RSMA-enabled Multi-UAV Secure Communication via MARL with Multi-Task Attention DRNN

Lijie Zheng, Ji He *Member, IEEE*, Xinghui Zhu *Member, IEEE,* Yuanyu Zhang *Member, IEEE*, Yulong Shen *Member, IEEE* and Tarik Taleb *Senior Member, IEEE*

***Abstract*—This paper investigates secure communication in multi-UAV networks, where each UAV employs rate-splitting multiple access (RSMA) to simultaneously deliver downlink data services to multiple ground terminals (GTs) under eavesdropping threats. To enhance network security, we propose a two-stage collaborative RSMA transmission scheme. Based on this scheme, we study the optimization of multi-UAV cooperative trajectory, time-step sharing and jamming power (MUCTSJ) to maximize the network's secrecy rate. Additionally, to ensure fairness in throughput allocation among GTs, we incorporate two typical UAV service principles—Channel Quality First (CQF) and Fair Service First (FSF)—into the optimization objectives. Given the non-convex and NP-hard nature of this optimization problem, we reformulate it as a Markov Decision Process (MDP) and introduce a multi-agent reinforcement learning (MARL) framework based on the Centralized Training and Decentralized Execution (CTDE) paradigm. To address the dynamic topological changes induced by UAV mobility and time-varying channel states, as well as the gradient interference among multiple learning tasks, we design a Multi-Task Attention Deep Recurrent Network (MTA-DRNN). This architecture effectively captures the distinct observed attributes of each UAV while enhancing the coordination between diverse actions, thereby improving the adaptability of the agent and the stability of training. Simulation results demonstrate the superiority of the proposed solution enhances the security of multi-UAV networks over other baseline schemes. Furthermore, deployment on corresponding hardware platforms confirms the solution's effectiveness and robustness in practical applications.**

***Index Terms*—Multi-UAV networks, RSMA, physical layer security, multi-agent reinforcement learning, MTA-DRNN.**

I. Introduction

UNMANNED aerial vehicles (UAVs) have gradually become a crucial component of modern communication infrastructure, particularly in B5G and 6G communication systems, owing to their flexibility, scalability [1]–[3]. For example, compared to ground-based channels, UAVs operating as aerial base stations (BSs) can offer enhanced communication services to ground terminals via line-of-sight (LoS) links. This makes them particularly useful in scenarios such as emergency disaster relief communications, providing network coverage in remote areas, and supporting smart city management [4], [5]. To improve spectrum efficiency in future networks, the deployment of multiple access technologies in UAV-assisted wireless networks has emerged as a key research direction [6], [7].

The broadcast nature of wireless mediums over LoS links makes UAV networks more vulnerable to interception by eavesdroppers and malicious jammers than ground-based channels, posing significant security and privacy risks [8]. Physical layer security (PLS) technologies, leveraging the random characteristics of wireless channels to ensure the security of wireless communication from an information-theoretic perspective, have been regarded as a promising complement to cryptographic methods [9]–[11]. Moreover, the inherent characteristics of UAVs, such as high maneuverability and time-varying network topology, provide natural random channel conditions for PLS, thus motivating increasing research on PLS in UAV networks [12], [13]. Zhao et al. [14] proposed an innovative cooperative secure transmission and computation strategy to counter mobile collusion eavesdroppers (Eves) in UAV-assisted mobile edge computing networks. Li et al. [15] designed a dual-UAV-assisted non-orthogonal multiple access (NOMA) communication architecture to optimize the PLS performance of the system while considering the outage probability constraint. Li et al. [16] proposed a collaborative beamforming scheme based on UAV virtual antenna arrays to achieve secure UAV communication with different base stations in the presence of known and unknown ground Eves. Guo et al. [17] studied the secrecy energy efficiency optimization problem of UAVs equipped with simultaneously transmissive and reflective reconfigurable intelligent surfaces in the NOMA uplink system.

In recent years, RSMA has been recognized as a promising physical layer transmission paradigm for non-orthogonal transmission, interference management and multiple access strategies in 6G [18]. Xiao et al. [19] studied the problem of joint 3D deployment and beamforming of UAV base station based on RSMA with the assistance of geographic information. With the growing adoption of RSMA in UAV communications for next-generation networks [20]–[22], the secrecy performance of

This work was supported in part by the National Key Program of China (No. 2023YFB3107500), in part by the Major Research Plan of the National Natural Science Foundation of China (No. 92267204 and U24A20238), in part by the Natural Science Foundation of China (No. 62302366, 92467201, 62502363, 62202354), in part by the China Postdoctoral Science Foundation (No. 2023M744309), in part by the Key R&D Program of Shandong Province of China (2023CXPT056, 2025CXPT089), in part by the Natural Science Basic Research Program of Shaanxi Province (No. 2025JC-YBQN-869), in part by the QinChuangyuan Innovation and Entrepreneurship Talent Project of Shaanxi (No. QCYRCXM-2022-144), in part by the Key Research and Development Program of Shaanxi Province (No. 2024GX-YBXM-073), in part by the Fundamental Research Funds for the Central Universities (No. QTZX25080, YJSJ25012), in part by the Shandong Province Key Laboratory of Autonomous and Reliable Computing Technology and Equipment (KT25500400-lab). (Corresponding author: Ji He)

L. Zheng, J. He, X. Zhu, Y. Zhang and Y. Shen are with the School of Computer Science and Technology, Xidian University, Xi'an, 710071 China (e-mail: lijzheng@stu.xidian.edu.cn; jihe@xidian.edu.cn; xhzhu@xidian.edu.cn;yyuzhang@xidian.edu.cn; ylshen@mail.xidian.edu.cn).

T. Taleb is with Ruhr University Bochum, Bochum, 44801 Germany (email: tarik.taleb@rub.de)

RSMA-enabled UAV systems has attracted significant research attention. Based on the PLS principles, Fu et al. [23] found that the common stream of RSMA can be used as an effective means to enhance the transmission rate of legitimate users, and can also be used as artificial noise (AN) to confuse Eves. It is proved that RSMA achieves an explicit secrecy rate gain over NOMA in multiple-input single-output broadcast channel. Taking advantage of this feature, Bastami et al. proposed a collaborative RSMA scheme to enhance secrecy rates in UAV-assisted cellular networks [24], and later extended their study to jointly optimize security and UAV energy efficiency by balancing secrecy performance against power consumption [25]. However, the effectiveness of the aforementioned PLS-based RSMA schemes is highly dependent on the accuracy of the eavesdropper's channel state information (CSI). To address this limitation, Bastami et al. [26] examined the robustness of RSMA-enabled UAV downlink networks under imperfect eavesdropper CSI. Building on this, security concerns under imperfect CSI have been extended to various networking scenarios [27]–[29].

The above research undoubtedly paves the way for applying RSMA in UAV networks and initially explores its security performance, laying a crucial foundation for the envisioned B5G and 6G wireless ecosystem. As shown in Table I, existing studies on secure RSMA-aided UAV networks mainly focuses on single-UAV scenarios (e.g., [19], [20], [24]), while considering multi-UAV scenarios does not consider security issues (e.g., [21], [22]). Compared to single-UAV scenarios, deploying multiple UAVs can improve communication efficiency, but it also increases the risk of eavesdropping under traditional rate-splitting schemes and induces exponential growth in computational complexity and solution space. Furthermore, the dynamic nature of UAVs leads to time-varying and high-dimensional GT information, where existing methods based on statistical approximations often fail to capture system dynamics accurately, resulting in reduced fairness among GTs in complex environments. Moreover, conventional single-output machine learning architectures force multiple tasks to share representations, causing gradient interference due to conflicting objectives. For instance, trajectory planning aims to maximize coverage and throughput, whereas jamming power allocation may reduce legitimate user throughput to enhance secrecy, thereby destabilizing training. Given the limited local observations available to each UAV, it is crucial to develop a robust distributed coordination framework that enables joint decision-making and strengthens overall network security.

While the security of RSMA-enabled multi-UAV networks encompasses a wide range of application scenarios, problem formulations, and optimization challenges, this work aims to make an attempt to address this issue. As a step forward in this direction, we tackle the practical security challenges in RSMA-enabled multi-UAV networks by proposing a novel two-stage collaborative RSMA (TS-CRSMA) scheme and an adaptive MARL framework to jointly optimize for both network secrecy and user fairness in dynamic environments. Specifically, the main contributions of our work are summarized as follows:

- Proposing a two-phase collaborative RSMA scheme for secure downlink multi-UAV networks with imperfect CSI: In the first phase, UAVs simultaneously transmit common and private messages, with the common message serving dual purposes: delivering intended data and acting as interference to hinder eavesdropping on private messages. In the second phase, the optimal relay employs a hybrid protocol to forward information, ensuring correct decoding by GTs. Additionally, idle UAVs emit jamming signals to further disrupt Eve's ability to intercept relayed information.
- Proposing a MARL framework for secure collaborative communication in RSMA-enabled multi-UAV networks: This paper formulates an optimization problem for multi-UAV cooperative trajectories, time-step sharing and jamming power based on the collaborative RSMA scheme. The primary objective is to maximize the network's secrecy rate while adhering to constraints on UAV position, speed, and power. To ensure fairness in throughput allocation among GTs, two typical service principles for UAVs (i.e., the CQF principle and the FSF principle) are incorporated into the optimization process. To address the non-convex and NP-Hard nature of the problem, we model it as a MDP and propose a MARL framework based on a multi-task attention decision network (MTA-MARL) to jointly optimize UAV trajectories, time-step sharing, and jamming power, which significantly enhances the network's secrecy and fairness performance.
- Introducing an adaptable neural network architecture to handle multi-scale inputs and multi-class action outputs: To address the dynamic changes in observation input dimensions caused by the UAV's dynamic topology during flight, a multi-head attention mechanism (MHA) is integrated into the input layer. Additionally, the output layer employs a multi-task output (MTO) structure, allowing each output head to focus on specific tasks and reducing gradient interference. Furthermore, the loss function is optimized to enhance task synergy, significantly improving learning efficiency and decision-making performance.
- We conducted extensive simulations across various scenarios with different numbers of UAVs and GTs, along with baseline comparisons, to evaluate the proposed solution's effectiveness. To assess its performance in realistic UAV deployment, we implemented it on Nvidia Jetson hardware. The results demonstrate the solution's robustness and applicability for real-world UAV operations.

The remainder of this paper is organized as follows. Section II defines the network model, elaborates on the proposed transmission scheme, and formulates the secrecy rate maximization problem. Section III introduces the principles and details of the proposed algorithm framework. Section IV presents simulation and numerical results and discusses deployment adaptability. Lastly, section V concludes this work. The main notations of this paper are presented in Table II.

II. Network Model and Preliminaries

A. Network Model

We consider a RSMA-enabled multi-UAV network as shown in Fig. 1, where UAVs act as aerial BSs positioned at a fixed

TABLE I: THE DIFFERENCE BETWEEN OUR WORK AND THE EXISTING WORKS.

Ref	**Scenario**	**Transmission scheme**	**Optimization parameters**	**Objective**	**Algorithm**
[4]	Multiple UAVs	-	UAV deployment	Coverage **(without security)**	Weighted voronoi distributed optimization
[6]	Multiple UAVs	-	Data offloading strategy	User Satisfaction **(without security)**	Game theoretic approach
[5]	Multiple UAVs	FDMA	UAV trajectory	Throughput and Fairness **(without security)**	MARL with GCVis & Comm
[14]	Single UAV	OFDMA	UAV trajectory, Jamming beamformers, Transmit power, Data offloading strategy	Secrecy Rate	CSTC Algorithm
[15]	Dual-UAV	NOMA	Communication resource, UAV trajectory, Artificial noise	Secrecy Energy Efficiency	SCA
[16]	Multiple UAVs	-	UAV positions, Excitation current weights, Communication order	Secrecy Rate, Sidelobe level and Energy consumption	IMODACH, P-IMODACH
[17]	Single UAV	NOMA	Power control, Reflection coefficients, UAV deployment	Secrecy Energy Efficiency	SCA and DDQN
[19]	Single UAV	RSMA	UAV deployment, Beamforming	Minimum Sum Rate **(without security)**	SDP and SCA
[20]	Single UAV	RSMA	UAV deployment, RSMA precoding, Rate splitting	Sum Rate **(without security)**	SCA and WMMSE
[21]	Multiple UAVs	RSMA	UAV network density, Power allocation	Eenergy Efficiency **(without security)**	Particle swarm optimization
[22]	Multiple UAVs	RSMA	Association, Beamforming	Sum Rate **(without security)**	Enhanced-MAPPO
[23]	-	RSMA	Beamforming	Secrecy Rate	SCA
[24]	Single UAV	Collaborative RSMA	Power allocation, Time-step sharing, Weighting factor	Worst-case Secrecy Rate	SPCA
[25]	Single UAV	Collaborative RSMA	Association, Precoding, Time-step sharing, Power allocation	Secrecy Energy Efficiency	SPCA
[26]	Single UAV	RSMA	Precoding	Worst-case Secrecy Rate	SPCA
[27]	Single UAV	RSMA	-	Close form of Secrecy Outage Probability	-
[28]	Single UAV	RSMA	-	Close form of Secrecy Rate	-
[29]	Single UAV	RSMA	Power allocation	Secrecy Rate	SPCA
Our work	**Multiple UAVs**	**Two-Phase Collaborative RSMA**	**Jamming power allocation**, **Time-step sharing**, **UAV trajectory**	**Secrecy Rate and Fairness**	**MTA-MARL**

'-' denotes this factor has not been considered.

altitude, H_{uav}, to deliver data services to GTs located in a $D \times D$ region while being subject to wiretapping attacks from Eves. The sets of UAVs, GTs and Eves are denoted by $\mathcal{K} = \{1, 2, \cdots, K\}$, $\mathcal{I} = \{1, 2, \cdots, I\}$, and $\mathcal{E} = \{1, 2, \cdots, E\}$, respectively. The three-dimensional coordinates of the i-th GT and the e-th Eve are represented $u_i = [x_i, y_i, 0]^T$, $\forall$ $i \in \mathcal{I}$ and $u_e = [x_e, y_e, 0]^T$, $\forall$ $e \in \mathcal{E}$, respectively. To model the UAVs' flight dynamics, the flight duration is segmented into T equal-length time steps of length Δt, denoted as $\mathcal{T} = \{1, 2, \cdots, T\}$. At any given time step t, the position of UAV k is denoted by $u_k(t) = [x_k(t), y_k(t), H_{\text{uav}}]^T$, where $x_k(t) \in [0, D]$ and $y_k(t) \in [0, D]$, $\forall$ $k \in \mathcal{K}$, $t \in \mathcal{T}$. The instantaneous velocity and flight direction of UAV k at time step t are $v_k(t) \in \mathbb{R}^+$ and $\omega_k(t) \in [0, 2\pi)$, respectively. To reflect real-world constraints, the UAV's velocity is bounded by a maximum velocity $v_{\max}$, i.e., $v_k(t) \leq v_{\max}$. Accordingly, the positions at the next time step $(t + 1)$ are updated according to $x_k(t + 1) = x_k(t) + v_k(t) \cos(\omega_k(t))\Delta t$ and $y_k(t + 1) = y_k(t) + v_k(t) \sin(\omega_k(t))\Delta t$.

