International
OPEN ACCESS Journal
Of Modern Engineering Research (IJMER)

Enhancing NOMA Network Security via RIS-UAV Systems: A DDPG Perspective

Syed Zain Ul Abideen[1*], Abdul Wahid[2], Mian Muhammad Kamal[3]
[1]College of Computer Science and Technology, Qingdao University, Qingdao 266071, China,
zain208shah@gmail.com
[2]College of Computer Science and Technology, Qingdao University, Qingdao 266071, China
wahidjan999@gmail.com
[3]Joint International Research Laboratory of Information Display and Visualization, School of Electronic Science and Engineering, Southeast University, Nanjing
210018, People's Republic of China
mmkamal@seu.edu.cn
[*]Corresponding author: zain208shah@gmail.com

Abstract— *In this work, we apply the deep deterministic policy gradient (DDPG) technique to improve the security of a non-orthogonal multiple access (NOMA) downlink network by enabling use of a reconfigurable intelligent surface (RIS) equipped unmanned aerial vehicles (UAV). Our main objective is to prevent eavesdroppers from accessing the network while preserving seamless communication for authorized users. The system is made up of a UAV integrated with an RIS which is essential for optimizing signal paths, and a Base Station. Our work aims to maximize secrecy rates for all users under possible eavesdropping scenarios by dynamically adjusting the RIS's phase shifts and power allocations. This method not only shows how flexible and successful the DDPG algorithm is at protecting wireless communications when used in conjunction with an RIS but it also highlights how much the algorithm has advanced secure communication systems.*

Keywords: *non-orthogonal multiple access, reconfigurable intelligent surface, unmanned aerial vehicles, deep reinforcement learning, deep deterministic policy gradient, Physical layer security.*

Date of Submission: 25-04-2024 Date of acceptance: 02-05-2024

I. INTRODUCTION

Next-level wireless communication systems, particularly those designed for the sixth generation (6G) networks, are expected to set new standards in connectivity characterized by high reliability, extremely fast data rates, and minimal latency. RIS and UAV assisted communication systems are identified as crucial components contributing to this significant advancement [1]. In such a scenario, NOMA is high- lighted as an impactful access method intended to enhance spectral efficiency and cater to the increasing demand for wireless connectivity [2]. RIS have emerged as a strategic approach to alter the propagation environment. Through the use of programmable meta surfaces containing numerous passive reflecting elements, RISs can manipulate the phase and amplitude of in- coming signals in order to tailor propagation paths according to specific communication needs [3]. The inclusion of this technology in UAV-assisted wireless networks is particularly promising due to its ability to provide a flexible and responsive solution for achieving comprehensive network coverage [4]. UAV introduce an additional level of dynamism to the telecommunications ecosystem. Their ability to be positioned and move in the air allows for rapid and specific deployment, making them extremely valuable tools for im- proving coverage, capacity, and addressing service gaps in ground-based networks [5]. When combined with RIS and aided by NOMA protocols [6], UAVs can further enhance

resource allocation efficiency and enhance user experiences throughout the network. The emergence of RIS as a cornerstone technology has marked a significant milestone in the progression of PLS, providing a dynamic and controllable domain for wireless communications [7]. RIS is architected from an assortment of passive components that are not only cost-efficient but also have the capacity for electromagnetic modulation, thanks to integrated PIN diodes [8]. With the strategic modulation of signal phases orchestrated by an advanced control system, RIS enhances signal integrity for approved receivers while concurrently disrupting those designated for unauthorized interceptors.

When compared to conventional PLS methodologies that utilize artificial noise [9] or advanced multi-antenna beamforming [10], RIS stands out for its passive operational nature, which bypasses the high costs associated with RF chains. Moreover, the flexible nature of RIS facilitates its straightforward incorporation into prevailing network systems and makes it ideal for attachment to a variety of structures within urban landscapes, as well as to wearable technology [11]. Nevertheless, the growing interconnections and complexity of systems also increase the risk of security breaches. The physical layer of wireless communication systems is particularly vulnerable to eavesdropping and other cyber threats, highlighting the need for new protective strategies [12]. In this respect RIS offer a valuable means to enhance the security of the physical layer. Through strategic manipulation of signal reflections [13], RISs can reduce signal reception at potential eavesdropper locations while strengthening it for authorized recipients, establishing a secure communication environment.

The field of industrial automation is experiencing a dynamic evolution due to the rise of innovative AI algorithm frameworks like RL, DL, and particularly DRL. These technologies are streamlining the path toward real- time automation and sustained advancement. DRL is at the forefront, distinguished by its capability to effectively process the intricate flow of data within communication systems and to navigate the complex management of system and resource control, often presented in non-linear and challenging non-convex scenarios. This level of adeptness is achieved even in the computationally rigorous task of de- ciphering and network formation for understanding wireless channels, conducted without reliance on established channel models or documented patterns of user movements [14]. Moreover, DRL’s strength lies in its strategic identification of optimal solutions for complex optimization issues, a skill honed by analyzing the patterns of rewards obtained from interactions in a wireless context, thus contributing vitally tothe advancement of cutting-edge algorithmic designs [11].

A. Related Work

RIS have rapidly emerged as a transformative technology within wireless networks gaining prominence in fields such as NOMA [15], [16], CoMP [17], [18], and UAV-based communications [19], [20], primarily due to their capacity to markedly enhance network functionalities. This development was made possible by the innovative research of [15], which provided an effective method for integrating RIS into NOMA systems while effectively managing the conflict between sum rate and power efficiency. This was achieved by carefully applying the SCA technique, which made it possible to improve beamforming and phase shifts on a periodic schedule.

Expanding on this research [16] extended the study to evaluate how RIS affects the performance of semi-grant- free NOMA systems, proposing algorithms that are adapted for various RIS setups. The investigation went on [17], who examined the role of RIS in CoMP communications, considering a spectrum of scenarios from ideal to less than ideal, and applying the dual close approach to optimize reflection coefficients. Simultaneously [18], focused on enhancing the long-term energy efficiency in STAR-RIS- facilitated CoMP networks by combining methods for active and passive beamforming optimization with a combination of partial programming and DRL approaches to achieve close to optimal results.

The discussion expands much more to include UAV communications, as [19], addressing the challenge of enhancing sum rates within RIS-supported multi-UAV NOMA frameworks. This entailed a holistic optimization strategy that encompassed UAV positioning, power management, RIS reflection matrix configurations, and the sequencing of NOMA decoding, all resolved through a BCD iterative methodology. Complementary to this [20], pioneered a novel DRRL algorithm designed to optimize both UAV flight patterns and beamforming processes active at the UAV and passive at the STAR-RIS simultaneously thus showcasing the expansive utility and significant advantages of RIS in advancing the capabilities of contemporary wireless communication systems.

Recent research incorporating RIS has significantly advanced the topic of Physical Layer Security (PLS). Important advancements include the [21], novel RIS configuration that protects downlink NOMA systems from attackers and the [22], method that enhances beamforming in secure wireless systems that use RIS. Robust optimization strategies were proposed by [23], to address the problem of imperfect eavesdropper CSI. By combining STAR-RIS with NOMA in a novel way. [24], improved security by creating artificial noise [25], explored the use of RIS to optimize secure energy efficiency in the context of UAVs and [26], used aerial RIS to adjust signal distributions for NOMA systems in order to improve secrecy rates, highlighting the revolutionary effect of RIS on improving wireless communication security.

