Robust and Secure Multi-User STAR-RIS-Aided Communications: Optimization vs Machine Learning

Sonia Pala, *Graduate Student Member, IEEE*, Keshav Singh, *Member, IEEE*, Omid Taghizadeh, *Member, IEEE*, Cunhua Pan, *Senior Member, IEEE*, Octavia A. Dobre, *Fellow, IEEE*, and Trung Q. Duong, *Fellow, IEEE*

***Abstract*—This paper investigates simultaneous transmitting and reflecting reconfigurable intelligent surface (STAR-RIS)-assisted multi-user downlink (dl) communications with a primary focus on maximizing information secrecy by considering the channel state information (CSI) error. Acquiring perfect CSI is particularly challenging due to the unavailability of radio frequency chains at the STAR-RIS, the inherent impact of noise and interference on the CSI estimation, as well as non-collaborative nature of the eavesdroppers. In particular, we tackle the worst-case robust beamforming design problem to maximize the sum secrecy rate of the system while considering transmit power limitations, quality of service requirements, and practical constraints on the STAR-RIS phase shifter array. To tackle the resulting non-convex problem, we employ the S-procedure as an initial step to approximate semi-infinite inequality constraints. Subsequently, we leverage the alternating optimization with a line search framework to update the precoder and phase shift matrix iteratively. Furthermore, we extend our solution to address the non-convexity by leveraging a deep reinforcement learning (DRL) multi-agent (MA) framework based on Markov decision process. We also analyze practical phase shifts and the effect of direct links to showcase the practicality of our approach. Simulation results confirm STAR-RIS's significant performance edge, exhibiting approximately 27.1% higher secrecy in conventional optimization and around 35.4% in the MA-DRL context compared over the conventional RIS. Moreover, our proposed MA-DRL approach surpasses single-agent schemes by about 8.6% in the case of proximal policy optimization and 19.9% in the case of deep deterministic policy gradient, emphasizing the benefits of the MA framework with STAR-RIS.**

***Index Terms*—Reconfigurable intelligent surface (RIS), robust beamforming, deep reinforcement learning (DRL).**

I. Introduction

Recently, the cutting-edge technology known as reconfigurable intelligent surfaces (RISs) has garnered substantial attention in both the research community and academia. These two-dimensional meta-surfaces consist of multiple low-cost passive reflection elements, each capable of applying a programmable phase shift to incoming signals, altering the direction of signal propagation. RISs offer a cost-effective and low-complexity approach to dynamically reconfigure wireless propagation conditions, significantly improving communication system performance [1]–[4]. This is achieved by enhancing signal-to-noise ratios (SNRs) for legitimate users while degrading SNRs for potential eavesdroppers, thus enhancing wireless communication security. Additionally, RISs mitigate fading, reduce interference, and enhance networks performance, providing a secure and energy-efficient alternative to traditional relaying methods.

Physical layer security (PLS) has emerged as a crucial technology for mitigating security concerns, leveraging the inherent physical properties of wireless channels. Given the broadcast nature of wireless channels, wireless users face vulnerability to potential eavesdropping attacks by malicious eavesdroppers within the network, leading to significant security concerns related to the potential leakage of confidential information. Moreover, with the proliferation of a vast number of connected wireless devices, network capacity increases but at the expense of higher costs, elevated energy consumption, and increased security risks [5], [6]. Hence, there is a critical need for innovative and resource-efficient solutions to bolster wireless network security. Within wireless networks, the integration of RIS offers an innovative spatial dimension, effectively diminishing signal strength for potential eavesdroppers and amplifying it for legitimate users. This leads to a substantial enhancement in security performance. Further, recent advancements have introduced the concept of simultaneous transmitting and reflecting RIS (STAR-RIS) which leverages energy splitting (ES), mode switching (MS), or time switching (TS) protocols to simultaneously transmit and reflect signals [7], [8]. In contrast to traditional RIS systems, STAR-RIS offers complete spatial coverage and a greater degree of freedom (DoFs) for controlling signal propagation, which can enhance the wireless performance in various scenarios.

One of the pivotal applications of RISs lies in bolstering PLS in wireless networks. PLS addresses vulnerabilities inherent in wireless channels by leveraging the physical properties of signals. Due to the broadcast nature of wireless transmissions, users are susceptible to eavesdropping attacks, posing security risks such as unauthorized data interception. Meanwhile, RISs play a crucial role in mitigating these risks by selectively enhancing signal strength for intended recipients while introducing significant degradation for potential eavesdroppers [9]–[12]. The evolution from conventional RISs to STAR-RIS represents a significant advancement as STAR-RIS introduces capabilities such as simultaneous transmission and reflection using ES, MS, or TS protocols. This innovation not only extends spatial coverage but also enhances control over signal propagation, thereby offering new avenues to improve system performance and security in wireless communications [7], [8].

The work of K. Singh was supported by the National Science and Technology Council of Taiwan under Grants NSTC 112-2221-E-110-038-MY3 and NSTC 113-2218-E-110-008. The work of O. A. Dobre was supported by Canada Research Chairs Program CRC-2022-00187. The work of T. Q. Duong was supported by Canada Excellence Research Chairs Program CERC-2022-00109.

S. Pala and K. Singh are with the Institute of Communications Engineering, National Sun Yat-sen University, Kaohsiung 804, Taiwan (Email: sony.pj12@gmail.com, keshav.singh@mail.nsysu.edu.tw).

O. Taghizadeh is with the 5G Wireless Research Group, Lenovo Deutschland GmbH, Germany (Email: smotlagh@lenovo.com).

C. Pan is with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (Email: cpan@seu.edu.cn).

O. A. Dobre is with the Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John's, NL A1C 5S7, Canada (E-mail: odobre@mun.ca).

T. Q. Duong is with the Faculty of Engineering and Applied Science, Memorial University, St. John's, NL A1C 5S7, Canada, and is also with the School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Belfast, U.K. (E-mail: tduong@mun.ca).

A. Related Works

Over the past years, numerous studies have addressed PLS and RIS-assisted PLS in wireless communication systems (refer to [9]–[16] and the included references). For instance, the authors in [13] addressed secrecy rate maximization and power minimization in a single user/eavesdropper multiple-input multiple-output (MIMO) system using Taylor series approximation. The secrecy rate maximization in single-cell multiple-input single-output networks while considering the constraint of minimum harvested energy was studied in [14]. In [15], an inexact block coordinate descent method was employed to tackle secrecy rate maximization in single-user MIMO simultaneous wireless information and power transfer systems. The application of the primal decomposition method to optimize secrecy throughput in wireless-powered communication networks was carried out in [16]. In [9], the authors focused on securing RIS-aided multi-user massive MIMO systems, optimizing artificial noise power and RIS phase shifts. Meanwhile, [10] suggested virtual partitioning of RIS elements to enhance physical layer security and optimize secrecy capacity under rate constraints. The authors of [11] jointly optimized transmit precoding, artificial noise covariance, and RIS phase shifts, confirming enhanced secrecy rates through RIS incorporation in the system. Moreover, the hybrid beamforming design as well as the RIS phase shift design to enhance the system sum secrecy rate was carried out in [12]. However, it is important to note that the works in [9]–[16] have delved into RIS-assisted PLS systems but exclusively under perfect channel state information (CSI) conditions. In contrast, the works in [17]–[22] have considered RIS-assisted PLS systems, particularly under varying imperfect CSI conditions. For instance, the authors of [17] proposed joint strategies for secure links using low-resolution programmable reflecting elements in a RIS for multi-antenna access points serving single-antenna users amidst multiple eavesdroppers. Meanwhile, [18] introduced a secure multicast communication system using RIS to combat eavesdroppers and jammers during multi-user transmission. The analysis conducted in [19], explored the application of active RIS to optimize worst-case secrecy rates and weighted sum-secrecy rates under varying imperfect CSI conditions. While in [20], the authors introduced a RIS-aided MIMO secure communication system, optimizing ergodic secrecy rates using random matrix theory-based derivations and a joint optimization algorithm under statistical CSI. The investigation on RIS-aided secure communication systems with hardware impairments for maximizing the ergodic secrecy rate was carried out in [21], while [22] addressed CSI errors to minimize transmit power. However, the aforementioned literature [9]–[12], [17]–[22] primarily focused on incorporating passive RIS over STAR-RIS for evaluating secure communication systems. More recent works have transitioned to investigating STAR-RIS in broader communication scenarios. For instance, works such as [23]–[25] have focused on incorporating STAR-RIS into conventional communication systems, showcasing its performance benefits over passive RIS. Specifically, the authors of [23] explored a STAR-RIS-aided MIMO network to maximize the weighted sum rate through an energy splitting (ES) scheme, while [24] optimized training patterns for the time switching (TS) protocol and customized schemes for the ES protocol to achieve efficient uplink channel estimation in STAR-RIS-aided two-user systems. In [25], the authors aimed to maximize the coverage range without addressing information secrecy or robust optimization under CSI errors. Following this, several works have extended STAR-RIS applications to secure communication systems, integrating them with PLS techniques. The authors of [26] introduced a STAR-RIS-aided secure communication system designed to mitigate full-space mutual eavesdropping by employing a penalty-based secrecy beamforming algorithm to optimize coupled phase-shift coefficients. Similarly, [27] explored various transmission protocols, including ES, mode switching (MS), and TS, and proposed joint optimization of beamforming and transmission/reflection coefficients to maximize the weighted sum secrecy rate. Furthermore, [28] leveraged STAR-RIS to reconfigure the electromagnetic environment, enabling secure communication between legitimate users and the base station (BS), taking into account both full and statistical eavesdropper's CSI. Secrecy performance, considering residual hardware impairments, was examined in [29], though without robust beamforming design. Finally, [30] integrated STAR-RIS with non-orthogonal multiple access (NOMA) and air-federated learning to mitigate interference and provide omnidirectional coverage, focusing on learning performance under non-ideal wireless channels.

Recently, there has been a surge in the adoption of deep reinforcement learning (DRL) methods. DRL involves iterative learning and decision-making within dynamic environments, presenting a promising alternative with its learning and decision-making process. For instance, the authors of [31] introduced a novel DRL-based secure beamforming approach, utilizing post-decision state (PDS) and prioritized experience replay schemes to enhance learning efficiency and secrecy performance. A novel learning-based approach, PDS-deep Q-network combined with fourier feature mapping algorithm, addressing the non-convex optimization problem and dynamic environment to improve secrecy rate and quality of service (QoS) satisfaction was carried out in [32]. While, [33] explored the integration of RIS in mobile edge computing enabled industrial internet of things (IIoT) networks to enhance task offloading security against eavesdroppers. Meanwhile, [34] introduced deep deterministic policy gradient (DDPG) and soft actor-critic algorithms to maximize the legitimate user's long-term security rate in a STAR-RIS-based integrated sensing and communication secure system, addressing the non-convex problems while ensuring echo SNR.

B. Motivation

The advent of 6G technology beckons a transformative era in wireless communication systems, marked by unprecedented demands for enhanced coverage, capacity, and security. Meeting these demands requires innovative strategies, where the synergy of STAR-RIS, secure communication protocols, and robust transmission designs emerges as a promising paradigm. The merits can be explained as follows:

- *Maximizing Security Amid Threats:* STAR-RIS contributes to preventing both jamming and eavesdropping attempts by manipulating RIS element phases and amplitudes, which can create artificial noise or beamforming, disrupting unwanted reception by eavesdroppers or jammers, while reinforcing the intended signal for the legitimate receiver [35].
- *Enhanced Performance Despite Imperfections:* In challenging scenarios with imperfect CSI and hardware limitations,

STAR-RIS uses diverse protocols to handle uncertainties. These methods optimize resource allocation, power consumption, and balance transmission and reflection modes, effectively improving system resilience and performance in adverse conditions [36].

Overall, the interplay of STAR-RIS, secure communication, and robust transmission design serves as a prime motivation for our study, offering a comprehensive and innovative framework for advancing the state-of-the-art of wireless communication systems.

Despite the promising potential of robust and secure transmission design in multi-user STAR-RIS-aided communications, there is a lack of comprehensive research exploring its full capabilities. Prior works primarily focused on PLS and passive RIS-aided PLS, assuming perfect CSI, as seen in [9]–[16]. While some research works considered imperfect CSI, such as [17]–[22], they often prioritized passive RIS over STAR-RIS for evaluating secure communication systems. The exploration of STAR-RIS in communication and PLS systems was carried out in [23]–[30]. However, these works often did not fully address the robust optimization under CSI errors, the practical challenges of imperfect CSI, and the comprehensive handling of downlink communication scenarios, as summarized in Table I. Specifically, they either focused on specific protocols ([23], [27]), uplink scenarios ([28]) or did not consider the non-convex nature of robust beamforming design under practical CSI constraints ([25], [26], [29]). Further, the authors in [30] analyzed learning performance under non-ideal wireless channels but faced complexities in handling nonconvex subproblems. Additionally, integrating DRL algorithms in PLS with either RIS or STAR-RIS has been explored [31]–[34], albeit in single-agent systems. Although certain works explored weighted sum-secrecy rate optimization under imperfect CSI scenarios, particularly emphasizing eavesdropper channel imperfections, these may not fully reflect practical scenarios, as in [19]. Nevertheless, leveraging STAR-RIS for enhanced security, especially in the presence of imperfect CSI, and analyzing optimization problems using both conventional robust optimization techniques and multi-agent (MA) DRL algorithms, remains an unexplored domain in the current literature.

C. Contribution

Motivated by the identified research gap, this paper explores robust transmission design in multi-user STAR-RIS-aided communication systems under imperfect CSI. Unlike previous studies that overlooked realistic links involving the base station, users, and eavesdroppers, our analytical framework addresses the sum secrecy rate maximization problem by integrating robust optimization techniques with a tailored multi-agent deep reinforcement learning (MA-DRL) approach. We adapt MA-RL for STAR-RIS, focusing on agent configuration, observation sharing, and policy updates to manage unique T-zone and R-zone requirements. Our hybridization of MA-RL with robust optimization methods, including the S-procedure and alternating optimization (AO), effectively handles non-convex challenges, while our customized proximal policy optimization (PPO) framework ensures stability and efficiency. These contributions advance secure transmission design in STAR-RIS systems, enhancing both security and computational performance.

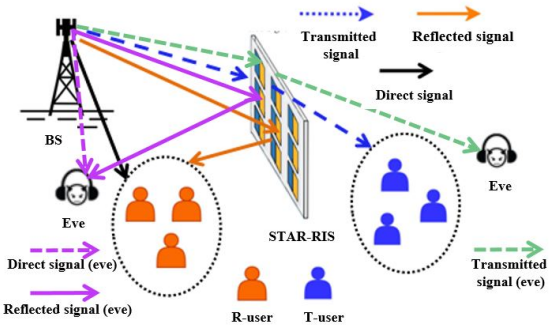

Fig. 1: STAR-RIS aided multi-user dl secure communication system.

- We tackle the worst-case robust beamforming design problems with the goal of maximizing the system sum secrecy rate while considering transmit power limitations, QoS requirements, and constraints on the STAR-RIS phase shifter array.
- The non-convex nature of this problem presents computational complexities. Initially, we employ the S-procedure to approximate semi-infinite inequality constraints. Following this, we utilize an alternating optimization (AO) framework along with line search methods for iterative updates to the precoder and phase shifter array. This approach also includes the analysis of practical phase shifts and the impact of direct links to enhance the practicality and robustness of our method.
- Furthermore, we extend our solution to address this non-convexity of the formulated optimization problem by adopting the MA-DRL solution tailored to STAR-RIS-aided communication systems with imperfect CSI. The contributions include the application of MA-DRL to handle distinct transmission and reflection zones (T-zone and R-zone) with secure communication requirements, integration with robust optimization techniques to address non-convexity, and adoption of the PPO framework for stability and efficiency.
- The computational complexity analysis is provided to specify the effectiveness of the proposed algorithms. Comparative analysis of the proposed multi-user STAR-RIS-aided secure communication scheme under imperfect-CSI against its counterparts (PPO and DDPG schemes) is provided, by varying transmit power budgets, the number of antennas at the BS, STAR-RIS elements, distance, number of users, and minimum rate requirements. We also introduced the perfect CSI assumption and conventional passive-RIS to compare with the proposed system.
- Simulation results highlight the superiority of STAR-RIS over conventional RIS configurations, particularly in scenarios involving random phase matrices at RIS and no-RIS setups. The STAR-RIS exhibits around 27.1% higher secrecy performance in conventional optimization and approximately 35.4% in the MA-DRL context compared to passive RIS. Additionally, the proposed MA-DRL approach outperforms single-agent schemes by approximately 8.6% (PPO) and 19.9% (DDPG), highlighting the benefits of employing the MA framework with STAR-RIS.

D. Structure of the Paper

Section II outlines the system model, Section III presents the formulated optimization problem and introduces the robust

TABLE I: Comparison of Performance Metrics with State-of-the-Art References [23]–[30].

