RIS-assisted SATINs with RSMA and DRL: A Trade-off between Spectral, Secrecy, and Energy Efficiency

Min Wu, Kefeng Guo, Xingwang Li, *Senior Member, IEEE*, Zhi Lin, Liang Yang, Theodoros A. Tsiftsis, *Senior Member, IEEE*, and Chau Yuen, *Fellow, IEEE*

***Abstract*—Given the rapid growth of diverse communication demands, future large-scale satellite-aerial-terrestrial integrated networks (SATINs) need to simultaneously provide services to users while guaranteeing spectral efficiency, secrecy and energy efficiency. This paper addresses the problem of maximising secrecy energy efficiency (SEE) in SATINs, which can accurately describe the effective trade-off between security, spectral efficiency and transmit power. Particularly, we investigate a secure beamforming (BF) scheme in cognitive SATINs that employs rate-splitting multiple access (RSMA) and reconfigurable intelligent surface (RIS) in the presence of multiple eavesdroppers (Eves) in a UAV-aided secondary network (SN). To optimize the SEE for secondary vehicle users while satisfying the constraints of primary users (PUs), we utilize deep reinforcement learning (DRL) to address the coupling between different optimized parameters based on the improved long short-term memory proximal policy optimization (LSTM-PPO) algorithm. The main innovation of this paper is to design a sophisticated reward function, action space, and state space according to each constraint to speed up the convergence. In addition, simulation results show that the proposed DRL-based optimization scheme exhibits significant advantages in terms of SEE compared with benchmark schemes, validating the effectiveness of this work.**

***Index Terms*—Satellite-aerial-terrestrial integrated networks (SATINs), reconfigurable intelligent surface (RIS), rate splitting multiple access (RSMA), deep reinforcement learning (DRL), secrecy energy efficiency (SEE).**

I. INTRODUCTION

This work was supported by the National Science Foundation of China under Grants 62001517. (*Corresponding author: Kefeng Guo*).

M. Wu is with the Institute of Artificial Intelligence, Beihang University, Beijing 100191, China, and also with the Bejing Advanced Innovation Center for Future Blockchain and Privacy Computing, Beihang University, Beijing, 100191, China (e-mail: 1800022837@pku.edu.cn).

K. Guo is with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China, also with the School of Space Information, Space Engineering University, Beijing 101407, China (e-mail: guokefeng.cool@163.com).

X. Li is with the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo, 454000, China, and also with the Jiaozuo Key Laboratory of Crow-Sensing Network, Jiaozuo 454003, China (e-mail: lixingwangbupt@gmail.com).

Z. Lin is with the College of Electronic Engineering, National University of Defense Technology, and also with Anhui Province Key Laboratory of Electronic Restriction, Hefei 230037, China (e-mail: linzhi945@163.com).

L. Yang is with the College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China (e-mail: liangy@hnu.edu.cn).

T. A. Tsiftsis is with the Department of Electrical and Electronic Engineering, University of Nottingham Ningbo China, Ningbo 315100, China, and also with the Department of Informatics and Telecommunications, University of Thessaly, Lamia 35100, Greece (e-mail: tsiftsis@uth.gr).

C. Yuen is with the School of Electrical and Electronics Engineering, Nanyang Technological University, 50 Nanyang Ave, Singapore 639798, Singapore (e-mail: chau.yuen@ntu.edu.sg).

TO further promote the development of the sixth-generation (6G) vision of low-latency connectivity, wide coverage and energy efficiency utilization, satellite-aerial-terrestrial integrated network (SATIN) has been recognized as a promising infrastructure to accelerate this goal. It involves the utilization of cognitive radio (CR) technology to enable the efficient sharing of limited frequency resources between the primary satellite network and the secondary unmanned aerial vehicle (UAV) network under specific constraints [1], [2]. Yuanzhi He et al. of [3] systematically analyzed the technical routes and challenges of Direct-to-Smartphone technology for 6G NTN, and proposed innovative solutions like high-dynamic air interface systems and ultra-large array antennas, providing important reference for the industry, highlighting their deep insight and forward-looking vision in satellite communication.

Despite the enhanced capabilities provided by SATIN, there are several significant challenges that still hinder its broader adoption. The challenge of limiting inevitable eavesdroppers (Eves) from intercept private signals, while ensuring the quality of service (QoS) of the primary network (PN), is a pressing task for the practical deployment of cognitive SATINs. As a complement to encryption methods, physical layer security (PLS) has received extensive investigation in traditional terrestrial networks [4]. However, the inherent broadcast nature of SATIN creates new challenges for the implementation of PLS, with all ground users being able to receive satellite signals, increasing the risk of illegal eavesdropping.

At the same time, the emergence of reconfigurable intelligent surfaces (RISs) provides a reliable paradigm for enhancing the PLS of SATINs by intelligently reconfiguring wireless propagation environment. In particular, by manipulating the phase shift of reflective elements, RIS can facilitate the constructive superimposition of direct and reflected signals at intended legitimate users, while inducing destructive interference for potential eavesdroppers [5]–[7]. Meanwhile, parallel developments include the rate-splitting multiple access (RSMA) scheme, which promises to mitigate multi-user interference by acting as a bridge between traditional multi-user linear precoding (MU-LP) and power-domain non-orthogonal multiple access (NOMA) [8]. By strategically separating the information into common and private parts, RSMA provides a more flexible approach to interference management. This capability not only enhances secrecy energy efficiency (SEE), but also significantly improves user fairness and system robustness under unpredictable network conditions.

However, while the use of RSMA and RIS technologies within the cognitive SATINs improves spectral efficiency (SE) and transmission security, it is important to recognize that there are complex trade-offs between spectral efficiency, secrecy and energy efficiency. For example, improving spectral efficiency may require more sophisticated signal processing techniques, which may increase energy consumption and reduce secrecy. While enhancing secrecy may require additional resource allocation, thus degrading spectral efficiency and energy efficiency. A preliminary exploration of these trade-offs is presented in [9]–[11], but an in-depth analysis for SATIN scenarios is still lacking.

To further exploit the advantages offered by RSMA and RIS techniques within SATINs, and taking into account the imperfect availability of eavesdropping channel state information (CSI), our study focuses on maximizing secondary network (SN) SEE while satisfying the primary user (PU) secrecy rate constraints. In view of the high dynamics of RIS-assisted cognitive SATIN with RSMA, traditional optimization algorithms require accurate mathematical models, which is very difficult for the proposed wireless communication scenario [12]–[14]. Coinciding with this, inspired by the recent popularity of model-free artificial intelligence (AI) frameworks, there is a growing trend to leverage these techniques such as deep learning (DL), reinforcement learning (RL) and deep reinforcement learning (DRL), to open wireless communication optimization solutions. In particular, DRL has shown significant advantages in dealing with complex wireless environments by dynamically interacting with the environment to process large amounts of communication data and optimizing long-term cumulative rewards, thereby fostering the development of robust communication strategies [15], [16]. This study aims to deeply explore the trade-off between spectral, secrecy, and energy efficiency in SATIN based on the DRL approach, providing strong support for the innovation of communication systems in the 6G era.

A. Related Works

Nowadays, various techniques have been applied to improve the spectral, secrecy, and energy efficiency of SATINs. Related works are discussed as follows:

1) RSMA-enabled SATINs: As a solution to the challenge of limited spectrum resources, the RSMA scheme has gained widespread adoption in SATINs due to its effectiveness in handling multi-user interference scenarios [17]. A key feature of RSMA in this context is the implementation of the power splitting (PS) receiver architecture, which divides the RF signal to be sent into common and private streams. For instance, the authors in [18] demonstrated that RSMA outperforms NOMA in terms of spectral efficiency (SE) in SATINs with multiple eavesdroppers. Additionally, in [19], the authors focused on maximizing the weighted sum data rate in the RSMA-based tethered UAV system, considering factors such as UAV altitude, RIS size, and positioning. Furthermore, in [20], a secure beamforming scheme for secondary networks was proposed in RSMA-enabled cognitive satellite terrestrial networks (CSTNs) to further reduce interference from eavesdropping users to legitimate users. Lastly, the authors in [21] looked into four non-orthogonal broadcast and unicast (NOBU) transmission schemes to find the best max-min rate for RSMA-enabled SATINs. After comparing the proposed benchmark schemes, the RSMA-based NOBU scheme was shown to have significant performance gains.

2) RIS-aided SATINs: With the rapid advancements in meta-surface technology, there is growing interest in the performance of RIS-assisted SATINs, leading to a plethora of research outcomes [22], [23]. The authors in [24] proposed an alternating optimization scheme to jointly transmit beamforming and RIS reflecting coefficients by taking the transmit power constraint and the interference threshold into account. In [25], the authors derived a theoretical expression of outage probability for all secondary vehicle users in RIS-aided SATINs. The authors of [26] also thought about using a double cascade correlation network (DCCN) to improve the UAV trajectory and RIS reflection coefficients, despite challenges posed by outdated channel state information (CSI) and interference from neighboring satellites. In [27], the authors proposed to mount the RIS on the UAV to facilitate the transmission effect from the satellite to the terrestrial users. The simulation results also confirm that the solution of installing the RIS in the UAV, due to the fact that the UAV can change its position rapidly, combined with the RIS ability to dynamically regulate the radio environment, the system can quickly adapt to different communication requirements and environmental changes, achieving on-demand coverage and optimization. More surprisingly, RIS not only enhances the desired signal but also mitigates Eve signals by adjusting its phase, making it a promising technology for enhancing PLS in SATINs. For example, the authors in [28] investigated the performance of RIS-aided hybrid automatic repeat request (HARQ) SATINs, pairing security-required users with QoS-sensitive users to implement the NOMA technique.

3) DRL-empowered SATINs: The DRL-empowered framework has demonstrated impressive performance across various wireless network systems, particularly in highly dynamic environments such as RIS-aided SATINs. The authors in [29] utilized historical CSI data and post-decision state to accelerate learning rate in RIS-aided SATINs based on multi-agent DRL algorithm. In [30], the authors introduced the proximal policy optimization (PPO) based on DRL to strike a balance between task execution latency and energy consumption within satellite-terrestrial integrated networks (STINs). Furthermore, in [31], the ergodic capacity was predicted by deep neural networks (DNNs), which was based on the assumption of imperfect CSI and uncertain line-of-sight (LoS) links in RIS-aided SATINs with NOMA scheme. In [32], the author investigated a low complexity DRL algorithm for optimizing the transmit precoding matrix, RIS phase shift matrix, and radar transmit waveform, aiming to minimize the impact of expected rewards on the target Q value through iterative factor adjustments. These advancements highlight the potential of DRL in optimizing the performance of SATIN scenarios.

B. Motivations

It is worth noting that current research on realizing physical layer security for large-scale SATINs predominantly focuses on the contribution of individual technology, such as employing RSMA to enhance spectral efficiency [33], [34], integrating RIS to strengthen PLS [35], [36], or leveraging cognitive radio techniques to facilitate flexible access. However, these studies often operate within isolated frameworks, neglecting the potential synergies among these advanced technologies. A unified optimization framework that harmoniously integrates RSMA, RIS, and cognitive radio is notably absent, despite the critical importance of both spectral utilization efficiency and PLS in the development of next-generation SATINs. The simultaneous introduction of these technologies could significantly elevate system performance at different levels. By applying RSMA strategies, spectrum and power allocation in multi-user scenarios can be optimized [37]. On the other hand, RIS enhances signal quality and mitigates interference, and cognitive radio ensures adaptability to dynamic electromagnetic environments, thereby maintaining robust communication performance. The convergence of these technologies presents a promising avenue for improving spectral efficiency at multiple levels and unlocking new application scenarios. Nevertheless, this integration also poses substantial challenges, particularly in the context of timing-related environmental states, which can lead to frequent shifts in the training phases of DRL-based mechanisms [38].

To further address these challenges and advance the development of secure communication strategies for SATINs, we propose the integration of LSTM networks into DRL frameworks. By incorporating LSTM modules, the DRL agent can consider long-term optimization goals when refining parameters, thereby mitigating the shortcomings associated with rapidly changing environmental conditions. Our research aims to bridge this gap by developing a comprehensive and adaptable framework that leverages the combined strengths of RSMA, RIS, and cognitive radio technology to maximize the secrecy-energy efficiency, while addressing the inherent challenges posed by dynamic SATIN environments.

C. Contributions

The main work of this paper can be summarized as follows

- We propose a new approach to integrate RSMA and RIS into a cognitive SATIN involving the presence of multiple Eves. In particular, we implement the RSMA scheme on auxiliary drones to enhance the SEE performance of SNs. RIS constructively enhances the signals reflected to legitimate users and destructively attenuates the signals leaked to Eves. To address dynamic eavesdropping behaviour, we adaptively optimise the RIS phase shift to suppress malicious intercepts, even in the presence of sudden changes in eve position.
- We establish an optimization problem focusing on enhancing SEE. This optimization involves a comprehensive design strategy that includes the transmit beamforming of secondary UAV, RIS reflecting phase shift matrix, and transmit power splitting ratio of secondary UAV, which needs to be guaranteed to be carried out under the QoS of the PNs as well as the common information rate constraints in the SNs. The problem explicitly incorporates robustness-oriented constraints, including, imperfect CSI with bounded estimation error and worst-case eavesdropping scenarios.
- The above optimization problem is a non-convex optimization problem characterized by non-linear constraints, finite phase shifts and coupled parameter variables. We then transform the optimization problem into an Markov Decision Process (MDP). Subsequently, we develop a unified PPO-empowered learning framework using DRL. Unlike existing learning frameworks, we introduce an LSTM mechanism for state feature extraction, which captures temporal correlations in dynamic environments to improve the robustness of the policy. The PPO agent is trained using normalised and feature-fused state information to adaptively balance confidentiality and energy efficiency under uncertainty.
- Finally, simulation results show that our proposed strategy is more adaptive to dynamic environments compared to several benchmark approaches. We also evaluate the impact of the maximum transmit power of the secondary SEE. Furthermore, our unified DRL approach outperforms other benchmark schemes in terms of secrecy efficiency and computation efficiency.