B. Motivation and Contribution

This work is motivated by the critical role that RIS play in the development of 6G wireless communication technology. We are at the beginning of a new age in wireless communications, and with it comes a growing need for improved energy efficiency, cost reduction and spectrum utilization. RIS presents a fresh solution to these problems because of its capacity to reorganize electromagnetic wave propagation especially in situations where direct communication routes are restricted.

Moreover, the escalating concerns regarding security breaches in wireless communications highlight the urgency of integrating robust PLS measures. RIS technology presents an innovative way to secure data transmission against illegal access and eavesdropping, while simultaneously enhancing service quality through innovative non-line-of-sight connections. By carefully modifying RIS elements one can boost QoS and strengthen security while simultaneously leveraging the unique spatial selectivity that comes with RIS to optimize signal confidentiality for authorized users and restrict it for potential eavesdroppers.

• The integration of RIS and NOMA along with the potential integration with UAVs highlight the revolutionary possibilities of these technologies in establishing a more adaptable, effective and safe wireless communication environment. The potential to leverage these state-of-the-art technologies to tackle the challenges of dynamic multi-user environments where traditional optimization techniques fall short because of their non- linear nature is what drives this research.

• Adaptability and Resource Efficiency: Our work employs the DDPG algorithm to simultaneously optimize power distribution, phase shifting, and UAV positioning allowing for dynamic adaptation to changing channel conditions and user requirements. With the help of this comprehensive optimization approach the system is able to efficiently allocate wireless resources in real-time maximizing spectrum efficiency and ensuring maximum performance in a variety of operational conditions. Moreover, the system may constantly im- prove its approaches responding to changing network dynamics and improving overall utilization of resources in wireless communication environments through using AI-driven learning algorithms.

• Securing Wireless Communication: Our work presents distinctive PLS approaches that attempt to enhance the system's security measures against potential eaves- dropping threats in addition to optimizing system performance. Our solution strengthens the security and integrity of wireless transmissions by incorporating PLS mechanisms like secure beamforming and trans- mission techniques into the optimization framework. This reduces the possibility of unauthorized access and information interception. The practical relevance and significance of our research in protecting wireless communication systems from malicious actors is high- lighted by the need of this proactive security approach which is especially important in wireless communication networks handling confidential information or operating in unsafe environments.

• Our aims are to demonstrate the efficacy of the DDPG- based strategy through comprehensive computer simulations demonstrating its ability to strengthen the physical layer against security attacks and improve overall network performance. The goal of this research is to fully utilize RIS, NOMA, and UAV technologies in order to influence the development of secure and efficient wireless communication networks in the future.

SYSTEM MODEL AND PROBLEM FORMULATION

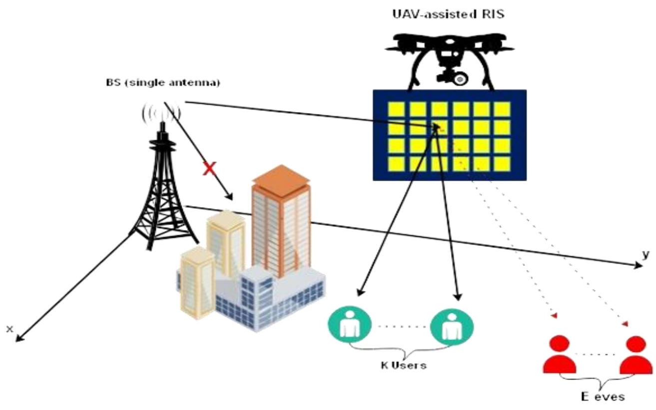

Figure 1: System Model for RIS-UAV Enhanced Security Architecture for NOMA Networks

In a complex communication system that utilizes UAV-assisted NOMA to enhance the confidentiality of the physical layer, the structure consists of a Base Transmission Station responsible for sending data to a set of K end-users represented as {1, 2, ..., K}. A UAV with an RIS comprising N passive modulators aids in reflecting and directing signals for this data relay. The UAV operates independently at a predetermined altitude above the specified area {A}, commencing its mission from an established charging point. The system is designed for quasi-static frequency flat-fading channels assuming perfect Channel State Information availability at both the BTS and the UAV-mounted IRS. Energy consumption and operational duration considerations are abstracted away for simplicity. This sophisticated setup is vulnerable to potential covert surveillance attempts by a group of eavesdroppers denoted as {E} = {1, 2, ..., E},
aiming to illicitly intercept communication. Notably, due to utilizing the DDPG method, the proposed algorithm can adjust to changing channel conditions across different time slots while maintaining consistency within each individual time slot ensuring resilience and dependability in dynamic communication environments. The linkage between the BTS and the RIS is characterized by $G \in \mathbb{C}^{N \times 1}$, representing the propagation path of the transmitted signal. Concurrently, the communication channels from the RIS to each k-th end-user and e-th eavesdropper are denoted by $h_{r,k} \in \mathbb{C}^{N \times 1}$ and $h_{r,e} \in \mathbb{C}^{N \times 1}$, respectively, capturing the nuances of signal reflection and potential attenuation or enhancement due to the RIS's modulation capabilities.

The reception at each k-th end-user is mathematically modeled as:

$$y_k = \left(\mathbf{h}_{r,k}^H \mathbf{\Phi} \mathbf{G}\right) \sum_{i=1}^{K} \sqrt{p_i} s_i + n_k,\ k \in \mathcal{K}\ (1)$$

where $\Phi = \mathrm{diag}\left(e^{j\theta_1}, \dots, e^{j\theta_N}\right)$ encapsulates the RIS's phase shift capabilities. p_i represents the power allocated to the i-th user's signal s_I, and n_k symbolizes the additive white Gaussian noise, inherent in wireless communication, at the k-th end-user's receiver. ρ_i denotes the BS's power allocation coefficient for the i-th user, constrained within the interval [0,1]. The sum of these coefficients across all K users equals 1. The transmitted signal for the i-th user is represented by s_I, designed such that its expected power equals 1, denoted as $E[s_i^2] = 1$. The noise affecting the k-th user's signal, denoted as n_k, follows a complex normal distribution with zero mean and variance σ^2. The RIS, positioned on a UAV, is located at v(x, y) with a height h_I, while the BS is at the origin (0,0) with height h_B. The horizontal position of each k-th user is given by $u_k(x_k,y_k)$,
The distance between the BS and RIS is calculated as

$$d_{BI} = \sqrt{x^2 + y^2 + (h_B - h_I)^2}$$

and the distance between the RIS and the $k-th$ user is
$d_{Iu_k} = \sqrt{(x - x_k)^2 + (y - y_k)^2 + h_I^2}$.
Surveillance by the $e-th$ eavesdropper yields the intercepted signal as:

$$y_e = \left(\mathbf{h}_{r,e}^H \mathbf{\Phi} \mathbf{G}\right) \sum_{i=1}^{K} \sqrt{p_i} s_i + n_e,\ e \in \mathcal{E}\ (2)$$

