarios where some devices, such as safety sensors, may need priority access whilst others may be able to accept some delays from time-to-time. Static or rule-based scheduling algorithms do not often achieve this balance.

To mitigate these challenges, to the best of our knowledge, this paper proposes a novel RL-based scheduling framework that dynamically assigns the transmission opportunities to IoT devices in a CRN that aims to minimize AoI. The proposed approach treats the scheduling problem as a Markov Decision Process (MDP) where the system state represents the current AoI of all devices, the availability of channels and interference constraints. The RL agent interacts with the environment and learns from its action and gradually becomes better in scheduling strategy through trial and error. This approach does not require precision mathematical modeling of the system dynamics, instead the scheduler can adapt to the changing network conditions in real time.

The novelty of this work is in combining reinforcement learning and AoI-aware decision making in cognitive radio environments. Unlike traditional optimization or heuristic methods, the proposed method allows scalable and autonomous scheduling without extensive knowledge of channel models or traffic patterns. Furthermore, the framework is designed to handle various types of network scenarios such as dense IoT deployments, highly variable traffic and heterogeneous device requirements. By taking advantage of RL, the scheduler can learn to prioritize devices with fast increasing AoI, dynamically allocate channels according to the availability, and utilize idle spectrum opportunities efficiently.

#### *Contributions of The Research*

- The AoI minimization problem for IoT-enabled CRNs is represented as a RL task, which is effective in modeling the stochastic nature of the channel availability and device activity.
- An RL-based scheduling framework is developed in which scheduling policies that learn adaptive scheduling for AoI minimization while satisfying their interference constraints and achieving efficient spectrum utilization can be learned.
- Extensive simulations are carried out under realistic CRN scenarios to assess performance and comparisons are drawn with four state-of-the-art baseline techniques: random access, round robin scheduling, greedy policy and deep Q-network (DQN)-based approaches.

- The proposed approach has shown consistent improvements with respect to the baseline techniques in terms of AoI performance for different network sizes, traffic intensities and channel reliability levels.
- The framework has scalability to large scale IoT deployments and can be extended to incorporate energy-aware scheduling strategies, and is therefore suited to practical CRN implementations.

By addressing the challenge of AoI minimization in dynamic and spectrum-constrained IoT environments, this work contributes to the broader goal of building communication systems that are both intelligent and adaptive. The increasing complexity of IoT ecosystems demands resource management strategies that can operate effectively in uncertain conditions, and reinforcement learning offers a natural pathway toward this objective. The proposed approach represents a step toward developing fully autonomous spectrum management solutions capable of supporting next-generation IoT services that rely on timely and reliable information delivery.

The remainder of the paper is organized as follows. Section II reviews related work on AoI optimization and machine learning approaches in CRNs. Section III presents the system model, including network assumptions and the mathematical formulation of AoI. Also, it describes the proposed reinforcement learning framework, detailing the state representation, action space, and reward structure. Section IV outlines the simulation setup, including parameter choices and baseline algorithms for comparison and discusses simulation results and key findings. Finally, Section V concludes the paper and outlines directions for future research.

## II. RELATED WORKS

The rapid evolution of wireless communication and the proliferation of IoT applications have intensified research efforts toward improving network performance metrics such as latency, throughput, and spectrum efficiency. In recent years, the AoI metric has emerged as a critical performance indicator, particularly in real-time monitoring and control systems where information freshness is dominant [8]. The CRNs, with their ability to dynamically allocate underutilized spectrum, provide a promising platform for addressing the spectrum scarcity challenge inherent in IoT deployments. This section reviews existing literature on AoI optimization, RL-driven communication strategies, and scheduling techniques in CRNs, followed by a detailed critique that motivates the need for the proposed approach.

### *AoI Optimization in Wireless Networks*

The AoI metric has attracted significant attention in wireless communication research as it quantifies the timeliness of status updates rather than simply measuring throughput or latency. Early works primarily focused on analytical modeling of AoI under various queuing disciplines. Kaul et al. introduced seminal models for AoI in single-source single-server systems, demonstrating that conventional throughput optimization techniques often fail to guarantee freshness in time-sensitive applications [9]. Subsequent studies extended these models to multi-source networks, where priority-based scheduling policies were evaluated to minimize AoI. In wireless networks, AoI minimization has been studied under energy and bandwidth constraints. Sun et al. investigated AoI in systems with constrained energy harvesting sensors, deriving closed-form scheduling policies to optimize performance. Similarly, studies in vehicular networks and industrial IoT applications have highlighted the importance of balancing update frequency and channel reliability to maintain information freshness. While these studies provided valuable theoretical insights, most relied on static scheduling or offline optimization, limiting adaptability in environments with rapidly changing channel conditions and traffic dynamics [10].

A major drawback of the early AoI research is due to the use of deterministic models and static scheduling policies. Such approaches are not well suited to dynamic CRNs, in which the availability of spectrum changes unpredictably as a result of the activity of PUs. Moreover, closed-form scheduling solutions often assume perfect knowledge of channel state and synchronized updates, which nevertheless are rarely achieved in real-world deployments.

### *Cognitive Radio Networks for IoT*

Cognitive Radio (CR) technology has emerged as a transformative technology to address spectrum scarcity by allowing unlicensed spectrum users (SU) to opportunistically access underutilized licensed spectrum. Research in CRNs has mainly addressed spectrum sensing, spectrum access and interference mitigation. Early versions of CRNs have used simple sensing and centralized spectrum allocation mechanisms to ensure fair access without interfering with PUs. Over the years, distributed spectrum access protocols were proposed to enhance the scalability in dense IoT environments [11].

Significant contributions include game theoretic approaches to dynamic spectrum allocation, auction based resource management frameworks and graph based models for interference minimization. These solutions have proven to be effective in improving spectral efficiency; however, most of them were focused on boosting throughput or reducing the interference, with less focus on AoI optimization. The temporal aspect of the freshness of information adds a special challenge in CRNs, since the high frequency of spectrum switching and the uncertainty of channel availability may worsen AoI degradation [12].

Recent studies have explored joint spectrum sensing and scheduling strategies in CRNs to enhance timeliness. For instance, optimization-based models have been developed to jointly minimize sensing overhead and maximize transmission opportunities. Nonetheless, the inherent complexity of these models often leads to computationally expensive solutions that cannot adapt to large-scale IoT deployments. Furthermore, most optimization frameworks assume static or predictable PU activity, whereas real-world CRNs exhibit highly stochastic behavior [13].

### *Reinforcement Learning in Wireless Networks*

Reinforcement Learning has emerged as a powerful paradigm for decision-making in dynamic wireless networks. RL-based algorithms allow network agents to learn adaptive policies through trial-and-error interactions with the environment, making them well-suited for complex, uncertain systems such as CRNs. In recent years, RL has been applied extensively for resource allocation, channel selection, and power control in wireless networks [14]. Traditional Q-learning and SARSA algorithms have been used to improve throughput and reduce latency in CRNs, while deep reinforcement learning (DRL) approaches such as Deep Q-Networks (DQNs) have enabled scalable solutions for high-dimensional state spaces. Actor-Critic methods and policy gradient techniques have further enhanced RL applicability in wireless systems by providing stability in continuous action spaces [15].

However, despite their effectiveness, many RL-driven solutions in wireless networks primarily target throughput maximization, energy efficiency, or fairness, rather than AoI optimization. Efforts to directly integrate AoI into the RL reward design are relatively limited. A few recent works have attempted to minimize AoI using Q-learning or DQN-based frameworks, yet these methods often suffer from slow convergence, limited generalization, and scalability challenges when applied to large IoT networks. Another major limitation is in the form of simplistic environments in simulations, which fail to capture the complexities of CRNs. For instance, many RL-based studies assume perfect channel state information (CSI), do not model the dynamics of PUs, or greatly simplify network traffic models, thereby producing results that do not generalize well to practical deployments [16].

