Multi-agent Attention Allocation Method for Optimizing Spectrum Resources of High-Speed Railway Vehicle-to-Ground Communication in Millimeter Wave

Yuan Zhuo
School of Information Engineering, Chongqing Industry and Information Vocational College, Chongqing 401233, China
Corresponding author: Yuan Zhuo (e-mail: cdmxzy2004@163.com)

ABSTRACT Millimeter-wave vehicle-to-ground communication is a key technology for high-capacity and low-latency services in intelligent high-speed railway systems, but high mobility causes time-varying channels, Doppler shifts, path loss, and frequent handovers. To improve spectrum utilization and anti-interference capability, this study proposes a multi-agent attention allocation method for spectrum-resource optimization in millimeter-wave high-speed railway communication. A vehicle-to-ground communication model is constructed with ground millimeter-wave base stations, high-speed train terminals, service links, and differentiated traffic requirements. The spectrum allocation problem is formulated as a nonconvex mixed-integer programming problem that maximizes throughput and spectrum efficiency while minimizing delay under power, channel, and service-priority constraints. Each communication link is modeled as an agent responsible for spectrum sensing and resource allocation, and an attention mechanism dynamically assigns cooperation weights according to channel quality, service priority, and resource demand. A PPO-based reinforcement learning algorithm solves the dynamic allocation problem. Simulations at 28 GHz and 350 km/h show that the proposed method increases throughput to 28.6 Gbps, reduces average delay to 3.2 ms, improves spectrum utilization to 82.3%, and maintains stronger anti-interference performance than game-theory, single-agent, and fixed-weight multi-agent baselines. The method is directly relevant to electromagnetic-wave propagation, millimeter-wave networking, and railway communication optimization.

INDEX TERMS millimeter wave communication, spectrum allocation, vehicle-to-ground communication, multi-agent reinforcement learning, high-speed railway

I. INTRODUCTION

WITH the improvement of my country's high-speed railway network and the continuous improvement of train running speed, emerging services such as automatic driving, on-board high-definition entertainment, and remote monitoring of equipment have put forward higher requirements for the capacity, delay, and reliability of vehicle-ground communication. Millimeter wave communication offers advantages such as ultra-wide frequency bands and low latency. It can effectively mitigate the bandwidth constraints of microwave communication and has become one of the best technical choices for high-speed rail vehicle-ground communication. However, the particularity of the high-speed rail scene has brought severe challenges to millimeter-wave vehicle-ground communication: the high-speed movement of the train leads to time-varying channels and obvious Doppler frequency shifts, the millimeter-wave signal propagation path has a large loss and is easily occluded, and different compartments, There are significant differences in the communication requirements of different services, which makes it difficult to allocate spectrum resources. The current spectrum resource optimization methods mostly use single-agent decision-making or fixed-weight collaboration, which is difficult to adapt to the dynamic and complex scenarios of high-speed rail. There are problems such as low spectrum utilization, weak anti-interference capabilities, and mismatch between resource allocation and business requirements, which restrict millimeter-wave high-speed rail vehicles. Improvement of ground communication system performance. Therefore, it is of great theoretical significance and engineering application value to carry out the research on spectrum resource optimization of millimeter-wave high-speed rail vehicle-to-ground communication under multi-agent attention allocation for improving vehicle-to-ground communication and promoting high-speed rail intelligence.

The analysis basis of this paper mainly includes the actual scene characteristics of high-speed rail vehicle-ground communication, millimeter wave channel propagation characteristics, multi-agent cooperation theory and reinforcement learning solution method, combined with existing research results and actual engineering needs, to solve the current spectrum resource optimization problems in a targeted manner. core pain points. The core problem to be solved in this paper is: how to achieve dynamic, efficient, and balanced allocation of spectrum resources through multi-agent cooperation and attention allocation in complex scenarios where high-speed rail is moving at high speed, channels are time-varying, and services are uneven, and the system is improved. Throughput, communication delay, and enhanced anti-jamming capabilities.

The innovations of this paper are mainly reflected in three aspects: first, break the limitations of single-agent decision-making, design a multi-agent cooperation framework, distribute spectrum sensing and allocation tasks, and improve the real-time and adaptability of decision-making; The second is to introduce an attention mechanism to realize adaptive weight distribution among agents, allowing each agent to dynamically cooperate with priorities according to channel status and business requirements, and solve the problem of insufficient flexibility in fixed-weight cooperation; The third is to combine the reinforcement learning algorithm and the spectrum optimization model to propose an efficient solution to break through the bottleneck of non-convex mixed integer programming problems, which are difficult to solve and slow to converge, and realize the dynamic optimization of resource allocation.

