

# Cooperative Jamming for Secure Air–Ground Integrated Networks: A Hierarchical Distributed Deep Reinforcement Learning Approach

Kakyeom Jeon<sup>1</sup>, Student Member, IEEE, Young-Seok Lee<sup>2</sup>, Student Member, IEEE,  
Bang Chul Jung<sup>1</sup>, Senior Member, IEEE, and Howon Lee<sup>1</sup>, Senior Member, IEEE

**Abstract**—Recently, Internet of Things (IoT) devices have been installed everywhere, and these devices are connected through wireless communication networks. In particular, uncrewed aerial vehicles (UAVs), one of the most promising IoT devices, are expected to be used actively in Internet of Battlefield Things (IoBT) networks due to their flexible 3-D mobility. To react to enemy UAV attacks in the IoBT networks, the ground-to-air (G2A) or air-to-air (A2A) radio jamming can be an effective counterattack technique that disrupts the communication and control signals of adversary equipment. That is, it can be a very effective means of coping with attacks by UAVs in modern battlefields characterized by electronic warfare. Accordingly, this article proposes a hierarchical distributed deep reinforcement learning-based cooperative jamming (HDRL-CJ) method for secure air–ground integrated networks. The proposed method uses two types of jammers: ground jammers (GJs) and UAV jammers (UJs). The GJ optimizes the beamwidth to maximize the effectiveness of jamming, and the UJ tracks the malicious UAV (MU) and controls the jamming power to minimize the MU’s signal-to-jamming-plus-noise ratio (SJNR) while considering the UJ’s limited battery capacity. Moreover, to reduce the computational complexity of reinforcement learning (RL) method, we devise a hierarchical RL architecture that separates the UJ’s movement control and transmit power control. Through extensive simulations, we demonstrate that the proposed HDRL-CJ method converges to the optimal solution obtained by the optimal exhaustive search algorithm. Furthermore, by comparing the jamming performance with several benchmark methods, we validate the proposed method’s superior performance under various 3-D network environments.

**Index Terms**—Cooperative jamming, deep reinforcement learning (DRL), distributed learning, Internet of Battlefield Things (IoBT), secure air–ground integrated network, signal-to-jamming-plus-noise ratio (SJNR), uncrewed aerial vehicle (UAV).

Received 6 June 2025; revised 22 August 2025; accepted 15 September 2025. Date of publication 29 September 2025; date of current version 8 December 2025. This work was supported in part by the National Research Foundation of Korea (NRF) through the Korean Government (MSIT) under Grant RS-2025-00563401 and Grant RS-2025-02303435 and in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) funded by MSIT (Augmented Beam-Routing: Carom-MIMO) under Grant 2021-0-00486. (Corresponding authors: Bang Chul Jung; Howon Lee.)

Kakyeom Jeon and Young-Seok Lee are with the Department of Artificial Intelligence Convergence Network, Ajou University, Suwon 16499, Republic of Korea (e-mail: kuem0803@ajou.ac.kr; youngseoklee@ajou.ac.kr).

Bang Chul Jung and Howon Lee are with the Department of Electrical and Computer Engineering, Ajou University, Suwon 16499, Republic of Korea (e-mail: bcjung@ajou.ac.kr; howon@ajou.ac.kr).

Digital Object Identifier 10.1109/JIOT.2025.3615825

## I. INTRODUCTION

### A. Background

ADVANCEMENTS in Internet of Things (IoT) technologies have significantly transformed modern warfare, leading to the development of the Internet of Battlefield Things (IoBT) [1], [2], [3]. IoBT integrates key military IoT components, such as sensors, UAVs, surveillance equipment, and wearables, to enable real-time data acquisition and monitoring. This enhances situational awareness and supports rapid, data-driven decision-making on the battlefield.

In addition, the importance of wireless communications using a broad frequency spectrum has gradually grown in the modern military to accomplish various critical tasks. These tasks include the command and control of forces through wired and wireless communications, tracking enemy activity, locating allies, and assessing the effectiveness of weapons [4]. To secure an initiative in network-centric warfare, intelligent data collection and analysis capabilities are required compared with the enemy’s ability to handle the vast amount of real-time data generated on a battlefield. This will enable proactively identifying enemy signals and locations, neutralizing their electronic weapon systems, and providing strategic decision-making. Specifically, electronic warfare can be classified into three categories: support, attack, and protection [5]. Among these, electronic attacks impair enemy combat capabilities by neutralizing their military communication equipment and non-communication systems, including radar and guided weapons. Because of the propagation and superposition properties of wireless channels, military communication systems are vulnerable to security threats such as eavesdropping and jamming attacks on wireless links [6]. Accordingly, extensive research has been conducted not only on protective countermeasures but also on the development of intelligent jamming techniques aimed at disrupting adversarial communications and data acquisition [7], [8]. Most conventional jamming techniques adopt the use of either ground jammers (GJs) or UAV jammers (UJs) independently to protect network communications and security, primarily targeting fixed eavesdropping devices or preidentified targets within mission areas [9], [10], [11], [12], [13], [14], [15], [16], [17], [18].

In particular, the UJ is regarded as a critical asset in modern electronic warfare owing to the unique advantages offered by

its aerial platform. Unlike a GJ fixed to a single location, the UJ possesses the decisive advantage of unrestricted mobility in 3-D space, free from terrain constraints. This enables the UJ to be positioned in real time at the optimal location to disrupt the adversary effectively. Moreover, leveraging real-time data collection and detection capabilities, the UJ can proactively respond to rapidly changing battlefield conditions such as frequency shifts or adversarial maneuvers. Such capabilities substantially enhance the efficiency and accuracy of electronic attack missions.

### B. Related Work

Recently, several studies have investigated jamming techniques that use GJs to disrupt communications between enemy base stations and adversaries who have infiltrated friendly territories to carry out their missions [9], [10], [11], [12], [13]. Feng et al. [9] proposed a jammer placement optimization algorithm to minimize the number of jammers required for jamming an entire network. They optimized the placement using integer linear programming and a metaheuristic algorithm. Similarly, Gezici et al. [10] optimized the placement of jammers based on Cramer–Rao lower bounds in an environment where the jammer was aware of the network topology, the number of nodes in the target, their location, and the signal-to-jamming-plus-noise ratio (SJNR). Meanwhile, Kim and Lim [11] proposed a reinforcement learning (RL)-based jamming strategy that uses beamforming to enhance the effectiveness of ground-based jammers. Specifically, a multiarmed bandit (MAB)-based algorithm was used to learn the beam direction, angle, and width to improve the jamming effectiveness. However, these studies considered only fixed enemy positions while determining the optimal solution. Huo et al. [12], proposed a learning-based intermittent cooperative jamming scheme for energy-constrained, nonslotted wireless transmission environments in IoT systems. By optimizing the jamming and energy-harvesting (EH) durations of the GJs through a learning-based approach, they effectively increased the bit-error rate of the eavesdropper while significantly reducing energy consumption compared with conventional methods. Li et al. [13] investigated the use of GJs to enhance the security of communication systems. Specifically, they considered a downlink cooperative nonorthogonal multiple access system with an untrusted relay, where a friendly jammer was deployed to improve physical-layer security and maximize the secrecy sum rate.

Meanwhile, with the advancements in UAV-related technologies, UAVs have begun to be used for military purposes, such as in electronic warfare and surveillance. For instance, by outfitting UAVs with specialized jamming equipment, UJs can be deployed on an actual battlefield. Many researchers have considered using friendly UJs to counter malicious ground node eavesdropping in our territory [14], [15], [16], [17], [18]. Such UJ-based electronic attacks can degrade hostile C3 (command, control, and communication) systems and radar capabilities across wide areas. Li et al. [14] used a UJ that could fly close to an eavesdropper and cooperatively transmit jamming signals during a specified flight time. The authors jointly optimized the moving paths and jamming power of

the UJs to maximize the average secrecy rate. Zhong et al. [15] investigated a cooperative jamming scheme in which the UJ generated jamming signals to prevent eavesdropping on UAV communications. This method is beneficial when the transmission signal from a UAV transmitter to a ground station is at risk of eavesdropping over a large network area. The authors used a successive convex approximation-based technique to plan the moving paths and control the transmission power of UAVs, which enhances communication quality and increases the secrecy rate. In addition, Nnamani et al. [16] proposed a UAV-assisted jamming technique to implement physical-layer keyless security in scenarios where an eavesdropper with an unknown location exists in a friendly territory. By jointly optimizing the moving path and jamming power of a UJ and the transmit power of a source node, they achieved secrecy rates comparable to scenarios in which the location of the eavesdropper is known. However, because these studies assumed static network environments where only the UJ could move, they are limited in their application to a dynamic environment. To mitigate eavesdropping, Dang-Ngoc et al. [17] proposed a cooperative friendly jamming framework using a UAV swarm for secure amplify-and-forward relaying networks with wireless EH. They proposed a time-switching relaying protocol that enables UAVs to harvest energy, relay data, and jam eavesdroppers and derived the secrecy outage probability under various detection techniques. Furthermore, they provided guidelines for optimizing EH duration and swarm size to achieve a target secrecy level. Similarly, Cabezas et al. [18] addressed the problem of efficiently adjusting the transmission directions (azimuth angles) of multiple cooperative UJs to enhance the security of legitimate ground transmissions, and proposed a precoder to suppress jamming at the legitimate receiver. They formulated the problem within an MAB framework and demonstrated that their approach achieved significant performance improvements and faster convergence compared with benchmarks based on gradient descent.

In addition, machine-learning (ML)-based algorithms suitable for dynamic network environments could be adopted for 3-D wireless networks. Several studies have been conducted on deep-Q-learning (QL)-based algorithms that continuously learn from dynamic networks and perform network optimization based on real-time decision-making [19], [20], [21]. Liu et al. [19] proposed a hierarchical deep Q-network (h-DQN) model for dynamic spectrum access. The h-DQN demonstrated faster convergence, higher performance, and higher channel utilization than conventional QL or deep RL (DRL). Lee et al. [20] introduced a hierarchical QL model that used an inner-loop RL to adjust the transmit power of UAVs and an outer-loop RL to control the frequency reuse factor of a multi-UAV wireless network. Their findings demonstrated that the proposed model effectively mitigates intercell interference and reduces transmit power consumption, even when UAVs move continuously. Furthermore, Kim et al. [21] proposed a hierarchical RL framework that exploits the inner-loop RL to determine the optimal transmit power of the UAV and ground base station (BS) and outer loop RL to determine the optimal UAV BS placement. The authors demonstrated

that this framework can maximize system throughput and minimize the number of outage users in a 3-D environment with mobility and traffic dynamics. Given that the UJ and GJ cooperatively jam the malicious UAV (MU), determining the optimal beamwidth of GJ, transmit power, and the position of UJ with a centralized RL causes enormous computational complexity. Thus, reducing the computational complexity and maximizing jamming performance is crucial with a hierarchical and distributed RL framework.

### C. Motivation and Contributions

This study aims to develop a cooperative jamming method that enables effective jamming against fast-moving MU in 3-D battlefields through the cooperation of GJ and UJ. Previous studies have focused solely on GJ-based configurations and demonstrated that GJs can provide jamming effects against MUs across multiple layers, ranging from hardware to software [22], [23], [24]. However, when GJs use omnidirectional antennas, their jamming footprint remains confined to a relatively small area. In contrast, antennas with pronounced directionality, such as the Yagi-Uda and helical designs, achieve higher gain along the boresight but inherently restrict the jammer to a narrow angular sector, further limiting overall coverage [25].

