Deep Reinforcement Learning Based Uplink Security Enhancement for STAR-RIS-Assisted NOMA Systems with Dual Eavesdroppers

Xintong Qin, Zhengyu Song, *Member, IEEE,* Jun Wang, Shengyu Du, Jiazi Gao, Wenjuan Yu, *Member, IEEE*, and Xin Sun

***Abstract*—This paper investigates the simultaneous transmitting and reflecting reconfigurable intelligent surface (STAR-RIS) assisted non-orthogonal multiple access (NOMA) systems with one cooperative jammer and dual eavesdroppers. To guarantee the uplink secure transmission, we maximize the sum secrecy rate under both the perfect and imperfect channel state information (CSI) by jointly optimizing the channel allocation, transmit power and coefficient matrices. For the problem with perfect CSI, a deep reinforcement learning algorithm is proposed based on the deep deterministic policy gradient (DDPG) framework. Then, by introducing the arbitrary distorted noise to the state space, the proposed algorithm is extended to solve the problem under imperfect CSI without causing additional computational complexity. Simulation results illustrate that: 1) The symmetry of STAR-RIS results in severe information leakage, and the sum secrecy rate further degrades when the dual eavesdroppers collaborate with each other. 2) The STAR-RIS with independent phase shift can achieve higher sum secrecy rate than that with coupled phase shift, while the performance gap is trivial when there are fewer STAR-RIS elements. 3) Our proposed algorithm can compensate for the impacts of the imperfect CSI, and the sum secrecy rate decreases with the increase of CSI uncertainty.**

***Index Terms*—Simultaneous transmitting and reflecting reconfigurable intelligent surface (STAR-RIS), physical layer security, resource allocation, deep reinforcement learning (DRL).**

I. Introduction

Recently, the Internet of Things (IoT) technology has swiftly advanced and seamlessly integrated into our daily routines. From the home automation [1] to smart cities [2], from the health-care [3] to smart manufacturing [4], the IoT has permeated various sectors, offering people a more intelligent, convenient, and efficient way of life. Nevertheless, the burgeoning number of IoT devices and diverse application requirements have placed unprecedented demands on the communication networks, such as higher system sum rate and more secure transmissions. In response to these challenges, the reconfigurable intelligent surface (RIS) is emerging as a promising technology to deal with the evolving demands of IoT and communication networks [5], [6].

The RIS, also called the intelligent reflecting surfaces (IRS), is an array composed of a numerous number of passive elements [7]. When a radio wave impinge upon the RIS, these elements can independently configure their phase shifts and amplitudes (also referred to as passive beamforming) through the attached controller, to make the induced surface currents generate reflected radio waves in a desired direction, and transform the propagation environment into a controllable element, i.e., smart radio environment [8], [9]. Owing to the ability to reshape the propagation environment, it is envisioned that the RIS-assisted networks are capable of obtaining higher system sum rate and more secure transmissions. Thus, countless efforts have been furnished to investigate how to achieve these significant performance enhancements by jointly designing the passive beamforming of RIS and the radio resources allocation for the RIS-assisted networks.

For example, in the earlier study aiming to increase the total rate of RIS-assisted system, C. Huang *et al.* [10] adopts the alternating optimization and majorization-minimization methods to optimize the transmit power and the phase shifts of RIS, where the formulated problem is simplified by neglecting the inter-user interference. In the simulation results, it is shown that the proposed scheme improves the sum rate by over 40% compared to the traditional systems without RIS, validating the significant potential of RIS in enhancing the system sum rate. Later in [11], J. Zuo *et al.* introduce NOMA into the RIS-assisted systems, where the inter-user interference can be handled by the successive interference cancellation (SIC) technique. In order to achieve the maximum sum rate, the channel and power allocation, decoding order, and reflection coefficients of RIS are jointly optimized by a three-step algorithm. Different from the ideal RIS in [11] whose phase shift is continuous, X. Mu *et al.* [12] design a quantization-based scheme for the non-ideal RIS-assisted NOMA systems to optimize the discrete phase shifts, and the simulations show that only 3-bit phase shifters are required to achieve a sum rate nearly equivalent to that of an ideal RIS.

The above-mentioned literatures rely on instantaneous CSI, where it is actually challenging to obtain in the RIS-assisted systems due to the passive nature of RIS. Thus, with the statistical CSI, H. Zhang *et al.* [13] optimize the beamforming

Manuscript received December 26, 2023; revised February 27 and May 10, 2024; accepted June 14, 2024. Date of publication June 14, 2024; date of current version June 14, 2024. This work was supported by the National Natural Science Foundation of China under Grant 61901027. *(Corresponding author: Jun Wang.)*

X. Qin, Z. Song, J. Wang, S. Du, J. Gao, and X. Sun are with the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing 100044, China (emails: 20111046@bjtu.edu.cn, songzy@bjtu.edu.cn, wangjun1@bjtu.edu.cn, 23120043@bjtu.edu.cn, 22120050@bjtu.edu.cn, xsun@bjtu.edu.cn).

W. Yu is with the School of Computing and Communications, InfoLab21, Lancaster University, Lancaster LA1 4WA, U.K. (e-mail: w.yu8@lancaster.ac.uk).

both at the BS and RIS by the water-filling and projected gradient ascent algorithms to maximize the sum rate for the RIS-assisted systems. Although a performance gap remains between the proposed scheme and the benchmark with instantaneous CSI, the practical implementation can benefit from the lower channel estimation overhead in the proposed scheme.

Through the dedicated passive beamforming and resource allocation, the RIS can enhances the sum rate by coherently combining the reflected signals at the legitimate users. Likewise, the RIS can also reduce the signal strengths of malicious eavesdroppers by configuring the passive beamforming of RIS to avoid the information leakage and enhance the physical layer security (PLS) [14]. However, it is important to mention that when the RIS is employed to enhance the PLS, the vision is to simultaneously improve the legitimate users' signal strength of and suppress that of malicious eavesdroppers, rendering the passive beamforming design and resources allocation more difficult. Therefore, starting from the simple networks with one user and one eavesdropper [15]–[17], the non-convex secrecy rate maximization problems for enhancing the PLS are solved by the alternating optimization algorithms to adjust the transmit beamforming of BS and the phase shift coefficient of RIS, where the simulation results validate the secrecy rate performance can be boosted by deploying the RIS. For more practical scenario with multiple users, Z. Zhang *et al.* [18] investigate the RIS-assisted NOMA systems and aim to reduce the transmit power in the conditions of satisfying the secrecy rate requirement. In order to better manage the inter-user interference, Y. Gao *et al.* [19] adopt the rate-splitting multiple access (RSMA) technique for the RIS-assisted systems with multiple users, and propose an iterative algorithm to maximize the minimum secrecy rate, which can enhance the max-min secrecy rate compared to benchmark approaches that either do not deploy RIS or utilize different multiple access techniques. Besides multiple users, it is also possible that there are multiple eavesdroppers. Therefore, Y. Wang *et al.* [20] study the RIS-assisted systems with multiple colluding eavesdroppers, where the sum secrecy rate is maximized via an SDR algorithm.

To further improve the PLS for the RIS-assisted systems without compromising the rate performance of legitimate users, X. Guan *et al.* [21] investigate the joint transmit beamforming with artificial noise (AN) and the passive beamforming of RIS, where the simulation results reveal the necessity of AN in the RIS-assisted secrecy communication systems. Motivated by these results, C. Zheng *et al.* [22] solve the secrecy rate maximization problems for the RIS-assisted systems with the integration of AN, where the semi-closed-form expressions for the transmit precoding and the AN matrices are derived via Lagrange dual method, and the closed-form expression for RIS's phase shift is obtained via the Majorization-Minimization (MM) algorithm. With the aid of AN, X. Yu *et al.* [23] design a robust AO algorithm to optimize the beamforming, phase shifts of RIS, as well as the covariance matrix of AN for secure transmissions in the multiple-RIS-assisted systems.

Nevertheless, if the AN is transmitted by a cooperative jammer located at the different locations with the BS, the above-mentioned algorithms for RIS's phase shifts optimization will not be effective. Additionally, when the RIS is deployed to improve the PLS, another challenge for passive beamforming and resource allocation design is the acquisition of correct CSI. It is challenging to obtain the perfect CSI in the RIS-assisted secure transmission systems since the RIS is a nearly-passive device without the inherent ability of channel sensing, and the eavesdroppers will endeavor to evade the channel measurement of BS through staying silence [14], [18]. Therefore, the impacts of the imperfect CSI are necessary to be taken into account to improve the robustness when designing the algorithms for passive beamforming and resource allocation. Assuming the channel estimation error is confined within a bounded region, the worst-case secrecy rate maximization problems are investigated in [24], [25] for the RIS-assisted systems, where the Cauchy-Schwarz inequality is utilized to derive the upper bound of channel gain between the BS and eavesdropper. In [18], [23], [26], the semi-infinite constraints of channel estimation errors are firstly transformed into the equivalent linear matrix inequalities (LMIs) by the S-procedure, and then the passive beamforming and resource allocation are optimized via the AO algorithms. Although the above-mentioned algorithms can transform the stochastic optimization problems of the RIS-assisted secure transmission systems with the imperfect CSI into a more tractable deterministic form, the upper bound method is conservative and may cause the waste of radio resources, while the equivalent constraint method owns a high computational complexity due to the large-dimensional LMIs. Therefore, further research is still essential to design efficient algorithms with lower complexity for the RIS-assisted PLS problems under imperfect CSI.

The above-mentioned literature showcases the superiority of RIS in enhancing the PLS. However, the RIS can only reflect the incident signals, which will cause the geographical restrictions and limit its deployment flexibility. Recently, the STAR-RIS is proposed to handle the drawbacks of reflecting-only RIS [27], [28], [29]. Compared to the reflecting-only RIS, each element of STAR-RIS can simultaneously transmit and reflect the incident signals by manipulating both the electric and magnetic currents. Meanwhile, besides the reflection coefficients, the transmission coefficients are also incorporated into the passive beamforming of STAR-RIS, which can provide more degrees of freedoms (DoFs) to configure the propagation environment and further enhance the PLS [30].

Despite the dramatic versatility of STAR-RIS, the investigation of STAR-RIS, notably in the domain of STAT-RIS-assisted PLS, is still at an early stage. For the simplified scenario with one legitimate user on each side of the STAR-RIS, the secrecy rate maximization problems are studied in [31] and [32], where it is unveiled that the STAR-RIS is superior than conventional RIS in the security enhancement. Y. Zhang *et al.* [33] introduce the multi-carrier NOMA to facilitate secure communications among numerous legitimate users in STAR-RIS-assisted systems, where the beamforming of STAR-RIS and BS, the power allocation coefficients, as well as the user pairing are jointly optimized to maximize the secrecy sum rate. Furthermore, Y. Han *et al.* [34] combine the

AN technique into the STAR-RIS-assisted NOMA systems, and observe that the influence of AN on enhancing the secrecy rate is not significant when there is a large number of STAR-RIS elements.

We note that existing studies related to the STAR-RIS-assisted PLS primarily concentrate on the downlink secure transmission, while there are more security risks in the uplink for the STAR-RIS-assisted systems. This is because the STAR-RIS is required to work in the energy splitting (ES) or mode switching (MS) protocol in order to simultaneously serve users located at the reflection and transmission spaces [35]. Nevertheless, each element of STAR-RIS is symmetrical and own the same transmission and reflection coefficients on both sides [36]. Thereby, it is inevitable that a portion of users' signals will leak to the other side of the STAR-RIS, where the base station cannot receive the leaked signals. More detrimentally, this portion of signals may also be wiretapped by malicious eavesdroppers. Although the uplink PLS for the STAR-RIS is investigated in [30] and [37], the considered scenarios compose of only one eavesdropper. Therefore, in order to mitigate the uplink security risks arising from the symmetry of STAR-RIS, the strategy of passive beamforming and resource allocation in the STAR-RIS-assisted systems with multiple eavesdroppers still require further investigation.