TABLE I
SYMBOL DEFINITION

Symbol	Meaning
$\mathbf{h}_{U,n}$	Channel gain coefficient from UAV to n-th VU
$\mathbf{h}_{U,m}$	Channel gain coefficient from UAV to m-th Eve
$\mathbf{h}_{R,n}$	Channel gain coefficient from RIS to n-th VU
$\mathbf{h}_{R,m}$	Channel gain coefficient from RIS to m-th Eve
$\mathbf{h}_{RG}$	Channel gain coefficient from RIS to GES
$\mathbf{H}_{UR}$	Channel gain coefficient from UAV to RIS
$\mathbf{g}$	Channel gain coefficient from satellite to GES
$\mathbf{g}_{c,n}$	Channel gain coefficient from satellite to n-th VU
$\mathbf{g}_{e,m}$	Channel gain coefficient from satellite to m-th Eve
$\mathbf{v}_c$	Secondary UAV transmit beamforming vector for common information
$\mathbf{v}_n$	Secondary UAV transmit beamforming vector for private information to n-th VU
$\mathbf{w}$	Primary satellite transmit beamforming vector
$\mathbf{\Phi}$	Passive beamforming matrix at RIS
P_S	Total transmit power at satellite
P_U	Total transmit power at UAV

The rest of this paper is organized as follows. In Section II, we describe the considered system model, including the optimization problem and corresponding constraints. In Section III, we transform the SEE optimization problem into an MDP, and focus on solving the problem using DRL. Section IV provides various benchmarks and simulation results. Finally, conclusions are given in Section V.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We explore the SEE maximization problem of secondary users in RIS-aided cognitive SATINs, considering the presence of a total of M Eves [39], [40]. In this setup, the whole integrated satellite-UAV-terrestrial network consists of two

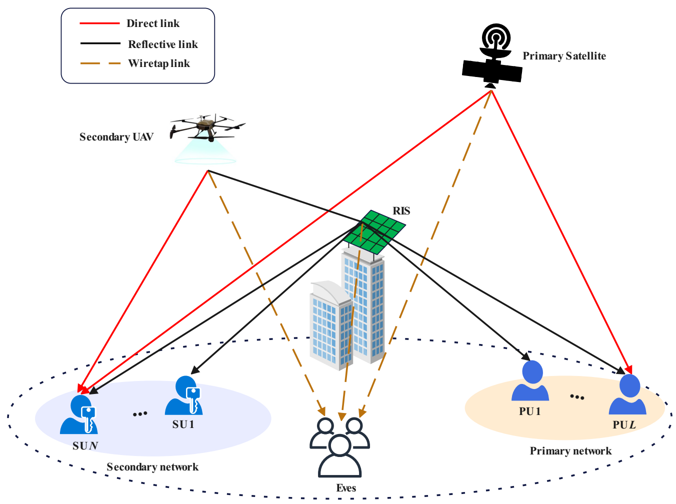

Fig. 1. Description of the proposed system model.

sub-networks, where the satellite sends multicast signal to a total of L GESs and shares spectrum with UAV under overlay mode, which is denoted as the PN, while the UAV adopts the RSMA to serve N vehicle users with the aid of RIS, which is denoted as SN. Here, we can assume that the satellite and UAV are outfitted with array fed reflector antennas comprising N_S feeds and a uniform linear array (ULA) consisting of N_U antennas, respectively [41]. [1] To represent symbols more concisely, some key system parameters are defined in Table I.

A. Channel Model Description

Considering the effects of path loss, rain attenuation, and satellite transmit beam gains, the satellite downlink channel model can be described as

$$\mathbf{g} = \sqrt{G_{s,t} G_{s,r} C_P} \odot \mathbf{B}^{\frac{1}{2}} \odot \varsigma^{-\frac{1}{2}} \odot \tilde{\mathbf{g}}, \tag{1}$$

where $G_{s,t}$ denotes the satellite transmit antenna gain [44]. In addition, $G_{s,r}$ is the parabolic antenna gain of the satellite serving users [45]. Besides, the symbol $\odot$ denotes the Hadamard Product. Besides, $C_P = (\lambda_w / 4\pi d_S)^2$ denotes the free-space loss between the satellite and terrestrial user with λ_w being the operating wavelength, and d_S being the transmission distance between the satellite and the signal receiver. $\mathbf{B}$ is the beam gain matrix of the transmitting antenna under satellite communication far-field characteristics, ς represents the vector of rain attenuation coefficients. $\tilde{\mathbf{g}} = [\tilde{g}_1, \tilde{g}_2, ..., \tilde{g}_{N_S}]^T$ denotes the composition of the phase differences of the arriving different beams from the satellite to the user, and its elements can be denoted as $\tilde{g}_{N_s} = e^{-j2\pi d_{N_s}/\lambda_w}$, where d_{N_s} denotes the distance from the terrestrial user to the N_s-th satellite feeder.

According to [46], it is assumed that $\mathbf{h}_{U,n} \in C^{N_U \times 1}$, $\mathbf{h}_{U,m} \in C^{N_U \times 1}$, $\mathbf{h}_{R,n} \in C^{K \times 1}$, $\mathbf{h}_{R,m} \in C^{K \times 1}$, $\mathbf{h}_{RG} \in C^{K \times 1}$, $\mathbf{H}_{UR} \in C^{N_U \times K}$ and $\mathbf{H}_{UG} \in C^{N_U \times K}$ are the channel gains from the UAV to the n-th VU, from the UAV to the m-th Eve, from the RIS to the n-th VU, from the RIS to the m-th Eve, from the RIS to the GES, from the UAV to the RIS, from the UAV to the GES, respectively. In the millimeter-wave (mmWave) communication transmission environment, all the above channels can be modelled according to the 3D Salen-Valenzula (SV) channel [47]. Here, we give the modelling process as an example of n-th VU in the SN, and other channels $\mathbf{h}_{U,m} \in C^{N_U \times 1}$, $\mathbf{h}_{R,m} \in C^{K \times 1}$ and $\mathbf{H}_{UG} \in C^{N_U \times K}$ can refer to this content. It is assumed that there are $L_{U,n}$, $L_{R,n}$ and $L_{U,R}$ transmission paths that characterize the UAV-VU link, the RIS-VU link and the UAV-RIS link, respectively. Then, we have

$$\mathbf{h}_{U,n} = \sqrt{\frac{1}{L_{U,n}}} \sum_{l=1}^{L_{U,n}} g_{n,l}^{U} \mathbf{a}_L \left(\omega_{n,l}^{AoD}\right), \tag{2a}$$

$$\mathbf{h}_{R,n} = \sqrt{\frac{1}{L_{R,n}}} \sum_{l=1}^{L_{R,n}} g_{n,l}^{R} \mathbf{a}_P \left(\omega_{n,l}^{AoD}, \vartheta_{n,l}^{AoD}\right), \tag{2b}$$

$$\mathbf{H}_{UR} = \sqrt{\frac{1}{L_{U,R}}} \sum_{l=1}^{L_{U,R}} g_{l}^{UR} \mathbf{a}_P \left(\omega_{n,l}^{AoA}, \vartheta_{n,l}^{AoA}\right) \mathbf{a}_L \left(\omega_{n,l}^{AoD}\right)^H, \tag{2c}$$

where the large-scale fading coefficients denoted by $g \in \left\{g_{n,l}^{U}, g_{n,l}^{R}, g_{l}^{UR}\right\}$ follow a complex Gaussian distribution as $\mathcal{CN}\left(0, 10^{\frac{PL}{10}}\right)$ with $PL\,[dB] = -C_0 - 10\alpha \lg D - \gamma_p$. Furthermore, C_0 denotes the path loss at a reference distance, α is the path-loss exponent, D represents the transmission link distance, and $\gamma_p \sim \mathcal{CN}\left(0, \sigma_{\gamma_p}^2\right)$ stands for the shadow attenuation component. Besides, $\mathbf{a}_L\left(\omega\right)$ and $\mathbf{a}_P\left(\omega, \vartheta\right)$ are the steering vectors with the type antenna about uniform linear array (ULA) and uniform planar array (UPA).

B. Signal Model Description

As mentioned above, the signal model can be divided into two main parts, the satellite transmit signals in PN and the UAV transmits signals in SN. In the transmission phase of the PN from the satellite to the GES as PUs, the private multicast signal x is mapped to the satellite transmit beamforming precoding matrix $\mathbf{w} \in C^{N_S \times 1}$ under the eavesdropping of Eves [48]. During the signaling process in SN, the UAV can be described as utilizing its flexibility to serve N single-antenna VUs in SN with the aid of RIS, while the RIS equipped with $K = K_y \times K_z$ reflective element to enhance the expected signals at the VUs while suppressing the signals received by the Eves. Specifically, the RIS reflecting phase shift matrix is given by $\mathbf{\Phi} = \mathrm{diag}\left\{\beta_1 e^{j\theta_1}, \beta_2 e^{j\theta_2}, ..., \beta_K e^{j\theta_K}\right\}$, where β_k and θ_k are the amplitude factor and phase shift at k-th RIS element [49].

By installing the RIS in the SN can enhance SN performance while reducing interference to the PN. Additionally, by identifying the ideal RIS phase-shift matrix, we can boost the intended user received signal while suppressing the Eves in the SN. At the same time, applying the RSMA technology to the UAV, the unicast signals s_n designated for the n-th VU gets divided into a common sub-signal s_c and a private sub-signal $s_{p,n}$. Utilizing a shared codebook among all VUs, the

[1] This paper focuses on fixed UAV as secondary network transmitter (e.g., UAVs as static relay nodes in emergency communications), and its core innovation lies in solving the spectrum sharing optimization problem under fixed deployment (e.g., static low altitude platforms as defined in 3GPP TR 36.777) [42], [43], where the UAV are powered via tethered power supply or regular rotation to achieve long-term stable service, dynamic path planning and energy consumption are not major constraints.

common sub-signals are collectively encoded into a common signal stream s_c. The same frequency band is permitted to be shared by both satellites and UAVs within a specified interference tolerance in the GES of the PN. Notably, a total of M Eves can also intercept signals transmitted by satellites and UAVs in a similar manner as legitimate VUs [50]. Furthermore, the transmit beamforming vectors $\mathbf{v}_c \in \mathbb{C}^{N_U \times 1}$ and $\mathbf{v}_n \in \mathbb{C}^{N_U \times 1}$, corresponding to the common signal s_c and private sub-signal $s_{p,n}$, respectively, are applied to the UAV. These vectors are pivotal in steering the signals toward their intended receivers while minimizing leakage to unintended recipients. The formulation of the received signals at various nodes, including the GES, the m-th Eve and the n-th VU, leverages these beamforming strategies. This approach ensures that the common signal and each private sub-signal are optimized for reception at the intended receiver while concurrently reducing the risk of interception by adversaries. In concrete terms, the received signals at the GES, the m-th Eve and the n-th VU can be expressed as

$$\begin{aligned} y_G =& \mathbf{g}^H \mathbf{w} x + \left(\mathbf{H}_{UG} + \mathbf{h}_{RG}^H \mathbf{\Phi} \mathbf{H}_{UR}\right) \mathbf{v}_c s_c \\ &+ \sum_{n=1}^{N} \left(\mathbf{H}_{UG} + \mathbf{h}_{RG}^H \mathbf{\Phi} \mathbf{H}_{UR}\right) \mathbf{v}_n s_{p,n} + n_G, \end{aligned} \tag{3}$$

$$\begin{aligned} y_{s,m} =& \mathbf{g}_{e,m}^H \mathbf{w} x + \left(\mathbf{h}_{U,m} + \mathbf{h}_{R,m}^H \mathbf{\Phi} \mathbf{H}_{UR}\right) \mathbf{v}_c s_c \\ &+ \sum_{n=1}^{N} \left(\mathbf{h}_{U,m} + \mathbf{h}_{R,m}^H \mathbf{\Phi} \mathbf{H}_{UR}\right) \mathbf{v}_n s_{p,n} + n_{e,m}, \end{aligned} \tag{4}$$

$$\begin{aligned} y_{c,n} =& \mathbf{g}_{c,n}^H \mathbf{w} x + \left(\mathbf{h}_{U,n} + \mathbf{h}_{R,n}^H \mathbf{\Phi} \mathbf{H}_{UR}\right) \mathbf{v}_c s_c \\ &+ \sum_{n=1}^{N} \left(\mathbf{h}_{U,n} + \mathbf{h}_{R,n}^H \mathbf{\Phi} \mathbf{H}_{UR}\right) \mathbf{v}_n s_{p,n} + n_{c,n}, \end{aligned} \tag{5}$$

where $n_G, n_{e,m}$ and $n_{c,n}$ represent independent identically complex Gaussian distribution with noise variance $\sigma_G^2, \sigma_{e,m}^2$ and $\sigma_{c,n}^2$, respectively. To streamline further, the equivalent channel from the UAV to the GES, the m-th Eve, the n-th VU after RIS reflection can be defined as

$$\mathbf{z}_G = \left(\mathbf{H}_{UG} + \mathbf{h}_{RG}^H \mathbf{\Phi} \mathbf{H}_{UR}\right)^H, \tag{6a}$$

$$\mathbf{z}_{e,m} = \left(\mathbf{h}_{U,m} + \mathbf{h}_{R,m}^H \mathbf{\Phi} \mathbf{H}_{UR}\right)^H, \tag{6b}$$

$$\mathbf{z}_{c,n} = \left(\mathbf{h}_{U,n} + \mathbf{h}_{R,n}^H \mathbf{\Phi} \mathbf{H}_{UR}\right)^H. \tag{6c}$$