This study assumes that the GJ is equipped with a permanent power supply, allowing for continuous and high-power jamming. However, since the GJ is deployed at a fixed position on the ground and operates with directional transmission characteristics, its jamming coverage can be limited. As a result, it faces distinct limitations when attempting to single-handedly jam a fast-moving MU. In contrast, the UJ can effectively leverage its high mobility to jam MU, but its limited battery capacity constrains its transmission power increment and flight time. The cooperative jamming method can address these issues by adjusting GJ's beamwidth and UJ's jamming power and movement. The GJ controls the jamming effectiveness by adjusting the jamming signal's beamwidth, which radiates upward in a direction perpendicular to the ground. This enables the determination of the optimal beamwidth depending on the location of the MU to maximize jamming performance. In contrast, the UJ can effectively jam by tracking the MU's movement and controlling the optimal jamming power to disrupt the MU's communication. Furthermore, in 3-D network environments where the MU moves continuously, finding an optimal solution—GJ beamwidth, UJ transmit power, and UJ moving path—is challenging.

Therefore, in this article, we propose a novel hierarchical distributed DRL-based cooperative jamming (HDRL-CJ) method. The proposed HDRL-CJ adopts a decentralized agent structure, where the GJ and UJ are defined as independent agents and perform distributed learning. In addition, to effectively respond to fast-moving MU in real-time, a hierarchical framework is designed to allow the UJ to sufficiently learn its mobility policy. In particular, HDRL-CJ takes into account the limited battery constraints of the UJ and a reward-sharing mechanism across different actions, enabling effective control not only of the GJ's beamwidth but also of the UJ's mobility

and jamming power. As a result, HDRL-CJ achieves effective jamming performance against MU while addressing the computational complexity of centralized RL frameworks and significantly reducing computational overhead.

### D. Article Organization

The remainder of this article is organized as follows: Section II presents the ground-to-air (G2A) and air-to-air (A2A) channel models and the two-step beam pattern design. Section III presents the proposed HDRL-CJ and each agent's MDP for optimal cooperative jamming. Section IV demonstrates the excellence of the proposed method through intensive simulations. Finally, the conclusions are presented in Section V.

## II. SYSTEM MODEL

This study considers a cooperative jamming system in a secure air-ground integrated network designed to respond to the intrusion of an MU into friendly airspace, as illustrated in Fig. 1. In this scenario, the MU is assumed to be unaware of the locations of the jammers and is controlled by command signals from a malicious ground base station (MG). If only a GJ is deployed, an MU equipped with 3-D mobility can easily evade the jamming coverage and carry out its mission. Therefore, to overcome the operational limitations of the GJ, a cooperative jamming strategy that incorporates a mobile jammer (i.e., UJ) capable of tracking the MU and providing additional jamming is required. In other words, the UJ leverages its 3-D mobility to track the MU in real-time and effectively neutralize it by jamming MU located in the GJ's coverage blind spots. In this study, the jammer actions are optimized through RL. In addition, we assume that external sensing devices can provide spectral information for command signaling between the MG and MU and the position of the MU. To analyze the SJNR of the MU from the GJ and UJ, we introduce the G2A and A2A channel models.

### A. G2A Channel Model

The G2A channel model can be used to model the MG-to-MU and GJ-to-MU wireless channels. This channel model separates the line-of-sight (LoS) and non-LoS (NLoS) propagations, considering ground buildings and other obstacles obstructing the signal propagation between the ground nodes ( $\mathcal{G} \in \{g_j, g_m\}$ ) and aerial nodes ( $\mathcal{U} \in \{u_j, u_m\}$ ). In this study, an elevation-angle-dependent probabilistic LoS model was adopted as the G2A channel model [26], [27]. The LoS probability between the ground node  $g \in \mathcal{G}$  and the aerial node  $u \in \mathcal{U}$  ( $\mathbb{P}_{(g,u)}^{\text{LoS}}$ ) can be calculated using a formula based on the sigmoid function as follows:

$$\mathbb{P}_{(g,u)}^{\text{LoS}}(\theta_{(g,u)}) = \frac{1}{1 + \alpha \times \exp(-\beta \times [\theta_{(g,u)} - \alpha])} \quad (1)$$

$$\theta_{(g,u)} = \frac{180^\circ}{\pi} \times \sin^{-1} \frac{h_u}{d_{(g,u)}}. \quad (2)$$

Here,  $\alpha$  and  $\beta$  are the G2A channel parameters for several urban environmental deployments, as listed in Table I. This

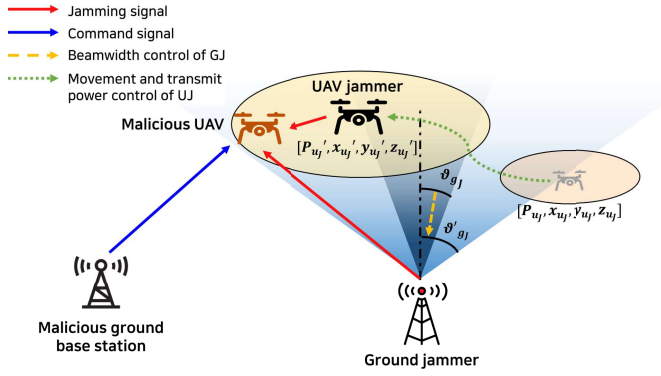


Fig. 1. System model of cooperative jamming using UJ and GJ.

TABLE I

G2A CHANNEL PARAMETERS FOR URBAN ENVIRONMENTAL DEPLOYMENTS [26], [27]

Environment	$\alpha$	$\beta$	$\zeta^{LoS}(dB)$	$\zeta^{NLoS}(dB)$
Suburban	4.88	0.429	0.1	21
Urban	9.61	0.16	1.0	20
Dense Urban	12.08	0.11	1.6	23
Highrise Urban	27.23	0.08	2.3	34

G2A channel model is based on the environmental deployment model suggested by the International Telecommunication Union-Radio Communication Sector (ITU-R). Although the elevation-angle-dependent probabilistic LoS model does not directly consider small-scale fading,  $\alpha$  and  $\beta$  reflect the statistical long- and short-term characteristics of G2A channels in different environments (suburban, urban, dense urban, and high-rise urban). The  $\theta_{(g,u)}$  and  $d_{(g,u)}$  represent the elevation angle and distance between  $g$  and  $u$ , respectively, and  $h_u$  denotes the altitude of aerial node  $u$ . From  $\mathbb{P}_{(g,u)}^{LoS}$ , the NLoS probability between ground node  $g$  and aerial node  $u$  ( $\mathbb{P}_{(g,u)}^{NLoS}$ ) can be calculated as  $\mathbb{P}_{(g,u)}^{NLoS} = 1 - \mathbb{P}_{(g,u)}^{LoS}$ . In addition, the LoS and NLoS path losses ( $L_{(g,u)}^{LoS}$ ,  $L_{(g,u)}^{NLoS}$ ) can be obtained as follows:

$$L_{(g,u)}^{LoS} = 20 \log \left( \frac{4\pi f d_{(g,u)}}{c} \right) + \zeta^{LoS} \quad (3)$$

$$L_{(g,u)}^{NLoS} = 20 \log \left( \frac{4\pi f d_{(g,u)}}{c} \right) + \zeta^{NLoS} \quad (4)$$

where  $f$  and  $c$  are the carrier frequency in GHz scale and the propagation speed of the radio waves, respectively.  $20 \log \left( \frac{4\pi f d_{(g,u)}}{c} \right)$  denotes the free-space path loss, which is equally included in both the equations. Using (3) and (4), we can express the overall path loss  $L_{(g,u)}^{G2A}$ , which combines the LoS and NLoS losses between  $g \in \mathcal{G}$  and  $u \in \mathcal{U}$  as follows:

$$L_{(g,u)}^{G2A} = \mathbb{P}_{(g,u)}^{LoS} \times L_{(g,u)}^{LoS} + \mathbb{P}_{(g,u)}^{NLoS} \times L_{(g,u)}^{NLoS}. \quad (5)$$

### B. A2A Channel Model

In an A2A communication environment, the signal transmitted by a UJ  $u_j$  to an MU  $u_m$  is affected by a strong LoS component and is subject to a free-space path loss model. Thus, the path loss of a jamming signal in an A2A environment

can be calculated using the Friis free-space equation, as expressed in [28]

$$L_{(u_j, u_m)}^{A2A} = 20 \log(f) + 20 \log \left( d_{(u_j, u_m)} \right) + 32.44. \quad (6)$$

Here,  $d_{(u_j, u_m)}$  and  $f$  are the distance between UJ  $u_j$  and MU  $u_m$  and the carrier frequency, respectively.

### C. 3-D Two-Step Beam Pattern Design

The GJ exploits a 3-D directional beam with main and side lobes to jam MU based on a two-step beam pattern design with 3-D upward radiation [29]. For effective jamming, the GJ mainly controls the beamwidth of the main lobe, and its radiation gain can be calculated using the law of energy conservation as follows:

$$\frac{1}{2\pi} \int_0^{\pi/2} \int_0^{2\pi} \eta(\vartheta, \phi) d\phi d\vartheta = 1 \quad (7)$$

where  $\vartheta$  is the width of the main lobe from the  $z$ -axis perpendicular to the reference plane and  $\phi$  is the azimuth angle. In addition, the beam gain is expressed as  $\eta(\vartheta, \phi)$  using  $\vartheta$  and  $\phi$ . The radiation area  $A_m$  of the main lobe is a part of the hemisphere and can be obtained as follows:

$$\int_0^{\vartheta} \int_0^{2\pi} \sin(\varphi) d\phi d\varphi = 2\pi(1 - \cos(\vartheta)) = A_m(0^\circ < \vartheta \leq 90^\circ). \quad (8)$$

In addition, the radiation area of the sidelobe ( $A_s$ ) is calculated as the area of the hemisphere in the upward direction minus that of the main lobe, which is equal to  $2\pi - A_m$ . The relationship between the radiation gain and the radiation area of each lobe with respect to the total radiation area can be represented as

$$1 = \eta_m \frac{A_m}{2\pi} + \eta_s \frac{A_s}{2\pi} = \eta_m \frac{A_m}{2\pi} + \eta_s \frac{2\pi - A_m}{2\pi} \quad (9)$$

where  $\eta_m$  and  $\eta_s$  are the radiation gains of the main and side lobes, respectively. In (9), the values of  $\eta_m$  and  $\eta_s$  can be determined using the interlobe gain ratio  $\delta$  as follows:

$$\eta_m = \frac{2\pi\delta}{A_m\delta - A_m + 2\pi} = \delta\eta_s \quad (10)$$

$$\eta_s = \frac{2\pi}{A_m\delta - A_m + 2\pi}. \quad (11)$$

Consequently, the channel gain of GJ ( $\eta_{g_j}$ ) is determined by  $\vartheta$  and the location of the MU  $u_m$  as follows:

$$\eta_{g_j} = \begin{cases} \eta_m, & \text{if } \vartheta \geq (90^\circ - \theta_{(g_j, u_m)}) \\ \eta_s, & \text{otherwise.} \end{cases} \quad (12)$$

### D. SJNR Calculation

Using the G2A channel model, A2A channel model, and 3-D two-step beam pattern model, the SJNR of the MU  $u_m$  ( $\Gamma_{u_m}$ ), which represents the effectiveness of the cooperative jamming

of GJ  $g_j$  and UJ  $u_j$  against the command and control signals from MG  $g_M$  to MU  $u_M$ , can be calculated as follows:

$$\Gamma_{u_M} = \frac{P_{g_M} \times \left( L_{(g_M, u_M)}^{G2A} \right)^{-1}}{P_{g_j} \times \left( L_{(g_j, u_M)}^{G2A} \right)^{-1} \times \eta_{g_j} + \Sigma_{u_j} \left( P_{u_j} \times \left( L_{(u_j, u_M)}^{A2A} \right)^{-1} \right) + \sigma^2}. \quad (13)$$

In particular,  $\Gamma_{u_M}$  is determined by three factors: the desired received signal strength from the MG, jamming signal strength from UJ and GJ, and additive white Gaussian noise.

