Energy Engineering

Tech Science Press

Doi:10.32604/ee.2025.071023

ARTICLE

Secure Rate Maximization for UAV-RIS-Aided IoT Network in Smart Grid

Jian Wu*, **Xiaowei Hao and Chao Han**

Information and Communication Branch, State Grid Shanxi Electric Power Company Limited, Taiyuan, 030021, China

*Corresponding Author: Jian Wu. Email: wujiansgcc@163.com

Received: 29 July 2025; Accepted: 23 September 2025; Published: 27 February 2026

ABSTRACT: Owing to the development of communication technologies and control systems, the integration of numerous Internet of Things (IoT) nodes into the power grid has become increasingly prevalent. These nodes are deployed to gather operational data from various distributed energy sources and monitor real-time energy consumption, thereby transforming the traditional power grid into a smart grid (SG). However, the openness of wireless communication channels introduces vulnerabilities, as it allows potential eavesdroppers to intercept sensitive information. This poses threats to the secure and efficient operation of the IoT-driven smart grid. To address these challenges, we propose a novel scenario that incorporates an Unmanned Aerial Vehicle (UAV) as a relay gateway for multiple authorized smart meters. This scenario is further enhanced by the integration of Reconfigurable Intelligent Surface (RIS) technology, which dynamically adjusts the direction of information transmission. Our objective is to maximize the secure rate within this UAV-RIS-aided system with multiple authorized smart meters and an eavesdropper based on physical layer security (PLS) techniques. We formulate the problem of secure rate maximization by jointly optimizing the active beamforming of the UAV, the passive beamforming of the RIS, and the UAV's trajectory. To solve this complex optimization problem, we introduce the Twin Soft Actor-Critic (TSAC) algorithm. This algorithm employs a dual-agent framework, where Agent 1 focuses on optimizing the beamforming for both the UAV and the RIS, while Agent 2 concurrently searches for the optimal trajectory of the UAV. Simulation results demonstrate the TSAC algorithm significantly enhances the secure rate of the system, achieving faster convergence and higher rewards under the worst communication conditions. The TSAC algorithm consistently outperforms the Twin Deep Deterministic Policy Gradient (TDDPG) and Twin Delayed Deep Deterministic Policy Gradient (TTD3) algorithms. Furthermore, the TSAC algorithm exhibits robust performance when the distribution of smart meters follows a Gaussian distribution, further validating its practical applicability and effectiveness in real-world scenarios.

KEYWORDS: Smart grid; Internet of Things; unmanned aerial vehicles; reconfigurable intelligent surface; eavesdropper; secure capacity

1 Introduction

Recent years have witnessed the evolution of communication and power control systems, various distributed energy sources (DESs), such as wind power, solar photovoltaics, and nuclear energy, are being increasingly integrated into the grid. Additionally, a multitude of Internet of Things (IoT) nodes, including sensors, actuators and smart meters, have been extensively introduced into the power grid [1]. These smart meters enable the bidirectional flow of operational and real-time energy consumption data within the grid, thereby transforming the traditional grid into a smart grid (SG) [2]. However, the openness of wireless communication channels may allow eavesdroppers to intercept sensitive information, such as customers' energy consumption habits, or to tamper with data. This can lead to privacy breaches or malicious activities against service providers, which in turn poses a threat to the normal operation of the SG [3]. Consequently, the importance of transmission security in the operation of SG has been growing [4].

Considering the transmission characteristics, Physical Layer Security (PLS) technology has emerged as an alternative to traditional encryption methods for SG. It provides secure communication for IoT devices from an information-theoretic perspective [5]. By establishing a secure communication channel between authorized users, the transmitter can provide the authorized receiver with an information rate, known as the channel secrecy rate, ensuring that eavesdroppers cannot obtain any useful information [6]. The authors proposed a Relay-Assisted Vectorized secure transmission (RAV) algorithm to mitigate active attacks. This RAV algorithm leverages the PLS framework to improve secure IoT communication [7]. Besides, a novel jamming scheme was proposed to confuse potential eavesdroppers in Wireless Sensor Networks (WSNs) for SG applications [8]. Furthermore, a gradient ascent algorithm was employed to boost the secure rate in IoT networks within the SG [9]. In combination with blockchain technology, a novel secure offloading algorithm was proposed to achieve energy-efficient services in IoT [10]. To achieve the secure communication economically, the authors used the power of 5G base stations to construct interference against eavesdroppers in SGs, thereby improving the secure rate [11]. However, interference, channel impairments, and the unpredictable wireless conditions arising from the dynamic and heterogeneous nature of the SG environment pose significant challenges to achieving robust PLS performance.

Reconfigurable Intelligent Surface (RIS) technology can effectively enhance the received signal power while suppressing signals intended for potential eavesdroppers [12]. Formulas for the secrecy outage probability (SOP) and the average secrecy capacity (ASC) of RIS-aided wireless communication systems have been derived [13]. Considering the phase errors within the system, the authors derived formulas for the SOP and secrecy capacity, demonstrating that system security can be enhanced by adjusting the RIS settings under phase errors [14]. When the wide area network (WAN) of the SG adopted a three-hop transmission mode, an RIS-assisted scheme was proposed. By optimizing the configuration of the RIS, the security of SG communications was significantly enhanced [15]. In complex scenarios with impulsive noise and interferers, reliable communication links were established in WSNs within SGs by optimizing the RIS configuration [16]. Furthermore, the security performance of RIS-aided communication systems under discrete phase control was investigated [17]. When faced with multiple eavesdroppers, approximate formulas for the SOP were derived for both suboptimal and optimal cases, and two distinct scheduling schemes were identified [18]. Furthermore, RIS was incorporated into a specific SG scenario involving smart meters, gateways and eavesdroppers to enhance the ASC of the wireless communications [19].

When gateways or base stations in SGs fail due to harsh environmental conditions, communication dead zones can emerge. Unmanned Aerial Vehicles (UAVs) can serve as data collectors or mobile gateways to expand coverage in SGs. However, the links between UAVs and ground nodes remain vulnerable to eavesdropping. Various algorithms have been proposed to ensure secure communication between UAVs and ground nodes via PLS. Lee et al. employed a cooperative UAV to transmit jamming signals, thereby improving the secure rate for ground users. By optimizing UAV's trajectory, transmit power and user scheduling based on block successive upper-bound minimization (BSUM), they maximized the minimum secure rate for ground users' [20]. Xiao et al. introduced UAVs as mobile relays for secure communications and optimized relay strategies and resource allocation together to enhance secrecy energy efficiency (SEE) [21]. Li et al. utilized partial eavesdropper information to design a cooperative jamming approach, enhancing security between UAVs and users [22]. Furthermore, the security and efficiency of multi-UAV communications have been improved through optimized resource allocation under a multi-eavesdropper scenario [23]. By combing RIS with UAVs to achieve two-hop communication, performance was enhanced by adjusting the phase of the RIS [24]. Furthermore, an optimized algorithm was proposed in a UAV-RIS network, where the security performance was enhanced by optimizing the trajectories of UAV and the passive beamforming of RIS [25]. In another complex scenario where eavesdroppers were mobile UAVs, secure communication was achieved by optimizing dynamic trajectories and communication strategies for authorized UAVs [26]. An efficient iterative algorithm was devised to enhance SEE in a UAV-aided IoT network by optimizing scheduling, power allocation, and UAV trajectories [27]. In another UAV-aided system with artificial noise injection, the suboptimal solution for the UAV's trajectory, transmit power, and artificial noise power was obtained via the successive convex approximation method, resulting in a faster convergence rate [28]. Moreover, an efficient iterative algorithm was designed to enhance the average security rate by optimizing the beamforming power, reflection phase shifts, and UAV trajectory [29]. Zhu et al. innovatively proposed a dual-UAV auxiliary communication scheme, employing continuous convex approximation and block coordinate descent (BCD) algorithms to achieve collaborative optimization of the UAV base station trajectory, the interference UAV's trajectory, transmit power control, and user scheduling, thereby maximizing the minimum average secure rate for ground users [30]. Additionally, a method based on BCD was devised to jointly optimize transmit power and UAV trajectory, maximizing the total secrecy capacity received by legitimate IoT devices [31].