Without loss of generality, the Eq. (3)-(5) can be converted correspondingly to the following

$$y_G = \mathbf{g}^H \mathbf{w} x + \mathbf{z}_G^H \mathbf{v}_c s_c + \sum_{n=1}^{N} \mathbf{z}_G^H \mathbf{v}_n s_{p,n} + n_G, \tag{7}$$

$$y_{s,m} = \mathbf{g}_{e,m}^H \mathbf{w} x + \mathbf{z}_{e,m}^H \mathbf{v}_c s_c + \sum_{n=1}^{N} \mathbf{z}_{e,m}^H \mathbf{v}_n s_{p,n} + n_{e,m}, \tag{8}$$

$$y_{c,n} = \mathbf{g}_{c,n}^H \mathbf{w} x + \mathbf{z}_{c,n}^H \mathbf{v}_c s_c + \sum_{n=1}^{N} \mathbf{z}_{c,n}^H \mathbf{v}_n s_{p,n} + n_{c,n}. \tag{9}$$

When the service of SN is provided by the UAV, the n-th VU first decodes the common signal s_c that the UAV sends out, disregarding the private sub-signals s_p as unwanted noise. After successfully decoding s_c, the SIC technique is applied. This process involves re-encoding and pre-coding s_c, followed by subtracting it from the received signal. This subtraction significantly reduces the interference caused by s_c, allowing the n-th VU to effectively decode its designated private sub-signal, denoted as $s_{p,n}$, while still treating other VUs private sub-signals as interference. Specifically, the rate of information transfer in s_c that the m-th Eve can potentially intercept is critical to understanding the security and efficacy of the communication strategy. Similarly, the ability of the n-th VU to decode $s_{p,n}$ in the remaining private sub-signals interferences is pivotal for assessing the robustness and reliability of the communication network. Therefore, the achievable rate of the GES, the m-th Eve, the common signal s_c and private sub-signal stream of the n-th VU can be expressed as

$$R_G = \log_2 \left(1 + \frac{\left|\mathbf{g}^H \mathbf{w}\right|^2}{\left|\mathbf{z}_G^H \mathbf{v}_c\right|^2 + \sum_{n=1}^{N} \left|\mathbf{z}_G^H \mathbf{v}_n\right|^2 + \sigma_G^2}\right), \tag{10a}$$

$$R_{s,m} = \log_2 \left(1 + \frac{\left|\mathbf{g}_{e,m}^H \mathbf{w}\right|^2}{\left|\mathbf{z}_{e,m}^H \mathbf{v}_c\right|^2 + \sum_{n=1}^{N} \left|\mathbf{z}_{e,m}^H \mathbf{v}_n\right|^2 + \sigma_{e,m}^2}\right), \tag{10b}$$

$$R_{c,n} = \log_2 \left(1 + \frac{\left|\mathbf{z}_{c,n}^H \mathbf{v}_c\right|^2}{\left|\mathbf{g}_{c,n}^H \mathbf{w}\right|^2 + \sum_{n=1}^{N} \left|\mathbf{z}_{c,n}^H \mathbf{v}_n\right|^2 + \sigma_{c,n}^2}\right), \tag{10c}$$

$$R_{p,n} = \log_2 \left(1 + \frac{\left|\mathbf{z}_{c,n}^H \mathbf{v}_n\right|^2}{\left|\mathbf{g}_{c,n}^H \mathbf{w}\right|^2 + \sum_{i \neq n}^{N} \left|\mathbf{z}_{c,n}^H \mathbf{v}_i\right|^2 + \sigma_{c,n}^2}\right). \tag{10d}$$

In fact, the common signal interference from secondary UAV transmitters can be regarded as a useful resource, i.e., beneficial interference, and its effective utilization then improves the transmission security of the primary network. To this end, the secure transmission problem in RSMA-based RIS-assisted cognitive SATINs is investigated and a joint secure beamforming design mechanism for the primary and secondary networks is proposed, which makes full use of the interference generated by the secondary UAV to suppress the potential eavesdroppers and enhance the security without having to additionally sacrifice the spectral efficiency, in order to take into account the efficiency. On this basis, we define $\chi_n \in (0,1)$ as the common-private rate-splitting ratio for the n-th VU, where the part of $\sqrt{\chi_n}$ is the received signal power used for transmitting the common information, while the remaining part $\sqrt{1-\chi_n}$ is used for transmitting the private signals [51], [52]. The above achievable rate of the common and private stream of the i-th VU can be rewritten as

$$R_c = \log_2 \left(1 + \frac{\chi_i \left|\mathbf{z}_{c,i}^H \mathbf{v}_c\right|^2}{\chi_i \sum_{n=1}^{N} \left|\mathbf{z}_{c,i}^H \mathbf{v}_n\right|^2 + \left|\mathbf{g}_{c,i}^H \mathbf{w}\right|^2 + \sigma_{c,i}^2}\right), \tag{11}$$

and

$$R_{p,i} = \log_2 \left(1 + \frac{\chi_i \left|\mathbf{z}_{c,i}^H \mathbf{v}_i\right|^2}{\chi_i \sum_{n=1, n \neq i}^{N} \left|\mathbf{z}_{c,i}^H \mathbf{v}_n\right|^2 + \left|\mathbf{g}_{c,i}^H \mathbf{w}\right|^2 + \sigma_{c,i}^2}\right). \tag{12}$$

On the other hand, it is also imperative to ascertain that all common information within the SN can be decoded successfully. This requirement is fulfilled if the achievable rate of the common information stream, denoted as

$\min_n R_{c,n}\left(\left\{\mathbf{v}_n, \mathbf{\Phi}\right\}, \chi_n\right)$, meets the established criteria. Here, let c_n represent the data rate of the common information received at the n-th VU, which needs to satisfy certain conditions as

$$\sum\nolimits_{n=1}^{N} c_n \leq \min_n R_{c,n}\left(\left\{\mathbf{v}_n, \mathbf{\Phi}\right\}, \chi_n\right). \tag{13}$$

Thus, the sum achievable rate of the UAV-enabled SN can be expressed as

$$R_U\left(\left\{\mathbf{v}_n, \mathbf{\Phi}\right\}, \chi_n\right) = \sum\nolimits_{i=1}^{N} \left(c_i + R_{p,i}\right). \tag{14}$$

C. CSI Estimation Error Model

In practical terms, the acquisition of perfect CSI for satellite downlink and UAV downlink transmissions presents substantial challenges. These challenges primarily stem from scattering, fast fading, and extensive propagation delays encountered during signal transmission, as noted in [53]. Furthermore, after implementing CR-empowered techniques, the PN remains unaware of the existence of the SN. This lack of awareness can further complicate the management and optimization of network resources. In other words, the CSI of the PN link is not perfectly available at the n-th VU because of the inability to exchange information properly between the PN and the SN [54]. Therefore, the worst case bounded CSI estimation error model is considered to estimate the perceived channel matrix in SN, which can be expressed as

$$\mathbf{h}_{U,n} = \tilde{\mathbf{h}}_{U,n} + \mathbf{e}_{U,n}, \eqsim_{U,n} \triangleq \left\{\mathbf{e}_{U,n}^H \mathbf{e}_{U,n} \leq \Delta_{U,n}^2\right\}, \tag{15}$$

$$\mathbf{h}_{R,n} = \tilde{\mathbf{h}}_{R,n} + \mathbf{e}_{R,n}, \eqsim_{R,n} \triangleq \left\{\mathbf{e}_{R,n}^H \mathbf{e}_{R,n} \leq \Delta_{R,n}^2\right\}, \tag{16}$$

where $\tilde{\mathbf{h}}_{U,n}$ and $\tilde{\mathbf{h}}_{R,n}$ denote the estimated CSI at VU, $\mathbf{e}_{U,n}$ and $\mathbf{e}_{R,n}$ represent the channel estimation error [55]. Besides, $\eqsim_{R,n}$ and $\eqsim_{R,n}$ are the uncertainty region limiting the CSI uncertainty, $\Delta_{U,n}^2$ and $\Delta_{R,n}^2$ indicate the radius of the region of uncertainty. Since Eves are not authenticated as legitimate users, it is can be assumed that the channel uncertainties $\mathbf{g}_{e,m}$ and $\mathbf{z}_{e,m}$ conform to the Euclidean norm bounded error models $\mathbf{g}_{e,m} = \tilde{\mathbf{g}}_{e,m} + \mathbf{e}_{g,m}$ and $\mathbf{z}_{e,m} = \tilde{\mathbf{z}}_{e,m} + \mathbf{e}_{z,m}$, where $\tilde{\mathbf{g}}_{e,m}$ and $\tilde{\mathbf{z}}_{e,m}$ represent the estimated CSI at the satellite and the UAV, $\left\|\mathbf{e}_{g,m}\right\| \leq \delta_g$ and $\left\|\mathbf{e}_{z,m}\right\| \leq \delta_z$.

As a result, the worst-case achievable secrecy rate (ASR) of the i-th VU can be expressed as

$$R_{c,i}\left(\left\{\mathbf{v}_i, \mathbf{\Phi}\right\}, \chi_i\right) = R_{c,i} - \max_{\mathbf{e}_{U,i}, \mathbf{e}_{R,n}} \max_{\mathbf{e}_{z,m}, \mathbf{e}_{g,m}} R_{s,m}. \tag{17}$$

From a practical deployment perspective, we only focus on scenarios where the worst-case ASR remains positive. This ensures that even in the least favorable conditions, the data rate for legitimate users still exceeds that of potential eavesdroppers. A negative ASR means that the data rate for Eves is higher than the legitimate user, which is not allowed.

D. Problem Description

In the system model described, the UAV-enabled SN has the capability to utilize the licensed spectrum of the satellite-enabled PN. This arrangement requires the UAV to maintain the interference with the PUs within acceptable levels, as discussed in [56], [57]. The ratio of the SN total achievable data rate to its overall power consumption quantifies the SEE. Our research focuses on enhancing the SEE of the SN, which is a critical factor for promoting secure and friendly communication practices. Working toward this goal, we need to work together to optimize the UAV active secure beamforming vector $\mathbf{v} = [\mathbf{v}_c, \mathbf{v}_1, ..., \mathbf{v}_N]$, where $\mathbf{v_c}$ is the common information stream beamforming vector, $\mathbf{\Phi}$ is the RIS phase shift matrix, and χ_n is the transmit power splitting ratio. The mathematical expression can be defined as

$$\mathrm{SEE} = \frac{R_U - \sum\limits_{m=1}^{M} \max R_{s,m}}{\left(\left\|\mathbf{v}_c\right\|^2 + \sum_{n=1}^{N} \left\|\mathbf{v}_n\right\|^2\right) + P_C}, \tag{18}$$

where P_C denotes the constant circuit power consumption at UAV. Here, we assume that the UAV provides communication services in a hovering state, and its energy consumption can refer to [58]. Mathematically, the whole optimization problem can be formulated as

$$\mathbf{P0}: \max_{\mathbf{v}_n, \mathbf{\Phi}, \chi_n} \max_{\mathbf{e}_{U,n}, \mathbf{e}_{R,n}} \mathrm{SEE}, \tag{19a}$$

$$\text{s.t. } C1: \min_{\mathbf{e}_{g,m}, \mathbf{e}_{z,m}} R_G - R_{s,m} \geq \Delta_U, \forall m, \tag{19b}$$

$$C2: R_{c,n} \geq \Delta_c, \forall n, \tag{19c}$$

$$C3: R_{p,n} \geq \Delta_p, \forall n, \tag{19d}$$

$$C4: \left\|\mathbf{w}\right\|^2 \leq P_S, \left\|\mathbf{v}_c\right\|^2 + \sum\nolimits_{n=1}^{N} \left\|\mathbf{v}_n\right\|^2 \leq P_U, \tag{19e}$$

$$C5: \left|\exp\left(j\theta^k\right)\right| = 1, \forall k, \tag{19f}$$

$$C6: \sum\nolimits_{n=1}^{N} c_n \leq \min_n R_{c,n}\left(\left\{\mathbf{w}_l\right\}, \mathbf{\Phi}, \chi_n\right), \tag{19g}$$

where P_C denotes the constant circuit power consumption. Meanwhile, $C1$ denotes the constraint that the achievable rate of the VUs must be less than the achievable rate of the GES, and $C2$ and $C3$ denote the minimum rate requirement for the transmit signals of SN. And, $C4$ is the transmit power limit for PN and SN, where P_S and P_U are the preset power caps for the satellite and UAV. $C6$ denotes the constraint that the VU of the SN is able to fully decode the common information. In addition, Δ_U denotes the secrecy rate limit for the UAV, whereas Δ_c and Δ_p signify the data rate limit corresponding to the common and private parts of the VUs, respectively.

