Energy-Efficient Deep Q-Learning Framework for Federated-inspired Cooperative Control in Overlapping Wi-Fi Networks*

Jin-Min Lee, Hye-Yeon Shim and Il-Gu Lee†

Sungshin Women's University, Seoul 02844, Korea
{csewa56579, holeman3838, iglee19} @gmail.com

Abstract

Overlapping Basic Service Set (OBSS) environments, where multiple wireless networks operate on the same or adjacent channels, suffer from interference that degrades network performance and can enable jamming attacks. Existing approaches rely on predefined models or focus only on TX power control, limiting their ability to adapt to dynamic network conditions and degrading Spatial Reuse (SR) efficiency. This paper proposed a Deep Q-Network (DQN)-based framework that jointly controls TX power and RX sensitivity to mitigate interference in OBSS environments. The framework supports both centralized and distributed architectures, with the distributed approach inspired by federated learning principles to share learned parameters among access points. The proposed method jointly optimizes network performance and energy efficiency through direct interaction with the network environment without requiring pre-collected training data.

Keywords: Overlapping Basic Service Set, Federated-Inspired Coordination, Energy Efficiency, Deep Q-Learning

1 Introduction

The recent rapid growth of Internet of Things (IoT) devices and wireless network equipment has led to the frequent occurrence of Overlapping Basic Service Set (OBSS) environments, where multiple wireless networks share the same or adjacent channels within the same space [1]. In these environments, the performance degradation due to channel interference increases significantly. Furthermore, attackers can intentionally generate interference signals or unnecessary traffic, resulting in jamming attacks that threaten the network availability and reliability [2]. Existing OBSS interference mitigation techniques primarily avoid interference issues by blocking certain links or applying time-division methods. However, these approaches have limitations: they degrade the overall network throughput and Spatial Reuse (SR) rates [3]. They often focus solely on TX power control (on sender side) or rely on predefined probability models, thereby failing to respond effectively to real-time changes in the network environment or dynamic traffic patterns. Furthermore, they do not consider controlling the RX

* Proceedings of the 9th International Conference on Mobile Internet Security (MobiSec'25), Article No. W8, December 16-18, 2025, Sapporo, Japan.

† Corresponding author

sensitivity (on receiver side), which can also affect the interference. Therefore, this study views the OBSS environment as a resource to be managed efficiently, and not merely as a constraint to avoid. This study proposed a Deep Q-Network (DQN)-based federated-inspired cooperative control framework that jointly controls TX power and RX sensitivity to mitigate interference in OBSS environments through centralized and distributed architectures. This study presented the design of centralized and distributed architectures for comparative analysis. In the centralized approach, a central Access Point (AP) collects network-wide information from all associated nodes and determines optimal control parameters. In contrast, the distributed approach allows each AP to make decisions autonomously using the state information collected only within its own coverage area. The main contributions of this study are as follows.

• A DQN-based reinforcement learning framework for joint control of TX power and RX sensitivity in OBSS environments is proposed.

• Centralized and distributed DQN architectures that adaptively respond to dynamic interference environments are proposed.

• Federated-inspired coordination mechanisms enable cooperative learning among distributed access points.

• Network performance and energy efficiency are jointly optimized through reward function design.

The structure of this paper is as follows: Section 2 analyzes the research related to OBSS interference mitigation and DQN-based network optimization. Section 3 describes the proposed technique. Finally, Section 4 presents conclusions and directions for future research.

2 Related Work

Jung et al. [4] proposed an OBSS packet detection SR technique based on an optimized TX power control to achieve high throughput in OBSS environments. The proposed technique derives the optimal TX power that maximizes the communication success probability through probabilistic geometric analysis and adjusts the clear channel assessment threshold accordingly to reduce interference and increase channel access opportunities. However, this technique has limitations in that it calculates the optimal values based on predefined probability models, making it difficult to adapt flexibly to real-time changes in the network environment or dynamic traffic patterns.

Zhu et al. [5] improved the performance of coordinated SR (CSR) in an IEEE 802.11be environment through TX power adjustment and distributed optimization using adaptive CSR and distributed CSR. However, this study did not address RX sensitivity control or adaptability to real-time environmental changes via ML, thereby limiting the comprehensive optimization of the transmit/receive parameters in dynamic traffic environments.