With the advancement of artificial intelligence (AI) technologies, deep reinforcement learning (DRL) algorithms, which obtain sequential decisions through interaction with the environment, have been extensive applied in wireless networks. Qian et al. employed a double deep Q-network (DDQN) algorithm to jointly control of UAV trajectory and transmit power, effectively improving the security rate of UAV-enabled systems [32]. Guo et al. constructed a twin deep deterministic policy gradient (DDPG) algorithm to maximize the secret sum rate (SSR) of authorized users under worst-case channel conditions in RIS-UAV aided networks [33]. To improve the DDPG algorithm, Tham et al. developed a twin delayed deep deterministic policy gradient (TD3) algorithm to further enhance the SSR [34]. Furthermore, a dual proximal policy optimization (DPPO) framework was designed to maximize the SEE for legitimate users in RIS-UAV aided networks [35].

In addition to these advancements, some researchers have designed anonymous and reliable authentication protocols to ensure secure and reliable information exchange between smart meters and grid operators [36]. Tanveer et al. developed a lightweight authenticated encryption access algorithm for smart meters, achieving secure communication with grid operators through lower resource consumption [37].

Based on the existing research results, this paper examines the enhancement of secure rate in UAV-RIS-aided SGs with multiple authorized SMs and a single eavesdropper. Taking into account the randomness and unknown channel state conditions under the worst communication scenarios, we construct problem aimed at maximizing secrecy rate. This non-convex problem can be formulated as a continuous Markov Decision Process (MDP). To address this, we design a twin Soft Actor Critic (SAC) algorithm to maximize secure rate. The contributions are as follows:

We consider the enhancement of secure capacity in UAV-RIS-aided SG systems with multiple authorized SMs and one eavesdropper. To enable more IoT nodes access the system, we also consider the millimeter-wave channels.

By jointly optimizing the active beamforming of the UAV, the passive beamforming of the RIS, and the UAV's trajectory, we formulate a secure rate maximizing problem within the SG context. The problem is treated as a continuous MDP.

We design the Twin Soft Actor Critic (TSAC) algorithm to solve the optimization problem. Two agents are introduced within the twin SAC framework: Agent 1 searches for optimal beamforming strategies while Agent 2 searches for the optimal trajectory of the UAV.

Simulation results demonstrate that the TSAC algorithm can significantly enhance the secure rate of the system, with faster convergence speeds under the worst communication conditions, outperforming both the TDDPG and TTD3 algorithms. Moreover, the TSAC algorithm performs well when the SMs distribute following Gaussian distribution.

2 System Model

The UAV-RIS-aided SG communication system studied in this paper is shown in Fig. 1. The UAV acts as a mobile gateway, and the RIS is adopted to enhance the PLS performance. There are V legitimate SMs and A eavesdroppers. All legitimate SMs and eavesdroppers are equipped with a single antenna.

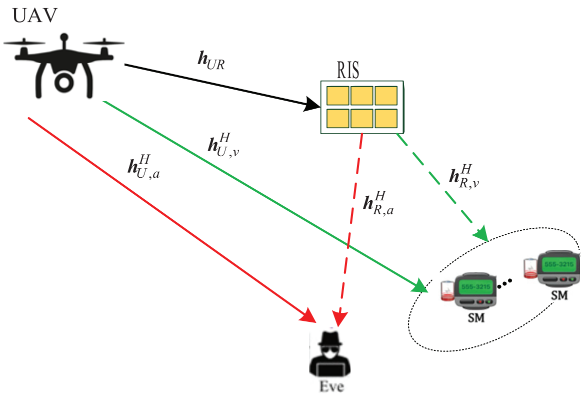
UAV
$\boldsymbol{h}_{UR}$
RIS
$\boldsymbol{h}_{U,v}^{H}$
$\boldsymbol{h}_{U,a}^{H}$
$\boldsymbol{h}_{R,a}^{H}$
$\boldsymbol{h}_{R,v}^{H}$
SM
SM
Eve

Figure 1: System model

There is a L-element Uniform Linear Array (ULA), which is used to achieve high-precision beamforming, on the UAV. The RIS is equipped with M reflective elements of a Uniform Planar Array (UPA), which can dynamically regulate the propagation of electromagnetic waves. $V = \{1, 2, \ldots, V\}$ and $A = \{1, 2, \ldots, A\}$ represent the sets of legitimate SMs and eavesdroppers, respectively. The entire flying time T is divided into N time slots, denoted as $T = N\delta_{\mathrm{n}}$, where δ_{n} represents each individual time slot. The coordinates of the RIS are $w_R = (x_R, y_R, z_R)^T$. The coordinates of the legitimate SMs and eavesdroppers in time slot n are represented as $w_i[n] = (x_i[n], y_i[n], z_i[n])^T, \forall i \in V \cup A$.

Assuming that the flying altitude is H_U, then the coordinates of UAV in time slot n are $q_U[n] = (x_U[n], y_U[n], H_U)^T$, which must follow the following constraints:

$$\begin{cases} q_U[0] \equiv (x_U[0], y_U[0], H_U)^T \\ |x_U[n]|, |y_U[n]| \le B, n = 1, \ldots, N \\ \sqrt{\|q_U[n+1] - q_U[n]\|^2} \le D_{\max}, n = 1, \ldots, N \end{cases} \tag{1}$$

Among them, $q_U[0]$ represents the initial coordinates of UAV, $D_{\max}$ is the maximum distance UAV can move within time slot n, and B is the movement boundary of the UAV. The speed of UAV in time slot n is

$$\|v_U[n]\| = \sqrt{\|q_U[n+1] - q_U[n]\|^2} / \delta_n \tag{2}$$

Considering a rotary-wing UAV, based on the speed $v_U[n]$, the energy consumed during propulsion process can be derived [38]:

$$\begin{aligned} E_p[n] &\approx \delta_n \left(P_0 + \frac{3P_0 v_U[n]^2}{U_{\text{tip}}^2} + \frac{1}{2} d_0 \rho s A_r v_U[n]^3 \right) \\ &+ \delta_n P_i \left(\sqrt{1 + \frac{v_U[n]^4}{4v_0^4}} - \frac{v_U[n]^2}{2v_0^2} \right)^{\frac{1}{2}} \end{aligned} \tag{3}$$

The constants P_0 and P_i represent blade profile power and induced power of UAV during hovering, respectively. U_{tip}^2 is the tip speed of rotor blade, v_0 is average rotor induced speed during hovering. d_0 is the fuselage drag ratio, s represents solidity of the rotor, ρ is air density, A_r is the area of the rotor disk.