Motivated by the above observations, in this paper, we investigate the sum secrecy rate maximization problems for the STAR-RIS-assisted uplink systems with one cooperative jammer and dual eavesdroppers, where both the perfect and imperfect CSI are considered. The key contributions of this paper are outlined below.

1) We propose a STAR-RIS-assisted uplink systems with one cooperative jammer and dual eavesdroppers, where multiple IoT devices transmit signals to the BS via NOMA. The STAR-RIS operates in the ES protocol and employs a more practical coupled phase shift model. The cooperative jammer transmit the AN to degrade the eavesdropping rate, and the dual eavesdroppers can cooperate with each other to form a malicious colluding. Aiming to enhance the PLS, we formulate the problems under both the perfect and imperfect CSI.
2) To solve the formulated sum secrecy rate maximization problem with perfect CSI, we propose a deep reinforcement learning algorithm based on the DDPG framework to jointly optimize the channel allocation, transmit power and coefficient matrices, where a novel mapping method is designed to reduce the dimension of action space. For the case of imperfect CSI, based on the sample average approximation (SAA) method, the arbitrary distorted noise is introduced to the state space to retrain the network, and the solution to the formulated problem with imperfect CSI can be obtained without causing additional computational complexity.
3) Simulation results reveal that: 1) The symmetry of STAR-RIS results in severe information leakage, and the sum secrecy rate further degrades when the dual eavesdroppers collaborate with each other. 2) The STAR-RIS with independent phase shift can achieve higher sum secrecy rate than that with coupled phase shift, while the performance gap is trivial when there are fewer STAR-RIS elements. 3) Our proposed algorithm can compensate for the impacts of the imperfect CSI, and the sum secrecy rate decreases with the increase of CSI uncertainty.

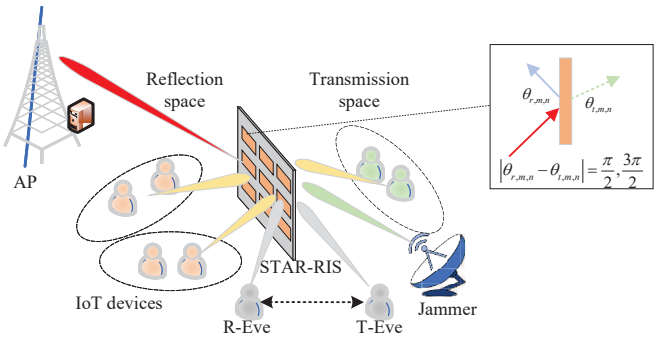


Fig. 1. The STAR-RIS-assisted secure NOMA system.

The structure of this paper is outlined as follows. The system model and the problems for maximizing the sum secrecy rate under both perfect and imperfect CSI are presented in Section II. In Section III, we detail the algorithms developed to address these problems. Section IV displays various numerical results, and Section V draws the conclusions.

II. System Model and Problem Formulation

A. System Model

As shown in Fig. 1, we consider a STAR-RIS-assisted secure NOMA system, including multiple legitimate IoT devices, two eavesdroppers, one BS, and one cooperative jammer. The eavesdroppers located at the reflection space and transmission space are represented as R-Eve and T-Eve, respectively. The STAR-RIS equipped with M elements operates in the ES protocol. Compared to the TS and MS protocols, the STAR-RIS in ES protocol can transmit and reflect the signals incident on both sides of the surface at the same time, and hence simultaneously assist the signal transmission of multiple IoT devices located at the reflection space and transmission space [28]. At time slot n, denote the amplitude coefficients of the m-th element of STAR-RIS for transmission and reflection as $\beta_{t,m,n}$ and $\beta_{r,m,n}$, with $\beta_{t,m,n}, \beta_{r,m,n} \in [0,1], \beta_{t,m,n}+\beta_{r,m,n}=1$. The phase shifts of the m-th element for transmission and reflection are given by $\theta_{t,m,n}$ and $\theta_{r,m,n}$, respectively. Considering the hardware constraint, a coupled phase shift model is adopted for the STAR-RIS, where we have [38]

$$|\theta_{r,m,n}-\theta_{t,m,n}|=\frac{\pi}{2}\text{or}\frac{3\pi}{2}, \forall m,n. \tag{1}$$

Thus, the coefficient matrices of STAR-RIS for transmission and reflection can be given by $\mathbf{u}_{r,n} = \text{diag}\left(\sqrt{\beta_{r,1,n}}e^{j\theta_{r,1,n}}, \sqrt{\beta_{r,2,n}}e^{j\theta_{r,2,n}}, ..., \sqrt{\beta_{r,M,n}}e^{j\theta_{r,M,n}}\right)$ and $\mathbf{u}_{t,n} = \text{diag}\left(\sqrt{\beta_{t,1,n}}e^{j\theta_{t,1,n}}, \sqrt{\beta_{t,2,n}}e^{j\theta_{t,2,n}}, ..., \sqrt{\beta_{t,M,n}}e^{j\theta_{t,M,n}}\right)$, respectively.

With the aid of STAR-RIS, IoT device i transmit signal $x_{i,n}$ to the BS with $\mathbb{E}|x_{i,n}|^2 = 1$. It is assumed that

obstacles block the direct links between the BS/Eves and IoT devices [30]. At time slot n, the channels from IoT device i and the jammer to the STAR-RIS are given by $\mathbf{h}_{i,\mathrm{RIS},n} \in \mathbb{C}^{1\times M}$ and $\mathbf{h}_{\mathrm{J},\mathrm{RIS},n} \in \mathbb{C}^{1\times M}$, respectively. The channels from the STAR-RIS to the BS, R-Eve, and T-Eve can be represented as $\mathbf{h}_{\mathrm{RIS},\mathrm{BS},n} \in \mathbb{C}^{1\times M}$, $\mathbf{h}_{\mathrm{RIS},\mathrm{R-Eve},n} \in \mathbb{C}^{1\times M}$, and $\mathbf{h}_{\mathrm{RIS},\mathrm{T-Eve},n} \in \mathbb{C}^{1\times M}$, respectively.

We adopt the Rician fading channel model for all involved channels, which includes both line-of-sight (LoS) and non-LoS (NLoS) elements. These channels are assumed to be stable within each time slot. For example, the channel from IoT device i to the STAR-RIS can be expressed as [30]

$$\mathbf{h}_{i,\mathrm{RIS},n}=\sqrt{\rho d_{i,\mathrm{RIS},n}^{-\alpha}}\left(\sqrt{\frac{\varepsilon}{1+\varepsilon}}\mathbf{h}_{i,\mathrm{RIS},n}^{\mathrm{LoS}}+\sqrt{\frac{1}{1+\varepsilon}}\mathbf{h}_{i,\mathrm{RIS},n}^{\mathrm{NLoS}}\right), \tag{2}$$

where ρ represents the path loss at a reference distance of 1 meter. $d_{i,\mathrm{RIS},n}$ is the distance between the i-th IoT device and the STAR-RIS. α is the path loss exponent, while ε indicates the Rician factor. $\mathbf{h}_{i,\mathrm{RIS,n}}^{\mathrm{LoS}}$ and $\mathbf{h}_{i,\mathrm{RIS},n}^{\mathrm{NLoS}}$ refer to the LoS and NLoS components, respectively. Notably, the STAR-RIS can simultaneously enhance the signals of IoT devices located at the reflection and transmission spaces, and the STAR-RIS will also leak the information to both the R-Eve and T-Eve.

The received signal of BS at time slot n is expressed as

$$\begin{aligned} y_{\mathrm{BS},n} = &\sum_{k=1}^{K}\sum_{i=1}^{I}\omega_{i,k,n}\sqrt{p_{i,n}}\mathbf{h}_{i,\mathrm{RIS},n}\mathbf{u}_{X,n}(\mathbf{h}_{\mathrm{RIS},\mathrm{BS},n})^{H}x_{i,n} \\ &+\sqrt{p_{\mathrm{J},n}}\mathbf{h}_{\mathrm{J},\mathrm{RIS},n}\mathbf{u}_{t,n}(\mathbf{h}_{\mathrm{RIS},\mathrm{BS},n})^{H}x_{\mathrm{J},n}+n_{\mathrm{BS}}, \end{aligned} \tag{3}$$

where $X \in \{r,t\}$. $p_{i,n}$ and $p_{\mathrm{J},n}$ are the transmit power of IoT device i and Jammer at time slot n. $n_{\mathrm{BS}} \sim \mathcal{CN}\left(0,\sigma^2\right)$ is the noise at the BS. $\omega_{i,k,n} \in \{0,1\}$ indicates whether the k-th channel is assigned to IoT device i at time slot n. The NOMA protocol is adopted when the IoT devices upload their information to the BS. The bandwidth B is divided into K channels, represented as $\mathcal{K}=\{1,2,...,K\}$. Assume that each IoT device is assigned only one channel and each channel is capable of supporting up to $L=I/K$ IoT devices. Therefore, it follows that

$$\sum_{k=1}^{K}\omega_{i,k,n}=1,\forall i,n,\sum_{i=1}^{I}\omega_{i,k,n}\le L,\forall k,n. \tag{4}$$

To mitigate the effects of jamming on the BS, the BS can perform zero-forcing (ZF) technique, and thus we have $\mathbf{h}_{\mathrm{J},\mathrm{RIS},n}\mathbf{u}_{t,n}(\mathbf{h}_{\mathrm{RIS},\mathrm{BS},n})^{H}x_{\mathrm{J},n}=0$. Hence, the achievable rate of IoT device i at the n-th time slot can be expressed as

$$R_{i,n}=\sum_{k=1}^{K}\frac{B}{K}\log(1+\frac{\omega_{i,k,n}p_{i,n}(\mathbf{h}_{i,\mathrm{RIS},n}\mathbf{u}_{X,n}(\mathbf{h}_{\mathrm{RIS},\mathrm{BS},n})^{H})^{2}}{\zeta_{i,k,n}+\sigma^2}), \tag{5}$$

where $\zeta_{i,k,n}=\sum_{\pi_k(i)\ge\pi_k(\tilde{i})}\omega_{\tilde{i},k,n}p_{\tilde{i},n}(\mathbf{h}_{\tilde{i},\mathrm{RIS},n}\mathbf{u}_{X,n}\mathbf{h}_{\mathrm{RIS},\mathrm{BS},n}^{H})^{2}$ $\tilde{i}\in I\backslash\{i\}$ represents the inter-user interference. The BS adopts the SIC technique to decode signals from multiple IoT devices. π_k is the given decoding order of IoT devices on the k-th channel [1].

For the R/T-Eve, the received signal can be given by

$$\begin{aligned} y_{\mathrm{R/T-Eve},n}=&\sqrt{p_{\mathrm{J},n}}\mathbf{h}_{\mathrm{J},\mathrm{RIS},n}\mathbf{u}_{X,n}\mathbf{h}_{\mathrm{RIS},\mathrm{R/T-Eve},n}^{H}x_{\mathrm{J},n}+n_{\mathrm{Eve}} \\ &+\sum_{k=1}^{K}\sum_{i=1}^{I}\omega_{i,k,n}\sqrt{p_{i,n}}\mathbf{h}_{i,\mathrm{RIS},n}\mathbf{u}_{X,n}\mathbf{h}_{\mathrm{RIS},\mathrm{R/T-Eve},n}^{H}x_{i,n}, \end{aligned} \tag{6}$$

where $n_{\mathrm{Eve}} \sim \mathcal{CN}\left(0,\sigma^2\right)$ is the AWGN at the R/T-Eve.