III. MDP Formulation for the Secrecy-Energy Efficiency Maximization

As modeled earlier, the problem $\mathbf{P0}$ is a high-dimensional problem that is difficult to solve directly, because it contains both discrete and continuous variables. In particular, continuous variables need to be specified in terms of their range of values and the amount of variation. Moreover, the given problem $\mathbf{P0}$ is only an optimization problem under a single time slot, which may converge to a sub-optimal solution and obtain a performance similar to that of greedy search due to the historical state of the neglected environment and long-term gains. Therefore, it is generally infeasible to employ conventional optimization techniques such as alternating optimization (AO), semidefinite program (SDP) or successive convex approximation (SCA) to achieve efficient beamforming strategies in dynamic RIS-aided secure SATINs [59].

In the model-free DRL family, the policy-based learning approach, represented prominently by PPO, solves the problem of difficult step size determination in policy gradient algorithms by using stochastic gradient ascent method to optimize instead of the objective function over, which can achieve small batch updates in multiple training steps. In this section, to model decision problems within a deterministic mathematical framework, we need to reformulate the problem $\mathbf{P0}$ as an MDP by taking into account the randomness of the environment and other incomplete information. Then, we choose a unified PPO-enabled learning approach to solve the problem to cope with the situation where the actions are mostly continuous in real-life situations [60].

The formulated MDP consists of an agent, i.e., the global decision maker, characterized by five essential elements:$\langle \mathcal{S}$, $\mathcal{A}$, $\mathcal{R}$, $\mathcal{T}$, $\gamma \rangle$. The agent can be considered as an au-tonomous entity that interacts with its environment by making a series of decisions. These decisions, or actions, are selected based on the policy, which is learned through the reinforce-ment learning process. The agent then receives feedback in the form of rewards or penalties, depending on how well the outcomes of its actions align with the predefined goals. This iterative process of decision-making and feedback is aimed at optimizing the agent's policy to maximize cumulative long-term rewards [61]. Here, $\mathcal{S}$ represents the state space, encompassing all possible situations the agent might encounter. $\mathcal{A}$ denotes the action space, which includes all potential actions the agent can take at any state. $\mathcal{R}$ is the reward function, which assigns a numerical value to each action taken by the agent in a particular state, guiding the agent's behavior by providing incentives for certain actions. $\mathcal{T}$ represents the transition probability, indicating the likelihood of moving from one state to another given an action. Lastly, γ, the discount factor ranging from 0 to 1, quantifies the present value of future rewards, reflecting the degree to which immediate rewards are prioritized over distant ones. In concrete terms, we will discuss the composition of each specifically as follows:

1) ***State space:*** In principle, the designed state space contains as much information as possible about the environment relevant to the problem $\mathbf{P0}$, guiding the agent to make meaningful decisions. Also, each state defined is explicit and there is no overlap between states.

The current information about all users mainly includes the corresponding channel information about all users $\mathbf{R}$, specifically the GES of the PN, the VU of the SN, and the Eve, as well as others. Thus, the achievable rate at the last time slot, which can be defined as

$$\mathbf{U} = \{\mathbf{g}, \mathbf{h}_n, \mathbf{h}_m, \mathbf{H}_{UR}, R_G, R_n, R_m\}, \tag{20}$$

where $\mathbf{g}$, $\mathbf{h}_n$ and $\mathbf{h}_m$ are the relevant channel coefficients of the GES, the n-th VU and m-th Eve, respectively. The chosen action vector $\mathbf{A} \in \mathcal{A}$, derived from the last time slot, is specifically designed based on how an agent's strategy update is influenced by the value function, particularly when assessing the advantage function. The instant rewards $(\mathbf{R})$ are determined based on state information and chosen actions. These rewards provide an intuitive measure of the agent's performance and effectiveness in addressing the problem $\mathbf{P0}$. In general, there are several ways to implement the instant reward, and the complete computation procedure for the instant reward is provided in the following section. As a consequence, the state space can be constructed as

$$\mathcal{S} = \{\mathbf{U}, \mathbf{A}, \mathbf{R}\}, \tag{21}$$

where $s^{(t)} \in \mathcal{S}$ is the state at the t-th time slot.

2) ***Action space:*** The action space can be designed as the UAV transmit beamforming vectors $\{\mathbf{v}_n\}$, the common-private rate splitting ratio χ_n, the RIS phase shift $\mathbf{\Phi}$. Since deep neural networks can only accept real parts rather than complex-valued parts as input or output, during the construction of the action $\mathcal{A}$, the real and imaginary parts are separated into separate input ports if complex numbers are involved. Given transmit symbols with unit variance, the transmit beamforming matrix $\{\mathbf{v}_n\}$ of the n-th VU can be decomposed into two parts, i.e.,

$$\mathbf{v}_n = \|\mathbf{v}_n\| \, \bar{\mathbf{v}}_n, \tag{22}$$

where $\|\mathbf{v}_n\|$ and $\bar{\mathbf{v}}_n$ are the transmit power of the corresponding data stream at UAV and the normalized beam assignment characterizing the beamforming direction, respectively.

Remark 1: In fact, with the availability of RSMA technology in the SN, the transmission of the common stream by the VU should not interfere with the reception of the private stream, thanks to the SIC technique. To address this, we employ the maximum ratio transmission (MRT) scheme to design the common stream transmission beamforming vector in SN. Additionally, to further mitigate interference between VUs and prevent potential information leakage, we utilize the zero-forcing (ZF) scheme to design the beamforming vector for the private stream, which has been not only adopted but also verified to be highly effective in [62]. By employing distinct beamforming strategies for common and private streams, we achieve a balance between maximizing data throughput and maintaining privacy and security.

Following *Remark 1*, our application of the above two methods involving beamforming directions separately also simplifies the action space based on the PPO-enabled method to facilitate the learning of the near-optimal mapping from MRT to actions by the agent as it is easier to implement, which can be expressed as

$$\bar{\mathbf{v}}_n = \begin{cases} \frac{\sum_{i=1}^{N} \mathbf{z}_{c,n}^H}{\left\| \sum_{i=1}^{N} \mathbf{z}_{c,n}^H \right\|}, n = 0, \\ \frac{\mathbf{V}_n}{\|\mathbf{V}_n\|}, n \neq 0, \end{cases} \tag{23}$$

where $\mathbf{V}_n$ denotes the n-th column of $\mathbf{V} = [\mathbf{V}_1, ..., \mathbf{V}_N]$ where $\mathbf{V} = \mathbf{Z}^H \left(\mathbf{Z}\mathbf{Z}^H\right)^{-1}$ and $\mathbf{Z} = [\mathbf{z}_{c,1}, ..., \mathbf{z}_{c,N}]$.

We initially compress the inputs to optimize them for processing within DNNs. This compression facilitates the normalization of input values, ensuring they fall within a fixed range suitable for neural network processing. This step is crucial in minimizing distortion and enhancing the performance of the DNN, especially in complex communication systems where the precision of input data significantly influences the network's output accuracy. Following this, the common-private rate splitting ratio χ_n, is determined using hyperbolic tangent functions. The use of hyperbolic tangent functions for this purpose is strategic, these functions efficiently map large input domain values to an output range of (-1,1), effectively stabilizing the learning process by preventing extreme value outputs that can destabilize the DNN. Specifically, i t c an be expressed as

$$\chi_n = \frac{1}{2}\left(\frac{\exp\left(x_n^{EC}\right) - \exp\left(-x_n^{EC}\right)}{\exp\left(x_n^{EC}\right) + \exp\left(-x_n^{EC}\right)} + 1\right), \tag{24}$$

where x_n^{EC} stands for the output processed by this activation function. Furthermore, the discrete phase shift values assigned to each reflective element of the RIS can be represented as

$$\Theta = \left\{0, \Delta\theta, ..., \left(2^b - 1\right)\Delta\theta\right\}, \tag{25}$$

where b is the number of unitized bits, denoting the phase shift level, and $\Delta\theta = \pi/2^{(b-1)}$. For this part, we can determine the phase shifts θ_k of each reflective from the Eq. (25). In this regard, the action space can be expressed as

$$\mathcal{A} = \left\{\left\{\mathbf{v}_n\right\}, \left\{\chi_n\right\}, \left\{c_n\right\}, \left\{\theta_k\right\}\right\}, \tag{26}$$

where we define $a^{(t)} \in \mathcal{A}$ to be the chosen action at the t-th time slot in a concise manner.

3) ***Reward function:*** In the reward function we design, the objective function **P0** may be only a part of the problem, while other aspects such as constraints or rules need to be considered in the decision-making process. Specifically, it can be composed of two terms, i.e., an instant reward term that represents the expression of unconstrained emotion and a penalty term that ensures the fulfillment of various constraints. The reward function can be defined as

$$\begin{aligned}\mathcal{R} = &\,\mathrm{SEE}\left(\left\{\mathbf{v}_n, \left\{\chi_n\right\}, \left\{c_n\right\}\right\}, \mathbf{\Phi}\right) \\ &\times \left(\varsigma_e \times \varsigma_c \times \varsigma_r\right),\end{aligned} \tag{27}$$

where

$$\varsigma_e = \begin{cases} 1, \min R_G - R_{s,m} \geq \Delta U, \\ 0, \min R_G - R_{s,m} < \Delta U, \end{cases} \tag{28a}$$

$$\varsigma_c = \begin{cases} 1, \left\|\mathbf{v}_c\right\|^2 + \sum_{n=1}^{N}\left\|\mathbf{v}_n\right\|^2 \leq P_U, \\ 0, \left\|\mathbf{v}_c\right\|^2 + \sum_{n=1}^{N}\left\|\mathbf{v}_n\right\|^2 > P_U, \end{cases} \tag{28b}$$

$$\varsigma_r = \begin{cases} 1, \sum_{n=1}^{N} c_n - \min_n R_{c,n} \leq 0, \\ 0, \sum_{n=1}^{N} c_n - \min_n R_{c,n} > 0, \end{cases} \tag{28c}$$

where ς_e, ς_c and ς_r stand for the penalties for the selected actions that does not satisfy the QoS requirements of the PN, the transmit power budget requirements of the SN, and the common information decoding requirements of the SN, respectively. The reward received at time t is denoted as $R^{(t)}$. It's important to note that the reward is non-zero only under the condition that all constraints specified in problem **P0** are fully met. This ensures that the agent not only strives for optimal performance but also adheres strictly to the constraints, integrating compliance directly into the reward structure. This mechanism promotes the development of solutions that are not only effective but also feasible within the defined problem parameters, ensuring that the approach is holistic. Furthermore, this design helps the system to quickly recognize and take the right actions to obtain more positive rewards and avoid negative consequences.

IV. Proposed Improved PPO-based Scheme

In general, the constructed MDP describes environments that are perfectly observable. This means that the observed state fully characterizes the environment features.One crucial element of MDP involves selecting actions solely based on the current state, without taking into account the past states in the decision-making process. The goal of MDP is to determine the optimal policy within a predefined framework. This framework directs the agent on how to maximize the long-term rewards by selecting the next action in the current state. Essentially, the framework establishes a mapping between the constructed state space and action space, which is governed by a probability distribution over the action space. In the constructed MDP, let's assume that $\pi\left(s^{(t)}\right)$, denoted by $s^{(t)} \in \mathcal{S}$, represents this policy function at the t-th time slot.

A. Basic Knowledge

In the scenario where the model is free from predefined state transition probabilities, the Q-learning algorithm is applied to manage decision-making in an unknown environment. The Q-learning Q-function $Q^\pi\left(s^{(t)}, a^{(t)}\right)$ within this framework is defined as the expected cumulative reward of selecting action $a^{(t)}$ in state $s^{(t)}$, guided by the policy $\pi\left(s^{(t)}\right)$, which can be expressed as

$$Q^\pi\left(s^{(t)}, a^{(t)}\right) = \mathbb{E}\left[\sum_{\tau=0}^{\infty}\gamma^\tau R\left(t+\tau+1\right)|s^{(t)}, a^{(t)}\right], \tag{29}$$

The Bellman equation iteratively updates the value of the Q-function in Q-learning based on the instant reward received and the maximum expected future rewards. This approach is independent of a specific model of the environment, making it versatile and robust for applications where the dynamics are complex or difficult to model. By employing this technique, the algorithm progressively refines its estimates of the Q-values through experience, essentially learning the best actions to take in various states without prior knowledge of the environmental dynamics. By iteratively evaluating the value of each possible action in a given state, this equation can compute the following optimal strategy

$$Q^\pi\left(s^{(t)}, a^{(t)}\right) = \mathbb{E}\left[R^{(t)} + \gamma Q^\pi\left(s^{(t+1)}, a^{(t+1)}\right)\right]. \tag{30}$$

As mentioned earlier, there are continuous and discrete variables in the objective optimization problem **P0**. Based on the state-of-the-art results in the field of DRL and considering the mutual coupling of the optimization parameters, we propose an improved PPO algorithm that incorporates LSTM mechanism, i.e., the policy network and the value network in the PPO algorithm use the LSTM architecture. The LSTM architecture enables the algorithm to maintain the memory of the past actions and outcomes, especially in the case of relayering problems. This is critical to understanding the time dependence and dynamics within the action space [43]. In the following, we will further introduce the traditional PPO framework, the LSTM mechanism, and the improved LSTM-PPO algorithm.