The channels are millimeter-wave channels [39]. Let $\boldsymbol{h}_{U,v} \in \mathbb{C}^{L\times 1}$, $\boldsymbol{h}_{R,v} \in \mathbb{C}^{M\times 1}$, $\boldsymbol{h}_{R,a} \in \mathbb{C}^{M\times 1}$ and $\boldsymbol{h}_{UR} \in \mathbb{C}^{M\times L}$ represent channel gains from UAV to the vth legitimate SM, from RIS to the vth legitimate SM, from RIS to the ath eavesdropper, and from UAV to RIS, respectively. Therefore, the channel from UAV to a legitimate SM or an eavesdropper is represented as $\boldsymbol{H}_{C,i} = diag\left(\boldsymbol{h}_{U,i}^H\right) \boldsymbol{h}_{UR}, \forall i \in A \cup P$.

Additionally, $\boldsymbol{\theta} = diag\left(\beta_1 e^{j\theta_1}, \beta_2 e^{j\theta_2}, \ldots, \beta_M e^{j\theta_M}\right)$ represents the beamforming matrix of RIS. Let $\theta_m \in [0, 2\pi)$ and $\beta_m \in [0,1]$ denote the phase shift and amplitude reflection coefficient of the mth reflective element, respectively, where $m \in \{1, 2, \ldots, M\}$. β_m is assumed to 1. By vectorizing the beamforming matrix of RIS as $\boldsymbol{\Psi} = vec(\boldsymbol{\theta})$, the channel coefficients from UAV to all signal receivers can be expressed as $\boldsymbol{H}_{C,i} = \left\{\boldsymbol{h}_{U,i}^H + \boldsymbol{\Psi}^H \boldsymbol{H}_{C,i} \,|\forall i \in V \cup A\right\}$. Thus, the signal received by the ith legitimate SM or eavesdropper from UAV is represented as:

$$y_i = \left(\boldsymbol{h}_{U,i}^H + \boldsymbol{\Psi}^H \boldsymbol{H}_{C,i}\right) \boldsymbol{G} s + n_i, \forall i \in V \cup A \tag{4}$$

In the above equation, s is the transmit symbol from UAV, and s is confined by

$$E\left[|s_v|^2\right] = 1 \tag{5}$$

$\boldsymbol{G} \in \mathbb{C}^{L\times V}$ is the beamforming matrix of UAV, n_i is the noise, g_v represents the vth row of $\boldsymbol{G}$. Therefore, the data communication rate that the vth legitimate SM can achieve in time slot n is

$$R_v^u[n] = \log_2\left(1 + \frac{\left|\left(\boldsymbol{h}_{U,v}^H + \boldsymbol{\Psi}^H \boldsymbol{H}_{C,v}\right) g_v\right|^2}{\sum_{v' \in V\backslash v} \left|\left(\boldsymbol{h}_{U,v}^H + \boldsymbol{\Psi}^H \boldsymbol{H}_{C,v}\right) g_{v'}\right|^2 + n_v^2}\right) \tag{6}$$

If the ath eavesdropper eavesdrops on the signal of the vth legitimate SM, the achievable rate is

$$R_{a,v}^e[n] = \log_2\left(1 + \frac{\left|\left(\boldsymbol{h}_{U,a}^H + \boldsymbol{\Psi}^H \boldsymbol{H}_{C,a}\right) g_v\right|^2}{\sum_{v' \in V\backslash v} \left|\left(\boldsymbol{h}_{U,a}^H + \boldsymbol{\Psi}^H \boldsymbol{H}_{C,a}\right) g_{v'}\right|^2 + n_a^2}\right) \tag{7}$$

Let $[x]^+ = \max(0, x)$, the secure rate from UAV to the vth legitimate SM is

$$R_v^{\sec}[n] = \left[R_v^u[n] - \max_{\forall a} R_{a,v}^e[n]\right]^+ \tag{8}$$

3 Problem Description

In order to enhance the secure rate between legitimate SMs and UAV, a joint optimization problem is constructed as

$$\begin{aligned} &\max_{\boldsymbol{G},\boldsymbol{\theta},\boldsymbol{Q}} \sum_{n=1}^{N} R_v^{\sec}[n] \\ s.t.\quad & C_1: \quad (1) \\ & C_2: \quad \Pr\left\{R_v^{\sec} \ge R_v^{sec,th}\right\} \ge 1-\rho_v, \forall v \in V \\ & C_3: \quad \mathrm{Tr}\left(\boldsymbol{G}\boldsymbol{G}^H\right) \le P_{\max} \\ & C_4: \quad \theta_m \in [0,2\pi), m=\{1,2,\dots,M\} \end{aligned} \tag{9}$$

Constraint C_2 represents that the secure rate of the vth legitimate SM is not lower than the probability threshold $1-\rho_v$. The $P_{\max}$ in constraint C_3 represents the maximum transmit power provided by UAV. The non-convex constraints in C_1, C_3, and C_4, as well as the dynamically changing wireless communication environment, make it complex to solve this problem. A method based on DRL is proposed to solve this problem.

3.1 Problem Solving

We equivalently transform the optimization problem to a Markov decision process for agent. The TSAC algorithm is proposed as shown in Fig. 2. TSAC adopts two SAC agents to separately optimize the two decoupled sub-problems. SAC Agent 1 is used to optimize the beamforming for UAV and RIS. SAC Agent 2 is used to optimize UAV's trajectory.

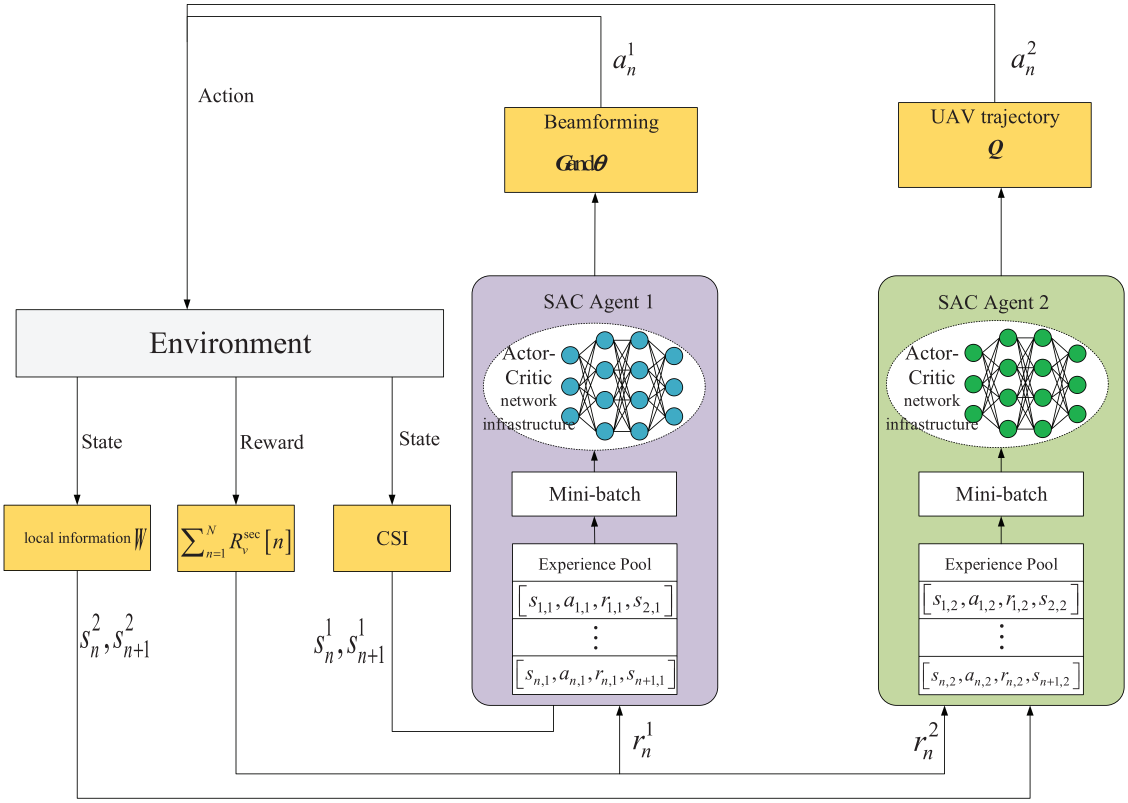
Action
a_n^1
a_n^2
Beamforming
G and θ
UAV trajectory
Q
Environment
SAC Agent 1
SAC Agent 2
Actor-Critic network infrastructure
Actor-Critic network infrastructure
State
Reward
State
Mini-batch
Mini-batch
local information W
$\sum_{n=1}^{N} R_v^{\sec}[n]$
CSI
Experience Pool
$[s_{1,1}, a_{1,1}, r_{1,1}, s_{2,1}]$
$[s_{n,1}, a_{n,1}, r_{n,1}, s_{n+1,1}]$
Experience Pool
$[s_{1,2}, a_{1,2}, r_{1,2}, s_{2,2}]$
$[s_{n,2}, a_{n,2}, r_{n,2}, s_{n+1,2}]$
s_n^2, s_{n+1}^2
s_n^1, s_{n+1}^1
r_n^1
r_n^2