For each legitimate user (k), the channel gain considering the path loss can be expressed as:

$$\boldsymbol{h}_k = \frac{\boldsymbol{h}_{rk}^H \mathbf{\Phi} \boldsymbol{g}}{(d_{BI} d_{Iu_k})^{\alpha}}\ (3)$$

where α is the path loss coefficient, and d_{BI} and d_{Iu_k} represent the distances from the base station (BS) to the RIS and from the RIS to the k user respectively.
To incorporate Eve into this model, we can extend the scenario to include the channel vector from the RIS to Eve, h_{re}, and the distance from the RIS to Eve, d_{Ie}. The channel gain for Eve would then be similar to that of the legitimate users, adjusted for Eve's position:

$$\boldsymbol{h}_e = \frac{\boldsymbol{h}_{re}^H \mathbf{\Phi} \boldsymbol{g}}{(d_{BI} d_{Ie})^{\alpha}}\ (4)$$

In the SINR calculations for implementing Successive Interference Cancellation (SIC) among NOMA users, it's important to account for the potential interception by Eve. The SINR for a user *j* decoding the signal intended for a weaker user *t* is given by:

$$\mathrm{SINR}_{t\to j} = \frac{|\boldsymbol{h}_j|^2 P_{\max}\rho_t}{\sum_{i=t+1}^{K} |\boldsymbol{h}_j|^2 P_{\max}\rho_i + \sigma^2} \quad (5)$$

For Eve attempting to decode the signal intended for user *t* the SINR would be:

$$\mathrm{SINR}_{t\to e} = \frac{|\boldsymbol{h}_e|^2 P_{\max}\rho_t}{\sum_{i=t+1}^{K} |\boldsymbol{h}_e|^2 P_{\max}\rho_i + \sigma^2} \quad (6)$$

This formulation allows the system to evaluate the risk posed by Eve's interception attempts and adjust the power allocation p_i and phase shifts θ_n accordingly to ensure secure communication. To ensure effective SIC and maintain the system's security, the data rates for the legitimate users must be optimized to maximize the difference between their SINRs and Eve's SINR, effectively increasing the secrecy rate and making the system robust against eavesdropping. To ensure physical layer confidentiality, the covert communication rate or secrecy rate for each $\mathrm{k-th}$ user, which quantifies the secure information transmission rate, is defined as:

$$R_{s,k} = \left[R_k - \max_{e\in\mathcal{E}} R_{e,k}\right]^+ \quad (7)$$

where R_k is the legitimate communication rate to the k-th user, and $R_{e,k}$ is the potential information rate accessible to the e-th eavesdropper. These rates are articulated by:

$$R_k = \log_2\left(1 + \frac{|\mathbf{h}_{r,k}^H \boldsymbol{\Phi}\mathbf{G}|^2 p_k}{\sigma^2 + \sum_{i\neq k} |\mathbf{h}_{r,k}^H \boldsymbol{\Phi}\mathbf{G}|^2 p_i}\right) \quad (8)$$

$$R_{e,k} = \log_2\left(1 + \frac{|\mathbf{h}_{r,e}^H \boldsymbol{\Phi}\mathbf{G}|^2 p_k}{\sigma^2 + \sum_{i\neq k} |\mathbf{h}_{r,e}^H \boldsymbol{\Phi}\mathbf{G}|^2 p_i}\right) \quad (9)$$

effectively capturing the dynamics of secure and potentially compromised communication paths. The principal objective in this advanced communication paradigm is to maximize the sum of all users' secrecy rates, which is pivotal for ensuring robust secure communication against eavesdropping threats.

Problem formulation

In our comprehensive RIS-UAV-NOMA downlink network model, focused on enhancing PLS alongside communication efficiency, our objective extends to not only maximizing the sum rate but also incorporating an aspect of security by optimizing several key parameters. These parameters include the power allocation $\boldsymbol{\Phi}$ at the Base Station (BS), the phase-shifting p_i of the RIS, and the horizontal positioning $\mathrm{v(x,y)}$ of the UAV. The integration of these aspects leads to a complex optimization problem that can be articulated as follows:

$$
\begin{aligned}
&\max_{\{\Theta,\rho,v\}} \sum_{t=1}^{K} R_{s,t} \quad (a)\\
\text{s.t.}\quad & R_{s,t} \geq R_{\min}, \forall t \in \mathcal{K} \quad (b)\\
& R_{t\to j} \geq R_{t\to t}, \forall t, j \in \mathcal{K}, t > j \quad (c)\\
& \sum_{k=1}^{K} \rho_k \leq 1 \quad (d)\\
& v(x,y) \in \mathcal{A} \quad (e)\\
& 0 \leq \theta_n \leq 2\pi, n = 1, \cdots, N \quad (f)
\end{aligned}
$$

In this formulation, $R_{s,t}$ represents the secrecy rate for user *t*, enhancing the PLS by considering the potential eavesdropping threats. Constraint (b) ensures the Quality of Service (QoS) for all users by guaranteeing that the secrecy rate for each user *t* is above a predefined minimum R_{min}. Constraint (c) is pivotal for the implementation of Successive Interference Cancellation (SIC), ensuring that the decoding process for NOMA can be executed effectively. The constraint (d) encapsulates the total transmission power limitation at the BS, ensuring efficient power usage. Constraint (e) specifies the operational area for the UAV, ensuring it remains within a designated feasible region A. Lastly, constraint (f) governs the phase shifts applied by the RIS, with each element θ_n confined within a 0 to 2π range. Given the non-convex nature of this optimization problem, primarily due to the intricate interplay between the variables $\{\theta, p, v\}$, finding a global optimal solution presents significant challenges. To navigate these complexities, we propose a sophisticated and efficient solution framework based on Deep Reinforcement Learning (DRL), specifically utilizing DDPG. This approach is designed to tackle the non-convexity and high dimensionality inherent in the problem, offering a robust and low complexity method to achieve near-optimal solutions, thereby ensuring an enhanced and secure communication network.

Proposed DDPG

The use of the DDPG algorithm is essential in the setting of the integrated RIS-equipped UAV-assisted NOMA downlink system with a focus on improving PLS [27]. This section begins with a brief overview of DDPG and then goes into great detail into how the DDPG framework was modified to fit the given optimization challenge.

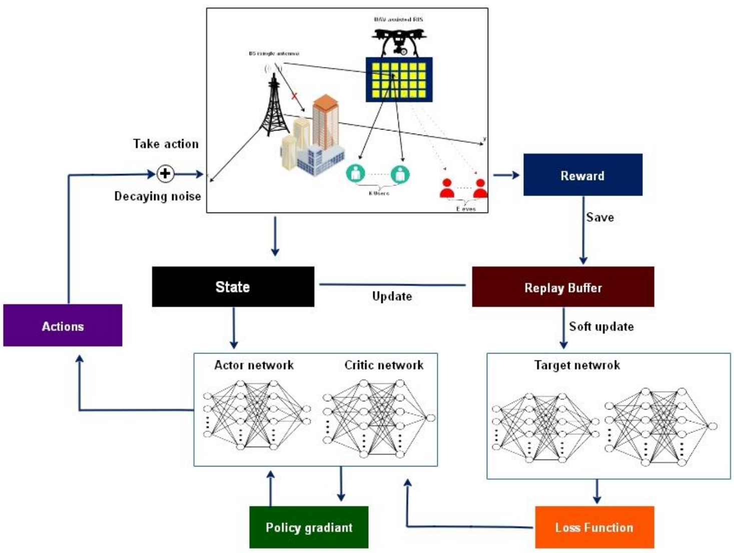

Figure 2: DDPG Diagram for RIS-UAV Enhanced Security Architecture for NOMA Networks

