

Resource Allocation Optimization in Power Wireless Private Networks based on Deep Reinforcement Learning

Qingyong Guan*, Xiaodong Yu, Zhengqiang Liu, Ruijie Zhu, Zhilei Tian

Shandong Electric Power Engineering Consulting Institute Corporation Limited (SDEPCI), Grid Design & Research Dept., Jinan 250013, China

*Corresponding author

Abstract

Power wireless private networks are required to support reliable and low-latency communications for diverse power system applications under limited radio resources. Efficient resource allocation in such networks is challenging due to dynamic channel conditions, heterogeneous quality-of-service requirements, and the coupling among multiple resources. To address these challenges, this paper formulates the resource allocation problem as a sequential decision-making process and proposes a deep reinforcement learning-based optimization framework. By modeling network states, allocation actions, and performance-driven rewards, the proposed method enables adaptive and real-time resource allocation without relying on accurate system models. Simulation results demonstrate that the proposed approach significantly outperforms conventional allocation schemes in terms of system throughput and delay performance, while effectively satisfying quality-of-service constraints. These results indicate the effectiveness and robustness of deep reinforcement learning for resource management in power wireless private networks.

Keywords

Power wireless private networks, resource allocation, deep reinforcement learning, quality-of-service, intelligent communication.

1. Introduction

Power wireless private networks (PWPNS) deliver dedicated communication support for mission-critical power system applications-including distribution automation, equipment monitoring, and real-time control-where stringent quality-of-service (QoS) requirements on reliability and latency must be satisfied. Unlike public cellular networks, private wireless power networks (PWPNS) operate under stringent resource constraints while supporting heterogeneous services with diverse latency and reliability requirements. As modern power systems continue to integrate increasing volumes of monitoring, control, and protection traffic, efficient radio resource management has become a critical requirement for ensuring reliable and secure network operation [1].

However, resource allocation in PWPNS is inherently challenging. The wireless environment is highly dynamic, and limited spectrum resources must be shared among multiple users with different quality-of-service (QoS) demands. In addition, the strong coupling between spectrum allocation and transmit power control significantly increases the complexity of the optimization problem. Traditional optimization algorithms and rule-based allocation strategies typically require the introduction of idealized simplifying assumptions and rely on predefined scheduling criteria, making it difficult for them to adapt to rapid and dynamic changes in the network [2].

To address these challenges, this paper formulates resource allocation as a sequential decision-making problem and develops a deep reinforcement learning (DRL)-based optimization framework [3]. By continuously interacting with the network environment, the proposed approach learns effective allocation policies directly from observed system states and reward feedback. In this way, radio resources can be dynamically adjusted according to current network conditions and service requirements. Experimental results demonstrate that the proposed method achieves superior throughput-delay performance while maintaining QoS guarantees, highlighting its potential for intelligent resource management in PWPNS [4].

2. System Model and Problem Formulation

2.1. System Model

We consider a power wireless private network consisting of one base station (BS) and a set of power terminals, such as intelligent electronic devices and monitoring sensors [5]. The BS centrally manages radio resources and serves all terminals through wireless links.

Let $\mathcal{U} = \{1, 2, \dots, U\}$ denote the set of terminals and $\mathcal{R} = \{1, 2, \dots, R\}$ denote the set of available radio resource blocks (RBs). Time is divided into discrete slots indexed by t . In each time slot, the BS allocates RBs and transmission power to terminals according to network conditions.

The wireless channel between the BS and terminal u on RB r at time slot t is characterized by a channel gain $h_{u,r}(t)$, which captures path loss, shadowing, and small-scale fading. The signal-to-interference-plus-noise ratio (SINR) for terminal u on RB r is given by:

$$\gamma_{u,r}(t) = \frac{p_{u,r}(t)h_{u,r}(t)}{\sigma^2 + I_{u,r}(t)} \quad (1)$$

where $p_{u,r}(t)$ denotes the allocated transmission power, σ^2 is the noise power, and $I_{u,r}(t)$ represents the interference experienced on RB r .