The dual colluding Eves can be deemed as a super Eve. Thus, to ease the expression of eavesdropping rate of this super Eve, we first define

$$\begin{aligned} \mathbf{H}_{\mathrm{RIS-Eve},n}=&\mathbf{u}_{X,n}\mathbf{h}_{\mathrm{RIS},\mathrm{R-Eve},n}(\mathbf{h}_{\mathrm{RIS},\mathrm{R-Eve},n})^{H}\mathbf{u}_{X,n} \\ &+\mathbf{u}_{\tilde{X},n}\mathbf{h}_{\mathrm{RIS},\mathrm{T-Eve},n}(\mathbf{h}_{\mathrm{RIS},\mathrm{T-Eve},n})^{H}\mathbf{u}_{\tilde{X},n}. \end{aligned} \tag{7}$$

Additionally, at the super Eve, the worst-case assumption is considered, which means it can decode signals without suffering from the inter-user interference. Hence, the eavesdropping rate of the super Eve for IoT device i can be expressed as

$$C_{i,n}=\sum_{k=1}^{K}\log\left(1+\frac{\omega_{i,k,n}p_{i,n}\mathbf{h}_{i,\mathrm{RIS},n}\mathbf{H}_{\mathrm{RIS-Eve},n}(\mathbf{h}_{i,\mathrm{RIS},n})^{H}}{\sigma^2+p_{\mathrm{J},n}\mathbf{h}_{\mathrm{J},\mathrm{RIS},n}\mathbf{H}_{\mathrm{RIS-Eve},n}(\mathbf{h}_{J,\mathrm{RIS},n})^{H}}\right). \tag{8}$$

B. Problem Formulation

1) Perfect CSI: In this paper, we focus on enhancing the total secrecy rate of IoT devices in the STAR-RIS-assisted NOMA system by jointly optimizing the channel allocation, transmit power and coefficient matrices. Under the assumption that the perfect CSI is available at both the BS and STAR-RIS, we can formulate the optimization problem as

$$\max_{\boldsymbol{\omega},\mathbf{p},\mathbf{u}_{\mathbf{X},\mathbf{n}}}\sum_{n=1}^{N}\sum_{i=1}^{I}[R_{i,n}-C_{i,n}]^{+} \tag{9a}$$

$$\text{s.t.}\sum_{k=1}^{K}\omega_{i,k,n}=1,\forall i,n,\sum_{i=1}^{I}\omega_{i,k,n}\le L,\forall k,n, \tag{9b}$$

$$\beta_{r,m,n}+\beta_{t,m,n}=1,\forall m,n, \tag{9c}$$

$$\beta_{r,m,n},\beta_{t,m,n}\in[0,1],\forall m,n, \tag{9d}$$

$$|\theta_{r,m,n}-\theta_{t,m,n}|=\frac{\pi}{2}\text{or}\frac{3\pi}{2},\forall m,n, \tag{9e}$$

$$\theta_{r,m,n},\theta_{t,m,n}\in[0,2\pi],\forall m,n, \tag{9f}$$

$$p_{i,n}\le P_{\max},\forall i,n, \tag{9g}$$

where $[\bullet]^{+}=\max(\bullet,0)$. (9b) are the channel allocation constraints. (9c) - (9f) are constrains on the amplitudes and phase shifts of STAR-RIS. (9g) limits the transmit power of IoT devices.

2) Imperfect CSI: In the STAR-RIS-assisted NOMA systems, it is challenging to obtain instantaneous and perfect CSI due to the passive feature of STAR-RIS and the time-varying nature of wireless channels. Besides, from the perspective of

[1]The objective function and constrains of the formulated problems in this paper are independent of decoding order, and the signals can be decoded correctly by the BS under any decoding orders when the achievable rate of IoT devices is larger than its target data rate [39]. Hence, we adopt a predefined decoding order and the decoding order optimization problem can be left as our future work.

PLS, the transmission of confidential data is more sensitive to the uncertainties of CSI. Consequently, it is crucial to consider channel estimation errors when designing strategies of passive beamforming and resource allocation. Denote the cascaded channels between IoT device i and BS/Eves as $\mathbf{G}_{i,\mathrm{BS},n} = \left(\mathbf{h}_{i,\mathrm{RIS},n}\right)^H \mathbf{h}_{\mathrm{RIS,BS},n}$ and $\mathbf{G}_{i,\mathrm{R/T-Eve},n} = \left(\mathbf{h}_{i,\mathrm{RIS},n}\right)^H \mathbf{h}_{\mathrm{RIS,R/T-Eve},n}$, respectively. The cascaded channel between Jammer and R/T-Eve is given by $\mathbf{G}_{\mathrm{J,R/T-Eve},n} = \left(\mathbf{h}_{\mathrm{J,RIS},n}\right)^H \mathbf{h}_{\mathrm{RIS,R/T-Eve},n}$. Then, the channel estimation errors can be expressed as

$$\begin{aligned} &\Delta\mathbf{G}_{i,\mathrm{BS},n} = \mathbf{G}_{i,\mathrm{BS},n} - \hat{\mathbf{G}}_{i,\mathrm{BS},n}, \\ &\Delta\mathbf{G}_{i,\mathrm{R/T-Eve},n} = \mathbf{G}_{i,\mathrm{R/T-Eve},n} - \hat{\mathbf{G}}_{i,\mathrm{R/T-Eve},n}, \\ &\Delta\mathbf{G}_{\mathrm{J,R/T-Eve},n} = \mathbf{G}_{\mathrm{J,R/T-Eve},n} - \hat{\mathbf{G}}_{\mathrm{J,R/T-Eve},n}, \end{aligned} \tag{10}$$

where $\hat{\mathbf{G}}_{i,\mathrm{BS},n}, \hat{\mathbf{G}}_{i,\mathrm{R/T-Eve},n}$, and $\hat{\mathbf{G}}_{\mathrm{J,R/T-Eve},n}$ represent corresponding estimated channels. Different from [24], [25], where the channel estimation error belongs to a bounded region. In this paper, considering the practical channel estimation error is generally unbounded, we adopt the statistical error model, which assumes that the channel estimation errors follow the circularly symmetric complex Gaussian (CSCG) distribution with zero mean. The error covariance matrices are represented as $\mathbf{\Theta}_{i,\mathrm{BS},n}$, $\mathbf{\Theta}_{i,\mathrm{R/T-Eve},n}$, and $\mathbf{\Theta}_{\mathrm{J,R/T-Eve},n}$, respectively. Define the sample space $\Psi \triangleq \left\{\mathbf{G}_{i,\mathrm{BS},n}(j), \mathbf{G}_{i,\mathrm{R/T-Eve},n}(j), \mathbf{G}_{\mathrm{J,R/T-Eve},n}(j)\right\}$, where j is the index of random realizations. It is found that the channel can be regarded as the sum of the estimated channel and channel estimation errors. If the CSI can be estimated perfectly, the samples in the space Ψ are equal to each other. Otherwise, the sample may be varying with j. Therefore, with imperfect CSI, problem (9) can be reformulated as a stochastic optimization problem to maximize the expectation of sum secrecy rate, i.e.,

$$\max_{\boldsymbol{\omega},\mathbf{p},\mathbf{u}\mathbf{x},\mathbf{n}} \mathbb{E}_j \left[\sum_{n=1}^{N}\sum_{i=1}^{I}\left[R_{i,n}(j) - C_{i,n}(j)\right]^+\right] \tag{11a}$$

$$\text{s.t.} \sum_{k=1}^{K}\omega_{i,k,n}(j) = 1, \forall i,n, \sum_{i=1}^{I}\omega_{i,k,n}(j) \le L, \forall k,n, \tag{11b}$$

$$\beta_{r,m,n}(j) + \beta_{t,m,n}(j) = 1, \forall m,n, \tag{11c}$$

$$\beta_{r,m,n}(j), \beta_{t,m,n}(j) \in [0,1], \forall m,n, \tag{11d}$$

$$|\theta_{r,m,n}(j) - \theta_{t,m,n}(j)| = \frac{\pi}{2} \text{or} \frac{3\pi}{2}, \forall m,n, \tag{11e}$$

$$\theta_{r,m,n}(j), \theta_{t,m,n}(j) \in [0,2\pi], \forall m,n \tag{11f}$$

$$p_{i,n}(j) \le P_{\max}, \forall i,n. \tag{11g}$$

III. Solution to the Formulated Problems

In this section, aiming to tackle the formulated problems, we first propose an improved DRL algorithm to address problem (9) with perfect CSI, where the DDPG framework is utilized to handle the action space with the hybrid of continuous actions ($\mathbf{p}$, $\boldsymbol{\theta}$, and $\boldsymbol{\beta}$) and discrete actions ($\boldsymbol{\omega}$), and a novel action mapping method is designed to further reduce the dimension of action space. Then, by introducing the arbitrary distorted noise to the state space and adjusting the sample process, the proposed algorithm for perfect CSI is adapted to address the problem with imperfect CSI.

A. Solution to the Perfect CSI

Problem (9) contains the optimization of binary variables and the optimization variables are highly coupled with each other. Although the traditional optimization methods can achieve the suboptimal solution, they require complicated iterations and own highly computational complexity, which render the execution time intolerable. Thus, in this paper, considering that the DRL algorithms can quickly respond to different environments after offline training, we first transform Problem (9) into a Markov decision process (MDP), and then propose an improved DDPG algorithm to maximize the sum secrecy rate. Different from proximal policy optimization (PPO) which outputs the probability distribution of different actions, the DDPG agent can directly output the action, which is suitable for the STAR-RIS-assisted NOMA systems with the requirement of real-time control. Additionally, as an off-policy algorithm, the experience replay technology in DDPG allows the algorithm to output the best policy from the whole training period, while the on-policy PPO algorithm overly focuses on the current rewards [40].

1) State space: In our proposed system, the state of time slot n requires to include the CSI of all links, i.e., $\mathbf{h}_{i,\mathrm{RIS},n}$, $\mathbf{h}_{\mathrm{J,RIS},n}$, $\mathbf{h}_{\mathrm{RIS,BS/R-Eve/T-Eve},n}$. Nevertheless, due to the passive feature of STAR-RIS, the CSI of links related to STAR-RIS is difficult to be estimated directly. To tackle this challenge, most of the existing literature pays attention to the channel estimation methods for cascaded CSI, and reveals that the cascaded CSI is sufficient for passive beamforming and resource allocation design in the STAR-RIS-assisted systems [41]. Thus, the state of time slot n comprises of the cascaded CSI among IoT devices/Jammer and Eves/BS, i.e.,

$$\mathbf{s}_n = \left\{\mathbf{G}_{i,\mathrm{BS},n}, \mathbf{G}_{i,\mathrm{R/T-Eve},n}, \mathbf{G}_{\mathrm{J,R/T-Eve},n}\right\}. \tag{12}$$

2) Action space: According to the state of time slot n, the agent takes actions to maximize its reward. All optimization variables in Problem (9) need to be included in the action space. Hence, the action space is given by

$$\mathbf{a}_n = \left\{\mathbf{a}_{\omega,n}, \mathbf{a}_{p,n}, \mathbf{a}_{\theta,n}, \mathbf{a}_{\beta,n}\right\}. \tag{13}$$

It is worth noting that the continuous and normalized actions output by the DDPG agent cannot be executed directly for the STAR-RIS-assisted NOMA system. Therefore, we design a mapping method to transform the output of agent into executable actions. To be more specific, action $\mathbf{a}_{\omega,n} = \{a^{\omega}_{1,1,n}, a^{\omega}_{1,2,n}, ..., a^{\omega}_{k,l,n} ..., a^{\omega}_{K,L,n}\}$ indicates the channel allocation, with $a^{\omega}_{k,l,n} \in [0,1]$. If $\forall \left\lceil a^{\omega}_{k,l,n} I \right\rceil = i, \omega_{i,k,n} = 1$; otherwise $\omega_{i,k,n} = 0$, where $\lceil \bullet \rceil$ is the ceiling operation. $\mathbf{a}_{p,n} = \{a^{p}_{1,n}, a^{p}_{2,n}, ..., a^{p}_{I,n}\}$, $a^{p}_{i,n} \in [0,1]$ represent the transmit power of IoT devices. We have $p_{i,n} = a^{p}_{i,n} P_{\max}$. $\mathbf{a}_{\theta,n} = \{a^{\theta}_{1,n}, a^{\theta}_{2,n} ..., a^{\theta}_{M,n}\}$ is the phase shift of STAR-RIS for transmission. Although the phase shifts of STAR-RIS for transmission and reflection satisfy equality constraint (9e), the phase shift for reflection cannot be directly obtained