The technical solution of this article is the focus of this article, as follows: First, build a millimeter-wave high-speed rail vehicle-ground communication system model, clarify the topology structure of trains, ground base stations, and communication links, and analyze the propagation loss and Doppler frequency shift of millimeter-wave channels and other characteristics, establish a business requirement model, and clarify the delay and throughput requirements of different services; Secondly, the spectrum resource optimization problem is transformed into a multi-objective non-convex mixed integer programming problem. The optimization objectives are to maximize the total system throughput, minimize the communication delay, and spectrum utilization, and consider channel time-varying, spectrum constraints, and service priorities. constraints; Thirdly, a multi-agent framework is designed. Each agent corresponds to a vehicle-ground communication link, which is responsible for the spectrum sensing, state collection and resource request of the link. By calculating the attention weight of each agent, collaborative decision-making is realized. To handle 'worst-case' scenarios where high-priority demands exceed total capacity, a preemptive scheduling logic is integrated: resources are prioritized for high-priority links via a hard constraint safeguard. If the system-wide capacity is insufficient, low-priority services are progressively throttled or suspended to ensure that all high-priority thresholds (R_1, T_1) are strictly maintained through a guaranteed resource reservation mechanism; Finally, the near-end strategy optimization (PPO) algorithm is used to solve the optimization problem. By training the decision-making model of the agent, the dynamic allocation of spectrum resources is realized, and the overall performance of the system is improved.

The practical implications of this research are fourfold: First, the proposed framework validates the feasibility of real-time spectrum management in high-mobility environments; Second, the attention mechanism quantifiably reduces coordination overhead compared to fixed-weight models; Third, the PPO solution provides a computationally efficient benchmark for non-convex optimization in HSR; Fourth, the results demonstrate a robust framework for multi-service coexistence under interference; Fourthly, the superiority of the proposed method is verified by simulation experiments, which provides a feasible technical solution and theoretical support for the optimization of millimeter-wave high-speed rail vehicle-ground communication spectrum resources. The following chapters are arranged as follows: The second part sorts out the current situation of relevant research and analyzes the shortcomings of existing research; The third part elaborates the specific realization of the proposed spectrum resource optimization method; In the fourth part, the effectiveness of the method is verified by experiments and the results are analyzed; The fifth part summarizes the full text, points out the research insufficiency and the future research direction.

II. RELATED WORK

In the field of multi-agent reinforcement learning, many scholars have carried out in-depth exploration. Mitra [1] took the lead in discussing the relevant methods and strategies to realize robust multi-agent reinforcement learning, which further provides a reliable guarantee for agent cooperation under complex conditions. Samieiyeganeh et al. [2] followed suit and studied the process and key points of the transition from deep reinforcement learning to multi-agent deep reinforcement learning, which further promoted the application of multi-agent systems to more complex scenarios. Qiu et al. [3] focused on cognition-oriented multi-agent reinforcement learning, and deeply explored its principles and applications. This research further expanded the cognitive dimension of multi-agent reinforcement learning. Shi et al. [4] proposed to use horizontal transfer learning to improve the performance and effect of multi-agent reinforcement learning, which further optimizes the efficiency of multi-agent learning. Karimzadeh [5] analyzed the application and characteristics of deep reinforcement learning in multi-agent systems, which further clarified the direction for subsequent research. Balla et al. [6] studied multi-agent reinforcement learning through table games, and explored its training and optimization. This unique research approach further enriched the training methods of multi-agent reinforcement learning. Shi et al. [7] are committed to achieving fault tolerance and improving system stability in multi-agent reinforcement learning, which further enhances the reliability of multi-agent systems. Sheng et al. [8] studied the method and significance of building a structured communication mechanism for multi-agent reinforcement learning, and further solved the communication problem between agents. Khan et al. [9] conducted a comprehensive investigation and review of communication problems in multi-agent reinforcement learning, which provided a comprehensive reference for further research in this field. Gharti et al. [10] analyzed and evaluated the performance of V2X services in 5G millimeter wave communication. Although it is not closely related to multi-agent reinforcement learning, it also provides a reference for related scenario applications. However, these studies still have shortcomings in complex scene adaptability and large-scale agent collaboration efficiency.