Performance metric	[23]	[24]	[26]	[27]	[28]	[25]	[29]	[30]	**Our Work**
Sum Secrecy Rate Maximization	✗	✗	✓	✓	✓	✗	✓	✗	✓
Energy Splitting (ES) Scheme	✓	✓	✗	✓	✗	✗	✗	✗	✓
Robust Beamforming Design	✗	✗	✗	✗	✗	✗	✗	✗	✓
Channel State Information (CSI) Errors	✗	✗	✗	✗	✗	✗	✗	✓	✓
Information Secrecy	✗	✗	✓	✓	✓	✗	✓	✗	✓
Multi-Agent DRL	✗	✗	✗	✗	✗	✗	✗	✗	✓
Markov Decision Processes	✗	✗	✗	✗	✗	✗	✗	✗	✓
Alternating Optimization	✗	✗	✗	✗	✓	✗	✗	✗	✓
S-Procedure for Semi-Infinite Constraints	✗	✗	✗	✗	✗	✗	✗	✗	✓
Practical Constraints on Phase Shifters	✗	✗	✗	✗	✗	✓	✗	✗	✓

optimization-based solution, while Section IV introduces the proposed MA-DRL-based framework. Numerical simulations are discussed in Section V, followed by conclusion remarks in Section VI.

II. System Model

Let us consider a STAR-RIS aided multi-user downlink (dl) system, depicted in Fig. 1. In this system, we have a multi-antenna BS equipped with N antennas, and a STAR-RIS comprising M elements. Their purpose is to efficiently support K single-antenna dl users. This system operates in the presence of an unauthorized receiver[1], referred to as an eavesdropper with a single antenna. To enhance clarity, we refer to the users located in the reflection zone as R-users and the users located in the transmission zone as T-users. The STAR-RIS divides the geometric space of the network into reflection and transmission regions, accommodating R-users and T-users, respectively. Notably, T-users are located in an area without direct links to the BS, often called the "dead zone[2]." Let us denote the number of R-users as K_r and the number of T-users as K_t. These sets are respectively represented as $\mathcal{K}_r = \{1, ..., K_r\}$ and $\mathcal{K}_t = \{K_r+1, ..., K_r+K_t\}$. It follows that the total user set $\mathcal{K}$ comprises all K users, which can be defined as $\mathcal{K} = K_t \cup K_r = \{1, ..., K\}$.

The transmission and reflection properties of the m^{th} RIS element are given by $\phi_m^t = \sqrt{\alpha_m^t} e^{j\theta_m^t}$ and $\phi_m^r = \sqrt{\alpha_m^r} e^{j\theta_m^r}$, where $\alpha_m^t, \alpha_m^r \in [0,1]$ and $\theta_{m,l} \in [0, 2\pi)$ denote the amplitude and phase shift response of the m^{th} element's transmission and reflection coefficients. Note that, for each element, the phase shifts for transmission and reflection, denoted as ϕ_m^t and ϕ_m^r, can typically be chosen independently from one another. However, it is essential to ensure that α_m^t and α_m^r adhere to the energy conservation constraint $\alpha_m^t + \alpha_m^r = 1$, applicable for all elements within the set $m \in \mathcal{M}$ [38]. Further, we consider the ES protocol, where all elements of the STAR-RIS simultaneously work in the two modes. Thus, the transmission and reflection coefficient matrices for the STAR-RIS are given as $\mathbf{\Phi}^t \triangleq \operatorname{diag}(\phi_1^t, \ldots, \phi_M^t) \in \mathbb{C}^{M \times M}$ and $\mathbf{\Phi}^r \triangleq \operatorname{diag}(\phi_1^r, \ldots, \phi_M^r) \in \mathbb{C}^{M \times M}$.

[1]While differential privacy and friendly jamming are valuable privacy solutions, our relay-based approach is specifically tailored to the unique capabilities and constraints of the STAR-RIS system, offering an optimal balance between privacy protection and system performance.

[2]Users in the transmission zone of STAR-RIS are referred to as being in a dead zone because these are areas where traditional RIS or direct signals from the BS are insufficient to provide reliable connectivity. STAR-RIS addresses this challenge by simultaneously transmitting and reflecting signals, effectively enhancing coverage and reducing the occurrence of dead zones. Our algorithms can also be readily extended to the case where the direct links exist [37].

A. Practial Phase-shift coupled STAR-RIS:

In many existing studies on STAR-RIS systems [39]–[42], it is assumed that phase-shift coefficients for transmission and reflection can be adjusted independently, a condition that involves allowing electric and magnetic impedances to take arbitrary values. However, such an assumption may not hold for passive STAR-RIS setups, where the realizable electric and magnetic impedances are limited to purely imaginary numbers [43].

In general, the practical STAR-RIS model takes into account the phase-shift coupled STAR-RIS model. In this model, the phase-shift coefficients for transmission and reflection are inter-dependent, reflecting real-world constraints and offering a more accurate representation of STAR-RIS behavior.

Proposition 1. *For given loss-less passive STAR-RISs, the transmission and reflection coefficient phase-shift for each STAR-RIS follow that*

$$\sqrt{\alpha_m^t}\sqrt{\alpha_m^r}\cos\left(\theta_m^t - \theta_m^r\right) = 0, \tag{1}$$

$$|\theta_m^t - \theta_m^r| = \frac{(2a+1)\pi}{2}, a \in \mathbb{Z}, \forall m \in \mathcal{M}, \tag{2}$$

such that the independent phase-shift model for transmission and reflection coefficient cannot be realized [44].

Proof. See proof of the Proposition 1 in [44]. ■

Thus, the signal received at the k^{th} dl T-user and eavesdropper in T-zone are given as

$$y_k^t = ({\mathbf{g}_k^t}^H \mathbf{\Phi}^t \mathbf{F})\mathbf{W}\mathbf{s} + n_k^t, \forall k \in \mathcal{K}_t, \tag{3}$$

$$y_{e,k}^t = ({\mathbf{g}_e^t}^H \mathbf{\Phi}^t \mathbf{F})\mathbf{W}\mathbf{s} + n_{e,k}^t, \forall k \in \mathcal{K}_t. \tag{4}$$

Here, $n_k^t \sim \mathcal{CN}\left(0, {\sigma_k^t}^2\right)$ and $n_{e,k}^t \sim \mathcal{CN}\left(0, {\sigma_{e,k}^t}^2\right)$ represent the additive white-Gaussian noise (AWGN) in T-zone with zero mean and variances ${\sigma_k^t}^2$ and ${\sigma_{e,k}^t}^2$, respectively. Similarly, the signal received at the k^{th} dl R-user and eavesdropper in R-zone are given as

$$y_k^r = (\mathbf{h}_k^H + {\mathbf{g}_k^r}^H \mathbf{\Phi}^r \mathbf{F})\mathbf{W}\mathbf{s} + n_k^r, \forall k \in \mathcal{K}_r, \tag{5}$$

$$y_{e,k}^r = (\mathbf{h}_e^H + {\mathbf{g}_e^r}^H \mathbf{\Phi}^r \mathbf{F})\mathbf{W}\mathbf{s} + n_{e,k}^r, \forall k \in \mathcal{K}_r, \tag{6}$$

where $\mathbf{h}_k \in \mathbb{C}^{N \times 1}$ is the direct channel between the BS and the k-th R-user, $\mathbf{g}_k^l \in \mathbb{C}^{M \times 1}, l \in \{t, r\}, \forall k \in \mathcal{K}$ is the channel between the k^{th} user and the STAR-RIS, and $\mathbf{F} \in \mathbb{C}^{M \times N}$ is the channel between the BS and the STAR-RIS. Moreover, $\mathbf{h}_e \in \mathbb{C}^{N \times 1}$ is the direct channel between the BS and the eavesdropper, $\mathbf{g}_e^l \in \mathbb{C}^{M \times 1}, l \in \{t, r\}$, is the channel between the STAR-RIS and the eavesdropper. Further, the precoder matrix is expressed

as $\mathbf{W} = [\mathbf{w}_1, ..., \mathbf{w}_K] \in \mathbb{C}^{N\times K}$. Moreover, $n_k^r \sim \mathcal{CN}\left(0, {\sigma_k^r}^2\right)$ and $n_{e,k}^r \sim \mathcal{CN}\left(0, {\sigma_{e,k}^r}^2\right)$ represent the AWGN in R-zone with zero mean and variances ${\sigma_k^r}^2$ and ${\sigma_{e,k}^r}^2$, respectively.

Let $\mathbf{G}_k^l = \text{diag}({\mathbf{g}_k^l}^H)\mathbf{F}$ be the cascaded channel between the BS and the k^{th} user via STAR-RIS. Similarly, $\mathbf{G}_e^l = \text{diag}({\mathbf{g}_e^l}^H)\mathbf{F}$ be the cascaded channel between the BS and the eavesdropper via STAR-RIS. In practical scenarios, CSI estimation can have errors due to factors such as imperfect channel estimation procedures. This paper assumes that the BS lacks perfect knowledge of the CSI, so the expressions for CSI are formulated as follows:

$$\mathbf{G}_k^l = \tilde{\mathbf{G}}_k^l + \Delta\mathbf{G}_k^l, \mathbf{G}_e^l = \tilde{\mathbf{G}}_e^l + \Delta\mathbf{G}_e^l; l \in \{t, r\}, \forall k \in \mathcal{K}, \quad (7)$$

$$\mathbf{h}_k = \tilde{\mathbf{h}}_k + \Delta\mathbf{h}_k, \forall k \in \mathcal{K}_r; \mathbf{h}_e = \tilde{\mathbf{h}}_e + \Delta\mathbf{h}_e, \quad (8)$$

where $\tilde{\mathbf{G}}_k^l, \tilde{\mathbf{h}}_k, \tilde{\mathbf{G}}_e^l, \tilde{\mathbf{h}}_e$, represents the estimated channel vectors and $\Delta\mathbf{G}_k^l, \Delta\mathbf{h}_k, \Delta\mathbf{G}_e^l, \Delta\mathbf{h}_e$ are the corresponding channel error vectors. Moreover, the channel uncertainties are modeled using the worst-case approach based on norm-bounded errors, as described in [45], which are given by

$$\begin{aligned} ||\Delta\mathbf{G}_k^l||_{Fro}^2 \leq (\xi_{g,k}^l)^2, ||\Delta\mathbf{h}_k||^2 \leq (\xi_{h,k})^2, \\ ||\Delta\mathbf{G}_e^l||_{Fro}^2 \leq (\xi_{g,e}^l)^2, ||\Delta\mathbf{h}_e||^2 \leq (\xi_{h,e})^2, \end{aligned} \quad (9)$$

where $\xi_{g,k}^l, \xi_{h,k}, \xi_{g,e}^l, \xi_{h,e}$ represent the uncertainty bounds.

Denote by $\boldsymbol{\theta}^l \triangleq \left[\phi_1^l, \ldots, \phi_M^l\right]^T \in \mathbb{C}^{M\times 1}$ the vector representing the diagonal elements of the matrix $\boldsymbol{\Phi}^l$, and by $\mathbf{W}_{-k} = [\mathbf{w}_1, .., \mathbf{w}_{k-1}, \mathbf{w}_{k+1}, .., \mathbf{w}_K]$. Then, the signal-to-interference plus noise ratio (sinr) at the k^{th} user can be expressed as

$$\gamma_k^l \triangleq \begin{cases} \frac{|({\boldsymbol{\theta}^l}^H \mathbf{G}_k^l)\mathbf{w}_k|^2}{||({\boldsymbol{\theta}^l}^H \mathbf{G}_k^l)\mathbf{W}_{-k}||_2^2 + {\sigma_k^l}^2}, & l = t, \forall k \in \mathcal{K}_t, \\ \frac{|(\mathbf{h}_k^H + {\boldsymbol{\theta}^l}^H \mathbf{G}_k^l)\mathbf{w}_k|^2}{||(\mathbf{h}_k^H + {\boldsymbol{\theta}^l}^H \mathbf{G}_k^l)\mathbf{W}_{-k}||_2^2 + {\sigma_k^l}^2}, & l = r, \forall k \in \mathcal{K}_r. \end{cases} \quad (10)$$

In a more stringent security scenario, we make the assumption that eavesdroppers have limitless computational resources, enabling them to effectively filter out all interference signals, as well as external noise during the decoding of each user's information [46]. Consequently, under these conditions, the sinr for eavesdroppers can be expressed as

$$\gamma_{e,k}^l \triangleq \begin{cases} |({\boldsymbol{\theta}^l}^H \mathbf{G}_e^l)\mathbf{w}_k|^2 / {\sigma_{e,k}^l}^2, & l = t, \forall k \in \mathcal{K}_t, \\ |(\mathbf{h}_e^H + {\boldsymbol{\theta}^l}^H \mathbf{G}_e^l)\mathbf{w}_k|^2 / {\sigma_{e,k}^l}^2, & l = r, \forall k \in \mathcal{K}_r. \end{cases} \quad (11)$$

Using the aforementioned sinr expressions, the achievable secrecy rates at the k^{th} T-user and R-user are respectively expressed as

$$R_{s,k}^t = \{\underbrace{\log_2\left(1 + \gamma_k^t\right)}_{R_k^t} - \underbrace{\log_2\left(1 + \gamma_{e,k}^t\right)}_{R_{e,k}^t}\}^+, \forall k \in \mathcal{K}_t, \quad (12)$$

and

$$R_{s,k}^r = \{\underbrace{\log_2\left(1 + \gamma_k^r\right)}_{R_k^r} - \underbrace{\log_2\left(1 + \gamma_{e,k}^r\right)}_{R_{e,k}^r}\}^+, \forall k \in \mathcal{K}_r. \quad (13)$$

Here, the notation $\{\cdot\}^+ \triangleq \max(0, \cdot)$ ensures that the secrecy rate can never be negative.

B. Problem Formulation

Taking into account that imperfect channels reside within a certain bounded region, the channel uncertainty is modeled as $\mathbf{G}_k^t \in \mathcal{E}_k^t \triangleq \{\forall\Delta\mathbf{G}_k^t, ||\Delta\mathbf{G}_k^t||^2 \leq (\xi_{g,k}^t)^2\}$ and $\mathbf{G}_k^r, \mathbf{h}_k \in \mathcal{E}_k^r \triangleq \{\forall\Delta\mathbf{G}_k^r, ||\Delta\mathbf{G}_k^r||^2 \leq (\xi_{g,k}^r)^2\}, \{\forall\Delta\mathbf{h}_k, ||\Delta\mathbf{h}_k||^2 \leq (\xi_{h,k})^2\}$. Similarly, at the eavesdroppers, the channel uncertainty is modeled as $\mathbf{G}_e^t \in \mathcal{E}_e^t \triangleq \{\forall\Delta\mathbf{G}_e^t, ||\Delta\mathbf{G}_e^t||^2 \leq (\xi_{g,e}^t)^2$ and $\mathbf{G}_e^r, \mathbf{h}_e \in \mathcal{E}_e^r \triangleq \{\forall\Delta\mathbf{G}_e^r, ||\Delta\mathbf{G}_e^r||^2 \leq (\xi_{g,e}^r)^2\}, \{\forall\Delta\mathbf{h}_e, ||\Delta\mathbf{h}_e||^2 \leq (\xi_{h,e})^2\}$. The sum secrecy rate maximization problem is formulated as

$$(\mathbf{P0}): \max_{\mathbf{W}, \boldsymbol{\Phi}^t, \boldsymbol{\Phi}^r} \min_{\substack{\mathbf{G}_k^l \in \mathcal{E}_k^l, \mathbf{h}_k \in \mathcal{E}_k^r \\ \mathbf{G}_e^l \in \mathcal{E}_e^l, \mathbf{h}_e \in \mathcal{E}_e^r}} \sum_{k\in\mathcal{K}_t} R_{s,k}^t + \sum_{k\in\mathcal{K}_r} R_{s,k}^r \quad (14a)$$

$$\text{s.t. } \text{Tr}(\mathbf{W}^H\mathbf{W}) \leq P_T, \quad (14b)$$

$$\left|\phi_m^t\right|^2 + \left|\phi_m^r\right|^2 = 1, \forall m \in \mathcal{M}, \quad (14c)$$

$$R_k^t \geq R_{\min}; \mathbf{G}_k^t \in \mathcal{E}_k^t; \forall k \in \mathcal{K}_t, \quad (14d)$$

$$R_k^r \geq R_{\min}; \mathbf{G}_k^r, \mathbf{h}_k \in \mathcal{E}_k^r; \forall k \in \mathcal{K}_r, \quad (14e)$$

where P_T is the maximum BS transmit power in (14b), representing the power budget at the BS, and (14c) is the energy conservation constraint of STAR-RIS. The minimal rate necessity at every user is specified by (14d) and (14e). The problem's non-convexity arises from the complex nature of the objective function, which aims to maximize the sum secrecy rate involving beamforming and phase-shift matrices. This complexity is compounded by non-linear and coupled constraints, such as the power constraint, energy conservation at the STAR-RIS, and rate requirements for users. Additionally, robustness considerations for channel uncertainties further contribute to the non-convexity. These challenges necessitate the use of advanced optimization techniques and heuristic methods like DRL to find approximate solutions.