Introduction to DDPG : Since conventional algorithms such Deep Q-Networks (DQN), are mostly designed to operate in discrete action spaces they are constrained in their application to continuous action spaces. In order to solve our stated problem, this constraint compels us to look into alternative strategies such DDPG. Using deep function approximators, DDPG which is characterized by its model free off-policy actor-critic mechanism efficiently explores the high-dimensional continuous action space [28]. The deterministic policy gradient technique which is a fundamental component of DDPG maps a particular action α systematically and assesses its effectiveness using a Q function, $Q(s,a,\theta^q)$ where θ^q stands for the parameters of the critic network. Optimizing the output Q value is the main goal of DDPG, which also improves system performance overall. Experience replay is a key component of DDPG, since it solves the problem of non-i.i.d. (independently and identically distributed) data that result from environment exploration in sequence [29]. To maintain stability and reduce variation in the learning process this is combined with the use of soft update approaches. The soft update technique uses the following equation to gradually adjust the target network parameters θ`, to match the learned evaluation network:

$$\theta' \leftarrow \tau\theta + (1-\tau)\theta' \quad (10)$$

Where τ is significantly less than 1, ensuring a conservative update approach. Moreover the investigation in the continuous action space presents an important challenge that

DDPG successfully resolves by introducing noise N into the policy. The following equation describes this approach:

$$\mu'(s_t) = \mu(s_t;\theta^{\mu}) + \mathcal{N} \quad (11)$$

where N is the environment dependent noise and $\mu'(s_t)$ is the target policy with extra noise encouraging exploration and helping in the identification of optimal policies. This comprehensive approach makes use of DDPG's capabilities to provide a solid basis for addressing the challenges of PLS optimization in the RIS-UAV-NOMA downlink system ensuring effective and secure network connectivity [30]. After introducing the DDPG algorithm and implementing it in a secure RIS-equipped UAV-assisted NOMA downlink communication system we then explore the details of processing Deep Reinforcement Learning (DRL) in this complicated environment.

The DRL Processing: In our advanced RIS-enhanced UAV-supported NOMA downlink system, the dynamic nature of wireless channels constitutes the environment in which our system operates. The collaborative setup of the RIS coupled with an UAV act as the central agent within this scenario, tasked with the dual objectives of maximizing transmission efficiency and ensuring robust security at the physical layer. Our deep reinforcement learning (DRL) framework is crafted to mirror the unique features and goals of our setup through several key elements:

• State Space Definition

For each discrete interval, labeled as t, the state space is meticulously defined to capture an exhaustive view of the system's present condition, incorporating aspects such as:

Prior timestep secrecy metrics ($R_{s,k}^{(t-1)}$) for individual users k, underscoring the emphasis on safeguarding data confidentiality.

Last known RIS modulation phases (θ) and power distribution profiles (ρ), key parameters for optimizing signal propagation and energy consumption.

The UAV's most recent positioning coordinates (x and y), vital for optimal aerial relay positioning and interference management. This state is mathematically captured as:

$$s_t = \Big[R_1^{(t-1)}, \cdots, R_K^{(t-1)}, R_{s,1}^{(t-1)}, \cdots, R_{s,K}^{(t-1)}, \theta_1^{(t-1)}, \cdots, \theta_{2N}^{(t-1)}, \rho_1^{(t-1)}, \cdots, \rho_k^{(t-1)}, x^{(t-1)}, y^{(t-1)}\Big] \quad (12)$$

• Action Space Configuration

Aligned with the state s_t, the action set a_t includes a suite of strategic decisions enacted by the RIS-UAV duo, designed to navigate the intricacies of the prevailing wireless communication landscape and user requirements. These actions encompass:

Fine-tuning of RIS reflective properties (θ) to enhance signal directionality and strength.

Strategic adjustments in transmission power allocations (ρ) tousers, ensuring optimal energy utilization.

Real-time repositioning of the UAV (x, y) to maintain superiorsignal quality and reduce potential interference.

$$a_t = \Big[\theta_1^{(t)}, \cdots, \theta_{2N}^{(t)}, \rho_1^{(t)}, \cdots, \rho_k^{(t)}, x^{(t)}, y^{(t)}\Big] \quad (13)$$

• Reward Function

At each interval t, the reward metric r_t is ingeniously formulated to resonate with the system's overarching aims of enhancing throughput and reinforcing data security across the wireless network. It amalgamates the total system throughput and the collective secrecy performance across all users:

$$r_t = \alpha \sum_{k=1}^{K} R_k^{(t)} + \beta \sum_{k=1}^{K} R_{s,k}^{(t)} \quad (14)$$

In this formulation, α and β are coefficients that calibrate the focus between operational efficiency and security enhancement, facilitating a balanced optimizationapproach.

Through this nuanced DRL processing methodology, our RIS-UAV framework adeptly learns to traverse the dynamic terrain of our NOMA downlink communication system. By judiciously making decisions that not only propel the throughput but also significantly elevatethe system's security posture, this adaptive learning paradigm, rooted in DDPG methodologies, marks a significant leap towards achieving secure, high-performing,and adaptive wireless communication networks.

Processing the Formulated Problem: In optimizing our sophisticated RIS-aided UAV-based NOMA downlink architecture, the DRL strategy is finely tuned to respect the system's Quality of Service (QoS) benchmarks and other operational intricacies. Here's how the DRL scheme, particularly leveraging the DDPG algorithm, methodically addresses the stipulated problem:

- Initial Data Rate Validation for QoS Alignment: At every decision point, marked as timestep *t*, the DRL algorithm diligently computes the transmission rate $R^{(t)}$ for every user in the setup. This step is pivotal for confirming adherence to predefined QoS criteria, a fundamental aspect delineated in our problem's constraint (b). Actions leading to satisfactory QoS fulfillment are stored within the model's memory buffer, reinforcing optimal behavior. In contrast, actions resulting in QoS breaches are either penalized or discarded to prevent recurrence, steering the DRL agent towards more effectivestrategies.
- Adaptive SIC Procedure Refinement: Navigating the complexities of efficient SIC execution, a critical facet of NOMA systems, our DRL framework opts for an agile approach. Post each action, it recalculates channel vectors, adjusting the decoding sequence to suit thecurrent channel dynamics. This flexibility ensures the SIC requirement is inherently met, underscoring theDRL algorithm's capacity to adjust to live changes within the communication framework, thus optimizing the SIC mechanism's effectiveness.
- Proposition: Guaranteeing SIC Compliance: The assurance of the SIC constraint within our DRL frameworkis showcased through an adaptive decoding sequence optimization, reliant on the immediate channel state.This assertion is supported by revisiting the SINR formulations:

For user *t* decoding at user *j*'s channel:

$$SINR_{t\to j} = \frac{P_{max}\rho_t}{\sum_{i=t+1}^{K} P_{max}\rho_i + \frac{\sigma^2}{|h_j|^2}} \quad (15)$$

And for user *t* decoding at its own channel:

$$SINR_{t\to t} = \frac{P_{max}\rho_t}{\sum_{i=t+1}^{K} P_{max}\rho_i + \frac{\sigma^2}{|h_t|^2}} \quad (16)$$

Given the channel condition

$$|h_j| \geq |h_t|$$

it's evident that the SINR for decoding at *j* consistently meets or surpasses the self-decoding SINR at *t* thereby upholding the SIC stipulation. This highlights the DRL algorithm's adeptness in dynamically tailoring to channel fluctuations while ensuring the SIC protocol's integrity within the NOMA framework.

