DRL Based Secure Optimization for RIS Aided SATINs with RSMA

Min Wu[1(✉)], Kefeng Guo[1], Zhi Lin[2], Huiyun Xia[3], Kang An[4,5], Liang Yang[5], and Jiangzhou Wang[6]

[1] School of Space Information, Space Engineering University, 101407 Beijing, China
1800022837@pku.edu.cn
[2] College of Electronic Engineering, National University of Defense Technology, 230037 Hefei, China
[3] College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, 210003 Nanjing, China
[4] Sixty third Research Institute, National University of Defense Technology, 210007 Nanjing, China
ankang89@nudt.edu.cn
[5] College of Computer Science and Electronic Engineering, Hunan University, 410082 Changsha, China
liangy@hnu.edu.cn
[6] School of Engineering, University of Kent, CT2 7NT Canterbury, UK
j.z.wang@kent.ac.uk

Abstract. Amid the escalating demand for accessible users and security insurance in satellite aerial terrestrial integrated networks (SATINs), security and energy efficiency emerge as pivotal indicators. This paper proposes a secure beamforming scheme in reconfigurable intelligent surface (RIS) aided SATINs, in presence with multiple eavesdroppers, where rate splitting multiple access (RSMA) and RIS are adopted at the secondary UAV networks for achieving multiuser diversity and antijamming. To optimize the secrecy energy efficiency (SEE) for secondary networks while adhering to constraints on ground earth station (GES) secrecy rate, a deep reinforcement learning (DRL) framework is proposed to address the coupling between optimization variables through the improved proximal policy optimization (PPO) method, of which from existing DRL scheme is that the proposed one builds a unified learning framework. Simulation results indicate that the SEE derived by the proposed DRL scheme is superior to that of benchmark schemes, which validate the advantage of this work.

Keywords: Satellite aerial terrestrial integrated networks · reconfigurable intelligent surface · rate splitting multiple access · deep reinforcement learning · security

H.-H. Chen and W. Meng (Eds.): WiSATS 2024, LNICST 605, pp. 77–88, 2025.
https://doi.org/10.1007/978-3-031-86196-3_7

1 Introduction

Recently, satellite aerial terrestrial integrated networks (SATINs) has been recognized as a promising infrastructure for the next generation networks (NGs), where satellite and unmanned aerial vehicle (UAV) subnetworks employ cognitive radio (CR) technology to share the limited frequency spectrum under certain constraints [1,2]. But in this way, the challenge of preventing eavesdroppers (Eves) from iniercepting private signals while ensuring the quality of service (QoS) of the primary network is a crucial one to be solved. As an effective supplement to cryptographic methods, physical layer security (PLS) has attracted extensive attention and investigation [3]. From this perspective, achieving an optimal balance between security and performance has become an important part of realizing the potential of the SATINs.

As well known that higher energy utilization efficiency is one of the inherent requirements of the NGs. Based on this consideration, reconfigurable intelligent surfaces (RISs) appear to tackle this issue. In contrast to traditional relays, RIS operates without the need for active radio frequency (RF) chains or complex signal processing components. This feature enables it to demonstrate significant capabilities in enhancing the PLS for SATINs. By strategically manipulating the phase shift of reflection elements, RIS facilitates constructive superimposition of direct and reflected signals at intended legitimate users, meanwhile inducing destructive interference for potential Eves [4–6]. Additionally, another method for energy efficiency improvement is multiple access, such as rate splitting multiple access (RSMA), which can split information into common and private parts for overlay transmission at the transmitter. This technique can leverage uses successive interference cancellation (SIC) at the receiver, which is considered a powerful approach for augmenting spectral efficiency (SE). Specifically, within the CR enabled NTNs, the design significance of this architecture lies in improving the performance of the unauthorized secondary network (SN). As a result, exploring the integration of RIS into SATINs to support RSMA in maximizing SEE of the SN while ensuring interference limitations becomes crucial.