B. Channel Model

In this work, we introduce the Air-to-Ground (A2G) and Ground-to-Ground (G2G) channel models to capture the communication dynamics within the UAV network. The A2G channel between UAV k to GT/Eve x can be expressed as $g^{\text{A2G}}_{k,x}(t) = \sqrt{d^{-\alpha}_{k,x}(t)} h^{\text{A2G}}_{k,x}(t), x \in \{\mathcal{I}, \mathcal{E}\}$, where $d_{k,x}(t) = \|u_k(t) - u_x\|_2$ is the Euclidean distance between UAV k and GT/Eve x, α represents the path exponent characterizing large-scale fading and $h^{\text{A2G}}_{k,x}(t)$ captures the small-scale fading component of the channel. Similarly, the forwarding link between GT i and GT/Eve x is modeled as a typical G2G

TABLE II: SUMMARY OF NOTATIONS.

Notation	**Description**
H_{uav}	Elevation of UAVs
D	Length of the flight area
$\mathcal{K}$, $\mathcal{I}$, $\mathcal{E}$	Set of UAVs, GTs and Eves
u_k, u_i, u_e	Position of UAVs, GTs, and Eves
$\mathcal{T}$	Set of time steps
$v_k(t)$, $w_k(t)$	Instant flight velocity and direction of UAV k
v_{max}	Maximum velocity of UAVs
$g_{k,x}^{\text{A2G}}(t)$	A2G channel between UAV k and GT/Eve x
$g_{i,x}^{\text{G2G}}(t)$	G2G channel between GT i and GT/Eve x
$d_{k,x}(t)$	Distance between UAV k and GT/Eve x
$d_{i,x}(t)$	Distance between GT i and GT/Eve x
$h_{k,x}^{\text{A2G}}(t)$	Small-scale fading component of the A2G channel between UAV k and GT/Eve x
$h_{i,x}^{\text{G2G}}(t)$	Small-scale fading component of the G2G channel between GT i and GT/Eve x
$\mathcal{R}^{\text{GT}}$, $\mathcal{R}^{\text{UAV}}$	Coverage and communication range of UAVs
$\mathbf{C}^{\text{GT}}(t)$, $\mathbf{C}^{\text{UAV}}(t)$	Sensing relations between UAVs and GTs/UAVs
$\mathbf{S}^{\text{GT}}(t)$	Scheduling relations between UAVs and GTs
$\mathbf{C}^{\text{Eve}}(t)$	Eavesdropping relations between UAVs and Eves
$\mathcal{W}_{k,i}^{c}$, $\mathcal{W}_{k,i}^{p}$	Common and private messages
$\mathbf{x}_k$	Signal transmitted by UAV k
$\mathbf{s}_k^c$, $\mathbf{s}_i^p$	Common and private signals
$R_k^c(t)$, $R_i^p(t)$	Achievable rate of common and private messages
$R_k^{\text{c,sr}}(t)$, $R_i^{\text{p,sr}}(t)$	Secrecy rate of common and private messages
$F(\hat{\mathbf{d}}_t)$	System fairness index

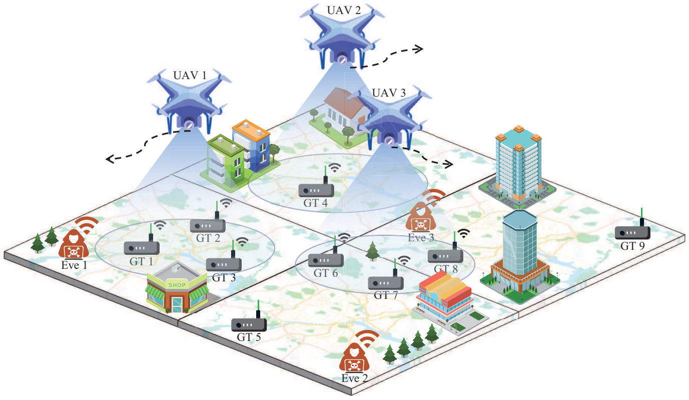


Fig. 1: Illustration of RSMA-enabled multi-UAV networks

model. The complex-value channel coefficient is determined by $g_{i,x}^{\text{G2G}}(t) = \sqrt{d_{i,x}^{-\alpha} h_{i,x}^{\text{G2G}}(t)}$, where $h_{i,x}^{\text{G2G}}(t)$ represents the small-scale fading component of the G2G channel.

In the considered multi-UAV network, unintended Eves typically do not frequently send pilot signals to update the CSI at the transmitter, leading to inaccurate CSI [30]. Moreover, since the UAVs can acquire the positions of Eves through synthetic aperture radar, they are capable of estimating large-scale fading components, such as path loss [15]. And path loss components vary slowly compared to small-scale fading [24]. Consequently, in this work, we consider imperfect CSI, particularly focusing on small-scale fading in both A2G and G2G channels. The small-scaling fading of the A2G channel between UAV k and Eve e can be modeled as $h_{k,e}^{\text{A2G}}(t) = \hat{h}_{k,e}^{\text{A2G}}(t) + \Delta h_{k,e}^{\text{A2G}}(t)$, where $\hat{h}_{k,e}^{\text{A2G}}(t) \sim \mathcal{CN}(0, 1 - \sigma_{err1}^2)$ denotes the estimated fading, and channel estimated error $\Delta h_{k,e}^{\text{A2G}}(t) \in \mathbb{C}$ follows a complex Gaussian distribution that has zero-mean and variance σ_{err1}^2. Similarly, the small-scale fading of the G2G channel between GT i and Eve e can be represented as $h_{i,e}^{\text{G2G}}(t) = \hat{h}_{i,e}^{\text{G2G}}(t) + \Delta h_{i,e}^{\text{G2G}}(t)$, where $\hat{h}_{i,e}^{\text{G2G}}(t) \sim \mathcal{CN}(0, 1 - \sigma_{err2}^2)$ is the estimated small-scale fading coefficient and $\Delta h_{i,e}^{\text{G2G}}(t) \sim \mathcal{CN}(0, \sigma_{err2}^2)$ denotes the corresponding channel estimated error.

C. UAV Sensing Model

In practical UAV networks, the maximum distance at which a UAV can detect a request from GTs is inherently constrained by signal attenuation. Similarly, each UAV can perceive other UAVs within a specific communication range. To incorporate these practical considerations, we introduce two key parameters: the coverage range $\mathcal{R}^{\text{GT}}$ for detecting nearby GTs and the communication range $\mathcal{R}^{\text{UAV}}$ for interacting with other UAVs. Correspondingly, GT i can be detected by UAV k at time step t and only if $\|u_k(t) - u_i\|_2 \leq \mathcal{R}^{\text{GT}}$. Likewise, UAV k and UAV l can communicate with each other if $\|u_k(t) - u_l(t)\|_2 \leq \mathcal{R}^{\text{UAV}}$. As a result, each UAV can obtain partial observations of the entire system. These spatial-temporal connectivity constraints are mathematically characterized using two binary adjacency matrices: $\mathbf{C}^{\text{GT}}(t) \in \{0,1\}^{K \times I}$ and $\mathbf{C}^{\text{UAV}}(t) \in \{0,1\}^{K \times K}$. Specifically, the (k,i)-th entry of $\mathbf{C}^{\text{GT}}(t)$, denoted by $\mathbf{C}_{k,i}^{\text{GT}}(t) = 1$ if GT i is detected by UAV k at time step t and 0 otherwise; Similarly, $\mathbf{C}_{k,l}^{\text{UAV}}(t) = 1$ if UAV l is in the communication range of UAV k at time step t and 0 if not.

Due to the limited payload and computational resources, the service capacity of a UAV is restricted. Let $\mathcal{C}$ represent the maximum number of GTs that a UAV can serve simultaneously. When the number of GTs within the coverage area exceeds this capacity, the UAV prioritizes serving users with better channel quality. Specifically, the user scheduling is governed by a deterministic, channel-aware heuristic rule that is executed at each time step t. The process is performed iteratively for each UAV $k \in \mathcal{K}$. For a given UAV, it first identifies the set of all detectable GTs within its coverage range. These GTs are then ranked in descending order based on their A2G channel gains, which are quantified by the by the squared magnitude of the channel coefficient $|g_{k,i}^{\text{A2G}}|^2$. Following this ranking, the UAV schedules for service the GTs with the best channel conditions, up to its maximum capacity $\mathcal{C}$. To model this scheduling behavior, we introduce a binary scheduling matrix $\mathbf{S}^{\text{GT}}(t) \in \{0,1\}^{K \times I}$, where the (k,i)-th entry $\mathbf{S}_{k,i}^{\text{GT}}(t) = 1$ if GT i is served by UAV k at time step t and 0 otherwise.

Additionally, potential eavesdropping threats are modeled using an eavesdropping incidence matrix $\mathbf{C}^{\text{Eve}}(t) \in \{0,1\}^{K \times E}$. Owning to the attenuation of the signal level, Eve can intercept a signal only if it lies within the coverage range of a UAV; otherwise, the signal is negligible. Thus, we define the element $\mathbf{C}^{\text{Eve}}(t)$ is 1 if Eve e lies within the *coverage range* $\mathcal{R}^{\text{GT}}$ of UAV k at time step t and 0 otherwise.

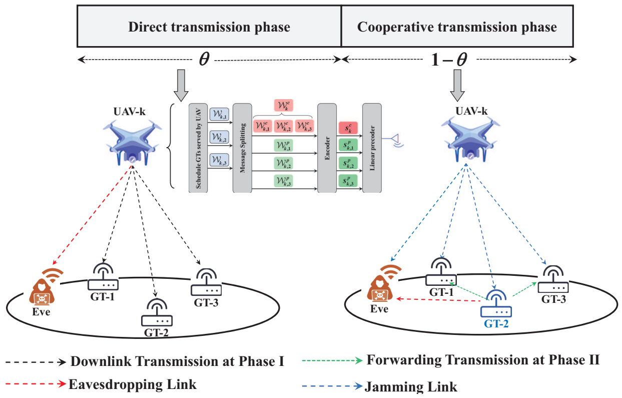


Fig. 2: Two-phase collaborative RSMA transmission scheme

D. Collaborative RSMA Transmission Scheme

To enhance the security of the considered network, we propose a two-phase collaborative RSMA transmission scheme, as illustrated in Fig. 2. In the first phase, the UAV employs the RSMA technique to encode both common and private messages into separate common and private streams, respectively, which are then transmitted to the GTs. The RSMA technique not only facilitates effective interference management but also allows the common message to serve as a jamming signal against potential Eves. In the second phase, the GT with the highest channel quality will forward the common messages using Decode-and-Forward (DF) scheme and the private messages using Amplify-and-Forward (AF) scheme. These transmissions are directed to other GTs under the same UAV schedule, thereby assisting them in further decoding the received message. To further enhance security against Eves, the UAV emits AN to disrupt Eves, which aims to disrupt the reception capabilities of Eves. The following contents provide a detailed explanation of each phase in the proposed two-phase transmission scheme. All derivations in the section are performed within a single time step, with the time symbol t omitted for simplicity.