### E. UAV Mobility Model

The MU  $u_M$  moves randomly in 3-D network space according to the 3-D random mobility model. The MU location  $u_M$  ( $(x_{u_M}(t), y_{u_M}(t), z_{u_M}(t))$ ) at time step  $t$  is obtained as

$$x_{u_M}(t) = x_{u_M}(t-1) + v_{u_M}(t) \sin \theta_{u_M}(t) \cos \phi_{u_M}(t) \quad (14)$$

$$y_{u_M}(t) = y_{u_M}(t-1) + v_{u_M}(t) \sin \theta_{u_M}(t) \sin \phi_{u_M}(t) \quad (15)$$

$$z_{u_M}(t) = z_{u_M}(t-1) + v_{u_M}(t) \cos \theta_{u_M}(t) \quad (16)$$

where  $t$  denotes the discrete time step at which the position of the malicious UAV (MU) is updated. Moreover,  $v_{u_M}(t)$  ( $0 \leq v_{u_M} \leq v_{u_M}^{\max}$ ),  $\theta_{u_M}(t)$  ( $0 \leq \theta_{u_M} \leq 180^\circ$ ), and  $\phi_{u_M}(t)$  ( $0^\circ \leq \phi_{u_M} \leq 360^\circ$ ) represent the velocity, elevation angle, and azimuth angle of the MU  $u_M$ , respectively. The equations use the velocity ( $v$ ), elevation angle ( $\theta$ ), and azimuth angle ( $\phi$ ) of the MU to compute its position iteratively at each time step. In addition,  $v_{u_M}(t)$ ,  $\theta_{u_M}(t)$ , and  $\phi_{u_M}(t)$  are randomly selected within the specified ranges.

## III. PROPOSED METHOD: HDRL-CJ

This study proposes HDRL-CJ, a novel cooperative jamming framework that integrates optimal beamwidth control of a GJ (DRL-GJ) with the movement and transmit power control of a UJ (HDRL-UJ), aiming to maximize jamming performance against MUs operating in friendly territories under the command and control of an MG. In the proposed HDRL-CJ, the GJ and UJ serve as the agents of DRL-GJ and HDRL-UJ, respectively. Each agent operates its own neural network and learns solely from its local state, which includes MU-related information and the corresponding control variables. Hence, no direct exchange of state information occurs between agents. Moreover, since transmit power and position control for the UJ involve considerable computational complexity, we design a hierarchical DRL architecture in which the learning process is divided into inner-loop RL for mobility control and outer-loop RL for transmit power control.

Fig. 2 shows the proposed HDRL-CJ framework, which integrates DRL-GJ and HDRL-UJ for cooperative jamming by GJ  $g_j$  and UJ  $u_j$ . In this framework, GJ  $g_j$  adjusts the beamwidth at each timestep. The UJ  $u_j$  performs mobility control at each time step using inner-loop RL and adjusts its transmission power every  $\tau$  time steps using outer-loop RL. To maintain analytical clarity and maintain the spotlight on communication-centric jamming performance, this article confines its energy budget to transmission, deferring propulsion-related considerations to future work. Accordingly,

we concentrate on the tradeoff between energy efficiency and jamming performance through the control of UJ's transmission power. This modeling choice allows us to reduce the overall system complexity while maintaining a clear focus on jamming performance as the core evaluation metric. The proposed HDRL-CJ optimizes jamming performance by assigning separate action spaces tailored to the specific roles of each agent. Specifically, the GJ is responsible for beamwidth control, while the UJ handles both mobility and transmission power control. This separation of action spaces allows for independent optimization of each jammer's functionalities within the hierarchical distributed framework, thereby enhancing overall learning efficiency and jamming effectiveness. A detailed explanation of the RL of GL, inner-loop RL of UJ, and outer-loop RL of UJ is as follows.

### A. MDP Design for DRL-GJ

The goal of DRL-GJ is to maximize the jamming effect by determining the optimal beamwidth. Specifically, as an agent, GJ  $g_j$  controls  $\vartheta$  at each time step. A detailed description of the state, action, and reward of DRL-GJ is as follows.

- 1) *State of DRL-GJ* ( $s_{g_j}(t)$ ): The state of the GJ  $s_{g_j}(t)$  comprises the beamwidth ( $\vartheta(t)$ ) and position ( $x_{u_M}(t), y_{u_M}(t), z_{u_M}(t)$ ) of MU  $u_M$

$$s_{g_j}(t) = \{\vartheta(t), x_{u_M}(t), y_{u_M}(t), z_{u_M}(t)\} \quad (17)$$

here, to enable the learning of dynamic network behavior,  $s_{g_j}(t)$  incorporates positional information, which is designed to implicitly capture directional changes or velocity of the MU as the position updates over time.

- 2) *Action of DRL-GJ* ( $a_{g_j}(t)$ ): As an action, GJ  $g_j$  controls its beamwidth. There are three actions such as increment, decrement, and stay between the minimum beamwidth  $\vartheta^{\min}$  and the maximum beamwidth  $\vartheta^{\max}$

$$a_{g_j}(t) \in \{\pm\Delta\vartheta, 0\}, \quad \vartheta^{\min} \leq \vartheta(t) \leq \vartheta^{\max}. \quad (18)$$

- 3) *Reward of DRL-GJ* ( $r_{g_j}(t)$ ): The reward of GJ  $g_j$  ( $r_{g_j}(t)$ ) is calculated based on  $\Gamma_{u_M}(t)$  obtained as a result of the proposed cooperative jamming. Basically, the GJ  $g_j$  performs DRL-based beamwidth control to minimize the achievable data rate “ $\text{BW} \times \log_2(1 + \Gamma_{u_M}(t))$ ” based on DRL-GJ. To convert it into a maximization problem, we consider “ $-\text{BW} \times \log_2(1 + \Gamma_{u_M}(t))$ ” as a reward as follows:

$$r_{g_j}(t) = -\text{BW} \times \log_2(1 + \Gamma_{u_M}(t)) \quad (19)$$

where BW is the frequency bandwidth.

### B. MDP Design for HDRL-UJ

1) *Inner-Loop RL of UJ*: The UJ  $u_j$  performs actions to optimize not only its movement via the inner-loop RL but also the amount of power transmitted via the outer-loop RL for maximizing its jamming effect against the MU  $u_M$ . The inner-loop RL in the proposed HDRL-UJ considers the UJ  $u_j$  as an agent and performs movement control to track the MU  $u_M$  at every time step, except when conducting the outer-loop

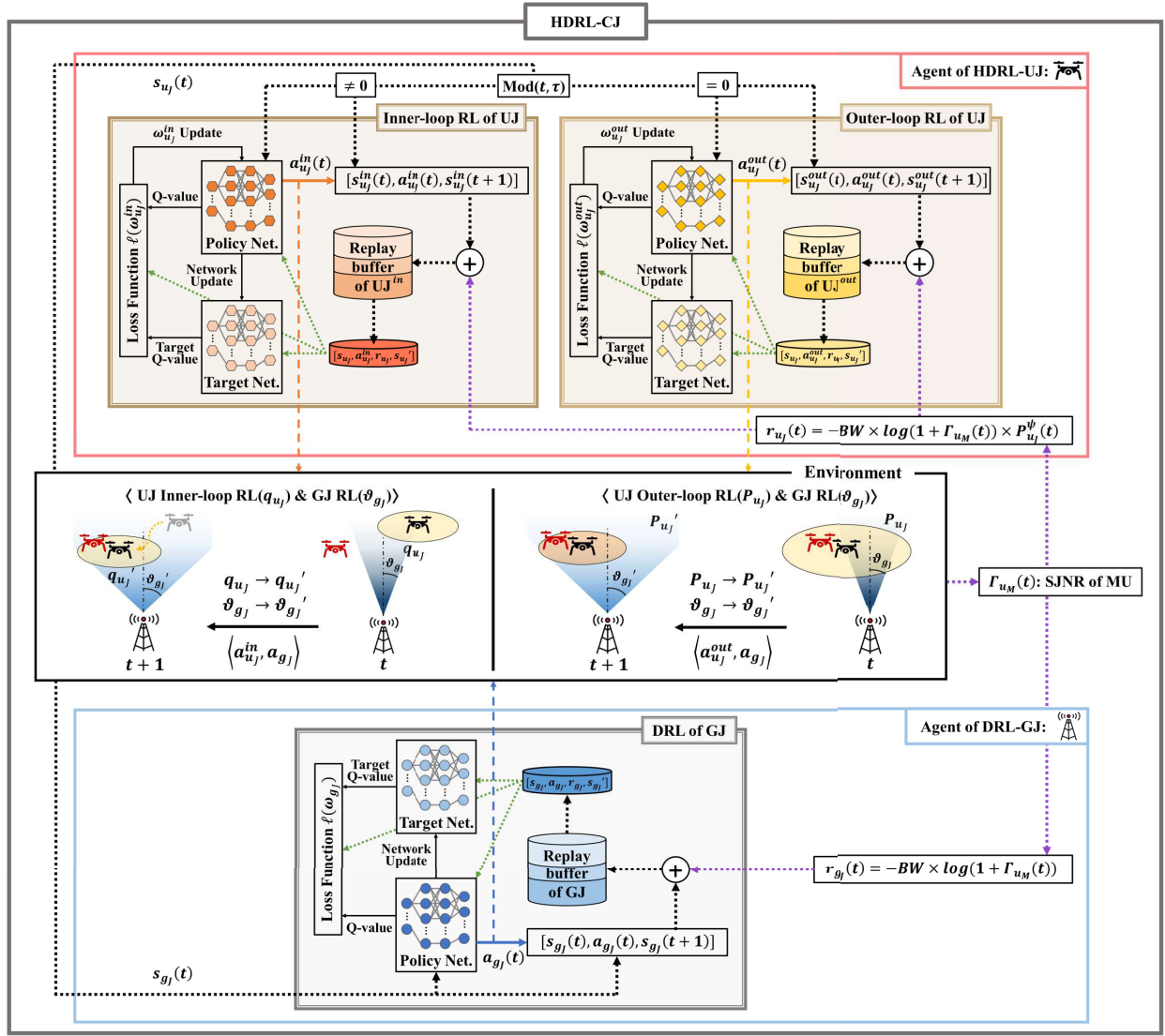


Fig. 2. HDRL-CJ framework for secure air-ground integrated networks. GJ receives the sender state from the environment and performs beamwidth control in DRL-GJ. UJ performs either movement control by UJ's inner-loop RL or transmit power control by UJ's outer-loop RL based on the state according to  $\text{Mod}(t, \tau)$ . When each RL executes, a sample set is stored in the corresponding replay buffer, which is used to update networks.

RL. A detailed explanation of the state and action of the UJ inner-loop RL is as follows.

