RIS Enabled Simultaneous Transmission and Key Generation with PPO: Exploring Security Boundary of RIS Phase Shift

FAN Kaiqing[1], YAO Yuze[1], GAO Ning[1], LI Xiao[2], JIN Shi[2]

(1. School of Cyber Science and Engineering, Southeast University, Nanjing 210096, China;
2. National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China)

DOI: 10.12142/ZTECOM.202501003

https://kns.cnki.net/kcms/detail/34.1294.TN.20250226.1147.002.html, published online February 26, 2025

Manuscript received: 2025-01-25

Abstract: Due to the broadcast nature of wireless channels and the development of quantum computers, the confidentiality of wireless communication is seriously threatened. In this paper, we propose an integrated communications and security (ICAS) design to enhance communication security using reconfigurable intelligent surfaces (RIS), in which the physical layer key generation (PLKG) rate and the data transmission rate are jointly considered. Specifically, to deal with the threat of eavesdropping attackers, we focus on studying the simultaneous transmission and key generation (STAG) by configuring the RIS phase shift. Firstly, we derive the key generation rate of the RIS assisted PLKG and formulate the optimization problem. Then, in light of the dynamic wireless environments, the optimization problem is modeled as a finite Markov decision process. We put forward a policy gradient-based proximal policy optimization (PPO) algorithm to optimize the continuous phase shift of the RIS, which improves the convergence stability and explores the security boundary of the RIS phase shift for STAG. The simulation results demonstrate that the proposed algorithm outperforms the benchmark method in convergence stability and system performance. By reasonably allocating the weight factors for the data transmission rate and the key generation rate, "one-time pad" communication can be achieved. The proposed method has about 90% performance improvement for "one-time pad" communication compared with the benchmark methods.

Keywords: reconfigurable intelligent surfaces; physical layer key generation; integrated communications and security; one-time pad; deep reinforcement learning

Citation (Format 1): FAN K Q, YAO Y Z, GAO N, et al. RIS enabled simultaneous transmission and key generation with PPO: exploring security boundary of RIS phase shift [J]. ZTE Communications, 2025, 23(1): 11 – 17. DOI: 10.12142/ZTECOM.202501003
Citation (Format 2): K. Q. Fan, Y. Z. Yao, N. Gao, et al., "RIS enabled simultaneous transmission and key generation with PPO: exploring security boundary of RIS phase shift," *ZTE Communications*, vol. 23, no. 1, pp. 11 – 17, Mar. 2025. doi: 10.12142/ZTECOM.202501003.

1 Introduction

With the advancement of the 6G wireless communication, we are gradually moving towards the era of comprehensive Internet of Things (IoT). This provides more solid technical support for applications like smart interaction, industrial control, and remote healthcare, which requires extremely low latency while ensuring high security[1]. However, the widespread access to diversified intelligent mobile terminals and the demand for Gbit/s-level ultra-high throughput highlight the crucial importance of data security, especially in the broadcast wireless channels. Traditional key encryption methods may not be able to meet such stringent security requirements[2]. Meanwhile, it is necessary to meet the requirements for rapid key generation to reduce communication latency and ensure key security to prevent from cracking by quantum computers. This necessitates an in-depth exploration of the key distribution mechanism to discover the optimal trade-off between latency and security. Hence, in the 6G era, constructing a secure and efficient confidential communication system is urgently demanded.

In recent years, physical layer key generation (PLKG) has garnered increasing attention in academics and industry. PLKG is based on the physical layer characteristics of wireless environments, including the wireless channels that inherently possess randomness and reciprocal features. PLKG leverages these characteristics to establish a key generation mechanism, thus circumventing the challenges of traditional key distribution and update approaches. Typically, PLKG encompasses four steps[3]. First comes channel sounding, where the transceiver sends a pilot sequence to detect the channel and obtain reciprocal characteristics. Next is the quantization step, where the channel reciprocity features are transformed

This work was supported in part by the National Science Foundation of China (NSFC) under Grant No. 62371131, in part by the National Key R&D Program of China under Grant No. 2024YFE0200700, and in part by the program of Zhishan Young Scholar of Southeast University under Grant No. 2242024RCB0030.