1) Phase I: All the messages $\mathcal{W}_{k,i}$ transmitted by UAV k are split into two parts: a common part $\mathcal{W}_{k,i}^c$ and a private part $\mathcal{W}_{k,i}^p$, where $i \in \mathcal{I}_k$. Here, $\mathcal{I}_k$ denotes the set of GTs assigned to UAV k. And let $N_{\mathcal{I}_k}$ denote the size of set $\mathcal{I}_k$. According to the RSMA principle, the common messages $\mathcal{W}_{k,i}^c$ are encoded into a common stream $\mathbf{s}_k^c$ using a codebook shared by all GTs [31], while the private messages $\mathcal{W}_{k,i}^p$ for GT i are encoded into the corresponding private stream $\mathbf{s}_i^p$. Thus, the signal $\mathbf{x}_k$ transmitted by UAV k is given by

$$\mathbf{x}_k = \sqrt{p_k^c}\mathbf{s}_k^c + \sum_{i=1}^{N_{\mathcal{I}_k}} \sqrt{\alpha_{k,i}^p p_k^p}\mathbf{s}_i^p, \tag{1}$$

where p_k^c and p_k^p are the power allocated by UAV k to the common stream and all private stream, respectively. The term $\alpha_{k,i}^p$ is the power allocation coefficient for the private stream of GT i. In this work, we adopt an equal power allocation strategy to simplify system model and set this coefficient to $\alpha_{k,i}^p = 1/N_{\mathcal{I}_k}$. Upon receiving the signal, each GT initially decodes the common stream $\mathbf{s}_k^c$ to retrieve the associated common message $\mathcal{W}_{k,i}^c$. Subsequently, each GT decodes its private stream $\mathbf{s}_i^p$ using a codebook distinct from the one employed at the transmitter, as outlined in [32]. Additionally, A2G inter-system interference occurs when a GT falls within the coverage range of multiple UAVs. The received signal at node $\tilde{i}$ (here, $\tilde{i} \in \{i, e\}$ denotes GT-i or Eve-e) from UAV-k is determined by

$$y_{k,\tilde{i}}^{(1)} = g_{k,\tilde{i}}^{\text{A2G}}\mathbf{x}_k + \sum\nolimits_{l \neq k} \mathbf{C}_{l,\tilde{i}}^{\Delta} g_{l,\tilde{i}}^{\text{A2G}}\mathbf{x}_l + n_{\tilde{i}}, \tag{2}$$

where $\Delta = \text{GT}$ when $\tilde{i} = i$, otherwise $\Delta = \text{Eve}$. The second term on the RHS of (2) is A2G inter-system interference, and $n_{\tilde{i}} \sim \mathcal{CN}(0, \sigma_{\tilde{i}}^2)$ represents the AWGN at node $\tilde{i}$.

During Phase I, each GT decodes the common stream $\mathbf{s}_k^c$ while treating the private streams as interference. Consequently, the Signal-to-Interference-plus-Noise Ratio (SINR) of GT-i when decoding the common stream $\mathbf{s}_k^c$ is given by

$$\gamma_i^{\text{c}(1)} = \frac{\mathbf{S}_{k,i}^{\text{GT}} \left| g_{k,i}^{\text{A2G}} \right|^2 p_k^c}{\mathbf{S}_{k,i}^{\text{GT}} \left| g_{k,i}^{\text{A2G}} \right|^2 p_k^p + I_i^{\text{in}(1)} + \sigma_i^2}, \tag{3}$$

where $I_i^{\text{in}(1)} = \sum_{l \neq k} \mathbf{C}_{l,i}^{\text{GT}} \left| g_{l,i}^{\text{A2G}} \right|^2 (p_l^c + p_l^p)$ represents the inter-system interference caused by other UAVs on GT i.

After removing the common part, each GT then decodes its private streams using SIC [33], [34]. Hence, the corresponding SINR at GT i when decoding its private stream $\mathbf{s}_i^p$ is given by

$$\gamma_i^{\text{p}(1)} = \frac{\mathbf{S}_{k,i}^{\text{GT}} \left| g_{k,i}^{\text{A2G}} \right|^2 \alpha_{k,i}^p p_k^p}{\mathbf{S}_{k,i}^{\text{GT}} \left| g_{k,i}^{\text{A2G}} \right|^2 \sum\limits_{j \neq i}^{N_{\mathcal{I}_k}} \alpha_{k,j}^p p_k^p + I_i^{\text{in}(1)} + \sigma_i^2}. \tag{4}$$

Similarly, the SINR for Eve e when attempting to decode the common stream $\mathbf{s}_k^c$ is given by

$$\gamma_{e,k}^{\text{c}(1)} = \frac{\mathbf{C}_{k,e}^{\text{Eve}} \left| g_{k,e}^{\text{A2G}} \right|^2 p_k^c}{\mathbf{C}_{k,e}^{\text{Eve}} \left| g_{k,e}^{\text{A2G}} \right|^2 p_k^p + I_e^{\text{in}(1)} + \sigma_e^2}, \tag{5}$$

where $I_e^{\text{in}(1)} = \sum_{l \neq k} \mathbf{C}_{l,e}^{\text{Eve}} \left| g_{l,e}^{\text{A2G}} \right|^2 (p_l^c + p_l^p)$ represents the inter-system interference when Eve lies within the coverage of multiple UAVs.

To ensure secure communication, the rate of the common stream from UAV to GTs is designed to be higher than the rate achievable by the Eves. Only in this way can the common message serve a dual purpose: conveying the intended data to GT while simultaneously serving as interference for Eve. Therefore, the SINR for Eve e when attempting to decode the private stream $\mathbf{s}_i^p$ of GT i is

$$\gamma_{e,i}^{\text{p}(1)} = \frac{\mathbf{C}_{k,e}^{\text{Eve}} \left| g_{k,e}^{\text{A2G}} \right|^2 \alpha_{k,i}^p p_k^p}{\mathbf{C}_{k,e}^{\text{Eve}} \left| g_{k,e}^{\text{A2G}} \right|^2 (p_k^c + \sum\limits_{j \neq i}^{N_{\mathcal{I}_k}} \alpha_{k,j}^p p_k^p) + I_e^{\text{in}(1)} + \sigma_e^2}. \tag{6}$$

2) Phase II: Note that the GT with the best channel quality is selected to relay the messages to the other GTs scheduled by the same UAV. Specifically, the relay will use the DF scheme with power $p^f_{j_k}$ to forward common stream and the AF scheme with the amplification factor β to forward private streams. Conveniently, let GT j_k denote the relay selected from all GTs served by UAV k. Therefore, the signal forwarded by the relay can be expressed as

$$\begin{aligned}\mathbf{x}_{j_k} &= \sqrt{p^f_{j_k}}\mathbf{s}^c_k \\ &+ \beta(g^{\mathrm{A2G}}_{k,j_k}\sum_{i=1}^{N_{\mathcal{I}_k}}\sqrt{\alpha^p_{k,i}p^p_k}\mathbf{s}^p_i + \sum_{l\neq k}^{\mathcal{K}}\mathbf{C}^{\mathrm{GT}}_{l,j_k}g^{\mathrm{A2G}}_{l,j_k}\mathbf{x}_l + n_{j_k})\end{aligned} \tag{7}$$

where the first, second and third term on the RHS of the equation denote common signal, private signals and interference, respectively. In this phase, idle UAVs also emit jamming signals to disrupt Eve's eavesdropping, which will also cause interference to GTs within the coverage range. Additionally, when GTs are within the effective communication range of other relay GTs, they are also subject to G2G inter-system interference. Consequently, the signal received at node $\tilde{i}$ ($\tilde{i} \in \{i_{i\in\mathcal{I}_k}, e_{e\in\mathcal{E}_k}\}$) is given as

$$\begin{aligned}y^{(2)}_{j_k,\tilde{i}} &= g^{\mathrm{G2G}}_{j_k,\tilde{i}}\mathbf{x}_{j_k} + \sum\nolimits_{l\neq k}\mathbf{C}^{\Delta}_{l,\tilde{i}}g^{\mathrm{G2G}}_{j_l,\tilde{i}}\mathbf{x}_{j_l} \\ &+ \sum\nolimits_{k=1}^{K}\mathbf{C}^{\Delta}_{k,\tilde{i}}g^{\mathrm{A2G}}_{k,\tilde{i}}\sqrt{p^J_k}\mathbf{s}^J_k + n_{\tilde{i}},\end{aligned} \tag{8}$$

where $\mathcal{E}_k$ represents the set of Eves eavesdropping on UAV k. The second and third terms on the RHS of the equation represent G2G inter-system interference and UAVs emit jamming emitted by UAVs, respectively.

Thus, the SINR at GT i for decoding the common stream $\mathbf{s}^c_k$ in Phase II is expressed as

$$\gamma^{\mathrm{c}(2)}_i = \frac{\left|g^{\mathrm{G2G}}_{j_k,i}\right|^2 p^f_{j_k}}{I^{\mathrm{in}(2)}_i + I^{\mathrm{p}(2)}_i + I^J_i + I^{\mathrm{in}(1)}_{j_k} + \beta^2\sigma^2_{j_k} + \sigma^2_i}, \tag{9}$$

where $I^{\mathrm{in}(2)}_i = \sum\limits_{l\neq k}^{\mathcal{K}}\mathbf{C}^{\mathrm{GT}}_{l,i}\left|g^{\mathrm{A2G}}_{j_l,i}\right|^2\left[p^f_{j_l} + \beta^2(|g^{\mathrm{A2G}}_{l,j_l}|^2p^p_l + I^{\mathrm{in}(1)}_{j_l} + \sigma^2_{j_l})\right]$ is G2G inter-system interference, $I^{\mathrm{p}(2)}_i = \beta^2|g^{\mathrm{G2G}}_{j_k,i}|^2|g^{\mathrm{A2G}}_{k,j_k}|^2p^p_k$ is the interference of private messages on GT when decoding common message, and $I^J_i = \sum_{k=1}^{K}\mathbf{C}^{\mathrm{GT}}_{k,i}|g^{\mathrm{A2G}}_{k,i}|^2p^J_k$ denotes the jamming signal emitted by the UAV. Upon successfully decoding the common message, the SINR for GT's decoding of private message $\mathbf{s}^p_i$ can be expressed as

$$\gamma^{\mathrm{p}(2)}_i = \frac{\beta^2\left|g^{\mathrm{G2G}}_{j_k,i}\right|^2\left|g^{\mathrm{A2G}}_{k,j_k}\right|^2\alpha^p_{k,i}p^p_k}{I^{\mathrm{in}(2)}_i + I^{\mathrm{op}(2)}_i + I^J_i + I^{\mathrm{in}(1)}_{j_k} + \beta^2\sigma^2_{j_k} + \sigma^2_i}, \tag{10}$$

where $I^{\mathrm{op}(2)}_i = \beta^2|g^{\mathrm{G2G}}_{j_k,i}|^2|g^{\mathrm{A2G}}_{k,j_k}|^2\sum_{j\neq i}^{I}\alpha^p_{k,j}p^p_k$ is the interference of other private messages.

Similarly, the SINR at Eve e at GT i for decoding the common stream $\mathbf{s}^c_k$ is expressed as

$$\gamma^{\mathrm{c}(2)}_{e,k} = \frac{\left|g^{\mathrm{G2G}}_{j_k,e}\right|^2 p^f_{j_k}}{I^{\mathrm{in}(2)}_e + I^{\mathrm{p}(2)}_e + I^J_e + I^{\mathrm{in}(1)}_{j_k} + \beta^2\sigma^2_{j_k} + \sigma^2_e}, \tag{11}$$

where $I^{\mathrm{in}(2)}_e = \sum\limits_{l\neq k}^{\mathcal{K}}\mathbf{C}^{\mathrm{Eve}}_{l,e}\left|g^{\mathrm{A2G}}_{j_l,e}\right|^2\left[p^f_{j_l} + \beta^2(|g^{\mathrm{A2G}}_{l,j_l}|^2p^p_l + I^{\mathrm{in}(1)}_{j_l} + \sigma^2_{j_l})\right]$ is G2G inter-system interference, $I^{\mathrm{p}(2)}_e = \beta^2|g^{\mathrm{G2G}}_{j_k,e}|^2|g^{\mathrm{A2G}}_{k,j_k}|^2p^p_k$ is the interference of private messages when decoding common message, and $I^J_e = \sum_{k=1}^{K}\mathbf{C}^{\mathrm{Eve}}_{k,e}|g^{\mathrm{A2G}}_{k,e}|^2p^J_k$ represents the jamming signal emitted by the UAV. Then, the SINR for Eve e decoding the private stream $\mathbf{s}^p_i$ of GT i is

$$\gamma^{\mathrm{p}(2)}_{e,i} = \frac{\beta^2\left|g^{\mathrm{G2G}}_{j_k,e}\right|^2\left|g^{\mathrm{A2G}}_{k,j_k}\right|^2\alpha^p_{k,i}p^p_k}{I^{\mathrm{in}(2)}_e + I^{\mathrm{op}(2)}_e + I^J_e + I^{\mathrm{in}(1)}_{j^k} + I^{\mathrm{c}(2)}_e + \beta\sigma^2_{j^k} + \sigma^2_e}, \tag{12}$$

where $I^{\mathrm{op}(2)}_e = \beta^2|g^{\mathrm{G2G}}_{j_k,e}|^2|g^{\mathrm{A2G}}_{k,j_k}|^2\sum_{j\neq i}^{I}\alpha^p_{k,j}p^p_k$ and $I^{\mathrm{c}(2)}_e = |g^{\mathrm{G2G}}_{j_k,e}|^2p^f_{j_k}$ represent the interference caused by other private streams and common stream, respectively.

E. Problem Formulation

Note that this work aims to maximize the secrecy rate of the network by ensuring the security of both common and private messages. According to Wyner's encoding scheme [35], the achievable secrecy rate is defined as the difference between the capacities of the legitimate channel (transmitter to intended receiver) and the eavesdropping channel (transmitter to eavesdropper). We introduce the secrecy rate for common and private messages as follows.