In addition, Haxhibeqiri et al. [6] proposed a centralized CSR approach to centrally optimize transmit parameters to resolve OBSS interference issues and enhance network throughput. This approach aims to optimize TX power and Modulation and Coding Scheme (MCS) index to avoid interference at the main receiver. However, centralized structures have limited SR efficiency in dynamic environments owing to structural constraints, such as scalability, overhead, and single points of failure. It also has the limitation of focusing solely on TX power without simultaneously considering RX sensitivity joint control.

Wojnar et al. [7] proposed a learning-based scheduling technique using multi-armed bandits (MABs) to optimize the TX power of multiple APs in an IEEE 802.11bn CSR environment. Specifically, they contributed to an 80% throughput improvement using hierarchical MAB (H-MAB) in a centralized manner. However, this study has limitations in terms of interference management, because it does not consider RX sensitivity control.

Yi et al. [8] proposed a Fair AI-STA, a Deep Q-Learning (DQN)-based intelligent station that dynamically adjusts TX power and RX sensitivity for fairness optimization in Wi-Fi networks. However, this study focuses on single-STA fairness optimization rather than multi-AP coordinated control for energy efficiency optimization. Furthermore, it does not address cooperative learning mechanisms that enable parameter sharing among multiple access points.

Previous studies have proposed various approaches to mitigate interference and enhance the SR efficiency in OBSS environments, such as TX power optimization and centralized or distributed parameter control. However, most of these approaches rely on predefined models, making them difficult to adapt flexibly to real-time changes in network environments and dynamic traffic patterns. Furthermore, comprehensive control strategies that jointly consider TX power and RX sensitivity are still lacking and fail to actively incorporate dynamic network factors.

3 DQN-Based Interference Mitigation Framework in OBSS Environments

This section describes a DQN-based reinforcement learning framework for interference mitigation in OBSS environments. Figure 1 illustrates the OBSS interference scenario addressed in this study. When AP1 communicates with its associated STA, AP2's transmission range overlaps with AP1's coverage area, causing interference. The proposed framework dynamically adjusts TX power and RX sensitivity of both APs to minimize such interference while maintaining communication quality.

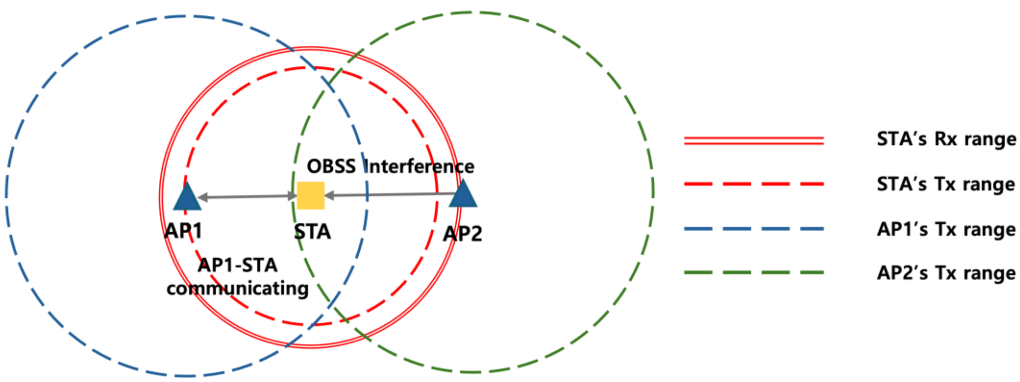

Figure 1: OBSS Interference Scenario

The proposed framework jointly controls TX power and RX sensitivity to optimize network performance and energy efficiency. The proposed DQN agent is designed to learn optimal control policies through direct interaction with the network environment. The DQN agent consists of two neural networks: a Q-network that estimates action values and a target network that provides stable learning targets. The agent stores experiences in a replay buffer and randomly samples mini-batches for learning, which stabilizes training and improves data efficiency. The Q-network estimates the expected cumulative reward for each state-action pair as shown in Equation (1):

$$Q(s,a;\ \theta)\ \approx\ E[\textstyle\sum_t \gamma^t r_t | s_0 = s,\ a_0 = a] \quad (1)$$

where θ represents the network parameters, γ is the discount factor, and r_t is the reward at time t. During training, the Q-network is updated by minimizing the loss function defined in Equation (2):

$$L(\theta) = E\left[\left(r + \gamma \max_{a'} Q(s', a';\ \theta') - Q(s, \alpha;\ \theta)\right)^2\right] \quad (2)$$

where θ' represents the target network parameters, which are periodically copied from the Q-network to stabilize training.