into a binary bit sequence, and this bit sequence is then generated as the raw key. Due to issues like quantization accuracy, noise, and incomplete synchronization, the original bit sequence might not match properly. The third step is information reconciliation, where the error correcting codes are employed for correction purposes. Finally, privacy amplification is utilized, which aims to eliminate the potential risks of information leakage within the original bit sequence and generate symmetric keys to safeguard data security. MAURER[4] first explores the problem of generating shared keys through public discussions when both parties are aware of the relevant random variables but do not have an initial shared key. PREMNATH et al. evaluate the effectiveness of extracting keys from changes in wireless signal strength through actual measurements in Ref. [5]. They find that there are some problems with key generation in poor scattering environments, e.g., the entropy of the key is relatively low and the attacker can easily crack the key. An adaptive key generation scheme has been proposed to address these issues. In Ref. [6], LI et al. focus on using principal component analysis (PCA) preprocessing to generate highly consistent uncorrelated keys. However, due to the low-key generation rate in static wireless environments like an indoor office, it seriously affects the key generation rate.

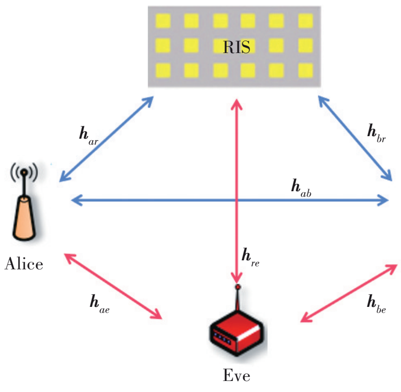
At present, some related studies begin to focus on reconfigurable intelligent surfaces (RIS) assisted PLKG to improve the key generation rate[7]. For example, Ref. [8] proposes a RIS assisted multi-carrier physical layer key generation framework to address the issue of insufficient randomness in wireless channels in static environments. Ref. [9] proposes the "Sem-Key" scheme, which utilizes the semantic drift phenomenon in semantic communication systems combined with RIS assistance to improve the key generation rate. The advantages and feasibility of this scheme have been experimentally verified. Ref. [10] proposes a RIS configuration method that utilizes channel state information (CSI) to control the activation of specific RIS units in the presence of eavesdroppers, thereby increasing the key capacity. However, the robust security of 6G enabled by the the RIS assisted PLKG, i. e, achieving "one-time pad" communications, still needs further study.

In 6G, the density of IoT devices per square kilometer can reach over 10 million. In such massive connection scenarios, communication security is extremely vulnerable. Integrated communications and security (ICAS) provides a potential solution to strong security, which shares communication resources and hardware resources and conducts an integrated design of communication functions and security functions. Specifically, the inherent by-products of communication are utilized to enhance the security abilities; at the same time, the improvement of security capabilities further ensures communication security, thereby enabling communication and security to mutually benefit and be internally generated with each other[2, 11]. Since the ICAS design focuses on real-time extreme security communication, artificial intelligence (AI) is an important endogenous power, especially deep learning (DL) and reinforcement learning (RL). In dynamic wireless environments, GAO et al. use deep Q-network (DQN) to optimize the RIS phase shift and for the first time demonstrate that the simultaneous transmission and key generation (STAG) can achieve "one-time pad" communication[11]. However, the existing DQN-based STAG method has some drawbacks, including the dimension explosion problem when the action space is large, and poor performance when there are many RIS units or high phase shift resolution. On the other hand, with the improvement of the RIS hardware manufacturing process, the high performance RIS with 3 bits or higher resolution, e.g., 360 degree RIS, has gradually emerged[12]. Motivated by these considerations, we propose a proximal policy optimization (PPO) based STAG method to study the security boundary of RIS phase shifts for STAG. The main contributions are summarized as follows.

- To improve the convergence stability of the deep reinforcement learning (DRL)-based STKG, a PPO-based STAG method is proposed. In particular, the RIS-assisted key generation rate is derived and the triple of the DRL, i.e., action, state, and reward, with respect to the STAG, is constructed.
- The continuous phase shift of RIS is optimized to explore the security boundary of RIS phase shifts. The upper bound of the RIS phase shifting capability for STAG is evaluated via the simulation. The continuous RIS phase shift yields over 5% higher reward than the 1-bit discrete RIS phase shift when the proposed algorithm converges.
- The simulation result shows that the "one-time pad" communication can be achieved by assigning suitable weight factors to STAG. Compared with the DQN-based method, the proposed PPO-based STAG method can obtain 90% performance improvement in "one-time pad" communication.