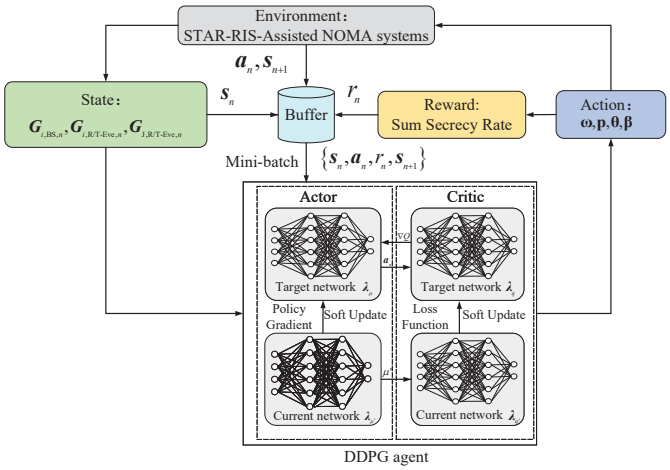


Fig. 2. The DDPG framework for solving problem (9) with perfect CSI.

through that for transmission, since the phase shift for reflection still need to determine whether to add $\pi/2$ or $3\pi/2$ on the phase shift for transmission. It is intuitively that an additional indicators can be employed to address this issue. However, in this case, there will be $2M$ actions for phase shift. Fortunately, based on the symmetry of trigonometric functions, it is observed that constraint (9e) can be written as

$$\theta_{r,m,n} - \theta_{t,m,n} = -\frac{\pi}{2} \text{or} \frac{\pi}{2}. \tag{14}$$

Through mapping $a_{m,n}^{\theta}$ into $[-1, 1]$, the problem of $\pi/2$ or $-\pi/2$ can be solved without increasing the number of actions. If $a_{m,n}^{\theta} \geq 0$, $\theta_{r,m,n} = \theta_{t,m,n} + \pi/2$; otherwise $\theta_{r,m,n} = \theta_{t,m,n} - \pi/2$. While for the phase shift for transmission, we have $\theta_{t,m,n} = 2\pi a_{m,n}^{\theta}$. $\mathbf{a}_{\beta,n} = \{a_{1,n}^{\beta}, a_{2,n}^{\beta} ..., a_{M,n}^{\beta}\}$, $a_{m,n}^{\beta} \in [0, 1]$ is the amplitude of STAR-RIS for transmission. According to constraint (9c), it is obtained that

$$\beta_{t,m,n} = a_{m,n}^{\beta}, \beta_{r,m,n} = 1 - a_{m,n}^{\beta}. \tag{15}$$

3) Reward: The reward can evaluate the agent's performance by mapping the states and actions to a scalar value. Specifically, when the agent executes action $\mathbf{a}_n$ after observing state $\mathbf{s}_n$, the agent will obtain reward r_n. The design of reward function should align with the objective function and constrains in Problem (9). The mapping of action space ensures the satisfaction of constraints (9c) - (9g). Therefore, to meet constraint (9b), the reward function can be formulated as

$$r_n = \begin{cases} 0, \text{if (9b) is violated;} \\ \sum_{i=1}^{I} [R_{i,n} - C_{i,n}]^+, \text{otherwise.} \end{cases} \tag{16}$$

As shown in Fig. 2, the DDPG agent is composed of the actor network and the critic network. Both of the actor and critic networks include the current and target networks. Denote the parameters for the current networks of the actor and critic as $\boldsymbol{\lambda}_{\mu}$ and $\boldsymbol{\lambda}_{q}$, respectively. Correspondingly, the parameters for the target networks of the actor and critic are given by $\boldsymbol{\lambda}_{\mu'}$ and $\boldsymbol{\lambda}_{q'}$.

The actor network produces actions based on the policy μ which depends on the parameters of actor networks. For example, at time slot n, the actions of agent generated by the actor's current network can be expressed as

$$\mathbf{a}_n = \mu(\mathbf{s}_n|\boldsymbol{\lambda}_{\mu}) + \eta, \tag{17}$$

where η is the action noise.

The critic network evaluates the actor's performance via Q value. The Q value represent the expected accumulated reward obtained by the agent when continuously executing policy μ. According to the Bellman equation, the Q value function can be defined as [42]

$$Q(\mathbf{s}_n, \mathbf{a}_n) = \mathbb{E}[r(\mathbf{s}_n, \mathbf{a}_n) + \gamma Q(\mathbf{s}_{n+1}, \mathbf{a}_{n+1})], \tag{18}$$

where γ stands for the discount factor. In the training of agent, the critic network updates $\boldsymbol{\lambda}_q$ through the loss function minimization. The loss function is given by

$$L(\boldsymbol{\lambda}_q) = \frac{1}{e} \sum_{e} [y_n - Q(\mathbf{s}_n, \mathbf{a}_n|\boldsymbol{\lambda}_q)]^2, \tag{19}$$

where e represents the size of mini-batch. y_n stands for the target value output by the critic's target network, which can be expressed as

$$y_n = r_n + \alpha Q'(\mathbf{s}_{n+1}, \mathbf{a}_{n+1}|\boldsymbol{\lambda}_{q'}). \tag{20}$$

Subsequently, with the gradient of Q value function, the actor network is trained by maximizing the function J. The function J evaluates the policy μ, which can be expressed as

$$J(\boldsymbol{\lambda}_{\mu}) = \frac{1}{e} \sum_{e} Q(\mathbf{s}_n, \mathbf{a}_n|\boldsymbol{\lambda}_q)|_{\mathbf{a}_n=\mu(\mathbf{s}_n|\boldsymbol{\lambda}_{\mu})}. \tag{21}$$

Then, the parameter of current network of actor can be updated as

$$\Delta\boldsymbol{\lambda}_{\mu} \approx \tau \nabla_{\boldsymbol{\lambda}_{\mu}} J, \tag{22}$$

where τ is a positive constant regarding the step size. $\nabla_{\boldsymbol{\lambda}_{\mu}} J$ is the gradient of function J. According to the chain rule [42], we have

$$\nabla_{\boldsymbol{\lambda}_{\mu}} J = \frac{1}{e} \sum_{e} \nabla_{\mathbf{a}_n} Q(\mathbf{s}_n, \mathbf{a}_n|\boldsymbol{\lambda}_q)|_{\mathbf{a}_n=\mu(\mathbf{s}_n)} \nabla_{\boldsymbol{\lambda}_{\mu}} \mu(\mathbf{s}_n|\boldsymbol{\lambda}_{\mu}), \tag{23}$$

where $\nabla_{\mathbf{a}_n} Q(\mathbf{s}_n, \mathbf{a}_n|\boldsymbol{\lambda}_q)|_{\mathbf{a}_n=\mu(\mathbf{s}_n)}$ is provided by the critic network. $\nabla_{\boldsymbol{\lambda}_{\mu}} \mu(\mathbf{s}_n|\boldsymbol{\lambda}_{\mu})$ is calculated by the actor's optimizer.

To improve the stability of learning, the agent softly update the parameters of target networks with a small constant ε, i.e.,

$$\begin{aligned} \boldsymbol{\lambda}_{q'} &= \varepsilon\boldsymbol{\lambda}_q + (1-\varepsilon)\boldsymbol{\lambda}_{q'} \\ \boldsymbol{\lambda}_{\mu'} &= \varepsilon\boldsymbol{\lambda}_{\mu} + (1-\varepsilon)\boldsymbol{\lambda}_{\mu'}. \end{aligned} \tag{24}$$

Based on the above analysis, the details of the developed DDPG algorithm for solving Problem (9) with perfect CSI are summarized in Algorithm 1.

With Algorithm 1, through continuously updating the parameters of actor and critic networks, we can obtain a trained actor network $\mu(\mathbf{s}_n|\boldsymbol{\lambda}_{\mu})$, which can output actions with respect to the channel states, and then map the actions to the channel allocation, transmit power and coefficient matrices for the STAR-RIS-assisted NOMA systems to achieve the maximum sum secrecy rate.

Algorithm 1 The DDPG algorithm for solving Problem (9) with perfect CSI

1. Initialize the environment, the network parameters $\boldsymbol{\lambda}_\mu$, $\boldsymbol{\lambda}_q$, $\boldsymbol{\lambda}'_\mu$, and $\boldsymbol{\lambda}'_q$, the buffer D;
2. **for** each episode **do**:
3. Reset the environment as $\mathbf{s}_0$.
4. **for** each step **do**:
5. Observe state $\mathbf{s}_n$ and execute action $\mathbf{a}_n$;
6. Calculate reward r_n based on (16) and observe state $\mathbf{s}_{n+1}$;
7. Store the transition $\{\mathbf{s}_n, \mathbf{a}_n, r_n, \mathbf{s}_{n+1}\}$ into the buffer;
8. Randomly sample a mini-batch of transitions from the buffer;
9. Update the current network of critic by minimizing (19);
10. Update the current network of actor by maximizing (21);
11. Update the target networks of actor and critic according to (24);
12. **end for**
13. **end for**
14. **Output:** The well-trained actor network $\mu\left(\mathbf{s}_n|\boldsymbol{\lambda}_\mu\right)$.

B. Solution to the Imperfect CSI

In this section, we extend Algorithm 1 to solve problem (11). The framework of proposed improved DDPG algorithm is shown in Fig. 3. Problem (11) is a stochastic optimization problem as the channel estimation errors are described by the probabilistic variables. The SAA method is an efficient solution to deal with the infinite possibilities of samples in Ψ by transforming the original problem into an approximate deterministic optimization problem [43]. More interestingly, we find that the basic idea of SAA method coincides with the training process of DDPG agent, which is approximating the objective function (11a) by the arithmetic mean of Z samples randomly chosen from Ψ. Motivated by this observation, we aim to extend the DDPG algorithm to solve problem (11).

Specifically, if the CSI can be perfectly estimated, the agent observes state $\mathbf{s}_n = \left\{\mathbf{G}_{i,\mathrm{BS},n}, \mathbf{G}_{i,\mathrm{R/T-Eve},n}, \mathbf{G}_{\mathrm{J,R/T-Eve},n}\right\}$, and executes action $\mathbf{a}_n$. However, it is rather challenging to obtain the perfect CSI. What the agent actually observed is one of the possible realizations of real channels. Therefore, in order to describe the uncertainties of CSI, the z-th possible realization of state $\mathbf{s}_n$ is given by

$$\mathbf{s}_n\left(z\right)=\left\{\mathbf{G}^e_{i,\mathrm{BS},n}\left(z\right), \mathbf{G}^e_{i,\mathrm{R/T-Eve},n}\left(z\right), \mathbf{G}^e_{\mathrm{J,R/T-Eve},n}\left(z\right)\right\}, \tag{25}$$

where

$$\begin{aligned}
&\mathbf{G}^e_{i,\mathrm{BS},n}\left(z\right) = \mathbf{G}_{i,\mathrm{BS},n} + \Delta\mathbf{G}_{i,\mathrm{BS},n}\left(z\right),\\
&\mathbf{G}^e_{i,\mathrm{R/T-Eve},n}\left(z\right) = \mathbf{G}_{i,\mathrm{R/T-Eve},n} + \Delta\mathbf{G}_{i,\mathrm{R/T-Eve},n}\left(z\right),\\
&\mathbf{G}^e_{\mathrm{J,R/T-Eve},n}\left(z\right) = \mathbf{G}_{\mathrm{J,R/T-Eve},n} + \Delta\mathbf{G}_{\mathrm{J,R/T-Eve},n}\left(z\right).
\end{aligned} \tag{26}$$