Table 1 shows the state space definition. The state space captures the current network conditions including SINR, throughput, energy efficiency, traffic load, and latency.

State Variable	Description
Node Density ($nodes/m^2$)	Number of nodes per area
Traffic Load ([0–1])	Current network traffic load
Collision Rate ([0–1])	Packet collision rate
Average PLR ([0–1])	Average packet loss rate
Average Latency (ms)	Average packet latency
Average Throughput ($Mbps$)	Average network throughput
Current TX Power (dBm)	Current transmission power
Current RX Sensitivity (dBm)	Current receiver sensitivity
Current CW ($slots$)	Current contention window size
Total Power (mW)	Total power consumption
Energy Efficiency ($Mbps/mW$)	Throughput per unit power
Average SINR (dB)	Average signal-to-interference-plus-noise ratio
SINR Standard (dB)	Standard deviation of SINR
Average Distance (m)	Average distance between nodes

Table 1: State Space Definition

Table 2 shows the action space definition. The action space includes discrete TX power levels within regulatory limits, RX sensitivity levels covering the standard receiver range, and contention window sizes following IEEE 802.11 standard values. The total number of actions is determined by the Cartesian product of these parameter settings.

Action Variable	Description
TX Power (dBm)	Adjustable transmission power
RX Sensitivity (dBm)	Adjustable receiver sensitivity
Contention Window ($slots$)	IEEE 802.11 contention window
Total Actions ($count$)	Combination of all parameter settings

Table 2: Action Space Definition

The reward function jointly optimizes network performance and energy efficiency. Energy efficiency (throughput-to-power ratio) is the primary objective, with additional consideration of packet loss rate, latency, and SINR, as defined in Equation (3):

$$Reward = 100 \times EnergyEfficiency - 10 \times PLR - 0.01 \times Latency + 0.1 \times SINR \quad (3)$$

The coefficients balance the different objectives. The highest weight is assigned to prioritize energy efficiency. Packet loss is penalized to maintain reliability. Latency and SINR receive smaller weights to guide the agent toward better channel conditions without excessively influencing the reward signal. Training uses an epsilon-greedy exploration strategy where the agent selects random actions with probability ε and selects the best action according to the Q-network with probability $1 - \varepsilon$. The epsilon value starts at 1.0 and decays toward 0.01 as training progresses. This allows the agent to explore various parameter combinations initially and exploit learned policies later. To ensure training stability, rewards are clipped to the range [-100, 100] and gradients are clipped to [-10, 10], preventing gradient explosion and training divergence caused by extreme values.

The proposed framework is designed as a federated-inspired cooperative control system and supports both centralized and distributed architectures. In the centralized architecture, a single DQN agent at the central AP collects state information from all APs and stations across the entire network. The central agent determines optimal TX/RX parameters based on the collected global state and distributes them to all nodes. This approach enables optimization considering the entire network.

Algorithm 1 shows the centralized DQN training process. The central agent collects global state information, selects actions using epsilon-greedy exploration, and updates the Q-network using experiences stored in the replay buffer. The target network is periodically updated to stabilize training.

Algorithm 1: Centralized DQN for Joint TX/RX Control

Input: Network state S, action space A, episodes N
Output: Optimal policy π^*
Initialize Q-network $Q(s, a; \theta)$ and target network $Q'(s, a; \theta')$
Initialize replay buffer D with capacity C
for episode = 1 to N do
  Collect global state s from all APs and STAs
  $\varepsilon \leftarrow \varepsilon_{start} \times (\varepsilon_{decay})^{episode}$
  for step = 1 to max_{steps} do
    if $random(\,) < \varepsilon$ then
      $a \leftarrow random\ action\ from\ A$
    else
      $a \leftarrow arg \max_a Q(s, a;\ \theta)$
    end if
    Execute action a for all nodes $(TX\ power, RX\ sensitivity, CW)$
    Observe reward r and next state s'
    $r \leftarrow clip(r, -100, 100)$
    Store transition (s, a, r, s') in D
    Sample mini-batch B from D
    for each transition in B do
      $y \leftarrow r + \gamma \times max_{a'} Q'(s', a';\ \theta')$
      Compute loss $L = (Q(s, a;\ \theta) - y)^2$
      $\nabla\theta \leftarrow clip(\nabla\theta, -10, 10)$
      $\theta \leftarrow \theta - \alpha\nabla\theta$
    end for
    if step mod $C_{target} == 0$ then
      $\theta' \leftarrow \theta$
    end if
    $s \leftarrow s'$
  end for
end for
return $Q(s, a;\ \theta)$