In the proposed transmission scheme, relay GT j_k receives common messages only during the first phase, while other GTs can receive the common messages during both phases. Therefore, the transmission rate for the common messages at time step t is determined as $R^c_k(t) = \min\{R^c_{j_k}(t), \{R^c_i(t)\}\}, i \neq j_k, i \in \mathcal{I}_k$, where $R^c_{j_k}(t) = \theta_k(t)\log_2(1 + \gamma^{\mathrm{c}(1)}_{j_k}(t))$ represents the achievable rate of the common messages at relay GT j_k, and $R^c_i(t) = \theta_k(t)\log_2(1 + \gamma^{\mathrm{c}(1)}_i(t)) + (1 - \theta_k(t))\log_2(1 + \gamma^{\mathrm{c}(2)}_i(t))$ is the achievable rate at any other GT served by UAV k. Here, $\theta_k(t)$ (resp. $1 - \theta_k(t)$) denotes the fractions of each time step allocated to phase I (resp. Phase II). For Eve e, the achievable rate of decoding the common stream $\mathbf{s}^c_k$ can be expressed as $R^c_{e,k}(t) = \theta_k(t)\log_2(1 + \gamma^{\mathrm{c}(1)}_{e,k}(t)) + (1 - \theta_k(t))\log_2(1 + \gamma^{\mathrm{c}(2)}_{e,k}(t))$. Therefore, the secrecy rate of the common message in each UAV cell, denoted as $R^{\mathrm{c,sr}}_k(t)$ can be calculated as

$$R^{\mathrm{c,sr}}_k(t) = [R^c_k(t) - \max\{R^c_{e,k}(t)|e \in \mathcal{E}\}]^+, \tag{13}$$

where $[x]^+ = \max\{x, 0\}$ ensures that the secrecy rate is non-negative.

Correspondingly, the transmission rate for the private messages of relay GT j_k and other GT i at time step t is given by $R^p_{j_k}(t) = \theta_k(t)\log_2(1 + \gamma^{\mathrm{p}(1)}_{j_k}(t))$ and $R^p_i(t) = \theta_k(t)\log_2(1 + \gamma^{\mathrm{p}(1)}_i(t)) + (1 - \theta_k(t))\log_2(1 + \gamma^{\mathrm{p}(2)}_i(t))$, respectively. The achievable rate of private messages at Eve can be expressed as $R^p_{e,i}(t) = \theta_k(t)\log_2(1 + \gamma^{\mathrm{p}(1)}_{e,i}(t)) + (1 - \theta_k(t))\log_2(1 + \gamma^{\mathrm{p}(2)}_{e,i}(t))$. Since the private message of each GT is unique, Eve must decode the private message of each GT individually. The secrecy rate of the private messages for GT i, denoted as $R^{\mathrm{p,sr}}_i(t)$, is then determined as

$$R^{\mathrm{p,sr}}_i(t) = [R^p_i(t) - \max\{R^p_{e,i}(t)|e \in \mathcal{E}\}]^+. \tag{14}$$

By analyzing the theoretical results, it becomes evident that to maximize the secrecy rate of the network, UAVs may prioritize serving GTs with high channel quality while potentially neglecting those with poorer channel quality. This strategy is referred to as the CQF principle. However, increasing the number of GTs served by the UAVs can enhance the overall network throughput and promote fairness, leading to the FSF principle.

To ensure fairness among GTs, it is desirable for UAVs to serve as many users as quickly as possible. Additionally, jamming power should be allocated preferentially to GTs with poor channel quality to counteract potential security vulnerabilities. This approach aligns with the concept of Jain's fairness index [36], which is used to evaluate the fairness of resource distribution among users. In this work, we define Jain's fairness index as

$$F(\hat{\mathbf{d}}_t) = \frac{(\sum_{i=1}^{I} \hat{d}_i(t))^2}{I \sum_{i=1}^{I} \hat{d}_i(t)^2} \tag{15}$$

where $\hat{\mathbf{d}}_t = [\hat{d}_1(t), ..., \hat{d}_I(t)]^T$ is a vector representing the average data rate for each GT up to time step t, and each element $\hat{d}_i(t)$ represents the average rate of the messages at GT i, defined as $\hat{d}_i(t) = \frac{1}{t} \sum_{t'=0}^{t} (R_i^c(t^{'}) + R_i^p(t^{'}))$. According to (15), the value of $F(\hat{\mathbf{d}}_t)$ ranges from $1/I$ (indicating extreme unfairness) to 1 (indicating perfect fairness). A higher value of $F(\hat{\mathbf{d}}_t)$ indicates more similar rates among GTs, thus representing better fairness in the network. To accommodate both the CQF and FSF principles, we define a piecewise function as

$$\mathbf{F}(\hat{\mathbf{d}}_t) = \begin{cases} 1 & \text{if CQF} \\ F(\hat{\mathbf{d}}_t) & \text{if FSF} \end{cases} \tag{16}$$

Performance evaluations reveal that UAV trajectory, time-step sharing, and jamming power planning are critical factors impacting the network's secrecy rate. Moreover, these elements—UAV trajectories, time-step sharing, and jamming power—are interdependent, meaning changes in one can influence the others. To address this challenge, we investigate the following fundamental problem: *Under the proposed transmission scheme, how should the trajectories, time-step sharing and jamming power of multiple UAVs be dynamically adjusted to maximize the overall network secrecy rate?* Leveraging the results discussed above, the problem of maximizing the secrecy rate can be mathematically formulated as the following optimization problem

$$\begin{align}
\max_{\mathbf{u}_t, \boldsymbol{\theta}_t, \mathbf{p}_t^J} \frac{1}{T} \sum_{t=0}^{T} \mathbf{F}(\hat{\mathbf{d}}_t) \left[\sum_{k=1}^{K} R_k^{\text{c,sr}}(t) + \sum_{i=1}^{I} R_i^{\text{p,sr}}(t) \right] \tag{17a} \\
\text{s.t. } u_k(t) \in [0, D]^2, v_k(t) \leq v_{\max}, d_{k,l} \geq d_c, \tag{17b} \\
p_k^J(t) \leq P_{\max}^J, p_k^c(t) \leq P_{\max}^c, p_k^p(t) \leq P_{\max}^p, \tag{17c} \\
R_{e,k}^c(t) \leq R_k^c(t), 0 < \theta_k(t) < 1, \tag{17d} \\
\sum\nolimits_{k=1}^{K} \mathbf{S}_{k,i}^{\text{GT}}(t) \leq 1, \sum\nolimits_{i=1}^{I} \mathbf{S}_{k,i}^{\text{GT}}(t) \leq \mathcal{C}, \tag{17e} \\
\forall e \in \mathcal{E}, \forall i \in \mathcal{I}, \forall l, k \in \mathcal{K}, t \in \mathcal{T}, \tag{17f}
\end{align}$$

where $\mathbf{u}_t = [u_1(t), u_2(t), \cdots, u_K(t)]^T$, $\boldsymbol{\theta}_t = [\theta_1(t), \theta_2(t), \cdots, \theta_K(t)]$ and $\mathbf{p}_t^J = [p_1^J(t), p_2^J(t), \cdots, p_K^J(t)]^T$ represent the real-time positions, time-step sharing and the jamming power of UAVs at time step t, respectively. The objective function (17a) includes the secrecy rate of both common messages $R_k^{\text{c,sr}}(t)$ and private messages $R_i^{\text{p,sr}}(t)$. Constraint (17b) limits the UAV's service range and speed, and stipulate a minimum distance d_c to avoid collisions. While (17c) imposes a limit on the maximum power for transmitting the jamming power, common message and private messages. (17d) ensures that Eve cannot successfully decode the common stream, which in turn reduces the likelihood that Eve will be able to decode the private messages. The first term in (17e) means that each GT can at most be served by one UAV, and the second term guarantees that the number of GTs scheduled to each UAV does not exceed its service capacity.

III. Proposed MARL Solution with MTA-DRNN

In this section, we first convert the MUCTSJ problem (17) into a MDP. Then, we develop a MTA-DRNN architecture adapted to the network dynamics. Finally, we propose a MARL framework with MTA-DRNN to the optimization problem.

A. MDP Formulation of MUCTSJ Problem

Considering the complex state space and time-varying scenarios, traditional reinforcement learning algorithms often struggle to achieve optimal security performance. In this work, we leverage the concept of MARL to address the MUCTSJ problem. To do this, we first formulate the optimization problem as a MDP, which is characterized by the tuple $(\mathcal{S}, \mathcal{O}, \mathcal{A}, \mathcal{R})$. In this formulation, $\mathcal{S}, \mathcal{O}, \mathcal{A}$, and $\mathcal{R}$ denote the state space, received partial observations, action space, and reward function, respectively. At each time step t, the UAVs, acting as agents, periodically gather the current state of the environment and select an optimal action based on a predefined policy. The ultimate goal is to enhance the decision-making process of the agents, allowing them to navigate the complex state space effectively and adapt to the time-varying nature of the environment. Next, we will detail the components of the defined MDP framework.

1) State Space $\mathcal{S}$: The state space of the MDP is defined as $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_T\}$, where each state $\mathbf{s}_t$ consists of three parts, i.e., $\mathbf{s}_t = \{\mathbf{s}_t^a, \mathbf{s}_t^b, \mathbf{s}_t^c\}$. Here, $\mathbf{s}_t^a$ denotes the location coordinates of UAVs, GTs, and Eves at each time step t, providing spatial awareness necessary for trajectory planning, time-step sharing and jamming strategies. $\mathbf{s}_t^b$ captures the trajectory information of all UAVs up to time t, which is essential for understanding the movement patterns and making future path decisions to achieve objectives such as coverage and security while avoiding collisions. Finally, $\mathbf{s}_t^c$ contains performance metrics, including the secrecy rate, throughput, and fairness index at time step t, offering a quantitative assessment of the system's performance in terms of communication security, data transmission efficiency, and resource distribution fairness. Together, these components enable the MDP framework to represent the dynamic environment accurately, guiding the optimization of problem (17).

2) Observation $\mathcal{O}$: Each UAV k only has access to partial observations of the environment at time step t, denoted as $\mathbf{o}_t^k$. These observations include information about the UAV itself, other UAVs, and GTs within its coverage range. Specifically, the attributes of UAV k are represented as $\mathcal{F}_k(t) = (u_k(t), R_k^{\text{sr}}(t))$, where $u_k(t)$ denotes the location coordinates of UAV k, and $R_k^{\text{sr}}(t) = \sum_{i=1}^{N_{\mathcal{I}_k}} R_i^{\text{sr}}(t)$ is the secrecy rate of the GTs served by UAV k. The attributes of other UAVs relative to UAV k are given by $\mathcal{F}_k^{\text{UAV}}(t) = \{(u_{l,k}(t), R_l^{\text{sr}}(t)) | \forall l \in \mathcal{K}, l \neq k\}$, where $u_{l,k}(t) = u_l(t) - u_k(t)$ represents the relative coordinates of UAV l to UAV k, and $R_l^{\text{sr}}(t)$ is the secrecy rate of the GTs served by UAV l. For the GTs within UAV k's coverage, the attributes are $\mathcal{F}_k^{\text{GT}}(t) = \{(u_{i,k}(t), R_i^{\text{sr}}(t)) | \forall i \in \mathcal{I}, \mathbf{C}_{k,i}^{\text{GT}}(t) = 1\}$, where $u_{i,k}(t) = u_i - u_k(t)$ indicates the relative coordinates of GT i to UAV k, and $R_i^{\text{sr}}(t) = R_i^c(t) + R_i^p(t)$ is the secrecy rate of GT i including both common and private rates. Thus, the overall observations of UAV k at time step t is given as

$$\mathbf{o}_t^k = \left\langle \mathcal{F}_k(t), \mathcal{F}_k^{\text{UAV}}(t), \mathcal{F}_k^{\text{GT}}(t) \right\rangle, \tag{18}$$

enabling each UAV to make decisions based on local environmental data it gathers.

3) Action $\mathcal{A}$: Action $\mathbf{a}_t$ for UAV k at time step t consists of three components: the flying direction $\omega_k(t)$, time-step sharing $\theta_k(t)$, and the jamming power $p_k^J(t)$. To accurately capture the flight strategy of UAVs, we define the UAV's flight action as either hovering "still") or moving in one of 16 discrete horizontal directions. Therefore, the set of possible directions is given by $\omega_k(t) \in \{\text{still}, \frac{\pi}{8}, \frac{2\pi}{8} \cdots, 2\pi\}$. The UAV's position is updated accordingly; if it hovers, it remains in place, otherwise, it moves in the chosen direction with a velocity. Correspondingly, after UAV k takes an action, its position is updated according to the following rule

$$u_k(t+1) = \begin{cases} u_k(t) + [0,0,0]^T & \text{if } \omega_k(t) = \text{still}, \\ u_k(t) + v_{\max}[\cos\frac{\pi}{8}, \sin\frac{\pi}{8}, 0]^T & \text{if } \omega_k(t) = \frac{\pi}{8}, \\ \vdots & \\ u_k(t) + v_{\max}[\cos(2\pi), \sin(2\pi), 0]^T & \text{if } \omega_k(t) = 2\pi. \end{cases} \tag{19}$$

In the proposed scheme, the time-step sharing factor $\theta_k(t)$ determines the proportion of time allocated to the two phases in each time step. The UAV dynamically adjust $\theta_k(t)$ to optimize the trade-off between the two phases each time step, maximizing the secrecy rate while mitigating eavesdropping risks. To adapt to dynamic channel conditions and improve computational efficiency, the set of possible values for the time-step sharing factor is defined as $\theta_k(t) \in \{0.2, 0.4, 0.5, 0.6, 0.8\}$.

The jamming power $p_k^J(t)$ is constrained by a maximum value $P_{\max}^J$ and is quantized into $|L|$ discrete levels using logarithmic normalization, which is appropriate since jamming power can vary over several orders of magnitude. The set of allowed jamming power levels is given by

$$p_k^J(t) = \left\{ 0, \left\{ P_{\min}^J \left(\frac{P_{\max}^J}{P_{\min}^J} \right)^{\frac{i}{|L|-2}} \middle| i = 0, \cdots, |L|-2 \right\} \right\} \tag{20}$$

where P_{min}^J is the minimum non-zero jamming power. This logarithmic quantization ensures that power levels are spaced appropriately across their range, allowing for finer control over lower power levels and progressively larger intervals as power increases, making it suitable for the wide dynamic range of jamming power typically encountered in practice [37].