Optimization problems in the RIS assisted SATINs under consideration are usually nonlinear and may contain a large number of variables. Traditional optimization methods often struggle to deal with this complexity. It is worth mentioning that in problems that require optimization of multiple variables, it is difficult to get optimized answers by traditional methods, and on this basis, model free artificial intelligence (AI) emerges. Based on the above, By establishing an appropriate Markov decision process (MDP), DRL has become a powerful technology for handling explosive communication data and finding optimal solutions through interaction with the environment continuously.

As discussed in the former paragraph, communication security is also the demand of the SATINs. This requirement highlights the significance of PLS in safeguarding transmissions against eavesdropping. Therefore, it can be seen that integrating multiple security enhancing technologies such as RIS and RSMA at different levels into SATINs will be a promising infrastructure. However, this amalgamation also presents several challenges that must be addressed to

optimize security performance effectively. First of all, the foremost challenge lies in the inadequacy of traditional model based wireless technologies to meet the requirements of emerging RIS assisted SATINs, such as excessively complex communication scenarios with accurate mathematical description. Second, with the incorporation of RIS, the dynamic adjustments of its reflective elements add to the unpredictability of the wireless environment, complicating real time sensing. Finally, the dynamic multi channel access of CR technology in the process of spectrum sharing requires independent selection of access free spectrum resources under the condition of time varying spectrum occupancy. As an effective method to solve dynamic problems, DRL has been widely used in the wireless communication optimization.

There is currently researches on utilizing DRL in RIS aided secure SATINs with RSMA to maximize secure transmission performance of secondary networks. Our objective is to maximize the SEE by jointly designing the transmit beamforming of the UAV, RIS reflecting matrix and power splitting ratio, while guaranteeing the service quality of the primary network. Since there are several mutually coupled parameters in the formulated optimization problem, which are difficult to be solved using traditional algorithms, we develop a unified proximal policy optimization (PPO) learning framework based on the DRL method by designing dynamic reward functions that can seamlessly handle both continuous and discrete variables. Finally, simulation results demonstrate that our proposed PPO enabled framework exhibits greater adaptability than several benchmark approaches.

2 System Model and Problem Formulation

In this paper, we investigate the SEE optimization of secondary users in RIS aided SATINs in the presence of multiple eavesdroppers (Eves). In this setup, the whole integrated networks consists of two subnetworks, where the satellite sends multicast signals to multiple GESs and shares spectrum to UAV with 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 (Fig. 1).

2.1 Signal Model with RSMA

According to [7,8], 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 vehicle user (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. Meanwhile, $\mathbf{g}_{e,m}$, $\mathbf{g}_{c,n}$ denotes the channel gains from satellite to the m th Eve, the satellite to the n th VU. We assume that the satellite and the UAV are outfitted with array fed reflector antennas comprising N_S feeds and a uniform linear array (ULA) consisting of N_U antennas, respectively. In the transmission phase of the primary network from the satellite to a total of L

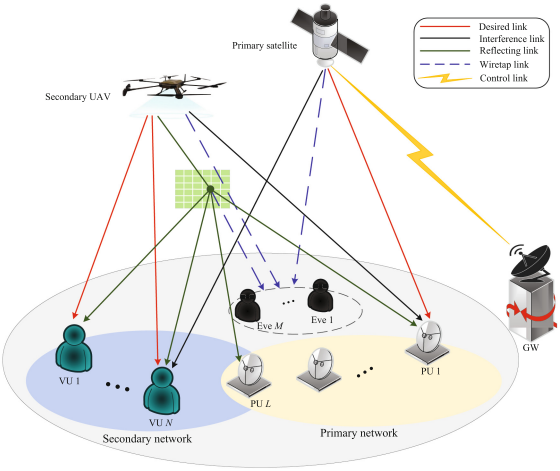

Fig. 1. Proposed system model.