III. METHOD

A. MODEL CONSTRUCTION OF MILLIMETER WAVE HIGH-SPEED RAILWAY VEHICLE-GROUND COMMUNICATION SYSTEM

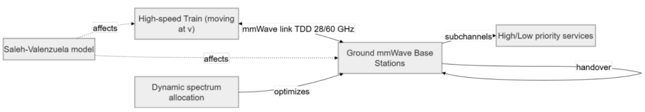

FIGURE 1. Construction of Millimeter Wave High Speed Rail Vehicle Ground Communication System Method

The millimeter-wave high-speed rail vehicle-ground communication system constructed in this paper is mainly composed of high-speed trains, ground millimeter-wave base stations, communication links and business terminals (as shown in Figure 1). All components work together to ensure the stability and efficiency of high-speed rail vehicle-ground communication. Among them, the high-speed train is used as a mobile communication terminal, traveling at a constant speed at speed v, and its on-board communication module is responsible for establishing a connection with the ground base station and transmitting various business data; Multiple ground millimeter-wave base stations are deployed with a 500m spacing. Given the 350 km/h speed, the train crosses a cell approximately every 5.1 seconds, necessitating high-frequency handovers. To address this, the system incorporates a dual-connectivity proactive handover trigger. The agents utilize the attention mechanism to predict the quality of the upcoming base station's channel, initiating data buffering and control-plane signaling transition 1.5 seconds before the estimated cell-edge crossing. This ensures that the dynamic spectrum allocation remains stable despite the rapid topology changes inherent in high-speed rail scenarios. The system adopts time division duplex (TDD), which can flexibly allocate uplink and downlink resources, adapt to the characteristics of unbalanced uplink and downlink traffic in vehicle-to-ground communication, and work in the 28GHz or 60GHz millimeter wave frequency band, which has the advantage of ultra-wideband. The spectrum resources are divided into multiple independent subchannels, and each subchannel can be allocated to different vehicle-to-ground communication links according to business needs to provide support for multi-service parallel transmission [11].

In order to accurately describe the communication characteristics of the system, a millimeter-wave channel model is first established, fully considering key influencing factors such as path loss, Doppler frequency shift, and shadow fading in the high-speed rail scene, and the Saleh-Valenzuela model is used to describe the multipath characteristics of the millimeter-wave channel. The influence law of key parameters such as the amplitude, Doppler frequency shift and delay of the multipath signal provides channel support for subsequent spectrum resource optimization. Secondly, establish a business demand model that fits the reality, and combine the business characteristics of high-speed rail vehicle-ground communication, divide the vehicle-ground communication business into high-priority services (such as automatic driving control signals, equipment failure warning signals) and low-priority services (such as In-vehicle entertainment, passenger Internet services), according to the service requirements of the two types of services, set differentiated throughput and delay constraints to ensure the real-time performance and reliability of high-priority services. Finally, the overall topology of the system and the connection relationship between each communication link are clarified, and the communication logic between trains, base stations, and business terminals is clarified, laying a solid foundation for the subsequent dynamic allocation and optimization of spectrum resources [12].

B. MODELING OF SPECTRUM RESOURCE OPTIMIZATION PROBLEM

This article models the spectrum resource optimization problem as a multi-objective non convex mixed integer programming problem, with the optimization objectives of maximizing the total system throughput, minimizing communication delay, and maximizing spectrum utilization, while considering channel time-varying constraints, spectrum resource constraints, business priority constraints, etc.

Define optimization variables: x_{ijk} indicate whether the i-th intelligent agent (corresponding to the i-th communication link) allocates the jth subchannel to the k-th type of service, $x_{ijk} \in \{0,1\}$, where $i = 1, 2, \ldots, M$ (where M is the number of agents), $j = 1, 2, \ldots, N$ (where N is the number of sub channels), $k = 1, 2$ (1 is a high priority business, 2 is a low priority business), p_{ijk} indicate the transmission power allocated by the i-th agent for the k-th type of service on the j-th subchannel, $p_{ijk} \geq 0$.

The optimization objective function is as follows:

(1) Maximizing the total throughput of the system:

$$\text{maximize} \sum \sum \sum x_{ijk} \cdot \log_2 \left(1 + \frac{p_{ijk} \cdot |h_{ij}|^2}{\sigma^2} \right) \quad (1)$$

Where h_{ij} is the channel gain of the i-th link on the jth subchannel, and σ^2 is the noise power.