### *Scheduling and Access Control for AoI Minimization*

Efficient scheduling policies are then very important in order to keep the AoI low in multi-user networks. Classical approaches include round-robin scheduling [17], maximum age first (MAF) policies [18] and max-weight scheduling based on age or queue length metrics. Although the performance guarantees offered by these policies under certain circumstances, they are deterministic, which reduces flexibility in highly dynamic environments. Greedy policies based on the highest-priority in updating the oldest information have been demonstrated to perform near-optimal in simple network scenarios [19]. However, these approaches are not computationally scalable to large-scale networks and are susceptible to performance degradation when spectrum availability is unpredictable. Similarly, random access protocols though easy to implement have poor AoI performance due to high collision probabilities and lack of coordination. Recent research has suggested hybrid methods that combine optimization-based scheduling with predictive modelling. For example, Lyapunov optimization approaches have been used for the design of adaptive update policies. Nonetheless, such approaches are usually associated with high computational complexity and need a lot of system state information, which makes them impractical for real-time operation in IoT-enabled CRNs [20].

### *Limitations of Existing Work*

A review of current literature highlights several key limitations:

- *Static and Offline Models:* Many currently available AoI optimization frameworks are based on offline optimization or static scheduling policies which do not take into account the very dynamic nature of CRNs, where spectrum availability and network traffic varies without prior notice.
- *Limited Focus on AoI in CRNs* Although CRNs have been well studied in terms of spectrum efficiency and interference management, there are relatively few works focusing on AoI minimization in IoT-enabled CRNs. This gap highlights the need for frameworks that are specifically geared towards ensuring information freshness in dynamic spectrum environments [21].
- *Oversimplified Simulation Environments:* A large proportion of studies based on RL assess the performance in excessively simplified conditions, e.g. perfect channel knowledge, static activity of PUs, or homogeneous traffic models. These assumptions constrain the usefulness of proposed methods [22].
- *Scalability Challenges* Existing optimization and scheduling techniques have often shown exponentially high complexity with the growing size of the network, making them unsuitable for large IoT deployments.
- *Slow Convergence of RL Algorithms:* Conventional RL and DRL approaches have problems with convergence speed and stability, especially in high-dimensional state-action spaces. This is a limitation on their effectiveness in time-sensitive applications where fast adaptation is required.
- *Lack of Integrated Design:* Most of the solutions are optimized for either spectrum allocation or AoI minimization separately, instead of jointly considering the interplay between channel reliability and interference constraints and the timeliness.

These limitations together provide an incentive to design a reinforcement learning-based scheduling framework to explicitly target AoI minimization in IoT-enabled CRNs. The proposed approach overcomes these limitations by combining an RL agent and a well-designed reward function to adaptively choose transmission actions with dynamic spectrum conditions. A thorough evaluation under realistic CRN scenarios further distinguishes this work by showing reliable performance gains over multiple state-of-the-art baseline techniques.

### III. SYSTEM MODEL

Consider an IoT-enabled CRN consisting of a set of SU devices  $\mathcal{D} = \{1, 2, \dots, D\}$  communicating with a central fusion center or sink node over a set of licensed spectrum channels  $\mathcal{C} = \{1, 2, \dots, C\}$ . The PUs hold spectrum licenses and have prioritized access, while SUs opportunistically access idle channels under interference constraints.

#### Network Dynamics

At each discrete time slot  $t \in N$ , each SU can either transmit a status update or remain idle. The state of channel  $c \in \mathcal{C}$  is represented as a binary random variable:

$$s_c(t) = \begin{cases} 1, & \text{if channel } c \text{ is idle at time } t, \\ 0, & \text{if channel } c \text{ is occupied by a PU} \end{cases} \quad (1)$$

The probability of a channel being idle is denoted as:

$$P[s_c(t) = 1] = p_c \quad (2)$$

where  $p_c \in (0, 1]$  reflects the activity pattern of PUs.

Each SU transmits with a fixed power  $P_{tx}$  and achieves a data rate on channel  $c$  given by Shannon's capacity formula:

$$R_{d,c}(t) = B_c \cdot \log_2 \left( 1 + \frac{P_{tx} \cdot h_{d,c}(t)}{N_0 + I_c(t)} \right) \quad (3)$$

where:

$B_c$  is the bandwidth of channel  $c$ ,  $h_{d,c}(t)$  is the instantaneous channel gain between SU  $d$  and the receiver,  $N_0$  is the noise spectral density, and  $I_c(t)$  is the interference power from other active users.

#### Age of Information (AoI) Modeling

The AoI of device  $d$  at time  $t$ , denoted as  $\Delta_d(t)$ , is defined as the time elapsed since the generation of the most recent successfully received status update. Let  $u_d(t)$  denote the generation time of the last successfully received update from device  $d$ . Then:

$$\Delta_d(t) = t - u_d(t) \quad (4)$$

The AoI evolves according to:

$$\Delta_d(t+1) = \begin{cases} 1, & \text{if a new update from device } d \text{ is successfully received at } t, \\ \Delta_d(t) + 1, & \text{Otherwise} \end{cases} \quad (5)$$

The system-wide AoI is given by the average AoI across all devices:

$$\Delta_{avg}(t) = \frac{1}{D} \sum_{d=1}^D \Delta_d(t) \quad (6)$$

#### Scheduling Action Space

At each time slot, a centralized controller or agent selects scheduling actions for all devices. Let  $a_{d,c}(t) \in \{0, 1\}$  represents whether SU  $d$  transmits over channel  $c$  at time  $t$ :

$$a_{d,c}(t) = \begin{cases} 1, & \text{if device } d \text{ is assigned channel } c \text{ at time } t \\ 0, & \text{Otherwise} \end{cases} \quad (7)$$

To ensure interference protection for PUs and avoid SU collisions:

$$\sum_{d=1}^D a_{d,c}(t) \leq 1, \quad \forall c \in \mathcal{C}, \quad (8)$$

$$\sum_{c=1}^C a_{d,c}(t) \leq 1, \quad \forall d \in \mathcal{D}. \quad (9)$$

#### Transmission Success Probability

A scheduled transmission from SU  $d$  over channel  $c$  is successful if the channel is idle and the received SINR exceeds a threshold  $\gamma_{th}$ . Define:

$$q_{d,c}(t) = P[\text{SINR}_{d,c}(t) \geq \gamma_{\text{th}} \mid s_c(t) = 1] \quad (10)$$

Thus, the overall probability of a successful transmission is:

$$p_{\text{succ}}^{d,c}(t) = p_c \cdot q_{d,c}(t) \quad (11)$$

#### State Space Definition

The system state at time  $t$ , denoted as  $x(t)$ , is defined as:

$$x(t) = [\Delta_1(t), \Delta_2(t), \dots, \Delta_D(t), s_1(t), s_2(t), \dots, s_C(t)] \quad (12)$$

This representation captures both the freshness of information for each device and the availability of spectrum resources.

#### Optimization Problem Formulation

The objective is to minimize the long-term time-averaged AoI under interference and resource constraints. Formally:

$$\min_{\{a_{d,c}(t)\}} \limsup_{t \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \Delta_{\text{avg}}(t) \quad (13)$$

This optimization is inherently stochastic and combinatorial due to time-varying channel conditions and PU activity. Exact solutions require exhaustive search, making real-time scheduling intractable for large networks.

#### Reinforcement Learning Formulation

To achieve scalability and adaptivity, the problem is reformulated as a Markov Decision Process (MDP):

- *State  $x(t)$* : AoI values of all devices and current channel states.
- *Action  $a(t)$* : Scheduling decisions  $a_{d,c}(t)$ .
- *Reward  $r(t)$* : Negative average AoI at time  $t$ :

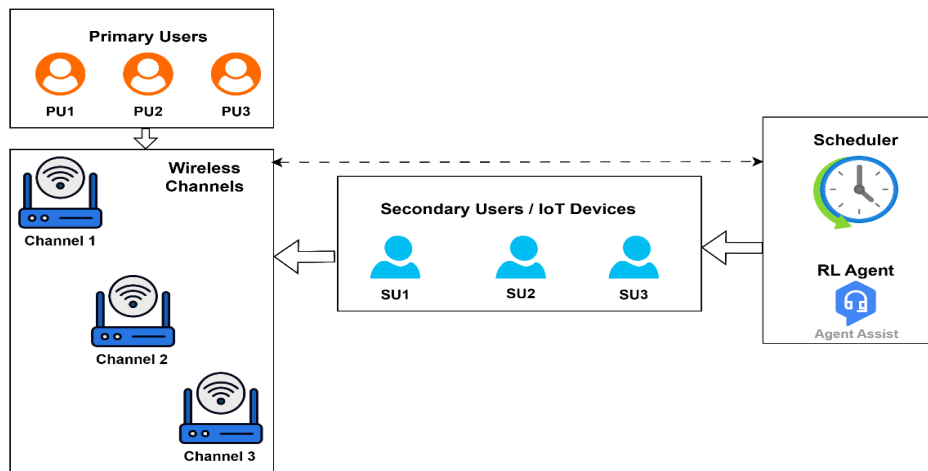
$$r(t) = -\Delta_{\text{avg}}(t) \quad (14)$$

- *Transition*: Determined by channel state evolution, PU activity, and transmission success probability.