The PPO algorithm is an optimization method that updates policies based on their gradients, aiming to find the optimal policy through gradient ascent. It achieves a favorable

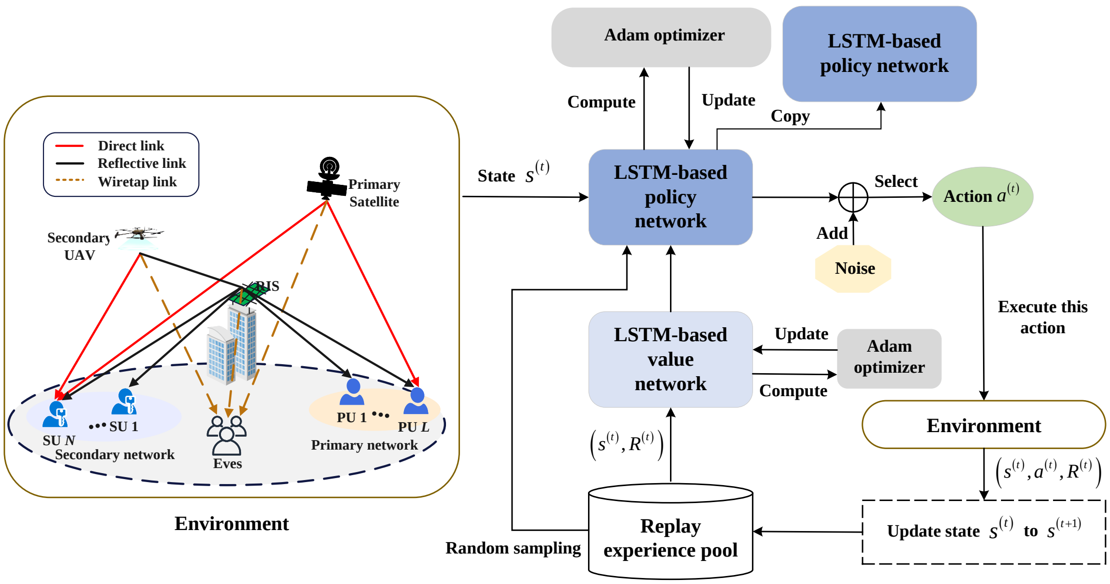

Fig. 2. The proposed LSTM-PPO-based framework.

balance between sample complexity, execution speed, and ease of parameter tuning, making it a preferred choice for optimization tasks in this paper. As previously mentioned, we define $\pi\left(a^{(t)}|s^{(t)}\right)$ as the policy network that selects actions $a^{(t)}$ given the current state $s^{(t)}$. Typically, this policy network is usually represented by a deep neural network with parameter η, which can be written $\pi_\eta\left(a^{(t)}|s^{(t)}\right)$. The state-value function, often denoted as $V_\pi\left(s^{(t)}\right)$, represents the expected rewards starting from state $s^{(t)}$ and following the policy π. One of the key features of PPO is its clipped surrogate objective, which ensures that policy updates are not too drastic, thereby maintaining learning stability. Within the PPO optimization framework, the surrogate objective function can be formulated as

$$J(\eta) = \mathbb{E}_\pi\left[\min\left(\frac{\pi_\eta\left(a^{(t)}|s^{(t)}\right)}{\pi_{\eta_{old}}\left(a^{(t)}|s^{(t)}\right)} A\left(s^{(t)}, a^{(t)}\right), c(\cdot)\right)\right], \tag{31}$$

where

$$A\left(s^{(t)}, a^{(t)}\right) = R\left(s^{(t)}, a^{(t)}\right) + \gamma V_\pi\left(s^{(t+1)}\right) - V_\pi\left(s^{(t)}\right), \tag{32}$$

and

$$c(\cdot) = \mathrm{clip}\left(\frac{\pi_\eta\left(a^{(t)}|s^{(t)}\right)}{\pi_{\eta_{old}}\left(a^{(t)}|s^{(t)}\right)}, 1-\varepsilon, 1+\varepsilon\right) A\left(s^{(t)}, a^{(t)}\right). \tag{33}$$

where $A\left(s^{(t)}, a^{(t)}\right)$ denotes the advantage function at each step, γ denotes the discount factor, $R\left(s^{(t)}, a^{(t)}\right)$ denotes the reward obtained for executing the current action $a^{(t)}$ in state $s^{(t)}$, and $c(\cdot)$ stands for the clipping function that restricts the correction weights to the interval $(1-\varepsilon, 1+\varepsilon)$ with ε being a small hyperparameter. This iterative process involves updating the associated parameter η based on gradients computed from a subset of transitions, facilitating efficient learning and convergence towards optimal values.

B. Description of the Proposed LSTM-PPO Algorithm

To enhance the exploration capability of the PPO algorithm, this paper first introduces Ornstein-Uhlenbeck (OU) noise to the output actions. As a type of time-correlated noise, the OU noise mechanism can simulate random disturbances, thus encouraging the agent to explore the environment more comprehensively and avoid falling into local optimal.

Furthermore, considering the highly dynamic nature, complex channel state information, and extensive coverage of the RIS-assisted SATINs communication environment, as well as the limitations of the fully connected neural network (FNN) used in traditional PPO algorithms for approximating policy and value functions, we propose an improved LSTM-PPO algorithm. Specifically, both the policy network and the value network in the PPO algorithm are designed based on an LSTM architecture, as shown in Fig. 2. LSTM networks are renowned for their excellent ability to process sequential data, enabling them to effectively extract key features from high-dimensional and time-correlated communication environment data. In particular, LSTM units, through their internal forget gate, input gate, and output gate mechanisms, can selectively retain and forget information, capturing long-term dependencies in channel state variations over time.

In our LSTM-PPO algorithm architecture shown in Fig. 3, the LSTM network acts as a feature extractor that receives the state of the high-dimensional communication environment as input. Subsequently, these features are passed to the fully-connected neural network layer, which further processes these features and eventually approximates the policy function (i.e., the probability distribution of actions output by the actor network) and the value function (i.e., the estimation of the state or action state values by the critic network). In this way,

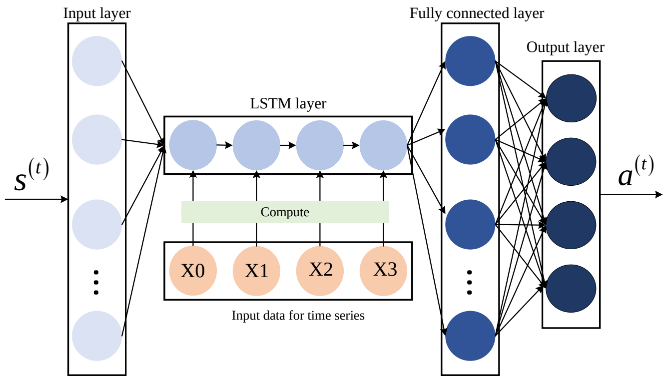

Fig. 3. Policy network architecture with integrated LSTM network.

the proposed LSTM-PPO algorithm not only enhances the ability to learn from complex continuous sample data, such that the ability to make efficient decisions in highly dynamic and uncertain wireless environments is also improved. The details of the design of the policy network and the value network are described below.

•**Policy network component**: For the policy network component, the input layer is set with 12 nodes corresponding to the dimensions of the designed state space, and the hidden layer is set with the LSTM network layer. Among them, the LSTM network layer is set up with 4 network cells. Following the LSTM network layer, three fully connected layers are implemented, each employing the Tanh activation function to enhance the network's nonlinear representation capability and facilitate feature integration. These fully connected layers are responsible for further processing the features extracted by the LSTM and provide the basis for the final decision variable generation. They correspond to the UAV transmit beamforming matrix variation $\Delta\mathbf{v}_n$, the common-private splitting ratio variation $\Delta\chi_n$, the common information transmission power variation Δc_n and RIS phase shift matrix variation $\Delta\theta_k$, respectively. The architecture of the policy network is shown in Fig. 3.

•**Value network component**: For the value network component, the input layer is set up with 16 nodes corresponding to the merging of the 12 state space quantities and the variations generated by the current policy network $(\Delta\mathbf{v}_n, \Delta\chi_n, \Delta c_n, \Delta\theta_k)$. Both the hidden and fully-connected layers are set up in the same way as the policy network. The output layer is set up with one node, corresponding to the state-value function $V_\pi\left(s^{(t)}\right)$.

•**Computational flow of the proposed algorithm**: The improved LSTM-PPO algorithm combines the feature extraction capability of LSTM and the PPO-enabled optimization method, which is suitable for dealing with complex sequence data and highly dynamic environments, and the algorithm flow can be summarized in **Algorithm 1.**

Algorithm 1: Proposed LSTM-PPO Approach

Input: Associated channels
- $\{\mathbf{h}_U\}, \{\mathbf{h}_R\}, \mathbf{H}_{UR}$, and $\mathbf{H}_{UG}$;
- Maximum number of episodes;
- Maximum step number per episode;
- Initial parameters of the LSTM and fully connected neural networks;
- Initial policy network π_η and value network V_π.

Output: Action set $a = \{\{\mathbf{v}_n\}, \{\chi_n\}, \{c_n\}, \{\theta_k\}\}$.

for *each episode $e = 1$ to maximum episodes* **do**
  Receive the initial state $s^{(0)} \in S$; set $S \leftarrow s^{(0)}$;
  for *step $t = 1$ to maximum time step* **do**
    Compute action $a^{(t)} \sim \pi_\eta(a^{(t)}|s^{(t)}, h^{(t)})$;
    Execute action $a^{(t)}$ in environment and observe reward $R^{(t)}$ and next state $s^{(t+1)}$;
    Store the tuple $(s^{(t)}, a^{(t)}, R^{(t)}, s^{(t+1)})$ into the memory replay pool;
    Update LSTM hidden state $h^{(t)}$ to $h^{(t+1)}$;
    if *memory buffer is sufficiently filled* **then**
      for *each mini-batch in replay buffer* **do**
        Compute advantages $\hat{A}_t$ using Generalized Advantage Estimation (GAE); Compute returns $\hat{R}^{(t)}$;
        Update policy network π_η by maximizing PPO-Clip objective Eq. (31);
        Update value network V_π by minimizing mean squared error:
        $$\pi \leftarrow \arg\min_{\pi} \mathbb{E}_\pi \left[(V_\pi(s^{(t)}, h^{(t)}) - \hat{R}_t)^2 \right]$$
      end
      Clear replay buffer;
    end
  end
end

C. Complexity and Convergence Analysis

The state information collected from the channel matrix associated with the optimization objective, as well as the previous moment's action including the UAV transmit beamforming scheme, the RIS phase shift matrix and the power splitting ratio, are used as inputs to the LSTM-PPO enabled framework. This is then processed through the LSTM mapping mechanism. During the training phase, the size of the input layer depends on the number of states and is denoted as Z_I. Note that the input layer is followed by the LSTM layer, and the computational complexity of the LSTM layer is usually related to the input size Z_I and the number of hidden units Z_L in the LSTM layer. However, in each iteration, the LSTM layer is processed time-step in time steps, the complexity of the LSTM layer at each time step is roughly $\mathcal{O}\left(Z_I Z_L + {Z_L}^2\right)$. Next is the fully connected layer, if there are F fully connected layers with Z_f neurons per layer, then the complexity of the fully connected layer can be calculated as $\mathcal{O}\left(\sum_{f=1}^{F} Z_f \cdot Z_{f+1}\right)$. Assuming that there are N^{epi} sets in each minibatch and that the time step for each set is $T_{\max}$, each iteration of the training model is completed until the algorithm converges. Therefore, the total computational complexity of the scheme can be calculated as $\mathcal{O}\left(N^{\text{epi}} T_{\max}\left(Z_I \cdot Z_L + {Z_L}^2 + \sum_{f=1}^{F} Z_f \cdot Z_{f+1}\right)\right)$.

Considering the subsequent resolution of potential scalability issues in large-scale SATIN deployment, we incorporate an early stopping mechanism into the LSTM-PPO framework. This mechanism leverages specially designed reward functions to terminate training when predefined c onvergence thresholds or constraint boundaries are reached, thereby preventing redundant iterations. However, both system performance and algorithmic complexity exhibit significant variations with changing numbers of secondary VUs or RIS reflecting elements. The dynamic nature of these parameters underscores challenges in maintaining operational consistency and efficiency across diverse scenarios. To this end, future work will focus on auxiliary optimization strategies such as agent reconfiguration mechanisms and model compression techniques, which adaptively adjust to spatial correlations in the retrained model to accommodate substantial parameter variations.