(2) Minimize communication latency

$$\text{minimize}\,(1/\,\text{TotalThroughput})\cdot\sum\sum\sum x_{ijk}\cdot L_k \quad (2)$$

Where L_k is the packet length of the k-th type of service.

(3) Maximizing Spectrum

$$\text{maximize}\,\frac{\sum\sum\sum x_{ijk}}{M\cdot N} \quad (3)$$

The constraints include:

(1) Spectrum resource constraint: Each subchannel can be allocated to a maximum of one communication link, i.e. $\sum\sum x_{ijk}\leq 1$

(2) Power constraint: The total transmission power of each agent shall not exceed the maximum transmission power, i.e. $\sum\sum p_{ijk}\leq P_{\max}$

(3) Business constraint: The latency and throughput of high priority businesses must meet preset thresholds, i.e.

$$\sum x_{ijk}\cdot\log_2\left(1+\frac{p_{ijk}\cdot|h_{ij}|^2}{\sigma^2}\right)\geq R_1 \quad (4)$$

$$\sum x_{ijk}\cdot\frac{L_1}{\log_2\left(1+\frac{p_{ijk}\cdot|h_{ij}|^2}{\sigma^2}\right)}\leq T_1 \quad (5)$$

(4) Channel time-varying constraint: Update x_{ijk} and p_{ijk} based on real-time channel status to ensure communication quality. Due to the presence of integer variables x_{ijk} in the objective function and constraints, and the non convex nature of the objective function, this problem belongs to non convex mixed integer programming and traditional methods are difficult to solve efficiently.

C. MULTI AGENT ATTENTION ALLOCATION AND PPO SOLVING ALGORITHM

In response to the above optimization problems, this paper designs a multi-agent attention allocation framework and combines it with the PPO algorithm to achieve efficient solutions. Firstly, the multi-agent framework consists of M agents, each corresponding to a vehicle ground communication link, responsible for spectrum sensing, channel state acquisition, business requirement statistics, and resource allocation decisions for that link. Each intelligent agent shares channel status, business requirements, and other information through information exchange to achieve collaborative decision-making.

Introducing attention mechanism to achieve adaptive weight allocation among intelligent agents: calculating the attention weight of each agent, the weight size is determined by the corresponding channel quality, business priority, and resource requirements of the agent. Specifically, first extract the feature vectors of each agent, including channel gain, service priority coefficient, resource demand level, etc; Then, the attention weights of each agent are calculated through the attention network. After normalizing the weights, each agent adjusts the cooperation priority based on the weight size. The higher the weight, the more spectrum resources and decision priorities the agent obtains. The expression for calculating attention weights is

$$w_i=\text{softmax}\,(f(h_i,s_i,d_i)) \quad (6)$$

Where h_i is the channel feature of the i-th agent, s_i is the service priority feature, d_i is the resource demand feature, $f(\cdot)$ is the attention network mapping function, and $\text{softmax}(\cdot)$ is the normalization function. The mapping function $f(\cdot)$ is implemented as a two-layer Multi-Layer Perceptron (MLP) with 128 and 64 neurons respectively, utilizing ReLU activation functions. Features h_i (normalized channel gain), s_i (one-hot encoded priority), and d_i (normalized queue length) are concatenated into a single input vector. The network learns to project these features into a scalar score representing agent importance before the softmax operation. The PPO algorithm is used to solve optimization problems, which has the advantages of fast convergence speed and good stability, and can effectively solve the sample efficiency problem in reinforcement learning. Treat each intelligent agent as an intelligent agent, with a state space consisting of channel status, business requirements, spectrum resource occupancy, etc., and an action space consisting of sub channel allocation and transmission power adjustment. The reward function is designed based on system throughput, communication delay, spectrum utilization, and constraint satisfaction

$$r=\alpha\cdot\text{Throughput}-\beta\cdot\text{Delay}+\gamma\cdot\text{SpectrumEfficiency}-\lambda\cdot\text{Penalty} \quad (7)$$

Among them, α, β, γ, and λ are weight coefficients, and Penalty is the penalty term for constraint violation. By training the strategy network of intelligent agents, dynamic optimization allocation of spectrum resources is achieved, and ultimately a spectrum allocation scheme that meets multi-objective optimization requirements is obtained [13].