Figure 2: TSAC framework

3.2 TSAC

SAC adopts actor-critic architecture containing an Actor network π_ϕ, two Critic networks with parameter Q_{θ_1} and Q_{θ_2}, two target Critic networks with parameter $Q_{\theta_1'}$ and $Q_{\theta_2'}$.

At each time step, the Actor network draws an action, the agent executes it, collects the reward, and observes the resulting successor state; this experience is promptly stored in the buffer. Then, a mini-batch experience is selected from the buffer. The policy network π_ϕ emits a stochastic action distribution. Critic network Q_{θ_1} and Q_{θ_2} estimate Q-values for this state–action pair. The lower Q-value is selected as target Q-value. The target Critics computes the target Q-value. The parameters Q_{θ_1} and Q_{θ_2} are updated by minimizing the squared Bellman error. π_ϕ is updated by maximizing the weighted sum of expected return and entropy, where the entropy regularization is used to balance the diversity and greediness of the policy. The entropy temperature parameter α is updated by optimizing a loss function related to the target entropy to better balance exploration and exploitation. Finally, $Q_{\theta_1'}$ and $Q_{\theta_2'}$ are updated by a soft update mechanism. By iteratively repeating the above process in a loop until the policy network converges and the agent's performance reaches the optimal level.

As shown in Fig. 2, the TSAC framework contains twin SAC network, with the name of SAC Agent 1 and SAC Agent 2. SAC Agent 1 takes the channel state information (CSI) as its state input and generates the optimal beamforming matrix $\boldsymbol{G}$ and $\boldsymbol{\theta}$ for both UAV and RIS. The state space, action space, and reward function for Agent 1 are as following:

State space

The state space for SAC Agent 1 is defined as S_1. It contains the aggregated CSI predicted from UAV to SMs and eavesdroppers in each time slot, denoted as C_n. For the state at time slot n, denoted as s_n^1, S_1 is expressed as follows:

$$S_1 = \{C_1, C_2, \ldots, C_n\} = \left\{s_1^1, s_2^1, \ldots, s_n^1\right\} \tag{10}$$

Action space

SAC Agent 1 will generate $\boldsymbol{G}$ and $\boldsymbol{\theta}$ as actions, denoted as (G_n, θ_n), and the action space is defined as A_1. To address the input of complex numbers, $\boldsymbol{G} = Re\{\boldsymbol{G}\} + Im\{\boldsymbol{G}\}$ and $\boldsymbol{\theta} = Re\{\boldsymbol{\theta}\} + Im\{\boldsymbol{\theta}\}$ are respectively the real and imaginary parts. For the action at time slot n, denoted as a_n^1, A_1 is expressed as follows:

$$A_1 = \{(G_1, \theta_1), (G_2, \theta_2), \ldots, (G_n, \theta_n)\} = \left\{a_1^1, a_2^1, \ldots, a_n^1\right\} \tag{11}$$

Reward function

The reward function is formulated as follows:

$$r_n = \tanh\left(\sum\nolimits_{v=1}^{V} R_v^{\sec}[n] - c_1 p_m - c_2 p_r - c_3 p_g - c_4 p_e\right) \tag{12}$$

In the above equation, p_m, p_r, and p_g are the penalty terms for the constraints C_1, C_2, and C_3 not being satisfied, respectively, and p_e is the penalty for high energy consumption. c_1, c_2, c_3 and c_4 are the corresponding weight coefficients.

As shown in Fig. 2, SAC Agent 2 is employed to compute the optimal UAV trajectory $\mathbf{Q}$ by taking the local information W as input. Here, W is $W \triangleq \{q[n]\} \cup \{w_i[n] \,|\, \forall i \in V \cup A\}$. Similarly, the state space, action space, and reward function are as follows:

State space

SAC Agent 2 takes W as input, denoted as W_n. The state space is defined as S_2 For the state at time slot n, denoted as s_n^2, S_2 is expressed as follows:

$$S_2 = \{W_1, W_2, \ldots, W_n\} = \left\{s_1^2, s_2^2, \ldots, s_n^2\right\} \tag{13}$$

Action space

In each time slot n, the SAC Agent 2 generates a flight direction$d\left[n\right]$, which is represented in the three-dimensional Cartesian coordinate system. Based on $d\left[n\right]$, the coordinates of UAV at the next time slot is calculated as $q_U\left[n\right] = q_U\left[n-1\right] + d\left[n\right]$. After N time slots, the complete UAV trajectory can be represented as $\boldsymbol{Q} = \{q_U\left[0\right], q_U\left[1\right], \ldots, q_U\left[N\right]\}$, denoted as $\mathrm{Q_n}$, and the action space is defined as A_2. For the action at time slot n, denoted as a_n^2, A_2 is expressed as follows:

$$A_2 = \{\mathrm{Q}_1, \mathrm{Q}_2, \ldots, \mathrm{Q}_n\} = \left\{a_1^2, a_2^2, \ldots, a_n^2\right\} \tag{14}$$

Reward function

The reward function of SAC Agent 2 is the same as that defined in Eq. (12).

3.3 Problem Solving Optimization of Secrecy Communication Rate Based on TSAC Algorithm

The TSAC algorithm designed in this chapter is shown in Algorithm 1. The algorithm initializes the Actor and Critic networks of two SAC algorithms to learn the optimal policy and value function for solving problems in continuous action spaces. During the training process, these networks are continuously updated to improve the quality and efficiency of decision-making. Meanwhile, the algorithm also initializes target networks and employs a soft update to bring parameters of the target network closer to those of main network, thereby enhancing training stability. In addition, the algorithm uses an experience replay buffer to store the interaction data with the environment, which is subsequently used for network updates. After setting the key hyper parameters such as the learning rate, discount factor, entropy regularization coefficient, and soft update coefficient, the algorithm enters the main training loop.