It is important to note that the objective function, as specified in (**P0**), seeks to maximize the minimum sum secrecy rate across all possible realizations of the channel uncertainties. This formulation aligns with robust optimization principles, ensuring that the system is optimized for the worst-case channel conditions within the defined uncertainty sets. It does not require precise CSI but instead leverages the bounds of uncertainty to ensure that the system performance is optimized across all potential states.

Remark 1: In instances where the eavesdropping attack is originated from user equipment or a node part of the network infrastructure (e.g., remote radio heads, etc.), the undesired receiver (i.e., eavesdropper) is indeed a known/authenticated entity part of (or subscribed to) the same communication network. In such instances, the communication behavior of the said authenticated user or infrastructure node as part of the communication network can be verified and hence trusted. Therefore, it is plausible to treat such an eavesdropping entity as a cooperative component regarding standard communication procedures like acquiring CSI, geographical positioning, and usual measurements [47], [48].

Lemma 1. *At the optimality of (**P0**), the $\{\cdot\}^+ \triangleq \max(0, \cdot)$ operator of the secrecy rate expressions for each user can be neglected without loss of optimality.*

Proof: The proof is obtained through contradiction, akin to [49], and is omitted due to space constraints. ■

III. Proposed Solution

The optimization problem in (14) is non-convex in nature and generally intractable. In this section, we address robust beamforming design with bounded CSI errors and aim to maximize the system sum secrecy rate by optimizing jointly the precoder matrix $\mathbf{W}$ and the phase shift matrix $\mathbf{\Phi}^l$. Moreover, the STAR-RIS introduces an additional layer of complexity by necessitating the joint optimization of both transmission and reflection phase shifts ($\mathbf{\Phi}^t$ and $\mathbf{\Phi}^r$). Unlike conventional RIS, which only controls the phase shifts for reflection, STAR-RIS must balance the dual functionality within the same physical device, making the optimization landscape more complex.

To tackle this non-convex problem involving semi-infinite inequality restrictions and coupling variables, we introduce an AO approach that leverages the S-procedure, second-order cone programming (socp), and penalty convex-concave procedure (p-ccp) techniques [50]. The S-procedure is chosen for efficiently transforming channel uncertainties, represented as quadratic forms, into tractable linear matrix inequalities (LMIs), without notably affecting solution feasibility or optimality. Firstly, we convert the infeasible problem into a solvable form by utilizing the epigraph form and introducing slack variables $\boldsymbol{\tau}^r = [\tau_1^r, ..., \tau_{K_r}^r]$ and $\boldsymbol{\tau}^t = [\tau_{K_r+1}^t, ..., \tau_{K_r+K_t}^t]$ as follows

$$\max_{\substack{\mathbf{W},\mathbf{\Phi}^t,\mathbf{\Phi}^r \\ \boldsymbol{\tau}^r,\boldsymbol{\tau}^t}} \sum_{k\in\mathcal{K}_t}(\tau_k^t-\tau_e^t)+\sum_{k\in\mathcal{K}_r}(\tau_k^r-\tau_e^r) \tag{15a}$$

$$\text{s.t. } (14b), (14c), \tag{15b}$$

$$R_k^t \geq \tau_k^t, \mathbf{G}_k^t \in \mathcal{E}_k^t, \forall k \in \mathcal{K}_t, \tag{15c}$$

$$R_k^r \geq \tau_k^r; \mathbf{G}_k^r, \mathbf{h}_k \in \mathcal{E}_k^r, \forall k \in \mathcal{K}_r, \tag{15d}$$

$$R_{e,k}^t \leq \tau_e^t, \mathbf{G}_e^t \in \mathcal{E}_e^t, \forall k \in \mathcal{K}_t, \tag{15e}$$

$$R_{e,k}^r \leq \tau_e^r; \mathbf{G}_e^r, \mathbf{h}_e \in \mathcal{E}_e^r, \forall k \in \mathcal{K}_r, \tag{15f}$$

$$\tau_k^t \geq R_{\min}, \mathbf{G}_k^t \in \mathcal{E}_k^t, \forall k \in \mathcal{K}_t, \tag{15g}$$

$$\tau_k^r \geq R_{\min}; \mathbf{G}_k^r, \mathbf{h}_k \in \mathcal{E}_k^r, \forall k \in \mathcal{K}_r, \tag{15h}$$

$$\tau_e^t < \tau_k^t, \mathbf{G}_e^t \in \mathcal{E}_e^t, \forall k \in \mathcal{K}_t, \tag{15i}$$

$$\tau_e^r < \tau_k^r; \mathbf{G}_e^r, \mathbf{h}_e \in \mathcal{E}_e^r, \forall k \in \mathcal{K}_r. \tag{15j}$$

To address the non-convexity of the rate constraint, (15c) is reformulated and split into worst-case desired and interference noise power, using auxiliary variables $\boldsymbol{\beta} = [\beta_1, ..., \beta_K]^T$, as

$$|(\boldsymbol{\theta}^{t\,H}\mathbf{G}_k^t)\mathbf{w}_k|^2 \geq \beta_k(2^{\tau_k^t}-1), \mathbf{G}_k^t \in \mathcal{E}_k^t, \forall k \in \mathcal{K}_t, \tag{16a}$$

$$||(\boldsymbol{\theta}^{t\,H}\mathbf{G}_k^t)\mathbf{W}_{-k}||_2^2 + {\sigma_k^t}^2 \leq \beta_k, \mathbf{G}_k^t \in \mathcal{E}_k^t, \forall k \in \mathcal{K}_t. \tag{16b}$$

To handle the non-convex semi-infinite inequalities (16a), we approximate non-convex elements and address semi-infinite inequalities using the S-Procedure. This involves a linear approximation of the useful signal power as follows (For a detailed explanation, please refer to the [51]). Substituting $\mathbf{G}_k^t = \tilde{\mathbf{G}}_k^t + \Delta\mathbf{G}_k^t$, then $|[\boldsymbol{\theta}^{t\,H}(\tilde{\mathbf{G}}_k^t + \Delta\mathbf{G}_k^t)]\mathbf{w}_k|^2$ is approximated linearly using its lower limit at $(\mathbf{w}_k^{(n)}, \boldsymbol{\theta}^{t(n)})$ as follows

$$\text{vec}^{\mathrm{T}}(\triangle\mathbf{G}_k^t)\mathbf{A}_k\text{vec}(\triangle\mathbf{G}_{\mathbf{k}}^{\mathbf{t}\,*})+2\text{Re}\left\{\mathbf{a}_k^{\mathrm{T}}\text{vec}(\triangle\mathbf{G}_{\mathbf{k}}^{\mathbf{t}\,*})\right\}+a_k, \tag{17}$$

where

$$\begin{aligned}\mathbf{A}_k = {} & \mathbf{w}_k\mathbf{w}_k^{(n),\mathrm{H}} \otimes \boldsymbol{\theta}^{t*}\boldsymbol{\theta}^{\mathbf{t}(n),\mathrm{T}} + \mathbf{w}_k^{(n)}\mathbf{w}_k^{\mathrm{H}} \otimes \boldsymbol{\theta}^{t(n),*}\boldsymbol{\theta}^{t\,\mathrm{T}} \\ & - (\mathbf{w}_k^{(n)}\mathbf{w}_k^{(n),\mathrm{H}} \otimes \boldsymbol{\theta}^{t(n),*}\boldsymbol{\theta}^{t(n),\mathrm{T}}),\end{aligned} \tag{18}$$

$$\begin{aligned}\mathbf{a}_k = {} & \text{vec}(\boldsymbol{\theta}^t\left(\boldsymbol{\theta}^{t(n),\mathrm{H}}\tilde{\mathbf{G}}_k^t\right)\mathbf{w}_k^{(n)}\mathbf{w}_k^{\mathrm{H}}) \\ & + \text{vec}(\boldsymbol{\theta}^{t(n)}\left(\boldsymbol{\theta}^{t\,\mathrm{H}}\tilde{\mathbf{G}}_k^t\right)\mathbf{w}_k\mathbf{w}_k^{(n),\mathrm{H}}) \\ & - \text{vec}(\boldsymbol{\theta}^{t(n)}\left(\boldsymbol{\theta}^{t(n),\mathrm{H}}\tilde{\mathbf{G}}_k^t\right)\mathbf{w}_k^{(n)}\mathbf{w}_k^{(n),\mathrm{H}}),\end{aligned} \tag{19}$$

$$\begin{aligned}a_k = {} & 2\text{Re}\left\{\left(\boldsymbol{\theta}^{t(n),\mathrm{H}}\widehat{\mathbf{G}}_k^t\right)\mathbf{w}_k^{(n)}\mathbf{w}_k^{\mathrm{H}}\left(\tilde{\mathbf{G}}_k^{t\,\mathrm{H}}\boldsymbol{\theta}^t\right)\right\} \\ & - \left(\boldsymbol{\theta}^{t(n),\mathrm{H}}\tilde{\mathbf{G}}_k^t\right)\mathbf{w}_k^{(n)}\mathbf{w}_k^{(n),\mathrm{H}}\left(\tilde{\mathbf{G}}_{\mathbf{k}}^{t\,\mathrm{H}}\boldsymbol{\theta}^{t(n)}\right).\end{aligned} \tag{20}$$

By substituting the signal power in (16a) and its linear approximation in (17), we reformulate (16a) as

$$\begin{aligned}&\text{vec}^{\mathrm{T}}(\triangle\mathbf{G}_k^t)\mathbf{A}_k\text{vec}(\triangle\mathbf{G}_{\mathbf{k}}^{\mathbf{t}\,*}) + 2\text{Re}\left\{\mathbf{a}_k^{\mathrm{T}}\text{vec}(\triangle\mathbf{G}_{\mathbf{k}}^{\mathbf{t}\,*})\right\} + a_k \\ &\qquad \geq \beta_k(2^{\tau_k^t}-1), \mathbf{G}_k^t \in \mathcal{E}_k^t, \forall k \in \mathcal{K}_t.\end{aligned} \tag{21}$$

Due to the coupling of variables on the right-hand side (R.H.S.) of (21), it is still intractable to solve. Thus, by applying Taylor's first-order approximation at ${\beta_k}^{(n)}$ and $\tau_k^{t(n)}$, (21) is transformed as

$$\begin{aligned}&\text{vec}^{\mathrm{T}}(\triangle\mathbf{G}_k^t)\mathbf{A}_k\text{vec}(\triangle\mathbf{G}_{\mathbf{k}}^{\mathbf{t}\,*}) + 2\text{Re}\left\{\mathbf{a}_k^{\mathrm{T}}\text{vec}(\Delta\mathbf{G}_k^{t\,*})\right\} + a_k \\ &\qquad \geq b_k, \mathbf{G}_k^t \in \mathcal{E}_k^t, \forall k \in \mathcal{K}_t,\end{aligned} \tag{22}$$

where $b_k = {\beta_k}^{(n)}2^{\tau_k^{t(n)}} + 2^{\tau_k^{t(n)}}({\beta_k} - {\beta_k}^{(n)}) + {\beta_k}^{(n)}[2^{\tau_k^{t(n)}}\log 2][\tau_k^t - \tau_k^{t(n)}]$. Then, (22) is transformed into the subsequent LMI as

$$\begin{bmatrix} \varpi_{g,k}\mathbf{I}_{MN} + \mathbf{A}_k & \mathbf{a}_k \\ \mathbf{a}_k^{\mathrm{T}} & C_k^t \end{bmatrix} \succeq \mathbf{0}, \forall k \in \mathcal{K}_t, \tag{23}$$

where $\boldsymbol{\varpi}_g = [\varpi_{g,1}, ..., \varpi_{g,K}] \geq 0$ are the slack variables and $C_k^t = a_k - b_k - \varpi_{g,k}(\xi_{g,k}^t)^2$. Further, through the application of Schur's complement method, (16b) is reformulated into an equivalent LMI as follows:

$$\begin{bmatrix} \beta_k^t - {\sigma_k^t}^2 - \mu_{g,k}M & \tilde{\mathbf{t}_k^t}^{\mathrm{H}} & \mathbf{0}_{1\times N} \\ \tilde{\mathbf{t}_k^t} & \mathbf{I}_{(K-1)} & \xi_{g,k}^t\mathbf{W}_{-k}^{\mathrm{H}} \\ \mathbf{0}_{N\times 1} & \xi_{g,k}^t\mathbf{W}_{-k} & \mu_{g,k}\mathbf{I}_N \end{bmatrix} \succeq \mathbf{0}, \forall k \in \mathcal{K}_t, \tag{24}$$

where $\tilde{\mathbf{t}_k^t} = [(\boldsymbol{\theta}^{t\,H}\tilde{\mathbf{G}_k^t})\mathbf{W}_{-k}]^{\mathrm{H}}$ and $\boldsymbol{\mu}_g = [\mu_{g,1}, ..., \mu_{g,K}]^{\mathrm{T}} \geq 0$. Similarly, for the case of eavesdropper in T-zone, substituting $\mathbf{G}_e^t = \tilde{\mathbf{G}}_e^t + \Delta\mathbf{G}_e^t$, then $[\boldsymbol{\theta}^{t\,H}(\tilde{\mathbf{G}}_e^t + \Delta\mathbf{G}_e^t)]\mathbf{w}_k|^2$ is linearly approximated similar to (22) by replacing $\mathbf{G}_k, \mathbf{a}_k$ and a_k with $\mathbf{G}_e, \bar{\mathbf{a}}_k$ and $\bar{a}_k$. By substituting channels and bounded error terms with eavesdropper channels, we can derive an equivalent LMI using a process similar to equations (18) to (23) as

$$\begin{bmatrix} \vartheta_{e,k}\mathbf{I}_{MN} + \mathbf{A}_k & \bar{\mathbf{a}}_k \\ \bar{\mathbf{a}}_k^{\mathrm{T}} & \bar{C}_k^t \end{bmatrix} \succeq \mathbf{0}, \forall k \in \mathcal{K}_t, \tag{25}$$

where $\boldsymbol{\vartheta}_e = [\vartheta_{e,1}, ..., \vartheta_{e,K}] \geq 0$ are the slack variables and $\bar{C}_k^t = \bar{a}_k - {\sigma_{e,k}^t}^2(2^{\tau_{e,k}^t} - 1) - \vartheta_{e,k}(\xi_{g,e}^t)^2$. Similarly, to tackle the non-convex rate constraint for R-zone, (15d) is reformulated as

$$|(\mathbf{h}_k^H + \boldsymbol{\theta}^{r\,H}\mathbf{G}_k^r)\mathbf{w}_k|^2 \geq \beta_k(2^{\tau_k^r} - 1); \mathbf{G}_k^r, \mathbf{h}_k \in \mathcal{E}_k^r, \forall k \in \mathcal{K}_r, \tag{26a}$$

$$||(\mathbf{h}_k^H + \boldsymbol{\theta}^{r\,H}\mathbf{G}_k^r)\mathbf{W}_{-k}||_2^2 + {\sigma_k^r}^2 \leq \beta_k; \mathbf{G}_k^r, \mathbf{h}_k \in \mathcal{E}_k^r, \forall k \in \mathcal{K}_r. \tag{26b}$$