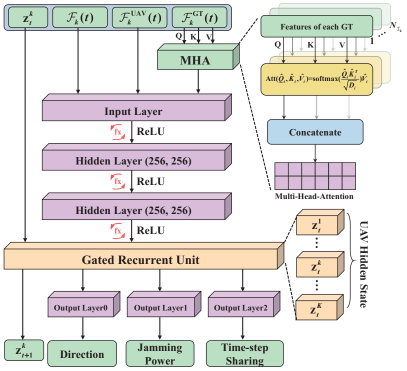


Fig. 3: Illustration of the proposed MTA-DRNN architecture.

4) Reward $\mathcal{R}$: The Reward $\mathbf{r}_t^k \in \mathcal{R}$ represents the immediate reward received by UAV k after executing action $\mathbf{a}_t^k$ in state $\mathbf{s}_t$. In this context, each UAV is tasked with the common objective of maximizing the global secrecy rate. In a cooperative setting, all agents share a global utility, which is defined based on (17a) as

$$r_t^{\text{glo}} = \zeta_r \times \mathbf{F}(\hat{\mathbf{d}}(t)) \left[\sum_{k=1}^{K} R_k^{\text{c,sr}}(t) + \sum_{i=1}^{I} R_i^{\text{p,sr}}(t) \right] \tag{21}$$

where ζ_r is a scaling factor that controls the numerical range, adjusted through trial-and-error. It is important to note that, in line with (16), it is crucial to separately consider the principles of CQF and FSF.

To monitor boundary violations, we introduce an indicator array $\eta^{\text{loc}} = \{0,1\}^K$, where $\eta_{k,t}^{\text{loc}} = 1$ if UAV k adheres to the spatial constraints at time step t, and $\eta_{k,t}^{\text{loc}} = 0$ otherwise. Besides, to avoid collisions during flight, we define another indicator array $\eta^{\text{col}} = \{0,1\}^K$, where $\eta_{k,t}^{\text{col}} = 1$ indicates a collision risk for UAV k at time step t, and $\eta_{k,t}^{\text{col}} = 0$ signifies safe operation. Similarly, we set an indicator array $\eta^{\text{sec}} = \{0,1\}^K$ and a constant penalty ϕ^{sec} for restriction (17d). Consequently, the reward function for UAV k is given by

$$\mathbf{r}_t^k = r_t^{\text{glo}} \times \eta_{k,t}^{\text{loc}} - \eta_{k,t}^{\text{col}} \times \phi^{\text{col}} - \eta_{k,t}^{\text{sec}} \times \phi^{\text{sec}}, \tag{22}$$

where ϕ^{col} is a constant representing the penalty imposed for collision risks. By incorporating these penalties into the reward function, the UAVs are incentivized to optimize their trajectories in a manner that balances secrecy performance, flight safety, and spatial compliance. This approach ultimately leads to robust and reliable system operation, ensuring both high secrecy rates and adherence to operational constraints.

B. Proposed MTA-DRNN Architecture

To address the challenges of a time-varying network environment, we developed a novel MTA-DRNN architecture. This design incorporates MHA to adapt to varying observation lengths and dimensions for each UAV, enabling more effective

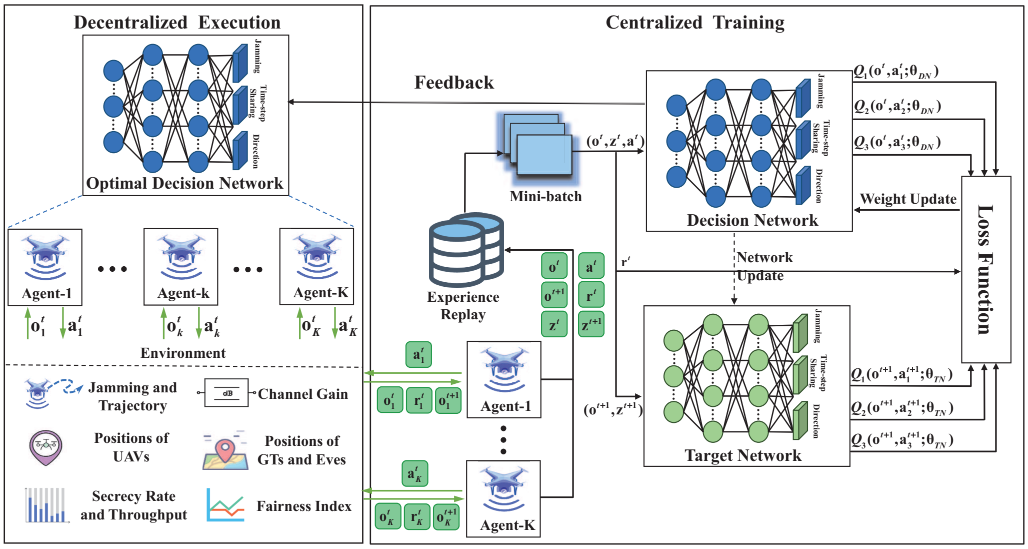


Fig. 4: Diagram of the multi-agent reinforcement learning framework utilizing MTA-DRNN for solving the MUCTSJ problem.

cluster control strategies. As shown in Fig. 3, the MTA-DRNN architecture includes an MHA layer, three Linear-layers with ReLU activation functions [38], a Recurrent Neural Network (RNN) layer, and three parallel output layers. Notably, all layers except the RNN hidden layer parameters, which are unique to each agent, share parameters across agents—referred to as the Parameter Sharing Layer. This parameter sharing allows multiple UAVs to use the same set of parameters for encoding and decoding environmental information, stabilizing training, and promoting cooperation between agents. Additionally, the RNN layer enhances decision-making by allowing each agent to leverage historical observations in a partially observable multi-agent environment [39].

Each UAV experiences varying observation lengths at different time steps, and the dimensions of these observations can also differ due to changes in the number of GTs being served. Traditional neural network models with a fixed input size struggle to effectively process this dynamic information. To address this, we incorporate a MHA mechanism before the model's input layer. The MHA module includes parallel scaled dot-product attention mechanisms equal to the maximum number of GTs that a UAV can serve. Scaled dot-product attention is calculated using the Query ($\hat{Q}$), Key ($\hat{K}$), and Value ($\hat{V}$) as

$$\mathrm{Att}(\hat{Q}_i, \hat{K}_i, \hat{V}_i) = \mathrm{softmax}\left(\frac{\hat{Q}_i \hat{K}_i^T}{\sqrt{D_i}}\right)\hat{V}_i \tag{23}$$

where D_i is the dimensionality of the attributes for each GT i. In our work, the attributes $\mathcal{F}_k^{\mathrm{GT}}(t)$ of GTs serve as the Query, Key, and Value for the MHA input. The attention mechanism normalizes the inputs of each GT into a fixed-sized vector, which are then combined through a concatenation function

$$\mathrm{MHA}(\mathcal{F}_k^{\mathrm{GT}}(t)) = \mathrm{Concat}(\mathrm{Att}_1, \cdots, \mathrm{Att}_{\mathrm{max}}). \tag{24}$$

By using attention, our model can handle information from a varying number of GTs based on their relative importance. Consequently, each agent receives the following observation at time step t, processed through MHA:

$$\mathbf{o}_t^k = \left\langle \mathcal{F}_k(t), \mathcal{F}_k^{\mathrm{UAV}}(t), \mathrm{MHA}(\mathcal{F}_k^{\mathrm{GT}}(t)) \right\rangle. \tag{25}$$

At the output end, we employ three parallel output layers, collectively referred to as the MTO structure, to separately handle jamming, time-step sharing and direction decisions. This design allows the model to fine-tune its parameters for each specific task independently, minimizing the gradient interference from unrelated tasks and thus leading to more accurate and effective decision-making. In addition, this decoupled schema helps to improve the scalability of MTA-DRNN. To integrate these objectives, we modify the loss function in Q-Learning to account for these output heads, assigning weights to the loss of each task. These weights are determined through a trial-and-error process. The detailed formulation of the loss function will be discussed in the following section.

C. MTA-MARL for MUTSJ Problem

We propose a MARL framework employing MTA-DRNN to tackle the MUCTSJ problem. This framework follows a centralized training and decentralized execution paradigm, as depicted in Fig. 4. During the centralized training phase, MTA-DRNN functions as both the Decision Network and the Target Network for the UAVs. By sharing an Experience Replay buffer across all UAVs, the training process becomes more efficient and stable. Conversely, in the decentralized execution phase, each UAV agent operates independently, making decisions based solely on its partial observations from the environment. This distinction allows the framework to leverage the benefits of centralized learning while enabling UAVs to act autonomously during execution. Next, we will delve into these two processes in detail.

1) Centralized Training: At the start of each episode, the initial locations of GTs and Eves are randomly set. At each time step t during training, UAV k obtains an observation $\mathbf{o}_t^k$ from the environment and then executes an action $\mathbf{a}_t^k$. After executing the action, UAV k receives a reward $\mathbf{r}_t^k$ and the next observation $\mathbf{o}_{t+1}^k$. This sequence forms the basis of the interaction loop. After each time step, the observations, actions, rewards, and subsequent observations from all agents are stored in the Experience Replay buffer. This experience is defined as $\tau = \{\tau_t^k = \{\mathbf{o}_t^k, \mathbf{a}_t^k, \mathbf{r}_t^k, \mathbf{o}_{t+1}^k\}, k \in \mathcal{K}, t \in \mathcal{T}\}$. To handle the partially observable nature of the problem, we employ RNNs within the MTA-DRNN architecture. The hidden states of the RNNs are also stored, and each experience is reformulated as $\tau = \{\tau_t^k = \{\mathbf{o}_t^k, \mathbf{z}_t^k, \mathbf{a}_t^k, \mathbf{r}_t^k, \mathbf{o}_{t+1}^k, \mathbf{z}_{t+1}^k\}, k \in \mathcal{K}, t \in \mathcal{T}\}$ where $\mathbf{z}_k^t$ represents the hidden state of the RNN for UAV k at time t. At the beginning of each episode, the hidden state for each agent is initialized to an all-zero vector, ensuring that the network starts without any prior context. Once a sufficient number of experiences are accumulated in the replay buffer, a mini-batch $\mathcal{B}$ of experiences is sampled for training. The elements of this mini-batch are then fed into the corresponding networks to update their parameters.

In the Decision Network, agents compute the corresponding Q-value $Q(\mathbf{o}_t, \mathbf{a}_t; \boldsymbol{\theta}_{\text{DN}})$ for the current state-action pair $(\mathbf{o}_t, \mathbf{a}_t)$, where $\boldsymbol{\theta}_{\text{DN}}$ denotes the network parameters of the Decision Network. Concurrently, the Target Network calculates the Q-value $Q(\mathbf{o}_{t+1}, \mathbf{a}_{t+1}; \boldsymbol{\theta}_{\text{TN}})$ based on the new state-action pair $(\mathbf{o}_{t+1}, \mathbf{a}_{t+1})$, where $\boldsymbol{\theta}_{\text{TN}}$ representing the parameters of the Target Network. As discussed in the previous section, our proposed MTA-DRNN has three parallel output heads to handle jamming, time-step sharing and direction decisions. Conveniently, let $Q_1(\mathbf{o}_t, \mathbf{a}_t^1; \boldsymbol{\theta}_{\text{DN}})$, $Q_2(\mathbf{o}_t, \mathbf{a}_t^2; \boldsymbol{\theta}_{\text{DN}})$ and $Q_3(\mathbf{o}_t, \mathbf{a}_t^3; \boldsymbol{\theta}_{\text{DN}})$ represent the decision Q-values for jamming, time-step sharing and direction decisions, respectively. Let $Q_1(\mathbf{o}_{t+1}, \mathbf{a}_{t+1}^1; \boldsymbol{\theta}_{\text{TN}})$, $Q_2(\mathbf{o}_{t+1}, \mathbf{a}_{t+1}^2; \boldsymbol{\theta}_{\text{TN}})$ and $Q_3(\mathbf{o}_{t+1}, \mathbf{a}_{t+1}^3; \boldsymbol{\theta}_{\text{TN}})$ denote the target Q-values for jamming, time-step sharing and direction, respectively.