- 1) *State of UJ Inner-Loop RL* ( $s_{u_j}^{\text{in}}(t)$ ): Considering the jamming power and the location of MU  $u_M$ ,  $s_{u_j}^{\text{in}}(t)$  is given by

$$s_{u_j}^{\text{in}}(t) = \left\{ P_{u_j}(t), x_{u_j}(t), y_{u_j}(t), z_{u_j}(t), \right. \\ \left. \times x_{u_M}(t), y_{u_M}(t), z_{u_M}(t) \right\}. \quad (20)$$

Specifically,  $s_{u_j}^{\text{in}}(t)$  comprises the jamming power of UJ  $u_j$  ( $P_{u_j}(t)$ ), Cartesian coordinate of UJ  $u_j$  ( $x_{u_j}(t), y_{u_j}(t), z_{u_j}(t)$ ), and Cartesian coordinate of MU  $u_M$  ( $x_{u_M}(t), y_{u_M}(t), z_{u_M}(t)$ ).

- 2) *Action of UJ Inner-Loop RL* ( $a_{u_j}^{\text{in}}(t)$ ): Here, the agent performs its movement control to track the MU movements

$$a_{u_j}^{\text{in}}(t) \in \left\{ \pm \Delta x_{u_j}, \pm \Delta y_{u_j}, \pm \Delta z_{u_j}, 0 \right\} \quad (21)$$

$$d_{(u_j, u_M)} \geq d^{\min} \quad (22)$$

$$X_u^{\min} \leq x_{u_j}(t) \leq X_u^{\max} \quad (23)$$

$$Y_u^{\min} \leq y_{u_j}(t) \leq Y_u^{\max} \quad (24)$$

$$Z_u^{\min} \leq z_{u_j}(t) \leq Z_u^{\max}. \quad (25)$$

Within a given altitude range, UJ  $u_j$  will move in the  $x$ -,  $y$ -, and  $z$ -axes by the specified moving speed  $v_{u_j}$  [m/iteration] or stay in its current position. If UJ  $u_j$  approaches the MU too closely, the UJ could be in danger on the battlefields. Therefore, UJ  $u_j$  should maintain a predefined minimum distance of  $d^{\min}$  from the MU  $u_M$ .

- 2) *Outer-Loop RL of UJ*: Every  $\tau$  time steps, the UJ  $u_j$  performs actions to optimize its transmit power via the outer-loop RL to maximize its jamming effect against MU  $u_M$ , where  $\tau$  is the predefined outer-loop triggering threshold. A detailed

explanation of the state and action of the UJ outer-loop RL is as follows.

- 1) *State of UJ Outer-Loop RL* ( $s_{u_j}^{\text{out}}(t)$ ): The state of UJ outer-loop RL ( $s_{u_j}^{\text{out}}(t)$ ) comprises the same as ( $s_{u_j}^{\text{in}}(t)$ ) as follows:

$$s_{u_j}^{\text{out}}(t) = \left\{ P_{u_j}(t), x_{u_j}(t), y_{u_j}(t), z_{u_j}(t), \right. \\ \left. \times x_{u_M}(t), y_{u_M}(t), z_{u_M}(t) \right\} = s_{u_j}^{\text{in}}(t). \quad (26)$$

- 2) *Action of UJ Outer-Loop RL* ( $a_{u_j}^{\text{out}}(t)$ ): Here, UJ  $u_j$  controls its jamming power ( $P_{u_j}(t)$ ) in the outer-loop RL. UJ  $u_j$  increases, decreases, or maintains the jamming power by  $\Delta P_{u_j}$  between the minimum jamming power ( $P_{u_j}^{\text{min}}$ ) and the maximum jamming power ( $P_{u_j}^{\text{max}}$ )

$$a_{u_j}^{\text{out}}(t) \in \left\{ \pm \Delta P_{u_j}, 0 \right\}, \quad P_{u_j}^{\text{min}} \leq P_{u_j}(t) \leq P_{u_j}^{\text{max}}. \quad (27)$$

3) *Unified Reward of UJ*: While GJ can receive permanent power through the power grid, UJ is a battery-powered device; therefore, the energy-efficient operation of UJ is essential. Therefore, HDRL-UJ aims to simultaneously maximize the term “ $-\text{BW} \times \log_2(1 + \Gamma_{u_M}(t))$ ” to enhance jamming performance and minimize  $P_{u_j}(t)$  to improve energy efficiency. To achieve this, we design a unified reward function that balances the tradeoff between SJNR and power consumption, enabling the simultaneous optimization of mobility and transmit power control. This optimization is carried out through a hierarchical learning framework, where both the inner-loop and outer-loop RL modules in the HDRL-UJ architecture use a shared unified reward  $r_{u_j}(t)$ , which is defined as follows:

$$r_{u_j}(t) = -\text{BW} \times \log_2(1 + \Gamma_{u_M}(t)) \times P_{u_j}^\psi(t). \quad (28)$$

If  $P_{u_j}$  is not included in the reward function, the UJ always transmits at maximum power to maximize jamming performance, which can lead to rapid battery depletion. Accordingly, the user can control the tradeoff between maximizing jamming performance and energy efficiency by adjusting the weighting factor  $\psi \in [0, 1]$  according to the network operational priority. In this article, we perform numerical simulations with various values of  $\psi$  to thoroughly analyze the tradeoff between jamming effectiveness and energy efficiency.

### C. Q-Function Update and Policy

To maximize the reward ( $r_{u_j}(t)$  and  $r_{g_j}(t)$ ) based on SJNR  $\Gamma_{u_M}(t)$  at time step  $t$ , the GJ determines the optimal beamwidth and the UJ determines the optimal position and transmit power. In HDRL-CJ, the Q function of the DNN is updated as follows:

$$Q(s(t), a(t)|\omega) \leftarrow (1 - \kappa)Q(s(t), a(t)|\omega) \\ + \kappa(r(t) + \rho \cdot \max_{a' \in \mathcal{A}(t+1)} Q(s(t+1), a'|\omega)) \quad (29)$$

where  $\kappa \in (0, 1]$  denotes the learning rate,  $\rho$  denotes the discount factor,  $\omega$  denotes the weight of the DNN, and  $a'$  denotes the action with the highest Q value in state  $s(t+1)$  at time step  $t+1$ . The reliability of the predicted Q value depends on the training degree of the neural network with weight  $\omega$ . To achieve the goal of HDRL-CJ, the target network trains

the learning network to approximate the predicted Q value to the target Q value. As the training progresses, the learning experience is stored in the form of  $(s(t), a(t), r(t), s(t+1))$ , and minibatches are sampled from the replay buffer. These minibatches were then used to update the weight  $\omega$  of the policy network. We define a loss function  $\ell(\omega)$  to approximate the predicted Q value to the target value, as follows:

$$\ell(\omega) = \sum_{t=0}^T \left[ Q(s(t), a(t)|\omega) - \left( r(t) + \rho \cdot \max_{a' \in \mathcal{A}(t+1)} Q(s(t+1), a'|\omega) \right) \right]^2. \quad (30)$$

The agents update their weights  $\omega$  to minimize the function  $\ell(\omega)$  using the values sampled at batch size  $T$ .

In addition, we used a decayed  $\varepsilon$ -greedy policy to allow each agent to fully explore different states early in learning [20], [21]. This policy can be expressed as

$$a(t) = \begin{cases} \text{Random action,} & \varepsilon(t) \\ \arg \max_{a' \in \mathcal{A}(t+1)} Q(s(t+1), a'|\omega), & 1 - \varepsilon(t). \end{cases} \quad (31)$$

The number of actions of an agent varies depending on the MDP design. The amount of exploration during the learning process was adjusted to ensure sufficient exploration. The  $\varepsilon(t)$ , which determines the ratio of an agent's exploration and exploitation at time step  $t$ , is calculated as

$$\varepsilon(t) = \varepsilon_{\text{init}} (1 - \varepsilon_{\text{init}})^{\frac{t}{\xi \times |\mathcal{A}|}} \quad (32)$$

where  $\varepsilon_{\text{init}}$  is the initial  $\varepsilon$  value. In addition,  $\xi$  denotes an exploration parameter that adjusts the attenuation rate of  $\varepsilon$ , and  $|\mathcal{A}|$  is the cardinality of the action set.

The HDRL-CJ agents are designed based on an MDP framework and use a deep neural network consisting of an input layer,  $N_H$  hidden layers, and an output layer. Each layer's nodes are defined  $\mathbf{l} = \{l_0, l_1, \dots, l_{N_H}, \dots, l_{N_H}, l_{N_H+1}\}$ . Here,  $l_0$  represents the input layer, corresponding to the size of the state space. The  $l_{N_H}$  hidden layers use rectified linear unit (ReLU) activation functions with a predefined number of nodes. The  $l_{N_H+1}$  represents the output layer, which corresponds to the size of the action space. The exact values for the number of each layer's nodes are determined based on the specific characteristics of each agent. This hierarchical distributed DRL framework, constructed using a deep neural network, allows GJ and UJ to effectively execute cooperative jamming operations tailored to dynamic mobility scenarios.

The proposed HDRL-CJ method randomly places MU  $u_M$  and UJ  $u_j$  in a given network area and initializes the jamming power of UJ  $u_j$  and the beamwidth of the GJ to random values within predefined ranges. During each iteration, the GJ  $g_j$  performs beamwidth control, whereas the UJ  $u_j$  performs movement or power control via the inner or outer-loop RL, depending on  $\tau$ . In each episode, MU  $u_M$  randomly moves based on the UAV mobility model with a maximum velocity of  $v_{u_M}^{\text{max}}$ . The detailed process of HDRL-CJ is given in Algorithm 1.

**Algorithm 1** Detailed Operational Procedure of HDRL-CJ

---

```

1: Initialization  $\vartheta$  and  $P_{u_j}$ , random distribution of UJ  $u_j$  and
   MU  $u_m$ .
2: for  $e = 1$ : Number of episodes do
3:   for  $i = 1$ : Number of iterations do
4:     Update time slot  $t = t + 1$ .
5:     Calculate state  $s_{u_j}(t)$  and  $s_{g_j}(t)$  of agent UJ  $u_j$  and
       GJ  $g_j$ .
6:     GJ  $g_j$  chooses actions  $a_{g_j}(t)$  using the decayed  $\varepsilon$ -
       greedy policy and adjusts its  $\vartheta(t)$  accordingly.
7:     if  $\text{Mod}(i, \tau) = 0$  then
8:       UJ  $u_j$  chooses the outer-loop action  $a_{u_j}^{\text{out}}(t)$  using
       the decayed  $\varepsilon$ -greedy policy and adjust its  $P_{u_j}(t)$ 
       accordingly.
9:       Calculate  $\Gamma_{u_m}(t)$  and reward  $r_{u_j}(t)$ ,  $r_{g_j}(t)$ .
10:      Update next state  $s_{u_j}(t+1)$ ,  $s_{g_j}(t+1)$  and weight
         $\omega_{u_j}^{\text{out}}, \omega_{g_j}$ .
11:      Save samples  $[s(t), a(t), r(t), s(t+1)]$  to the replay
        memory of UJ  $u_j$  and GJ  $g_j$ , respectively.
12:     else
13:       UJ  $u_j$  chooses inner-loop action  $a_{u_j}^{\text{in}}(t)$  using the
       decayed  $\varepsilon$ -greedy policy and moves its position
       accordingly.
14:       Calculate  $\Gamma_{u_m}(t)$  and reward  $r_{u_j}(t+1)$ ,  $r_{g_j}(t+1)$ .
15:       Update next state  $s_{u_j}(t+1)$ ,  $s_{g_j}(t+1)$  and weight
         $\omega_{u_j}^{\text{in}}, \omega_{g_j}$ .
16:       Save samples  $[s(t), a(t), r(t), s(t+1)]$  to the replay
        memory of UJ  $u_j$  and GJ  $g_j$ , respectively.
17:     end if
18:   end for
19:   MU  $u_m$  moves randomly with UAV mobility model.
20: end for