At the beginning of each round, the positions of UAV and all legitimate SMs are reset. In the main training loop, the algorithm interacts with the environment over multiple episodes, each consisting of multiple time steps. At each time step, the agent chooses action based on current policy and adds random noise to enhance exploration. These actions are then executed to obtain rewards and the next state information from the environment. The experience data are stored in the experience replay. When the number of samples in the buffer reaches a certain batch size, the algorithm randomly samples from it to update the networks. This includes updating critic networks to optimize the value function predictions, updating actor networks to improve policy, and smoothly updating the parameters of the target networks via soft updating mechanism. As training progresses, the agent continuously learns and improves, eventually finding an optimal policy that maximizes cumulative rewards. Entropy regularization helps the algorithm balance exploration and exploitation. This process iterates continuously. Training begins when the amount of stored data exceeds the capacity of the experience replay buffer and continues until the task is completed.

The details of TSAC algorithm are shown in Algorithm 1.

Algorithm 1: UAV trajectory planning, beamforming for UAV and RIS optimization based on the TSAC algorithm

```
Initialize the Actor Networks, Critic Networks, Target Critic Networks, Replay Buffer, and noise generator for two SAC algorithms.
Set the learning rate, discount factor, entropy regularization coefficient, and soft update coefficient.
for episode = 1, 2, . . ., Nep do
    Reset the positions of UAV and all SMs.
    for step = 1, 2, . . ., $N_{step}$ do
        Interact with the environment to obtain states $s_n^1$ and $s_n^2$.
        Select actions $a_n^1$ and $a_n^2$ according to the current policy, add random noise, execute the actions, and obtain the rewards $r_n$ and the next states ($s_{n+1}^1$ and $s_{n+1}^2$).
        Place $\left(s_n^1, a_n^1, r_n, s_{n+1}^1\right)$ and $\left(s_n^2, a_n^2, r_n, s_{n+1}^2\right)$ into their respective experience replay buffers.
        if the size of the experience replay buffer reaches the batch size
            Randomly sample a batch of experiences from the experience replay buffer
            Calculate the target values and use the loss function to update the critic network
            Update the actor network based on the feedback from the critic network
            Update the parameters of the target critic network
        end if
        Update the current state to the next state.
    end for
end for
```

4 Results and Discussion

To verify the effectiveness of TSAC algorithm, this section evaluates the performance through simulation experiments. Based on the parameter model in [34], the initial positions of the legitimate SMs are (25, 25, 0 m) and (4, 47, 0 m), respectively. The UAV's initial position is (0, 25, 50 m), while the positions of RIS and eavesdropper are (0, 50, 12.5 m) and (47, −4, 0 m), respectively. The remaining parameter configurations are $D_{\max} = 0.25$ m, $\delta_n = 0.1$ ms, $P_{\max} = 30$ dBm, $M = 4$, $L = 4$, $V = 2$, $A = 1$, $\beta_0 = -30$ dB. The path loss factors for each link are $\alpha_{U,v} = \alpha_{U,a} = 3.5$, $\alpha_{R,v} = \alpha_{R,a} = 2.8$, $c_{UR} = 2.2$ The specific parameter settings are shown in Tables 1 and 2.

Table 1: Parameter settings

Parameters	Value
Blade profile power P_0	582.65 W
Induced power P_i	790.67 W
Rotor blade tip speed U_{tip}	200 m/s
Air density $\boldsymbol{\rho}$	1.225 kg/m^3
Fuselage drag ratio d_0	0.3
Rotor solidity s	0.05
Rotor disk area A_r	0.97 m^2
Average rotor induced speed v_0	2.567 m/s

Table 2: Hyperparameters of TSAC

Hyperparameters	Value
Size of SAC Agent 1	27 × 800 × 600 × 515 × 256 × 20
Size of SAC Agent 2	3 × 400 × 300 × 256 × 128 × 2
Actor learning rate	0.0001
Critic learning rate	0.001
Training episode length N_{ep}	300
Time steps N	100
Batch size N_b	64
Experience pool size	30,000
Actor update interval	2

The TSAC algorithm is compared with the following two schemes.

(1) The TTD3 algorithm employed in [34].

(2) The TDDPG algorithm employed in [33].

Fig. 3 compares average reward over different algorithms. The simulation results show that although the TSAC algorithm experiences significant fluctuations in the early stages of training, its reward value quickly stabilizes at nearly 80 episodes. In contrast, the TTD3 algorithm shows faster initial reward growth but converges slightly slower than TSAC, stabilizing after about 100 episodes. The TDDPG algorithm experiences rapid initial reward growth but begins to decline around episode 80, with significant fluctuations. The efficiency of TSAC is attributed to the balance between its optimization strategy and exploration mechanism.

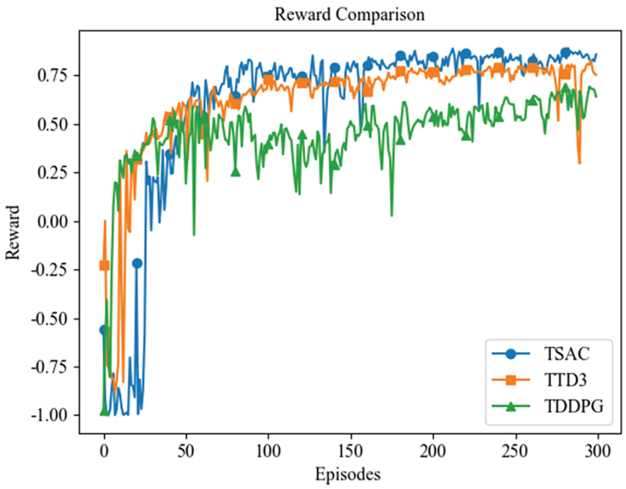

Figure 3: Average reward over different algorithms

Figs. 4 and 5 show the performance of user (SM) capacity and secure capacity over three algorithms, respectively. The two figures exhibit the same trend with different capacity value. As shown in Figs. 4 and 5, TSAC algorithm demonstrates the best performance: although there are minor fluctuations in the early stages of training, it maintains stable growth and eventually reaches the highest level, reflecting its effectiveness in capacity optimization. The TTD3 algorithm exhibits different learning characteristics: although there is a brief decline in the initial stage, it gradually recovers with the increase of training rounds, showing a significant acceleration phenomenon, but the final user capacity value is still lower than that of TSAC. In comparison, the optimization effect of the TDDPG algorithm is the most limited. Although it grows rapidly in the early stage, its growth rate slows down in the middle stage and it always lags behind the other two algorithms, with a significantly lower final performance, reflecting its insufficient convergence efficiency. Through the comparative analysis of the system performance, the experimental results verify the superior performance of the TSAC algorithm in improving secure capacity.