2 System Model

In Fig. 1, we consider a static RIS-assisted key generation scenario, which consists of four components: the legitimate transmitter and the receiver, namely Alice and Bob, the RIS,



RIS: reconfigurable intelligent surface

Figure 1. System model schematic diagram

and the malicious eavesdropper Eve. To simplify the analysis, we assume that Eve is in the middle of the legitimate users. Each participant is location-fixed and equipped with one antenna, and RIS has N reflection units.

2.1 Channel Model

The transmitter and the receiver intend to simultaneously generate keys and transmitted data, and the eavesdropper passively eavesdrops on the channel information. The signal received by Alice can be represented as:

$$r_a = \underbrace{\left(h_{br}^T \boldsymbol{\Phi} h_{ra} + h_{ba}\right)}_{h_{bra}} s_b + n_a \tag{1},$$

where $h_{br} \in C^{N \times 1}$ is the channel from Bob to RIS, $h_{ra} \in C^{1 \times N}$ is the channel from RIS to Alice, $h_{ba} \in C$ is the direct channel from Bob to Alice, h_{bra} is the equivalent channel, s_b is the transmission signal of Bob, $\boldsymbol{\Phi} = \mathrm{diag}\left[\alpha_1 e^{j\theta_1}, \alpha_2 e^{j\theta_2}, \cdots, \alpha_N e^{j\theta_N}\right]$ is the phase-shift matrix of RIS with $\phi_{n,n} = \alpha_{\mathrm{n}} e^{j\theta_n}, \alpha_n = 1, \theta_n \in [0, 2\pi), n = 1,2,\cdots,N$, and n_a is the channel noise following complex Gaussian distribution with zero mean and σ^2 variance.

Similarly, we can obtain the received signals of Bob and Eve, respectively, which is given by Eqs. (2) and (3). Therein, $h_{re} \in C^{N \times 1}$ is the channel from RIS to Eve, $h_{ae} \in C^{N \times 1}$ is the channel from Alice to Eve, and n_a and n_e are the channel noise.

$$r_b = \underbrace{\left(h_{ar}^T \boldsymbol{\Phi} h_{rb} + h_{ab}\right)}_{h_{arb}} s_a + n_{\mathrm{b}} \tag{2},$$

$$r_e = \underbrace{\left(h_{ar}^T \boldsymbol{\Phi} h_{re} + h_{ae}\right)}_{h_{are}} s_a + n_e \tag{3}.$$

2.2 Channel Estimation

During the coherent time, Alice exchanges the pilot signal with Bob for channel estimation. Let Alice be the communication initiator, and Bob estimates CSI through least squares.*

$$\hat{h}_{arb} = h_{ar}^T \boldsymbol{\Phi} h_{rb} + h_{ab} + n_0 s_a^{p*} \tag{4},$$

where s_a^p is the pilot signal from Alice to Bob and the pilot signal satisfies $s_a^p s_a^{p*} = 1$; the channel estimation error is $n_0 s_a^{p*}$. Next, symmetric keys are generated through quantization, information reconciliation and privacy amplification[13]. Since these steps are not the key point of this paper, they are not elaborated on any further.

2.3 Key Generation Rate

The mutual information between the channel observations of the legitimate parties is an important factor in determining the key generation rate. Due to quantization error in bit representation, we consider the mutual information as the upper bound of the key generation rate, which is the mutual information of CSI under Eve's observation. With the eavesdropper Eve, the key generation rate can be formulated as[14]:

$$\begin{aligned}
&\mathcal{R}_{\text{key}} = \frac{1}{T} I\left(\hat{h}_{arb}; \hat{h}_{bra} | \hat{h}_{are}\right) = \\
&\frac{1}{T}\left[H\left(\hat{h}_{arb} | \hat{h}_{are}\right) - H\left(\hat{h}_{arb} | \hat{h}_{bra}, \hat{h}_{are}\right)\right] = \\
&\frac{1}{T} \log_2 \frac{\det\left(R_{ae}\right)\det\left(R_{\mathrm{be}}\right)}{\det\left(R_e\right)\det\left(R_{abe}\right)}
\end{aligned} \tag{5},$$

where $\det(R)$ is the matrix determinant, while $\boldsymbol{R}_{ae}, \boldsymbol{R}_{be}, \boldsymbol{R}_e$, and $\boldsymbol{R}_{abe}$ are the covariance matrices. T represents the observation time. Specifically, the covariance matrix is as follows:

$$\boldsymbol{R}_{A_1, \cdots, A_n} = E\begin{bmatrix} a_1 a_1^* & \cdots & a_1 a_n^* \\ \vdots & \ddots & \vdots \\ a_n a_1^* & \cdots & a_n a_n^* \end{bmatrix} \tag{6}.$$