```

---

## IV. SIMULATION RESULTS

In this study, we consider a scenario in which a single GJ is located at (70, 70, 0) [m], and one or more UJs operate within a 140 [m]  $\times$  140 [m] area to jam the signal received by the MU from the MG positioned at (20, 20, 0) [m]. To ensure consistency in the learning environment and to enable a clear analysis of the performance of the proposed HDRL-CJ method, both  $P_{g_j}$  and  $P_{g_m}$  are fixed at 10 [W] [30], [31], [32]. Specifically, in this network, the MG functions as an adversarial ground node that transmits command signals to the MU. By fixing  $P_{g_m}$ , we eliminate ambiguity in the MU's received signal that could otherwise arise simultaneously from both MU mobility and fluctuations in  $P_{g_m}$ . Meanwhile, the GJ can dynamically adjust its jamming coverage and intensity through beamwidth control based on a two-step beam pattern, and thus operates with a fixed transmit power  $P_{g_j}$ . This configuration enables a clear isolation of the cooperative jamming control effects during training and ensures the stability of the learning process. In particular, the MU moves according to a 3-D random mobility model and receives command signals from the MG. In addition, UJ and MU adjusted their altitudes to within the range of  $[h^{\min}, h^{\max}]$ . The other simulation parameters are listed in Table II. Furthermore, to evaluate the jamming performance

TABLE II  
SIMULATION PARAMETERS

Parameter	Value
Carrier frequency ( $f$ )	2.4 GHz
Bandwidth size ( $BW$ )	20 MHz
Thermal noise power ( $\sigma^2$ )	-110 dB
Excessive path loss ( $\zeta^{LoS}, \zeta^{NLoS}$ )	0.1, 21 dB
Suburban's parameter ( $\alpha, \beta$ )	4.88, 0.429
Beamwidth angle interval ( $\Delta\vartheta$ )	5 °
Maximum beamwidth ( $\vartheta^{\max}$ )	45 °
Minimum beamwidth ( $\vartheta^{\min}$ )	20 °
Lobe gain factor ( $\delta$ )	40
Transmission power of MG ( $P_{g_m}$ )	10 W
Transmission power of GJ ( $P_{g_j}$ )	10 W
Transmission power of UJ ( $P_{u_j}$ )	1~5 W
Maximum altitude ( $h^{\max}$ )	160 m
Minimum altitude ( $h^{\min}$ )	100 m
Minimum distance btw. UJ and MU ( $d^{\min}$ )	20 m
Outer-loop RL period ( $\tau$ )	5 iterations
Initial epsilon value ( $\varepsilon_{\text{init}}$ )	0.99
Exploration parameter ( $\xi$ )	12
Learning rate ( $\kappa$ )	0.0001
Discount factor ( $\rho$ )	0.5

of HDRL-CJ, several methods are considered as benchmarks. These include random action (RAND), GJ only, UJ only, a predefined trajectory of UJ (PT), without reward-sharing (w/o R-S), centralized DRL (C-DRL), and QL. A detailed description of the benchmark methods is provided below.

- 1) *(BM1) RAND*: The RAND randomly controls the beamwidth of the GJ and the transmit power and position of the UJ at each iteration. RAND is helpful in analyzing the convergence behavior of HDRL-CJ.
- 2) *(BM2) GJ Only*: A conventional method using only GJ, where GJ performs jamming through RL-based beamwidth control according to the location of the MU.
- 3) *(BM3) UJ Only*: A conventional method using only UJ, where UJ performs jamming through RL-based transmit power and movement control according to the location of the MU.
- 4) *(BM4) PT*: The UJ is executing cooperative jamming maneuvers within a predefined moving path with dimensions of 100 [m]  $\times$  100 [m]. Here, UJ performs only transmit power control.
- 5) *(BM5) w/o R-S*: A variant of HDRL-CJ without reward sharing, where the DRL-GJ and HDRL-UJ learn independently of their respective local rewards using separate neural networks, without considering each other's jamming influence.
- 6) *(BM6) C-DRL*: Using a centralized DRL framework, a central controller manages both GJ and UJ based on the network information.
- 7) *(BM7) QL*: A QL-based method where each agent shares the same MDP elements as HDRL-CJ.

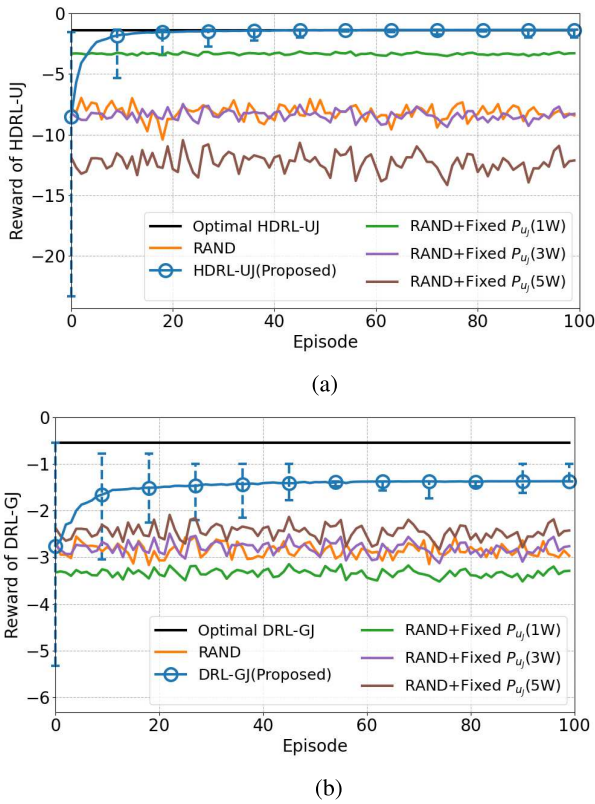


Fig. 3. Accumulated average reward of the proposed and benchmark methods. (a) Average reward of HDRL-UJ versus episode. (b) Average reward of DRL-GJ versus episode.

As with other DRL-based approaches, the proposed HDRL-CJ relies on high-dimensional nonlinear function approximation, making it challenging to formally guarantee convergence to a globally optimal solution in dynamic environments [33], [34]. Accordingly, we first rigorously evaluate the fundamental performance of HDRL-CJ in a static environment, including convergence and objective-driven behavior, and verify its optimality by comparing the results with the true optimum obtained through exhaustive search. Then, we demonstrate the practical superiority of the proposed framework in more complicated and dynamic environments through comparisons to representative benchmark methods. Fig. 3 demonstrates the convergence of the proposed jamming techniques in a static environment. Simulations were conducted over 100 episodes with 5000 iterations per episode. The initial positions of the UJ  $u_j$  and MU  $u_m$  were randomly assigned, while the MU's position was kept fixed during training. To verify the convergence of HDRL-UJ and DRL-GJ, the ground-truth optimum was obtained through an exhaustive search procedure. As baselines, RAND refers to a policy that randomly controls the GJ beamwidth and the UJ's transmit power and position at each iteration. In contrast, RAND + fixed  $P_{u_j}$  fixes the outer-loop transmit power of HDRL-UJ while randomizing the remaining control variables. In Fig. 3(a), the error bars of HDRL-UJ decrease progressively as training proceeds, indicating reduced variance and stable convergence. In Fig. 3(b), DRL-GJ converges to a value slightly below the exhaustive-search

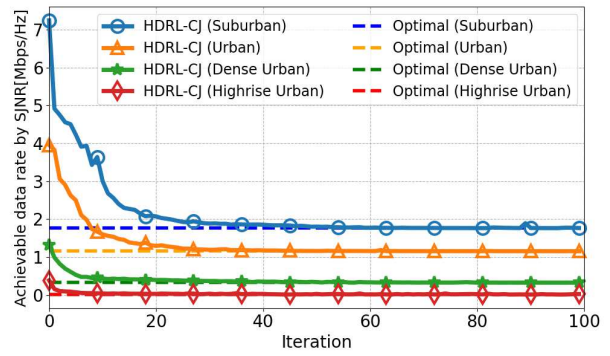


Fig. 4. Achievable data rate by SJNR of HDRL-CJ according to various network environments.

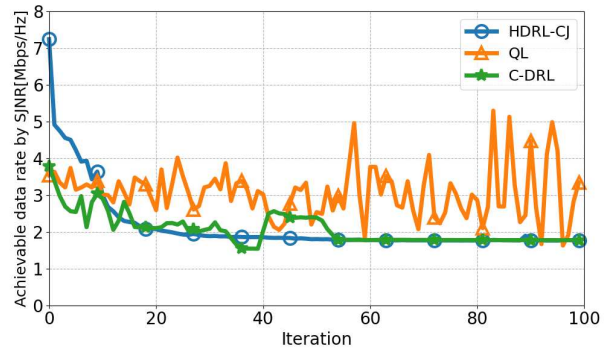


Fig. 5. Achievable data rate by SJNR of HDRL-CJ, C-DRL, and QL.

optimum, since its reward function accounts for the achievable data rates by SJNR for both the GJ and UJ. Consequently, HDRL-UJ and DRL-GJ outperform RAND and RAND + fixed  $P_{u_j}$  in terms of achievable data rates by SJNR.

Fig. 4 shows the simulation results of the proposed HDRL-CJ scheme in four representative environments: suburban, urban, dense urban, and high-rise urban. In all the scenarios, the achievable data rate of HDRL-CJ gradually converged to the optimal value through the learning process, demonstrating the robustness and adaptability of the proposed cooperative jamming strategy across diverse channel conditions. Notably, in environments with higher density, the data rate under SJNR conditions was observed to be relatively lower, while also exhibiting faster convergence to the optimal value. This effect can be attributed to the multipath propagation environment, which acts not only as a source of interference but also as a factor that partially enhances the jamming effectiveness.