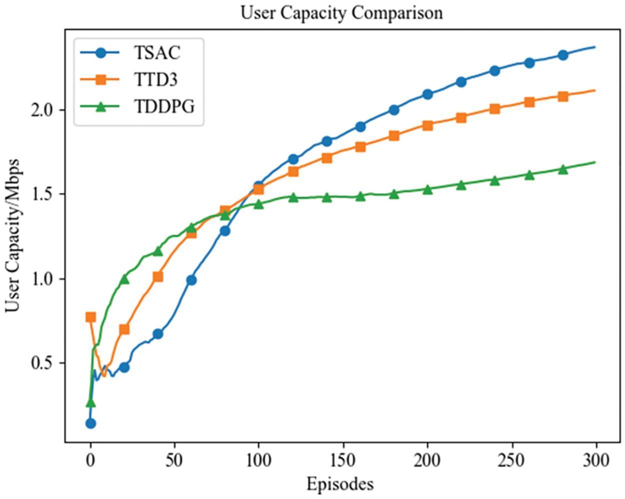

Figure 4: User capacity over different algorithms

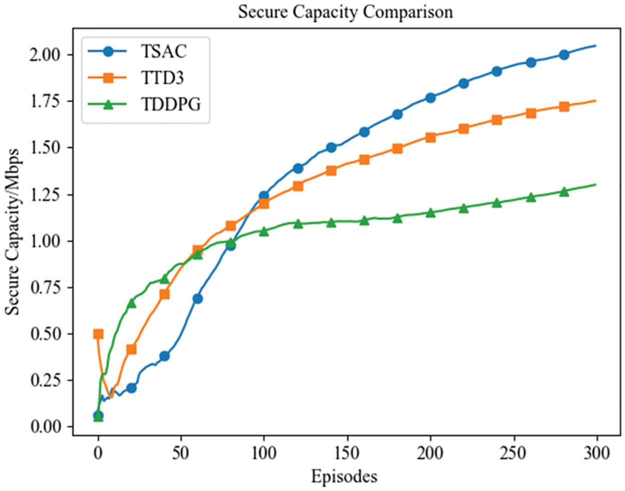

Figure 5: Secure capacity over different algorithms

Fig. 6 shows the trend of SSR with the number of training episodes, with and without the introduction of an energy penalty term. TSAC and TTD3 represent the cases without the energy penalty term, while TSAC (EP) and TTD3 (EP) represent the cases with the energy penalty term introduced. It can be seen from Fig. 6 that throughout the training process, the TSAC algorithm shows higher secrecy rate whether or not the energy penalty term is introduced, demonstrating strong robustness and stability. Further observation reveals that both TSAC (EP) and TTD3 (EP) exhibit higher secrecy rate across the training process, indicating that the energy penalty term significantly enhances the performance of the algorithms to some extent.

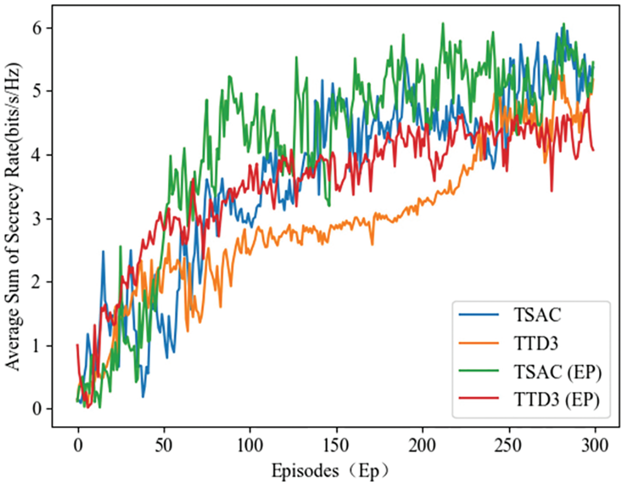

Figure 6: Comparison of the SSR

To further verify the effectiveness of the algorithms, considering legitimate SMs follow a Gaussian distribution (GD) which can more realistically reflect the randomness and clustering characteristics in practical applications. Fig. 7 shows the Reward comparison over different algorithms. The results indicate that in the initial training stage, the average reward values of all three algorithms exhibit significant fluctuations, reflecting that the strategies have not yet been effectively formed and the reward values are unstable. As the training rounds increases, the TSAC algorithm performs the best, with the highest average reward value and the smallest fluctuation, indicating its strongest generalization ability and adaptability. The TTD3 algorithm performs stably, with an average reward value slightly lower than that of TSAC. The TDDPG algorithm has the lowest average reward value and the largest fluctuation, suggesting that its strategy exploration efficiency is relatively low during the training process.

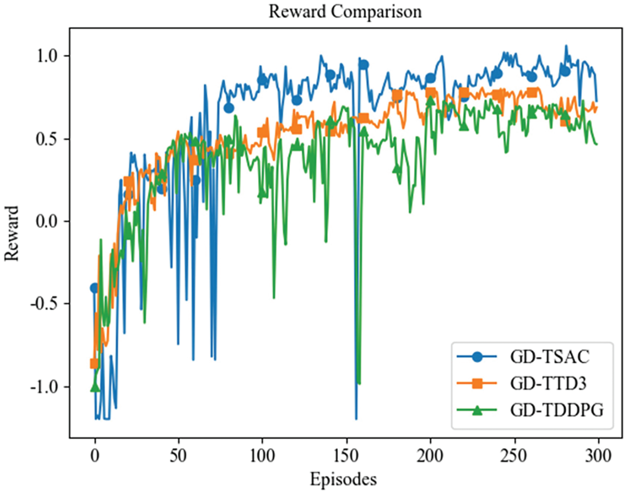

Figure 7: Reward comparison over different algorithms under GD

Figs. 8 and 9 show the performance of user (SM) capacity and secure capacity over three algorithms, under Gaussian distribution, respectively. The two figures exhibit the same trend with different capacity value. The results show that all three algorithms exhibit fluctuating trends in the early stages of training, with

TSAC and TTD3 showing greater volatility. It is found that in Fig. 8, TSAC algorithm demonstrates the best generalization ability, with its capacity value steadily increasing as the training rounds increase and reaching the highest value at the end of training, reflecting its strong adaptability and learning ability. The TTD3 algorithm has a relatively fast growth rate in the early stage of training, but is overtaken by TSAC in the middle stage. Eventually, its user capacity value is close to that of TSAC, indicating its good potential and adaptability. In contrast, the growth rate of the TDDPG algorithm is relatively slow throughout the training process, and its final user capacity value is significantly lower than the other two algorithms. As shown in Fig. 9, as the algorithms converge and enter the middle stages of training, TSAC begins to show significant growth, TTD3 maintains an upward trend despite a slowdown in growth rate, and TDDPG continues to increase at a relatively slow pace. In the later stages of training, TSAC ultimately achieves the highest secure capacity value, TTD3's performance approaches that of TSAC, while TDDPG, due to its sustained low growth rate, performs relatively poorly. Overall, in terms of algorithm generalization ability, the TSAC algorithm demonstrates the best secure capacity enhancement capability throughout the training process, a result that validates the effectiveness of the algorithm.