Substituting $\mathbf{h}_k = \tilde{\mathbf{h}}_k + \Delta\mathbf{h}_k$ and $\mathbf{G}_k^r = \tilde{\mathbf{G}}_k^r + \Delta\mathbf{G}_k^r$, then $|[(\tilde{\mathbf{h}}_k + \Delta\mathbf{h}_k)^H + \boldsymbol{\theta}^{r\,H}(\tilde{\mathbf{G}}_k^r + \Delta\mathbf{G}_k^r)]\mathbf{w}_k|^2$ is linearly approximated at its

lower bound at $(\mathbf{w}_k^{(n)}, \boldsymbol{\theta}^{r(n)})$ as $\mathbf{x}_k^{\mathrm{H}}\widetilde{\mathbf{A}}_k\mathbf{x}_k + 2\mathrm{Re}\left\{\widetilde{\mathbf{a}}_k^{\mathrm{H}}\mathbf{x}_k\right\} + \widetilde{a}_k$, where

$$\widetilde{\mathbf{A}}_k = \mathbf{D}_k - \mathbf{Z}_k + \mathbf{D}_k^{\mathrm{H}},$$
$$\mathbf{D}_k = \begin{bmatrix} \mathbf{w_k^{(n)}} \\ \mathbf{w}_k^{(n)} \otimes \boldsymbol{\theta}^{r(\mathrm{n}),*} \end{bmatrix}\begin{bmatrix} \mathbf{w}_k^{\mathrm{H}} & \mathbf{w}_k^{\mathrm{H}} \otimes \boldsymbol{\theta}^{r\mathrm{T}} \end{bmatrix},$$
$$\mathbf{Z}_k = \begin{bmatrix} \mathbf{w_k^{(n)}} \\ \mathbf{w}_k^{(n)} \otimes \boldsymbol{\theta}^{r(\mathrm{n}),*} \end{bmatrix}\begin{bmatrix} \mathbf{w}_k^{(n)\mathrm{H}} & \mathbf{w}_k^{(n)\mathrm{H}} \otimes \boldsymbol{\theta}^{r(n)\mathrm{T}} \end{bmatrix},$$
$$\widetilde{\mathbf{a}}_k = \mathbf{d}_{2,k} + \mathbf{d}_{1,k} - \mathbf{z}_k,$$
$$\mathbf{d}_{1,k} = \begin{bmatrix} \mathbf{w_k}\mathbf{w_k^{(n)},H}\left(\tilde{\mathbf{h}}_\mathbf{k} + \tilde{\mathbf{G}}_\mathbf{k}^{\mathbf{r}\,\mathrm{H}}\boldsymbol{\theta}^{\mathbf{r(n)}}\right) \\ \mathrm{vec}^*(\boldsymbol{\theta}^r\left(\tilde{\mathbf{h}}_k^{\mathrm{H}} + \boldsymbol{\theta}^{r(n),\mathrm{H}}\tilde{\mathbf{G}}_k^r\right)\mathbf{w}_k^{(n)}\mathbf{w}_k^{\mathrm{H}}) \end{bmatrix},$$
$$\mathbf{d}_{2,k} = \begin{bmatrix} \mathbf{w_k^{(n)}}\mathbf{w}_\mathbf{k}^{\mathrm{H}}\left(\tilde{\mathbf{h}}_\mathbf{k} + \tilde{\mathbf{G}}_\mathbf{k}^{\mathbf{r}\,\mathrm{H}}\boldsymbol{\theta}^{\mathbf{r}}\right) \\ \mathrm{vec}^*(\boldsymbol{\theta}^{r(n)}\left(\tilde{\mathbf{h}}_k^{\mathrm{H}} + \boldsymbol{\theta}^{r\mathrm{H}}\tilde{\mathbf{G}}_k^r\right)\mathbf{w}_k\mathbf{w}_k^{(n),\mathrm{H}}) \end{bmatrix},$$
$$\mathbf{z}_k = \begin{bmatrix} \mathbf{w_k^{(n)}}\mathbf{w_k^{(n),H}}\left(\tilde{\mathbf{h}}_\mathbf{k} + \tilde{\mathbf{G}}_\mathbf{k}^{\mathbf{r}\,\mathrm{H}}\boldsymbol{\theta}^{\mathbf{r(n)}}\right) \\ \mathrm{vec}^*(\boldsymbol{\theta}^{r(n)}\left(\tilde{\mathbf{h}}_k^{\mathrm{H}} + \boldsymbol{\theta}^{r(n),\mathrm{H}}\tilde{\mathbf{G}}_k^r\right)\mathbf{w}_k^{(n)}\mathbf{w}_k^{(n),\mathrm{H}}) \end{bmatrix},$$
$$\widetilde{a}_k = 2\mathrm{Re}\left\{d_k\right\} - z_k;\ \mathbf{x}_k = \begin{bmatrix} \triangle\mathbf{h}_k^{\mathrm{H}} & \mathrm{vec}^{\mathrm{H}}(\triangle\mathbf{G_k^{r}}^*) \end{bmatrix}^{\mathrm{H}},$$
$$d_k = \left(\tilde{\mathbf{h}}_k^{\mathrm{H}} + \boldsymbol{\theta}^{r(n),\mathrm{H}}\tilde{\mathbf{G}}_k^r\right)\mathbf{w}_k^{(n)}\mathbf{w}_k^{\mathrm{H}}\left(\tilde{\mathbf{h}}_k + \tilde{\mathbf{G}}_k^{r\,\mathrm{H}}\boldsymbol{\theta}^r\right),$$
$$z_k = \left(\tilde{\mathbf{h}}_k^{\mathrm{H}} + \boldsymbol{\theta}^{r(n)\mathrm{H}}\tilde{\mathbf{G}}_k^r\right)\mathbf{w}_k^{(n)}\mathbf{w}_k^{(n)\mathrm{H}}\left(\tilde{\mathbf{h}}_k + \tilde{\mathbf{G}}_k^{r\mathrm{H}}\boldsymbol{\theta}^{r(n)}\right).$$

However, due to the coupling of variables on the R.H.S. of (24), it is still intractable to solve. Thus, by applying Taylor's first-order approximation at ${\beta_k}^{(n)}$ and $\tau_k^{r(n)}$, (24) is transformed as

$$\mathbf{x}_k^{\mathrm{H}}\widetilde{\mathbf{A}}_k\mathbf{x}_k + 2\mathrm{Re}\left\{\widetilde{\mathbf{a}}_k^{\mathrm{H}}\mathbf{x}_k\right\} + \widetilde{a}_k \geq f_k; \mathbf{G}_k^r, \mathbf{h}_k \in \mathcal{E}_k^r, \forall k \in \mathcal{K}_r, \tag{27}$$

where $f_k = {\beta_k}^{(n)}2^{\tau_k^{r(n)}} + 2^{\tau_k^{r(n)}}(\beta_k - {\beta_k}^{(n)}) + {\beta_k}^{(n)}[2^{\tau_k^{r(n)}}\log 2][\tau_k^r - \tau_k^{r(n)}]$. Moreover,

$$\mathcal{E}_k^r \triangleq \begin{cases} \mathbf{x}_k^{\mathrm{H}}\begin{bmatrix} \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\mathbf{x}_k - \xi_{\mathrm{h},k}^2 \leq 0, \\ \mathbf{x}_k^{\mathrm{H}}\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{MN} \end{bmatrix}\mathbf{x}_k - {\xi_{g,k}^r}^2 \leq 0. \end{cases} \tag{28}$$

Then, the corresponding LMI is expressed as

$$\begin{bmatrix} \widetilde{\mathbf{A}}_k + \begin{bmatrix} \varpi_{\mathrm{h},k}^r\mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \varpi_{\mathrm{g},k}\mathbf{I}_{MN} \end{bmatrix} & \widetilde{\mathbf{a}}_k \\ \widetilde{\mathbf{a}}_k^{\mathrm{H}} & C_k^r \end{bmatrix} \succeq \mathbf{0}, \forall k \in \mathcal{K}_r, \tag{29}$$

where $\boldsymbol{\varpi}_h^r = [\varpi_{h,1}^r, ..., \varpi_{h,K_r}^r] \geq 0$ are the slack variables and $C_k^r = \widetilde{a}_k - f_k - \varpi_{h,k}^r(\xi_{h,e})^2 - \varpi_{g,k}(\xi_{g,e}^r)^2$. Next, by substituting $\mathbf{h}_k = \tilde{\mathbf{h}}_k + \Delta\mathbf{h}_k$ and $\mathbf{G}_k^r = \tilde{\mathbf{G}}_k^r + \Delta\mathbf{G}_k^r$, (26b) is transformed into the equivalent matrix inequality as

$$\mathbf{0} \preceq \begin{bmatrix} \beta_k - {\sigma_k^r}^2 & \widetilde{\mathbf{t}}_k^{r\mathrm{H}} \\ \widetilde{\mathbf{t}}_k^r & \mathbf{I} \end{bmatrix} + \begin{bmatrix} 0 & \left(\triangle\mathbf{h}_k^{\mathrm{H}} + \boldsymbol{\theta}^{r\mathrm{H}}\triangle\mathbf{G}_k^r\right)\mathbf{W}_{-k} \\ \mathbf{W}_{-k}^{\mathrm{H}}\left(\triangle\mathbf{h}_k + \triangle\mathbf{G}_k^{r\mathrm{H}}\boldsymbol{\theta}^r\right) & \mathbf{0} \end{bmatrix}$$
$$\preceq \begin{bmatrix} \mathbf{0} \\ \mathbf{W}_{-k}^{\mathrm{H}} \end{bmatrix}\begin{bmatrix} \triangle\mathbf{h}_k & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \triangle\mathbf{h}_k^{\mathrm{H}} \\ \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{0} & \mathbf{W}_{-k} \end{bmatrix}$$
$$+ \begin{bmatrix} \mathbf{0} \\ \mathbf{W}_{-k}^{\mathrm{H}} \end{bmatrix}\triangle\mathbf{G}_k^{r\mathrm{H}}\begin{bmatrix} \boldsymbol{\theta}^r & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\theta}^{r\mathrm{H}} \\ \mathbf{0} \end{bmatrix}\triangle\mathbf{G}_k^r\begin{bmatrix} \mathbf{0} & \mathbf{W}_{-k} \end{bmatrix}$$
$$+ \begin{bmatrix} \beta_k - {\sigma_k^r}^2 & \widetilde{\mathbf{t}}_k^{r\mathrm{H}} \\ \widetilde{\mathbf{t}}_k^r & \mathbf{I} \end{bmatrix}. \tag{30}$$

With $m_k = \beta_k - {\sigma_k^r}^2 - \mu_{g,k}M - \mu_{h,k}^r$, $\tilde{\mathbf{t}}_k^r = [(\tilde{\mathbf{h}_k}^{\mathrm{H}} + {\boldsymbol{\theta}^t}^H\tilde{\mathbf{G}}_k^t)\mathbf{W}_{-k}]^{\mathrm{H}}$ and further adopting the Schur's complement method, the power inequalities in (26b) is reformulated into an equivalent LMI as follows:

$$\begin{bmatrix} m_k & \widetilde{\mathbf{t}}_k^{r\,\mathrm{H}} & \mathbf{0}_{1\times N} & \mathbf{0}_{1\times N} \\ \widetilde{\mathbf{t}}_k^r & \mathbf{I}_{(K-1)} & \xi_{g,k}^r\mathbf{W}_{-k}^{\mathrm{H}} & \xi_{\mathrm{h},k}\mathbf{W}_{-k}^{\mathrm{H}} \\ \mathbf{0}_{N\times 1} & \xi_{g,k}\mathbf{W}_{-k} & \mu_{g,k}\mathbf{I}_N & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times 1} & \xi_{\mathrm{h},k}\mathbf{W}_{-k} & \mathbf{0}_{N\times N} & \mu_{h,k}^r\mathbf{I}_N \end{bmatrix} \succeq \mathbf{0}, \forall k \in \mathcal{K}_r, \tag{31}$$

where $\boldsymbol{\mu}_h^r = [\mu_{h,1}^r, ..., \mu_{h,K_r}^r]^{\mathrm{T}} \geq 0$ and $\tilde{\mathbf{t}}_k^t = [({\boldsymbol{\theta}^t}^H\tilde{\mathbf{G}}_k^t)\mathbf{W}_{-k}]^{\mathrm{H}}$. Similarly, for the case of the eavesdropper in R-zone, substituting $\mathbf{G}_e^r = \tilde{\mathbf{G}}_e^r + \Delta\mathbf{G}_e^r$ and $\mathbf{h}_e = \tilde{\mathbf{h}}_e + \Delta\mathbf{h}_e$, then $|(\tilde{\mathbf{h}}_e + \Delta\mathbf{h}_e)^H + {\boldsymbol{\theta}^r}^H(\tilde{\mathbf{G}}_e^r + \Delta\mathbf{G}_e^r)^r)\mathbf{w}_k|^2$ is approximated linearly by its upper boundary at $(\mathbf{w}_k^{(n)}, \boldsymbol{\theta}^{r(n)})$ as $\mathbf{u}_e^{\mathrm{H}}\widetilde{\mathbf{A}}_k\mathbf{u}_e + 2\mathrm{Re}\left\{\widehat{\mathbf{a}}_k^{\mathrm{H}}\mathbf{u}_e\right\} + \widehat{a}_k$, where $\mathbf{u}_e = \begin{bmatrix} \triangle\mathbf{h}_e^{\mathrm{H}} & \mathrm{vec}^{\mathrm{H}}(\triangle\mathbf{G_e^{r}}^*) \end{bmatrix}^{\mathrm{H}}$. It is important to note that the rate constraint (and hence the corresponding sinr) has an upper bound, setting the upper bound for these constraints within this scenario. Further, $\widetilde{\mathbf{A}}_k, \widehat{\mathbf{a}}_k$, and $\widehat{a}_k$ and the equivalent LMI can be derived in a similar manner as described in equations (27) to (29). However, in this case, we replace the channels and norm-bounded error terms with the corresponding R-zone eavesdropper channels and errors. Therefore, the respective LMI can be expressed as follows

$$\begin{bmatrix} \widetilde{\mathbf{A}}_k + \begin{bmatrix} \varrho_{\mathrm{e},k}\mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \vartheta_{\mathrm{e},k}\mathbf{I}_{MN} \end{bmatrix} & \widehat{\mathbf{a}}_k \\ \widehat{\mathbf{a}}_k^{\mathrm{H}} & \widehat{C}_k^r \end{bmatrix} \succeq \mathbf{0}, \forall k \in \mathcal{K}_r, \tag{32}$$

where $\boldsymbol{\varrho}_e = [\varrho_{e,1}, ..., \varrho_{e,K_r}] \geq 0$ are the slack variables and $\widehat{C}_k^r = \widehat{a}_k - {\sigma_{e,k}^r}^2(2^{\tau_{e,k}^r} - 1) - \varrho_{e,k}(\xi_{h,e})^2 - \vartheta_{e,k}(\xi_{g,e}^r)^2$.

However, the problem remains non-convex and poses a challenge to jointly optimize $\mathbf{W}$ and $\boldsymbol{\Phi}^l$ due to the coupling of variables. To address this, we employ the AO method, optimizing $\mathbf{W}$ and $\boldsymbol{\Phi}^l$ sequentially in an iterative fashion. We begin by maximizing the worst-case sum secrecy rate while keeping $\boldsymbol{\Phi}^l$ fixed. This step transforms the problem into a convex one concerning $\mathbf{W}$, which is efficiently solved using the CVX tool. Specifically, given a fixed $\boldsymbol{\Phi}^l$, the subproblem for $\mathbf{W}$ is formulated as

$$\max_{\substack{\mathbf{W},\boldsymbol{\tau}^r,\boldsymbol{\tau}^t \\ \boldsymbol{\varpi}_g,\boldsymbol{\mu}_g,\boldsymbol{\vartheta}_e, \\ \boldsymbol{\varpi}_h^r,\boldsymbol{\mu}_h^r,\boldsymbol{\varrho}_e,\boldsymbol{\beta}}} \sum_{k\in\mathcal{K}_t}(\tau_k^t - \tau_e^t) + \sum_{k\in\mathcal{K}_r}(\tau_k^r - \tau_e^r) \tag{33a}$$
$$\text{s.t. (14b), (15g)}-\text{(15j), (23)}-\text{(25), (29), (31), (32),} \tag{33b}$$
$$\{\boldsymbol{\varpi}_g, \boldsymbol{\mu}_g, \boldsymbol{\vartheta}_e, \boldsymbol{\varpi}_h^r, \boldsymbol{\mu}_h^r, \boldsymbol{\varrho}_e\} \geq 0. \tag{33c}$$

Then, for a given value of $\mathbf{W}$, the subproblem concerning $\boldsymbol{\Phi}^l$ becomes a feasibility check. To enhance the convergence of the $\boldsymbol{\Phi}^l$ optimization, we introduce the slack variables $\boldsymbol{\lambda} = [\lambda_1, .., \lambda_K]$, $\tilde{\boldsymbol{\lambda}} = [\tilde{\lambda}_1, .., \tilde{\lambda}_K]$ and further neglecting the portion of LMI independent of $\boldsymbol{\theta}^l$, the respective power inequalities in (16a), (16b) and (25) are modified as

$$\begin{bmatrix} \varpi_{g,k}\mathbf{I}_{MN} + \mathbf{A}_k & \mathbf{a}_k \\ \mathbf{a}_k^{\mathrm{T}} & C_k^t - \lambda_k \end{bmatrix} \succeq \mathbf{0}, \forall k \in \mathcal{K}_t, \tag{34}$$

$$\begin{bmatrix} \beta_k^t - {\sigma_k^t}^2 - \mu_{g,k}M & \tilde{\mathbf{t}}_k^{t\,\mathrm{H}} \\ \tilde{\mathbf{t}}_k^t & \mathbf{I}_{(K-1)} \end{bmatrix} \succeq \mathbf{0}, \forall k \in \mathcal{K}_t, \tag{35}$$

Algorithm 1 AO and Line Search Algorithm

1: Set $t_e^l > 0$.
2: **Repeat** (line search algorithm)
3: **Initialize** $\tilde{R} = 0$, $r = 0$, select $\boldsymbol{\theta}^{l(0)}$ randomly.
4: **Repeat** AO algorithm
5: Under fixed $\boldsymbol{\theta}^l$, optimize $\mathbf{w}_k$ by solving (33)
6: Under fixed $\mathbf{w}_k$, optimize $\boldsymbol{\theta}^l$ by solving (40)
7: Calculate the objective $\tilde{R}_r = \min_k R_{s,k}^l$ given $\mathbf{w}_k$ and
8: $\boldsymbol{\theta}^l$; $r = r + 1$
9: **until** convergence of $\tilde{R}_r$
10: Update $t_e^l = t_e^l + \Delta t_e^l$ and proceed to step 3.
11: **until** $t_e^l \geq t_{e,max}^l$
12: **Output:** $(\mathbf{w}_k, \boldsymbol{\theta}^l)$ relies on the local optimal t_e^l.