Fig. 5 resents a performance comparison between the proposed HDRL-CJ, designed with a hierarchical distributed framework, and conventional methods such as C-DRL and QL. The simulation results show that the achievable data rate by SJNR of HDRL-CJ rapidly converges to the optimal value within the given training period. While C-DRL also converges to the same optimal value, its convergence rate is slower than that of HDRL-CJ. In contrast, QL fails to converge within the same training time due to the exponential growth of the Q-table with the network size. The superior performance of HDRL-CJ originates from its hierarchical design, which effectively addresses the joint optimization

TABLE III  
SJNR AND ACHIEVABLE DATA RATE BY SJNR FOR HDRL-CJ  
ACCORDING TO PATH OF MU

Path of MU	SJNR[dB]	Achievable data rate by SJNR[Mbps/Hz]
Random path	-11.33	2.08
Pre-defined path 1	-10.77	2.41
Pre-defined path 2	-9.75	2.68

of the UJ's mobility and jamming power. By separating these physically distinct control actions into inner and outer loops, the proposed model not only captures their interdependence more precisely but also inherently reduces the dimensionality of the learning space, thereby mitigating the high computational complexity that would arise if they were treated as a single flat optimization problem. In contrast, QL suffers from the exponential growth of the Q-table, while C-DRL incurs substantial computational overhead due to its centralized structure. Consequently, both C-DRL and QL exhibit significantly higher computational complexity than HDRL-CJ, leading to much longer training times. In addition, we provide a quantitative analysis of the computational complexity of HDRL-CJ. The complexity is given by  $O(|g_J|(4l_1 + \sum_{k=1}^{N_H-1} l_k l_{k+1} + 3l_{N_H}) + |u_J| \times (7l_1 + \sum_{k=1}^{N_H-1} l_k l_{k+1} + 3l_{N_H}) (\frac{\tau-1}{\tau}) + (7l_1 + \sum_{k=1}^{N_H-1} l_k l_{k+1} + 7l_{N_H}) \frac{1}{\tau})$ . This expression shows that the complexity of HDRL-CJ scales linearly with the number of agents when considering the hierarchical learning cycle. In contrast, nonhierarchical structures have the following complexity:  $O(|g_J|(4l_1 + \sum_{k=1}^{N_H-1} l_k l_{k+1} + 3l_{N_H}) + |u_J|(7l_1 + \sum_{k=1}^{N_H-1} l_k l_{k+1} + 9l_{N_H}))$ , which imposes a greater computational burden than hierarchical designs. Moreover, the centralized nature of C-DRL leads to a complexity of  $O(8(|g_J| + |u_J|)l_1 + \sum_{k=1}^{N_H-1} l_k l_{k+1} + 2|g_J|8^{|u_J|}l_{N_H})$  which grows exponentially with the number of agents. These results confirm that HDRL-CJ achieves efficient optimization of heterogeneous control actions with significantly lower computational complexity. This makes it a practical cooperative jamming solution for scalable network environments.

Table III presents the SJNR and the corresponding achievable data rate of the proposed HDRL-CJ technique under different MU mobility patterns. The MU is configured to move at a maximum speed of  $v_{u_M}^{\max} = 10$ , with three types of mobility patterns considered: a random path, a rectangular path (predefined path 1), and a zigzag path (predefined path 2). In predefined path 1, the MU maintains a constant altitude while following a square trajectory of 80 [m]  $\times$  80 [m], whereas in predefined path 2, it moves through a 90 [m]  $\times$  90 [m] area in a zigzag pattern. The simulation results show achievable data rates by SJNR of 2.08, 2.41, and 2.68 [Mbps/Hz] for the random path, predefined path1, and predefined path2, respectively. These results demonstrate that the proposed HDRL-CJ consistently maintains high jamming performance while effectively adapting to both the structured and unpredictable MU mobility scenarios, thereby proving its flexibility and robustness in dynamic environments.

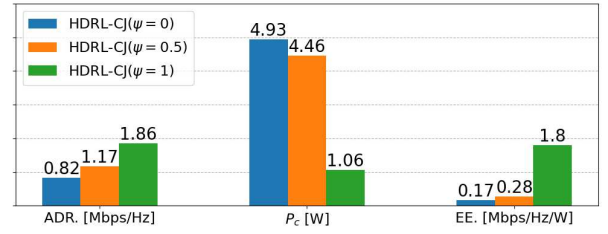


Fig. 6. Simulation results by weight of unified reward for UJ  $\psi$ .

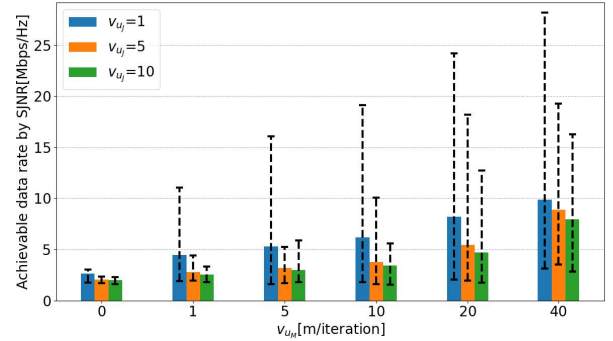


Fig. 7. Achievable data rate by SJNR of HDRL-CJ according to variation  $v_{u_J}$  versus  $v_{u_M}^{\max}$ .

Fig. 6 presents the simulation results according to the weighting parameter  $\psi$  in the unified reward function of the UJ. As  $\psi$  approaches 0, HDRL-CJ prioritizes maximizing jamming performance, whereas as  $\psi$  approaches 1, energy efficiency becomes the primary objective. Simulation results show that when  $\psi = 1$ , the achievable data rate increased by 1.04 [Mbps/Hz] compared with the case when  $\psi = 0$ , while power consumption decreased by 3.87 [W]. Consequently, energy efficiency improved by 1.63 [Mbps/Hz/W]. These findings demonstrate that the proposed HDRL-CJ framework can effectively perform cooperative jamming while adequately considering the battery constraints of the UJ. Therefore, depending on the tactical network environment, the proposed unified reward function enables flexible adjustment of the tradeoff between jamming performance and power consumption.

To analyze the cooperative jamming performance of the trained HDRL-CJ, Fig. 7 shows the average achievable data rate of the MU using the UJ. Simulations were performed for 100 episodes, with each episode comprising 100 iterations. In every episode, the position of UJ, the jamming power of UJ, and beamwidth of GJ are randomly initialized. In addition, at each iteration, the MU  $u_M$  randomly moves within the 3-D network area, the UJ performs either movement or transmit power control, and the GJ performs beamwidth adjustment. The MU  $u_M$  and UJ  $u_J$  moves with a maximum velocity  $v_{u_M}^{\max} \in \{0, 1, 5, 10, 20, 40\}$  [m/iteration] and  $v_{u_J} \in \{1, 5, 10\}$  [m/iteration] When  $v_{u_M}^{\max} = 0$ , the achievable data rate of  $v_{u_J} = 5$  was 2.5% higher than the achievable data rate of  $v_{u_J} = 10$ . However, the achievable data rate by SJNR for  $v_{u_J} = 1$  is 27.9% higher. This gap can be because when  $v_{u_J} = 1$ , the time required for a randomly placed UJ to approach the minimum

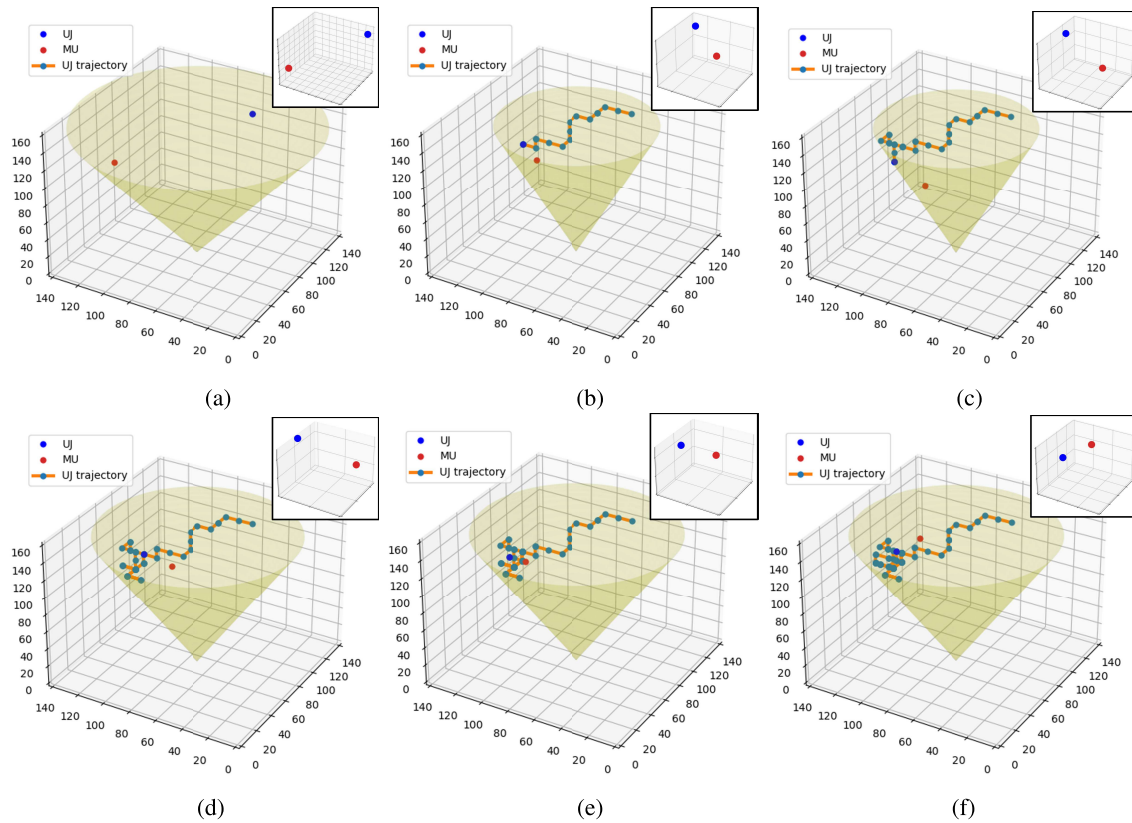


Fig. 8. Illustration of trajectory of UJ, beamwidth of GJ, and location of MU. (a) 1st iteration. (b) 20th iteration. (c) 40th iteration. (d) 60th iteration. (e) 80th iteration. (f) 100th iteration.

distance to the MU increases, which subsequently increases the overall achievable data rate by SJNR. Moreover, because  $v_{u_M}^{\max}$  increases, the performance gap between the achievable data rates by SJNR of  $v_{u_j} = 5$  and  $v_{u_j} = 10$  increases gradually. When  $v_{u_M}^{\max} = 10$ , the achievable data rate of  $v_{u_j} = 1$  is 78.33% higher than the achievable data rate of  $v_{u_j} = 10$ , and the achievable data rate of  $v_{u_j} = 5$  is 5.83% higher. As the value of  $v_{u_M}^{\max}$  increases, it becomes increasingly challenging to maintain the minimum distance between them. In other words, the gap in the achievable data rate according to  $v_{u_j}$  increases when  $v_{u_M}^{\max}$  increases. In summary, when  $v_{u_j}$  exceeds  $v_{u_M}^{\max}$ , a stable jamming performance is achieved. Moreover, as the MU's speed exceeds and further increases beyond  $v_{u_j}$ , HDRL-CJ's performance tends to decline gradually. Nonetheless, when the MU's speed increased to  $v_{u_M}^{\max} = 40$ , the achievable data rate was 9.89 [Mbps/Hz], which is 64.89% lower than the maximum value of 28.17 [Mbps/Hz]. This demonstrates that HDRL-CJ still delivers meaningful jamming performance even in more complex environments.

Fig. 8 illustrates the position of the MU, the movement trajectory of the UJ, and the beamwidth adjustments of the GJ during a specific episode where  $v_{u_j} = 10$  and  $v_{u_M}^{\max} = 10$ . In this scenario, the MU and UJ are initially located at (30, 110, 120) [m] and (90, 40, 160) [m], respectively, as shown in Fig. 8(a). Throughout the iterations, the UJ and GJ continuously optimize jamming performance based on the real-time movement of the MU. At iteration 20, the MU moves to (36, 83, 127) [m], and the UJ repositions to (50, 100, 130) [m], reducing the distance between them to 22.7 [m],

as shown in the inset of Fig. 8(b). Simultaneously, the GJ adjusts its beamwidth to an optimal value of  $25^\circ$ , centered at  $15.8^\circ$ . At iteration 40, the MU further moves to (41, 80, 101) [m], and the GJ reconfigures its beamwidth to  $25^\circ$ , now centered at  $16.6^\circ$ . Meanwhile, the UJ is located at (40, 100, 120) [m], resulting in an increased distance of 27.5 [m] from the MU. Despite these dynamic changes, the MU remains within the main-lobe region of the GJ for 86.67% of the simulation duration, and the average distance between the MU and UJ is maintained at 29.3 [m]. These results demonstrate the effectiveness of the proposed HDRL-CJ framework in adapting to previously unseen MU movement patterns while maintaining robust cooperative jamming performance.