ground earth stations (GESs) as PUs, the private multicast signal x is mapped to the satellite transmit beamforming precoding matrix $\mathbf{w} \in C^{N_S \times 1}$ in the presence of M Eves [9]. During the secondary network signalling process, the UAV can be described as utilizing its flexibility to serve N single antenna ground VUs, while the RIS equipped with $K = K_y \times K_z$ reflective elements is also deployed in the SN to enhance the expected signals at the VUs while suppressing the signals received by the Eves. Specifically, the reflecting phase shift matrix of RIS is given by $\mathbf{\Phi} = diag\left\{\beta_1 e^{j\theta_1}, \beta_2 e^{j\theta_2}, ..., \beta_K e^{j\theta_K}\right\}$, where β_k and θ_k are the amplitude factor and phase shift at the k th RIS element [10]. Meanwhile, applying the RSMA technology at the UAV, the unicast signal s_n designated for the n th VU gets divided into a common subsignal s_c with common transmit beamforming vector $\mathbf{v}_c \in \mathbb{C}^{N_U \times 1}$ and a private subsignal $s_{p,n}$ with the private beamforming vector $\mathbf{v}_n \in \mathbb{C}^{N_U \times 1}$. Utilizing a shared codebook among all VUs, the common subsignals are collectively encoded into a common signal stream s_c. The same frequency band is permitted to be shared by both satellites and the UAV within a specified interference tolerance in the GES of the primary network. Notably, a total of M Eves can also intercept signals transmitted by satellites and the UAV in a similar manner as legitimate VUs [11]. Consequently, the received signals at the GES, the m th Eve and the n th VU can be expressed as

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

$$y_{s,m} = \mathbf{g}_{e,m}^H \mathbf{w} x + \mathbf{z}_{e,m}^H \mathbf{v}_c s_c + \sum_{n=1}^{N} \mathbf{z}_{e,m}^H \mathbf{v}_n s_{p,n} + n_{e,m}, \tag{2}$$

$$y_{c,n} = \mathbf{g}_{c,n}^H \mathbf{w} x + \mathbf{z}_{c,n}^H \mathbf{v}_c s_c + \sum_{n=1}^{N} \mathbf{z}_{c,n}^H \mathbf{v}_n s_{p,n} + n_{c,n}, \tag{3}$$

where

$$\mathbf{z}_G = \left(\mathbf{H}_{UG} + \mathbf{h}_{RG}^H \mathbf{\Phi} \mathbf{H}_{UR}\right)^H, \tag{4}$$

$$\mathbf{z}_{e,m} = \left(\mathbf{h}_{U,m} + \mathbf{h}_{R,m}^H \mathbf{\Phi} \mathbf{H}_{UR}\right)^H, \tag{5}$$

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

Thus, the achievable rate of the GES, the m th Eves, the common signal s_c and the private subsignal stream of the n th VU can be expressed as

$$R_G = \log_2 \left(1 + \frac{\left|\mathbf{g}^H \mathbf{w}\right|^2}{\left|\mathbf{z}_G^H \mathbf{v}_c\right|^2 + \sum_{n=1}^{N} \left|\mathbf{z}_G^H \mathbf{v}_n\right|^2 + \sigma_G^2}\right), \tag{7a}$$

$$R_{s,m} = \log_2 \left(1 + \frac{\left|\mathbf{g}_{e,m}^H \mathbf{w}\right|^2}{\left|\mathbf{z}_{e,m}^H \mathbf{v}_c\right|^2 + \sum_{n=1}^{N} \left|\mathbf{z}_{e,m}^H \mathbf{v}_n\right|^2 + \sigma_{e,m}^2}\right), \tag{7b}$$

$$R_{c,n} = \log_2 \left(1 + \frac{\left|\mathbf{z}_{c,n}^H \mathbf{v}_c\right|^2}{\left|\mathbf{g}_{c,n}^H \mathbf{w}\right|^2 + \sum_{n=1}^{N} \left|\mathbf{z}_{c,n}^H \mathbf{v}_n\right|^2 + \sigma_{c,n}^2}\right), \tag{7c}$$

$$R_{p,n} = \log_2 \left(1 + \frac{\left|\mathbf{z}_{c,n}^H \mathbf{v}_n\right|^2}{\left|\mathbf{g}_{c,n}^H \mathbf{w}\right|^2 + \sum_{i \neq n}^{N} \left|\mathbf{z}_{c,n}^H \mathbf{v}_i\right|^2 + \sigma_{c,n}^2}\right). \tag{7d}$$