In this case, the reward by executing action $\mathbf{a}_n$ can be expressed as

$$r_n\left(z\right) = \begin{cases} 0, \text{if } \ (11b) \ \text{ is violated} \\ \sum\limits_{i=1}^{I}\left[R_{i,n}(z) - C_{i,n}(z)\right]^+, \text{otherwise.} \end{cases} \tag{27}$$

where

$$R_{i,n}\left(z\right) = \sum_{k=1}^{K}\log\left(1+\frac{\omega_{i,k,n}p_{i,n}\left(\mathbf{u}_{X,n}\mathbf{G}^e_{i,\mathrm{BS},n}\left(z\right)\right)^2}{\zeta_{i,k,n}\left(z\right)+\sigma^2}\right), \tag{28}$$

$$\zeta_{i,k,n}\left(z\right) = \sum\nolimits_{\pi_k(i)\geq\pi_k(\tilde{i})}\omega_{\tilde{i},k,n}p_{\tilde{i},n}\left(\mathbf{u}_{X,n}\mathbf{G}^e_{\tilde{i},\mathrm{BS},n}\left(z\right)\right)^2, \tag{29}$$

$$C_{i,n}\left(z\right) = \sum_{k=1}^{K}\log\left(1+\frac{\omega_{i,k,n}p_{i,n}\Gamma_i\left(z\right)}{\sigma^2+p_{\mathrm{J},n}\Gamma_i\left(z\right)}\right), \tag{30}$$

$$\Gamma_i\left(z\right) = \left(\mathbf{u}_{X,n}\mathbf{G}^e_{i,\mathrm{R-Eve},n}\left(z\right)\right)^2 + \left(\mathbf{u}_{\tilde{X},n}\mathbf{G}^e_{i,\mathrm{T-Eve},n}\left(z\right)\right)^2. \tag{31}$$

Then, as can be seen from Fig. 3, the transition stored into the buffer is reformulated as $\{\hat{\mathbf{s}}_n, \mathbf{a}_n, \hat{r}_n, \mathbf{s}_{n+1}\}$, where

$$\hat{\mathbf{s}}_n = \frac{1}{Z}\sum_{z=1}^{Z}\mathbf{s}_n\left(z\right), \hat{r}_n = \frac{1}{Z}\sum_{z=1}^{Z}r_n\left(z\right). \tag{32}$$

With the modified transitions, the loss function for critic and the function J for actor can be derived as

$$L\left(\hat{\boldsymbol{\lambda}}_q\right) = \frac{1}{e}\sum_{e}\left[\hat{y}_n - Q\left(\hat{\mathbf{s}}_n, \mathbf{a}_n|\hat{\boldsymbol{\lambda}}_q\right)\right]^2, \tag{33}$$

$$J\left(\hat{\boldsymbol{\lambda}}_\mu\right) = \frac{1}{e}\sum_{e}Q\left(\hat{\mathbf{s}}_n, \mathbf{a}_n|\hat{\boldsymbol{\lambda}}_q\right)|_{\mathbf{a}_n=\mu\left(\hat{\mathbf{s}}_n|\hat{\boldsymbol{\lambda}}_\mu\right)}. \tag{34}$$

Similar to the training process of the above subsection for perfect CSI, through updating the parameters of actor and critic networks, a well-trained actor network which can compensate for the effects of imperfect CSI is finally obtained. The details of the improved DDPG algorithm for solving Problem (11) with imperfect CSI are summarized in Algorithm 2.

Remark 1: It should be noted that action $\mathbf{a}_n$ is the response to state $\mathbf{s}_n$ rather than $\hat{\mathbf{s}}_n$, while the state stored into the butter is $\hat{\mathbf{s}}_n$, i.e., $(\hat{\mathbf{s}}_n, \mathbf{a}_n)$. This is because the DDPG agent reacts to its observed states, which indicates the agent is unable to recognize the difference between the observed state and the actual state. If we input the imperfect CSI $\hat{\mathbf{s}}_n$, the agent will generate action $\hat{\mathbf{a}}_n$. In this case, the transition $(\hat{\mathbf{s}}_n, \hat{\mathbf{a}}_n)$ owns the same effect with $(\mathbf{s}_n, \mathbf{a}_n)$, and cannot reflects the influences of channel estimation errors. Besides, it is also found that the complexity can be further reduced by replacing $\hat{r}_n$ with $r_n\left(\hat{\mathbf{s}}_n\right)$. Nevertheless, the error is inevitable due to the complicated non-linear relations between $\hat{r}_n$ and $\hat{\mathbf{s}}_n$.

C. Discussion

Based on the DDPG framework and the carefully designed mapping function, Algorithm 1 can achieve the optimization of channel allocation, transmit power and coefficient matrices under the perfect CSI. To tackle the uncertainties caused by

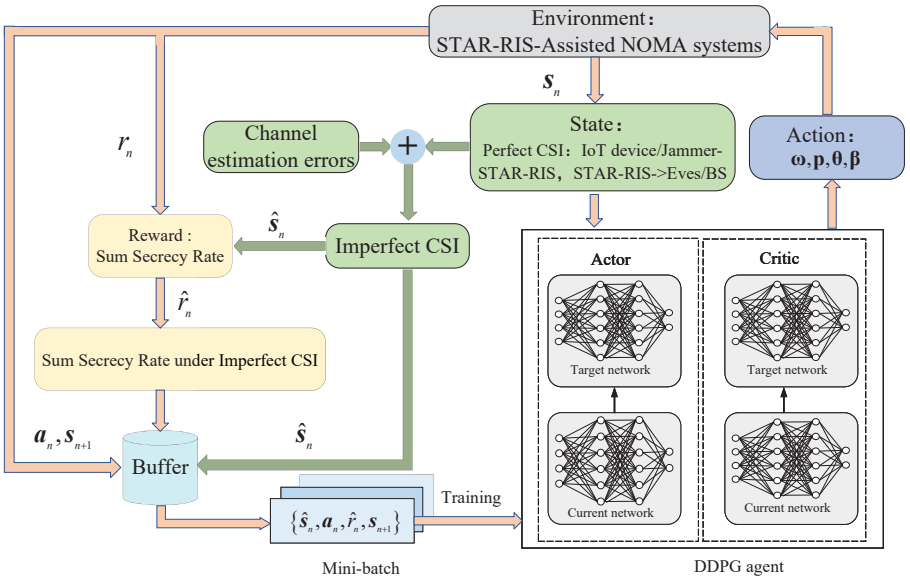


Fig. 3. The framework of improved DDPG algorithm for solving problem (11) with imperfect CSI.

Algorithm 2 The improved DDPG algorithm for solving Problem (11) with imperfect CSI

1. Initialize the environment, the network parameters $\hat{\boldsymbol{\lambda}}_{\mu}$, $\hat{\boldsymbol{\lambda}}_{q}$, $\hat{\boldsymbol{\lambda}}_{\mu'}$, and $\hat{\boldsymbol{\lambda}}_{q'}$, the buffer D;
2. **for** each episode **do**:
3. Reset the environment as $\mathbf{s}_0$.
4. **for** each step **do**:
5. Observe state $\mathbf{s}_n$ and execute action $\mathbf{a}_n$;
6. **repeat:**
7. Calculate reward r_n based on (16) and observe state $\mathbf{s}_{n+1}$;
8. Update the iterative index $z = z + 1$;
9. **Until:** $z > Z_{\max}$
10. Calculate $\hat{\mathbf{s}}_n$ and $\hat{r}_n$ based on (32);
11. Store the transition $\{\hat{\mathbf{s}}_n, \mathbf{a}_n, \hat{r}_n, \mathbf{s}_{n+1}\}$ into the buffer;
12. Randomly sample a mini-batch of transitions from the buffer;
13. Update the current network of critic by minimizing (33);
14. Update the current network of actor by maximizing (34);
15. Update the target networks of actor and critic;
16. **end for**
17. **end for**
18. **Output:** The well-trained actor network $\mu\left(\hat{\mathbf{s}}_n|\hat{\boldsymbol{\lambda}}_{\mu}\right)$.

channel estimation errors, benefitting from the SAA method, an improved DDPG algorithm, i.e., Algorithm 2, is designed to perform the joint optimization of passive beamforming and resource allocation under the imperfect CSI. Although both Algorithm 1 and Algorithm 2 are based on the DDPG framework, their training process and sample process are quite different.

Since the sample process in Algorithm 2 does not cause extra complexity, the complexity of both Algorithm 1 and Algorithm 2 depends on the DDPG algorithm. Denote the numbers of fully connected layers in the actor and critic networks as ς and υ, respectively. The complexity of Algorithm 1 and Algorithm 2 can be expressed as [42]

$$\mathcal{O}\left(\sum_{i=0}^{\varsigma-1} w_i^{\text{actor}} w_{i+1}^{\text{actor}} + \sum_{i=0}^{\upsilon-1} w_i^{\text{critic}} w_{i+1}^{\text{critic}}\right), \tag{35}$$

where w_i^{actor} and w_i^{critic} are the numbers of neurons in the i-th layer of the actor and critic networks.

IV. Simulation Results

A. Simulation Setting

In this section, we demonstrate the effectiveness of our proposed algorithms for STAR-RIS-assisted NOMA systems through simulation results. As shown in Fig. 4, a three-dimensional coordinate network setup is considered, where the BS and the Jammer are deployed at $(0, 0, 10)$ m and $(10, 45, 0)$ m, respectively. The STAR-RIS is located at $(0, 30, 20)$ m. Based on the location of STAR-RIS, the half space where the BS is located is defined as reflection space, and the other half space is transmission space. More specifically, if IoT device i's y coordinate is less than the y coordinate of STAR-RIS, then IoT device i is considered to be in the reflection space. Conversely, if it is greater, IoT device i is in the transmission space. $I/2$ IoT devices are randomly distributed on the reflection/transmission space. The R-Eve and T-Eve are located at $(20, 25, 0)$ m and $(20, 35, 0)$ m, respectively.

To evaluate the performance of the proposed algorithms, we consider the following benchmark schemes:

- Actor critic (AC) algorithm [44], Alternative optimization (AO) algorithm [45]: To illustrate the performance of

TABLE I
SIMULATION PARAMETERS [42], [46]

Parameters	Default Values
batch size, e	32
discount factor, γ	0.99
maximum transmit power, $P_{\max}$	0.01 W
buffer size, D	10000
noise power, σ^2	-90 dBm
pathloss factor, α	2.6

the proposed DDPG-based algorithm, the basic reinforcement learning algorithm, i.e., AC algorithm, and the optimization-based method, i.e., AO algorithm are implemented as benchmarks.

- Conventional reflecting/transmitting-only RIS (C-RIS) [12], Independent phase shift STAR-RIS (IPS-STAR-RIS) [28], Coupled phase shift STAR-RIS (CPS-SATR-RIS): In order to present the performance degradation caused by different constraints of STAR-RIS, we compare the sum secrecy rate of C-RIS, IPS-STAR-RIS, and CPS-SATR-RIS. Each element of the IPS-STAR-RIS can independently adjust the phase shift for transmission and reflection. While the CPS-SATR-RIS is required to satisfy constraint (9e). Unless specified otherwise, it is assumed that the CPS-SATR-RIS is deployed to assist the communication between the IoT devices and the BS in the following benchmark schemes.
- Without Eve-pCSI/ipCSI (WE-pCSI/ipCSI): Whether under the perfect or imperfect CSI, the scheme without Eve aims to maximize the achievable rate of IoT devices, which can be regarded as an upper bound for the other schemes.
- Only R-Eve/T-Eve, Non-colluding Eves, Colluding Eves: To further investigate the impacts of Eves, we consider the scheme with only one Eve located either at the reflection space or the transmission space, which is referred to as Only R-Eve/T-Eve. For the scheme with two Eves, if they intercept the information independently, i.e., Non-colluding Eves, the equivalent channel gain from IoT device i to the Eves will equal to the channel gain from IoT device i to the Eve with the best channel conditions. Otherwise, if these two Eves can exchange the information with each other, i.e., Colluding Eves, the equivalent channel gain is expressed as (31).
- OMA-pCSI/ipCSI [2], NOMA-Optimized Channel Allocation (OCA)-pCSI/ipCSI, NOMA-Fixed Channel Allocation (FCA)-pCSI/ipCSI: We also shed light on the performance comparison between NOMA and OMA in the STAR-RIS-assisted uplink systems. Besides, to verify the necessity of channel allocation optimization in the NOMA systems, the NOMA-FCA-pCSI/ipCSI scheme is designed as a benchmark scheme for the NOMA-OCA-pCSI/ipCSI.