Problem Tackling and Constraints Adherence: The DRL approach refocuses the optimization challenge towards elevating each user's data rate, while stringently adhering to the system's QoS requisites and the physical constraints tied to the RIS and UAV:

$$\max_{\{\Theta,a,v\}} \sum_{t=1}^{K} R_{t\to t} \quad (17)$$

Subject to: Maintaining QoS by ensuring $R_{t\to t}$ $R_{\min}$ for all network users t, Observing power distribution boundaries and UAV spatial constraints for optimal resource utilization, Confirming RIS phase adjustments $>$ remain within practical limits, facilitating peak system operation.
This structured DRL methodology, powered by the nuanced capabilities of the DDPG algorithm, proficiently navigates through the multifaceted constraints and goals inherent in the RIS-enhanced UAV-assisted NOMA downlink system. It not only propels network efficiency but also substantially fortifies the physical layer's security, illustrating the potential for ground-breaking progress in secure and high-performing wire- less communication networks.

Working Procedure:

The foundation of our DDPG-based learning mechanism is established upon four critical neuralnetworks: the target and evaluation networks for both the actor and critic components. Specifically, the target actor network (θ_μ') and the evaluation actor network (θ_μ), alongside their critic counterparts ($\theta_q'$and θ_q), are meticulously constructed with parallel architectures to ensure consistency in learning and prediction dynamics. An experience replay buffer (B), with a predefined capacity (C), serves as the memory infrastructure, storing transitions that encapsulate the state-action-reward sequences experienced by the agent. Within each episode of the learning process, a fresh initialization of channel gain *(h_k)*, RIS phase shifts (Φ), and user positions *(u)* within the designated area (A) is conducted. The UAV's horizontal position *(v)* is set at a predetermined point, and power allocations *(ρ)* are uniformly distributed as initial conditions. Following this setup, the system computes the data rates ($R_i^{(t)}$) for all users, setting the stage for the initial state (s_t). The evaluation actor network then processes s_t to generate the corresponding action (a_t), which encompasses decisions on phase shifts, power allocations, and UAV positioning. The immediate reward (r_t), reflective of the system's performance as per equation [9], and the subsequent state (s_{t+1}), as determined by equation [8], are determined thereafter.
These elements form a transition ({s_t, a_t, r_t, s_{t+1}}) that is archived within the replay buffer (D). Upon filling the replay buffer, the training phase commences, wherein each episode entails updating the current transition within D, followed by the extraction of a mini-batch consisting of N_B transitions ({s_i, a_i, r_i, s_{i+1}}) for processing. The target Q values (y_i) are computed for each transition in the minibatch, employing the equation:

$$y_i = \left\{ r_i r_i + \lambda Q'\left(s_{i+1}, \mu'\left(s_{i+1}; \theta_{\mu}'; \theta_{q}'\right)\right) \right. \quad (18)$$

Here, λ denotes the discount factor, emphasizing the value of future rewards. The critic evaluation network (θ_q) is updated by minimizing the loss function:

$$L(\theta_q) = \frac{1}{N_B} \sum_{i=1}^{N_B} (y_i - Q(s_i, a_i; \theta_q))^2 \quad (19)$$

Subsequently, the actor evaluation network (θ_u) is refined through gradient ascent, leveraging the gradient of the Q function with respect to the actor parameters:

$$\nabla_{\theta_\mu} J = \frac{1}{N_B} \sum_{i=1}^{N_B} \left(\nabla_a Q(s_i, \mu(s_i; \theta_\mu); \theta_q) | \nabla_{\theta_\mu} \mu(s_i; \theta_\mu)\right) \quad (20)$$

The algorithm iterates through these steps, periodically employing soft updates (as per equation (6)) to gradually align the target networks ($\theta\mu'$ and $\theta q'$) with their evaluation counterparts, thereby ensuring a stable convergence towards optimal policy and value functions

Complexity Analysis: The computational complexity of the DDPG algorithm, within this secure communication framework, is predominantly dictated by the dimensions of the input

$$(D_{in} = 2(K + N + 1))$$

rendering the overall complexity to

$$2\mathcal{O}(D_{in}^4 + D_{in}^5)$$

This complexity profile is notably more efficient than traditional semidefinite relaxation (SDR)-based optimization methods, which typically exhibit a higher computational burden

$(\mathcal{O}((6K^2 + N^2 + M)^{3.5}))$
thereby highlighting the efficacy and computational viability of employing DDPG in enhancing the security and performance of RIS-equipped UAV-assisted NOMA downlink systems.

State Space (s): Comprises UAV's position $v(x,y)$,
RIS phase shifts Φ, and channel state information (CSI) G, $h_{r,k}$, $h_{r,e}$ for all users and eavesdroppers.

Action Space (a): Includes power allocation coefficients p_k for each user, RIS phase shift matrix ϕ and UAV's next position $v`(x,y)$. Reward (r): Based on secrecy rates of users, calculated as

$$r_t = \sum_{k=1}^{K} \left[R_k - \max_{e\in\mathcal{E}} R_{e,k}\right]^+$$

where R_k and $R_{e,k}$ are defined by our system model equations

Transition Dynamics: The transition from s_t to s_{t+1} incorporates the UAV's movement, changes in the channel state, and adjustments in the RIS phase shifts according to the system model equations.

Constraints Handling: Ensure actions a_t (power allocations, phase shifts, UAV positions) adhere to physical constraints and system model requirements.
Termination: Episodes terminate based on predefined conditions such as maximum steps, achievement of a desired secrecy rate level, or stability in the learning process.

Notes for Integration: The reward function r_t directly incorporates the secrecy rate equations from our system model, ensuring that the

optimization is closely aligned with the objective of enhancing PLS The state and action representations are designed to capture the essential elements of the RIS-equipped UAV-NOMA system, including the UAV's mobility, the RIS's configurability, and the dynamic wireless environment. Adjustments in the algorithm (e.g., network architectures, exploration strategies) might be necessary to accommodate the complexity and specifics of our communication system model.

This structured DDPG algorithm, tailored to our RIS-UAV-NOMA system, provides a mathematical framework for optimizing the system's performance with respect to PLS leveraging deep reinforcement learning techniques to address the high-dimensional and non-convex nature of the problem.

Algorithm 1 Proposed DDPG-based algorithm

Initialization:

Randomly initialize the critic network $Q(s,a;\theta^Q)$ and the actor network $\mu(s;\theta^\mu)$ with weights θ^Q and θ^μ, respectively.

Initialize the target networks Q' and μ' with weights $\theta^{Q'} \leftarrow \theta^Q$ and $\theta^{\mu'} \leftarrow \theta^\mu$.

Prepare the experience replay buffer $\mathcal{D}$ with capacity C.