Meanwhile, we define $\chi_n \in (0, 1)$ as the common private power splitting ratio for the n th VU [12]. In this form, the above achievable rate of the common and private stream of the i th VU can be rewritten as

$$R_{c,i} = \log_2 \left(1 + \frac{\chi_i \left|\mathbf{z}_{c,i}^H \mathbf{v}_c\right|^2}{\chi_i \sum_{n=1}^{N} \left|\mathbf{z}_{c,i}^H \mathbf{v}_n\right|^2 + \left|\mathbf{g}_{c,i}^H \mathbf{w}\right|^2 + \sigma_{c,i}^2}\right), \tag{8}$$

and

$$R_{p,i} = \log_2 \left(1 + \frac{\chi_i \left|\mathbf{z}_{c,i}^H \mathbf{v}_i\right|^2}{\chi_i \sum_{n=1, n \neq i}^{N} \left|\mathbf{z}_{c,i}^H \mathbf{v}_n\right|^2 + \left|\mathbf{g}_{c,i}^H \mathbf{w}\right|^2 + \sigma_{c,i}^2}\right). \tag{9}$$

2.2 Problem Description

For secure and green communication, we need to codesign the UAV active secure beamforming vector $\mathbf{v} = [\mathbf{v}_c, \mathbf{v}_1, ..., \mathbf{v}_N]$ with $\mathbf{v_c}$ being the common stream beamforming vector, the RIS phase shift matrix $\mathbf{\Phi}$, and the transmit power splitting ratio χ_n to maximize the SN SEE, which can be defined as the ratio of the sum

rate of the SN to the power consumption [13]. Besides, the achievable sum rate of the UAV enabled SN can be expressed as

$$R_U\left(\{\mathbf{v}_n, \mathbf{\Phi}\}, \chi_n\right) = \sum\nolimits_{i=1}^{N} (c_i + R_{p,i}), \tag{10}$$

While satisfying the GES secrecy constraint, the mathematical expression for the SEE can be defined as

$$\mathrm{SEE} = \frac{R_U - \sum\limits_{m=1}^{M} \max R_{s,m}}{\left(\|\mathbf{v}_c\|^2 + \sum_{n=1}^{N} \|\mathbf{v}_n\|^2\right) + P_C}, \tag{11}$$

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 [14]. Mathematically, the whole optimization problem can be formulated as

$$\mathbf{P0}: \max_{\mathbf{v}_n, \mathbf{\Phi}, \chi_n} \max_{\mathbf{e}_{U,n}, \mathbf{e}_{R,n}} \mathrm{SEE}, \tag{12a}$$

$$\text{s.t. } C1: \min_{\mathbf{e}_{g,m}, \mathbf{e}_{z,m}} R_G - R_{s,m} \geq \Delta_U, \forall m, \tag{12b}$$

$$C2: R_{c,n} \geq \Delta_c, \forall n, \tag{12c}$$

$$C3: R_{p,n} \geq \Delta_p, \forall n, \tag{12d}$$

$$C4: \|\mathbf{w}\|^2 \leq P_S, \|\mathbf{v}_c\|^2 + \sum\nolimits_{n=1}^{N} \|\mathbf{v}_n\|^2 \leq P_U, \tag{12e}$$

$$C5: \left|\exp\left(j\theta^k\right)\right| = 1, \forall k, \tag{12f}$$

$$C6: \sum\nolimits_{n=1}^{N} c_n \leq \min_n R_{c,n}\left(\{\mathbf{w}_l\}, \mathbf{\Phi}, \chi_n\right), \tag{12g}$$

where $\mathbf{e}_{U,n}$ and $\mathbf{e}_{R,n}$ denote the channel estimation error about $\mathbf{h}_{U,n}$ and $\mathbf{h}_{R,n}$. And, P_C denotes the constant circuit power consumption. Meanwhile, $C1$ denotes that the achievable rate of the secondary network users (SUs) must be less than the achievable rate of the PUs, and $C2$ and $C3$ denote the minimum rate requirement for the SN transmit signals. $C4$ is the transmit power limit for PN and SN, where P_S and P_U are the preset power caps for satellites and UAVs. $C6$ denotes that the VU of the SN is able to fully decode the common stream.