V. Numerical Results and Analysis

In this section, we present the simulation results to further demonstrate the performance of our proposed LSTM-PPO-based joint resource allocation algorithm. We obtain the numerical results through extensive simulations, utilizing a set of system parameters and DNN hyperparameters, as detailed in Table I. To maintain generality without sacrificing realism, we focus our analysis on users within the satellite's and UAV's beam coverage range. Typical deployment scenarios justify this assumption, as they often place users requiring connectivity within the coverage areas of these aerial platforms [63]. Furthermore, the homogeneous poisson point process models the strategic deployment of potential Eves near legitimate users to intercept sensitive data. Taking the satellite beam center as the origin, the PUs are evenly distributed at a distance of 50m from the beam center. The UAV beam center's position coordinates are (5,100), and the VUs position coordinates are (5,125), (5,145), and (5,165). Besides, RIS is located near VUs, with coordinates (20,125).

According to [64], satellites are limited in their service range to users within specific altitude bands. While the satellite downlink channel is generalized, it's crucial to consider the impact of altitude on the statistical characteristics of the satellite channel. Typically, this is achieved through empirical expression transformations. As noted in literature [65], satellite users typically operate within elevation angles ranging from approximately 80° and 20°, accounting for terrain effects. Satellite downlink channels are often classified based on fading severity, such as frequent light fading (FLS), average fading (AS), and frequent heavy fading (FHS), which consider factors like fading severity, multipath component power averages, and LoS components [66].

A. Convergence Analysis

To further illustrate the convergence properties of the proposed LSTM-PPO-based method, we compare its training results with other well-known DRL strategies, such as PPO, DDQN, DDPG, and Soft Actor-Critic (SAC) [67]. The DNN in 'PPO' is a traditional CNN architecture, as can be seen in [68]. The complexity of each algorithm at different time slots is tabulated in Table 3 and the performance within a fixed total time slot ($T = 100$) is evaluated as given in Fig. 3. It can be seen that the PPO-based approach shows a clear superiority, obtaining significantly higher average rewards. This advantage mainly stems from the unique ability of the PPO-based framework to adaptively modify the magnitude of the gradient update. Better learning results are achieved by effectively fine-tuning the strategy in both discrete and continuous action spaces. Computational complexity is another concern, as illustrated by the running time of the programme shown in Table III. It can be seen that the LSTM-PPO scheme has the shortest running time, this is because LSTM naturally encodes historical information through the hidden state, and only the current state and the hidden state are needed for inference, which reduces the amount of data inputted in a single step, and thus reduces the computational complexity. Furthermore, it should be noted that Fig. 4 shows the convergence of the different algorithms, while Fig. 5 shows the performance of the PPO-based algorithm for different total time slots T. It can be seen that as the total time slot increases, the performance improvement is obvious. This may be due to the fact that in RIS-assisted SATIN, the number of time slots may represent the length of the network transmission or the time granularity of resource allocation. In general, more time slots may mean more transmission opportunities, thus potentially improving the overall performance.

TABLE II
Simulation Parameters Setting

Simulation Parameters	**Value**
Satellite type	GEO
Maximum antenna gain of UAV	48 dB
Maximum transmit power of UAV	30 dBm
Circuit power consumption of UAV P_C	5 dBm
Altitude of UAV	5 km
Carrier frequency	28 GHz
Noise variance at Eves	100 dBm
Secrecy rate constraint	$\Delta_U = 0.2$ bit/s/Hz
Frequent heavy shadowing, FHS	(1,0.063,0.0007)
Average shadowing, AS	(5,0.251,0.279)
Simulation DNN Parameters	**Value**
Reward discount factor	0.9
Size about the experience replay buffer	100000
Learning rate	0.0001
Amount of training episodes	10000
Size about mini-sample	128
Amount of experience in the mini-batch	64
Number of neurons in LSTM layer	32
Dropout rate for LSTM layer	0.6
Clipping parameter ε	0.2

TABLE III
Comparison of computational complexity (s)

T	LSTM-PPO(s)	PPO(s)	DDQN(s)	DDPG(s)	SAC(s)
60	780.41	810.11	1351.51	1647.12	2988.36
80	847.89	933.11	1653.17	1985.55	3622.71
100	1026.93	1066.11	2331.98	2633.72	4489.32

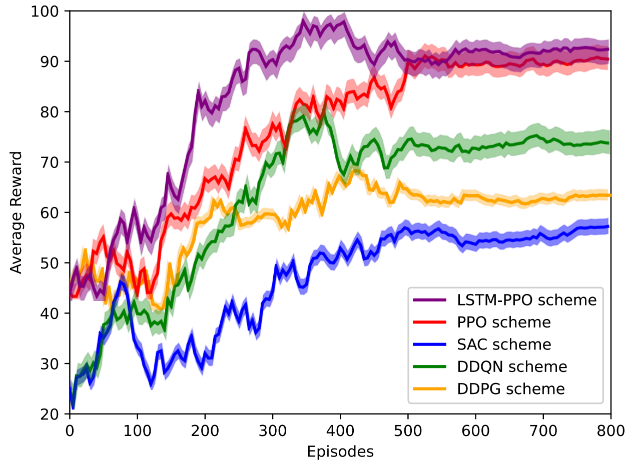

Fig. 4. Convergence comparison between different algorithms.

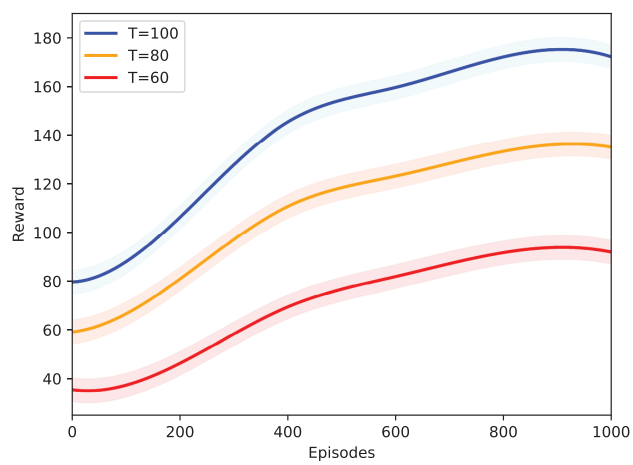

Fig. 5. Convergence of the LSTM-PPO-based algorithm under different time slots.

B. Variable Ablation Experiments

Furthermore, ablation experiments with different system model variable settings are also simulated to confirm the validity of each optimization variable and the comprehensiveness of the action space construction.

• Fixed active beamforming scheme: This scheme can be abbreviated to “FAB scheme” [69]. Denoting that only the RIS phase shift matrix $\mathbf{\Phi}$ and power splitting ratio χ_n are jointly optimized but the UAV transmit beamforming matrix $\mathbf{v}$ is fixed by MRT scheme.

• Fixed passive beamforming scheme: This scheme can be abbreviated to “FPB scheme”. Denoting that only the UAV transmit beamforming matrix $\mathbf{v}$ and power splitting ratio χ_n are jointly optimized but the RIS phase shift matrix $\mathbf{\Phi}$ is fixed without optimization.

• Random passive beamforming scheme - RPB scheme: This scheme can be abbreviated to “RPB scheme”. Denoting that only the UAV transmit beamforming matrix $\mathbf{v}$ and power splitting ratio χ_n are jointly optimized but the RIS phase shift matrix $\mathbf{\Phi}$ is randomly generated.

The advantages of the PPO-based algorithm have already been demonstrated, thus, Fig. 5 depicts the variation of the SEE with the number of episodes based on the PPO algorithm for different ablation experimental settings of $N=3$, $K=10$, $N_U=4$, $\Delta_c=2$ bit/s/Hz, and $T=100$. It can be seen that all the algorithms converge to a state of convergence around 450 episodes. Meanwhile, because the PPO-based algorithm optimizes the three parameters combinations $(\Delta\mathbf{v}_n, \Delta\chi_n, \Delta\theta_k)$ simultaneously, the SEE obtained is significantly higher than the other schemes. In addition, the FAB scheme outperforms the FPB scheme because when the active transmit beamforming matrix is fixed, the RIS can adapt the wireless propagation environment by adjusting its phase shift matrix to maximize the system sum rate. The advantages of the FPB scheme over the RPB scheme, on the other hand, are mainly in stability and predictability. In the FPB scheme, although the RIS phase shift matrix is fixed, the system is still able to optimize the transmission performance to a certain extent through the designed active transmit beamforming matrix. This is because the adjustment of the active transmit beamforming compensates for the limitation due to the fixed RIS phase shift matrix, making the signal transmission more efficient. In contrast, the RPB scheme is often difficult to achieve effective signal transmission optimization due to the random nature of the phase shift matrix, which leads to unstable performance.

C. Simulations Settings with different system parameters

Fig. 6 plots the variation of SEE of SN versus different maximum transmit power of UAV employing various algorithms, where $N=3$, $K=10$, $N_U=4$, $\Delta_c=2$ bit/s/Hz, and $T=100$. The graph reveals a progressive increase in the SEE for the RIS-assisted SN scheme as the maximum transmit power escalates, with a convergence evident at a threshold of 22 dBm. Notably, the PPO-based algorithm consistently achieves a higher SEE compared to other reference schemes. This excellent performance is attributed to the algorithm’s comprehensive optimization of multiple parameters, even in the presence of multiple Eves. These findings underscore that simply raising the maximum transmit power is not invariably beneficial for SEE improvement. This complex interplay suggests a diminishing return on SEE beyond certain transmit power levels, emphasizing the need for strategic power management in RIS-assisted secure communications.

Fig. 7 demonstrates the relationship between the average worst-case ASR and the progressive scaling of RIS reflective elements under configuration parameters $N=3$, $N_U=4$, $\Delta_c=2$ bit/s/Hz, and $T=100$. A pronounced increase in average worst-case ASR is observed across all evaluated methodologies as the RIS element count grows, fundamentally driven by the augmented signal propagation capabilities enabled by expanded reflective surfaces. Notably, the PPO-based and FAB schemes exhibit significantly steeper performance growth compared to FPB and RPB baselines. This divergence originates from the advanced capacity of PPO-based and FAB frameworks to dynamically optimize phase shift configurations across RIS elements in real time, enabling precise alignment of reflection parameters to maximize secure transmission efficiency. Such dynamic adaptation not only enhances coherent

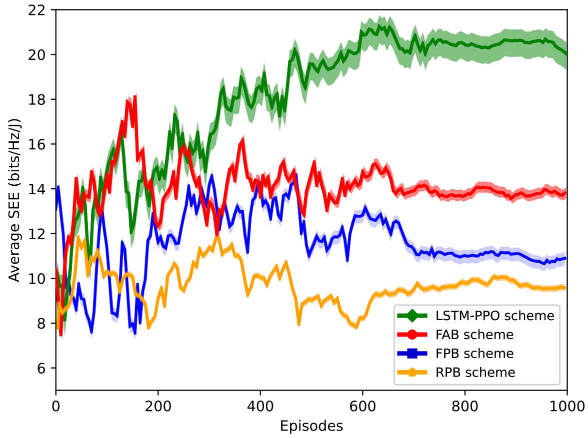

Fig. 6. Convergence comparison between different schemes based on PPO algorithm.

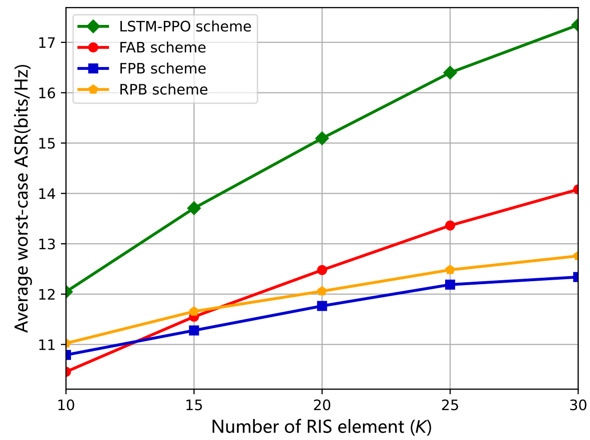

Fig. 8. Comparison of average worse-case ASR and the number of RIS elements.

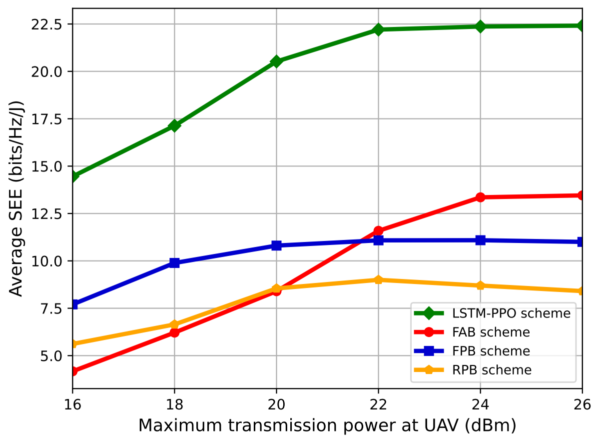

Fig. 7. Comparison of SEE and UAV maximum transmission power.