[2]The frequency division multiple access (FDMA) scheme is adopted in OMA-pCSI/ipCSI where each IoT device occupies the bandwidth of B/I.

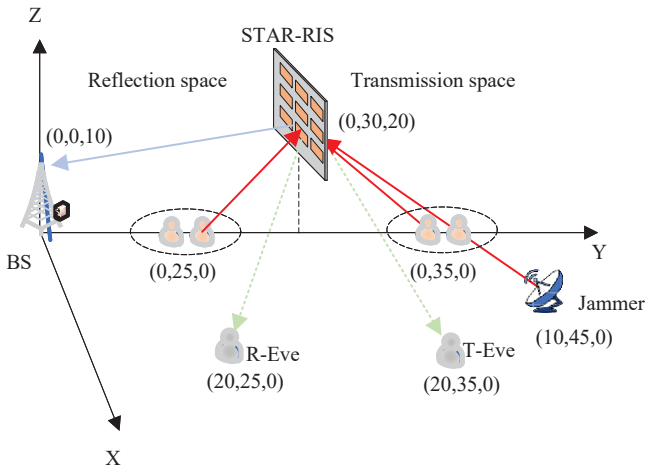


Fig. 4. Simulation setup.

B. Convergence Performance of the Proposed Algorithm

Fig. 5 illustrates the convergence performance of Algorithm 1 under different learning rate (Lr). When $\mathrm{Lr} = 0.1$, the reward cannot converge. This is because with a larger learning rate, the change of actor network's policy will be rapid, and hence may lead to the overshooting of optimal policy, making it difficult to converge to a stable solution. The reward of $\mathrm{Lr} = 0.01$ converges to a local optimal solution, and the reward of $\mathrm{Lr} = 0.0001$ requires more episodes to converge. Thus, based on the above results, we set the learning rate as 0.001 in the following simulations. Under the learning rate of 0.001, Fig. 6 illustrates the impacts of discount factor on the rewards. It can be seen that the proposed algorithm performs less effectively with $\gamma = 0.01$ or 0.1 compared to when $\gamma = 0.99$ or 1, since a tiny γ will cause the network to be incapable of anticipating the future practice in time, and resulting in a lower reward [44]. Although the performance of $\gamma = 0.99$ is close to that of $\gamma = 1$, the reward might grow infinitely with an improper number of time slot N when $\gamma = 1$. Therefore, in order to avoid potential issues like unstable training caused by the infinite reward, γ is set as 0.99 in the following simulations.

C. Performance Comparison

Fig. 7 shows the comparison of our proposed algorithm with AC and AO algorithms. It can be seen that as the number of STAR-RIS's elements increases, the sum secrecy rate of all algorithms increases. This is because a larger number of elements can provide more potentials to design STAR-RIS's passive beamforming strategies which are aimed at enhancing the cascaded channels between IoT devices and the BS while simultaneously suppressing those between the IoT devices and Eves. Besides, we can find that our proposed algorithm based on DDPG can achieve higher sum secrecy rate than the AC algorithm since the unique current and target networks in DDPG can enhance the robustness of training process. We also observe that our proposed algorithm performs close to the AO algorithm, while it is worth mentioning that the

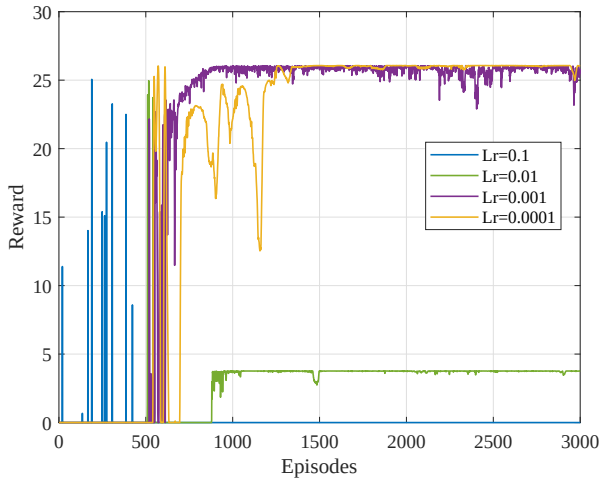


Fig. 5. Convergence performance of Algorithm 1.

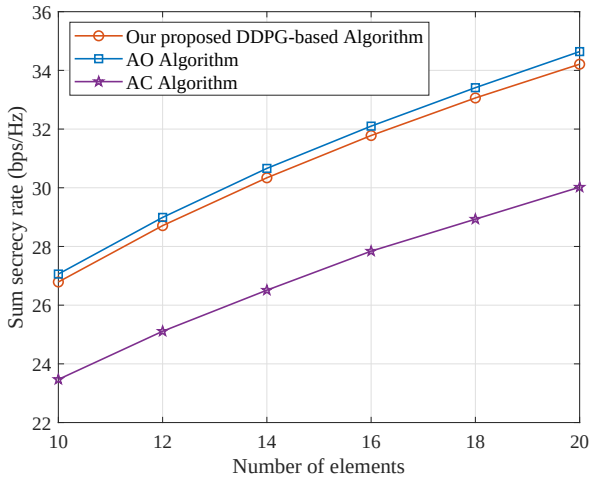


Fig. 7. Comparison of our proposed algorithm with AC and AO algorithms.

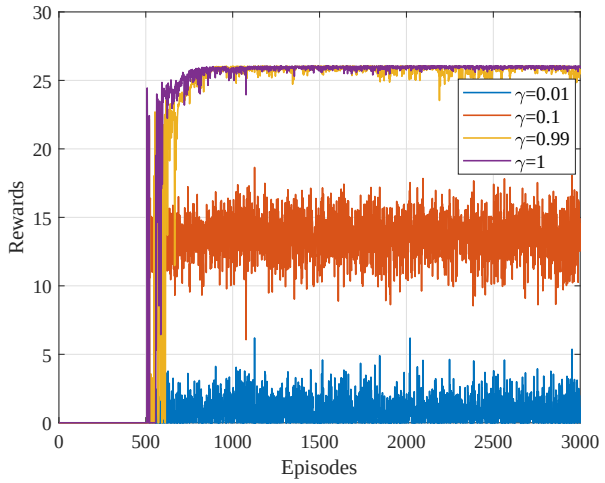


Fig. 6. The impacts of discount factor on the rewards.

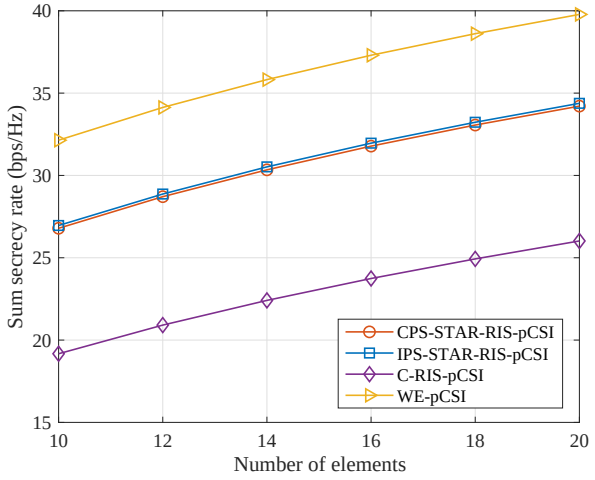


Fig. 8. The impacts of different constraints of STAR-RIS on the sum secrecy rate.

computational complexity of AO algorithm will be unacceptable when there are a large number of STAR-RIS elements, since the computational complexity of AO algorithm depends on the number of iterations and the number of decision variables. On the contrast, the computational complexity of our proposed DDPG-based algorithm is independent of the number of STAR-RIS elements, which is more suitable for the practical implementation.

Fig. 8 shows the impacts of different constraints of STAR-RIS on the sum secrecy rate. Due to the fact that there is no information leakage, the WE-pCSI scheme outperforms the other three schemes with Eves, which is in line with the theoretical expectations. The STAR-RISs, including the CPS-STAR-RIS and the IPS-STAR-RIS, can achieve higher sum secrecy rate than the C-RIS. This is because the amplitude coefficient of C-RIS for transmission and reflection can only be 0 or 1, which heavily restricts the DoFs for passive beamforming design. We also find that the IPS-STAR-RIS outperforms the CPS-STAR-RIS in terms of security enhancement, but the performance loss of CPS-STAR-RIS over IPS-STAR-RIS caused by the coupled phase shift is trivial when there are fewer STAR-RIS elements.

Fig. 9 presents the sum secrecy rate versus the pathloss factor between STAR-RIS and Eves ξ. With the increase of ξ, the channel gains between the STAR-RIS and Eves decrease, resulting in a lower eavesdropping rate, and hence boosting the secrecy rate. It can be found that the Only R-Eve and Only T-Eve schemes perform close to each other, and they outperform the schemes with two Eves. For the schemes with two Eves, the Non-colluding Eves is better than the scheme of Colluding Eves. This is because the ES protocol of STAR-RIS indicates that the incident signals will be divided into reflected and transmitted signals by the STAR-RIS. Consequently, when the IoT devices transmit signals to the BS, only one portion of the signals can be received by the BS via STAR-RIS's reflecting/transmitting, and the other portion of signals will be transmitted/reflected to the opposite side of the BS. If there is only one malicious R-Eve/T-Eve, it can only wiretap either the reflected signals or transmitted signals. While if

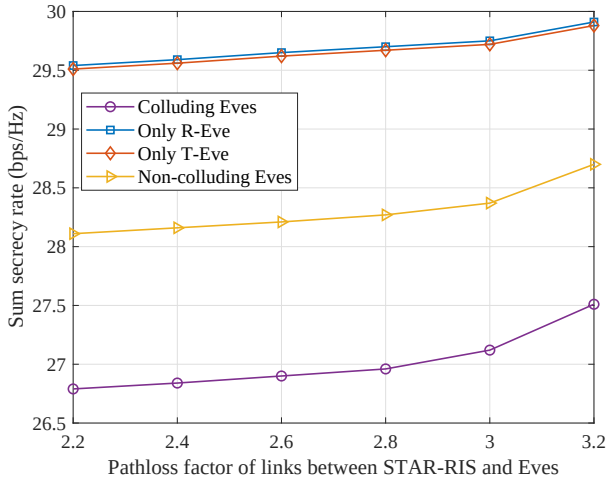


Fig. 9. Sum secrecy rate versus pathloss factor between STAR-RIS and Eves.

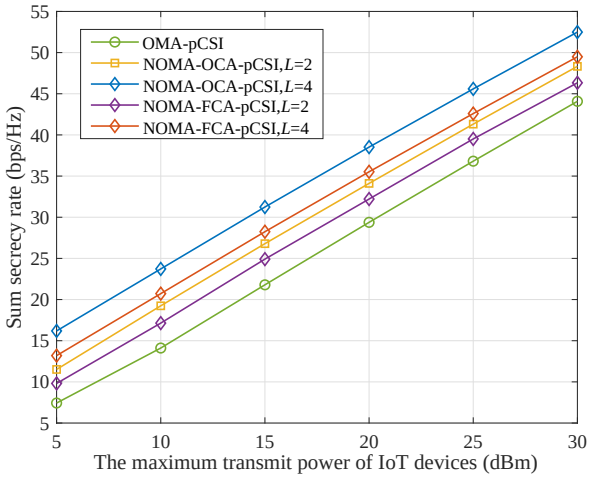


Fig. 10. Sum secrecy rate versus the maximum transmit power of IoT devices.

there are two Eves respectively located at the reflection space and transmission space, both of the reflected signals and transmitted signals will be gleaned. Even if these two Eves operate independently without cooperation, the eavesdropping rate is not less than the scenario with only R-Eve/T-Eve. On the contrary, for the Colluding Eves, the equivalent channel gain from IoT device i to Eves is larger than that to any of individual Eve, resulting in the most serious information leakage.