Algorithm 2 Penalty CCP Algorithm

1: **Initialize** $\boldsymbol{\Phi}^{l[0]}, \gamma^{[0]} > 1$, and set $t = 0$.
2: **Repeat**.
3: **if** $t < t_{max}$
4: Update $\boldsymbol{\Phi}^{l[t+1]}$ from Problem (40);
5: $\alpha^{[t+1]} = \min\left\{\gamma\alpha^{[t]}, \alpha_{\max}\right\}$;
6: $t = t + 1$
7: **else**
8: Initialize with a new random $\boldsymbol{\Phi}^{l[0]}$, set up $\alpha^{[0]} > 1$ again,
9: and set $t = 0$
10: **end if**
11: **Until** $\|\mathbf{c}\|_1 \leq \zeta$ and $\left\|\boldsymbol{\Phi}^{l[t]} - \boldsymbol{\Phi}^{l[t-1]}\right\|_1 \leq \chi$
12: **Output:** $\boldsymbol{\Phi}^{l[t+1]} = \boldsymbol{\Phi}^{l[t]}$

and

$$\begin{bmatrix} \vartheta_{e,k}\mathbf{I}_{MN} + \mathbf{A}_k & \bar{\mathbf{a}}_k \\ \bar{\mathbf{a}}_k^{\mathrm{T}} & \bar{C}_k^t - \tilde{\lambda}_k \end{bmatrix} \succeq \mathbf{0}, \forall k \in \mathcal{K}_t. \tag{36}$$

Similarly, the corresponding power inequalities in (26a), (26b) and (32) are modified as

$$\begin{bmatrix} \widetilde{\mathbf{A}}_k + \begin{bmatrix} \varpi_{\mathrm{h},k}^r \mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \varpi_{\mathrm{g},k}\mathbf{I}_{MN} \end{bmatrix} & \widetilde{\mathbf{a}}_k \\ \widetilde{\mathbf{a}}_k^{\mathrm{H}} & C_k^r - \lambda_k \end{bmatrix} \succeq \mathbf{0}, \forall k \in \mathcal{K}_r, \tag{37}$$

$$\begin{bmatrix} m_k & \widetilde{\mathbf{t}}_k^{r\ \mathrm{H}} \\ \widetilde{\mathbf{t}}_k^r & \mathbf{I}_{(K-1)} \end{bmatrix} \succeq \mathbf{0}, \forall k \in \mathcal{K}_r, \tag{38}$$

and

$$\begin{bmatrix} \widetilde{\mathbf{A}}_k + \begin{bmatrix} \varrho_{\mathrm{e},k}\mathbf{I}_N & \mathbf{0} \\ \mathbf{0} & \vartheta_{\mathrm{e},k}\mathbf{I}_{MN} \end{bmatrix} & \widehat{\mathbf{a}}_k \\ \widehat{\mathbf{a}}_k^{\mathrm{H}} & \widehat{C}_k^r - \tilde{\lambda}_k \end{bmatrix} \succeq \mathbf{0}, \forall k \in \mathcal{K}_r. \tag{39}$$

However, the problem remains non-convex due to the presence of unit-modulus constraints. Therefore, similar to the approach followed in [50], we employ the p-ccp approach to address these non-convex constraints. Under the p-ccp framework, we express the constraint (14c) in an equivalent form $1 \leq |\phi_m^t|^2 + |\phi_m^r|^2 \leq 1, \forall m \in \mathcal{M}$. The non-convex components of these constraints are then linearized through $|\phi_m^{t\ [t]}|^2 - 2\,\mathrm{Re}(\phi_m^{t\ *}\phi_m^{t\ [t]}) + |\phi_m^{r\ [t]}|^2 - 2\,\mathrm{Re}(\phi_m^{r\ *}\phi_m^{r\ [t]}) \leq -1, \forall m \in \mathcal{M}$, at fixed $\phi_m^{l\ [t]}$. Thus, the convex subproblem for $\boldsymbol{\Phi}^l$ is expressed as

$$\max_{\substack{\boldsymbol{\Phi}^t, \boldsymbol{\Phi}^r, \boldsymbol{\tau}^r, \boldsymbol{\tau}^t \\ \boldsymbol{\varpi}_g, \boldsymbol{\mu}_g, \boldsymbol{\vartheta}_e, \boldsymbol{\varpi}_h^r, \boldsymbol{\mu}_h^r \\ \boldsymbol{\varrho}_e, \boldsymbol{\beta}, \mathbf{c}, \tilde{\boldsymbol{\lambda}}, \boldsymbol{\lambda}}} \sum_{k \in \mathcal{K}_t} (\tau_k^t - \tau_e^t) + \sum_{k \in \mathcal{K}_r} (\tau_k^r - \tau_e^r) - \alpha^{[t]} \sum_{m=1}^{2M} c_m \tag{40a}$$

$$\text{s.t. } (15\text{g}) - (15\text{j}), (33\text{c}), (34) - (39), \tag{40b}$$

$$|\phi_m^{t\ [t]}|^2 - 2\,\mathrm{Re}(\phi_m^{t\ *}\phi_m^{t\ [t]}) + |\phi_m^{r\ [t]}|^2 - 2\,\mathrm{Re}(\phi_m^{r\ *}\phi_m^{r\ [t]}) \leq c_m - 1, \forall m, \tag{40c}$$

$$|\phi_m^{t\ [t]}|^2 + |\phi_m^{r\ [t]}|^2 \leq 1 + c_{M+m}, \forall m, \tag{40d}$$

$$\tilde{\boldsymbol{\lambda}} \geq 0, \boldsymbol{\lambda} \geq 0, \mathbf{c} \geq 0, \tag{40e}$$

where $\mathbf{c} = [c_1, \ldots, c_{2M}]^T$ represents the slack variables introduced to enforce the equivalent linear constraints for the unit-modulus constraints. The objective function includes the penalty term $||\mathbf{c}||_1$, which is adjusted by the regularization coefficient $\alpha^{[t]}$ to control the constraints' feasibility.

After optimizing $(\mathbf{w_k}, \boldsymbol{\Phi}^t)$, we update t_e^l and reiterate the optimization using the AO algorithm. The range for t_e^l is limited to $t_e^l \in (0, t_{e,max}^l)$ to ensure the eavesdropper's achievable rate remains below the worst-case rate $R_{s,k}^l$ for users without eavesdroppers. Going beyond $t_{e,max}^l$ violates security principles and jeopardizes secure transmission. To find the optimal t_e^l, we use a uniform sampling-based line search algorithm within the range $(0, t_{e,max}^l)$ [52]. Once determined, this leads to a local optimum solution for the original problem **P0**. By employing the AO framework, we iteratively tackle the problem in (15) by addressing problems (33) and (40). Specifically, $\phi_m^{l\ [t]}$ in constraint (40c) and $\alpha^{[t]}$ are updated iteratively using CCP, while $\boldsymbol{\theta}^{l(n)}$ undergoes iterative updates within the outer AO framework. Further, the combined AO and line search algorithms are summarized in **Algorithm 1**.

Problem (40) is a semidefinite program (SDP) and can be solved by the CVX tool. The steps of finding a feasible solution of $\boldsymbol{\Phi}^l$ are summarized in **Algorithm 2**. We remark that: a) When ζ is sufficiently low, constraints (14c) in the original problem is guaranteed by $||c||_1 \leq \zeta$; b) The maximum value α_{max} is imposed to avoid a numerical problem, that is, a feasible solution satisfying $||c||_1 \leq \zeta$ may not be found when the iteration converges to the stopping criterion $||\boldsymbol{\Phi}^{l[t]} - \boldsymbol{\Phi}^{l[t-1]}||1 \leq \chi$ with the increase of $\alpha^{[t]}$; c) Stopping criteria $||\boldsymbol{\Phi}^{l[t]} - \boldsymbol{\Phi}^{l[t-1]}||1 \leq \chi$ controls the convergence of Algorithm 2; d) As mentioned in [53], a feasible solution to Problem (40) is guaranteed by imposing a maximum number of iterations t_{max} and, in case it is reached, we restart the iteration based on a new initial point. Further, as proved in [54], **Algorithm 1** generates a sequence of $\{\mathbf{W}^{(\star)}, \boldsymbol{\Phi}^{t(\star)}, \boldsymbol{\Phi}^{r(\star)}\}$ which corresponds to non-decreasing values of problem (14) objective function. As a result, the stationary point of the original problem is obtained after a sufficient number of iterations.

A. Extension to Phase-shift coupled Mode

With simple mathematical derivations, it can be proved that the phase-shift coupled constraint in (1) and (2) for the ES protocol can be used as [44]

$$\theta_m^{t\ [t]}\theta_m^r + \theta_m^t\theta_m^{r\ [t]} - \theta_m^{t\ [t]}\theta_m^{r\ [t]} + 1 \geq 0. \tag{41}$$

Owing to (41), the beamforming design problem in (40) can be rewritten for phase-shift coupled ES mode. To this end, we

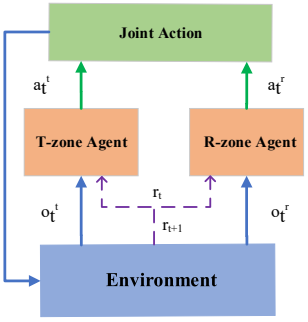

Fig. 2: Representation of MA-DRL framework.

adopt a penalty-based convex approximation framework to solve it as follows

$$\max_{\substack{\boldsymbol{\Phi}^t,\boldsymbol{\Phi}^r,\boldsymbol{\tau}^r,\boldsymbol{\tau}^t \\ \boldsymbol{\varpi}_g,\boldsymbol{\mu}_g,\boldsymbol{\vartheta}_e,\boldsymbol{\varpi}_h^r,\boldsymbol{\mu}_h^r \\ \boldsymbol{\varrho}_e,\boldsymbol{\beta},\boldsymbol{c},\tilde{\boldsymbol{\lambda}},\boldsymbol{\lambda}}} \sum_{k\in\mathcal{K}_t}(\tau_k^t-\tau_e^t)+\sum_{k\in\mathcal{K}_r}(\tau_k^r-\tau_e^r)-\alpha^{[t]}\sum_{m=1}^{2M}c_m \tag{42a}$$

$$\text{s.t.} \quad (15g)-(15j),(33c),(34)-(39),(40e),(41). \tag{42b}$$

IV. Multi-Agent DRL Based Framework

DRL holds significant promise for addressing complex challenges in wireless communication systems by enabling agents to acquire optimal policies through continuous interactions with their environment. However, many existing DRL-based approaches predominantly center on single-agent systems, potentially resulting in inefficiencies when dealing with a growing number of network nodes. Our study proposes DRL, particularly in an MA framework, to address these challenges. By leveraging multiple DRL-based agents, we achieve superior adaptability, scalability, and robustness to uncertainties compared to single-agent DRL and traditional optimization techniques [55]. This approach enhances real-time performance, effectively manages high-dimensional optimization problems, and provides resilient solutions for dynamic environments. In our system model depicted in Fig. 1, we categorize communication links into different zones. The R-zone includes links connecting the BS, STAR-RIS, and R-users, encompassing both the links from the BS to R-users and the links from the BS to STAR-RIS to R-users. Conversely, the T-zone encompasses links associated with the BS, STAR-RIS, and T-users. In our framework, we introduce two agents, one for the R-zone and another for the T-zone, with each agent responsible for making decisions within its respective zone. These agents interact with the environment, comprising the BS, STAR-RIS, and users, and enhance their decision-making policies through learning from their experiences.

In an MA system, each individual agent faces the challenge of acquiring precise knowledge of the entire trained model, including information about the states and rewards of other agents. Each agent takes on the role of a policy maker, guiding an agent's learning and experience updates within the environment until an optimal policy is achieved. For training the agents, we utilize the PPO algorithm, a state-of-the-art method in DRL. PPO ensures stable and efficient learning by iteratively adjusting the policy parameters based on the experiences collected. The agents within this distributed MA-DRL framework communicate, exchange information, and coordinate their actions to derive the optimal policy. To tackle the potential instability inherent in the MA approach, we employ the PPO algorithm, which aids in training neural networks. Our choice of MA-PPO addresses key challenges in MA-DRL: non-stationarity, scalability with joint action spaces, and partial observability. PPO's actor-critic architecture mitigates non-stationarity, while its decentralized approach and shared policy parameters enhance scalability [56]. Additionally, integrating local observations with shared experiences fosters robust policy development.

A. MDP Based Problem Formulation

In the depicted MA-DRL framework (Fig. 2), the Markov decision process (MDP) framework is structured with five components, forming a five-tuple $\mathcal{S},\mathcal{A},\mathcal{R},O,\mathcal{P}$. When the agent takes action $a_t \in \mathcal{A}$ based on the policy $\pi(a_t|s_t)$, $\mathcal{P}$ denotes the probability of transitioning from the current state $s_t \in \mathcal{S}$ to the next state $s_{t+1} \in \mathcal{S}$. Furthermore, the agent obtains an observation o_t from the set O rather than directly acquiring s_t from the domain $\mathcal{S}$. At time t, the immediate reward of the agent is denoted by the variable r_t. Specifically, the definitions of $\mathcal{S},\mathcal{A}$, $\mathcal{R}$, and O are as follows:

1) State Space and Local Observations: The overarching concept is to include maximal environmental data related to problem (**P0**) within the state space. Let $\mathcal{S}$ represent the system state space, encompassing global channel conditions and behaviors of all agents. The formulation of this state space relies on the information available to the BS, obtained either directly or indirectly, and holds a crucial role in defining the reward function expressed as:

$$\mathcal{S}=\{\mathbf{J}_k^t,\mathbf{J}_k^r,\mathbf{V}_{e,k}^t,\mathbf{V}_{e,k}^r\}, \tag{43}$$

where

$$\mathbf{J}_k^l \triangleq \begin{cases}\boldsymbol{\theta}^{lH}\mathbf{G}_k^l, & l=t,\forall k\in\mathcal{K}_t,\\ \mathbf{h}_k^H+\boldsymbol{\theta}^{lH}\mathbf{G}_k^l, & l=r,\forall k\in\mathcal{K}_r,\end{cases} \tag{44}$$

and

$$\mathbf{V}_{e,k}^l \triangleq \begin{cases}\boldsymbol{\theta}^{lH}\mathbf{G}_e^l, & l=t,\forall k\in\mathcal{K}_t,\\ \mathbf{h}_e^H+\boldsymbol{\theta}^{lH}\mathbf{G}_e^l, & l=r,\forall k\in\mathcal{K}_r.\end{cases} \tag{45}$$

Please note that $\mathbf{J}_k^l$ and $\mathbf{V}_{e,k}^l$ denote the effective channels in the T-zone and R-zone, respectively, representing communication between the BS and users or eavesdroppers. The phase-shift matrices at STAR-RIS and channel information of all the users are components of the current state $s_t \in \mathcal{S}$. Due to imperfect channel state information, the feedbacks $\mathbf{J}_k^l$ and $\mathbf{V}_{e,k}^l$ account for channel error vectors [57]. However, it is essential to note that the system state $\mathcal{S}$ remains unknown to each individual agent. In the MA-DRL, agents operate based on their limited local observations and experiences, lacking access to complete global state information. Each agent maintains its partial state representation, including relevant information specific to the agent. Since we have two distinct agents, one dedicated to the T-zone and the other to the R-zone, the individual observation spaces are denoted $o_t^t=\{\mathbf{J}_k^t,\mathbf{V}_{e,k}^t\}$, and $o_t^r=\{\mathbf{J}_k^r,\mathbf{V}_{e,k}^r\}$, respectively. Hence, the observation space set is given by $O=\{o_t^l \mid l=\{t,r\}\}$. The observations from both the T and R agents are stored in a centralized memory buffer, which is implemented at the BS. Each agent can access this memory through its dedicated control channel. During training, the neural network undergoes offline updates by randomly sampling observations from this memory. Subsequently, each agent leverages these stored observations to inform its decision-making for the next actions.

This methodology ensures efficient learning and adaptation in the dynamic network environment, progressively enhancing the decision-making capabilities of the agents.