IV. RESULTS AND DISCUSSION

A. EXPERIMENTAL ENVIRONMENT AND PARAMETER SETTINGS

To verify the effectiveness of the proposed method (as shown in Figure 2), a millimeter wave high-speed rail vehicle-to-ground communication environment was developed using MATLAB R2023b, specifically leveraging the Communication Toolbox and Reinforcement Learning Toolbox. The millimeter-wave channel was implemented by integrating the Saleh-Valenzuela model with a custom Doppler shift function (max 1200Hz). Multi-agent interactions were simulated via a centralized training and decentralized execution (CTDE) architecture within a custom-defined Gymnasium-compatible environment class. The experimental parameters are set as follows: the train speed is 350km/h, the millimeter wave operating frequency band is 28GHz, the number of sub channels is 16, and the bandwidth of each sub channel is 100MHz; The distance between ground

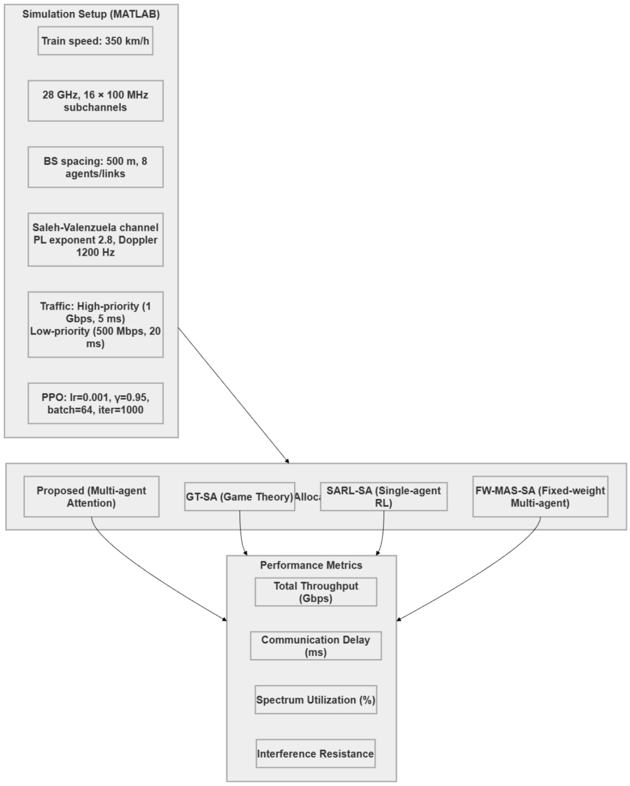

FIGURE 2. Experimental environment and parameter settings

base stations is 500m, and the number of intelligent agents and communication links is the same, both being 8; The channel model adopts the Saleh Valenzuela model, with a path loss index of 2.8 and a maximum Doppler frequency shift of 1200Hz. Business parameters: high priority business throughput threshold of 1Gbps and delay threshold of 5ms, low priority business throughput threshold of 500Mbps and delay threshold of 20ms.

PPO algorithm parameters: learning rate of 0.001, discount factor of 0.95, batch size of 64, and training iterations of 1000 times.

Select three existing mainstream methods as comparison methods:

(1) GT-SA: A non-cooperative game theory approach where agents reach a Nash Equilibrium via iterative best-response dynamics;

(2) SARL-SA: A single-agent Deep Q-Network (DQN) that treats the entire system state as a global input;

(3) FW-MAS-SA: A multi-agent PPO framework where cooperation is mediated by static, equal-weighted interaction matrices.

Baseline parameters were optimized using a grid search to ensure optimal performance for a fair comparison. Comparative indicators include total system throughput, communication latency, spectrum utilization, and anti-interference performance.

B. ANALYSIS OF EXPERIMENTAL RESULTS

Analysis of total system throughput: As the number of training iterations increases (as shown in Figure 3), the total system throughput of both the proposed method and the comparative method gradually improves and tends to stabilize. When the number of iterations reaches 800, the throughput of each method tends to stabilize. The total throughput of the system proposed in this paper reaches 28.6 Gbps (SD = 0.42), representing a 25.7% improvement over the GT-SA method, 21.4% higher than the SARL-SA method, and 18.3% higher than the FW-MAS-SA method. To verify these gains, paired t-tests were conducted across 50 independent simulation runs, yielding p-values < 0.01, which confirms that the observed improvements in throughput and delay are statistically significant and not due to random variance. The reason is that the method proposed in this article can dynamically allocate spectrum resources based on channel status and business requirements through multi-agent collaboration and attention allocation, avoiding the limitations of single decision-making and fixed weight collaboration, and improving resource utilization efficiency.