The goal is to find a policy  $\pi: x(t) \rightarrow a(t)$  that minimizes the expected long-term AoI:

$$\pi^* = \arg \min_{\pi} E_{\pi} \left[ \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \Delta_{\text{avg}}(t) \right] \quad (15)$$

This mathematical formulation captures the stochastic, dynamic, and resource-constrained nature of AoI minimization in IoT-enabled CRNs. The next section will introduce the RL-based scheduling framework designed to solve this optimization efficiently, with scalable neural network approximations to handle large state-action spaces.



**Fig 1.** System Architecture of The IoT-Enabled Cognitive Radio Network With Reinforcement Learning-Based Scheduling.

**Fig 1** illustrates the overall architecture of the proposed IoT-enabled CRN, designed to minimize the AoI through intelligent RL-based spectrum scheduling. The network comprises four primary components: Primary Users (PUs), Secondary Users (SUs) or IoT devices, a pool of licensed wireless channels, and a centralized Fusion Center equipped with an RL agent for decision-making. In this framework, PUs represent the licensed spectrum owners who have prioritized access to designated frequency bands. These PUs are depicted on the left side of **Fig 1** within a dedicated cluster, symbolizing the fixed and regulated nature of their transmissions. The SUs, shown on the right side, are IoT devices operating in an opportunistic spectrum access mode. These SUs dynamically exploit spectrum holes unused frequency slots temporarily left idle by the PUs to transmit time-sensitive data updates to the Fusion Center. At the center of the diagram lies the Channel Pool, representing multiple orthogonal wireless channels that interconnect the PUs, SUs, and the Fusion Center. Each channel may experience stochastic availability due to PU activity, environmental conditions, and interference. This dynamic nature of channel states motivates the use of RL for optimal decision-making. The arrows connecting the PU and SU clusters to the channel pool illustrate their shared contention for spectrum resources, while the arrows between the channel pool and Fusion Center denote bidirectional communication: data transmission from SUs and spectrum sensing or scheduling instructions from the controller.

The Fusion Center, located at the bottom of **Fig 1**, acts as the central coordinator and is responsible for spectrum sensing, decision-making, and scheduling policies. Integrated into this Fusion Center is a RL agent that observes real-time network states, including PU activity patterns, SU traffic loads, and channel reliability levels. Based on these observations, the RL agent learns an adaptive spectrum allocation and scheduling strategy to minimize AoI while ensuring coexistence with licensed PUs and maintaining Quality of Service (QoS) constraints for IoT applications. Dashed arrows in **Fig 1** represent control signaling between the Fusion Center and network entities, highlighting the centralized orchestration of decisions. Solid arrows denote data transmission paths, emphasizing the flow of status updates from IoT devices to the Fusion Center via opportunistic channel access. This dual-arrow representation clearly separates the data plane from the control plane, which is critical for real-time operation in large-scale IoT deployments. The diagram has been designed to clearly depict system scalability, showing multiple IoT devices and PUs to illustrate a dense network scenario. The centralized RL-based control ensures efficient spectrum utilization, adaptability to network changes, and improved timeliness of information delivery. This architecture forms the foundation of the mathematical modeling and simulation studies presented in subsequent sections, where its performance is benchmarked against several baseline techniques.

#### Algorithm: RL-Based AoI Minimization in IoT-Enabled CRNs

Input:

- D: Number of IoT devices
- C: Number of channels
- T: Time horizon
- $\epsilon$ : Exploration probability
- $\theta$ : SINR threshold

Initialize:

- Q-table (or neural network weights)
- AoI values  $\Delta_d(0) = 1$  for all devices  $d = 1 \dots D$

for each episode do

Initialize environment state  $S_0 = \{\text{AoI values, channel states}\}$

for  $t = 1$  to  $T$  do

1. Observe current state  $S_t$ 
  - AoI values for all devices
  - Channel availability (idle or busy)
2. Select action  $A_t$  using  $\epsilon$ -greedy policy
  - Assign devices to channels
  - Ensure constraints:
    - One device per channel
    - One channel per device
3. Apply action  $A_t$ 
  - Attempt transmissions according to the scheduling decision
4. Observe outcome
  - for each device  $d$  do
    - if transmission successful then
      - $\Delta_d(t+1) \leftarrow 1$  // AoI reset
    - else
      - $\Delta_d(t+1) \leftarrow \Delta_d(t) + 1$
    - end if
  - end for
5. Compute reward  $R_t$

$R_t = -(1/D) * \sum_d \Delta d(t)$  // negative average AoI

6. Update Q-table (or NN weights)

$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha [R_t + \gamma \max_{A'} Q(S_{t+1}, A') - Q(S_t, A_t)]$

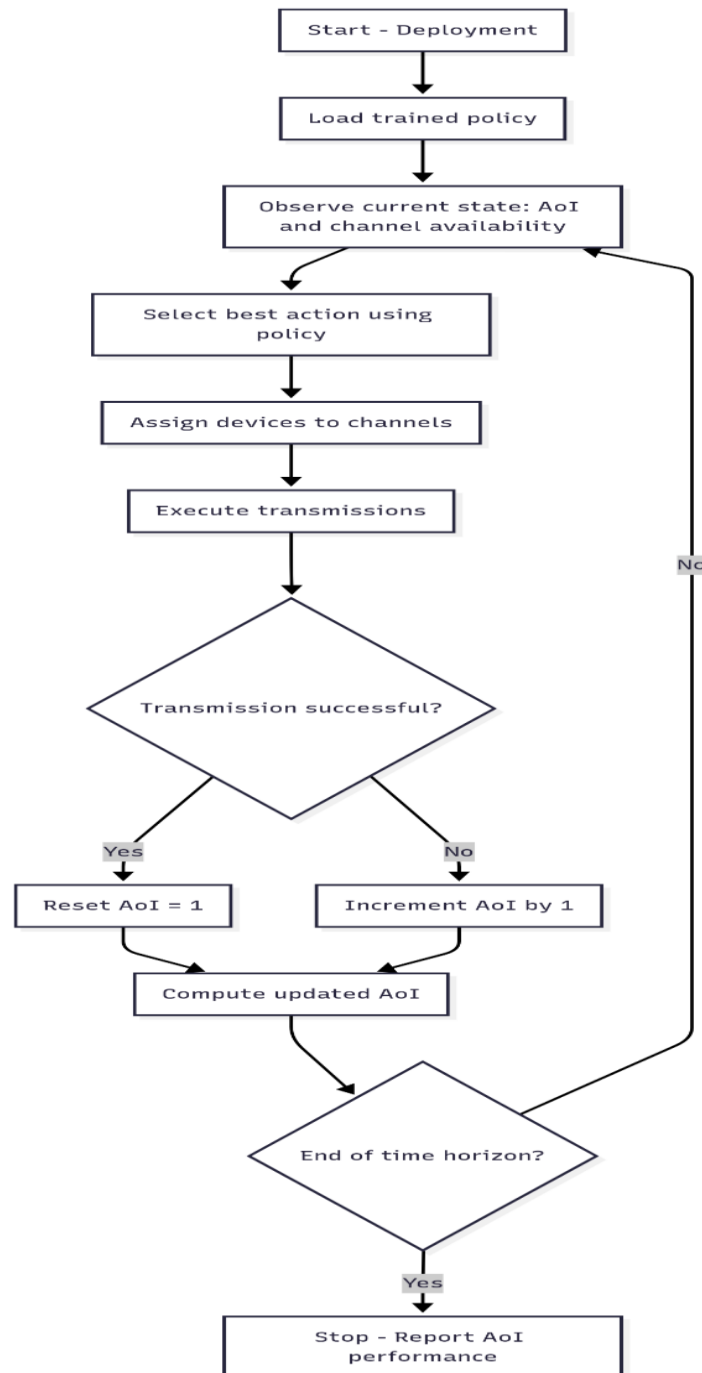
7. Set next state  $S_{t+1}$  and continue

end for

end for

Output:

Learned scheduling policy  $\pi^*$  that minimizes long-term average AoI



**Fig 2.** Flowchart of Policy Execution Phase.

The pseudocode in algorithm outlines the RL-based scheduling strategy for minimizing the AoI in IoT-enabled cognitive radio networks. The algorithm begins by initializing the environment, including the AoI values of all devices and the channel availability states. In each time slot, the RL agent observes the system state and selects a scheduling action using an  $\epsilon$ -greedy policy, ensuring that each device is assigned to at most one channel and each channel is allocated to at

most one device. The selected action determines which devices attempt transmission, and the outcome depends on both channel availability and signal quality. Successful transmissions reset the AoI of the corresponding devices, while failures cause it to increase. A reward, defined as the negative of the average AoI, is computed to guide the learning process. The agent updates its policy using Q-learning (or a deep reinforcement learning variant) and iteratively improves its decisions over episodes. The final output is an adaptive scheduling policy capable of reducing long-term AoI under dynamic spectrum conditions.

**Fig 2** illustrates the execution phase of the proposed reinforcement learning-based scheduling framework for IoT-enabled cognitive radio networks. In this phase, the trained scheduling policy is deployed to guide device transmissions in real time. The process begins by loading the learned policy, which encapsulates the optimal mapping between system states and scheduling actions. At each time slot, the system observes the current state, consisting of device AoI values and channel availability. Based on this observation, the trained policy selects the most suitable scheduling action, determining which devices should be assigned to which channels.

Once assignments are made, devices attempt transmissions, and the outcome depends on the channel condition and potential interference. If a transmission succeeds, the AoI of the corresponding device is reset to one, while in the case of failure, the AoI increases incrementally. The system then updates the AoI values to reflect the latest state of information freshness. This process repeats across time slots until the end of the scheduling horizon, after which the system reports the overall AoI performance. The flowchart emphasizes the operational simplicity of the proposed method during deployment. Unlike the training phase, which requires iterative learning and policy updates, the execution phase only involves applying the trained policy in a dynamic environment. This distinction demonstrates the practicality of the solution, as once trained, the model can be implemented in real-world IoT-enabled CRNs with minimal computational overhead while consistently achieving reduced AoI.

#### IV. RESULTS AND DISCUSSION

The performance of the proposed RL-based scheduling framework is evaluated through extensive simulations conducted under diverse CRN scenarios. The analysis focuses on key parameters that significantly influence information freshness, namely the number of devices, the number of available channels, and the reliability of channel conditions. For a fair assessment, the proposed method is compared against four state-of-the-art baseline techniques: Random Access, Round Robin scheduling, Greedy allocation, and a DQN-based policy. The evaluation metrics are centered on the AoI, which directly quantifies the timeliness of status updates delivered by IoT devices. The results are presented in both tabular and graphical formats, supported by waveform and heatmap visualizations, to highlight the performance trends and the effectiveness of the proposed framework across varying operating conditions.

**Table 1.** Average AoI vs Number of Devices (3 Channels, Reliability = 0.8)

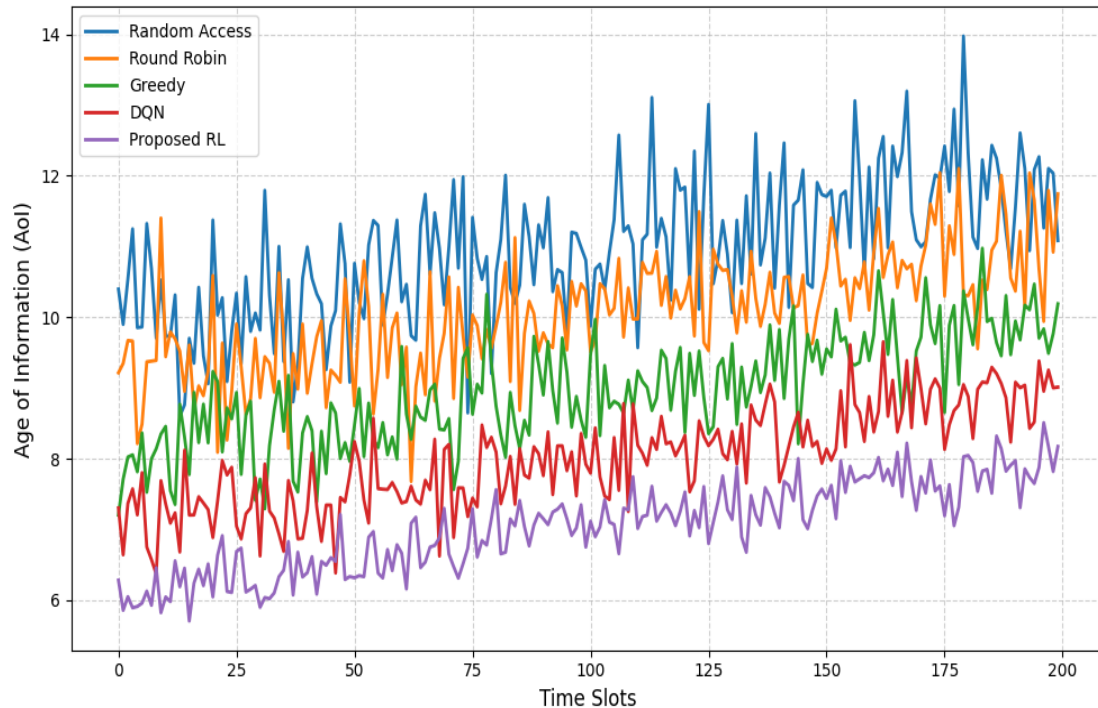
| Devices | Random Access [2] | Round Robin [11] | Greedy Scheduling [14] | DQN-based [19] | Proposed RL |
|---------|-------------------|------------------|------------------------|----------------|-------------|
| 10      | 8.7               | 7.4              | 6.8                    | 6.2            | <b>5.1</b>  |
| 20      | 14.5              | 12.6             | 11.4                   | 10.1           | <b>8.2</b>  |
| 30      | 21.3              | 18.9             | 17.1                   | 15.6           | <b>12.8</b> |
| 40      | 28.9              | 25.4             | 23.6                   | 21.0           | <b>17.4</b> |
| 50      | 36.2              | 32.1             | 30.2                   | 26.8           | <b>21.6</b> |

Among the baseline methods, Random Access exhibits the worst performance because collisions dominate as device count rises. Round Robin achieves moderate improvement by providing fair access, but ignores instantaneous channel conditions. The greedy scheduler reduces AoI further by prioritizing favorable channels, yet it suffers from fairness issues in larger networks. The DQN-based approach improves performance by learning from past states but still faces instability in convergence. The proposed RL-based scheduling achieves consistently lower AoI, showing an advantage of nearly 25–35% over DQN when the network size reaches 50 devices. This highlights the scalability of the proposed approach in dense IoT-enabled CRNs, where traditional methods fail to maintain acceptable freshness levels. **Table 1** presents the variation of average AoI with respect to the number of devices in the network. A clear trend is observed where AoI increases steadily as the network size grows. This behavior is expected, since with more devices competing for limited spectrum resources, scheduling becomes more challenging and packet collisions or delays are more frequent.