2) Action Space: The action space is constructed by the transmit beamforming matrix $\mathbf{W}$, the phase-shifts matrix $\mathbf{\Phi}^l$, and the channel error vectors $\Delta\mathbf{G}_k^l, \Delta\mathbf{h}_k, \Delta\mathbf{G}_e^l$, and $\Delta\mathbf{h}_e$ for the corresponding T-zone and R-zone channels. Thus, the action space can be given by

$$\mathcal{A}=\{\{\mathbf{w}_k\},\{\phi_m^l\},\{\Delta\mathbf{G}_k^l, \Delta\mathbf{h}_k, \Delta\mathbf{G}_e^l, \Delta\mathbf{h}_e\}\}, l=\{t,r\}, \forall k\in\mathcal{K}. \tag{46}$$

Given the presence of multiple agents, the action space is formulated to incorporate the independent actions of each agent. In particular, the action space $\mathcal{A}^t \in \mathcal{A}$ for the T-agent is expressed as:

$$\mathcal{A}^t=\{\{\mathbf{w}_k\},\{\phi_m^t\},\{\Delta\mathbf{G}_k^t, \Delta\mathbf{G}_e^t\}\}, \forall k \in \mathcal{K}_t. \tag{47}$$

Similarly, for the R-agent, the action space $\mathcal{A}^r \in \mathcal{A}$ can be defined as

$$\mathcal{A}^r=\{\{\mathbf{w}_k\},\{\phi_m^r\},\{\Delta\mathbf{G}_k^r, \Delta\mathbf{h}_k, \Delta\mathbf{G}_e^r, \Delta\mathbf{h}_e\}\}, \forall k\in\mathcal{K}_r. \tag{48}$$

As part of the learning process, the agent advances a step at a time. The environment transitions from the current state, s_t, to the subsequent state s_{t+1} when the agent takes action $a_t \in \mathcal{A}$ at time step t. Consequently, the agent receives a reward r_t. With distinct action spaces for each agent, the T and R agents independently make decisions based on their observations and policies. This decentralized approach enables both agents to optimize their modes, considering the overarching system objectives, ultimately enhancing performance in terms of throughput, coverage, and overall system efficiency.

3) Reward: The reward function in the paper quantifies the immediate return obtained by taking action in a given state s_t, with a design to maximize the sum rate of the system model described in (14) through joint optimization of the action space. In DRL, the agent aims to choose actions that maximize the cumulative reward over time by interacting with the environment in discrete time steps. To achieve this, we designate the reward for each agent as $r_t^l,\ l=\{t,r\}$. To this end, the reward function at the time step t is defined as $r_t = \sum_l r_t^l$.

The objective of learning is to determine an optimal policy π^* that maximizes the expected overall reward from any initial state. By defining reward functions for both agents, the proposed MA-DRL framework aims to maximize the accumulated rewards obtained by the T and R-zone agents throughout the interaction horizon. In this context, the term reward function denotes the cumulative value of discounted rewards denoted as $\mathcal{R}_t = \sum_{t=0}^{T} \gamma_t\ r_t$, where $\gamma \in (0,1]$ is the discount factor representing the impact of the reward at time step t.

Thus, our MA-DRL framework ensures adherence to constraints (14a) to (14e) through a synergistic integration of observations, action spaces, and reward functions tailored for each agent.

- *Observations:* Agents receive localized observations that are pivotal for informed decision-making. These observations include effective channel information and network feedback, crucial for dynamic adjustments within the framework's constraints.
- *Action Space and Constraints Compliance:* The action space is carefully designed to allow for adjustments that remain within the power budget and phase shift regulations ((14b) and (14c)), as well as ensuring the satisfaction of the minimum rate requirements ((14d) and (14e)). The real-time observations enable agents to assess the impact of their actions and adjust strategies to comply with these constraints.
- *Reward Function:* The reward structure is designed to penalize actions that risk constraint violations, guiding agents toward strategies that optimize system performance while respecting the defined limits. This mechanism relies on observations to identify and reinforce constraint-compliant behaviors.

In summary, the effective use of observations within our MA-DRL framework enables each agent to act in a manner that is not only aligned with achieving high sum secrecy rates but also in strict adherence to the system's operational constraints. This ensures a balanced approach to optimizing network performance while maintaining compliance with all specified requirements.

B. PPO Based Algorithm

PPO algorithms belong to a category of DRL methods recognized for their stability and superior training performance. Typically, the actor network receives the system state s_t as input. Given that the state involves varying magnitudes of channel information, Z-score normalization is applied to standardize the inputs throughout the entire episode. Further, the actor network produces two components in its output: the mean vector and standard deviation of the policy.

Certainly, PPO is a policy optimization technique that seeks to maximize the expected cumulative reward by employing a surrogate objective function. This function ensures that the updated policy remains close to the old one, preventing drastic changes that may lead to instability. This adaptability makes PPO effective in handling diverse environments. Moreover, PPO addresses the challenge of selecting an appropriate step size by enabling modifications to the objective function during training.

The inherent randomness in the policy $\pi_\mu(a_t|s_t)$ implies that the probability of future trajectories $\tau = \{s_t, a_t\}$ relies on parameters governing the action sampling probability. Consequently, the objective function, dependent on μ, can be expressed as $J(\mu) = \underset{\tau\sim\pi_\mu}{\mathbb{E}}[\mathcal{R}(\tau)]$, where μ represents the neural network parameter of the policy function π. In an algorithm iterating between sampling and optimization, $\mathbb{E}$ denotes the empirical average calculated over a finite batch of samples. Further, by utilizing the gradient ascent method the parameter μ is updated as follows

$$\mu_{t+1} \leftarrow \mu_t + \omega\nabla_\mu J(\mu), \tag{49}$$

where ω is the learning rate or step size. The primary objective is to find the optimal policy π^* for the BS, maximizing $J(\mu)$ through repeated gradient estimation:

$$\nabla_\mu J(\mu) = \mathbb{E}_{\pi_\mu}\left[\nabla_\mu \log \pi_\mu(s_t,a_t) A_{\pi_\mu}(s_t,a_t)\right], \tag{50}$$

where $A_{\pi_\mu}(s_t,a_t)$ is the advantage function assessing whether an action is superior or inferior to the policy's default behavior. The advantage function at time step t is defined as $A_{\pi_\mu}(s_t,a_t) = Q_{\pi_\mu}(s_t,a_t) - V_{\pi_\mu}(s_t)$, with $Q_{\pi_\mu}(s_t,a_t) = \mathbb{E}_{\pi_\mu}[\mathcal{R}_t|s=s_t, a=a_t]$ and $V_{\pi_\mu}(s_t) = \mathbb{E}_{\pi_\mu}[\mathcal{R}_t|s=s_t]$ representing the action value and state value functions, respectively. However, estimating

the advantage function is susceptible to bias, and an improper learning rate can lead to instability or slow convergence. To address this, PPO uses $\chi_t(\mu) = \pi_\mu(a_t|s_t)/\pi_{\mu_{old}}(a_t|s_t)$, a probability ratio constraining the update range, reducing sensitivity to learning rates and enhancing efficiency.

To satisfy the trust region constraint, this PPO-based approach maximizes a clipping surrogate objective function. Thus, the clipping surrogate objective, limiting substantial weight updates, is expressed as:

$$J_t^{\text{clip}}(\mu) = \mathbb{E}_\pi \Big[\min\Big(\chi_t(\mu) A_{\pi_{\mu_{old}}}(s_t, a_t), \text{clip}\left(\chi_t(\mu), 1-\epsilon, 1+\epsilon\right) A_{\pi_{\mu_{old}}}(s_t, a_t)\Big)\Big], \quad (51)$$

where ϵ is the hyperparameter that tunes the fraction used for clipping within the specified range. The second term in the clipping surrogate objective adjusts the probability ratio within the range $[1-\epsilon, 1+\epsilon]$. This approach creates a lower bound and pessimistic estimate of the unclipped objective by selecting the minimum of the clipped and unclipped objectives [58]. To enhance the objective, a value function error term and an entropy bonus are included to ensure sufficient exploration. Combining these terms, the final objective of the proposed algorithm is formulated as

$$J_t^{\text{PPO}}(\mu) = \mathbb{E}_\pi \left[J_t^{\text{clip}}(\mu) - c_1 L_t^{\text{VF}}(\mu) + c_2 \mathbb{S}_{\pi_\mu}(s_t)\right], \quad (52)$$

where c_1, c_2 are coefficients, $L_t^{\text{VF}} = (V_{\pi_\mu}(s_t) - V_t^{\text{targ}})^2$ is the value function error, and $\mathbb{S} = -\sum_a \pi_\mu(a|s_t)\log\pi_\mu(a|s_t)$ is the entropy bonus term. The generalized advantage estimation function is defined as

$$A_t = r_t + \gamma V_{\pi_\mu}(s_{t+1}) - V_{\pi_\mu}(s_t). \quad (53)$$

Moreover, in the realm of the MA-DRL system model, the PPO-based algorithm is applied individually to each individual agent, facilitating tailored policy optimization and adaptation. The proposed MA-PPO framework involves a structured exchange of information that includes local observations, shared experiences through a centralized memory buffer, reward signals, and coordinated actions. The detailed procedure is outlined in **Algorithm 3**.

The convergence of our proposed PPO algorithm is rigorously established. According to Theorem 1 and Corollary 1 in [59], under suitable conditions on learning rates and the characteristics of the loss functions, our PPO algorithm converges to a local minimum of the associated objective function. These conditions ensure stability and convergence of the algorithm, making it suitable for optimizing policies in dynamic environments. For detailed proofs and additional algorithmic specifics, please refer to Sect. 3.1 and 3.2 of the [59].

C. DDPG Based Algorithm

DDPG enhances the actor-critic framework using deep neural networks (DNNs) for modeling policy and value functions, addressing high-dimensional state and action spaces effectively. Notably, DDPG handles continuous action spaces, making it suitable for such environments. The architecture includes:

Critic Network: Also known as the Q-network with parameter ϱ_c, it processes state s and action a inputs to yield $Q(s^t, a^t; \varrho_c)$. The Q-function is defined as

$$Q_\pi\left(s^t, a^t\right) = \mathbb{E}_\pi\left[R^t | s^t = s, a^t = a\right], \quad (54)$$

Algorithm 3 Proposed MA-DRL Based PPO Algorithm

```
Initialize the parameter settings for the proposed STAR-RIS aided secure communication system, neural networks at t = 0
Input: Environment, observation space O
Output: A={{w_k}, {φ_m^l}}
Initialize: μ, π_μ, V_{π_μ}, memory buffer
for episode = 1 → E do
    Get the initial observation state s_t, t = 0 and memory buffer;
    for t = 1 → S_T do
        for each agent do
            Identify o_t^l and determine the action a_t^l by employing sampling of the corresponding density function;
            Choose action a_t^l based on current policy;
        end for
        Obtain r_t^l and the next state s_{t+1};
        Each agent takes actions and receives reward r_{t+1}
        for each agent do
            Observe o_{t+1}^l and calculate r_{t+1}^l
            Store the transition (o_t^l, a_t^l, r_{t+1}^l, done) in the memory buffer;
        end for
    end for
    for each agent do:
        Compute the advantage function using (53);
        Compute the final objective of the proposed PPO algorithm using (52);
        Compute gradient according to (50);
        Update μ according to (49) via gradient ascent method;
    end for
end for
```

and updated via the Bellman expectation equation. The optimal Q-function is

$$Q^*\left(s^t, a^t\right) = r^t + \bar{\gamma} \max_{a^{t+1}\in\mathcal{A}} Q^*\left(s^{t+1}, a^{t+1}\right), \quad (55)$$

with the optimal action derived by

$$a^* = \arg\max_{a\in\mathcal{A}} Q^*(s, a). \quad (56)$$

Actor Network: Known as the policy network, it maps state s to continuous action a, denoted as $a^t = \pi(s^t; \varrho_\mu)$. The actor optimizes the state-value function using the policy gradient method:

$$\nabla_{\varrho_\mu} J(\varrho_\mu) \approx \mathbb{E}[\nabla_a Q(s^t, a; \varrho_c)|_{a=\pi(s^t;\varrho_\mu)} \nabla_{\varrho_\mu}\pi(s^t; \varrho_\mu)], \quad (57)$$

where $J(\varrho_\mu) = \mathbb{E}_{s\sim\varrho_c, a\sim\varrho_\mu} R(s, a)$. The critic minimizes the loss function:

$$L(\varrho_c) = \mathbb{E}\left[(y^t - Q(s^t, a^t; \varrho_c))^2\right], \quad (58)$$

with $y^t = R^t + \bar{\gamma} Q(s^{t+1}, \pi(s^{t+1}; \varrho'_\mu); \varrho'_c)$. Target networks $\pi(s^t; \varrho'_\mu)$ and $Q(s^t, a^t; \varrho'_c)$ stabilize training, updated as $\varrho'_\mu = \zeta\varrho_\mu + (1-\zeta)\varrho'_\mu$ and $\varrho'_c = \zeta\varrho_c + (1-\zeta)\varrho'_c$.

Unlike Q-learning, policy gradient methods optimize the policy directly, avoiding overestimation bias. During testing, the best policy is selected deterministically. In MA-DRL, each agent uses DDPG for personalized policy optimization. This specific methodology is described in **Algorithm 4**.

D. Complexity Analysis of Algorithm 1 and Algorithm 4

In this section, we conduct an analysis of the computational complexity associated with the proposed methods for robust

TABLE II: Computational complexity of conventional robust optimization problem, MA, PPO, and DDPG algorithms.

Robust Optimization	$\mathcal{O}\left(\left(\sum_{i=1}^{I} c_i\right)^{1/2} NK\left[N^2K^2+NK\sum_{i=1}^{I} c_i^2+\sum_{i=1}^{I} c_i^3\right]\right)+\mathcal{O}\left(\left(\sum_{i=1}^{I} c_i\right)^{1/2} M\left[M^2+M\sum_{i=1}^{I} c_i^2+\sum_{i=1}^{I} c_i^3\right]\right)$
Proposed MA	$\mathcal{O}\Big(W^2\big(10M\sqrt{\frac{a}{\epsilon}M}+12M\sqrt{\frac{a}{\epsilon M}}+10\sqrt{\frac{a}{\epsilon}M}+6\sqrt{\frac{a}{\epsilon M}}+(3M+6)\frac{a}{\epsilon}+4M+4\big)^2+W\big(2M\sqrt{\frac{a}{\epsilon}M}+4M\sqrt{\frac{a}{\epsilon M}}+2\sqrt{\frac{a}{\epsilon}M}+3\sqrt{\frac{a}{\epsilon M}}+(2M+2)\frac{a}{\epsilon}+5\big)\Big)$
PPO	$\mathcal{O}\left(W\left(12M\sqrt{\frac{a}{\epsilon}M}+16M\sqrt{\frac{a}{\epsilon M}}+13\sqrt{\frac{a}{\epsilon}M}+8\sqrt{\frac{a}{\epsilon M}}+\big(4M+8\big)\frac{a}{\epsilon}+4M+5\right)\right)$
DDPG	$\mathcal{O}\left(W\left(12M\sqrt{aM}+16M\sqrt{\frac{a}{M}}+13\sqrt{aM}+8\sqrt{\frac{a}{M}}+\big(4M+8\big)a+4M+5\right)\right)$

Algorithm 4 DDPG Algorithm for Each Agent

Input: Initialize the parameter settings for the proposed system model, neural networks at $t=0$
Input: Exploration parameter ϵ, learning rate Ω, number of episodes E
Initialize the actor-network, $\pi(s^t;\varrho_\mu)$ and the critic network $Q(s^t,a^t;\varrho_c)$ with the weights ϱ_μ and ϱ_c.
Create the target DNNs by setting $\varrho'_\mu \leftarrow \varrho_\mu$ and $\varrho'_c \leftarrow \varrho_c$
Initialize a replay buffer
Initialization: get initial ϱ_μ from server
for $ep=1\rightarrow E$ **do**
    Initialize a random process η for action exploration
    Receive initial observation state s^1
    for $t=1\rightarrow T$ **do**
        Obtain action a^t from the actor-network;
        Add exploration noise to a^t as $a^t=a^t+\eta$
        Calculate the instant reward r^t
        Observe the new state s^{t+1}
        Store experiences in the buffer and sample random mini-batches of experiences to train the DNNs
        Set the expected return y^t
        Update the actor policy via (57) and critic via (58)
        Update the target actor ϱ'_μ and the target critic ϱ'_c
    end for
end for

transmission design in Algorithm 1. Given that all resulting convex problems involve LMIs, soc constraints, and linear constraints solvable through standard interior point methods [51], we aim to compare the computational complexities of different methods based on their worst-case runtime. The general expression for this runtime, excluding the complexities of linear constraints and soc, is presented in Table I. Specifically, for problem (33), $n_1 = NK$ represents the number of variables, while $I = 3K_t + 3K_r$ denotes the count of LMIs, each sized as c_i. Similarly, for problem (40), $n_2 = M$ signifies the number of variables involved. Additionally, we define $c_i = 3MNK_r + 4NK_r + 2MNK_r + 2K_r + 2NK_t + 3K_t + K_t^2 + K_r^2$.