3 MDP for the Secrecy Energy Efficiency Maximization

As shown in Sect. 2, the problem **P0** is a high dimension complex problem that is difficult to solve directly, because it contains both discrete and continuous variables. In addition, in realistic RIS aided secure SATINs, the capabilities about obtaining information of VUs in the SN, the channel quality, and the service demand will change dynamically. Moreover, **P0** is an optimization problem confined to a single time slot. Its solution might converge to a suboptimal

one, akin to a greedy search, given the overlooked historical environmental state and long term gains. Therefore, it is generally infeasible to employ conventional optimization techniques such as AO, SDP or SCA to achieve efficient and secure beamforming strategies in uncertain dynamic satellite wireless environments.

Model free reinforcement learning is a dynamic programming tool which can be continuously adopted to tackle decision making problems by learning optimal solutions in dynamic environments. Therefore, we model the secure beamforming optimization in RIS aided SATINs as an RL problem. In addition, in the model free RL family, the policy based learning approach, represented prominently by PPO, which 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. Here, the optimization problem **P0** is first transformed into MDP and then solved through by the proposed approach that supports a unified PPO framework [15]. By utilizing the PPO algorithm, the gradient is updated by employing a method that trims advantage functions. This is done to control the magnitude of each update, with the goal of maximizing the cumulative reward for agents across various states.

1) ***State space:*** For the designed state space, it contains, in principle, as much information as possible about the environment relevant to the problem **P0**. The effectiveness of the DRL algorithm is largely determined by the state space, which needs to include all participating states of the entire system, such as the current channel information of all users, all action vectors to be selected $\mathbf{A}$, and rewards $\mathbf{R} \in \mathcal{R}$ obtained after interaction. The current information about all users mainly includes the corresponding channel information, the achievable rate can be defined as

$$\mathbf{U} = \{\mathbf{g}, \mathbf{h}_n, \mathbf{h}_m, R_G, R_n, R_m\}, \tag{13}$$

where $\mathbf{g}$, $\mathbf{h}_n$ and $\mathbf{h}_m$ are the relevant channel coefficients of the GES, the n th VU and the mth Eve, respectively. As a consequence, the state space can be constructed as

$$\mathcal{S} = \{\mathbf{U}, \mathbf{A}, \mathbf{R}\}, \tag{14}$$

where we define $s^{(t)} \in \mathcal{S}$ to be the state of the representation 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.,

$$\mathbf{v}_n = \|\mathbf{v}_n\| \, \bar{\mathbf{v}}_n, \tag{15}$$

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

Following [12], we simplified the beamforming direction to facilitate the learning of approximate optimal mapping from state to action by agents, as it is easier to implement and can be represented as

$$\bar{\mathbf{v}}_n = \begin{cases} \frac{\sum_{i=1}^{N} \mathbf{z}_{c,n}^H}{\left\|\sum_{i=1}^{N} \mathbf{z}_{c,n}^H\right\|}, n = 0, \\ \frac{\mathbf{V}_n}{\|\mathbf{V}_n\|}, n \neq 0, \end{cases} \tag{16}$$

where $\mathbf{V}_n$ denotes the V 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 $\mathbf{Z} = [\mathbf{z}_{c,1}, ..., \mathbf{z}_{c,N}]$.

In this regard, the action space can be expressed as

$$\mathcal{A} = \left\{ \{\mathbf{v}_n\}, \{\chi_n\}, \{c_n\}, \{\theta_k\} \right\}, \tag{17}$$

where χ_n is the generation of hyperbolic tangent functions, θ_k indicates the amount of phase shift change. Besides, we define $a^{(t)} \in \mathcal{A}$ to be the chosen action at the tth time slot of the subsequent representation.