Fig. 10 presents the sum secrecy rate versus the maximum transmit power of IoT devices. As expected, the sum secrecy rate increases with the increase of uses' maximum transmit power. Compared to the scheme of OMA-pCSI where each IoT device occupies different frequency, the NOMA-OCA-pCSI and NOMA-FCA-pCSI can achieve higher sum secrecy rate since multiple IoT devices can simultaneously share the same frequency via power domain multiplexing in the NOMA systems. As can be seen from (5), the bandwidth dominates the achievable rate of IoT devices. Hence, although the inter-user interference increases when there are more IoT devices in the same channel, the sum secrecy rate for $L = 4$ outperforms that for $L = 2$. This is because the IoT devices in $L = 4$ own double the bandwidth compared to those in $L = 2$. In addition, through channel allocation optimization, the NOMA-OCA-pCSI demonstrates superior performance compared to NOMA-FCA-pCSI. This is because in the NOMA systems, the achievable sum rate of IoT devices can be improved by pairing the IoT devices with more distinctive channel conditions into the same channel.

Fig. 11 presents the sum secrecy rate versus the number of IoT devices, where $L = 2$. It can be seen that the sum secrecy rate rises with the increasing number of IoT devices. Interestingly, we also observe that when the number of IoT devices increases, the sum secrecy rate does not increase proportionally. This is because STAR-RIS aims to maximize the performance of overall system while simultaneously serving multiple IoT devices, rather than maximizing the rate for individual IoT devices. Therefore, compared to the secrecy

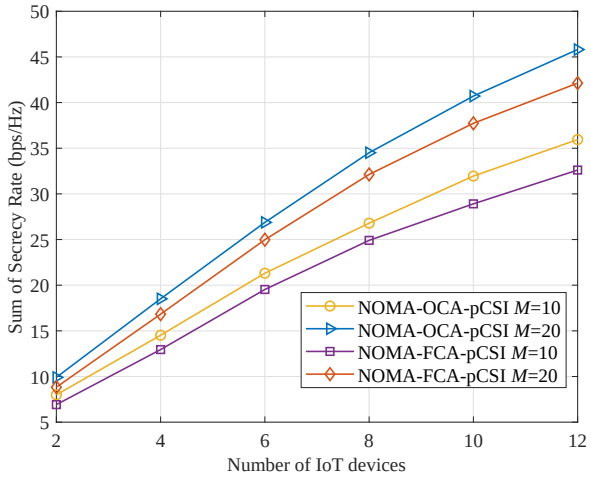


Fig. 11. Sum secrecy rate versus the number of IoT devices.

rate when there are only two IoT devices, the average secrecy rate for multiple IoT devices tends to decrease.

Fig. 12 depicts the sum secrecy rate versus the transmit power of Jammer. The Jammer's signal can confuse the Eves and thereby reduce the eavesdropping rate. Hence, besides the scheme of WE-pCSI, the sum secrecy rates of the other schemes improve as the Jammer's transmit power increases. It is observed that when the Jammer's transmit power exceeds 40 dBm, the performance of the Only R/T Eve, Colluding Eves, and Non-colluding Eves schemes is close to the WE-pCSI scheme. In this situation, the strength of the IoT devices' signals intercepted by Eves is much weaker than that of the intercepted jamming signals, i.e., $\omega_{i,k,n}p_{i,n}\mathbf{h}_{i,\mathrm{RIS},n}\mathbf{H}_{\mathrm{RIS-Eve},n}(\mathbf{h}_{i,\mathrm{RIS},n})^H \ll p_{\mathrm{J},n}\mathbf{h}_{\mathrm{J,RIS},n}\mathbf{H}_{\mathrm{RIS-Eve},n}(\mathbf{h}_{J,\mathrm{RIS},n})^H$, resulting in the eavesdropping rate close to zero. We also note that there is still an inescapable performance gap between the C-RIS-pCSI scheme and the other schemes when $p_{\mathrm{J},n} = 50$ dBm, which is caused by the restricted DoFs of C-RIS. These results verify the superiority of STAR-RIS over the traditional RIS.

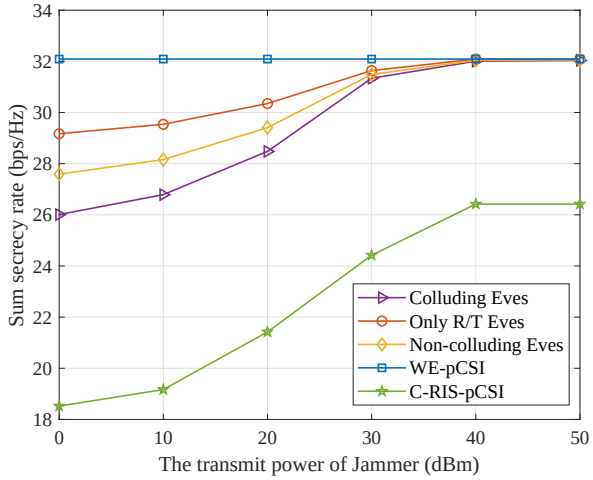


Fig. 12. Sum secrecy rate versus the transmit power of Jammer.

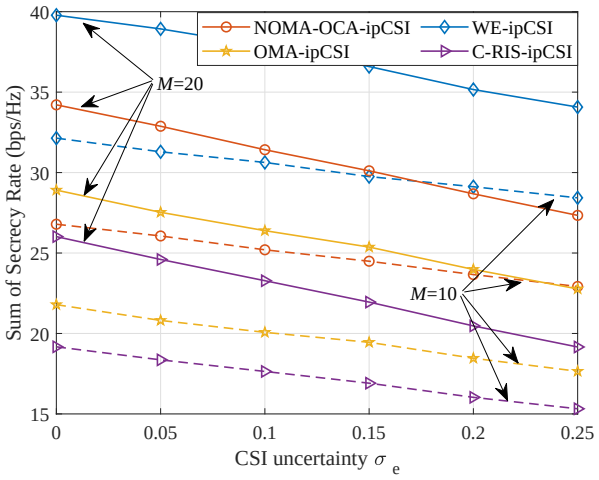


Fig. 14. Sum secrecy rate versus the CSI uncertainty.

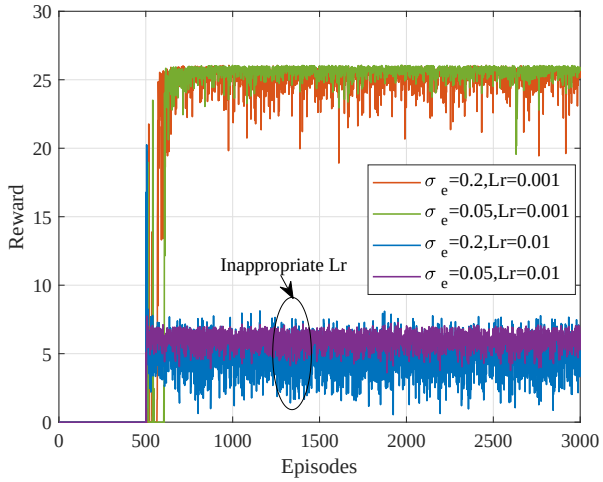


Fig. 13. Convergence performance of Algorithm 2.

D. Impacts of the Imperfect CSI

In this part, we study the impacts of imperfect CSI on the sum secrecy rate. The CSI uncertainty is measured by $\sigma_e = |\mathbf{\Theta}_{i,\mathrm{BS},n}|^2 \big/ |\mathbf{G}_{i,\mathrm{BS},n}|^2 = |\mathbf{\Theta}_{i/\mathrm{J,R/T-Eve},n}|^2 \big/ |\mathbf{G}_{i/\mathrm{J,R/T-Eve},n}|^2$. Fig. 13 presents the convergence performance of Algorithm 2 under different learning rates and CSI uncertainties. It can be seen that the reward converges after 1000 episodes when $\mathrm{Lr} = 0.001$. Compared to the case of perfect CSI, the fluctuation of Algorithm 2 is more severe due to the uncertainties of CSI.

Fig. 14 illustrates the sum secrecy rate versus the CSI uncertainty σ_e. It is found that the sum secrecy rate decreases when the CSI uncertainty increases. The channel estimation errors not only lead to the decrease of cascaded channel gains, but also result in a reduction in the signal enhancement effect of STAR-RIS. This is because the signals, via reflecting/transmitting of STAR-RIS, combine at the BS following the principle of vector addition. If the CSI cannot be estimated correctly, the destructive combination will result in a reduction of signal strength. We can observe that, with the CSI uncertainty of $\sigma_e = 0.25$, our proposed algorithm for STAR-RIS-assisted NOMA systems can still achieve security enhancement by 28.4 % and 44.5% compared to the OMA-ipCSI and C-RIS-ipCSI, respectively. In addition, with the growth of σ_e, the performance gain brought by increasing the number of elements gradually diminishes, which reveals the necessity of more accurate CSI for passive beamforming and resource allocation in the STAR-RIS-assisted NOMA systems.

V. Conclusions

In this paper, we investigated the security issues present in the STAR-RIS assisted NOMA uplink systems with one cooperative jammer and dual eavesdroppers, where the coupled phase shift model was employed at the STAR-RIS. Aiming to maximize the sum secrecy rate under the perfect CSI, a DDPG algorithm was proposed to optimize the channel allocation, transmit power and the coefficient matrices. Then based on the SAA method, the proposed algorithm was further adapted to tackle the sum secrecy rate maximization problem under imperfect CSI without causing additional computational complexity. Numerical results demonstrated that our proposed algorithms can achieve higher sum secrecy rate than the scheme with C-RIS and that with OMA under both the perfect and imperfect CSI. More importantly, we observed that the symmetry of STAR-RIS leads to severe information leakage in the uplink transmissions, and the sum secrecy rate degrades further when the dual eavesdroppers collaborate with each other. Additionally, we also noted that when STAR-RIS served multiple IoT devices simultaneously, it may prevent individual IoT devices from achieving maximum performance. Consequently, one focus of our future work would be addressing issues with different QoS constraints.

References

[1] T. Sauter and A. Treytl, "IoT-enabled sensors in automation systems and their security challenges," *IEEE Sensors Letters*, DOI: 10.1109/LSENS.2023.3332404 2023.

[2] C.-Y. Huang, Y.-H. Chiang, and F. Tsai, "An ontology integrating the open standards of city models and internet of things for smart-city applications," *IEEE Internet of Things Journal*, vol. 9, no. 20, pp. 20 444–20 457, Oct. 2022.

[3] S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwak, "The internet of things for health care: A comprehensive survey," *IEEE Access*, vol. 3, pp. 678–708, 2015.

[4] K. Haricha, A. Khiat, Y. Issaoui, A. Bahnasse, and H. Ouajji, "Recent technological progress to empower smart manufacturing: Review and potential guidelines," *IEEE Access*, vol. 11, pp. 77 929–77 951, 2023.

[5] T. M. Hoang, S. Dinh-Van, B. Barn, R. Trestian, and H. X. Nguyen, "RIS-aided smart manufacturing: Information transmission and machine health monitoring," *IEEE Internet of Things Journal*, vol. 9, no. 22, pp. 22 930–22 943, Nov. 2022.

[6] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface-aided wireless communications: A tutorial," *IEEE Transactions on Communications*, vol. 69, no. 5, pp. 3313–3351, May 2021.

[7] Y. Liu, X. Liu, X. Mu, T. Hou, J. Xu, M. Di Renzo, and N. Al-Dhahir, "Reconfigurable intelligent surfaces: Principles and opportunities," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1546–1577, 3rd Quarter 2021.

[8] M. Di Renzo, A. Zappone, M. Debbah, M.-S. Alouini, C. Yuen, J. de Rosny, and S. Tretyakov, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.

[9] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 106–112, Jan. 2020.