Set learning parameters: learning rate β, discount factor γ, soft update coefficient τ, and minibatch size N_B.

for *each episode* $j = 1$ ***to*** J **do**

- Initialize the UAV's position $v(x,y)$, RIS phase shifts Φ, and obtain the channel state information $G^{(j)}$ and $h_{rk}^{(j)}$ for all users and eavesdroppers.
- Set power allocation coefficients $\rho_k = \frac{1}{K}$ for each user k.
- Calculate the secrecy rates for each user using $R_{s,k}$ equations and determine the initial state s_1.
- **for** *each timestep* $t = 1$ ***to*** T **do**
  - Select action $a_t = \mu(s_t;\theta^\mu) + \mathcal{N}_t$ based on the current policy and exploration noise, including adjustments to ρ_k, Φ, and $v'(x,y)$.
  - Implement the selected actions, adjusting RIS phase shifts, power allocations, and UAV's position.
  - Calculate the new secrecy rates $R_{s,k}^{(t)}$ for each user and the total secrecy rate to obtain the reward r_t.
  - Observe the new state s_{t+1} based on the updated UAV's position, RIS phase shifts, and CSI.
  - Store the transition (s_t, a_t, r_t, s_{t+1}) in the replay buffer $\mathcal{D}$.
  - Sample a minibatch of N_B transitions from $\mathcal{D}$ for training.
  - Update the critic network by minimizing the loss function based on the target Q values.
  - Update the actor network using the sampled policy gradient.
  - Apply soft updates to the target networks Q' and μ' using the coefficient τ.

Proposed DDPG Algorithm for RIS-UAV Enhanced Security Architecture for NOMA Networks

II. Simulation Results

In our investigation, we employ a DDPG-driven framework tailored for an RIS-supported UAV-NOMA communication setup to scrutinize its efficacy in boosting system performance and security. The simulation setup positions the Base Station (BS) at the coordinate origin while situating the RIS-equipped UAV at the initial location of (50,0). The designated user area, defined by the vertices (45,45), (55,45), (55,55), and (45,55) hosts users whose positions are predetermined and remain constant throughout each simulation episode.

The system assumes Line-of-Sight (LoS) connectivity for both the BS-to-RIS and RIS-to-user links adopting a Rician fading model articulated as:

$$\boldsymbol{G} = \sqrt{\frac{\Omega}{\Omega+1}}\bar{\boldsymbol{H}} + \sqrt{\frac{1}{\Omega+1}}\boldsymbol{H}_{\text{Rayleigh}}$$

Here, $\bar{\boldsymbol{H}}$ represents the line-of-sight component, $\boldsymbol{H}_{Rayleigh}$ signifies the non-line-of-sight component subject to Rayleigh fading, and $\boldsymbol{\Omega}$ is the Rician K-factor set to 10 for our simulations. The path loss exponent is denoted by α and is chosen as 2. Channel conditions are randomly generated at exponent is denoted by α and is chosen as 2. Channel conditions are randomly generated at the onset of each episode and remain static for the episode's duration. The BS and the RIS-equipped UAV are fixed at altitudes of 20 meters and 30 meters, respectively.

Further parameters include a noise power setting of σ^2 = -60dB and a baseline user rate requirement of R_{min}= 1.2 bps/Hz. The network architecture for the Actor network comprises a dual-layer fully connected neural network for both the evaluation and target networks, with input and output layers sized according to the dimensions of the state and action vectors. Activation functions include ReLU for the initial layer and tanh for the output layer to ensure a strong gradient signal.

The Critic network adopts a similar two-layered structure, processing state and action inputs through separate pathways before merging and applying ReLU activation, leading into the final output layer. Batch normalization is applied across both networks to stabilize learning.

Hyperparameters for the simulation include an evaluation network learning rate (β) of 0.0001, a discount factor (λ) of 0.95, a soft update rate (τ) of 0.005, a replay buffer capacity (C) of 50,000, 1500 episodes, 300 steps per episode, and a minibatch size (N_B) of 16. Exploration noise, introduced to promote policy diversity, follows a complex Gaussian distribution with zero mean and a variance of 0.1.

To incorporate the potential threat posed by eavesdroppers (Eve) into the system, the simulation contemplates the RIS-to-Eve channel, denoted as $h_{r,e}$, and the RIS-to-Eve distance, $d_{I,e}$. The channel gain experienced by Eve, while analogous to that of legitimate users, is adjusted for Eve's specific location, expressed as:

$$\mathbf{h}_e = \frac{\mathbf{h}_{r,e}^H \mathbf{\Phi} \mathbf{G}}{(d_{B,I} \cdot d_{I,e})^\alpha}$$

This incorporation enables the system to evaluate and mitigate the risks associated with Eve's interception attempts, fine-tuning power allocations p_i and phase adjustments θ_n to safeguard communication channels.

In Figure. 3, the graph showcases the progression of secrecy rates over a sequence of episodes for various learning rates in a UAV-aided communication network using a DDPG algorithm. The system, aimed at bolstering PLS adapts and refines its performance across episodes. Learning rates are a critical factor here: a high rate (LR=0.1) leads to rapid learning but with considerable volatility, potentially due to over-adjustments. Conversely, lower learning rates, such as LR=0.001, demonstrate a gradual but stable enhancement in secrecy rates, suggesting a more methodical learning approach that might converge more reliably to optimal strategies for secure communications. This balance between speed and stability in learning rates is crucial for the algorithm to efficiently navigate and optimize the complex dynamics of the secure communication environment.

In Figure. 4 depicts the impact of different numbers of eavesdroppers (denoted by E) on the secrecy rate in a UAV-assisted secure communication system, as episodes progress. A clear pattern emerges: as the number of eavesdroppers increases from E=2 to E=5, there is a general trend of decreasing secrecy rate, indicating that more eavesdroppers make it challenging to maintain high levels of secure communication. The plot with E=2 reaches the highest secrecy rate more quickly and maintains it with less fluctuation, suggesting that fewer eavesdroppers make it easier for the system to optimize security. Conversely, with E=5, the system takes longer to reach a stable and high secrecy rate, and it experiences more significant variance, reflecting the increased complexity of optimizing secure communication in the presence of more eavesdropping threats. This illustrates the system's adaptive learning in response to the number of eavesdroppers, which directly affects the secrecy rate over time.

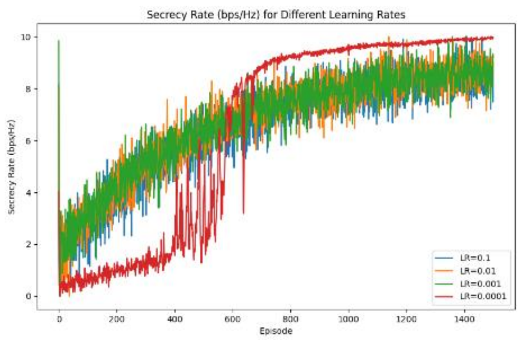

Figure 3: Different Learning rates

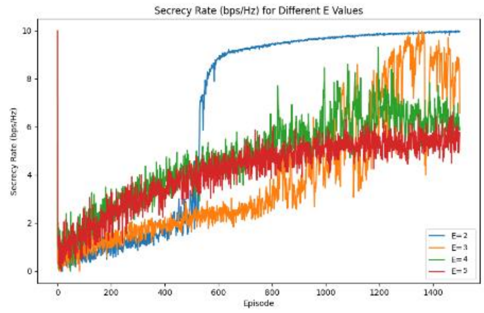

Figure 4: Different Number of Eves

In Figure. 5 displays the evolution of the secrecy rate, measured in bits per second per Hertz (bps/Hz), across different episodes for varying numbers of users in a secure communication system aided by an UAV. As the episodes increase, the secrecy rate for each scenario with different user counts ranging from K=2 to K=5 tends to rise and eventually stabilizes. Initially, the learning algorithm quickly enhances the secrecy rate for lower user scenarios (K=2 and K=3), indicating that the system can more easily optimize for fewer users. As the user count increases (K=4 and K=5), the rate at which the secrecy rate improves slows down, suggesting that a greater number of users presents additional challenges for the system to optimize secure communication. However, all scenarios reach a point of convergence, indicating that despite the complexity introduced by more users, the system's learning algorithm can adapt and enhance the security over time.