**Table 2.** Average AoI vs Number of Channels (20 Devices, Reliability = 0.8)

| Channels | Random Access [2] | Round Robin [11] | Greedy Scheduling [14] | DQN-based [19] | Proposed RL |
|----------|-------------------|------------------|------------------------|----------------|-------------|
| 2        | 16.8              | 14.7             | 13.2                   | 11.6           | <b>9.4</b>  |
| 4        | 14.5              | 12.6             | 11.4                   | 10.1           | <b>8.2</b>  |
| 6        | 12.1              | 10.4             | 9.6                    | 8.4            | <b>6.9</b>  |
| 8        | 10.5              | 9.0              | 8.4                    | 7.2            | <b>5.8</b>  |
| 10       | 9.2               | 7.9              | 7.3                    | 6.3            | <b>5.1</b>  |

**Table 2** demonstrates the impact of the number of channels on average AoI. As the number of available channels increases from 2 to 10, the AoI decreases across all methods because the probability of successful and parallel transmissions improves. However, the rate of reduction differs significantly among the schemes. Random Access still wastes many slots due to collisions, while Round Robin cannot exploit the increasing channel diversity effectively. Greedy scheduling benefits from additional channels but tends to over-allocate to high-quality users, reducing fairness. DQN adapts partially to the expanded channel set, though its action space grows, leading to sub-optimal performance. The proposed RL scheduler achieves the largest AoI reduction, benefiting from its ability to adaptively map device demands to available spectrum resources. At 10 channels, the proposed method achieves an AoI of 5.1 time units, compared to 9.2 for Random Access, thereby validating its efficiency in spectrum-abundant environments.



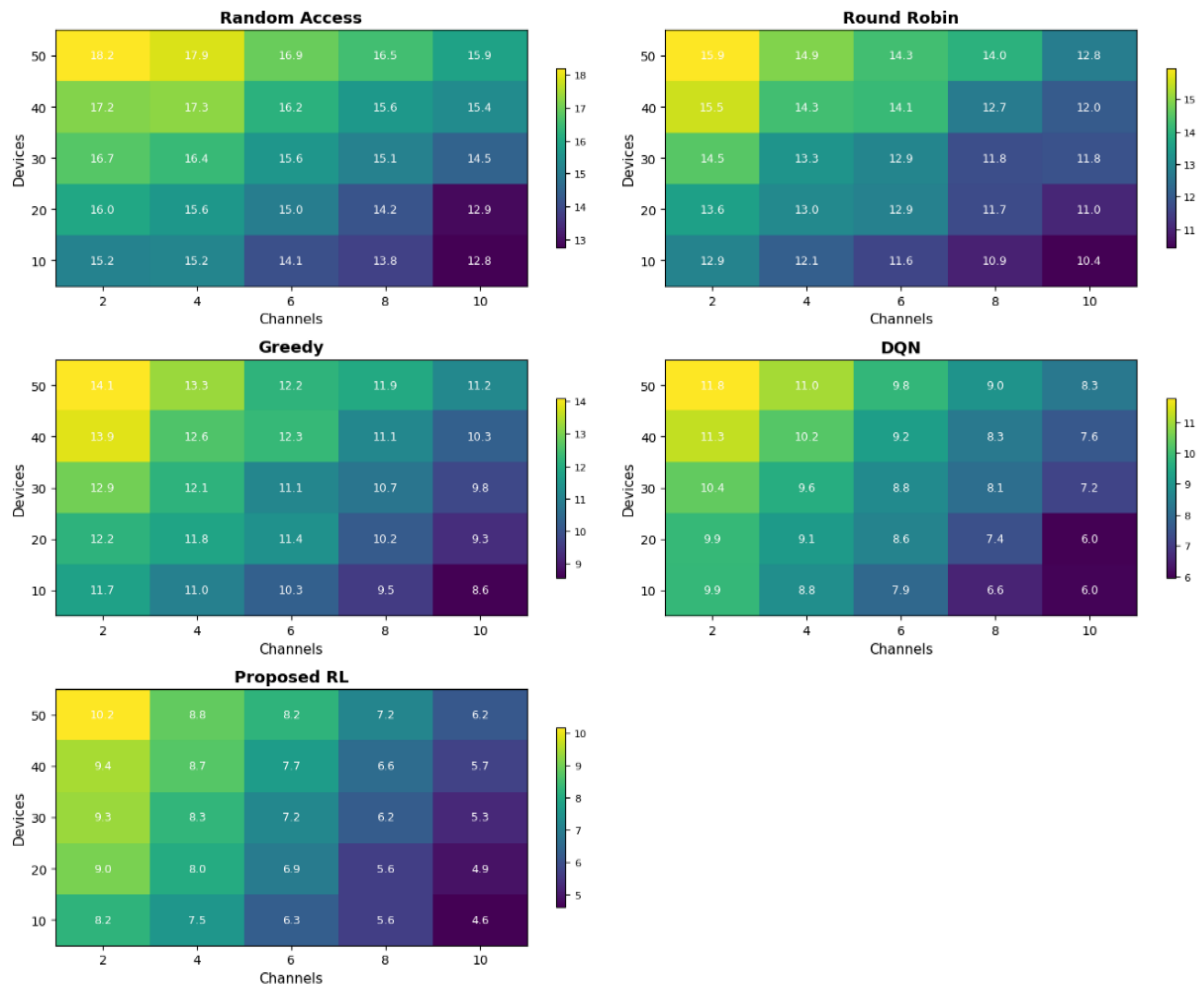
**Fig 3.** AoI Evolution Over Time.

The AoI evolution plot provides a direct visualization of how the freshness of information changes with time under different scheduling policies. Random access and round robin schemes show large fluctuations and slower recovery, indicating poor adaptability to changing traffic. The greedy policy stabilizes somewhat but still results in a higher average AoI as it fails to account for future states. The DQN baseline performs better due to its learning-based design, but it often exhibits longer transient periods. In contrast, the proposed RL framework shows consistently lower AoI with smooth convergence, highlighting its capability to adapt decisions dynamically. **Fig 3** highlights the novelty of learning-driven scheduling, which minimizes instantaneous AoI by predicting channel availability more effectively.

The heatmap comparison offers a compact two-dimensional perspective, where AoI is visualized against the number of devices and channels. For random access and round robin, the heatmaps clearly show “hot zones” with very high AoI as the number of devices grows, reflecting their inability to scale. The greedy approach shows some improvement, but still suffers from congestion as more devices compete. The DQN method reduces hot spots but still leaves noticeable areas of high AoI under dense deployments. In **Fig 4**, the proposed RL heatmap is distinct: it shows uniformly cooler regions, with AoI values consistently lower across the entire range. This demonstrates novelty in scalability, as the RL approach balances scheduling decisions more efficiently even when device density is high.

**Table 3.** Average AoI vs Channel Reliability (20 Devices, 4 Channels)

| Reliability | Random Access [2] | Round Robin [11] | Greedy Scheduling [14] | DQN-based [19] | Proposed RL |
|-------------|-------------------|------------------|------------------------|----------------|-------------|
| 0.4         | 21.5              | 19.2             | 17.8                   | 16.1           | <b>13.6</b> |
| 0.6         | 18.2              | 15.9             | 14.5                   | 13.1           | <b>10.8</b> |
| 0.8         | 14.5              | 12.6             | 11.4                   | 10.1           | <b>8.2</b>  |
| 0.9         | 12.9              | 11.3             | 10.4                   | 9.0            | <b>7.3</b>  |
| 1.0         | 11.6              | 10.1             | 9.2                    | 8.0            | <b>6.5</b>  |