Communication delay analysis: At different train speeds (250km/h, 300km/h, 350km/h, 400km/h), the communication delay of the method proposed in this paper is lower than that of the comparative method. When the train speed is 350km/h, the average communication delay of our method is 3.2ms, which is 27.5% lower than the GT-SA method, 24.2% lower than the SARL-SA method, and 20.1% lower than the FW-MAS-SA method. This is because the method

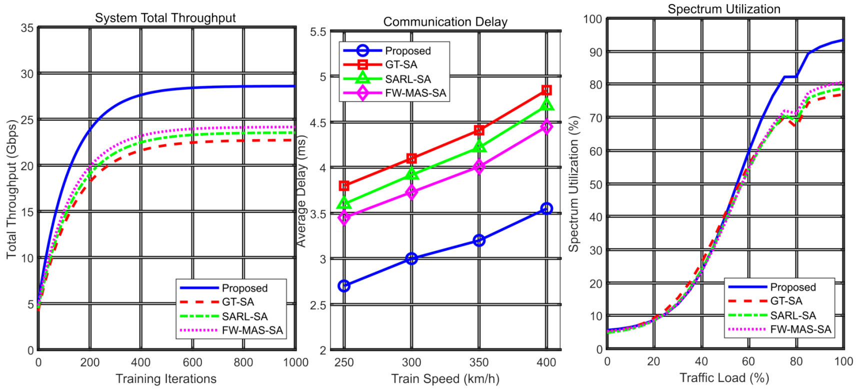

FIGURE 3. Performance comparison between proposed method and baseline

proposed in this article introduces an attention mechanism that can quickly respond to channel changes, adjust spectrum allocation schemes in a timely manner, reduce data transmission latency, and meet the latency constraints of high priority services.

Spectrum utilization analysis: The spectrum utilization of the method proposed in this article is always higher than that of the comparison method. When the business load is 80%, the spectrum utilization of this method reaches 82.3%, which is 22.8% higher than the GT-SA method, 19.5% higher than the SARL-SA method, and 15.6% higher than the FW-MAS-SA method. This indicates that the multi-agent attention allocation framework can achieve balanced allocation of spectrum resources, avoid resource waste, and improve spectrum utilization efficiency.

C. ANTI INTERFERENCE PERFORMANCE AND DISCUSSION

Anti interference performance test: In the presence of external interference (interference power is 0.5 times the signal power), the system throughput of the proposed method in this paper decreases by 8.7% and the communication delay increases by 10.2%, while the throughput decreases by more than 15% and the delay increases by more than 18% compared to the other methods. The method described in this article can quickly perceive interference signals through attention mechanisms, adjust spectrum allocation schemes, avoid interfering sub channels, and improve the system's anti-interference performance.

The experimental results show that the spectrum resource optimization method proposed in this paper under multi-agent attention allocation is superior to existing mainstream methods in terms of system throughput, communication delay, spectrum utilization, and anti-interference performance. It can effectively adapt to complex scenarios such as high-speed rail, time-varying channels, and uneven business. Meanwhile, it was also found in the experiment that when the number of intelligent agents is too large, the decision-making complexity of the system will increase, resulting in slower convergence speed. In the future, the system performance can be further improved by optimizing the attention network structure and introducing distributed training methods. In addition, the experiment used simulation scenarios, and the engineering practicality of the method can be further verified by combining actual high-speed rail test data in the future.

V. CONCLUSION

This article proposes a spectrum resource optimization method under multi-agent attention allocation to address the problems of low spectrum utilization, weak anti-interference ability, and imbalanced resource allocation in millimeter wave high-speed railway vehicle ground communication. By constructing a millimeter wave high-speed railway vehicle ground communication system model, the spectrum optimization problem is modeled as a multi-objective non convex mixed integer programming problem. A multi-agent attention cooperation framework is designed, and attention mechanism is introduced to achieve adaptive weight allocation between agents. The PPO algorithm is combined to efficiently solve the optimization problem. The experimental results show that the proposed method can significantly improve the total throughput and spectrum utilization of the system, reduce communication latency, enhance anti-interference performance, and is superior to existing mainstream optimization methods. This study solves the problem of efficient allocation of millimeter wave vehicle ground communication spectrum resources in dynamic scenarios of high-speed rail, providing theoretical support and technical solutions for the engineering application of millimeter wave technology in high-speed rail vehicle ground communication. There are still certain shortcomings in this study, such as slow convergence speed when there are a large number of intelligent agents, and insufficient consideration of spectrum allocation in multi base station collaborative switching scenarios. Future research can focus on optimizing the attention network structure, introducing distributed training methods, and combining multi base station collaborative switching technology to further enhance the adaptability and practicality of the methods, promoting the development of high-speed rail vehicle ground communication towards higher quality, efficiency, and reliability.