3) ***Reward function:*** The constraints of the objective problem P0 need to be considered simultaneously in the reward function we have designed. It can be composed of two items, i.e., the instant reward term that represents the expression of unconstrained emotion and the penalty term that ensures various constraints can be satisfied. The main purpose of the reward function is to guide the agent to learn towards the desired goal. Instant rewards can help the agent to get positive feedback quickly and accelerate the learning process. The penalty terms, on the other hand, can avoid the agent from adopting bad behaviors or actions with large errors, thus helping the agent to better explore and optimize the strategy. Thus, to further strike a balance between the instant reward and penalty terms, the reward function can be defined as

$$\begin{aligned} \mathcal{R} = &\operatorname{SEE}\left(\{\mathbf{v}_n, \{\chi_n\}, \{c_n\}\}, \mathbf{\Phi}\right) \\ &\times \left(\varsigma_e \times \varsigma_c \times \varsigma_r\right), \end{aligned} \tag{18}$$

where

$$\varsigma_e = \begin{cases} 1, \min R_G - R_{s,m} \geq \Delta U, \\ 0, \min R_G - R_{s,m} < \Delta U, \end{cases} \tag{19a}$$

$$\varsigma_c = \begin{cases} 1, \|\mathbf{v}_c\|^2 + \sum_{n=1}^{N} \|\mathbf{v}_n\|^2 \leq P_U, \\ 0, \|\mathbf{v}_c\|^2 + \sum_{n=1}^{N} \|\mathbf{v}_n\|^2 > P_U, \end{cases} \tag{19b}$$

$$\varsigma_r = \begin{cases} 1, \sum_{n=1}^{N} c_n - \min_n R_{c,n} \leq 0, \\ 0, \sum_{n=1}^{N} c_n - \min_n R_{c,n} > 0, \end{cases} \tag{19c}$$

where ς_e, ς_c and ς_r are the penalties for the selected actions that do not satisfy the QoS requirements of the PN, the transmit power budget requirements of the SN, and the common message decoding requirements of the SN, respectively.

Within the PPO enabled framework, the collected state information from the UAV transmit beamforming matrix, RIS phase shift matrix, and power splitting ratio serves as input. During each learning iteration, the agent undergoes alternating phases of sampling and optimization over T time slots. For the training phase, let us denote the size of the training layer as L. The input layer size, which is dependent on the number of states, is denoted as Z_I. Additionally, Z_l represents the number of neurons in the l th layer of the DNN. It should be noted that at this point, the computational complexity of the agent at each time step can be calculated as $\mathcal{O}\left(Z_I Z_l + \sum_{l=1}^{L} Z_l Z_{l+1}\right)$. Assuming N^{epi} episodes with each episodes in each mini batch being $T_{\max}$ time steps, each training model iteration is completed when the algorithm reaches convergence. As a result, the total computational complexity in this proposed scheme can be computed as $\mathcal{O}\left(N^{\text{epi}} T_{\max}\left(Z_I Z_l + \sum_{l=1}^{L} Z_l Z_{l+1}\right)\right)$. Clarifying this streamlined time complexity is crucial for ensuring efficient processing, especially in real time applications where fast decision making is crucial.

4 Numerical Results

In this Sect. 4, we present the simulation results to further demonstrate the performance and validate the superiority of our proposed algorithm. Taking the satellite beam center as the origin, the PUs are evenly distributed at a distance of 50 m from the beam center. The position coordinate of the UAV beam center is (5,100), and that of VUs are (5,125), (5,145) and (5,165) [16]. RIS is located near VUs, with the coordinate being (20,125).

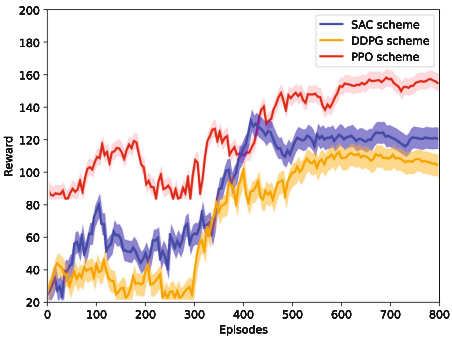

Fig. 2. Convergence comparison between different algorithms.