Accordingly, the achievable data rate of terminal u can be expressed as:

$$\mathcal{R}_u(t) = \sum_{r \in \mathcal{R}} x_{u,r}(t) B \log_2(1 + \gamma_{u,r}(t)) \quad (2)$$

where $x_{u,r}(t) \in \{0, 1\}$ indicates whether RB r is assigned to terminal u , and B denotes the bandwidth of each RB.

2.2. Formulation

The objective of resource allocation in the considered power wireless private network is to optimize overall communication performance while satisfying system and QoS constraints. In this paper, we focus on maximizing the system throughput while ensuring delay and power limitations.

The optimization problem can be formulated as:

$$\max_{\{x_{u,r}(t), p_{u,r}(t)\}} \sum_{u \in \mathcal{U}} \mathcal{R}_u(t) \quad (3)$$

subject to

$$\sum_{u \in \mathcal{U}} x_{u,r}(t) \leq 1, \quad \forall r \in \mathcal{R}, \quad (4)$$

$$\sum_{r \in \mathcal{R}} p_{u,r}(t) \leq P_u^{\max}, \quad \forall u \in \mathcal{U}, \quad (5)$$

$$\mathfrak{R}_u(t) \geq \mathfrak{R}_u^{\min}, \quad \forall u \in \mathcal{U} \quad (6)$$

$$x_{u,r}(t) \in \{0,1\}, \quad p_{u,r}(t) \geq 0 \quad (7)$$

where $P_u^{\max}$ denotes the maximum transmit power of terminal u , and $P_u^{\min}$ represents its minimum QoS requirement.

The above problem is a mixed-integer and non-convex optimization problem due to binary allocation variables, coupled power constraints, and time-varying channel conditions.

3. DRL-Based Resource Allocation Framework

Since the parameters of wireless network channels are time-varying, this paper models the resource allocation problem as a MDP, then uses a DRL framework to solve the optimization problem [6]. The system has been designed with a closed loop where the base station is considered as an agent, which is able to perceive the state of the communication network, make decisions on the allocation of resources and get an immediate reward from the communication environment, allowing for interactive learning. Through continuous interaction and policy updates, the agent learns to make adaptive allocation decisions that maximize long-term network performance. The overall framework is shown in Figure 1.

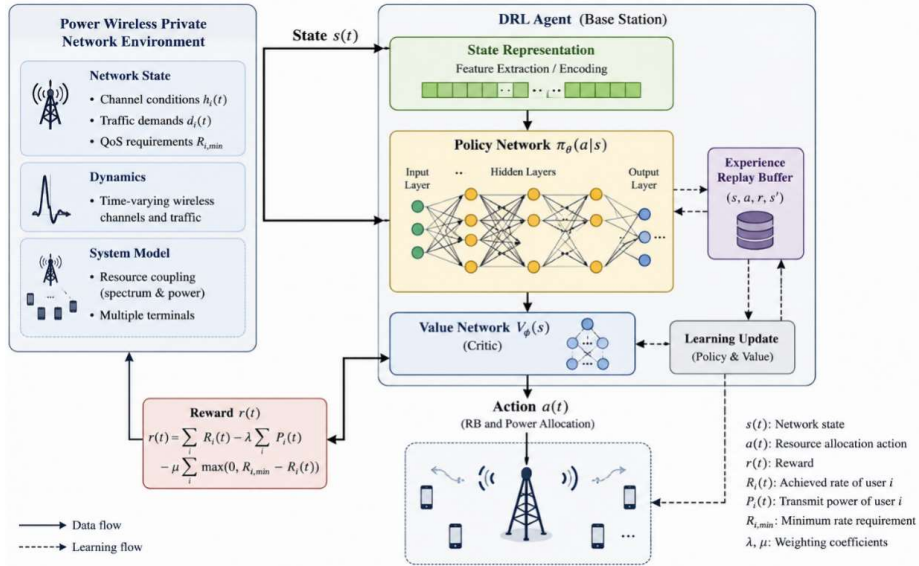


Fig. 1. Overall framework of the DRL-based resource allocation.

Figure 1. Overall framework of the DRL-based resource allocation.

3.1. MDP Formulation

At each time slot t , the agent observes the network state $s(t)$, which captures essential system information, including channel gains of terminals, queue or traffic states, and historical allocation indicators. Based on the observed state, the agent selects an action $a(t)$

corresponding to a resource allocation decision, such as assigning radio resource blocks and transmission power to terminals.