The loss function for updating the Decision Network can be formulated as follows

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) =& \frac{1}{|\mathcal{B}|K} \sum_{b=1}^{|\mathcal{B}|} \sum_{k=1}^{K} \Bigg[\mu_1 \left(\delta_t^{k,b,1} - Q_1(\mathbf{o}_t^{k,b}, \mathbf{a}_t^{k,b,1}; \boldsymbol{\theta}_{\text{DN}}) \right)^2 \\ &+ \mu_2 \left(\delta_t^{k,b,2} - Q_2(\mathbf{o}_t^{k,b}, \mathbf{a}_t^{k,b,2}; \boldsymbol{\theta}_{\text{DN}}) \right)^2 \\ &+ \mu_3 \left(\delta_t^{k,b,3} - Q_3(\mathbf{o}_t^{k,b}, \mathbf{a}_t^{k,b,3}; \boldsymbol{\theta}_{\text{DN}}) \right)^2 \Bigg], \end{aligned} \tag{26}$$

where $\delta_t^{k,b,1} = \mathbf{r}_t^{k,b} + \gamma \arg\max_{\mathbf{a}_{t+1}^{k,b,1}} Q_1(\mathbf{o}_{t+1}^{k,b}, \mathbf{a}_{t+1}^{k,b,1}; \boldsymbol{\theta}_{\text{TN}})$, $\delta_t^{k,b,2} = \mathbf{r}_t^{k,b} + \gamma \arg\max_{\mathbf{a}_{t+1}^{k,b,2}} Q_2(\mathbf{o}_{t+1}^{k,b}, \mathbf{a}_{t+1}^{k,b,2}; \boldsymbol{\theta}_{\text{TN}})$ and $\delta_t^{k,b,3} = \mathbf{r}_t^{k,b} + \gamma \arg\max_{\mathbf{a}_{t+1}^{k,b,3}} Q_3(\mathbf{o}_{t+1}^{k,b}, \mathbf{a}_{t+1}^{k,b,3}; \boldsymbol{\theta}_{\text{TN}})$ are the Temporal-Difference (TD) errors for jamming, time-step sharing and direction decisions. Here, μ_1, μ_2 and μ_3 are weights that balance the importance of the different decision losses, and the superscript b indexes the samples. The parameters of the Decision Network are updated using the gradient of the loss function as

$$\boldsymbol{\theta}_{\text{DN}} \leftarrow \boldsymbol{\theta}_{\text{DN}} + \lambda \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}) \tag{27}$$

where λ is the learning rate. For the Target Network, its parameters $\boldsymbol{\theta}_{\text{TN}}$ are updated from Decision Network $\boldsymbol{\theta}_{\text{DN}}$ using polyak averaging as

$$\boldsymbol{\theta}_{\text{TN}} = \varphi \boldsymbol{\theta}_{\text{DN}} + (1 - \varphi) \boldsymbol{\theta}_{\text{TN}} \tag{28}$$

Here, φ is a factor that determines the update rate for the Target Network parameters.

2) Decentralized Execution: During the decentralized execution phase, each UAV operates autonomously using the trained optimal decision network. This network enables each UAV to make decisions based on its local observations $\mathbf{o}_t^k$ derived from the current state of the environment $\mathbf{s}_t$. The decentralized nature of this process means that each UAV does not rely on centralized control or global state information, which is crucial in dynamic and potentially communication-constrained UAV networks. The entire distributed execution process can be described as Algorithm 1.

Algorithm 1: Proposed MTA-MARL Algorithm.

```
Input: Initial network state $\mathcal{S}$; number of training
       episodes $N_{epi}$; episode length $T$; mini-batch
       size $|\mathcal{B}|$; learning rate $\lambda$.
Output: Optimized UAV action set $\mathcal{A} = \{\mathbf{a}_t =
        \{\omega_k(t), \theta_k(t), p_k^J(t) \mid \forall k \in \mathcal{K}\}, t \in \mathcal{T}\}$;
        trained decision network.
Initialization: Parameters of decision network $\boldsymbol{\theta}_{\text{DN}}$;
  parameters of target network $\boldsymbol{\theta}_{\text{TN}}$; experience replay
  memory $\mathcal{D}$; RNN hidden states $\{\mathbf{z}_k = \mathbf{0} \mid \forall k \in \mathcal{K}\}$
for episode $= 0$ to $N_{epi} - 1$ do
    Reset positions of UAVs $\{u_k(0) \mid \forall k \in \mathcal{K}\}$, GTs
      $\{u_i \mid \forall i \in \mathcal{I}\}$, and Eves $\{u_e \mid \forall e \in \mathcal{E}\}$
    for time step $t = 0$ to $T$ do
        for each UAV $k \in \mathcal{K}$ do
            Get the observations $\mathbf{o}_t^k$ based on (18), then
              encode GT attributes $\mathcal{F}_k^{\text{GT}}(t)$ using MHA
              to obtain MHA$(\mathcal{F}_k^{\text{GT}}(t))$ and form
              enhanced observation $\mathbf{o}_t^k$ via Eq. (25);
            Select actions $\mathbf{a}_t^k$ using $\epsilon - greedy$ policy;
            Receive reward $\mathbf{r}_t^k$ and observe state $\mathbf{o}_{t+1}^k$;
        end
        Environment transitions from state $\mathbf{s}_t$ to $\mathbf{s}_{t+1}$;
        Store $\tau = \{\{\mathbf{o}_k^t, \mathbf{z}_t^k, \mathbf{a}_t^k, \mathbf{r}_t^k, \mathbf{o}_{t+1}^k, \mathbf{z}_{t+1}^k\}, k \in \mathcal{K}\}$
          into experience replay memory $\mathcal{D}$;
        if $|\mathcal{D}| \geq |\mathcal{B}|$ then
            Sample a mini-batch $\mathcal{B}$ from $\mathcal{D}$;
            Obtain $Q_1$, $Q_2$, $Q_3$ from Decision Network
              and Target Network;
            Minimize loss function Eq. (26) and then
              update weights of Decision Network $\boldsymbol{\theta}_{\text{DN}}$
              according to Eq. (27);
        end
        Update target network parameters $\boldsymbol{\theta}_{\text{TN}}$ using
          Polyak averaging (Eq. 28)
    end
end
```

At the beginning of each episode, the environment is reset to establish initial conditions, including the take-off points of UAVs and the positions of GTs and Eves. At each time step $t \in \mathcal{T}$, each UAV agent collects an observation $\mathbf{o}_t^k$, which includes partial information about the environment. This observation is processed using MHA to encode the attributes of GTs, making the input suitable for the MTA-DRNN architecture. Based on this encoded information, each UAV selects an action $\mathbf{a}_t^k$ using an ϵ-greedy policy derived from the decision network.

After all UAVs execute their actions, the environment transitions from state $\mathbf{s}_t$ to $\mathbf{s}_{t+1}$, reflecting the changes brought about by the UAVs' actions. Each UAV then receives new observations $\mathbf{o}_{t+1}^k$, which provide updated information about the environment's current state. As the UAVs interact with the environment over multiple time steps, their experiences—comprising observations, actions, rewards, and subsequent observations—are stored in the experience replay buffer $\mathcal{D}$. Once sufficient experiences have been collected, the UAVs update the parameters of both Decision Network $\boldsymbol{\theta}_{\text{DN}}$ and Target Network $\boldsymbol{\theta}_{\text{TN}}$ by sampling a mini-batch $\mathcal{B}$ from the replay buffer. This batch is used to perform gradient descent, improving the policy by minimizing the TD error. Training continues until the rewards converge, indicating that the UAVs have learned a stable policy. At this point, each UAV is capable of executing actions at any time step t based solely on its partial observations, without requiring additional information, thus achieving effective decentralized decision-making.

D. Complexity Analysis

In this subsection, we analyze the computational complexity of the proposed algorithm during both centralized training and decentralized execution phases. The key computational component of the MARL framework is the MTA-DRNN, which we examine in detail. Following the analysis in [40], the time complexity of the MHA mechanism is $O(I \times |\mathcal{F}_k^{\text{GT}}(t)|^2 + I^2 \times |\mathcal{F}_k^{\text{GT}}(t)|)$, where I denotes the number of attention heads, and $|\mathcal{F}_k^{\text{GT}}(t)|$ is the dimensionality of the GT features. The three fully connected layers contribute a complexity of $O(3 \times |\boldsymbol{\theta}|^2)$, where $|\boldsymbol{\theta}|$ is the number of neurons per layer. Since each UAV maintains an independent RNN hidden state, the RNN layer scales with the number of agents, resulting in a complexity of $O(K \times |\boldsymbol{\theta}|^2)$ where K is the total number of UAVs. In addition, the three parallel output layers, which generate direction and jamming decisions, contribute $O(|\mathbf{a}_k^t| \times |\boldsymbol{\theta}|)$, where $|\mathbf{a}_k^t|$ denotes the dimensionality of the action space for each agent. Consequently, the total time complexity for a single inference step using the MTA-DRNN architecture is $O_I(I \times |\mathcal{F}_k^{\text{GT}}(t)|^2 + I^2 \times |\mathcal{F}_k^{\text{GT}}(t)| + (3+K) \times |\boldsymbol{\theta}|^2 + |\mathbf{a}_k^t| \times |\boldsymbol{\theta}|)$. Based on this analysis, the overall time complexity of MTA-MARL algorithm during centralized training and decentralized execution is summarized as follows:

- ***Centralized Training:*** The computational complexity is primarily determined by the number of training episodes N_{epi}, the number of time steps per episode T, the batch-size $|\mathcal{B}|$ and the training interval N_d. So, the total training complexity is $O(N_{epi} \times T \times O_I + N_{epi}/N_d \times |\mathcal{B}| \times O_I)$.
- ***Decentralized Execution:*** During decentralized execution, each agent collects observations $\mathbf{o}^t$ from the environment $\mathcal{S}$ and selects optimal actions $\mathbf{a}^t$ at each time step t. Thus, The overall complexity for a single episode $\mathcal{T}$ is $O(T \times O_I)$.

TABLE III: Parameters Settings.

Parameters	Values (Unit)
Coverage and serve ranges of UAV ($\mathcal{R}^{\text{GT}}$,$\mathcal{R}^{\text{UAV}}$)	50,50 (m)
Maximum velocity and altitude of UAV ($v_{\max}, H_{\text{uav}}$)	20 (m/s),100m
Central carrier frequency (f_c)	2.4 (GHz)
Maximum power of common stream of UAV (p_k^c)	30 (dBm)
Maximum power of private stream of UAV (p_k^p)	10 (dBm)
Maximum power of forward stream of GT j_k ($p_{j_k}^f$)	10 (dBm)
Maximum jamming Power of UAV ($P_{\max}^J$)	15 (dBm)
PSD of AWGN at GTs (σ^2)	-170 (dBm/HZ)
Channel S-curve parameters (δ, f)	9.61, 0.15
Maximum number of GTs served ($\mathcal{C}$)	5

IV. Performance Evaluation and Discussion

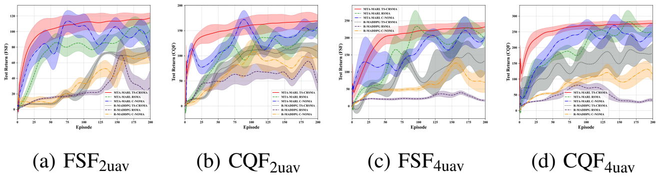
We first evaluate the experiments to evaluate the performance of our proposed MTA-MARL algorithm through experiments conducted on a computing host equipped with an NVIDIA GeForce RTX 4080 GPU, using PyTorch 2.2.2 for deep learning computations. We then test the algorithm on the Nvidia Jetson Xavier NX module (TensorRT 8.5.2) to assess its performance in embedded systems.

A. Experiment Setup

Parameter Settings: In our simulations, we consider a multi-UAV network where GTs and Eves are randomly distributed over a 400 m $\times$ 400 m area. Each UAV employs the MTA-DRNN architecture as its own decision network to adapt to complex environmental in real time. Key parameters are summarized in Table III. The number of attention heads in MHA is set to 2. The discount factor for each agent's reward is $\gamma = 0.99$, and the reward scaling factor is set to $\zeta_r = 0.1$. The decision network is trained over approximately $N_{epi} = 100,000$ episodes, with each episode consisting of $\mathcal{T} = 100$ time steps. For training process, $\epsilon - greedy$ exploration rate decreases from 1.0 to a final value of 0.05 over N_{epi} episodes to balance exploration and exploitation. The replay buffer size is set to $|\mathcal{D}| = 100,000$. A mini-batch size of $|\mathcal{B}| = 128$ is used, and the initial learning rate is set to 2.5×10^{-4}. To improve training stability, we use the LambdaLR learning rate scheduler, which decays the learning rate by 1% after each parameter update, with a minimum of 40% of the initial learning rate. To accelerate convergence and mitigate overfitting, we adopt the AdamW optimizer with a weight decay coefficient of 0.01 to update the network parameters. The target network is updated from the decision network using Polyak averaging, with a smoothing coefficient of $\varphi = 1 \times 10^{-4}$. Furthermore, to prevent gradient explosion, gradients are clipped within the range $[-1, 1]$.

Baseline Settings: To comprehensively evaluate the performance of the proposed solution with TS-CRSMA, we implement five baseline methods for comparison. The details of each baseline are as follows:

- MTA-MARL with TS-CRSMA: The proposed algorithm integrated with the two-stage collaborative RSMA in this work.
- MTA-MARL with RSMA: The proposed algorithm combined with the single-antenna RSMA described in [31].
- MTA-MARL with C-NOMA: The proposed algorithm paired with the cooperative NOMA from [41].



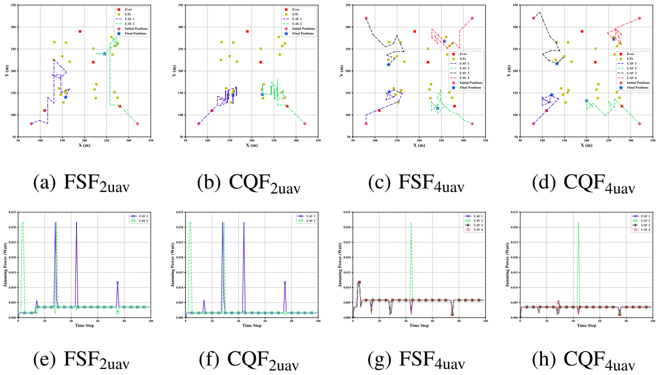
(a) $\text{FSF}_{2\text{uav}}$ (b) $\text{CQF}_{2\text{uav}}$ (c) $\text{FSF}_{4\text{uav}}$ (d) $\text{CQF}_{4\text{uav}}$

Fig. 5: Training and testing performance of different baselines.

- R-MADDPG with TS-CRSMA: The recurrent Multi-Agent Deep Deterministic Policy Gradient (R-MADDPG) algorithm [42] applied with TS-CRSMA scheme.
- R-MADDPG with RSMA: The R-MADDPG algorithm combined with the original RSMA scheme.
- R-MADDPG with C-NOMA: The R-MADDPG algorithm applied alongside the C-NOMA scheme.