The key rate is expressed in Eq. (7), where $\|\cdot\|$ is the Euclidian norm operator. E represents mathematical expectation. For convenience, we simplify the variance of the noise to 1. Thus, we can obtain the key generation rate, which is expressed in Eq. (8).

$$\begin{aligned}
&\mathcal{R}_{\text{key}} = \\
&\log_2\left(\frac{\left(\left(R_a + \sigma^2\right)\left(R_e + \sigma^2\right) - \|R_{ae}\|^2\right)^2}{\left(R_e + \sigma^2\right)\left(\left(2R_a\sigma^2 + \sigma^4\right)\left(R_e + \sigma^2\right) - 2\sigma^2\|R_{ae}\|^2\right)}\right)
\end{aligned} \tag{7},$$

$$\mathcal{R}_{\text{key}} = \log_2\left(\frac{\left(\left(R_a + 1\right)\left(R_e + 1\right) - \|R_{ae}\|^2\right)^2}{\left(R_e + 1\right)\left(\left(2R_a + 1\right)\left(R_e + \sigma^2\right) - 2\|R_{ae}\|^2\right)}\right) \tag{8}.$$

According to Shannon's formula, the maximum channel capacity is the theoretical maximum transmission rate, which can be obtained by calculating the signal-to-noise ratio (SNR). Thus, we can obtain the ergodic data transmission rate at Alice as:

$$\mathcal{R}_{\text{data}} = B\log_2\left(1 + E\|h_{br}^T \boldsymbol{\Phi} h_{ra} + h_{ba}\|^2\right) \tag{9},$$

where B is the signal bandwidth.

3 Problem Formulation and Proposed Solution

3.1 Problem Formulation

We consider jointly optimizing the key generation rate and data transmission rate, that is, to ensure the data transmission

* The channel estimation considered in this paper has no error, and the analysis is based on perfect channel state information. The research based on imperfect channel state information will be carried out in the future.

rate reaches a high level while maximizing the key generation rate to enhance the confidentiality of wireless communication. For the trade-off between the key generation rate and the data transmission rate, we make decisions based on the specific application scenarios, such as in real-time communication prioritized applications about the voice calls and the video conferences, which increases the weight of the data transmission rate and appropriately reduces the key generation rate. For financial transaction scenarios and military communication scenarios that focus on high security and confidentiality, we increase the weight of the key generation rate accordingly. In short, we can first evaluate the security and the quality of service (QoS) requirements of the scenario and allocate corresponding weight reasonably to the specific scenario. Therefore, we can formulate the optimization problem as

$$\begin{aligned}&\mathbb{P}: \quad \max\ w_d \mathcal{R}_{\text{data}} + w_k \mathcal{R}_{\text{key}} \\ &\text{s.t.}\ \ 0 \leqslant \theta_n < 2\pi, \forall n \in \{1, \cdots, N\} \\ &|\phi_{n,n}| = 1\end{aligned} \qquad (10),$$

where $w_d \in [0,1]$, $w_k = 1 - w_d \in [0,1]$ is the weight that balances the priority level of the key generation rate and data transmission, n represents the number of reflection units of RIS, θ_n represents the phase shift unit of RIS, and $|\phi_{n,n}|$ is the phase-shift unit of RIS with a constant modulus constraint.

Due to the non-convex nature of the optimization problem, it is hard for the traditional convex optimization to obtain the optimal solution in real-time. Considering the dynamic wireless environments, we construct the time series of the dynamic channel as a Markov decision process. This indicates that DRL is a potent instrument for resolving the Markov decision process. PPO is a model-free reinforcement learning algorithm, which belongs to the family of strategy gradient algorithms. It is mainly used to optimize the strategy network so that the agents can take optimal actions in the environment to maximize the cumulative rewards. Due to the increasing demand for efficient and stable algorithms, PPO has emerged where the action space is continuous. It not only performs well in the large dimensional action space but also has the advantages of high training efficiency and easy convergence. Therefore, we use the PPO algorithm to jointly optimize the transmission rate and the key generation rate with the continuous RIS phase shift[15].