Next, we analyze the complexity analysis of the DRL algorithm which primarily hinges on several parameters, including dataset size, state and action dimensions, complexities associated with forward and backward propagations within the neural networks, and the structural configuration of the fully connected neural network architecture. Specifically, within the actor network, weight adjustments occur through both forward and backward propagations, while in the critic network, only forward propagation is utilized. Additionally, as outlined in [60], the dimensions of the first fully connected layer (f_1) and second fully connected layer (f_2) in the actor-network are determined based on various factors such as the number of actor-network learning samples denoted as a and the count of reflecting elements represented by N. These dimensions are mathematically defined as $f_1 = \sqrt{aN} + 2\sqrt{a/N}$ and $f_2 = \sqrt{aN}$ and $f_2 = \sqrt{aN}$.

On a similar note, within the MA-DRL framework in Algorithm 2, the utilization of multiple agents results in diverse actions undertaken by each agent. Consequently, the incorporation of multiple agents in the MA-DRL algorithm, coupled with the global critic's oversight across the actor networks, leads to an exponential increase in actor-network complexity relative to the number of agents. In our proposed MA-DRL framework, we specifically operate with two distinct agents. With this context in mind, we examine the computational complexities associated with the robust optimization problem, MA-DRL-based algorithm, as well as those of the PPO, and DDPG, presented in **Table II**.

V. Numerical Simulations and Discussion

In this section, the exhaustive simulation-based results are presented to verify the convergence and effectiveness of the proposed algorithm. The location of the BS is fixed at $(0,0)$m. The STAR-RIS is fixed at $(50,10)$ m. The R-zone and T-zone users are randomly and uniformly distributed in a circle centered at (30,0) m and (70,0) m with a radius of 5 m, respectively. The R-zone and T-zone eavesdroppers are fixed at (20,0) m and (65,0) m, respectively. All the considered channels are assumed to include large-scale and small-scale fading. The large-scale path-loss model is $\chi_l = -\chi_0 - 10\varsigma\log_{10}(d)$, where ς represents the path-loss exponent, d denotes the link distance in meters, χ_0 is the reference path-loss at 1m distance which is defined as 40 dB utilizing the 3GPP-UMi model at a carrier frequency of 3.5 GHz [61]. The small-scale fading is assumed to be Rayleigh fading distribution. Furthermore, the neural network parameters are updated using the Adam optimizer and the activation function used is ReLU. The proposed PPO and DDPG frameworks employ two hidden layers each, with 256 neurons [62]. Moreover, we set the hyperparameters as $\gamma = 0.9$ [63], critic and actor-network learning rates are given as $c = 0.0002$ and $a = 0.0001$, respectively [64], memory buffer $W = 10000$ [65], $\epsilon = 0.2$ [58], size of buffer = 32, episodes $\mathcal{E} = 2000$, steps $S_T = 500$. We set $M=16$, $K_t=K_r=3$ such that $K=K_t+K_r=6$, $N=6$, $R_{\min}=2$ bits/s/Hz [51], $\xi^l_{g,k}=\xi_{h,k}=\xi^l_{g,e}=\xi_{h,e}=-15$ dB, ${\sigma^l_k}^2={\sigma^l_{e,k}}^2=-100$dBm [33], $P_T=30$ dBm, convergence factor as 10^{-3}, and sampling interval $\Delta t^l_e=0.1$ [51]. For the robust beamforming design, we considered 100 channel realizations. The results are then averaged over these 100 channel realizations. This averaging process helps to ensure that the performance metrics reported are reflective of typical system behavior under a variety of conditions. Unless stated otherwise, the parameters follow the mentioned specifications.

We label our proposed MA-DRL framework and the conventional robust transmission design, utilized for secure transmission design in a multi-user STAR-RIS-aided communication system, as "MA" and "Con," respectively. Additionally, we conduct comparisons with the following benchmark schemes:

1) **Perfect:** This analysis compares our proposed robust transmission design under imperfect CSI conditions with a scenario where the transmitter possesses perfect CSI [66].
2) **Passive-RIS**: In this approach, we apply our proposed beamforming design at the BS while employing conventional passive beamforming at the RIS for secure communication. This approach substitutes STAR-RIS with passive-RIS in all comparative simulations [67], [68].
3) **Random-RIS**: This scheme implements our proposed beamforming design at the BS while incorporating random passive beamforming at the RIS [69].
4) **No-RIS**: This scheme represents secure communication without employing RIS, essentially the scenario without any RIS usage. Comparing our proposed beamforming design against this case allows us to emphasize the advantages brought by RIS deployment in secure communication [69].
5) **Non-Secrecy**: Combining beamforming design at the BS with passive beamforming at STAR-RIS, this scheme focuses on non-secure communication without considering eavesdroppers' effects. The aim is to analyze the system's performance without security considerations [70].

The convergence pattern of the proposed MA-DRL framework and the robust optimization problem are shown in Fig. 3 and Fig. 4 respectively. In Fig. 3, rewards indicating the overall system sum secrecy rate across episodes are illustrated. This evaluation contrasts the proposed MA approach with benchmark PPO and DDPG single-agent algorithms in a configuration involving $M = 16$ RIS elements. The MA algorithm achieves a peak reward of 29.8 bits/s/Hz, surpassing PPO's 27.9 bits/s/Hz and DDPG's 25.2 bits/s/Hz. This is because of MA's consideration of system-wide objectives while optimizing individual agent policies. By comprehensively evaluating actions' global impact, MA allocates resources more coherently and efficiently. Employing PPO within the MA framework allows broader action exploration, aiding in reaching optimal solutions and enhancing overall performance. Regarding convergence, MA converges at around 600 episodes, PPO at roughly 360 episodes, and DDPG within the initial 200 episodes due to its deterministic policy output. The MA-PPO scheme demands agents to consider others' behaviors, necessitating more episodes to discover optimal policies that maximize system performance. Conversely, DDPG's lack of

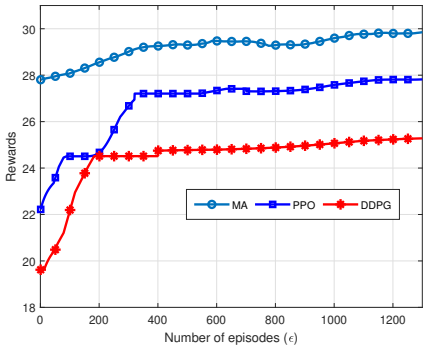

Fig. 3: Convergence behaviour of proposed DRL framework.

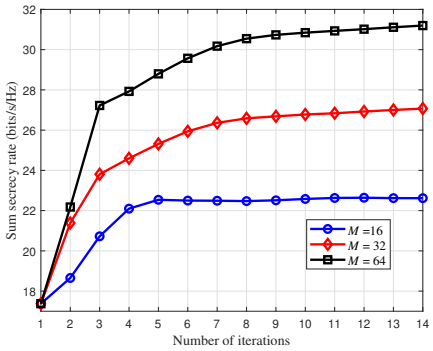

Fig. 4: Convergence behaviour of robust optimization problem.

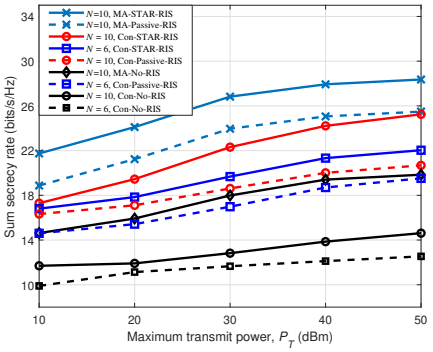

Fig. 5: Impact of P_T over varying N.

action exploration leads to faster convergence but lower rewards. PPO's exploration during training causes slower convergence, potentially achieving higher rewards.

Fig. 4 demonstrates the convergence behavior of the robust optimization problem. It showcases the system's achievable sum secrecy rate against iteration count for varying STAR-RIS elements ($M = \{16, 32, 64\}$) at P_T =30 dBm. Notably, as the number of STAR-RIS elements increases, the attainable sum secrecy rate also increases under the proposed solution. Initially, performance exhibits steady enhancement with more iterations, yet it eventually stabilizes after a certain threshold. For instance, at $M = 64$, the algorithm peaks at a sum secrecy rate of 31.1 bits/s/Hz. Similar trends occur for $M=\{16, 32\}$, achieving sum secrecy rates of 22.4 and 26.5 bits/s/Hz, respectively. Specifically, for $M = \{16, 32, 64\}$, the sum secrecy rate stabilizes around $\{6, 8, 10\}$ iterations, indicating the algorithm's convergence to a stable solution within a relatively small iteration count. This behavior highlights the effectiveness of the proposed algorithm in optimizing the system sum secrecy rate.

In Fig. 5, the impact of adjusting maximum transmit power on secrecy performance across various BS antenna configurations is shown for both the MA framework and conventional robust optimization. As expected, higher power and larger antenna arrays lead to increased degrees of freedom (DoF), resulting in better system performance, as validated in Fig. 5. Consistently, the MA framework outperforms the conventional method across all scenarios. Moreover, we compare these against passive-RIS and No-RIS schemes. The proposed STAR-RIS approach, optimizing the phase-shift matrix, outperforms the other scenarios by providing comprehensive spatial coverage and greater control over signal propagation compared to passive RIS. Notably, using STAR-RIS brings a significant performance improvement compared to operating the system without RIS, underlining its substantial

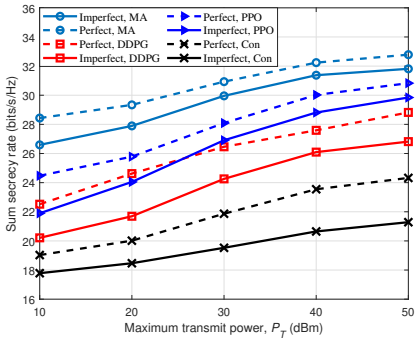

Fig. 6: Comparison of various algorithms performance.

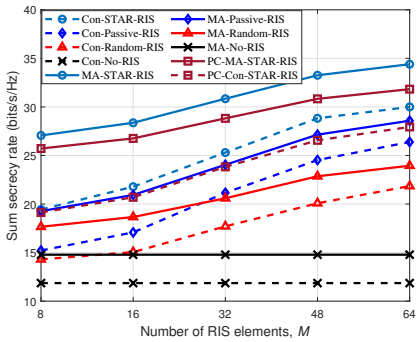

Fig. 7: Impact of M.

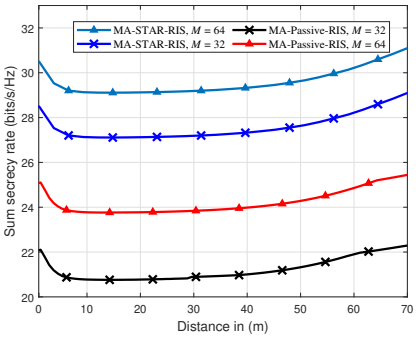

Fig. 8: Impact of STAR-RIS distance concerning BS and users.

impact.

In Fig. 6, we compare the proposed system's performance, considering robust transmission design with both imperfect and perfect CSI variations across varying maximum transmit power at the BS. At a transmission power of 40 dBm, the proposed MA algorithm and conventional robust design achieve substantial secrecy sum rates of 31.3[3] and 20.6 under imperfect CSI conditions, respectively. Meanwhile, the PPO and DDPG algorithms yield sum secrecy rates of approximately 28.8 and 26.09. We further compare the proposed imperfect CSI design with perfect CSI scenarios. Under perfect CSI assumptions, the MA, PPO, DDPG, and Con schemes achieve enhanced sum secrecy rates of 32.2, 30.02, 27.5, and 23.5, respectively. These findings highlight the effectiveness of the MA approach within our model, showcasing its superior performance compared to other considered DRL schemes like PPO and DDPG frameworks. Additionally, this comparison underscores the consistent superiority of the MA-DRL framework over conventional robust designs, highlighting its inherent benefits and notably improved performance within the considered system.

Fig. 7 shows how varying STAR-RIS element numbers impact the overall sum secrecy rate for both the proposed MA-DRL framework and conventional robust optimization, comparing them with passive-RIS, random phase settings, and No-RIS scenarios. Overall, increasing RIS elements improves the sum secrecy rate except in the No-RIS case. In the conventional robust optimization scenario, employing 16 RIS elements in Con-STAR-RIS yields a sum secrecy rate of 21.7, surpassing passive RIS at 17.07. Meanwhile, random RIS phase initialization results in 15.05, while the absence of RIS reaches 11.8. Similarly, in the MA-DRL framework with 16 RIS elements, MA-STAR-RIS achieves a sum secrecy rate of 28.3, outperforming passive RIS at 20.9. Random RIS phase initialization leads to 18.6, while the absence of RIS results in 14.7, a pivotal role of RIS implementation, particularly STAR-RIS, in enhancing performance. Therefore at $M = 16$, STAR-RIS demonstrates approximately 27.1% higher secrecy performance than passive RIS in the conventional case and around 35.4% higher performance in the MA-DRL case. This advantage stems from STAR-RIS offering spatial diversity and simultaneous transmission and reflection modes, amplifying SNR and secrecy rates for legitimate users while reducing these metrics for eavesdroppers. Moreover, we have also examined the performance of the phase-shift coupled STAR-RIS design, i.e., PC-MA-STAR-RIS and PC-Con-STAR-RIS with all other schemes. The results in Fig. 7 demonstrate that the consideration of phase-shift coupled design performs slightly inferior to the ideal ES mode and achieves a significant performance gain over the conventional RIS, as validated in existing works [71], [72].

Fig. 8 explores the impact of varying deployment positions of the STAR-RIS on network performance concerning the sum secrecy rate while identifying the optimal deployment location. It discusses the sum secrecy rate achieved by deploying a STAR-RIS and a Passive-RIS in different locations, specifically adjusting the distance from the STAR-RIS to the BS while maintaining fixed positions for the BS and users. For this particular simulation, the BS is positioned in the XY-plane at (70, 0), and two users (K= 2) are deployed at (0, 0) and (1, 0), respectively, assuming their proximity. The distance of the STAR-RIS is varied concerning the BS and users. The observed trend reveals that as the distance between the STAR-RIS and both the BS and users increases, there is a notable decrease in the sum secrecy rate. This decline is attributed to increased path loss resulting from greater distances, consequently degrading the sum secrecy rate, as confirmed in Fig. 8. Consequently, the findings suggest that the network demonstrates superior sum secrecy rate performance when the deployment positions of the STAR-RIS are nearer to either the BS or the users.

Fig. 9 illustrates the impact of minimum data rate ($R_{\min}$) constraints on the system's sum secrecy rate under different configurations. The results demonstrate that as $R_{\min}$ increases, the sum secrecy rate decreases. This trend is due to the stricter quality-of-service (QoS) requirements imposed by higher $R_{\min}$, which reduce the feasible region for the optimization problem. Specifically, a higher $R_{\min}$ forces the BS to allocate more power to users with poor channel conditions to meet their QoS demands. This reallocation reduces the power available for other users and compromises the system's overall secrecy performance. For instance, when $R_{\min}$ is low, the BS can optimize power allocation more freely, achieving a higher sum secrecy rate. However, as $R_{\min}$ increases, the sum secrecy rate decreases more rapidly, especially for users with poor channel conditions. The figure also compares the proposed MA-based scheme with a conventional optimization method. The MA-based scheme demonstrates superior performance in maintaining higher secrecy rates under varying $R_{\min}$, validating the efficacy of the proposed algorithm in handling stringent QoS constraints.

Fig. 10 illustrates the system performance of STAR-RIS,

[3]The sum secrecy rate is expressed in bits/s/Hz. However, for brevity, the unit of measure is omitted in the text.

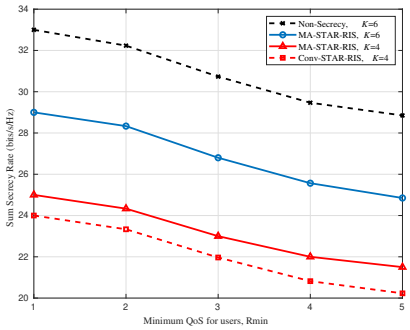

Fig. 9: Impact of $R_{\min}$.

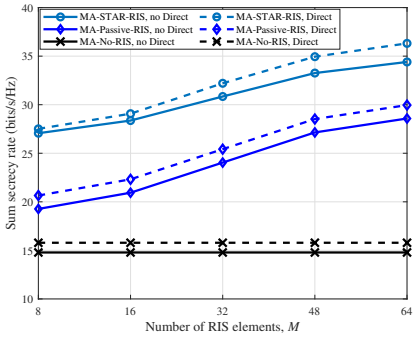

Fig. 10: Impact of Weak Direct Links.