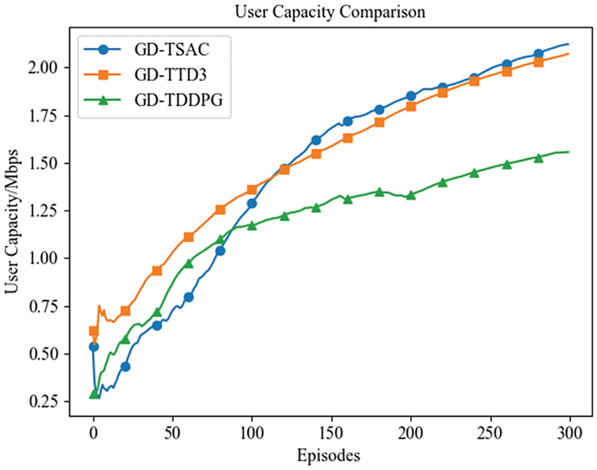

Figure 8: Capacity over different algorithms under GD

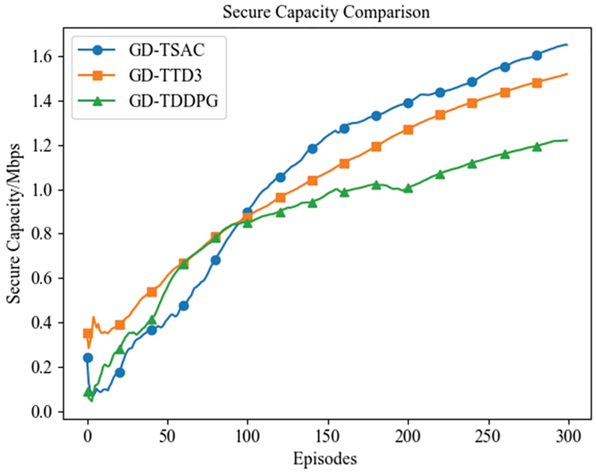

Figure 9: Secure capacity over different algorithms

Fig. 10 shows average SSR over different algorithms with and without the use of energy penalties. TSAC and TTD3 represent the cases without energy penalties, while TSAC (EP) and TTD3 (EP) represent the cases with energy penalties. It can be seen that throughout the training process, the TSAC algorithm shows higher sum of secrecy rate whether or not the penalty is used. In contrast, the SSR of the TTD3 algorithm are relatively lower under the same conditions. The higher rates exhibited by TSAC (EP) and TTD3 (EP) compared to TSAC and TTD3 indicate that the energy penalty term has enhanced the performance of the algorithms to some extent. By introducing the energy penalty term, the energy consumption of the UAV is effectively constrained, thereby ensuring high SSR. In summary, the TSAC algorithm demonstrates better performance in terms of secrecy communication rates.

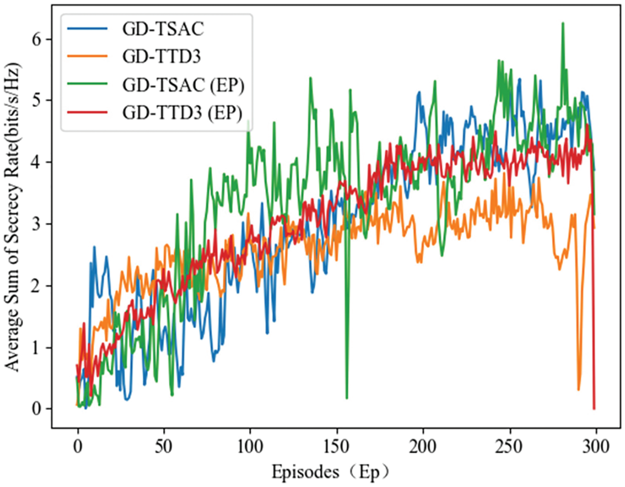

Figure 10: Comparison of average SSR under GD

5 Conclusions

We address the challenge of secure rate maximization in a UAV-RIS-aided smart grid (SG) system, which includes multiple authorized smart meters (SMs) and an eavesdropper. The Twin Soft Actor-Critic (TSAC) algorithm is proposed to tackle this issue. TSAC algorithm employs a dual-agent framework with Agent 1 focusing on optimizing the beamforming for both the UAV and the RIS, while Agent 2 concurrently searching for the optimal trajectory of the UAV. Simulation results show that the TSAC algorithm can significantly enhance the system's secure rate. It achieves faster convergence and higher rewards, even under the most challenging communication conditions, outperforming both the Twin Deep Deterministic Policy Gradient (TDDPG) and Twin Delayed Deep Deterministic Policy Gradient (TTD3) algorithms. Additionally, the TSAC algorithm shows strong performance when the distribution of smart meters follows a Gaussian distribution. In the future, we intend to study UAV-RIS-assisted downlink transmission scenarios in smart grids that involve multiple eavesdroppers. We will focus on developing more efficient deep reinforcement learning algorithms to address these challenges and facilitate the practical applications of secure communication in smart grids.

Acknowledgement: The authors would like to thank all of the authors cited.

Funding Statement: This work was supported by State Grid Shanxi Electric Power Company's Science and Technology Projects (No. 52051C230102).

Author Contributions: Conceptualization, Jian Wu; data curation, Jian Wu, Xiaowei Hao; writing—original draft, Jian Wu, Chao Han; writing—review and editing, Jian Wu, Xiaowei Hao, Chao Han; supervision, Jian Wu; funding acquisition, Chao Han. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data and materials supporting the findings of this study are fully presented within the manuscript.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Kong P-Y. Optimal configuration of interdependence between communication network and power grid. IEEE Trans Ind Informat. 2019;15:4054–65. doi:10.1109/TII.2019.2893132.
2. Ali SS, Choi BJ. State-of-the-art artificial intelligence techniques for distributed smart grids: a review. Electronics. 2020;9:1030. doi:10.3390/electronics9061030.
3. Rawat DB, Ghafoor KZ. Smart cities cybersecurity and privacy. Amsterdam, The Netherlands: Elsevier; 2018.
4. Bou-Harb E, Fachkha C, Pourzandi M, Debbabi M, Assi C. Communication security for smart grid distribution networks. IEEE Commun Mag. 2013;51:42–9. doi:10.1109/MCOM.2013.6400437.
5. Wang N, Wang P, Alipour-Fanid A, Jiao L, Zeng K. Physical-layer security of 5G wireless networks for IoT: challenges and opportunities. IEEE Internet Things J. 2019;6:8169–81. doi:10.1109/JIOT.2019.2927379.
6. Li X, Li J, Liu Y, Niyato D. Physical layer security of cognitive ambient backscatter communications for green internet-of-things. IEEE Trans Green Commun Netw. 2021;5:1066–76. doi:10.1109/TGCN.2021.3062060.
7. Zhang N, Wu R, Yuan S, Yuan C, Chen D. RAV: relay aided vectorized secure transmission in physical layer security for internet of things under active attacks. IEEE Internet Things J. 2019;6(5):8496–506. doi:10.1109/JIOT.2019.2919743.
8. Atallah M, Alam MS, Kaddoum G. Secrecy analysis of wireless sensor network in smart grid with destination assisted jamming. IET Commun. 2019;13:1748–52. doi:10.1049/iet-com.2018.5344.
9. Mensi N, Rawat DB, Balti E. Gradient ascent algorithm for enhancing secrecy rate in wireless communications for smart grid. IEEE Trans Green Commun Netw. 2022;6:107–16. doi:10.1109/TGCN.2021.3093821.
10. Liu Y, Su Z, Wang Y. Energy-efficient and physical-layer secure computation offloading in blockchain-empowered internet of things. IEEE Internet Things J. 2023;10:6598–610. doi:10.1109/JIOT.2022.3159248.
11. Liu P, Zou Y, Guo Q, Ma K, Tian N, Zhang Y. A secure transmission strategy for smart grid communications assisted by 5G base station. IEEE Trans Ind Appl. 2025;61(1):1695–703. doi:10.1109/TIA.2024.3522487.
12. ElMossallamy MA, Zhang H, Song L, Seddik KG, Han Z, Li GY. Reconfigurable intelligent surfaces for wireless communications: principles, challenges, and opportunities. IEEE Trans Cogn Commun Netw. 2020;6:990–1002. doi:10.1109/COMST.2021.3077737.
13. Yang L, Yang J, Xie W, Hasna MO, Tsiftsis T, Renzo MD. Secrecy performance analysis of RIS-aided wireless communication systems. IEEE Trans Veh Technol. 2020;69(10):12296–300. doi:10.1109/TVT.2020.3007521.
14. Sánchez JDV, Espinosa PR, Martínez FJL. Physical layer security of large reflecting surface aided communications with phase errors. IEEE Wireless Commun Lett. 2021;10:325–9. doi:10.1109/LWC.2020.3029816.
15. Padhan AK, Sahu HK, Sahu PR, Samantaray SR. Performance analysis of smart grid wide area network with RIS assisted three hop system. IEEE Trans Signal Inf Process Netw. 2023;9:48–59. doi:10.1109/TSIPN.2023.3239652.
16. Sikri A, Selim B, Kaddoum G, Au M, Agba BL. RIS-aided wireless sensor network in the presence of impulsive noise and interferers for smart-grid communications. IEEE Commun Lett. 2023;27:2501–5. doi:10.1109/LCOMM.2023.3299510.
17. Shi W, Xu J, Xu W, Di Renzo M, Zhao C. Secure outage analysis of RIS-assisted communications with discrete phase control. IEEE Trans Veh Technol. 2023;72:5435–40. doi:10.1109/TVT.2022.3224967.
18. Wafai B, Ghose S, Kundu C, Dubey A, Flanagan MF. Opportunistic user scheduling for secure RIS-aided wireless communications. IEEE Transact Vehic Technol. 2025;74:9194–209. doi:10.1109/TVT.2025.3537652.