3.2 Sample Collection

Firstly, we use the current strategy network to interact with the environment and collect a series of state-action-reward samples $\{(s_i, a_i, r_i)\}$[16]. These samples form an experience replay buffer. Then, the advantage function and target value are calculated based on the collected samples, and the state value function is estimated to calculate the advantage function $A(s,a)$. Here, we use Monte Carlo estimation to calculate the value function, where the advantage function can be calculated by subtracting the state value function from the cumulative reward of the trajectory[17]. For time difference learning, it can be denoted as:

$$A(s,a) = r + \gamma V(s') - V(s) \qquad (11),$$

where r is the instant reward, γ is the discount factor, and s' is the next state.

3.3 Strategy Network Update

In this step, the gradient descent algorithm is employed to minimize the loss function and optimize the policy network. The loss function $L^{\text{CLIP}}(\theta)$ is calculated to obtain the gradient of the policy network parameter. Then, the gradient descent is used to update via the formula $\theta = \theta - \beta \nabla_\theta L^{\text{CLIP}}(\theta)$, where β represents the learning rate[18]. The specific settings of the state space, the action space, and the reward function in the Markov decision process are as follows.

State: The state space is defined as the CSI of the communication environment observed by Alice. Therefore, at time step i, the state is denoted as:

$$s^i = \left\{ h^i_{bar,\Phi^{i-1}}, h^i_{bae,\Phi^{i-1}} \right\} \qquad (12).$$

The state information is the basis for intelligent agents to make decisions.

Action: Since we train the network by continuously adjusting the phase shift of RIS, the action space at time step i can be represented as:

$$a^i = \{\boldsymbol{\Phi}^i\} \qquad (13),$$

where $\boldsymbol{\Phi} = \text{diag}[\alpha_1 e^{j\theta_1}, \alpha_2 e^{j\theta_2}, \cdots, \alpha_N e^{j\theta_N}]$ and the phase shift of RIS is $\theta_N \in [0, 2\pi)$.

Reward: As the formulated optimization problem, the reward function can be established in the form of the optimization objective, which can be expressed as:

$$r = w_d \mathcal{R}_{\text{data}} + w_k \mathcal{R}_{\text{key}} \qquad (14).$$

3.4 Computational Complexity

The computational complexity of the proposed algorithm includes training complexity and deployment complexity, which will be analyzed as follows.

Training complexity: Firstly, we calculate the computational complexity of the activation layers. The computational complexity of the ReLU layer is "1", that of the sigmoid layer is "2", and that of the tanh layer is "2". Assume that the total number of nodes in the state normalization layer, ReLU layer, sigmoid layer, and tanh layer are $|\mathcal{S}|, n_r, n_s$, and n_t. Thus, the training complexity for node computation is $O(|\mathcal{S}| + n_r + 2n_s + 2n_t)$. Furthermore, we assume that both the evaluated network and the target network consist of L fully connected layers and the l-th layer has n_l nodes. The training complexity of one forward propagation and two backward propagations can be calculated by $O\left(\sum_{l=0}^{L-1} 3n_l n_{l+1}\right)$. In the PPO algorithm, multiple trajectories need to be sampled from the environment for learning. Supposing that the trajectories sample is N and the length of each trajectory is T, the complexity of the sampling and the update process can be expressed as $O\left(N\cdot T\cdot\left(3\sum_{l=0}^{L-1} n_l n_{l+1}\right)\right)$. The total complexity of the PPO algorithm in the training phase is $O\left(K\cdot N\cdot T\cdot\left(|\mathcal{S}| + n_r + 2n_s + 2n_t + 3\sum_{l=0}^{L-1} n_l n_{l+1}\right)\right)$, where K represents the total number of iterations.

Deployment complexity: Since we only use the policy network π_θ for action selection, sampling and update operations are not involved. Therefore, only the computational complexity of state normalization and one forward propagation needs to be considered. Similar to the above analysis, the complexity of the deployment phase can be expressed as $O\left(\sum_{l=0}^{L-1} n_l n_{l+1}\right) + O(|\mathcal{S}|)$.