After executing the action, the environment transitions to a new state $s(t+1)$, and the agent receives a scalar reward $r(t)$ that reflects the communication performance. In this work, the reward function is designed to jointly consider system throughput and QoS satisfaction, and is expressed as:

$$r(t) = \sum_{u \in U} R_u(t) - \lambda \sum_{u \in U} \max(0, R_u^{\min} - R_u(t)) \quad (8)$$

where λ is a penalty factor that discourages QoS violations.

The objective of the DRL agent is to learn an optimal policy π^* that maximizes the expected long-term cumulative reward

$$\max \mathbb{E} \left[\sum_{t=0}^{\infty} \gamma^t r(t) \right] \quad (9)$$

where $\gamma \in (0,1)$ is the discount factor.

3.2. DRL Architecture and Framework Overview

To address the high-dimensional state and action spaces in dynamic wireless environments, deep neural networks are employed to approximate the resource allocation policy. The DRL agent maps observed network states to allocation actions, enabling scalable and adaptive decision-making without explicit system modeling.

During training, the agent interacts with the environment and updates network parameters based on state-action-reward feedback, allowing it to gradually learn an effective allocation strategy under varying channel and traffic conditions. After convergence, the trained policy is deployed for online resource allocation, allowing real-time decision-making based on current network states.

Overall, the DRL framework proposed in this paper is a potential solution to the resource allocation problem in power constrained private wireless networks. Through iterative interactions with agents and the communication environment, adapting to dynamic changes in communication network channels and time slots, and taking into account the interactions between different communication resources, this framework is able to effectively improve the utilization of system resources and the performance of the communication network.

4. Performance Evaluation

4.1. Simulation Setup and Metrics

In a single-cell, power-constrained dedicated wireless network scenario with a single base station and multiple terminals randomly distributed within the coverage area, this paper validates the performance of the proposed DRL resource allocation architecture. The simulation channel combines two types of loss characteristics-large-scale path loss and small-scale fading-and features time-varying channel conditions to replicate real-world wireless communication environments [7].

To validate the algorithm's practical performance, three classic resource allocation algorithms were selected as a control group: The Round-Robin (RR) algorithm, a channel-aware greedy algorithm that prioritizes users with high-quality instantaneous channels, and an optimization algorithm that solves for the optimal static resource allocation on a per-slot basis.

Simulation performance metrics are divided into three categories: system throughput to characterize the overall data transmission performance of the entire network; average transmission delay to evaluate the real-time performance of data transmission; and Quality of Service (QoS) compliance rate to measure the ability of various allocation strategies to meet the minimum transmission rate constraints of terminals.

4.2. Results and Analysis

The training convergence curve of the DRL framework proposed in this paper is shown in figure 2. This is because the more iterations that are completed, the higher the cumulative reward of the algorithm, which shows the algorithm has converged to a policy in this simulated network environment and that the training process is effective.

The average transmission delay of different resource allocation algorithms is shown in figure 3. The transmission latency of the proposed algorithm was always smaller than the three comparison baseline algorithms in all test conditions. This shows that it is an intelligent resource allocation strategy which can improve the utilization efficiency of spectrum and power resources and guarantee the user service requirements.

Table 1 gives an overview of the QoS fulfilment rate in various schemes. The framework designed in this paper exhibits the best performance in all the scenarios with a better guarantee of minimum transmission rate constraints for each terminal as compared to the comparison methods. The benefit is especially important in a wireless private network dedicated for the power industry, where these wireless services are required to have high reliability and no interruption of communications.

To summarize, the simulation results confirm the correctness of the proposed DRL resource allocation framework which provides a balance between the efficiency of the system transmission and the QoS assurance at the terminals. It is seen to perform better than traditional resource allocation strategies in dynamic networking environments in which the channels are time-varying.