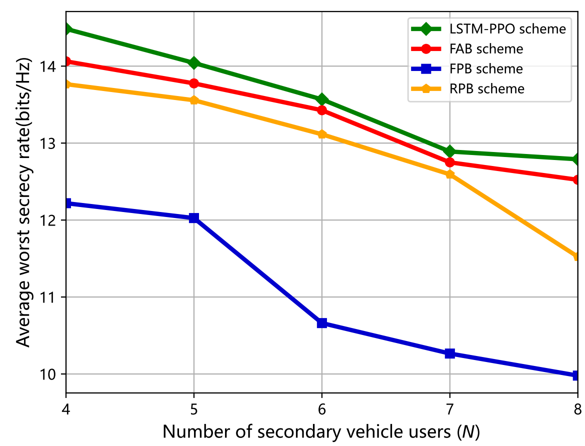

Fig. 9. Comparison of average worse-case ASR and the number of VUs.

signal superposition but also ensures robust environmental adaptability, allowing these methods to mitigate interference and maintain performance stability under varying channel conditions. In contrast, the static or restricted phase adjustment mechanisms inherent to FPB and RPB approaches limit their ability, resulting in suboptimal reflective beamforming and diminished secrecy rate gains.

Fig. 8 illustrates the average worst-case ASR as a function of the number of VUs, with parameters set as $K = 10$, $N_U = 4$, $\Delta_c = 2$ bit/s/Hz, and $T = 100$. The figure shows a clear trend that as the number of VUs increases, the average worst-case ASR decreases for all inspection scenarios. This decrease is mainly due to the escalation of inter-user interference caused by a large number of VUs sharing the communication medium [70]. Interestingly, whilst inter-user interference typically results in signal overlap and obfuscation, thereby degrading communication performance, it provides a unique opportunity to enhance security protocols. RSMA technology cleverly exploits this interference to obfuscate transmissions, but in practice the implementation of such a technique requires a delicate balancing act between effectively exploiting the interference to increase the data rate and not unduly affecting the overall system performance. In our proposed scheme, although the main focus is on the optimisation of passive RIS, we also consider key factors such as signal processing and energy efficiency, which are also applicable to the design of active RIS [71], [72].

VI. Conclusion

This paper employed RSMA and RIS for improving connectivity and enhancing security, thus providing a brand new infrastructure of cognitive SATINs in the presence of eavesdroppers. Different from existing works focusing on secrecy efficiency (SE) or energy efficiency (EE) separately, we formulated a secrecy-energy efficiency (SEE) maximization problem to achieve a tradeoff in secondary UAV network, and then proposed a unified LSTM-PPO framework based on DRL to further handle both continuous and discrete optimization variables. This framework designed state-action pairs that dealt with secondary UAV transmit beamforming, RIS reflected beamforming, and the power splitting ratio. Simulation results showed that by reconfiguring the wireless communication environment, the role of RIS clearly achieves an adjustable trade-off among spectral-, secrecy- and energy-efficiency. By synergistically leveraging RIS and RSMA technologies, the proposed LSTM-PPO framework demonstrates superior SEE performance in SATINs compared to other benchmark schemes, offering a promising direction for the future development of SATINs.

References

[1] R. Liu *et al.*, "RIS-empowered satellite-aerial-terrestrial networks with PD-NOMA," *IEEE Commun. Surv. Tutorials*, doi: 10.1109/COMST.2024.3393612.

[2] K. An *et al.*, "Secure transmission in cognitive satellite terrestrial networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 11, pp. 3025-3037, Nov. 2016.

[3] Y. He, Y. Xiao, S. Zhang, M. Jia, and Z. Li, "Direct-to-smartphone for 6G NTN: Technical routes, challenges, and key technologies," *IEEE Network*, vol. 38, no. 4, pp. 128-135, Jul. 2024.

[4] R. Tang *et al.*, "Deep reinforcement learning-based resource allocation for multi-UAV-assisted full-duplex wireless-powered IoT networks," *IEEE Trans. Cognit. Commun. Networking*, doi: 10.1109/TCCN.2024.3407096.

[5] Q. Wu *et al.*, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106-112, Jan. 2020.

[6] J. Xu *et al.*, "Sum secrecy rate maximization for IRS-aided multi-cluster MIMO-NOMA Terahertz systems," *IEEE Trans. Inf. Forensics Secur.*, vol. 18, pp. 4463-4474, 2023.

[7] X. Cao *et al.*, "Reconfigurable intelligent surface-assisted aerial-terrestrial communications via multi-task learning," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 10, pp. 3035-3050, Oct. 2021

[8] X. Li *et al.*, "Hardware impaired ambient backscatter NOMA systems: Reliability and security," *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2723-2736, April 2021.

[9] E. Basar *et al.*, "Reconfigurable intelligent surfaces for 6G: Emerging hardware architectures, applications, and open challenges," *IEEE Veh. Technol. Mag.*, vol. 19, no. 3, pp. 27-47, Sept. 2024.

[10] F. Naeem, M. Ali, G. Kaddoum, C. Huang, C. Yuen, "Security and privacy for reconfigurable intelligent surface in 6G: A review of prospective applications and challenges," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 1196-1217, 2023.

[11] X. Li *et al.*, "Enhancing secrecy performance for STAR-RIS NOMA networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 2684-2688, Feb. 2023.

[12] K. Guo *et al.*, "On the secrecy performance of NOMA-based integrated satellite multiple-terrestrial relay networks with hardware impairments,"*IEEE Trans. Veh. Technol.*, vol. 70, no. 4, pp. 3661-3676, Apr. 2021.

[13] R. Kaur, B. Bansal, S. Majhi, S. Jain, C. Huang and C. Yuen, "A Survey on reconfigurable intelligent surface for physical layer security of next-generation wireless communications," *IEEE Open J. Veh. Technol.*, vol. 5, pp. 172-199, 2024.

[14] X. Li *et al.*, "Secure communication of active RIS assisted NOMA networks," *IEEE Trans. Wireless Commun.*, vol. 23, no. 5, pp. 4489-4503, May 2024.

[15] Z. Feng *et al.*, "Graph attention-based reinforcement learning for trajectory design and resource assignment in multi-UAV assisted communication," *IEEE Internet Things J.,* vol. 11, no. 16, pp. 27421-27434, 15 Aug.15, 2024.

[16] C. Huang *et al.*, "Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1839-1850, Aug. 2020.

[17] J. Huang, Y. Yang, J. Lee, D. He and Y. Li, "Deep reinforcement learning-based resource allocation for RSMA in LEO satellite-terrestrial networks," *IEEE Trans. Commun.*, vol. 72, no. 3, pp. 1341-1354, Mar. 2024.

[18] Z. Lin *et al.*, "Secure and energy efficient transmission for RSMA-based cognitive satellite-terrestrial networks," *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 251-255, Feb. 2021.

[19] M. Rivera *et al.*, "Optimization of RIS-assisted RSMA-enabled tethered-UAV communications," *2023 IEEE Globecom Workshops (GC Wkshps)*, Kuala Lumpur, Malaysia, 2023, pp. 227-232.

[20] X. Li *et al.*, "Rate-splitting multiple access-enabled security analysis in cognitive satellite terrestrial networks," *IEEE Trans. Vehi. Tech.*, vol. 71, no. 11, pp. 11756-11771, Nov. 2022.

[21] S. Han *et al.*, "Joint broadcast and unicast transmission based on RSMA and spectrum sharing for integrated satellite-terrestrial network," *IEEE Trans. Cognit. Commun. Networking*, vol. 10, no. 3, pp. 1090-1103, June 2024.

[22] K. Guo *et al.*, "STAR-RIS-empowered cognitive non-terrestrial vehicle network with NOMA," *IEEE Trans. Intell. Veh.*, vol. 8, no. 6, pp. 3735-3749, Jun. 2023.

[23] C. Pan *et al.*, "An overview of signal processing techniques for RIS/IRS-aided wireless systems," *IEEE J. Sel. Top. Signal Process.*, vol. 16, no. 5, pp. 883-917, Aug. 2022.

[24] H. Niu *et al.*, "Active RIS-assisted secure transmission for cognitive satellite terrestrial networks," *IEEE Trans. Vehi. Tech.*, vol. 72, no. 2, pp. 2609-2614, Feb. 2023.

[25] K. Guo *et al.*, "Outage performance of RIS-assisted cognitive non-terrestrial network with NOMA," *IEEE Trans. Vehi. Tech.*, vol. 73, no. 4, pp. 5953-5958, Apr. 2024.

[26] Y. Liu *et al.*, "Deep learning empowered trajectory and passive beamforming design in UAV-RIS enabled secure cognitive non-terrestrial networks," *IEEE Wireless Commun. Lett.*, vol. 13, no. 1, pp. 188-192, Jan. 2024.

[27] K. Guo *et al.*, "Joint trajectory and beamforming optimization for federated DRL-aided space-aerial-terrestrial relay networks with RIS and RSMA," *IEEE Trans. Wireless Commun.*, vol. 23, no. 12, pp. 18456-18471, Dec. 2024.

[28] X. Sheng *et al.*, "Performance analysis of STAR-RIS assisted secure cognitive NOMA-HARQ networks," *IEEE Wireless Commun. Lett.*, vol. 13, no. 3, pp. 696-700, Mar. 2024.

[29] Q. T. Ngo *et al.*, "Multi-agent DRL-based RIS-assisted spectrum sensing in cognitive satellite-terrestrial networks," *IEEE Wireless Commun. Lett.*, vol. 12, no. 12, pp. 2213-2217, Dec. 2023.

[30] H. Wu, X. Yang and Z. Bu, "Deep reinforcement learning for computation offloading and resource allocation in satellite-terrestrial integrated networks," *2022 IEEE 95th Vehi. Tech. Conf.: (VTC2022-Spring)*, Helsinki, Finland, 2022, pp. 1-5.

[31] W. Xu *et al.*, "Time-varying channel prediction for RIS-assisted MU-MISO networks via deep learning," *IEEE Trans. Cognit. Commun. Networking*, vol. 8, no. 4, pp. 1802-1815, Dec. 2022.

[32] P. Saikia *et al.*, "Meta reinforcement learning-based spectrum sharing between RIS-assisted cellular communications and MIMO radar," *IEEE Trans. Cognit. Commun. Networking*, vol. 10, no. 1, pp. 164-179, Feb. 2024.

[33] D. Li *et al.*, "A unified rate-splitting framework for secure spectrum sharing via joint precoding optimization," *IEEE Systems J.*, vol. 17, no. 4, pp. 5580-5591, Dec. 2023

[34] D. Li *et al.*, "Precoding optimization assisted secure transmission for rate-splitting multiple access," *ICC 2022 - IEEE Inter. Conf. Commun.*, Seoul, Korea, Republic of, 2022, pp. 673-678.

[35] K. Guo *et al.*, "Deep reinforcement learning and NOMA-based multi-objective RIS-assisted IS-UAV-TNs: Trajectory optimization and beamforming design," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 9, pp. 10197-10210, Sept. 2023.

[36] J. Xu, C. Yuen, C. Huang, N. Ul Hassan, G. C. Alexandropoulos, M. D. Renzo, M. Debbah, "Reconfiguring wireless environments via intelligent surfaces for 6G: Reflection, modulation, and security," *Sci. China Inf. Sci.*, Nov 2022.

[37] Y. Liang, K. Huang, L. Yang, X. Li, H. Liu and Y. Bian, "Covert communications for active STAR-RIS-aided RSMA systems with hardware impairments," *IEEE Trans. Intell. Transp. Syst.*, doi: 10.1109/TITS.2025.3542808.

[38] A. Bansal *et al.*, "Rate splitting multiple access for intelligent reflecting surface aided multiuser communications," *IEEE Trans. Veh. Technol.*, vol. 70, no. 9, pp. 9217-9229, Sep. 2021.

[39] Y. He, B. Sheng, H. Yin, D. Yan, and Y. Zhang, "Multi-objective deep reinforcement learning based time-frequency resource allocation for multi-beam satellite communications," *China Commun.*, vol. 19, no. 1, pp. 77-91, Jan. 2022.

[40] Y. He, Y. Liu, C. Jiang, and X. Zhong, "Multiobjective anti-collision for massive access ranging in MF-TDMA satellite communication system," *IEEE Internet Things J.*, vol. 9, no. 16, pp. 14655-14666, Aug. 2022.

[41] F. Wen, J. Shi and Z. Zhang, "Joint 2D-DOD, 2D-DOA, and polarization angles estimation for bistatic EMVS-MIMO radar via PARAFAC analysis," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1626-1638, Feb. 2020,

[42] 3GPP TR 36.777, "Study on Enhanced LTE Support for Aerial Vehicles", Release 15, V15.0.0, 2018-12.

[43] M. Wu *et al.*, "Deep reinforcement learning-based energy efficiency optimization for RIS-aided integrated satellite-aerial-terrestrial relay networks," *IEEE Trans. Commun.*, vol. 72, no. 7, pp. 4163-4178, July 2024.

[44] K. Guo *et al.*, "Performance analysis of hybrid satellite-terrestrial cooperative networks with relay selection," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 9053-9067, Aug. 2020.

[45] G. Zheng *et al.*, "Physical layer security in multibeam satellite systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 2, pp. 852–863, Feb. 2012.

[46] Y. Sun *et al.*, "Energy-efficient hybrid beamforming for multilayer RIS-assisted secure integrated terrestrial-aerial networks," *IEEE Trans. Commun.*, vol. 70, no. 6, pp. 4189-4210, Jun. 2022.

[47] T. S. Rappaport *et al.*, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029-3056, Sept. 2015.