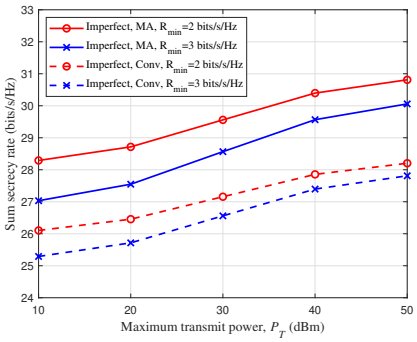

Fig. 11: Impact of power allocation and R_{min}.

passive RIS, and no RIS configurations, with and without direct links between the BS and users/eavesdroppers. Since users are far from the BS, they primarily rely on STAR-RIS-assisted links. Our analysis shows minimal performance variation between scenarios with weak direct links and no direct links, indicating that direct links have negligible impact. This underscores the significant role of STAR-RIS in enhancing system performance. Regardless of the presence of direct links, STAR-RIS consistently improves efficiency and robustness, demonstrating that the majority of performance gains are thanks to STAR-RIS, not the marginal influence of direct links.

Fig. 11 illustrates the relationship between maximum transmit power and sum secrecy rate under different $R_{\min}$ constraints and system conditions. Higher transmit power enhances secrecy rates, but stricter QoS constraints ($R_{\min} = 3$ vs. $R_{\min} = 2$) limit performance due to reduced power allocation flexibility. The analysis also examines the impact of imperfect system conditions, where inaccuracies in parameters like channel state information degrade secrecy rate performance. However, the MA-based scheme consistently outperforms the conventional method, demonstrating its robustness and adaptability in maintaining higher secrecy rates across varying transmit power levels and $R_{\min}$ constraints, even under imperfections.

VI. Conclusions

In this paper, we explored the realm of multi-user STAR-RIS-assisted dl communication with a primary focus on maximizing information secrecy. We tackled the worst-case robust beamforming design problem to maximize the overall system sum secrecy rate while considering constraints related to transmit power limitations, specified QoS requirements, and practical constraints on the STAR-RIS phase shifter array. To address the non-convex problem, we employed the S-procedure within the AO framework, incorporating a line search to iteratively update the precoder and phase shift matrix. Further, we extended our solution by utilizing an MA-DRL framework based on MDP to tackle non-convexity. We also analyzed practical phase shifts and the effect of direct links to showcase the practicality of our approach. Simulation results underscore the significant advantage of STAR-RIS, showing approximately 27.1% higher secrecy in conventional optimization and about 35.4% in the MA-DRL context compared to the conventional RIS. Moreover, our MA-DRL approach outperforms single-agent schemes by approximately 8.6% of PPO and 19.9% of DDPG, highlighting the substantial benefits of coupling of the proposed framework.

References

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

[2] C. Pan *et.al.*, "Reconfigurable intelligent surfaces for 6G systems: Principles, applications, and research directions," *IEEE Commun. Mag.*, vol. 59, no. 6, pp. 14–20, Jun. 2021.

[3] S. Pala, M. Katwe, K. Singh, B. Clerckx, and C.-P. Li, "Spectral-efficient RIS-aided RSMA URLLC: Toward mobile broadband reliable low latency communication (mBRLLC) system," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2023.

[4] S. Pala *et.al*, "Joint optimization of URLLC parameters and beamforming design for multi-RIS-aided MU-MISO URLLC system," *IEEE Wireless Commun. Lett.*, vol. 12, no. 1, pp. 148–152, Jan. 2023.

[5] N. Yang *et.al*, "Safeguarding 5G wireless communication networks using physical layer security," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 20–27, Apr. 2015.

[6] F. Tariq, M. R. A. Khandaker, K.-K. Wong, M. A. Imran, M. Bennis, and M. Debbah, "A speculative study on 6G," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 118–125, Aug. 2020.

[7] Y. Liu *et.al*, "STAR: Simultaneous transmission and reflection for 360° coverage by intelligent surfaces," *IEEE Wireless Commun.*, vol. 28, no. 6, pp. 102–109, Dec. 2021.

[8] M. Ahmed, A. Wahid, S. S. Laique, W. U. Khan, A. Ihsan, F. Xu, S. Chatzinotas, and Z. Han, "A survey on STAR-RIS: Use cases, recent advances, and future research challenges," *IEEE Internet Things J.*, vol. 10, no. 16, pp. 14 689–14 711, Aug. 2023.

[9] D. Yang *et.al*, "Secure communication for spatially correlated RIS-aided multiuser massive MIMO systems: Analysis and optimization," *IEEE Commun. Lett.*, vol. 27, no. 3, pp. 797–801, Mar. 2023.

[10] S. Arzykulov *et.al*, "Artificial noise and RIS-aided physical layer security: Optimal RIS partitioning and power control," *IEEE Wireless Commun. Lett.*, vol. 12, no. 6, pp. 992–996, Jun. 2023.

[11] S. Hong *et al.*, "Artificial-noise-aided secure MIMO wireless communications via intelligent reflecting surface," *IEEE Trans. Commun.*, vol. 68, no. 12, pp. 7851–7866, Dec. 2020.

[12] S. Pala, O. Taghizadeh, M. Katwe, K. Singh, C.-P. Li, and A. Schmeink, "Secure RIS-assisted hybrid beamforming design with low-resolution phase shifters," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2024.

[13] K. Cumanan *et.al*, "Secrecy rate optimizations for a MIMO secrecy channel with a multiple-antenna eavesdropper," *IEEE Trans. Veh. Technol.*, vol. 63, no. 4, pp. 1678–1690, May 2014.

[14] L. Liu *et.al*, "Secrecy wireless information and power transfer with MISO beamforming," *IEEE Trans. Signal Process.*, vol. 62, no. 7, pp. 1850–1863, Apr. 2014.

[15] Q. Shi *et.al*, "Secure beamforming for MIMO broadcasting with wireless information and power transfer," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2841–2853, May 2015.

[16] X. Liu *et al.*, "Secrecy throughput optimization for the WPCNs with non-linear EH model," *IEEE Access*, vol. 7, pp. 59 477–59 490, 2019.

[17] J. Wang, J. Zhang, J. Lu, J. Wang, Q. Zhang, and D. Wang, "Secrecy rate analysis for RIS-aided multi-user MISO system over rician fading channel," *J. Commun. Inf. Netw.*, vol. 8, no. 1, pp. 48–56, Mar. 2023.

[18] S. Lin, Y. Xu, H. Wang, J. Gu, J. Liu, and G. Ding, "Secure multicast communications via RIS against eavesdropping and jamming with imperfect CSI," *IEEE Trans. Veh. Technol.*, pp. 1–6, 2023.

[19] L. Dong, Y. Li, W. Cheng, and Y. Huo, "Robust and secure transmission over active reconfigurable intelligent surface aided multi-user system," *IEEE Trans. Veh. Technol.*, vol. 72, no. 9, pp. 11 515–11 531, Sep. 2023.

[20] J. Liu *et al.*, "Secrecy rate analysis for reconfigurable intelligent surface-assisted mimo communications with statistical CSI," *China Commun.*, vol. 18, no. 3, pp. 52–62, Mar. 2021.

[21] Z. Peng, R. Weng, C. Pan, G. Zhou, M. D. Renzo, and A. L. Swindlehurst, "Robust transmission design for RIS-assisted secure multiuser communication systems in the presence of hardware impairments," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 7506–7521, Nov. 2023.

[22] S. Hong *et al.*, "Robust transmission design for intelligent reflecting surface-aided secure communication systems with imperfect cascaded CSI," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2487–2501, Apr. 2021.

[23] H. Niu, Z. Chu, F. Zhou, P. Xiao, and N. Al-Dhahir, "Weighted sum rate optimization for STAR-RIS-assisted MIMO system," *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 2122–2127, Feb. 2022.

[24] C. Wu, C. You, Y. Liu, X. Gu, and Y. Cai, "Channel estimation for STAR-RIS-aided wireless communication," *IEEE Commun. Lett.*, vol. 26, no. 3, pp. 652–656, Mar. 2022.

[25] C. Wu, Y. Liu, X. Mu, X. Gu, and O. A. Dobre, "Coverage characterization of STAR-RIS networks: NOMA and OMA," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 3036–3040, Sep. 2021.

[26] Z. Zhang, Z. Wang, Y. Liu, B. He, L. Lv, and J. Chen, "Security enhancement for coupled phase-shift STAR-RIS networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 6, pp. 8210–8215, Jun. 2023.

[27] H. Niu *et al.*, "Simultaneous transmission and reflection reconfigurable intelligent surface assisted secrecy MISO networks," *IEEE Commun. Lett.*, vol. 25, no. 11, pp. 3498–3502, Nov. 2021.

[28] 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 Trans. Wireless Commun.*, vol. 21, no. 12, pp. 11 207–11 221, Dec. 2022.

[29] X. Li, Y. Zheng, M. Zeng, Y. Liu, and O. A. Dobre, "Enhancing secrecy performance for STAR-RIS NOMA networks," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 2684–2688, Feb. 2023.

[30] W. Ni, Y. Liu, Y. C. Eldar, Z. Yang, and H. Tian, "STAR-RIS integrated nonorthogonal multiple access and over-the-air federated learning: Framework, analysis, and optimization," *IEEE Internet Things J.*, vol. 9, no. 18, pp. 17 136–17 156, Sep. 2022.

[31] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao, and Q. Wu, "Deep reinforcement learning-based intelligent reflecting surface for secure wireless communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 375–388, Jan. 2021.

[32] H. Yang, S. Liu, L. Xiao, Y. Zhang, Z. Xiong, and W. Zhuang, "Learning-based reliable and secure transmission for UAV-RIS-assisted communication systems," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2023.

[33] J. Xu *et.al*, "Deep reinforcement learning for RIS-aided secure mobile edge computing in industrial internet of things," *IEEE Trans. Ind. Informat.*, pp. 1–10, 2023.

[34] Z. Zhu *et.al*, "DRL-based STAR-RIS-assisted ISAC secure communications," in *Proc. Ucom*, Jul. 2023, pp. 127–132.

[35] T. Zhou, K. Xu, G. Hu, X. Xia, W. Xie, and C. Li, "Robust beamforming design for STAR-RIS-assisted anti-jamming and secure transmission," *IEEE Trans. Green Commun. Netw.*, pp. 1–1, 2023.

[36] Y. Wen, G. Chen, S. Fang, Z. Chu, P. Xiao, and R. Tafazolli, "STAR-RIS-assisted-full-duplex jamming design for secure wireless communications system," *arXiv preprint arXiv:2309.04566*, 2023.

[37] C. Wu *et.al*, "Resource allocation in STAR-RIS-aided networks: OMA and NOMA," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7653–7667, Sep. 2022.

[38] J. Xu *et.al*, "STAR-RISs: Simultaneous transmitting and reflecting reconfigurable intelligent surfaces," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 3134–3138, Sep. 2021.

[39] Y. Liu *et.al*, "STAR: Simultaneous transmission and reflection for 360° coverage by intelligent surfaces," *IEEE Wireless Commun.*, vol. 28, no. 6, pp. 102–109, 2021.

[40] C. Wu, Y. Liu, X. Mu, X. Gu, and O. A. Dobre, "Coverage characterization of STAR-RIS networks: NOMA and OMA," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 3036–3040, 2021.

[41] J. Xu, Y. Liu, X. Mu, J. T. Zhou, L. Song, H. V. Poor, and L. Hanzo, "Simultaneously transmitting and reflecting (STAR) intelligent omni-surfaces, their modeling and implementation," *arXiv preprint arXiv:2108.06233*, 2021.

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

[43] B. O. Zhu, K. Chen, N. Jia, L. Sun, J. Zhao, T. Jiang, and Y. Feng, "Dynamic control of electromagnetic wave propagation with the equivalent principle inspired tunable metasurface," *Scientific reports*, vol. 4, no. 1, p. 4971, 2014.

[44] M. Katwe, K. Singh, B. Clerckx, and C.-P. Li, "Improved spectral efficiency in STAR-RIS aided uplink communication using rate splitting multiple access," *IEEE Trans. Wireless Commun.*, 2023.

[45] Y. Zhang *et.al*, "Distributed optimal beamformers for cognitive radios robust to channel uncertainties," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6495–6508, Dec. 2012.

[46] X. Yu *et.al*, "Robust and secure wireless communications via intelligent reflecting surfaces," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2637–2652, Nov. 2020.

[47] O.Taghizadeh *et.al.*, "Private uplink communication in C-RAN with untrusted radios," *IEEE Trans. Veh. Technol.*, vol. 69, no. 7, pp. 8034–8039, Jul. 2020.

[48] ——, "Quantization-aided secrecy: FD C-RAN communications with untrusted radios," *IEEE Trans. Wireless Commun.*, vol. 21, no. 10, pp. 8522–8535, Oct. 2022.

[49] O. Taghizadeh *et.al*, "Secrecy energy efficiency of MIMOME wiretap channels with full-duplex jamming," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5588–5603, Aug. 2019.

[50] T. Lipp and S. Boyd, "Variations and extension of the convex–concave procedure," *Optim. Eng.*, Jun. 2016.

[51] G. Zhou *et.al*, "A framework of robust transmission design for IRS-aided MISO communications with imperfect cascaded channels," *IEEE Trans. Signal Process.*, vol. 68, pp. 5092–5106, 2020.

[52] Q. Li and W.-K. Ma, "Spatially selective artificial-noise aided transmit optimization for MISO multi-eves secrecy rate maximization," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2704–2717, May 2013.

[53] T. Lipp and S. Boyd, "Variations and extension of the convex–concave procedure," *Optimization and Engineering*, vol. 17, pp. 263–287, 2016.

[54] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Operations research*, vol. 26, no. 4, pp. 681–683, 1978.

[55] M. Everett, B. Lütjens, and J. P. How, "Certifiable robustness to adversarial state uncertainty in deep reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4184–4198, Sep. 2022.

[56] T. Li *et.al*, "Applications of multi-agent reinforcement learning in future internet: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 1240–1279, Secondquarter 2022.

[57] Z. Peng *et.al*, "Deep reinforcement learning for RIS-aided multiuser full-duplex secure communications with hardware impairments," *IEEE Internet Things J.*, vol. 9, no. 21, pp. 21 121–21 135, Nov. 2022.

[58] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017.

[59] M. Holzleitner *et.al*, "Convergence proof for actor-critic methods applied to PPO and RUDDER," in *Transactions on Large-Scale Data-and Knowledge-Centered Systems XLVIII*. Springer, 2021, pp. 105–130.

[60] G.-B. Huang, "Learning capability and storage capacity of two-hidden-layer feedforward networks," *IEEE Trans. Neural Netw.*, vol. 14, no. 2, pp. 274–281, Mar. 2003.

[61] 3GPP, "Technical specification group radio access network; study on 3d channel model for LTE (release 12)," TR 36.873 V12.7.0, Dec. 2017.

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

[63] L. N. Smith, "A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay," *arXiv preprint arXiv:1803.09820*, 2018.

[64] ——, "Cyclical learning rates for training neural networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Mar. 2017, pp. 464–472.

[65] Y. Dai, N. Fei, and Z. Lu, "Improvable gap balancing for multi-task learning," in *Uncertainty in Artificial Intelligence*. PMLR, 2023, pp. 496–506.

[66] Z. Sheng, H. D. Tuan, T. Q. Duong, and H. V. Poor, "Beamforming optimization for physical layer security in MISO wireless networks," *IEEE Trans. Sig. Process.*, vol. 66, no. 14, pp. 3710–3723, Jul. 2018.

[67] C. Liu, J. Zhou, Y. Gao, D. Qiao, and H. Qian, "IRS-aided secure communications over an untrusted af relay system," *IEEE Trans. Wireless Commun.*, pp. 1–1, 2023.

[68] P. Saikia *et.al*, "Proximal policy optimization for RIS-assisted full duplex 6G-V2X communications," *IEEE Trans. Intell. Veh.*, pp. 1–16, 2023.

[69] L. Chai, L. Bai, T. Bai, J. Shi, and A. Nallanathan, "Secure RIS-aided MISO-NOMA system design in the presence of active eavesdropping," *IEEE Internet Things J.*, vol. 10, no. 22, pp. 19 479–19 494, Nov. 2023.

[70] H. Niu and X. Liang, "Weighted sum-rate maximization for STAR-RISs-aided networks with coupled phase-shifters," *IEEE Sys. J.*, vol. 17, no. 1, pp. 1083–1086, 2023.

[71] J. Xu *et.al*, "STAR-RISs: A correlated T&R phase-shift model and practical phase-shift configuration strategies," *IEEE J. Sel. Top. Signal Process.*, vol. 16, no. 5, pp. 1097–1111, 2022.

[72] Z. Wang *et.al*, "Coupled phase-shift STAR-RISs: A general optimization framework," *IEEE Wireless Communi. Lett.*, vol. 12, no. 2, pp. 207–211, 2022.