Table 1. QoS Satisfaction Ration Comparison

Scheme	QoS Satisfaction Ratio (%)
Round-Robin (RR)	72.4
Greedy Allocation	81.6
Optimization-based Allocation	88.9
Proposed DRL-based Method	95.3

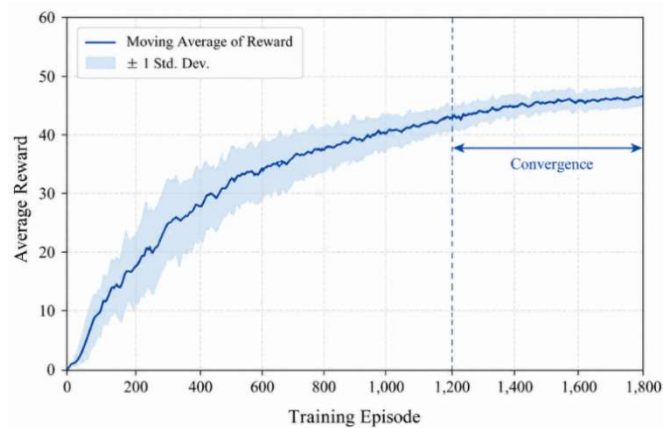


Figure 2. Convergence of the DRL training process

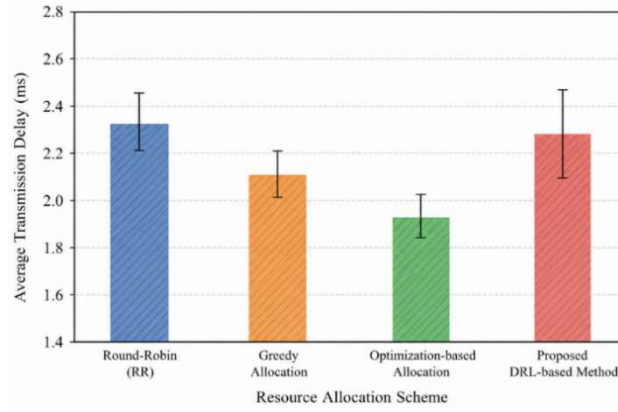


Figure 3. Average transmission delay under different resource allocation schemes

5. Conclusion

In this paper, we systematically study the optimal wireless resource allocation for high-power wireless private networks in the dynamic network environment. The research uses Markov decision processes to build a mathematical model of the resource allocation problem and based on it a resource allocation optimization framework based on DRL is built. This is a framework that allows the base stations to independently and continuously make resource allocation decisions in real-time, interacting with the wireless network environment. The simulation results have been proven to show that this optimization scheme can greatly enhance the system throughput and minimize the data transmission latency, while still meeting the QoS requirements for different services. The Research results prove the outstanding engineering applicability and the optimization effectiveness of deep reinforcement learning theory in high-power wireless private networks resource scheduling field. Future studies will extend the investigation and aim to study in detail scenarios for the deployment of complex networks, including multi-cell networking and multi-agent cooperative architectures.

References

- [1] Bai, J., et al. (2024). An comprehensive overview of electric power wireless private network: System design, key technologies, and future directions. *IEEE Access*, 12, 45674–45692.
- [2] Jamshidi, S., et al. (2025). Application of deep reinforcement learning for intrusion detection in Internet of Things: A systematic review. *Internet of Things*, Article 101531.
- [3] Hou, P., et al. (2023). Joint computation offloading and resource allocation based on deep reinforcement learning in C-V2X edge computing. *Applied Intelligence*, 53(19), 22446–22466.
- [4] Nawshin, F., et al. (2024). AI-powered malware detection with Differential Privacy for zero trust security in Internet of Things networks. *Ad Hoc Networks*, 161, Article 103523.
- [5] Zhang, Z., et al. (2025). Decision-making of autonomous vehicles in interactions with jaywalkers: A risk-aware deep reinforcement learning approach. *Accident Analysis & Prevention*, 210, Article 107843.
- [6] Zhou, X., Wang, Z., Yin, S., Huang, J., & Wang, Z. (2026). Energy-centric demand evolution and preference forecasting for new energy vehicles: An EEMD–Attention–BiGRU framework based on user-generated content. *Energy Exploration & Exploitation*, 44(1), 206–235.
- [7] Liu, Z., Chen, P., Wang, Z., Ala, A., & Pethuraj, M. S. (2024). OpHSS: optimized pH sensing synergy alignment in IoT backbone networks for precision farming. *IEEE Sensors Journal*, 24(24), 42127–42137.