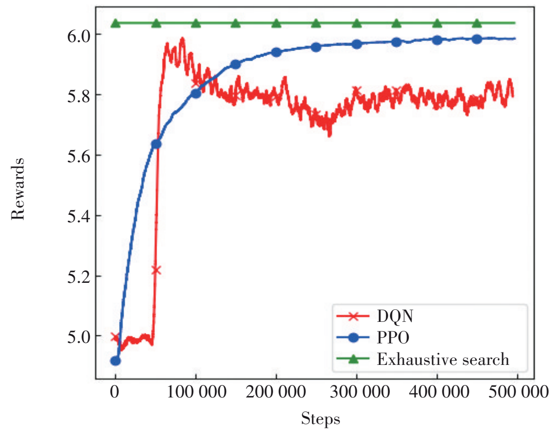
4 Simulation Results

In terms of weight factors, the weights of both the data transmission rate w_d and the key generation rate w_k are set to 0.5, which means that the two tasks have equal priority. The learning rate β is set to 0.000 3. This small value ensures that the model parameter updates are relatively stable during the training process, thereby reducing the risk of missing the optimal solution or making the training diverge due to overly large update steps. The discount factor γ is set to 0.99, indicating that the agent places great emphasis on relatively long-term returns. The batch size *batch_size* is set to 64. When parameters are updated each time, 64 samples are extracted from the sample data for calculation. This value can maintain a reasonable computational efficiency while taking into account a certain degree of stability in gradient estimation. In the generalized advantage estimation, the parameter *gae_λ* is set to 0.95, biasing the advantage estimation towards prioritizing the long-term temporal difference error information.

The DQN-based STAG method is proposed to optimize the key generation rate[11]. However, this method makes it difficult to handle continuous action space problems, thereby leading to a dimensional disaster for the large action space or the loss of some action information. The proposed PPO-based STAG method can effectively handle continuous action space and the convergence is stable. Thus, we use the DQN-based method as a benchmark and study the security boundary of the RIS phase shift for STAG with the PPO-based method.

Specifically, the DQN algorithm selects (discrete) phase shift values for the 8 elements in the action space of RIS from [0, 2π), and the resolution of the RIS phase-shift is 1 bit. The PPO-based STAG method selects continuous phase shift values for the 8 elements in the continuous action space of RIS from [0, 2π). Fig. 2 shows although the DQN-based STAG method converges slightly faster than the PPO-based method, the reward of the former is unstable and not as high as the reward of the latter. The reward of PPO can reach 6.0, while DQN is only 5.7, which has improved the reward by more than 5%. To analyze the optimal solution, we use an exhaustive search optimization method and compare it with the optimization results of the PPO algorithm. In Fig. 2, the optimization results of the PPO algorithm are very close to the optimal result of the exhaustive search optimization, which is demonstrated to be optimally achieved in dynamic wireless environments.

To prove the convergence stability of the PPO-based STAG method in large dimensional action space, we explore the relationship between the reward and the number of RIS reflection units with the continuous phase shift in Fig. 3. When the number of RIS reflection units increases, the key generation rate increases obviously. When the number of RIS reflection units



DQN: deep Q-network PPO: proximal policy optimization

Figure 2. Comparison between DQN algorithm and PPO algorithm

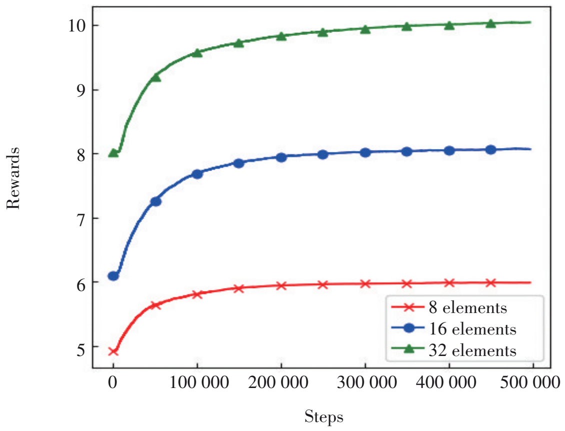


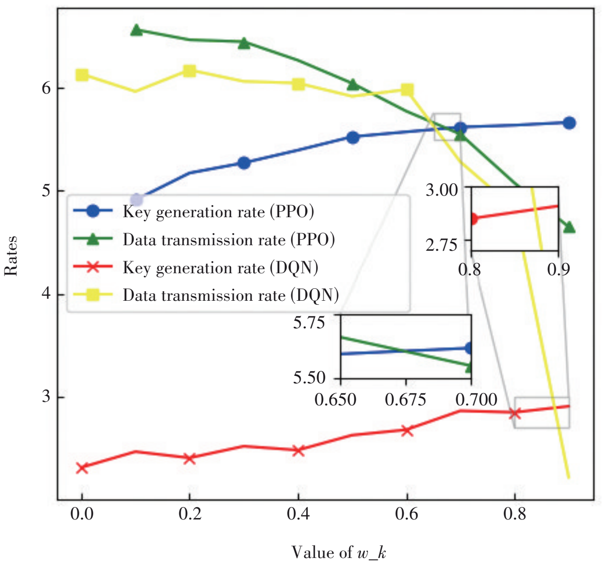
Figure 3. Comparison of different RIS components by using PPO algorithm

is 32, the reward is close to 10, which is twice as much as when the number of RIS reflection units is 8. Specifically, as the number of reflection units rises, the channel gain increases with the assistance of the RIS, thereby improving the STAG performance.