Figure 2 demonstrates how the proposed PPO based approach converges over the course of the training phase compared to DDPG and SAC algorithms, with a total time slot T set to 100. It can be seen that the proposed PPO enabled algorithm achieves higher rewards, followed by SAC, with DDPG yielding the lowest

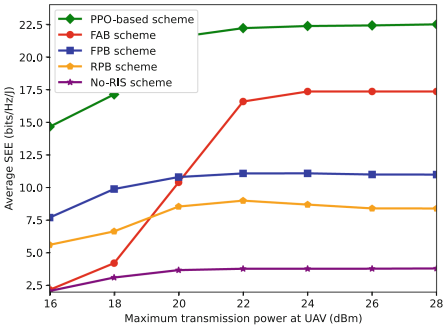

Fig. 3. Comparison of SEE with UAV maximum transmit power.

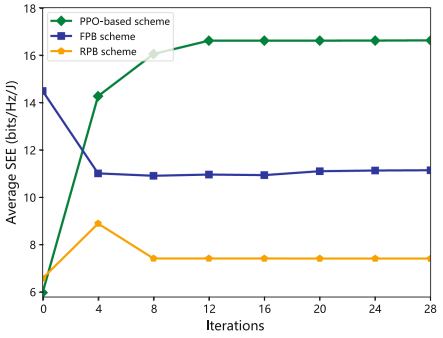

Fig. 4. Comparison of average SEE under different algorithms.

rewards. This visualization highlights the algorithm's steady progress toward optimal performance, clearly showing the effectiveness of our method in enhancing learning and decision making accuracy.

Furthermore, the comparisons between the four benchmark algorithms and our proposed algorithm are provided. In particular, our proposed algorithm is labeled as PPO based scheme. The first benchmark algorithm is marked as FAB (fixed active beamforming) scheme, where the UAV transmit active beamforming is fixed. The second and third benchmark algorithms are labeled as FPB (fixed passive beamforming) and RPB (random passive beamforming), which refer to the optimization problem solved by designing active beamforming and power splitting ratio with fixed and random phase shift matrix. Finally, the fourth benchmark algorithm is marked as No RIS scheme, i.e., without the aid of RIS. Figure 3 shows the variation of the SEE with the maximum transmit power of the UAV. It can be seen that the SEE increases with the increase of the maximum transmit power and tends to converge when the maximum transmit power reaches 22 dBm [17].

Figure 4 shows the variation of different SEE with the number of algorithm iterations under different algorithms. It can be seen that after the 8th iteration,

all the proposed algorithms can converge to a fixed value. Here, the main purpose of this article is to verify the improvement in system performance after introducing RIS. Therefore, we focus on studying three phase shift schemes. We only list three RIS phase shift change schemes, which are optimizing RIS phase shift using PPO algorithm, FPB scheme, and RPB scheme. It can be seen that the SEE of the PPO algorithm mentioned is higher than the other two algorithms, which verifies the superiority of this algorithm [18].

Figure 5 clearly illustrates the trend of the average worst secrecy rate as the number of secondary vehicle users varies. The graph reveals that, as the user count increases, the average worst secrecy rate corresponding to all optimization strategies exhibits a decreasing trend. The average worst secrecy rate is defined as the expected secrecy rate between legitimate users and eavesdropping users, computed across all possible channel conditions. This phenomenon underscores the challenges posed by secure communication in multi user environments, specifically the overall deterioration in security performance stemming from user interference and competition for channel resources.

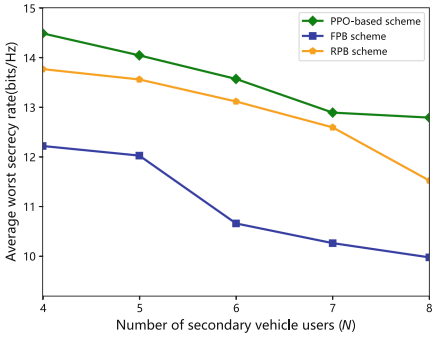

Fig. 5. Comparison of the average worst secrecy rate under different numbers of secondary vehicle users.