In Figure. 6 plots the secrecy rate in bits per second per Hertz (bps/Hz) against the transmit power in decibels (dB) for different configurations of an RIS using both a proposed DDPG approach and a random RIS orientation. It compares the performance of a 50-element and 100-element RIS under both methodologies. As the transmit power increases, all configurations demonstrate an increase in the secrecy rate. The proposed DDPG algorithm with 100 RIS elements achieves the highest secrecy rate, indicating that the DDPG approach efficiently optimizes the phase shifts of the RIS for enhanced secure communication. In contrast, a random RIS orientation, even with the same number of elements, results in a significantly lower secrecy rate, highlighting the benefits of intelligent phase shift design in RIS-aided communication systems. The difference in performance between the proposed DDPG method and the random approach is more pronounced at higher element counts, suggesting that the advantages of the DDPG algorithm become more substantial as the number of RIS elements increases.

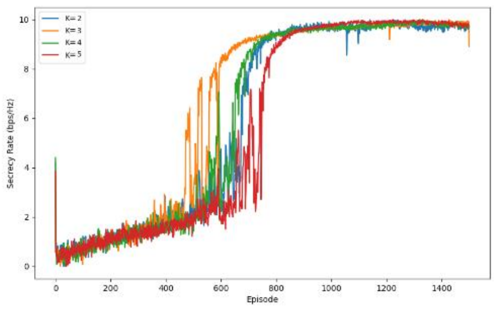

Figure 5: Different Number of Users

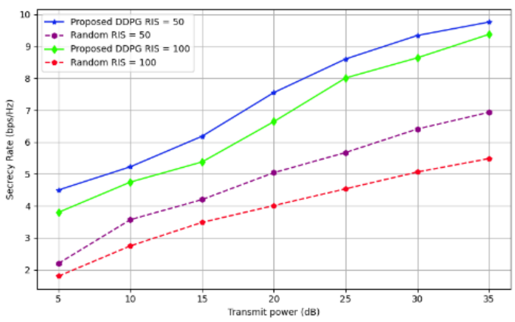

Figure 6: Transmit Power

In Figure. 7 depicts a comparison of secrecy rates achieved by employing an RIS with different numbers of elements. It compares a proposed DDPG optimization strategy against a random RIS element orientation at two transmit power levels, 10dB and 30dB. The DDPG strategy outperforms the random orientation at both power levels, with its advantage more pronounced at the higher number of RIS elements.

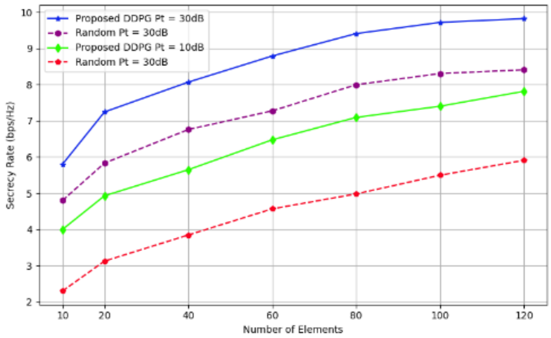

Figure 7: Number of RIS Elements

The DDPG method maintains superior performance even at the lower transmit power, highlighting the effectiveness of intelligent optimization in enhancing secure wireless communication.

III. Conclusion

Secure data transmission in a downlink network for NOMA by utilizing an Intelligent RIS in conjunction with an UAV. The basis for fine-tuning the IRS's phase shifts, the BS power distribution, and the UAV's spatial coordination is the DDPG algorithm. The main objective is to increase the total data rate while maintaining the confidentiality of SIC, which requires a decoding order that is flexible enough to change as the channel conditions do. The simulation results confirm that the implemented DDPG-centric approach increases network performance overall and remains robust in dealing with of fluctuating IRS elements and user counts. Significantly, it highlights an advanced defensive approach against unauthorized access, strengthening the network's defense against eavesdropping while maintaining broad access for authorized users and therefore providing a significant advancement in the enhancement of wireless networks.

References

[1]. J. Li, S. Xu, J. Liu, Y. Cao, and W. Gao, "Reconfigurable Intelli- gent Surface Enhanced Secure Aerial-Ground Communication," IEEE Transactions on Communications, vol. 69, no. 9, pp. 6185–6197, Jun, 2021.

[2]. V. N. Vo, C. So-In, H. Tran, D.-D. Tran, and T. P. Huu, "Performance Analysis of an Energy-Harvesting IoT System Using a UAV Friendly Jammer and NOMA Under Cooperative Attack," IEEE Access, vol. 8, pp. 221 986–222 000, Dec, 2020.

[3]. A. S. Abdalla, T. F. Rahman, and V. Marojevic, "UAVs with reconfigurable intelligent surfaces: Applications, challenges, and opportunities," arXiv preprint arXiv:2012.04775, 2020.

[4]. W. U. Khan, A. Mahmood, C. K. Sheemar, E. Lagunas, S. Chatzinotas, and B. Ottersten, "Reconfigurable Intelligent Surfaces for 6G Non- Terrestrial Networks: Assisting Connectivity from the Sky," IEEE Internet of Things Magazine, vol. 7, no. 1, pp. 34–39, Jan, 2024.

[5]. T. Alladi, Naren, G. Bansal, V. Chamola, and M. Guizani, "Secauthuav: A novel authentication scheme for uav-ground station and uav- uav communication," IEEE Transactions on Vehicular Technology, vol. 69, no. 12, pp. 15 068–15 077, Oct, 2020.

[6]. S. Jiao, X. Xie, and Z. Ding, "Deep reinforcement learning based optimization for irs based uav-noma downlink networks," arXiv preprint arXiv:2106.09616, 2021.

[7]. J. Yao and J. Xu, "Joint 3D Maneuver and Power Adaptation for Secure UAV Communication With CoMP Reception," IEEE Transactions on Wireless Communications, vol. 19, no. 10, pp. 6992–7006, July, 2020.

[8]. M. Najafi, V. Jamali, R. Schober, and H. V. Poor, "Physics-Based Modeling and Scalable Optimization of Large Intelligent Reflecting Surfaces," IEEE Transactions on Communications, vol. 69, no. 4, pp. 2673–2691, Dec, 2021.

[9]. F. Zhou, Z. Li, J. Cheng, Q. Li, and J. Si, "Robust AN-Aided Beamforming and Power Splitting Design for Secure MISO Cognitive Radio With SWIPT," IEEE Transactions on Wireless Communications, vol. 16, no. 4, pp. 2450–2464, Mar, 2017.

[10]. X. Zhang, X. Zhou, and M. R. McKay, "Enhancing Secrecy with Multi-Antenna Transmission in Wireless Ad Hoc Networks," IEEE Transactions on Information Forensics and Security, vol. 8, no. 11, pp. 1802–1814, Aug, 2013.

[11]. Y. Liu, X. Liu, X. Mu, T. Hou, J. Xu, M. Di Renzo, and N. Al-Dhahir, "Reconfigurable Intelligent Surfaces: Principles and Opportunities," IEEE Communications Surveys Tutorials, vol. 23, no. 3, pp. 1546– 1577, May, 2021.