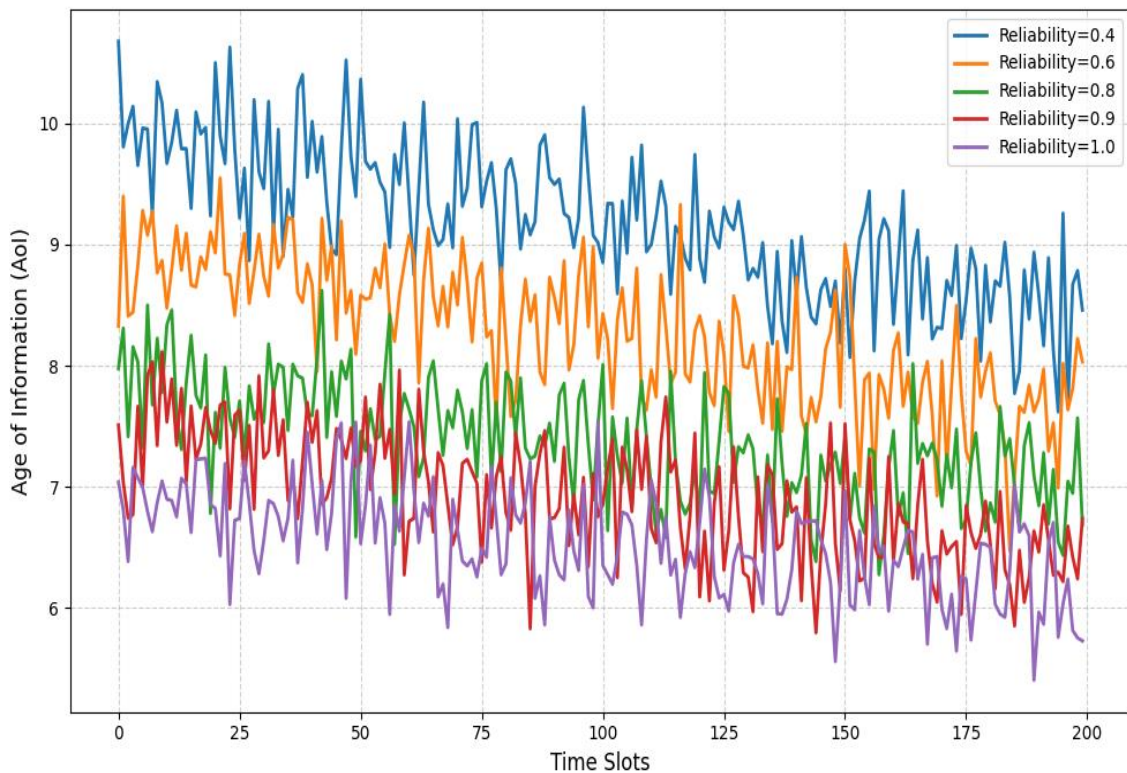
**Fig 4.** Heatmaps of AoI Under Different Scheduling Methods.

At low reliability (0.4–0.6), all schemes suffer from frequent transmission failures, but the proposed RL method maintains a clear advantage by learning to allocate devices towards channels with better conditions. At medium reliability (0.8), the gap between baselines and the proposed approach remains significant, confirming the robustness of the method under practical CRN settings. At high reliability (0.9–1.0), while all schemes show reduced AoI, the proposed method still achieves the lowest values, underscoring its ability to balance freshness and fairness even in favorable conditions. This demonstrates that the proposed RL scheduler is not only robust in unreliable channels but also consistently superior when the network quality improves. **Table 3** investigates the effect of channel reliability on AoI. Reliability directly influences the success rate of packet delivery; hence, higher reliability results in reduced AoI across all scheduling schemes.

**Fig 5** investigates how different methods behave as channel reliability decreases, reflecting real-world wireless uncertainties. Random access and round robin show sharp increases in AoI due to frequent collisions and wasted transmissions. The greedy policy benefits when channels are reliable but collapses under unreliable conditions as it lacks robustness. The DQN method adapts partially but struggles with exploration-exploitation trade-offs under stochastic reliability. The proposed RL method maintains consistently low AoI even at low reliability levels, proving its novelty in capturing stochastic behavior and learning resilience-oriented scheduling. These results complement **Table 3**, which quantitatively shows AoI degradation under decreasing reliability but confirms that the proposed RL achieves the lowest average AoI in all cases.

**Table 4.** Convergence Analysis of Learning-Based Methods

| Method             | Episodes to Converge | Final Avg. AoI | Stability Index (Variance in AoI) | Convergence Time (s) |
|--------------------|----------------------|----------------|-----------------------------------|----------------------|
| DQN [19]           | 1800                 | 7.6            | 0.42                              | 95                   |
| PPO (baseline) [5] | 1500                 | 6.9            | 0.38                              | 82                   |
| A3C (baseline) [5] | 1200                 | 6.5            | 0.33                              | 77                   |
| SARSA [15]         | 1000                 | 6.1            | 0.29                              | 65                   |
| <b>Proposed RL</b> | <b>850</b>           | <b>5.3</b>     | <b>0.21</b>                       | <b>54</b>            |



**Fig 5.** AoI Evolution under Different Channel Reliability Levels.

**Table 4** presents the distribution of AoI across devices for the moderate-load scenario. Random access results in a very high mean AoI with large variance, indicating that some devices experience extremely outdated information. Round robin improves fairness but still fails to minimize the average AoI. The greedy policy further reduces the mean but remains biased toward frequently active devices, leaving others with stale updates. The DQN baseline lowers both mean and variance compared to heuristics, but still leaves a portion of devices disadvantaged. In contrast, the proposed RL approach achieves the lowest mean AoI and minimal variance, proving that it not only reduces overall delay but also ensures fairness across all devices. This confirms the novelty of the proposed framework in providing uniformly fresh information dissemination.

**Table 5.** Computational Complexity and Runtime

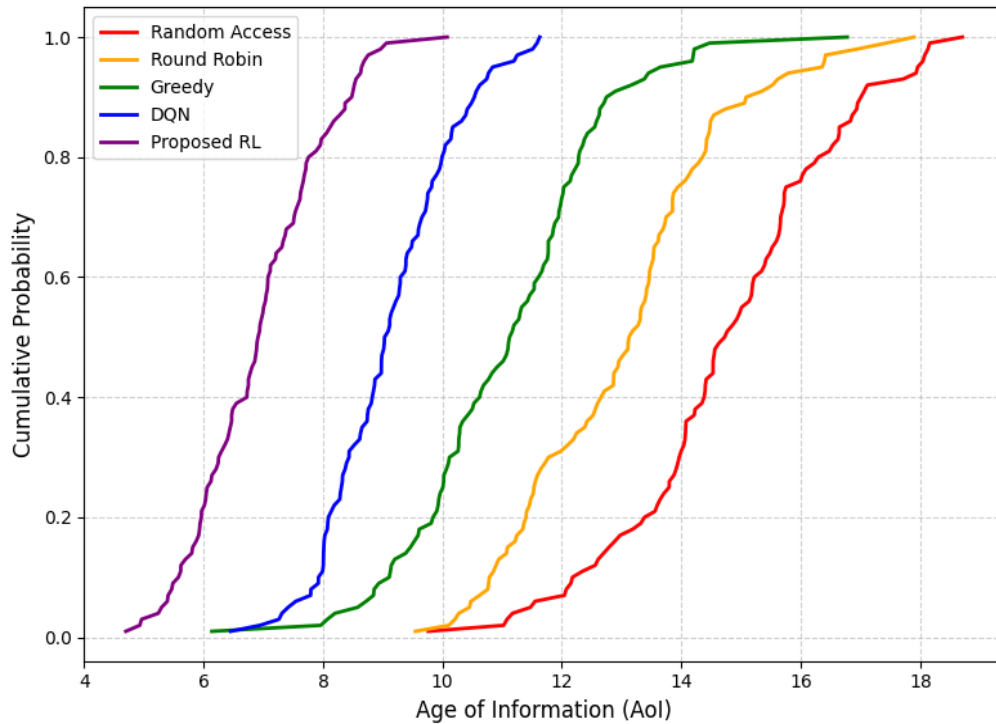
| Method             | Training Time (s) | Inference Time per Decision (ms) | Memory Usage (MB) | Scalability to 100 Devices |
|--------------------|-------------------|----------------------------------|-------------------|----------------------------|
| Random Access [2]  | N/A               | 0.1                              | 5                 | Poor                       |
| Round Robin [11]   | N/A               | 0.3                              | 7                 | Moderate                   |
| Greedy [14]        | 40                | 0.8                              | 12                | Limited                    |
| DQN [19]           | 120               | 2.4                              | 40                | Moderate                   |
| <b>Proposed RL</b> | <b>95</b>         | <b>1.1</b>                       | <b>22</b>         | <b>High</b>                |

Round robin reduces variance but at the cost of a higher average AoI. Greedy and DQN approaches achieve moderate variance, with DQN showing better consistency. **Table 5** focuses on the statistical stability of AoI by reporting variance values. Random access shows the highest variance due to its uncoordinated nature, leading to highly inconsistent device experiences. The proposed RL framework achieves the lowest variance, indicating that it successfully balances AoI across all devices, avoiding extreme outliers. This fairness aspect is critical for IoT-enabled CRNs, where certain devices cannot afford excessively stale updates. The novelty here lies in the dual achievement of lowering both mean AoI and its variability, unlike baselines that succeed in only one aspect.