To explore the security boundary of the RIS phase shift and validate the effect of the "one-time pad" with STAG, we study the optimal transmission rate and key generation rate in different weights. Fig. 4 illustrates the relationship between weight and the rate change based on the PPO algorithm. It can be found that as the weight w_k increases, the data transmission rate decreases and the key generation rate increases. The PPO-based STAG method outperforms the DQN-based method both in key generation rate and the data transmission rate. Importantly, the key generation rate and the data generation rate are equal for the proposed PPO-based STAG and the DQN-based STAG when the weight is about 0.675 and 0.92, respectively. It suggests that this weight can achieve "one-time pad" communication via STAG design. Specifically, there is about 90% performance improvement for "one-time pad" communication than that of DQN-based STAG method, which shows the security boundary of the RIS phase shift.

5 Conclusions

In this paper, we study the potential of ICAS to attain perfectly secure communication with the presence of the eavesdropper via the STAG design. Specifically, we consider the dynamic wireless environments and propose a policy gradient algorithm based on PPO, which is to improve the convergence stability of STAG in large-scale action space and explore the security boundary of the RIS phase shift. The simulation results indicate that the proposed PPO-based STAG method has a better performance than the DQN-based STAG method and approaches the optimal exhaustive search, which shows the security boundary of the RIS phase shift. By setting a suitable weight to balance the data transmission rate and communication security, "one-time pad" communication can be achieved.



DQN: deep Q-network PPO: proximal policy optimization

Figure 4. Variation of the data transmission rate and the key generation rate with different weights

References

[1] SANG J, YUAN Y F, TANG W K, et al. Coverage enhancement by deploying RIS in 5G commercial mobile networks: field trials [J]. IEEE wireless communications, 2024, 31(1): 172 – 180. DOI: 10.1109/MWC.011.2200356

[2] GAO N, HAN Y, LI N N, et al. When physical layer key generation meets RIS: opportunities, challenges, and road ahead [J]. IEEE wireless communications, 2024, 31(3): 355 – 361. DOI: 10.1109/MWC.013.2200538

[3] MOARA-NKWE K, SHI Q, LEE G M, et al. A novel physical layer secure key generation and refreshment scheme for wireless sensor networks [J]. IEEE access, 2018, 6: 11374 – 11387. DOI: 10.1109/ACCESS.2018.2806423

[4] MAURER U M. Secret key agreement by public discussion from common information [J]. IEEE transactions on information theory, 1993, 39(3): 733 – 742. DOI: 10.1109/18.256484

[5] PREMNATH S N, JANA S, CROFT J, et al. Secret key extraction from wireless signal strength in real environments [J]. IEEE transactions on mobile computing, 2013, 12(5): 917 – 930. DOI: 10.1109/TMC.2012.63

[6] LI G Y, HU A Q, ZHANG J Q, et al. High-agreement uncorrelated secret key generation based on principal component analysis preprocessing [J]. IEEE transactions on communications, 2018, 66(7): 3022 – 3034. DOI: 10.1109/TCOMM.2018.2814607

[7] JI Z J, YEOH P L, ZHANG D Y, et al. Secret key generation for intelligent reflecting surface assisted wireless communication networks [J]. IEEE transactions on vehicular technology, 2021, 70(1): 1030 – 1034. DOI: 10.1109/TVT.2020.3045728

[8] GU J, OUYANG C J, ZHANG X, et al. RIS-assisted multi-carrier secret key generation in static environments [J]. IEEE wireless communications letters, 2024, 13(10): 2777 – 2781. DOI: 10.1109/LWC.2024.3445268

[9] ZHAO R, QIN Q, XU N Y, et al. SemKey: boosting secret key generation for RIS-assisted semantic communication systems [C]//The 96th Vehicular Technology Conference. IEEE, 2022: 1 – 5. DOI: 10.1109/VTC2022-Fall57202.2022.10013083

[10] XU N Y, NAN G S, TAO X F. Passive eavesdropping can significantly slow down RIS-assisted secret key generation [C]//IEEE Global Communications Conference. IEEE, 2023: 3294 – 3299. DOI: 10.1109/GLOBECOM54140.2023.10437788