[12]. X. Sun, D. W. K. Ng, Z. Ding, Y. Xu, and Z. Zhong, "Physical layer security in uav systems: Challenges and opportunities," IEEE Wireless Communications, vol. 26, no. 5, pp. 40–47, Oct, 2019.

[13]. Z. Zhang, C. Zhang, C. Jiang, F. Jia, J. Ge, and F. Gong, "Improving physical layer security for reconfigurable intelligent surface aided Noma 6g networks," IEEE Transactions on Vehicular Technology, vol. 70, no. 5, pp. 4451–4463, Mar, 2021.

[14]. C. Huang, G. C. Alexandropoulos, A. Zappone, C. Yuen, and M. Deb- bah, "Deep Learning for UL/DL Channel Calibration in Generic Massive MIMO Systems," in ICC 2019 - 2019 IEEE International Conference on Communications (ICC), Shanghai, China, July, 2019, pp. 1–6.

[15]. T. Wang, F. Fang, and Z. Ding, "An SCA and Relaxation Based Energy Efficiency Optimization for Multi-User RIS-Assisted NOMA Networks," IEEE Transactions on Vehicular Technology, vol. 71, no. 6, pp. 6843–6847, Mar, 2022.
[16]. J. Chen, L. Guo, J. Jia, J. Shang, and X. Wang, "Resource Allocation for IRS Assisted SGF NOMA Transmission: A MADRL Approach," IEEE Journal on Selected Areas in Communications, vol. 40, no. 4, pp. 1302–1316, 2022.
[17]. J. Chen, Y. Xie, X. Mu, J. Jia, Y. Liu, and X. Wang, "Energy Efficient Resource Allocation for IRS Assisted CoMP Systems," IEEE Transactions on Wireless Communications, vol. 21, no. 7, pp. 5688– 5702, Jan, 2022.
[18]. J. Chen, Z. Ma, Y. Zou, J. Jia, and X. Wang, "DRL-based Energy Efficient Resource Allocation for STAR-RIS Assisted Coordinated Multi-Cell Networks," in GLOBECOM 2022 - 2022 IEEE Global Communications Conference, Jan, 2022, pp. 4232–4237.
[19]. X. Mu, Y. Liu, L. Guo, J. Lin, and H. V. Poor, "Intelligent Reflecting Surface Enhanced Multi-UAV NOMA Networks," IEEE Journal on Selected Areas in Communications, vol. 39, no. 10, pp. 3051–3066, Jun, 2021.
[20]. J. Zhao, Y. Zhu, X. Mu, K. Cai, Y. Liu, and L. Hanzo, "Simultaneously Transmitting and Reflecting Reconfigurable Intelligent Surface (STAR-RIS) Assisted UAV Communications," IEEE Journal on Selected Areas in Communications, vol. 40, no. 10, pp. 3041–3056, Aug, 2022.
[21]. Z. Tang, T. Hou, Y. Liu, J. Zhang, and C. Zhong, "A Novel Design of RIS for Enhancing the Physical Layer Security for RIS-Aided NOMA Networks," IEEE Wireless Communications Letters, vol. 10, no. 11, pp. 2398–2401, Aug, 2021.
[22]. M. Cui, G. Zhang, and R. Zhang, "Secure Wireless Communication via Intelligent Reflecting Surface," IEEE Wireless Communications Letters, vol. 8, no. 5, pp. 1410–1414, May, 2019.
[23]. X. Yu, D. Xu, Y. Sun, D. W. K. Ng, and R. Schober, "Robust and Secure Wireless Communications via Intelligent Reflecting Surfaces," IEEE Journal on Selected Areas in Communications, vol. 38, no. 11, pp. 2637–2652, July, 2020.
[24]. Y. Han, N. Li, Y. Liu, T. Zhang, and X. Tao, "Artificial Noise Aided Secure NOMA Communications in STAR-RIS Networks," IEEE Wireless Communications Letters, vol. 11, no. 6, pp. 1191– 1195, Mar, 2022.
[25]. H. Long, M. Chen, Z. Yang, Z. Li, B. Wang, X. Yun, and M. ShikhBahaei, "Joint Trajectory and Passive Beamforming Design for Secure UAV Networks with RIS," in 2020 IEEE Globecom Workshops (GC Wkshps, Mar, 2020, pp. 1–6.
[26]. D. Wang, Y. Zhao, Y. Lou, L. Pang, Y. He, and D. Zhang, "Se- cure NOMA Based RIS-UAV Networks: Passive Beamforming and Location Optimization," in GLOBECOM 2022 - 2022 IEEE Global Communications Conference, Rio de Janeiro, Brazil, Jan, 2022, pp. 3168–3173.
[27]. L. Li, Q. Cheng, X. Tang, T. Bai, W. Chen, Z. Ding, and Z. Han, "Resource Allocation for NOMA-MEC Systems in Ultra-Dense Net- works: A Learning Aided Mean-Field Game Approach," IEEE Trans- actions on Wireless Communications, vol. 20, no. 3, pp. 1487–1500, Nov, 2021.
[28]. D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Ried- miller, "Deterministic policy gradient algorithms," in international conference on machine learning. Pmlr, 2014, pp. 387–395.
[29]. S. Zhang and R. S. Sutton, "A deeper look at experience replay," arXiv preprint arXiv:1712.01275, 2017.
[30]. H. Peng and X. S. Shen, "DDPG-based Resource Management for MEC/UAV-Assisted Vehicular Networks," in 2020 IEEE 92nd Vehicular Technology Conference (VTC2020-Fall), Victoria, BC, Canada, Feb, 2020, pp. 1–6.

BIOGRAPHIES OF AUTHORS

SYED ZAIN UL ABIDEEN received the B.Sc. degree in software engineering from University of Science and Technology Bannu, kpk Pakistan in 2018. Since 2018, he has worked as a Full Stack Developer in various companies, mastering a wide range of programming languages. He is currently pursuing the Master degree with College of Computer Science and Technology, Qingdao University, Qingdao 266071, China. his main research interests include wireless communication, machine learning, deep learning, physical layer security, reconfigurable intelligent surface and backscatter communication. E-mail: zain208shah@gmail.com.

ABDUL WAHID received the B.S. degree in Computer Science from the University of Peshawar, Pakistan, and the M.S. degree in Computer Science and Technology from the Southwest University of Science and Technology (SWUST), Mianyang, Sichuan, China, in 2017 and 2020, respectively. He is currently pursuing the Ph.D. degree with the College of Computer Science and Technology, Qingdao University, Qingdao, Shandong, China. His current research interests include wireless communications, physical layer security, reconfigurable intelligent surface, machine learning, and 6G technology.
E-mail: wahidjan999@gmail.com

Mian Muhammad Kamal received the M.S. degree from Northwestern Polytechnical University, in 2018, and the Ph.D. degree from the Zhengzhou University in 2023. He was a part of the Next Generation Ubiquitous Network Laboratory, College of Electronic and Information, Northwestern Polytechnical University, China, from Sep 2016 to June 2019. He is currently a Postdoctoral Fellow at Southeast University. His research interests include MIMO antenna systems, smart-phone antennas, slot antennas, 5G antennas, millimeter-wave, terahertz components, origami and kirigami antennas, multiband antennas, and meta-surfaces. He is the author or coauthor of more than 25 technical journals and conference papers