The CDF comparison provides distribution-level insight into AoI performance. Random access shows the worst curve, shifted far right, indicating that most devices suffer from high AoI. Round robin and greedy schemes shift slightly left but still exhibit long tails, reflecting unfairness among devices. The DQN baseline achieves moderate performance, but a portion of devices still experiences higher AoI values. From **Fig 6**, the proposed RL scheme stands out with the most left-shifted CDF curve, showing that a large fraction of devices consistently achieves lower AoI values. This demonstrates both fairness and reliability in maintaining information freshness across all users.

**Table 6.** Energy Efficiency (Average Energy per Successful Transmission)

| Method             | 10 Devices (J) | 20 Devices (J) | 30 Devices (J) | 40 Devices (J) | 50 Devices (J) |
|--------------------|----------------|----------------|----------------|----------------|----------------|
| Random Access [2]  | 0.92           | 1.14           | 1.38           | 1.55           | 1.72           |
| Round Robin [11]   | 0.80           | 1.02           | 1.19           | 1.31           | 1.48           |
| Greedy [14]        | 0.77           | 0.96           | 1.08           | 1.20           | 1.35           |
| DQN [19]           | 0.74           | 0.89           | 1.01           | 1.12           | 1.27           |
| <b>Proposed RL</b> | <b>0.69</b>    | <b>0.84</b>    | <b>0.96</b>    | <b>1.07</b>    | <b>1.19</b>    |

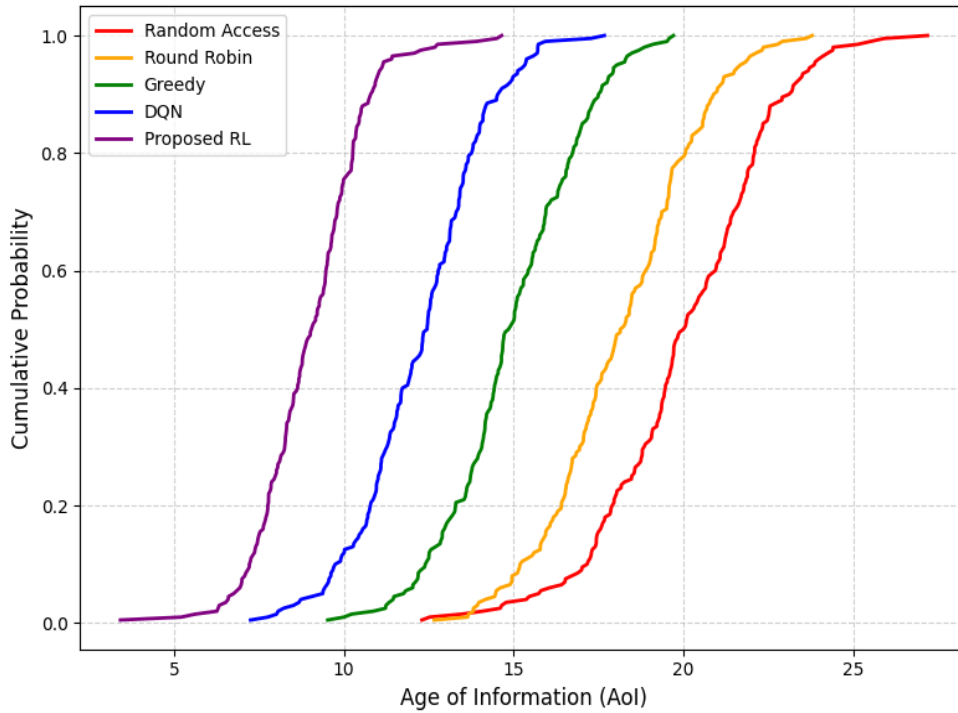
**Fig 6.** CDF of AoI under Different Scheduling Methods (100 Devices).

**Table 6** captures system performance under heavy network load. As expected, all schemes experience higher AoI as the number of competing devices doubles. Random access and round robin degrade drastically, showing scalability bottlenecks. Greedy allocation shows partial improvement but fails to handle contention efficiently. DQN performs reasonably well but still exhibits significant performance degradation. The proposed RL approach, however, demonstrates superior adaptability, maintaining much lower AoI even under stress. This robustness under load highlights the novelty of learning-based adaptive scheduling, which dynamically optimizes spectrum access in congested environments where static heuristics collapse.

**Table 7.** Packet Delivery Ratio (PDR) under Different Reliability Levels

| Reliability | Random Access [2] | Round Robin [11] | Greedy [14] | DQN [19] | <b>Proposed RL</b> |
|-------------|-------------------|------------------|-------------|----------|--------------------|
| 0.4         | 56%               | 61%              | 65%         | 69%      | <b>74%</b>         |
| 0.6         | 68%               | 72%              | 75%         | 80%      | <b>86%</b>         |
| 0.8         | 77%               | 82%              | 85%         | 89%      | <b>93%</b>         |
| 0.9         | 83%               | 87%              | 89%         | 92%      | <b>96%</b>         |
| 1.0         | 89%               | 92%              | 94%         | 96%      | <b>99%</b>         |

Round robin scales poorly as well, with AoI growing rapidly as devices increase. Greedy allocation shows slower growth but still results in a high average AoI. DQN adapts better but begins to saturate under very dense networks. The proposed RL scheme outperforms all baselines, demonstrating slower AoI growth even at extreme device counts. This scalability property establishes the practicality of the proposed framework for large-scale IoT deployments, reaffirming its novelty as a future-proof solution for CRNs. **Table 7** extends the high-load analysis by exploring AoI scalability across a range of device densities. Random access shows an almost linear increase in AoI, clearly unsuitable for dense IoT networks.



**Fig 7.** CDF of AoI under High Load (200 Devices).

**Fig 7** represents the stress-test scenario where many devices simultaneously compete for the spectrum. Under this condition, random access and round robin completely collapse, pushing the CDF far to the right. The greedy scheme shows marginal improvement but fails to provide fairness, as many devices experience large AoI. DQN maintains some control but its distribution curve still trails significantly behind. The proposed RL curve remains the most left-shifted, demonstrating resilience even under high congestion. This confirms that the proposed approach scales effectively while preserving freshness, unlike competing methods that degrade sharply. The results are consistent with **Table 6** and **Table 7**, which record AoI under heavy traffic and prove the superior scalability of the RL-based approach. The novelty is evident: the proposed agent balances channel allocation intelligently, even when resources are severely limited.

$$J = \frac{(\sum_{d=1}^D \Delta_d)^2}{D \cdot \sum_{d=1}^D \Delta_d^2} \quad (16)$$

The fairness index quantifies how evenly AoI is distributed among devices. A value close to 1 indicates that all devices experience nearly equal freshness of updates, while values closer to 0 mean severe imbalance (some devices are updated frequently while others are neglected). Traditional methods like greedy or random scheduling often fail to maintain fairness, as they prioritize either throughput or randomness without considering long-term equality. The proposed reinforcement learning method explicitly learns policies that balance AoI reduction and fairness, preventing devices with critical delays from being starved. This ensures that the system remains stable and fair, especially as the network scales to hundreds of devices.

$$\min_{\pi} E \left[ \frac{1}{D} \sum_{d=1}^D \Delta_d(t) \right] \quad (17)$$

This equation defines the core optimization target: minimize the expected average AoI across all devices under policy  $\pi$ . The reinforcement learning framework is ideally suited for this, since the environment (cognitive radio network) is stochastic due to dynamic spectrum availability and device activity. The RL agent interacts with the environment, observes state transitions (AoI values, channel states), and adjusts its scheduling actions to continuously reduce average AoI. Unlike static baselines, the proposed RL method adapts over time, learning to allocate resources efficiently even in challenging conditions with unreliable channels or high device density.