[11] GAO N, YAO Y Z, JIN S, et al. Integrated communications and security: RIS-assisted simultaneous transmission and generation of secret keys [J]. IEEE transactions on information forensics and security, 2024, 19: 7573 – 7587. DOI: 10.1109/TIFS.2024.3436885

[12] TANG J W, XU S H, YANG F, et al. Recent developments of transmissive reconfigurable intelligent surfaces: a review [J]. ZTE Communications, 2022, 20(1): 21 – 27. DOI: 10.12142/ZTECOM.202201004

[13] LIU Y W, LIU X, MU X D, et al. Reconfigurable intelligent surfaces: principles and opportunities [J]. IEEE communications surveys & tutorials, 2021, 23(3): 1546 – 1577. DOI: 10.1109/COMST.2021.3077737

[14] GAO N, QIN Z J, JING X J, et al. Anti-intelligent UAV jamming strategy via deep Q-networks [J]. IEEE transactions on communications, 2020, 68 (1): 569 – 581. DOI: 10.1109/TCOMM.2019.2947918

[15] LUONG N C, HOANG D T, GONG S M, et al. Applications of deep rein-

forcement learning in communications and networking: a survey [J]. IEEE communications surveys and tutorials, 2019, 21(4): 3133 – 3174

[16] QIAN X, DI RENZO M, LIU J, et al. Beamforming through reconfigurable intelligent surfaces in single-user MIMO systems: SNR distribution and scaling laws in the presence of channel fading and phase noise [J]. IEEE wireless communications letters, 2021, 10(1): 77 – 81. DOI: 10.1109/LWC.2020.3021058

[17] ZHANG H Q, LI X, GAO N, et al. A deep reinforcement learning approach to two-timescale transmission for RIS-aided multiuser MISO systems [J]. IEEE wireless communications letters, 2023, 12(8): 1444 – 1448. DOI: 10.1109/LWC.2023.3278171

[18] LU T Y, CHEN L Q, ZHANG J Q, et al. Joint precoding and phase shift design in reconfigurable intelligent surfaces-assisted secret key generation [J]. IEEE transactions on information forensics and security, 2023, 18: 3251 – 3266. DOI: 10.1109/TIFS.2023.326888

Biographies

FAN Kaiqing received his BS degree in computer science from Nanjing University of Finance and Economics, China in 2023. He is currently pursuing his MS degree with the School of Cyber Science and Engineering, Southeast University, China. His research interests are RIS-assisted physical layer security and deep reinforcement learning.

YAO Yuze received his BS degree in information security from China University of Mining and Technology, China in 2023. He is currently pursuing his MS degree with the School of Cyber Science and Engineering, Southeast University, China. His research interests include wireless communication security and deep reinforcement learning.

GAO Ning (ninggao@seu.edu.cn) received his PhD degree in information and communications engineering from Beijing University of Posts and Telecommunications, China in 2019. From 2017 to 2018, he was a visiting PhD student with the School of Computing and Communications, Lancaster University, UK. From 2019 to 2022, he was a research fellow with the National Mobile Communications Research Laboratory, Southeast University, China. He is currently an associate professor with the School of Cyber Science and Engineering, Southeast University. His research interests include AI enabled wireless communications and security, reconfigurable intelligent surfaces (RIS), and UAV communications.

LI Xiao received her PhD degree in communication and information systems from Southeast University, China in 2010. Then, she joined the School of Information Science and Engineering, Southeast University, where she has been a professor of information systems and communications since July 2020. From January 2013 to January 2014, she was a postdoctoral fellow at The University of Texas at Austin, USA. Her current research interests include massive MIMO, reconfigurable intelligent surface assisted communications, and intelligent communications. She was a recipient of the 2013 National Excellent Doctoral Dissertation of China for her PhD dissertation.

JIN Shi received his PhD degree in communications and information systems from Southeast University, China in 2007. From June 2007 to October 2009, he was a research fellow with the Adastral Park Research Campus, University College London, UK. He is currently a faculty member with the National Mobile Communications Research Laboratory, Southeast University. His research interests include wireless communications, random matrix theory, and information theory. He was an associate editor of *IEEE Transactions on Wireless Communications*, *IEEE Communications letters*, and *IET Communications*. He serves as an area editor of *IEEE Transactions on Communications* and *IET Electronics Letters*.