Table IV presents the simulation results of the SJNR and achievable data rate by SJNR of the MU when the number of episodes is 100, the number of iterations is 100,  $v_{u_M}^{\max} = 10$ , and  $v_{u_j} = 10$ . The average achievable data rates by SJNR for the GJ only, UJ only, and w/o R-S methods were 46.15%, 348.08%, and 136.54% higher than those of HDRL-CJ, respectively. This result empirically demonstrates the superiority of HDRL-CJ's cooperative jamming performance over benchmark methods that use only a single type of jammer or operate without reward sharing. In contrast, the PT method achieved a 37.50% higher achievable data rate by SJNR despite using both UJ and GJ. This result shows the importance of optimizing the movement of the UJ in A2A jamming.

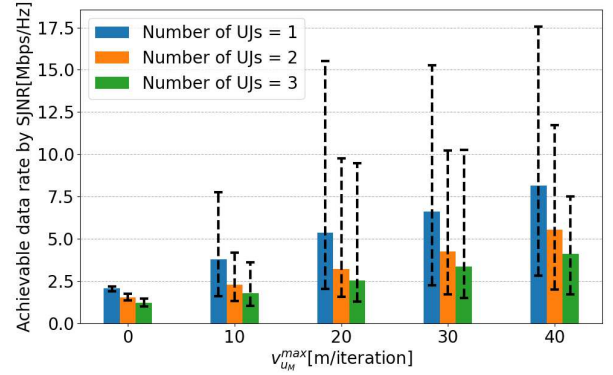
As the moving speed of the MU increases, it becomes increasingly challenging to adjust the beamwidth of the GJ and

TABLE IV

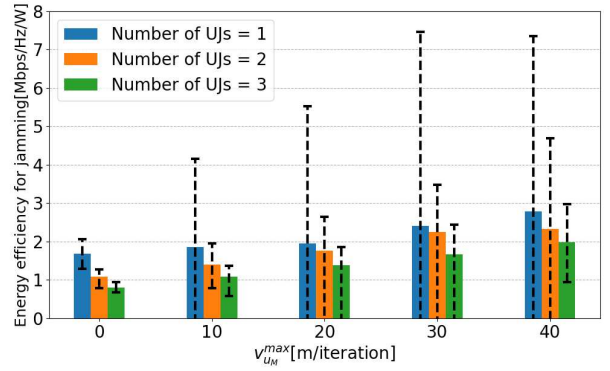
SJNR AND ACHIEVABLE DATA RATE BY SJNR FOR HDRL-CJ AND BENCHMARK METHODS UNDER  $v_{u_j} = 10$  [M/ITERATION],  $v_{u_M}^{\max} = 10$  [M/ITERATION]

Benchmark method	SJNR[dB]	Achievable data rate by SJNR[Mbps/Hz]
HDRL-CJ	-11.33	2.08
GJ only	-9.57	3.04
UJ only	-5.60	9.17
w/o R-S	-7.39	4.92
PT	-10.77	2.41

control the movement of the UJ to respond in real time to the movement of the MU. One potential solution to this problem is to increase the number of UJs, provided their moving speeds are limited. Fig. 9 illustrates the average achievable data rate of the MU based on SJNR and energy efficiency, according to the variations in the MU's maximum speed, under the conditions of  $v_{u_j} = 10$ , 100 episodes, and 100 iterations. Since the proposed HDRL-CJ is based on a distributed learning framework, each UJ uses a pretrained model learned in a single-UJ environment and dynamically optimizes its jamming strategy in multi-UJ scenarios using its individual jamming power and positional information. Fig. 9(a) shows the performance stability of HDRL-CJ under dynamic scenarios where  $v_{u_M}^{\max}$  exceeds  $v_{u_j}$ . First, when the MU is stationary (i.e.,  $v_{u_M}^{\max} = 0$ ), the achievable data rate by SJNR with a single UJ and GJ is 2.08 [Mbps/Hz]. When the number of UJs is increased to two and three while keeping the number of GJs fixed at one, the achievable data rate decreases by 0.54 and 0.87 Mbps/Hz, resulting in 1.54 and 1.21 Mbps/Hz, respectively. This clearly indicates that increasing the number of UJs enhances the jamming performance, thereby more effectively degrading the MU's communication capability. Meanwhile, when there is a single UJ and the speeds of the MU and UJ are equal (i.e.,  $v_{u_M}^{\max} = v_{u_j}$ ), the achievable data rate by SJNR increases to 3.80 [Mbps/Hz], indicating a degradation in jamming performance compared with the stationary MU case. As  $v_{u_M}^{\max}$  increases further, the distance between the MU and UJ grows, resulting in a gradual decline in jamming effectiveness for a single UJ. Nevertheless, even when the MU moves at four times the speed of the UJ, the data rate remains at 8.16 [Mbps/Hz], which is 2.1 times higher than in the equal-speed case and corresponds to 46.5% of the maximum value, 17.56 [Mbps/Hz]. These results demonstrate that despite limitations imposed by speed disparities between the MU and UJ, HDRL-CJ can still achieve a significant level of jamming performance within the given environment. When  $v_{u_M}^{\max} = 40$ , the achievable data rates by SJNR with one, two, and three UJs increase by 3.9, 3.6, and 3.4 times, respectively, compared with the case where  $v_{u_M}^{\max} = 0$ . As shown in Fig. 9(b), at  $v_{u_M}^{\max} = 40$ , the energy efficiency decreases by 16.94% and 29.10% when using two and three UJs, respectively. However, the achievable data rates by SJNR with two and three UJs are observed to decrease by 31.96% and 49.69%, respectively, compared with that with a single UJ, indicating an improvement in jamming performance. These results confirm that increasing



(a)



(b)

Fig. 9. Simulation results of HDRL-CJ according to variation number of UJs versus  $v_{u_M}^{\max}$ . (a) Achievable data rate by SJNR of HDRL-CJ. (b) Energy efficiency of HDRL-CJ.

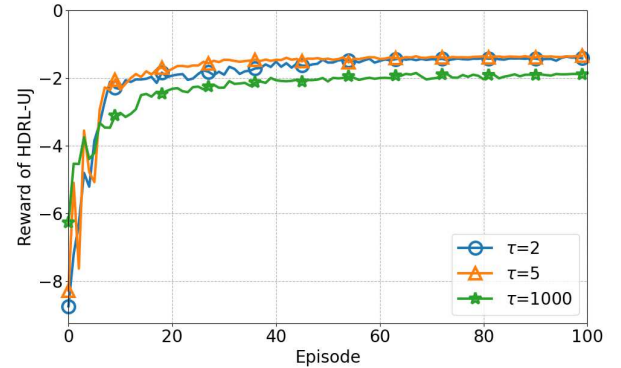


Fig. 10. Accumulated reward of UJ for HDRL-CJ versus episode according to the variation in  $\tau$ .

the number of UJs not only provides an effective countermeasure against high-speed MU mobility but also offers practical advantages in terms of expanding jamming coverage and maintaining performance, even when some UJs are detected.

Fig. 10 shows the reward of UJ  $u_j$  with a hierarchical RL framework based on the variations in  $\tau$ . The hierarchical RL framework can reduce the overall computational complexity when determining the optimal movement and transmit power control. When  $\tau = 2$ , more time is required for the UJ  $u_j$  until the optimal position is found in the entire 3-D space; this is

because the transmit power control and movement control are performed alternately once each. However, when  $\tau = 1000$ , UJ  $u_j$  may not be able to determine the optimal transmit power, meaning that the UJ herein should only perform movement control. Therefore, determining the optimal  $\tau$  value depending on the network situation is very important.

## V. CONCLUSION

In this study, HDRL-CJ was proposed for cooperative jamming based on a hierarchical DRL architecture in secure air-ground integrated networks. HDRL-CJ uses G2A and A2A jamming via the GJ and UJ to maximize the jamming performance against the MUs operating within friendly territories. To achieve this goal, the GJ controls the beamwidth by considering a tradeoff between jamming coverage and strength. In addition, the UJ performs optimal movement control to track the MU and optimal transmit power control to overcome limited battery power. The simulation results demonstrated that HDRL-CJ could find the optimal solution during the simulation time, whereas nonhierarchical RL-based methods failed to achieve this goal. Furthermore, HDRL-CJ outperformed several benchmark methods, and it was found that additional UJs and their faster movement can enhance the jamming performance. Based on these results, we can conclude that the proposed HDRL-CJ method can be adaptively applied to various military network environments. We hope that this study will help in developing robust and secure air-ground integrated networks in the future.

In future work, we plan to more realistically reflect the power consumption characteristics of battery-constrained UJs and the mobility patterns of MUs. For example, by incorporating movement-related energy consumption and MU speed into the MDP design, we aim to develop an effective jamming strategy that considers energy efficiency even in complex environments.