5 Conclusion

In the paper, we proposed a new infrastructure by combining SATINs with RSMA for supporting massive connectivity, and with RIS for enhancing the security. Different from existing works focusing on SE or EE separately, we formulated a SEE maximization problem, and then proposed a unified PPO learning framework based on DRL to further handle both continuous and discrete optimization variables. Specifically, this framework designed state action pairs to determine UAV transmit beamforming, RIS reflected beamforming, and power splitting ratio. The simulation results showed that the proposed scheme outperforms the benchmark scheme in terms of achievable SEE performance and computational complexity.

References

1. Liu, R., et al.: RIS-empowered satellite-aerial-terrestrial networks with PD-NOMA. IEEE Commun. Surv. Tutor. **26**, 2258–2289 (2024). https://doi.org/10.1109/COMST.2024.3393612
2. An, K., et al.: Secure transmission in cognitive satellite terrestrial networks. IEEE J. Sel. Areas Commun. **34**(11), 3025–3037 (2016)
3. Ma, R., et al.: Covert mmWave communications with finite blocklength against spatially random wardens. IEEE Internet Things J. **11**(2), 3402–3416 (2024)
4. Xu, J., et al.: Sum secrecy rate maximization for IRS-aided multi-cluster MIMO-NOMA Terahertz systems. IEEE Trans. Inf. Forensics Secur. **18**, 4463–4474 (2023)
5. Wu, Q., Zhang, R.: Towards smart and reconfigurable environment: intelligent reflecting surface aided wireless network. IEEE Commun. Mag. **58**(1), 106–112 (2020)
6. Li, X., et al.: Exploiting benefits of IRS in wireless powered NOMA networks. IEEE Trans. Green Commun. Netw. **6**(1), 175–186 (2022)
7. Guo, K., et al.: Outage performance of RIS-assisted cognitive non-terrestrial network with NOMA. IEEE Trans. Veh. Tech. **73**(4), 5953–5958 (2024)
8. Sun, Y., et al.: Energy-efficient hybrid beamforming for multilayer RIS-assisted secure integrated terrestrial-aerial networks. IEEE Trans. Commun. **70**(6), 4189–4210 (2022)
9. Li, X., et al.: Secure communication of active RIS assisted NOMA networks. IEEE Trans. Wirel. Commun. **23**(5), 4489–4503 (2024)
10. Huang, C., et al.: Reconfigurable intelligent surfaces for energy efficiency in wireless communication. IEEE Trans. Wirel. Commun. **18**(8), 4157–4170 (2019)
11. Dong, R., et al.: Secure transmission design of RIS enabled UAV communication networks exploiting deep reinforcement learning. IEEE Trans. Veh. Tech. **73**, 8404–8419 (2024). https://doi.org/10.1109/TVT.2024.3357821
12. Zhang, R., et al.: Energy efficiency maximization in RIS-assisted SWIPT networks with RSMA: a PPO-based approach. IEEE J. Sel. Areas Commun. **41**(5), 1413–1430 (2023)
13. Lin, Z., et al.: Refracting RIS-aided hybrid satellite-terrestrial relay networks: joint beamforming design and optimization. IEEE Trans. Aerosp. Electron. Syst. **58**(4), 3717–3724 (2022)
14. Mozaffari, M., Saad, W., Bennis, M., Debbah, M.: Wireless communication using unmanned aerial vehicles (UAVs): optimal transport theory for hover time optimization. IEEE Trans. Wirel. Commun. **16**(12), 8052–8066 (2017)
15. An, H., Wang, L.: Robust topology generation of internet of things based on PPO algorithm using discrete action space. IEEE Trans. Ind. Inform. **20**(4), 5406–5414 (2024)
16. Hao, W., et al.: Securing reconfigurable intelligent surface-aided cell-free networks. IEEE Trans. Inf. Forensics Secur. **17**, 3720–3733 (2022)
17. Zhou, C., et al.: Energy-efficient maximization for RIS-aided MISO symbiotic radio systems. IEEE Trans. Veh. Technol. **72**(10), 13689–13694 (2023)
18. Wu, M., et al.: Deep reinforcement learning-based energy efficiency optimization for RIS-aided integrated satellite-aerial-terrestrial relay networks. IEEE Trans. Commun. **72**(7), 4163–4178 (2024)