19. Kawh M, Yun Z, Janti R. Secrecy performance analysis of RIS-aided smart grid communications. IEEE Trans Ind Inform. 2023;20:5415–27. doi:10.1109/TII.2023.3333842.
20. Lee H, Eom S, Park J, Lee I. UAV-aided secure communications with cooperative jamming. IEEE Trans Veh Technol. 2018;67:9385–92. doi:10.1109/TVT.2018.2853723.
21. Xiao L, Xu Y, Yang D, Zeng Y. Secrecy energy efficiency maximization for UAV-enabled mobile relaying. IEEE Trans Green Commun Netw. 2019;4(1):180–93. doi:10.1109/TGCN.2019.2949802.
22. Li Y, Zhang R, Zhang J, Gao S, Yang L. Cooperative jamming for secure UAV communications with partial eavesdropper information. IEEE Access. 2019;7:94593–603. doi:10.1109/ACCESS.2019.2926741.
23. Li R, Wei Z, Yang L, Ng DWK, Yuan J, An J. Resource allocation for secure multi-UAV communication systems with multi-eavesdropper. IEEE Trans Commun. 2020;68:4490–506. doi:10.1109/TCOMM.2020.2983040.
24. Yang L, Meng F, Zhang J, Hasna MO, Renzo MD. On the performance of RIS-assisted dual-hop UAV communication systems. IEEE Trans Veh Technol. 2020;69:10385–90. doi:10.1109/TVT.2020.3004598.
25. Long H, Chen M, Yang Z, Li Z, Wang B, Yun X, et al. Joint trajectory and passive beamforming design for secure UAV networks with RIS. In: Proceedings of the 2020 IEEE Globecom Workshops (GC Wkshps); 2020 Dec 7–11; Taipei, Taiwan. p. 1–6. doi:10.1109/GCWkshps50303.2020.9367542.
26. Wen C, Qiu L, Liang X. Securing UAV communication with mobile UAV eavesdroppers: joint trajectory and communication design. In: Proceedings of the 2021 IEEE Wireless Communications and Networking Conference (WCNC); 2021 Mar 29–31; Nanjing, China. p. 1–6. doi:10.1109/WCNC49053.2021.9417318.
27. Tao Q, Su G, Chen B, Dai M, Lin X, Wang H. Secrecy energy efficiency maximization for UAV enabled communication systems. In: Proceedings of the 2021 30th Wireless and Optical Communications Conference (WOCC); 2021 May 28–29; Shanghai, China. p. 240–4. doi:10.1109/WOCC53213.2021.9603100.
28. Mamaghani MT, Hong Y. Joint trajectory and power allocation design for secure artificial noise aided UAV communications. IEEE Trans Veh Technol. 2021;70:2850–5. doi:10.1109/TVT.2021.3057397.
29. Wang D, Zhao Y, He Y, Tang X, Li L, Zhang R, et al. Passive beamforming and trajectory optimization for reconfigurable intelligent surface-assisted UAV secure communication. Remote Sens. 2021;13:4286. doi:10.3390/rs13214286.
30. Zhu Z, Su G, Chen B, Dai M, Lin X, Wang H. Joint trajectory and power control for secure dual-UAV communications against air and ground eavesdropper. In: Proceedings of the 2022 31st Wireless and Optical Communications Conference (WOCC); 2022 May 14–15; Chongqing, China. p. 175–80. doi:10.1109/WOCC55104.2022.9880601.
31. Tian W, Ding X, Liu G, Dai Y, Han Z. A UAV-assisted secure communication system by jointly optimizing transmit power and trajectory in the internet of things. IEEE Trans Green Commun Netw. 2023;7:2025–37. doi:10.1109/TGCN.2023.3235887.
32. Qian Z, Deng Z, Cai C, Li H. Reinforcement learning based dual-UAV trajectory optimization for secure communication. Electronics. 2023;12:2008. doi:10.3390/electronics12092008.
33. Guo X, Chen Y, Wang Y. Learning-based robust and secure transmission for reconfigurable intelligent surface aided millimeter wave UAV communications. IEEE Wireless Commun Lett. 2021;10:1795–9. doi:10.1109/LWC.2021.3081464.
34. Tham M, Wong YJ, Iqbal A, Bin Ramli N, Zhu Y, Dagiuklas T. Deep reinforcement learning for secrecy energy-efficient UAV communication with reconfigurable intelligent surface. In: Proceedings of the 2023 IEEE Wireless Communications and Networking Conference (WCNC); 2023 Mar 26–29; Glasgow, UK. p. 1–6. doi:10.1109/WCNC55385.2023.10118891.
35. Zhang W, Zhao R, Xu Y. Aerial reconfigurable intelligent surface-assisted secrecy energy-efficient communication based on deep reinforcement learning. In: Proceedings of the 2024 12th International Conference on Intelligent Computing and Wireless Optical Communications (ICWOC); 2024 May 17–19; Chongqing, China. p. 60–5. doi:10.1109/ICWOC62055.2024.10684922.
36. Tanveer M, Khan AU, Shah H, Alkhayyat A, Chaudhry SA, Ahmad M. ARAP-SG: anonymous and reliable authentication protocol for smart grids. IEEE Access. 2021;9:143366–77. doi:10.1109/access.2021.3121291.
37. Tanveer M, Khan AU, Kumar N, Naushad A, Chaudhry SA. A robust access control protocol for the smart grid systems. IEEE Inter Things J. 2022;9:6855–65. doi:10.1109/jiot.2021.3113469.

38. Zeng Y, Xu J, Zhang R. Energy minimization for wireless communication with rotary-wing UAV. IEEE Trans Wireless Commun. 2019;18:2329–45. doi:10.1109/TWC.2019.2902559.

39. Zhou G, Pan C, Ren H, Wang K, Elkashlan M, Renzo MD. Stochastic learning-based robust beamforming design for RIS-aided millimeter-wave systems in the presence of random blockages. IEEE Trans Veh Technol. 2021;70:1057–61. doi:10.1109/TVT.2021.3049257.