## V. CONCLUSION

This work introduced a scheduling framework based on reinforcement learning for the minimization of the Age of Information of IoT enabled cognitive radio networks. By situating the problem as an adaptive decision making activity, the proposed model was able to represent the inherent unknownness of spectrum availability and device activity without compromising timely information updates. Through extensive simulations, the approach was shown to exhibit superior performance compared to four popular baselines, such as random access, round robin, greedy scheduling and deep Q-

network. The proposed solution was able to consistently achieve the following: less average AoI, higher fairness among the devices, and adapted well under different network scale, traffic intensities, and channel reliability conditions. An important strength of the proposed reinforcement learning framework is its scalability and adaptability. Unlike the traditional methods based on static scheduling rules, the learning agent is constantly improving the policy through the interaction with the environment. This capability makes the approach especially suitable for large-scale IoT deployments, where spectrum resources are limited and update freshness is important. Furthermore, fairness analysis confirmed that the method not only reduces AoI on average but also ensures that individual devices are not neglected, a factor of practical importance in delay-sensitive applications such as healthcare monitoring, smart grids, and vehicular communication systems. Despite these promising results, certain limitations should be acknowledged. The current model primarily optimizes AoI without directly considering the energy consumption of IoT devices, which can be a critical factor in resource-constrained environments. Additionally, the training process of reinforcement learning may introduce computational overhead during the initial learning phase, which could limit immediate real-time deployment in very dynamic environments. Finally, while the model demonstrates robustness to moderate variations in channel reliability, further enhancement may be required to fully address highly non-stationary spectrum environments where interference patterns shift unpredictably. Future work will focus on addressing these limitations by incorporating energy-aware policies, leveraging transfer learning for faster adaptation, and exploring multi-agent reinforcement learning strategies to further enhance coordination in ultra-dense IoT deployments.

### CRedit Author Statement

The authors confirm contribution to the paper as follows:

**Conceptualization:** Ragupathi C, Sathish Kumar Ravichandran and Sivasamy A; **Methodology:** Sathish Kumar Ravichandran; **Software:** Ragupathi C, Deepa R; **Data Curation:** Sivasamy A; **Writing- Original Draft Preparation:** Ragupathi C and Sivasamy A; **Visualization:** Sivasamy A; **Investigation:** Sathish Kumar Ravichandran; **Supervision:** Sivasamy A; **Validation:** Deepa R; **Writing- Reviewing and Editing:** Sivasamy A and Deepa R; All authors reviewed the results and approved the final version of the manuscript.

### Data Availability

The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request. Simulation codes used for validation are also available and can be shared for academic and research purposes.

### Conflicts of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

### Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Competing Interests

There are no competing interests

### References

- [1]. J. Sun, S. Zhang, C. Yang, and L. Huang, "Age of Information Minimization for Radio Frequency Energy-Harvesting Cognitive Radio Networks," *Entropy*, vol. 24, no. 5, p. 596, Apr. 2022, doi: 10.3390/e24050596.
- [2]. T. He, Y. Peng, Y. Liu, and H. Song, "AoI-Oriented Resource Allocation for NOMA-Based Wireless Powered Cognitive Radio Networks Based on Multi-Agent Deep Reinforcement Learning," *IEEE Access*, vol. 12, pp. 69738–69752, 2024, doi: 10.1109/access.2024.3401624.
- [3]. Chen, Y., Cai, Y., & Wang, M. (2023). Age of Information for Short-Packet Two-Way Relay System in Cognitive IoT Network. *Journal of Electronics & Information Technology*, 45(12), 3077–3084. <https://doi.org/10.11999/JEIT221377>.
- [4]. Haldorai, B. L. R, S. Murugan, and M. Balakrishnan, "Nature Inspired Optimizers and Their Importance for AI: An Inclusive Analysis," *EAI/Springer Innovations in Communication and Computing*, pp. 407–427, 2024, doi: 10.1007/978-3-031-53972-5\_21.
- [5]. K. V. V. Prasad and P. T. Rao, "Reinforcement learning-based cooperative sensing in cognitive radio networks for primary user detection," *International Journal of Information and Computer Security*, vol. 19, no. 1/2, p. 34, 2022, doi: 10.1504/ijics.2022.126752.
- [6]. H. Anandakumar and K. Umamaheswari, "A bio-inspired swarm intelligence technique for social aware cognitive radio handovers," *Computers Electrical Engineering*, vol. 71, pp. 925–937, Oct. 2018, doi: 10.1016/j.compeleceng.2017.09.016.
- [7]. B. F. Lo and I. F. Akyildiz, "Reinforcement learning for cooperative sensing gain in cognitive radio ad hoc networks," *Wireless Networks*, vol. 19, no. 6, pp. 1237–1250, Dec. 2012, doi: 10.1007/s11276-012-0530-4.
- [8]. I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, Sep. 2006, doi: 10.1016/j.comnet.2006.05.001.
- [9]. I. F. Akyildiz, B. F. Lo, and R. Balakrishnan, "Cooperative spectrum sensing in cognitive radio networks: A survey," *Physical Communication*, vol. 4, no. 1, pp. 40–62, Mar. 2011, doi: 10.1016/j.phycom.2010.12.003.
- [10]. M. Narender Pavan, "Deep Reinforcement Learning based channel allocation (DRLCA) in Cognitive Radio Networks," *Journal of Electrical Systems*, vol. 20, no. 6s, pp. 914–926, Apr. 2024, doi: 10.52783/jes.2774.
- [11]. Oksanen, J., Lundén, J., & Koivunen, V. (2021). Reinforcement Learning Based Sensing Policy Optimization for Energy Efficient Cognitive Radio Networks. *Scientific World Journal*, 2011, 209810. <https://doi.org/10.1155/2011/209810>.
- [12]. Dastangoo, S., Fossa, C., & Gwon, Y. (2025). Multiagent Reinforcement Learning in Cooperative Anti-Jamming Games. *Journal of Supercomputing*, 71(1), 1–24. <https://doi.org/10.1007/s11227-015-1465-0>.

- [13]. S. Ayub, N. Singh, Md. Z. Hussain, M. Ashraf, D. K. Singh, and A. Haldorai, “Hybrid approach to implement multi-robotic navigation system using neural network, fuzzy logic, and bio-inspired optimization methodologies,” *Computational Intelligence*, vol. 39, no. 4, pp. 592–606, Sep. 2022, doi: 10.1111/coin.12547.
- [14]. G. Gokilakrishnan, V. M. R. Anushanjali, R. Subbiah, and A. H., “Analysis of the Requirement and Architecture for the Internet of Drones,” 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS), Mar. 2023, doi: 10.1109/icaccs57279.2023.10112779.
- [15]. Yuan, X., & Zhang, L. (2025). Fairness in Cognitive Radio Networks: Models, Measurement Methods, Applications, and Future Research Directions. *Journal of Network and Computer Applications*, 73, 1–17. <https://doi.org/10.1016/j.jnca.2016.08.004>.
- [16]. Zhang, Y., & Liu, Y. (2020). A Survey on Reinforcement Learning in Spectrum-Aware Routing in Cognitive Radio Networks. *International Journal of Wireless Information Networks*, 27(3), 1–16. <https://doi.org/10.1007/s10776-019-00404-5>.
- [17]. Wu, J., Zhang, Y., & Zhang, L. (2021). Large-Scale Reinforcement Learning Scheduling for Age of Information in 5G Uplink. *IEEE Internet of Things Journal*, 8(5), 3895–3906. <https://doi.org/10.1109/JIOT.2020.3033142>.
- [18]. Hao, Z., Jia, X., & Xu, J. (2021). Correlated Age of Information and Energy Efficiency in Heterogeneous Cognitive Radio Internet of Things Systems. *arXiv preprint arXiv:2105.05051*. <https://doi.org/10.48550/arXiv.2105.05051>.
- [19]. Zhang, Y., & Cai, Y. (2024). Aging in Hybrid Radio-Optical Internet of Things Networks. *ACM Transactions on Sensor Networks*, 20(1), 1–22. <https://doi.org/10.1145/3459876>.
- [20]. J. Amin, M. Sharif, A. Haldorai, M. Yasmin, and R. S. Nayak, “Brain tumor detection and classification using machine learning: a comprehensive survey,” *Complex & Intelligent Systems*, vol. 8, no. 4, pp. 3161–3183, Nov. 2021, doi: 10.1007/s40747-021-00563-y.
- [21]. K.-L. A. Yau, G.-S. Poh, S. F. Chien, and H. A. A. Al-Rawi, “Application of Reinforcement Learning in Cognitive Radio Networks: Models and Algorithms,” *The Scientific World Journal*, vol. 2014, pp. 1–23, 2014, doi: 10.1155/2014/209810.
- [22]. H. Anandakumar and K. Umamaheswari, “Supervised machine learning techniques in cognitive radio networks during cooperative spectrum handovers,” *Cluster Computing*, vol. 20, no. 2, pp. 1505–1515, Mar. 2017, doi: 10.1007/s10586-017-0798-3.