[48] X. Li *et al.*, "Exploiting benefits of IRS in wireless powered NOMA networks," *IEEE Trans. Green Commun. Networking*, vol. 6, no. 1, pp. 175-186, Mar. 2022.

[49] C. Huang *et al.*, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157-4170, Aug. 2019.

[50] R. Dong *et al.*, "Secure transmission design of RIS enabled UAV communication networks exploiting deep reinforcement learning," *IEEE Trans. Vehi. Tech.*, vol. 73, no. 6, pp. 8404-8419, June 2024.

[51] R. Zhang *et al.,* "Energy efficiency maximization in RIS-assisted SWIPT networks with RSMA: A PPO-based approach," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 5, pp. 1413-1430, May 2023.

[52] H. Niu *et al.,* "Robust design for intelligent reflecting surface-assisted secrecy SWIPT network," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 4133-4149, Jun. 2022.

[53] X. Yan *et al.*, "Ergodic capacity of NOMA-based uplink satellite networks with randomly deployed users," *IEEE Syst. J.*, vol. 14, no. 3, pp. 3343–3350, Sep. 2020.

[54] X. Gao *et al.*, "Rate-splitting multiple access-based cognitive radio network with ipSIC and CEEs," *IEEE Trans. Vehi. Tech.*, vol. 73, no. 1, pp. 1430-1434, Jan. 2024.

[55] H. Niu *et al.*, "Joint beamforming design for secure RIS-assisted IoT networks," *IEEE Internet of Things J.*, vol. 10, no. 2, pp. 1628-1641, Jan., 2023.

[56] Z. Lin *et al.*, "Refracting RIS-aided hybrid satellite-terrestrial relay networks: Joint beamforming design and optimization," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 4, pp. 3717-3724, Aug. 2022.

[57] H. Niu *et al.*, "Active RIS assisted rate-splitting multiple access network: Spectral and energy efficiency tradeoff," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 5, pp. 1452-1467, May 2023.

[58] M. Mozaffari *et al.*, "Wireless communication using unmanned aerial vehicles (UAVs): Optimal transport theory for hover time optimization," *IEEE Trans. Wireless Commun.*, vol. 16, no. 12, pp. 8052-8066, Dec. 2017.

[59] T. Gong *et al.*, "Holographic MIMO communications: Theoretical foundations, enabling technologies, and future directions," *IEEE Commun. Surv. Tutorials*, vol. 26, no. 1, pp. 196-257, Firstquarter 2024.

[60] H. An and L. Wang, "Robust topology generation of internet of things based on PPO algorithm using discrete action space," *IEEE Trans. Ind. Informat.*, vol. 20, no. 4, pp. 5406-5414, Apr. 2024.

[61] N. Zhao, Z. Ye, Y. Pei, Y. -C. Liang and D. Niyato, "Multi-agent deep reinforcement learning for task offloading in UAV-assisted mobile edge computing," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 6949-6960, Sept. 2022.

[62] B. Clerckx *et al.*, "Rate-splitting unifying SDMA, OMA, NOMA, and multicasting in MISO broadcast channel: A simple two-user rate analysis," *IEEE Wireless Commun. Lett.*, vol. 9, no. 3, pp. 349-353, Mar. 2020.

[63] K. Guo *et al.*, "Ergodic capacity of NOMA-based overlay cognitive integrated satellite-UAV-terrestrial networks," *Chinese J. Electron.*, vol. 32, no. 2, pp. 273-282, Mar. 2023.

[64] H. Zhang *et al.*, "Outage analysis of cooperative satellite-aerial-terrestrial networks with spatially random terminals," *IEEE Trans. Commun.*, vol. 70, no. 7, pp. 4972-4987, Jul. 2022.

[65] K. An *et al.*, "Outage performance of cognitive satellite terrestrial networks with interference constraint," *IEEE Trans. Vehi. Tech.*, vol. 65, no. 11, pp. 9397-9404, Jan. 2016.

[66] Z. Lin *et al.*, "Secrecy-energy efficient hybrid beamforming for satellite-terrestrial integrated networks," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 6345-6360, Sept. 2021.

[67] C. Meng *et al.*, "Sum-rate maximization in STAR-RIS-assisted RSMA networks: A PPO-based algorithm," *IEEE Internet Things J.*, vol. 11, no. 4, pp. 5667-5680, Feb., 2024.

[68] Y. Meng, S. Kuppannagari, R. Kannan and V. Prasanna, "PPOAccel: A high-throughput acceleration framework for proximal policy optimization," *IEEE Trans. Parallel Distrib. Syst.*, vol. 33, no. 9, pp. 2066-2078, 1 Sept. 2022.

[69] Y. Wu *et al.*, "Spectrum and energy efficiency tradeoff in IRS-assisted CRNs with NOMA: A multi-objective optimization framework," *ICC 2022 - IEEE Inter. Conf. Commun.*, Seoul, Korea, Republic of, 2022, pp. 3930-3935.

[70] C. Zhou *et al.*, "Energy-efficient maximization for RIS-aided MISO symbiotic radio systems," *IEEE Trans. Veh. Technol.*, vol. 72, no. 10, pp. 13689-13694, Oct. 2023.

[71] K. Zhi, C. Pan, H. Ren, K. K. Chai and M. Elkashlan, "Active RIS versus passive RIS: Which is superior with the same power budget?," *IEEE Commun. Lett.*, vol. 26, no. 5, pp. 1150-1154, May 2022.

[72] T. Wu *et al.*, "Employing high-dimensional RIS information for RIS-aided localization systems," *IEEE Commun. Lett.*, vol. 28, no. 9, pp. 2046-2050, Sept. 2024.

Min Wu received the Ph.D. degree from Space Engineering University, Beijing, China, in 2024, where she is currently a research associate at Beihang University. Her research interests include satellite-terrestrial covert networks, multiuser communication systems, and federated reinforcement learning. Moreover, she was the recipient of the Best Paper Award of EAI WiSATS 2024.

Kefeng Guo received his B.S. degree from Beijing Institute of Technology, Beijing, China in 2012, and the Ph.D. degree in Army Engineering University, Nanjing, China in 2018. He is a Lecturer in School of Space Information, Space Engineering University. He is also an associate professor in the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics. He has authored or coauthored nearly 100 research papers in international journals and conferences. His research interests focus on cooperative relay networks, MIMO communications systems, multiuser communication systems, satellite communication, hardware impairments, cognitive radio, NOMA technology and physical layer security. He was a recipient of exemplary Reviewer for **IEEE Transactions on Communications** in 2022. He was the recipient of the Outstanding Ph.D. Thesis Award of Chinese Institute of Command and Control in 2020. He also was the recipient of the Excellent Ph.D. Thesis Award of Jiangsu Province, China in 2020. He was the recipient of the Best Paper Award of WiSATS 2024. He was listed in the World's Top 2% Scientists identified by Stanford University in 2022-2024. He also serves as an Editor on the Editorial Board for the **EURASIP Journal on Wireless Communications and Networking** and **IEEE Open Journal of the Communications Society.**

Dr. Guo has been the TPC member of many IEEE sponsored conferences, such as IEEE ICC, IEEE GLOBECOM and IEEE WCNC.

Xingwang Li (S'12-M'15-SM'20) received the M. S. and Ph. D. degrees from University of Electronic Science and Technology of China and Beijing University of Posts and Telecommunications in 2010 and 2015. From 2010 to 2012, he worked at Comba Telecom Ltd. In Guangzhou China, as an engineer. He spent one year from 2017 to 2018 as a visiting scholar at Queen's University Belfast, Belfast, UK. He is currently an Associated Professor with the School of Physics and Electronic Information Engineering, Henan Polytechnic University, Jiaozuo China. His research interests span wireless communication, intelligent transport system, artificial intelligence, Internet of things.

Dr. Li has served as many TPC members, such as the IEEE Globecom, IEEE ICC, IEEE WCNC, IEEE VTC, IEEE ICCC etc. He has also served as the Co-Chair for the IEEE/IET CSNDSP 2020 of the Green Communications and Networks Track. He also serves as an Editor on the Editorial Board for IEEE Transactions on Intelligent Transportation Systems, IEEE Transactions on Vehicular Technology, IEEE Systems Journal, Physical Communication. He was the Guest Editor for the special issue on Computational Intelligence and Advanced Learning for Next-Generation Industrial IoT of IEEE Transactions on Network Science and Engineering, "Recent Advances in Physical Layer Technologies for 5G-Enabled Internet of Things" of the Wireless Communications and Mobile Computing.

Lin Zhi received the B.E. and M.E. degrees in information and communication engineering from the PLA University of Science and Technology, and the Ph.D. degree in electronic science and technology from the Army Engineering University of PLA, Nanjing, China, in 2013, 2016, and 2020, respectively. From March 2019 to June 2020, he was a visiting Ph.D. student with the Department of Electrical and Computer Engineering, McGill University, Montreal, Canada. Since January 2021, he has been with the College of Electronic Engineering, National University of Defense Technology, Hefei, China, where he is currently an Associate Professor.

Dr. Lin's research interests include array signal processing, physical layer security, reconfigurable intelligent surface and satellite-aerial-terrestrial integrated networks. He was listed in the 2024 Highly Cited Chinese Researchers by Elsevier, World's Top 2% Scientists identified by Stanford University in 2022-2024. He was the recipient of the Young Elite Scientist Sponsorship Program of China Association for Science and Technology in 2021, the Macao Young Scholars Fellowship in 2022, the Outstanding Ph.D. Thesis Award of Chinese Institute of Electronics in 2022, and the Best Paper Awards from IEEE/CIC ICCC 2024, EAI WiSATS 2024, NCIC 2024, IEEE ICCT 2023, and IWCMC 2023. He was the Symposium Co-Chair of IEEE WCSP 2022 and WCSP 2024, and TPC members of IEEE flagship conferences, including IEEE ICC, Globecom, Infocom, VTC, and so on. He has been serving as an Editor for the Physical Communication since 2024 and a Lead Guest Editor of the IET Communications Special Issues on Reconfigurable Intelligent Surfaces Aided Physical Layer Security in 6G Wireless Networks.

Liang Yang was born in Hunan, China. He received the Ph.D. degree in electrical engineering from Sun Yat-sen University, Guangzhou, China, in 2006. From 2006 to 2013, he was a Teacher with Jinan University, Guangzhou. In 2013, he joined the Guangdong University of Technology, Guangzhou. He is currently a Professor with Hunan University, Changsha, China. His current research interests include the performance analysis of wireless communications systems.

Theodoros A. Tsiftsis (S'02-M'04-SM'10) received the PhD degree in electrical engineering from the University of Patras, Greece, in 2006. He is is with the Department of Electrical and Electronic Engineering, University of Nottingham Ningbo China, Ningbo 315100, China, and also with the Department of Informatics and Telecommunications, University of Thessaly, Lamia 35100, Greece. His research interests fall into the broad areas of communication theory and wireless communications, with an emphasis on wireless communications theory, reconfigurable intelligent surfaces, optical wireless communications, and physical layer security.

Dr. Tsiftsis served on the Editorial Boards of the IEEE Transactions on Communications, IEEE Transactions on Vehicular Technology, IEEE Communications Letters, and IEEE Transactions on Mobile Computing. He is currently an Associate Editor of the IEEE Transactions on Wireless Communications, and Specialty Chief Editor for Networks and Communications of Frontiers in Computer Science. Prof. Tsiftsis was appointed as an IEEE Vehicular Technology Society Distinguished Lecturer for two terms (2018–2022) and was recently appointed as an IEEE Communications Society Distinguished Lecturer (2024–2025).

Chau Yuen received the B.Eng. and Ph.D. degrees from Nanyang Technological University, Singapore, in 2000 and 2004, respectively. He was a Post-Doctoral Fellow with Lucent Technologies Bell Labs, Murray Hill, in 2005. From 2006 to 2010, he was with the Institute for Infocomm Research, Singapore. From 2010 to 2023, he was with the Engineering Product Development Pillar, Singapore University of Technology and Design. Since 2023, he has been with the School of Electrical and Electronic Engineering, Nanyang Technological University, currently he is Provost's Chair in Wireless Communications, Assistant Dean in Graduate College, and Cluster Director for Sustainable Built Environment at ERIN. Dr. Yuen received IEEE Communications Society Leonard G. Abraham Prize (2024), IEEE Communications Society Best Tutorial Paper Award (2024), IEEE Communications Society Fred W. Ellersick Prize (2023), IEEE Marconi Prize Paper Award in Wireless Communications (2021), IEEE APB Outstanding Paper Award (2023), and EURASIP Best Paper Award for JOURNAL ON WIRELESS COMMUNICATIONS AND NETWORKING (2021). Dr Yuen current serves as an Editor-in-Chief for Springer Nature Computer Science, Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, where he was awarded as IEEE TNSE Excellent Editor Award 2024 and 2022, and Top Associate Editor for TVT from 2009 to 2015. He also served as the guest editor for several special issues, including IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS MAGAZINE, IEEE COMMUNICATIONS MAGAZINE, IEEE VEHICULAR TECHNOLOGY MAGAZINE, IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, and ELSEVIER APPLIED ENERGY.

He is listed as Top 2% Scientists by Stanford University, and also a Highly Cited Researcher by Clarivate Web of Science from 2022. He has 4 US patents and published over 500 research papers at international journals.