[10] C. Huang, A. Zappone, M. Debbah, and C. Yuen, "Achievable rate maximization by passive intelligent mirrors," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 3714–3718.

[11] J. Zuo, Y. Liu, Z. Qin, and N. Al-Dhahir, "Resource allocation in intelligent reflecting surface assisted NOMA systems," *IEEE Transactions on Communications*, vol. 68, no. 11, pp. 7170–7183, Nov. 2020.

[12] X. Mu, Y. Liu, L. Guo, J. Lin, and N. Al-Dhahir, "Exploiting intelligent reflecting surfaces in NOMA networks: Joint beamforming optimization," *IEEE Transactions on Wireless Communications*, vol. 19, no. 10, pp. 6884–6898, Oct. 2020.

[13] H. Zhang, S. Ma, Z. Shi, X. Zhao, and G. Yang, "Sum-rate maximization of RIS-aided multi-user MIMO systems with statistical CSI," *IEEE Transactions on Wireless Communications*, vol. 22, no. 7, pp. 4788–4801, Jul. 2023.

[14] Y. Liu, H.-H. Chen, and L. Wang, "Physical layer security for next generation wireless networks: Theories, technologies, and challenges," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 347–376, 1st Quarter 2017.

[15] M. Cui, G. Zhang, and R. Zhang, "Secure wireless communication via intelligent reflecting surface," *IEEE Wireless Communications Letters*, vol. 8, no. 5, pp. 1410–1414, Oct. 2019.

[16] J. Qiao and M.-S. Alouini, "Secure transmission for intelligent reflecting surface-assisted mmwave and terahertz systems," *IEEE Wireless Communications Letters*, vol. 9, no. 10, pp. 1743–1747, Oct. 2020.

[17] L. Dong and H.-M. Wang, "Secure MIMO transmission via intelligent reflecting surface," *IEEE Wireless Communications Letters*, vol. 9, no. 6, pp. 787–790, Jun. 2020.

[18] Z. Zhang, L. Lv, Q. Wu, H. Deng, and J. Chen, "Robust and secure communications in intelligent reflecting surface assisted NOMA networks," *IEEE Communications Letters*, vol. 25, no. 3, pp. 739–743, Mar. 2021.

[19] Y. Gao, Q. Wu, W. Chen, and D. W. K. Ng, "Rate-splitting multiple access for intelligent reflecting surface-aided secure transmission," *IEEE Communications Letters*, vol. 27, no. 2, pp. 482–486, Feb. 2023.

[20] Y. Wang, W. Shi *et al.*, "Intelligent reflecting surface aided secure transmission with colluding eavesdroppers," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 9, pp. 10 155–10 160, Sep. 2022.

[21] X. Guan, Q. Wu, and R. Zhang, "Intelligent reflecting surface assisted secrecy communication: Is artificial noise helpful or not?" *IEEE Wireless Communications Letters*, vol. 9, no. 6, pp. 778–782, Jun. 2020.

[22] Z. Chu, W. Hao, P. Xiao, D. Mi, Z. Liu, M. Khalily, J. R. Kelly, and A. P. Feresidis, "Secrecy rate optimization for intelligent reflecting surface assisted MIMO system," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1655–1669, 2021.

[23] X. Yu, D. Xu, Y. Sun, D. W. K. Ng, and R. Schober, "Robust and secure wireless communications via intelligent reflecting surfaces," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 11, pp. 2637–2652, Nov. 2020.

[24] J. Qiao, C. Zhang, A. Dong, J. Bian, and M.-S. Alouini, "Securing intelligent reflecting surface assisted terahertz systems," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 8, pp. 8519–8533, Aug. 2022.

[25] X. Lu, W. Yang, X. Guan, Q. Wu, and Y. Cai, "Robust and secure beamforming for intelligent reflecting surface aided mmwave MISO systems," *IEEE Wireless Communications Letters*, vol. 9, no. 12, pp. 2068–2072, Dec. 2020.

[26] Y. Ge and J. Fan, "Robust secure beamforming for intelligent reflecting surface assisted full-duplex MISO systems," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 253–264, 2022.

[27] J. Xu, Y. Liu, X. Mu, and O. A. Dobre, "STAR-RISs: Simultaneous transmitting and reflecting reconfigurable intelligent surfaces," *IEEE Communications Letters*, vol. 25, no. 9, pp. 3134–3138, Sep. 2021.

[28] X. Mu, Y. Liu, L. Guo, J. Lin, and R. Schober, "Simultaneously transmitting and reflecting (STAR) ris aided wireless communications," *IEEE Transactions on Wireless Communications*, vol. 21, no. 5, pp. 3083–3098, May 2022.

[29] Q. Zhang, Y. Zhao, H. Li, S. Hou, and Z. Song, "Joint optimization of STAR-RIS assisted UAV communication systems," *IEEE Wireless Communications Letters*, vol. 11, no. 11, pp. 2390–2394, Nov. 2022.

[30] Z. Zhang, J. Chen, Y. Liu, Q. Wu, B. He, and L. Yang, "On the secrecy design of STAR-RIS assisted uplink NOMA networks," *IEEE Transactions on Wireless Communications*, vol. 21, no. 12, pp. 11 207–11 221, Dec. 2022.

[31] H. Jia, L. Ma, and S. Valaee, "STAR-RIS enabled downlink secure NOMA network under imperfect CSI of eavesdroppers," *IEEE Communications Letters*, vol. 27, no. 3, pp. 802–806, Mar. 2023.

[32] H. Niu, Z. Chu, F. Zhou, and Z. Zhu, "Simultaneous transmission and reflection reconfigurable intelligent surface assisted secrecy MISO networks," *IEEE Communications Letters*, vol. 25, no. 11, pp. 3498–3502, Nov. 2021.

[33] Y. Zhang, Z. Yang, J. Cui, P. Xu, G. Chen, Y. Wu, and M. D. Renzo, "STAR-RIS assisted secure transmission for downlink multi-carrier NOMA networks," *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 5788–5803, 2023.

[34] Y. Han, N. Li, Y. Liu, T. Zhang, and X. Tao, "Artificial noise aided secure NOMA communications in STAR-RIS networks," *IEEE Wireless Communications Letters*, vol. 11, no. 6, pp. 1191–1195, Jun. 2022.

[35] X. Qin, Z. Song, T. Hou, W. Yu, J. Wang, and X. Sun, "Joint resource allocation and configuration design for STAR-RIS-enhanced wireless-powered MEC," *IEEE Transactions on Communications*, vol. 71, no. 4, pp. 2381–2395, Apr. 2023.

[36] J. Xu, X. Mu, J. T. Zhou, and Y. Liu, "Simultaneously transmitting and reflecting (STAR)-RISs: Are they applicable to dual-sided incidence?" *IEEE Wireless Communications Letters*, vol. 12, no. 1, pp. 129–133, Jan. 2023.

[37] Y. Liu, K. Huang *et al.*, "Secure wireless communications for STAR-RIS-assisted millimetre-wave NOMA uplink networks," *IET Communications*, vol. 17, pp. 1127–1139, Mar. 2023.

[38] Y. Liu, X. Mu *et al.*, "Simultaneously transmitting and reflecting (STAR)-RISs: A coupled phase-shift model," in *ICC 2022 - IEEE International Conference on Communications*, 2022, pp. 2840–2845.

[39] Z. Ding, R. Schober, and H. V. Poor, "Unveiling the importance of SIC in NOMA systems—part 1: State of the art and recent findings," *IEEE Communications Letters*, vol. 24, no. 11, pp. 2373–2377, Nov. 2020.

[40] S. Siboo, A. Bhattacharyya *et al.*, "An empirical study of DDPG and PPO-based reinforcement learning algorithms for autonomous driving," *IEEE Access*, vol. 11, pp. 125 094–125 108, 2023.

[41] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "A framework of robust transmission design for IRS-aided MISO communications with imperfect cascaded channels," *IEEE Transactions on Signal Processing*, vol. 68, pp. 5092–5106, 2020.

[42] R. Zhong, Y. Liu, X. Mu, Y. Chen, X. Wang, and L. Hanzo, "Hybrid reinforcement learning for STAR-RISs: A coupled phase-shift model based beamformer," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 9, pp. 2556–2569, Sep. 2022.

[43] D. Bertsimas, V. Gupta, and N. Kallus, "Robust sample average approximation," *Mathematical Programming*, vol. 171, pp. 217–282, 2018.

[44] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[45] Z. Zhang, J. Chen, Q. Wu, Y. Liu, L. Lv, and X. Su, "Securing NOMA networks by exploiting intelligent reflecting surface," *IEEE Transactions on Communications*, vol. 70, no. 2, pp. 1096–1111, Feb. 2022.

[46] X. Hu, C. Masouros *et al.*, "Reconfigurable intelligent surface aided mobile edge computing: From optimization-based to location-only learning-based solutions," *IEEE Transactions on Communications*, vol. 69, no. 6, pp. 3709–3725, Jun. 2021.

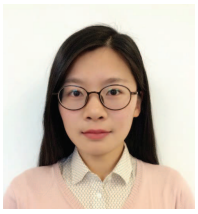
Xintong Qin received the B.Sc. degree from the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China, in 2020, where he is currently pursuing the Ph.D. degree. His research interests are in the field of nonorthogonal multiple access, mobile edge computing, unmanned aerial vehicle, and reconfigurable intelligent surfaces.

Zhengyu Song (Member, IEEE) received the B.Sc. and M.Sc. degrees in information and communication engineering from Beijing Jiaotong University, Beijing, China, in 2008 and 2011, respectively, and the Ph.D. degree in information and communication engineering from the Beijing Institute of Technology, Beijing, in 2016. He is currently with the School of Electronic and Information Engineering, Beijing Jiaotong University. His main research interests include nonorthogonal multiple access, mobile-edge computing, reconfigurable intelligent surfaces, unmanned aerial vehicle, and terrestrial-satellite-integrated communications.

Jun Wang received the B.Sc. degree in electronic information engineering and the Ph.D. degree in signal processing from Beijing Institute of Technology in 2004 and 2011, respectively. Then, he was a postdoctoral researcher in Beijing Jiaotong University. He is currently an associate professor with Beijing Jiaotong University, Beijing, China. His research interests include signal processing in wireless communications and satellite navigation systems.

Shengyu Du received the B.Sc. degree from the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China, in 2023, where she is currently pursuing the Ph.D. degree. Her research interests include the wireless communications, mobile Ad Hoc network, and resource allocation.

Jiazi Gao received the B.Sc. degree from the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China, in 2022, where she is currently pursuing the Ph.D. degree. Her research interests include mobile edge computing, low earth orbit satellite communications, and reconfigurable intelligent surfaces.



Wenjuan Yu (Member, IEEE) received her Ph.D. degree in Communication Systems from Lancaster University, Lancaster, U.K. She is currently a Lecturer with the School of Computing and Communications (SCC), InfoLab21, Lancaster University. She was a Research Fellow with the 5G Innovation Centre (5GIC), Institute for Communication Systems, University of Surrey, UK, from 2018 to 2020. Prior to that, she worked as a part-time Research Officer at the School of Computer Science and Electronic Engineering, University of Essex, UK, from Aug 2017 to Jan 2018. Her research interests include radio resource management, low latency communications, and machine learning for communications. She was an Executive Editor of the Transactions on Emerging Telecommunications Technologies from 2019 to 2022. She served as the lead Co-Chair for NGMA workshop in IEEE VTC2023-Spring and also a TPC Member for many conferences such as IEEE GLOBECOM, IEEE ICC, and IEEE VTC. She is the Conference Symposium and Workshop Officer for the Next Generation Multiple Access Emerging Technology Initiative (NGMA-ETI).

Xin Sun received the Ph.D. degree in electromagnetic measurement technology and instrument from Harbin Institute of Technology, Harbin, China, in 1998. She is currently a Professor with the School of Electronic and Information Engineering, Beijing Jiaotong University, Beijing, China. Her main research interests are professional mobile communications, wireless personal communications, green networking, and satellite communications.