Algorithm 1: Centralized DQN for Joint TX/RX Control

The centralized training process begins by initializing the Q-network, target network, and replay buffer. During each episode, the agent collects global state from all network nodes and selects actions using epsilon-greedy exploration, where the exploration rate ε decays exponentially over time. Selected actions adjust TX power, RX sensitivity, and contention window parameters for all nodes (APs and STAs) across the network. Each experience is stored in the replay buffer, and mini-batches are randomly sampled for learning. The Q-network is updated by minimizing the difference between its predictions and target values computed using the target network. Reward and gradient clipping ensure training stability, while the target network is periodically synchronized with the Q-network. While the centralized architecture enables optimization considering the entire network, it has limitations. All computational burden concentrates at the central AP, creating a single point of failure, and communication overhead and computational complexity increase with network size, limiting scalability. To address these limitations, the distributed architecture is proposed.

In the distributed architecture, each AP operates an independent DQN agent. Each agent observes and controls only nodes within a 40-meter radius, which corresponds to typical Wi-Fi coverage range in indoor environments. Agents learn using local state information, distributing the computational burden and operating independently to resolve single point of failure issues. However, local optimization alone has limitations in reaching globally optimal performance. To overcome this challenge, a federated-inspired coordination mechanism is applied, derived from the parameter aggregation concept of federated learning. This mechanism enables distributed agents to share knowledge while maintaining local autonomy. Each AP agent learns independently but exchanges learned Q-network parameters with other APs at regular intervals (every 10 episodes), balancing learning convergence and communication overhead. The collected parameters are averaged as shown in Equation (4):

$$\theta_{global} = \frac{1}{M}\sum_{i=1}^{M}\theta_i \quad (4)$$

where M is the number of APs and θ_i represents the parameters of agent i. The averaged parameters are then redistributed to all agents. This allows each agent to learn indirectly from network situations experienced by other APs, achieving performance closer to global optimization while maintaining local autonomy. Our approach differs from traditional federated learning by focusing on energy efficiency rather than privacy protection, offloading the learning burden from resource-constrained devices to APs.

Algorithm 2 shows the distributed DQN training with federated-inspired coordination. Each AP trains its local agent independently. Every 10 episodes, the agents share parameters, compute the averaged parameters, and update their local networks with the global parameters.

Algorithm 2: Distributed DQN with Federated-Inspired Coordination

```
Input: Local states S_i for each AP i, episodes N
Output: Federated policy π_fed
for each AP i do
    Initialize local Q-network Q_i(s, a; θ_i)
    Initialize local replay buffer D_i
end for
for episode = 1 to N do
    for each AP i in parallel do
        Collect local state s_i (within 40m radius)
        ε ← ε_start × (ε_decay)^episode
        for step = 1 to max_steps do
            if random() < ε then
                a ← random action from A
```

else
    $a \leftarrow arg \max_{a} Q_i(s_i, a;\ \theta_i)$
end if
Execute action a for local nodes ($TX\ power, RX\ sensitivity, CW$)
Observe reward r_i and next state s'_i
$r_i \leftarrow clip(r_i, -100, 100)$
Store (s_i, a, r_i, s'_i) in D_i
Sample mini-batch B_i from D_i
for each transition in B_i from D_i
    $y_i \leftarrow r_i + \gamma \times max_{a'} Q'_i(s'_i, a';\ \theta'_i)$
    $L_i \leftarrow (Q_i(s_i, a;\ \theta_i) - y_i)^2$
    $\nabla\theta_i \leftarrow clip(\nabla\theta_i, -10, 10)$
    $\theta_i \leftarrow \theta_i - \alpha\nabla\theta_i$
end for
$s_i \leftarrow s'_i$
end for
end for

if episode $mod\ 10\ ==\ 0$ then
    $\theta_{global} \leftarrow \frac{1}{M}\sum_{i=1}^{M}\theta_i$
    for each AP i do
        $\theta_i \leftarrow \theta_{global}$
    end for
end if
end for
return $\{Q_i(s, a;\ \theta_i)\}$ for all i

Algorithm 2: Distributed DQN with Federated-Inspired Coordination

The distributed training process enables parallel learning across multiple APs. Each AP maintains its own Q-network and replay buffer, collecting state information only from nodes within its 40-meter coverage radius. This reduces communication overhead and eliminates the single point of failure. Each AP independently performs local learning cycles following the same procedure as the centralized approach but using only local experiences. The key innovation is the federated-inspired coordination mechanism: every 10 episodes, all APs share their learned parameters, which are aggregated through simple averaging to create a global model. The global parameters are redistributed to all APs, allowing each agent to benefit from diverse network conditions experienced by other APs while maintaining local autonomy.