## REFERENCES

- [1] A. Kott, A. Swami, and B. J. West, "The Internet of Battle Things," *Computer*, vol. 49, no. 12, pp. 70–75, Dec. 2016.
- [2] I. Agadacos et al., "Application of trust assessment techniques to IoBT systems," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Nov. 2019, pp. 833–840.
- [3] S. Russell and T. Abdelzaher, "The Internet of Battlefield Things: The next generation of command, control, communications and intelligence (C3I) decision-making," in *Proc. IEEE Mil. Commun. Conf. (MILCOM)*, Oct. 2018, pp. 737–742.
- [4] (2022). *Markets and Markets (2022) 'Electronic Warfare Market by Capability, Platform, Product, End Use and Region • Global Forecast to 2027'*. [Online]. Available: <https://www.marketsandmarkets.com/Market-Reports/electronic-warfare-market-1301.html>
- [5] J. R. Hoehn, "Defense primer: Electronic warfare," Nov. 2022, Art. no. IF11118. [Online]. Available: <https://www.congress.gov/crs-product/IF11118>
- [6] W. Khalid, M. A. U. Rehman, T. Van Chien, Z. Kaleem, H. Lee, and H. Yu, "Reconfigurable intelligent surface for physical layer security in 6G-IoT: Designs, issues, and advances," *IEEE Internet Things J.*, vol. 11, no. 2, pp. 3599–3613, Jan. 2024.
- [7] K. Grover, A. Lim, and Q. Yang, "Jamming and anti-jamming techniques in wireless networks: A survey," *Int. J. Ad Hoc Ubiquitous Comput.*, vol. 17, no. 4, p. 197, Dec. 2014.
- [8] H. Guo, J. Li, J. Liu, N. Tian, and N. Kato, "A survey on space-air-ground-sea integrated network security in 6G," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 53–87, 1st Quart., 2022.
- [9] J. Feng, E. L. Pasilio, W. E. Dixon, and J. M. Shea, "An optimal jamming strategy to partition a wireless network," in *Proc. IEEE Mil. Commun. Conf.*, Oct. 2015, pp. 978–984.
- [10] S. Gezici, S. Bayram, M. N. Kurt, and M. R. Gholami, "Optimal jammer placement in wireless localization systems," *IEEE Trans. Signal Process.*, vol. 64, no. 17, pp. 4534–4549, Sep. 2016.
- [11] G. Kim and H. Lim, "Reinforcement learning based beamforming jammer for unknown wireless networks," *IEEE Access*, vol. 8, pp. 210127–210139, 2020.
- [12] Y. Huo, Y. Wu, R. Li, Q. Gao, and X. Luo, "A learning-aided intermittent cooperative jamming scheme for nonslotted wireless transmission in an IoT system," *IEEE Internet Things J.*, vol. 9, no. 12, pp. 9354–9366, Jun. 2022.
- [13] J. Li, X. Lei, P. D. Diamantoulakis, L. Fan, and G. K. Karagiannidis, "Security optimization of cooperative NOMA networks with friendly jamming," *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 13422–13428, Dec. 2022.
- [14] A. Li, Q. Wu, and R. Zhang, "UAV-enabled cooperative jamming for improving secrecy of ground wiretap channel," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 181–184, Feb. 2019.
- [15] C. Zhong, J. Yao, and J. Xu, "Secure UAV communication with cooperative jamming and trajectory control," *IEEE Commun. Lett.*, vol. 23, no. 2, pp. 286–289, Feb. 2019.
- [16] C. O. Nnamani, M. R. A. Khandaker, and M. Sellathurai, "UAV-aided jamming for secure ground communication with unknown eavesdropper location," *IEEE Access*, vol. 8, pp. 72881–72892, 2020.
- [17] H. Dang-Ngoc et al., "Secure swarm UAV-assisted communications with cooperative friendly jamming," *IEEE Internet Things J.*, vol. 9, no. 24, pp. 25596–25611, Dec. 2022.
- [18] X. A. F. Cabezas, D. P. M. Osorio, and M. Juntti, "A multi-armed bandit framework for efficient UAV-based cooperative jamming coverage," *IEEE Trans. Veh. Technol.*, vol. 72, no. 12, pp. 16893–16898, Dec. 2023.
- [19] S. Liu, J. Wu, and J. He, "Dynamic multichannel sensing in cognitive radio: Hierarchical reinforcement learning," *IEEE Access*, vol. 9, pp. 25473–25481, 2021.
- [20] S. Lee, S. Lim, S. H. Chae, B. C. Jung, C. Y. Park, and H. Lee, "Optimal frequency reuse and power control in multi-UAV wireless networks: Hierarchical multi-agent reinforcement learning perspective," *IEEE Access*, vol. 10, pp. 39555–39565, 2022.
- [21] E. Kim, J. Kim, J.-H. Kim, and H. Lee, "HiMAQ: Hierarchical multi-agent Q-learning-based throughput and fairness improvement for UAV-aided IoT networks," *J. Netw. Comput. Appl.*, vol. 223, Mar. 2024, Art. no. 103813.
- [22] R. Curpen, T. Bălan, I. A. Micioș, and I. Comănici, "Assessment of signal jamming efficiency against LTE UAVs," in *Proc. Int. Conf. Commun. (COMM)*, Jun. 2018, pp. 367–370.
- [23] B. J. Willner, "Methods and apparatuses for detecting and neutralizing remotely activated explosives," U.S. Patent 12 126 570, Nov. 26, 2009.
- [24] S. Nava. (Mar. 7, 2023). *101st Airborne Division Conducts Counter-UAS Training in Preparation for Future Fight*. Accessed: Apr. 15, 2025. [Online]. Available: [https://www.army.mil/article/264312/101st\\_airborne\\_division\\_conducts\\_counter\\_uas\\_training\\_in\\_preparation\\_for\\_future\\_fight](https://www.army.mil/article/264312/101st_airborne_division_conducts_counter_uas_training_in_preparation_for_future_fight)
- [25] J. Shi, L. Zhu, N.-W. Liu, and W. Wu, "A microstrip Yagi antenna with an enlarged beam tilt angle via a slot-loaded patch reflector and pin-loaded patch directors," *IEEE Antennas Wireless Propag. Lett.*, vol. 18, pp. 679–683, 2019.
- [26] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 569–572, Dec. 2014.
- [27] *Propagation Data and Prediction Methods Required for the Design of Earth-Space Telecommunication Systems*, document ITU-R P.618-13, 2017.
- [28] *The Concept of Transmission Loss for Radio Links*, document ITU-R P.341-6, 2016.
- [29] J. S. Yeom, G. Noh, H. Chung, I. Kim, and B. C. Jung, "Performance analysis of satellite and terrestrial spectrum-shared networks with directional antenna," *ETRI J.*, vol. 42, no. 5, pp. 712–720, Nov. 2020.
- [30] L. Hu et al., "Cooperative jamming for physical layer security enhancement in Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 219–228, Feb. 2018.
- [31] P. Mu, X. Hu, B. Wang, and Z. Li, "Secrecy rate maximization with uncoordinated cooperative jamming by single-antenna helpers under secrecy outage probability constraint," *IEEE Commun. Lett.*, vol. 19, no. 12, pp. 2174–2177, Dec. 2015.

- [32] J. Yang, D. Bai, X. Yuan, and D. Li, "Anti-jamming UAV relaying based on joint geographic-electromagnetic domain optimization," *IEEE Access*, vol. 12, pp. 53928–53943, 2024.
- [33] J. Fan, Z. Wang, Y. Xie, and Z. Yang, "A theoretical analysis of deep Q-learning," 2019, *arXiv:1901.00137*.
- [34] T. P. Lillicrap et al., "Continuous control with deep reinforcement learning," 2015, *arXiv:1509.02971*.



**Kakyom Jeon** (Student Member, IEEE) received the B.S. degree in electronic and electrical engineering from Hankyong National University, Anseong, South Korea, in 2021, and the M.S. degree in AI convergence network from Ajou University, Suwon, South Korea, in 2025.

Since 2025, he has been a master's degree Researcher with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea. His current research interests include AI-RAN, O-RAN architecture, O-CU/O-DU, machine

learning for wireless communications, intelligent wireless networks, and 6G wireless communications.

Mr. Jeon is a recipient of the Korea Photonics Technology Institute (KOPTI) Best Paper Award at the KICS Summer Conference in 2023; the Best Paper Award at the KICS Summer Conference in 2024; and the Best Paper Award at the KICS Winter Conference in 2025.



**Young-Seok Lee** (Student Member, IEEE) received the B.S. degree and the combined master's degree and doctoral program coursework in electronics engineering from Chungnam National University, Daejeon, South Korea, in 2020 and 2024, respectively. He is currently pursuing the Ph.D. degree with the Department of Artificial Intelligence Convergence Network, Electronics Engineering Major, Ajou University, Suwon, South Korea.

His research interests include 5G+/6G Mobile & IoT communication systems, 6G MIMO beamforming (CAP-MIMO/holographic MIMO/orbital angular momentum), integrated sensing and communication (ISAC) systems, reconfigurable intelligent surface-based nonorthogonal multiple access, military communications, classical and electromagnetic information theory, statistical signal processing, low-earth orbit satellite networks, and GNSS receiver signal processing.

Mr. Lee received the Best Paper Award at IEIE Symposium in December 2022, the Grand Prize of the KICS Winter Conference Undergraduate Student Paper Contest in February 2024, and the Grand Prize of the KICS Haedong Best Paper Award in January 2024.



**Bang Chul Jung** (Senior Member, IEEE) received the B.S. degree in electronics engineering from Ajou University, Suwon, South Korea, in 2002, and the M.S. and Ph.D. degrees in electrical and computer engineering from Korea Advanced Institute for Science and Technology, Daejeon, South Korea, in 2004 and 2008, respectively.

He was a Post-Doctoral Fellow/Research Professor with the KAIST Institute for Information Technology Convergence, Daejeon, from 2009 to 2010. From 2010 to 2025, he was a Faculty Member with Chungnam National University, Daejeon, and Gyeongsang National University, Jinju, South Korea. He is currently a Professor with the Department of Electrical and Computer Engineering, Ajou University. His research interests include wireless communication systems, Internet of Things communications, statistical signal processing, information theory, interference management, radio resource management, spectrum sharing techniques, and machine learning.

Prof. Jung was a recipient of the Bronze Prize of the Intel Student Paper Contest in 2005, the First Prize of the KAIST Invention Idea Contest in 2008, the Bronze Prize of the Samsung Humantech Paper Contest in 2009, and the Fifth IEEE Communication Society Asia-Pacific Outstanding Young Researcher Award in 2011. He was selected as a Winner of the Haedong Young Scholar Award in 2015, which is sponsored by the Haedong Foundation and given by KICS. He has been selected as a winner of the 29th Science and Technology Best Paper Award in 2019, which is sponsored by the Korean Federation of Science and Technology Societies. He was honored with the Ministerial Commendation from the Ministry of Science and ICT of the Republic of Korea in 2024. He was the TPC Chair of IEEE CCNC 2023 and the General Chair of IEEE CCNC 2025. He was an Associate Editor of *IEEE Vehicular Technology Magazine* from 2020 to 2022, and is now a Senior Editor of *IEEE Vehicular Technology Magazine*. He has been an Associate Editor of the IEEE Open Journal of the Communication Society since 2025.



**Howon Lee** (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical and computer engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2003, 2005, and 2009, respectively.

From 2009 to 2012, he was a Senior Research Staff/Team Leader of the Knowledge Convergence Team at the KAIST Institute for Information Technology Convergence (KI-ITC). From 2012 to 2024, he was with the School of Electronic and Electrical Engineering and the Institute for IT Convergence (IITC) at Hankyong National University (HKNU), Anseong, South Korea. Since 2024, he has been with the Department of Electrical and Computer Engineering, Ajou University, Suwon, South Korea. He has also experienced as a Visiting Scholar with the University of California, San Diego (UCSD), La Jolla, CA, USA, in 2018. His current research interests include B5G/6G wireless communications, ultradense distributed networks, in-network computations for 3-D images, cross-layer radio resource management, reinforcement learning for UAV and satellite networks, unsupervised learning for wireless communication networks, and Internet of Things.

Dr. Lee is the recipient of the Joint Conference on Communications and Information (JCCI) 2006 Best Paper Award and the Bronze Prize at Intel Student Paper Contest in 2006. He is also the recipient of the Telecommunications Technology Association (TTA) Paper Contest Encouragement Award in 2011, the Best Paper Award at the Korean Institute of Communications and Information Sciences (KICS) Summer Conference in 2015, the Best Paper Award at the KICS Fall Conference in 2015, the Honorable Achievement Award from 5G Forum Korea in 2016, the Best Paper Award at the KICS Summer Conference in 2017, the Best Paper Award at the KICS Winter Conference in 2018, the Best Paper Award at the KICS Summer Conference in 2018, the Best Paper Award at the KICS Winter Conference in 2020, the Haedong Grand Prize at the KICS Summer Conference in 2022, the Haedong Grand Prize at the KICS Winter Conference in 2023, the Haedong Grand Prize at the KICS Summer Conference in 2024, and the Haedong Grand Prize at the KICS Winter Conference in 2025. He received the Minister's Commendation by the Minister of Science and ICT in 2017. He is the Technical Program Committee (TPC) Chair of IEEE CCNC 2026. Since 2024, he has served as an Associate Editor for IEEE INTERNET OF THINGS JOURNAL.