AUTHOR CONTRIBUTIONS

Yuan Zhuo designed, collected and analyzed the data, and drafted the manuscript. Yuan Zhuo conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Yuan Zhuo participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] A. Mitra, "Towards Robust Multi-Agent Reinforcement Learning," Proceedings of the AAAI Symposium Series, vol. 3, no. 1, pp. 308-308, 2024, doi: 10.1609/aaaiss.v3i1.31222.
[2] M. Samieiyeganeh, K. Rahmat R W B O, B. Khalid F, and A. Bin Kasmiran K, "Deep Reinforcement Learning to Multi-Agent Deep Reinforcement Learning," Journal of theoretical and applied information technology, vol. 100, no. 4, pp. 990-1003, 2022.
[3] T. Qiu, S. Wu, Z. Liu, et al., "Cognition-Oriented Multiagent Reinforcement Learning," Neural Networks and Learning Systems, IEEE Transactions on, vol. 36, no. 6, pp. 10736-10748, 2025, doi: 10.1109/TNNLS.2024.3502176.
[4] H. Shi, J. Li, J. Mao, et al., "Lateral Transfer Learning for Multiagent Reinforcement Learning," IEEE Transactions on Cybernetics, vol. 53, no. 3, pp. 1699-1711, 2023, doi: 10.1109/TCYB.2021.3108237.
[5] D. Karimzadeh, "Deep Reinforcement Learning Multi-Agent Systems," International journal of Modern Achievement in Science, Engineering and Technology, vol. 2, no. 1, pp. 56-62, 2024, doi: 10.63053/ijset.60.
[6] M. Balla, M. Long G E, J. Goodman, et al., "PyTAG: Tabletop Games for Multiagent Reinforcement Learning," IEEE Transactions on Games, vol. 16, no. 4, pp. 993-1002, 2024, doi: 10.1109/TG.2024.3404133.
[7] Y. Shi, H. Pei, L. Feng, et al., "Toward Fault Tolerance in Multi-Agent Reinforcement Learning," IEEE Transactions on Automation Science and Engineering, vol. 22, no. 22, pp. 19007-19024, 2025, doi: 10.1109/TASE.2025.3592721.
[8] J. Sheng, X. Wang, B. Jin, et al., "Learning structured communication for multi-agent reinforcement learning," Autonomous Agents and Multi-Agent Systems, vol. 36, no. 2, pp. 1-31, 2022, doi: 10.1007/s10458-022-09580-8.
[9] R. Khan, N. Khan, and T. Ahmad, "Communication in Multi-Agent Reinforcement Learning: A Survey," The Nucleus, vol. 60, no. 2, pp. 174-184, 2023, doi: 10.71330/thenucleus.2023.1303.
[10] P. Gharti M and R. Joshi S, "Performance Analysis of V2X Services in 5G Millimeter Wave Communication," Journal of Science and Technology, vol. 3, no. 1, pp. 1-5, 2023, doi: 10.3126/jost.v3i1.69057.
[11] A. Madankar E A, "Performance Estimation for 5G Millimeter-Wave Communication," International Journal on Recent and Innovation Trends in Computing and Communication, vol. 11, no. 9, pp. 3487-3493, 2023, doi: 10.17762/ijritcc.v11i9.9560.
[12] X. Li, S. Guo, and S. Malik, "Sensing-aided peer-to-peer millimeter-wave communication," International Journal of Informa tion Technology, vol. 16, no. 4, pp. 2069-2075, 2024, doi: 10.1007/s41870-024-01768-3.
[13] C. Redondi A E, C. Innamorati, S. Gallucci, et al., "A Survey on Future Millimeter-Wave Communication Applications," Access, IEEE, vol. 12, pp. 133165-133182, 2024, doi: 10.1109/ACCESS.2024.3438625.