4 Conclusion

As wireless networks become increasingly dense with IoT devices, OBSS environments suffer from severe interference that degrades network performance and enables jamming attacks. Conventional interference mitigation techniques that rely on link blocking or time-division approaches sacrifice throughput and spatial reuse efficiency. Moreover, existing methods focus only on transmitter-side control or depend on predefined models, limiting their ability to adapt to dynamic network conditions and failing to exploit receiver-side control opportunities. To address these limitations, this paper proposed a DQN-based reinforcement learning framework that jointly controls TX power and RX sensitivity for interference mitigation in OBSS environments. The framework supports both centralized and distributed architectures. The centralized approach provides global optimization using network-wide information, while the distributed approach offers scalability and robustness through local decision-making. Federated-inspired coordination mechanisms enable cooperative learning among distributed agents while maintaining local autonomy. The proposed framework learns optimal control policies through direct interaction with the network environment without requiring pre-collected training data. The reward function jointly optimizes network performance and energy efficiency, and training stability is ensured through reward and gradient clipping mechanisms. Future work includes implementing and evaluating the proposed framework using the ns-3 network simulator, validating scalability across varying network scales, and assessing robustness under diverse traffic patterns and interference conditions.

Acknowledgments. This work is supported by the Ministry of Trade, Industry and Energy (MOTIE) under Training Industrial Security Specialist for High-Tech Industry [grant number RS-2024-00415520] supervised by the Korea Institute for Advancement of Technology (KIAT), the Ministry of Science and ICT (MSIT) under the ICAN (ICT Challenge and Advanced Network of HRD) program [grant number IITP-2022-RS-2022-00156310] and National Research Foundation of Korea (NRF) grant [RS-2025-00518150], and the Information Security Core Technology Development program [grant number RS-2024-00437252] supervised by the Institute of Information & Communication Technology Planning & Evaluation (IITP).

References

Shafique, K., Khawaja, B. A., Sabir, F., Qazi, S., & Mustaqim, M. (2020). Internet of things (IoT) for next-generation smart systems: A review of current challenges, future trends, and prospects for emerging 5G-IoT scenarios. *IEEE Access, 8*, 23022–23040.

Pirayesh, H., & Zeng, H. (2022). Jamming attacks and anti-jamming strategies in wireless networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials, 24*(2), 767–809.

Lanante, L., & Roy, S. (2021). Performance analysis of the IEEE 802.11ax OBSS_PD-based spatial reuse. *IEEE/ACM Transactions on Networking, 30*(2), 616–628.

Jung, J., Baik, J., Kim, Y., Park, H.-S., & Chung, J.-M. (2023). OTOP: Optimized transmission power controlled OBSS PD-based spatial reuse for high throughput in IEEE 802.11be WLANs. *IEEE Internet of Things Journal, 10*(19), 17110–17123.

Zhu, D., Wang, L., Pan, G., & Luan, S. (2025). Two enhanced schemes for coordinated spatial reuse in IEEE 802.11be: Adaptive and distributed approaches. *Computer Networks, 258*, Article 111060.

Haxhibeqiri, J., Jiao, X., Shen, X., Pan, C., Jiang, X., Hoebeke, J., & Moerman, I. (2024). Coordinated spatial reuse for Wi-Fi networks: A centralized approach. In *Proceedings of the IEEE International Workshop on Factory Communication Systems* (pp. 1–6). IEEE.

Wojnar, M., et al. (2025). Coordinated spatial reuse scheduling with machine learning in IEEE 802.11 MAPC networks. *arXiv preprint* arXiv:2505.07278v2.

Yi, P., Cheng, W., Ju, Z., Wang, J., Pan, J., Ouyang, Y., & Zhang, W. (2025). Fair AI-STA for legacy Wi-Fi: Enhancing sensing and power management with deep Q-learning. *arXiv preprint* arXiv:2412.10874.