B. Performance Results

To evaluate the training performance of the proposed method, we present Fig. 5 which illustrates the convergence behavior for varying numbers of UAVs. The initial horizontal positions of UAVs 1 through 4 are set to $[80, 80]$, $[320, 80]$, $[80, 320]$, and $[320, 320]$, respectively. During training, model performance is tested every $N_{epi}/200$ episodes. A performance comparisons was performed between the proposed MTA-MARL with the benchmarks. The results show that the proposed method has faster convergence, smaller variance and higher stability than other schemes.

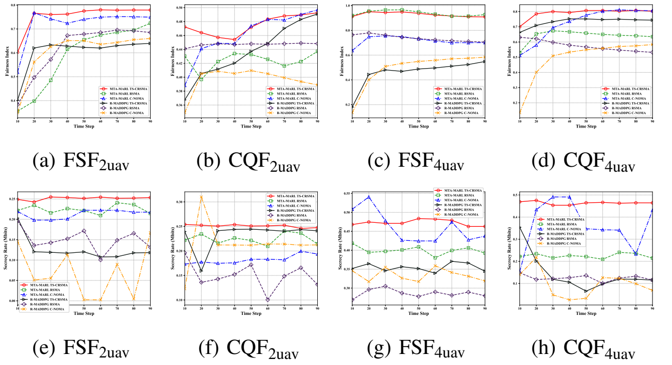


(a) $\text{FSF}_{2\text{uav}}$ (b) $\text{CQF}_{2\text{uav}}$ (c) $\text{FSF}_{4\text{uav}}$ (d) $\text{CQF}_{4\text{uav}}$

(e) $\text{FSF}_{2\text{uav}}$ (f) $\text{CQF}_{2\text{uav}}$ (g) $\text{FSF}_{4\text{uav}}$ (h) $\text{CQF}_{4\text{uav}}$

Fig. 6: Trajectories and jamming power with various UAVs' number.

Fig. 6 illustrates the flight trajectories and jamming power levels of UAVs serving GTs under the FSF and CQF principles. As shown in Fig. 6a and Fig. 6b, UAVs consistently move away from Eve to enhance overall network security. A comparison between the FSF and CQF principles reveals distinct operational strategies. Under the FSF principle, the two UAVs prioritize serving the global GTs to ensure fairness in system throughput. In contrast, under the CQF principle, UAVs focus on serving nearby GTs with better channel quality. This behavioral difference is similarly evident in Fig. 6c and Fig. 6d. With a larger UAV fleet, each UAV can focus on a smaller, dedicated area, reducing the need for long-distance movement and thereby improving overall service efficiency. The optimal jamming power decisions corresponding to the flight trajectories are depicted in Fig. 6e-6h. When a UAV serves GTs located near Eve, it employs higher jamming power to mitigate potential eavesdropping. In other cases, the UAV maintains lower jamming power to reduce interference with legitimate GTs. By comparing Fig. 6e with 6f and Fig. 6g with 6h, it is evident that the FSF principle incurs higher interference power than CQF, primarily for two reasons. First, to enhance throughput fairness, the UAV deliberately transmits a small amount of interference to slightly suppress the high throughput of central GTs—where channel conditions are strong—while this same interference has negligible impact on edge GTs with poorer links. Second, the FSF-driven trajectory, which prioritizes serving distant or underserved users, often brings the UAV closer to Eve's threat zone, triggering more frequent protective jamming; in contrast, CQF keeps the UAV in safer regions, reducing the need for interference.



(a) $\text{FSF}_{2\text{uav}}$ (b) $\text{CQF}_{2\text{uav}}$ (c) $\text{FSF}_{4\text{uav}}$ (d) $\text{CQF}_{4\text{uav}}$

(e) $\text{FSF}_{2\text{uav}}$ (f) $\text{CQF}_{2\text{uav}}$ (g) $\text{FSF}_{4\text{uav}}$ (h) $\text{CQF}_{4\text{uav}}$

Fig. 7: Performance comparison in terms of fairness index ((a)-(d)) and secrecy rate ((e)-(h)).

Moreover, the FSF principle incurs higher interference power than CQF, primarily for two reasons. First, to improve fairness, the UAV reduces interference toward high-throughput ground terminals (GTs) in the coverage center—where channels are strong—while this has minimal impact on edge GTs. Second, FSF drives UAVs to follow longer, more diverse trajectories to serve distant users, increasing exposure to Eve's threat zone and thus triggering more frequent protective jamming, whereas CQF keeps UAVs in safer regions with lower interference needs.

To provide a more intuitive demonstration of the proposed solution's performance, Fig. 7 presents a detailed comparison of six baseline methods in terms of secrecy rate and fairness index. Overall, the proposed MTA-MARL with TS-CRSMA consistently outperforms other methods across different UAV counts and service principles. Specifically, as shown in Fig. 7a-7d, the fairness index under the FSF principle stabilizes at a higher level than under the CQF principle. This is because the FSF principle prioritizes serving a larger number of users, resulting in improved fairness compared to the CQF principle. Moreover, the fairness index is higher in the four-UAV scenario than in the two-UAV scenario. This improvement is attributed to the increased number of UAVs, which enables each UAV to serve users within its designated region, thereby minimizing the likelihood of neglecting GTs due to long-distance travel. This behavior is further corroborated by the trajectory results shown in Fig. 6.

Fig. 7e-7h illustrate the secrecy rate achieved by the entire network at each time step for both two-UAV and four-UAV network scenarios. We can observe a very interesting phenomenon: under the same scenario, the secrecy rate that our proposed MTA-MARL with TS-CRSMA consistently converges to a similar value smoothly, regardless of the service principle in effect. In contrast, the secrecy performance of other baseline methods is more significantly influenced by the choice of service principle. Furthermore, we observe that the secrecy rate increases as the number of UAVs grows, and the fluctuation of the curve is less obvious when the number of drones is large compared to the case with a small number of drones. By systematically comparing the metrics of the proposed algorithms and transmission schemes while controlling variables, it is evident that our proposed solution exhibits strong adaptability to different scenarios.

C. Deployment Adaptability Discussion

To more realistically simulate the actual performance of the proposed solution when deployed on the UAV platform, we expanded the area to 800m × 800m, and then utilized both Nvidia Jetson and the Host machine to run the decision model and test the algorithm's performance concurrently.

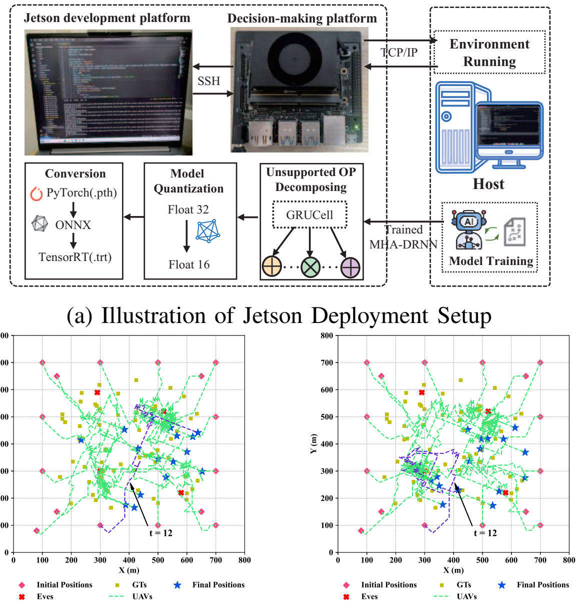
As illustrated in Fig. 8a, the Jetson platform is remotely developed using VSCode over a Secure Shell (SSH) connection. On the Jetson platform, we first quantize the trained decision network to half-precision (FP16) offline to optimize inference speed and reduce memory usage. After this quantization step, we convert the network to the `.onnx` format and subsequently transform it into the `.trt` format, which is loaded using Nvidia's TensorRT framework. It is important to note that, for operators that are not natively supported by ONNX, such as `GRUCell`, we decompose them using supported ONNX operators to ensure compatibility. Once these preparations are complete, the Jetson platform can be employed as the decision-making platform. The agents on the decision-making platform interact with the simulation environment $\mathcal{S}$ on the Host through the TCP/IP protocol.

Table IV summarizes the inference time on Jetson and the Host across one episode $\mathcal{T}$, with varying numbers of GTs and UAVs. From the table, it is evident that as the number of UAVs increases, the inference time increases significantly. This is due to the larger hidden state parameters of the RNN. However, for the same number of UAVs, the inference time does not necessarily increase with the number of GTs. This is likely because the MHA mechanism in the decision network allows UAVs to better encode the dynamic environmental information, reducing the impact of GT quantity on inference time.

In the following, we will demonstrate the algorithm's performance by analyzing various metrics on both the Jetson and Host platforms. Fig. 8b and Fig. 8c depict the trajectories inferred by the Host and Jetson platforms, respectively, during episode $\mathcal{T}$ in a scenario with 16 UAVs and 60 GTs. Interestingly, we observe that the trajectories of the UAVs are identical in the early stages on both Jetson and Host platforms. However, after a certain time step (e.g., at step 12), the trajectories start to diverge. The reason for this phenomenon lies in the FP16 quantization applied to the decision network on the Jetson platform, which causes a loss in inference precision. This loss of precision at time step t results in a decision $\hat{\mathbf{a}}_t^k$ made by the UAVs on Jetson that differs from the decision $\mathbf{a}_t^k$ made on the Host. As a consequence, the next environment state $\hat{\mathbf{s}}_{t+1}$ on Jetson deviates from the corresponding environment state $\mathbf{s}_{t+1}$ on the Host, leading to different subsequent trajectories.

TABLE IV: INFERENCE TIME (MS)

Number of UAVs	Host			Jetson		
	20	40	60	20	40	60
4	14.57	14.67	15.34	214.24	217.73	224.26
8	14.78	15.25	15.56	232.66	223.05	243.54
12	15.22	15.38	15.29	248.21	255.44	267.86
16	15.24	15.57	15.88	251.57	269.91	293.34



(a) Illustration of Jetson Deployment Setup

(b) Trajectory on Jetson

(c) Trajectory on Host

Fig. 8: The inference results of the proposed algorithm on Jetson and Host in terms of the UAV trajectory.

We further provide the performance comparisons with varying numbers of UAVs and GTs on Host and Jetson in Table V. As the number of GTs increases (e.g., from 20 to 60), we can observe an expected rise in both secrecy rate and throughput on both platforms. However, this improvement comes at a cost: an increased number of GTs results in 'fatigue' effect on UAVs, where the fairness index tends to decline slightly, especially when UAVs are tasked with handling larger volumes of GTs. For example, when the number of UAVs is 12, the fairness index on Host drops from 0.54 for 20 GTs to 0.48 for 60 GTs, indicating that UAVs struggle to maintain equal service distribution among all GTs. From Table V, we see that although the Jetson platform suffers from some precision loss, both platforms exhibit similar overall performance metrics

TABLE V: PERFORMANCE COMPARISON RESULTS ON HOST AND JETSON

Number of GTs	Device Type	Secrecy Rate (MBits)				Throughput (MBits)				Fairness Index			
		4	8	12	16	4	8	12	16	4	8	12	16
20	Host	96.34	124.70	136.76	142.32	326.12	515.11	559.63	578.81	0.31	0.30	0.57	0.54
	Jetson	105.32	130.79	133.91	145.31	355.42	547.24	518.91	561.32	0.30	0.32	0.54	0.53
40	Host	115.32	155.60	179.67	197.14	572.91	1372.01	1527.47	1579.35	0.27	0.35	0.49	0.48
	Jetson	110.25	180.39	170.77	200.09	600.31	1497.63	1473.41	1629.21	0.29	0.36	0.47	0.48
60	Host	123.61	172.01	188.41	211.69	1445.02	2234.23	2465.70	2990.24	0.29	0.36	0.45	0.48
	Jetson	135.72	163.42	191.44	238.89	1554.48	2118.23	2327.81	3019.61	0.27	0.35	0.47	0.49

in terms of secrecy rate, throughput, and fairness index. The difference in performance between the Jetson and Host platforms is minimal, demonstrating the robustness of the proposed MTA-MARL algorithm across different hardware environments.

V. CONCLUSION

This paper proposed a novel two-phase collaborative RSMA transmission scheme to enhance communication security in multi-UAV networks. To further improve network secrecy, we developed a MARL framework based on MTA-DRNN, which jointly optimized UAV trajectories, time-step sharing and jamming power to maximize secrecy rate. Simulation results demonstrated that the proposed framework effectively captured various high-dimensional probability distributions of decisions, enabling agents to make optimal policy decisions in scenarios with varying numbers of UAVs and GTs. This capability facilitated high levels of UAV coordination, ensuring mission consistency and reliability. Additionally, an evaluation on the Nvidia Jetson platform confirmed the framework's robustness and adaptability, underscoring its potential for deployment on practical UAV systems.

In this paper, we address the security challenges of single-antenna RSMA-enabled multi-UAV networks. Extending the proposed framework to multi-antenna RSMA-enabled multi-UAV systems would harness spatial degrees of freedom via advanced beamforming, thereby enabling simultaneous improvements in PLS and spectral efficiency. Therefore, a particularly promising research direction is the investigation of PLS in multi-antenna RSMA-enabled multi-UAV networks. Moving in this direction requires redesigning the RSMA framework to incorporate precoding, which correspondingly increases the complexity of system modeling and optimization.

REFERENCES

[1] G. Geraci *et al.*, "What will the future of UAV cellular communications be? A flight from 5G to 6G," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 3, pp. 1304–1335, 3rd Quart., 2022.

[2] X. Cao *et al.*, "Exploring LLM-based multi-agent situation awareness for zero-trust space-air-ground integrated network," *IEEE J. Sel. Areas Commun.*, vol. 43, no. 6, pp. 2230–2247, Jun. 2025.

[3] T. Taleb, A. Ksentini, H. Hellaoui, and O. Bekkouche, "On supporting UAV based services in 5G and beyond mobile systems," *IEEE Netw.*, vol. 35, no. 4, pp. 220–227, Aug. 2021.

[4] H. Huang and A. V. Savkin, "Deployment of heterogeneous UAV base stations for optimal quality of coverage," *IEEE Internet Things J.*, vol. 9, no. 17, pp. 16 429–16 437, Sep. 2022.

[5] X. Zhang, H. Zhao, J. Wei, C. Yan, J. Xiong, and X. Liu, "Cooperative trajectory design of multiple UAV base stations with heterogeneous graph neural networks," *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 1495–1509, Mar. 2022.

[6] P. A. Apostolopoulos, G. Fragkos, E. Tsiropoulou, and S. Papavassiliou, "Data offloading in UAV-assisted multi-access edge computing systems under resource uncertainty," *IEEE Trans. Mob. Comput.*, vol. 22, no. 1, pp. 175–190, Jan. 2023.

[7] Z. Ning *et al.*, "Dynamic computation offloading and server deployment for UAV-enabled multi-access edge computing," *IEEE Trans. Mob. Comput.*, vol. 22, no. 5, pp. 2628–2644, Sep. 2023.

[8] A. Fotouhi *et al.*, "Survey on UAV cellular communications: Practical aspects, standardization advancements, regulation, and security challenges," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3417–3442, 4th Quart., 2019.

[9] P. Angueira *et al.*, "A survey of physical layer techniques for secure wireless communications in industry," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 810–838, 2nd Quart., 2022.

[10] H. Fang, L. Xu, G. Nan, D. Zheng, H. Zhao, and X. Wang, "Accountable distributed access control with privacy preservation for blockchain-enabled internet of things systems: A zero-trust security scheme," *IEEE Internet Things J.*, vol. 12, no. 11, pp. 17 936–17 947, Jun. 2025.

[11] Y. Li, D. Zheng, H. Fang, H. Xing, X. Chen, and X. Cao, "Towards prompt chain deployment in zero trust-enabled compute first networks," *IEEE Trans. Consum. Electron.*, Jun. 2025, early access.

[12] X. Sun, D. W. K. Ng, Z. Ding, Y. Xu, and Z. Zhong, "Physical layer security in UAV systems: Challenges and opportunities," *IEEE Wireless Commun.*, vol. 26, no. 5, pp. 40–47, Oct. 2019.

[13] H.-M. Wang, X. Zhang, and J.-C. Jiang, "UAV-involved wireless physical-layer secure communications: Overview and research directions," *IEEE Wireless Commun.*, vol. 26, no. 5, pp. 32–39, Oct. 2019.

[14] M. Zhao, Z. Wang, K. Guo, R. Zhang, and T. Q. Quek, "Against mobile collusive eavesdroppers: Cooperative secure transmission and computation in UAV-assisted MEC networks," *IEEE Trans. Mobile Comput.*, vol. 24, no. 6, pp. 5280–5297, Jun. 2025.

[15] Y. Li, H. Zhang, and K. Long, "Joint resource, trajectory, and artificial noise optimization in secure driven 3-D UAVs with NOMA and imperfect CSI," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 11, pp. 3363–3377, Nov. 2021.

[16] J. Li *et al.*, "Multi-objective optimization approaches for physical layer secure communications based on collaborative beamforming in UAV networks," *IEEE/ACM Trans. Netw.*, vol. 31, no. 4, pp. 1902–1917, Aug. 2023.

[17] L. Guo, J. Jia, J. Chen, and X. Wang, "Secure communication optimization in NOMA systems with UAV-mounted STAR-RIS," *IEEE Trans. Inf. Forensics Security*, vol. 19, pp. 2300–2314, 2023.

[18] Y. Mao *et al.*, "Rate-splitting multiple access: Fundamentals, survey, and future research trends," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 4, pp. 2073–2126, 4th Quart. 2022.

[19] M. Xiao, H. Cui, Z. Zhao, X. Cao, and D. O. Wu, "Joint 3D deployment and beamforming for RSMA-enabled UAV base station with geographic information," *IEEE Trans. Wireless Commun.*, vol. 23, no. 4, pp. 2547–2559, Apr. 2023.

[20] W. Jaafar, S. Naser, S. Muhaidat, P. C. Sofotasios, and H. Yanikomeroglu, "On the downlink performance of RSMA-based UAV communications," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 16 258–16 263, Dec. 2020.

[21] L. Liu, L. Wu, and L. Qiu, "Performance analysis and ee optimization in UAV-assisted dual-layer heterogeneous communication network based

on RSMA," *IEEE Trans. Veh. Technol.*, vol. 74, no. 7, pp. 10 912–10 926, Jul. 2025.

[22] J. Ji, L. Cai, K. Zhu, and D. Niyato, "Decoupled association with rate splitting multiple access in UAV-assisted cellular networks using multi-agent deep reinforcement learning," *IEEE Trans. Mobile Comput.*, vol. 23, no. 3, pp. 2186–2201, Mar. 2023.

[23] H. Fu, S. Feng, W. Tang, and D. W. K. Ng, "Robust secure beamforming design for two-user downlink MISO rate-splitting systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 12, pp. 8351–8365, Dec. 2020.

[24] H. Bastami, M. Letafati, M. Moradikia, A. Abdelhadi, H. Behroozi, and L. Hanzo, "On the physical layer security of the cooperative rate-splitting-aided downlink in UAV networks," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 5018–5033, Oct. 2021.

[25] H. Bastami, M. Moradikia, A. Abdelhadi, H. Behroozi, B. Clerckx, and L. Hanzo, "Maximizing the secrecy energy efficiency of the cooperative rate-splitting aided downlink in multi-carrier UAV networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 11, pp. 11 803–11 819, Nov. 2022.

[26] H. Bastami, M. Moradikia, M. Letafati, A. Abdelhadi, and H. Behroozi, "Outage-constrained robust and secure design for downlink rate-splitting UAV networks," in *Proc. IEEE Int. Conf. Commun. Workshops.*, 2021, pp. 1–7.

[27] X. Pei, X. Wang, J. Liu, J. Li, Y. Chen, and M. Wen, "On the physical-layer security of RSMA-based UAV communications with imperfect CSI over nakagami-m fading channels," in *2024 IEEE/CIC Int. Conf. Commun. China*, 2024, pp. 705–710.

[28] X. Pei, Y. Chen, M. Wen, T. Pei, X. Wang, and T. A. Tsiftsis, "Secrecy performance analysis of RSMA-based communications under partial CSIT against randomly located eavesdroppers," *IEEE Trans. Veh. Technol.*, vol. 73, no. 10, pp. 15 727–15 732, Oct. 2024.

[29] H. Bastami, H. Behroozi, M. Moradikia, A. Abdelhadi, D. W. K. Ng, and L. Hanzo, "Large-scale rate-splitting multiple access in uplink UAV networks: Effective secrecy throughput maximization under limited feedback channel," *IEEE Trans. Veh. Technol.*, vol. 72, no. 7, pp. 9267–9280, Jul. 2023.

[30] S. Lin, Y. Xu, H. Wang, and G. Ding, "Multi-antenna covert communication assisted by UAV-RIS with imperfect CSI," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 13 841–13 855, Jun. 2024.

[31] Z. Yang, M. Chen, W. Saad, and M. Shikh-Bahaei, "Optimization of rate allocation and power control for rate splitting multiple access (RSMA)," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 5988–6002, Sep. 2021.

[32] M. Dai, B. Clerckx, D. Gesbert, and G. Caire, "A rate splitting strategy for massive MIMO with imperfect CSIT," *IEEE Trans. Wireless Commun.*, vol. 15, no. 7, pp. 4611–4624, Jul. 2016.

[33] B. Clerckx, H. Joudeh, C. Hao, M. Dai, and B. Rassouli, "Rate splitting for MIMO wireless networks: A promising PHY-layer strategy for LTE evolution," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 98–105, May. 2016.

[34] H. Joudeh and B. Clerckx, "Robust transmission in downlink multiuser MISO systems: A rate-splitting approach," *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6227–6242, Dec. 2016.

[35] A. D. Wyner, "The wire-tap channel," *Bell Syst. Tech. J.*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.

[36] R. K. Jain, D.-M. W. Chiu, W. R. Hawe *et al.*, "A quantitative measure of fairness and discrimination," *Eastern Res. Lab., Digit. Equip. Corporat., Hudson, MA, USA*, vol. 21, p. 1, 1984.

[37] F. Meng, P. Chen, L. Wu, and J. Cheng, "Power allocation in multi-user cellular networks: Deep reinforcement learning approaches," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6255–6267, Oct. 2020.

[38] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statistics.*, Apr. 2011, pp. 315–323.

[39] M. Hausknecht and P. Stone, "Deep Recurrent Q-Learning for Partially Observable MDPs," in *Proc. AAAI Fall Symp. Ser.*, 2015, pp. 29–37.

[40] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.

[41] A. Salem and L. Musavian, "NOMA in cooperative communication systems with energy-harvesting nodes and wireless secure transmission," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 1023–1037, Oct. 2020.

[42] R. E. Wang, M. Everett, and J. P. How, "R-MADDPG for partially observable environments and limited communication," 2020, arXiv:2002.06684.

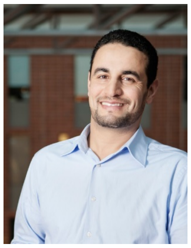
Lijie Zheng received the BS degree from Shandong University of Science and Technology, China, in 2023. He is currently pursuing his PhD degree at the School of Computer Science and Technology, Xidian University, China. His research interests include multi-agent reinforcement learning and physical layer security.

Ji He received the B.S. and M.S. degrees from Xidian University, Xi'an, Shaanxi, China, in 2014 and 2018, respectively, and the Ph.D. degree from Future University Hakodate, Hakodate, Hokkaido, Japan, in 2020. He is currently an associate with the School of Computer Science and Technology, Xidian University. His research interests lie in the field of wireless network security, AI-based cyber-security. He has published over 30 technical papers at premium international journals and conferences, such as IEEE TIFS, IEEE TDSC, IEEE TWC, IEEE TCOM. He is the (co-) recipient of the 2021 IEEE Sapporo Section Based Award and 2020 IEEE Sapporo Section Encouragement Award.

Xinghui Zhu received the Ph.D. degree in computer science and technology from Xidian University, Xi'an, Shaanxi, China, in 2023. He is a Lecturer with the School of Computer Science and Technology at Xidian University. His research interests include data security and IoT security.

Yuanyu Zhang received the B.S. and M.S. degrees from Xidian University, Xi'an, China, in 2011 and 2014, respectively, and the Ph.D. degree from Future University Hakodate, Hakodate, Hokkaido, Japan, in 2017. He is currently an Associate Professor with the School of Computer Science and Technology, Xidian University, Xi'an, Shaanxi, China. Prior to joining Xidian University, he was an Assistant Professor with the Graduate School of Science and Technology, Nara Institute of Science and Technology, Ikoma, Japan. His research interests include Satellite Internet Security, physical layer authentication, physical layer security of wireless communications.

Yulong Shen received the B.S. and M.S. degrees in computer science and Ph.D. degree in cryptography from Xidian University, Xi'an, China, in 2002, 2005, and 2008, respectively. He is currently a Professor with the School of Computer Science and Technology, Xidian University, where he is also an Associate Director of the Shaanxi Key Laboratory of Network and System Security and a member of the State Key Laboratory of Integrated Services Networks. His research interests include wireless network security and cloud computing security. He has also served on the technical program committees of several international conferences, including ICEBE, INCoS, CIS, and SOWN.



Tarik Taleb received the BE degree in information engineering with distinction and the MSc and PhD degrees in information sciences from Tohoku University, Sendai, in 2001, 2003, and 2005, respectively. He is currently a full professor with the Faculty of Electrical Engineering and Information Technology, Ruhr University Bochum. He is the founder and director of the MOSA!C Lab. Between 2018 and 2023, he was a full professor with the Center for Wireless Communications, University of Oulu, Oulu, Finland. Between 2014 and 2021, he was a professor with the School of Electrical Engineering, Aalto University, Finland. Prior to that, he was working as senior researcher and 3GPP standards expert with NEC Europe Ltd., Heidelberg, Germany. Before joining NEC and till March 2009, he worked as a assistant professor with the Graduate School of Information Sciences, Tohoku University, Japan, in a lab fully funded by KDDI, the second largest mobile operator, in Japan. From October 2005 till March 2006, he worked as research fellow with the Intelligent Cosmos Research Institute, Sendai. His research interests lie in the field of telco cloud, network softwarization and network slicing, AI-based software defined security, immersive communications, mobile multimedia streaming, and next generation mobile networking. He has been also directly engaged in the development and standardization of the Evolved Packet System as a member of 3GPP's System Architecture working group 2. He served as the general chair of the 2019 edition of the IEEE Wireless Communications and Networking Conference (WCNC'19) held in Marrakech, Morocco. He was the guest editor in chief of the IEEE JSAC Series on Network Softwarization and Enablers. He was on the editorial board of IEEE Transactions on Wireless Communications, IEEE Wireless Communications Magazine, IEEE Journal on Internet of Things, IEEE Transactions on Vehicular Technology, IEEE Communications Surveys and Tutorials, and a number of Wiley journals. Till December 2016, he served as chair of the Wireless Communications Technical Committee