



ARTICLE

Computational Optimization of RIS-Enhanced Backscatter and Direct Communication for 6G IoT: A DDPG-Based Approach with Physical Layer Security

Syed Zain Ul Abideen¹, Mian Muhammad Kamal^{2,*}, Eaman Alharbi³, Ashfaq Ahmad Malik⁴,
Wadee Alhalabi⁵, Muhammad Shahid Anwar^{6,*} and Liaqat Ali⁷

¹College of Computer Science and Technology, Qingdao University, Qingdao, 266071, China

²School of Electronic Science and Engineering, Southeast University, Nanjing, 210018, China

³Computer Science Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, 80200, Saudi Arabia

⁴Department of Quality Assurance, Al-Kawthar University, Karachi, 75300, Pakistan

⁵Department of Computer Science, Immersive Virtual Reality Research Group, King Abdulaziz University, Jeddah, 80200, Saudi Arabia

⁶Department of AI and Software, Gachon University, Seongnam-si, 13120, Republic of Korea

⁷Department of Electrical Engineering, University of Science and Technology, Bannu, 28100, Pakistan

*Corresponding Authors: Mian Muhammad Kamal. Email: mmkamal@seu.edu.cn;

Muhammad Shahid Anwar. Email: shahidanwar786@gachon.ac.kr

Received: 02 December 2024; Accepted: 05 February 2025; Published: 03 March 2025

ABSTRACT: The rapid evolution of wireless technologies and the advent of 6G networks present new challenges and opportunities for Internet of Things (IoT) applications, particularly in terms of ultra-reliable, secure, and energy-efficient communication. This study explores the integration of Reconfigurable Intelligent Surfaces (RIS) into IoT networks to enhance communication performance. Unlike traditional passive reflector-based approaches, RIS is leveraged as an active optimization tool to improve both backscatter and direct communication modes, addressing critical IoT challenges such as energy efficiency, limited communication range, and double-fading effects in backscatter communication. We propose a novel computational framework that combines RIS functionality with Physical Layer Security (PLS) mechanisms, optimized through the algorithm known as Deep Deterministic Policy Gradient (DDPG). This framework adaptively adapts RIS configurations and transmitter beamforming to reduce key challenges, including imperfect channel state information (CSI) and hardware limitations like quantized RIS phase shifts. By optimizing both RIS settings and beamforming in real-time, our approach outperforms traditional methods by significantly increasing secrecy rates, improving spectral efficiency, and enhancing energy efficiency. Notably, this framework adapts more effectively to the dynamic nature of wireless channels compared to conventional optimization techniques, providing scalable solutions for large-scale RIS deployments. Our results demonstrate substantial improvements in communication performance setting a new benchmark for secure, efficient and scalable 6G communication. This work offers valuable insights for the future of IoT networks, with a focus on computational optimization, high spectral efficiency and energy-aware operations.

KEYWORDS: Computational optimization; reconfigurable intelligent surfaces (RIS); 6G networks; IoT and DDPG; physical layer security (PLS); backscatter communication



1 Introduction

The dynamic advancement in wireless technology is driving the transition toward sixth-generation (6G) networks which aim to meet the demands of revolutionary data rates with minimal latency and secure communication for emerging Internet of Things (IoT) applications. With IoT devices proliferating across diverse industries, the networks face critical challenges in scalability, bandwidth efficiency and security [1,2]. Addressing these difficulties requires innovative communication frameworks that can efficiently manage resources while safeguarding against security threats.

BackCom is increasingly recognized as a groundbreaking energy efficient solution by reflecting incident signals from a carrier emitter without generating new radio waves [3,4]. This technique is particularly attractive for resource constrained IoT devices, as it significantly reduces power consumption. However, BackCom suffers from inherent limitations such as signal degradation caused by double-fading effects and restricted communication range, which hinder its scalability and practical deployment in dense IoT networks [5]. Overcoming these drawbacks requires advanced technologies capable of optimizing signal propagation and enhancing communication reliability.

Reconfigurable Intelligent Surface (RIS) have recently emerged as a transformative technology in wireless networks [6]. RIS is composed of an extensive array of passive reflective elements, capable of adaptively manipulating the phase and magnitude of approaching electromagnetic waves. This ability makes RIS a powerful enabler for enhancing both backscatter and direct communication by improving signal propagation, mitigating interference and boosting spectral efficiency in complex IoT environments [7]. Unlike conventional systems that treat the environment as a passive channel, RIS-enabled systems actively shape the propagation environment enabling intelligent signal control and reducing the impact of path loss [8].

To fully exploit the transformative potential of RIS, this study introduces an innovative RIS-aided Joint Backscattering and Communication (JBAC) system setting a new benchmark for 6G networks. The proposed system seamlessly integrates RIS with traditional backscatter communication, addressing inherent limitations by enhancing transmission efficiency and optimizing spectral utilization [9]. Beyond boosting communication performance, the study prioritizes data privacy and security by embedding PLS mechanisms [10]. By leveraging the inherent randomness of wireless channels, the PLS provides robust protection against eavesdropping serving as a lightweight, energy efficient alternative to conventional encryption techniques [11,12].

To achieve real-time optimization the system employs the DDPG algorithm a reinforcement learning method well-suited for continuous action spaces [13]. DDPG dynamically optimizes RIS phase shifts and transmitter beamforming, enabling adaptive responses to changing channel conditions. By seamlessly coordinating RIS configurations and power allocation at the transmitter, the proposed framework enhances communication performance while strengthening security measures [14]. Furthermore, the integration of reinforcement learning ensures robust operation in the context of practical issues like incomplete channel state information (CSI) and RIS phase shift quantization

This work combines RIS, PLS and reinforcement learning to address the critical challenges of security and efficiency in IoT networks. By synergizing these advanced technologies the proposed framework sets new benchmarks for 6G communication protocols. It achieves significant improvements in secrecy rate and transmission efficiency, enabling the development of interconnected, secure and energy-conscious IoT ecosystems. The results of this research contribute substantially to the growing body of knowledge around RIS-enabled systems providing a foundation for high-performance IoT networks in the 6G era.

1.1 Related Work

The emergence of RIS technology has redefined the concept of wireless communication by enabling the control of electromagnetic waves through programmable surfaces, as summarized in Table 1, which dynamically manipulate phase shifts and amplitudes [15,16]. RIS technology is increasingly being integrated into 6G communication systems to meet the demands of high transmission rates, incredibly low latency and ubiquitous connectivity especially for IoT applications [17]. Through RIS, communication systems can reduce interference, improve spectral efficiency, and mitigate fading effects, thereby extending the operational range and robustness of both backscatter communication (BackCom) and direct transmissions [18,19].

Backscatter communication has gained prominence for its energy-efficient data transmission capabilities by using passive reflection of continuous carrier waves. This makes it particularly suitable for IoT devices with constrained energy resources [3]. However, BackCom faces key challenges such as the double-fading effect, which severely impacts the quality of transmission and limits communication range [20,21]. To overcome these limitations, researchers have proposed RIS-aided BackCom systems that enhance transmission efficiency by optimizing signal paths through intelligent reflections [22]. By introducing additional propagation channels, RIS helps mitigate path loss and extend coverage, facilitating seamless communication in dense IoT environments [23,24].

In parallel, the rise of PLS has been motivated by the growing need for lightweight security solutions in IoT and wireless systems. PLS takes advantage of the inherent randomness in wireless channels to secure data transmission without the computational overhead associated with cryptographic methods [25,12]. RIS-assisted PLS frameworks have demonstrated the ability to boost secrecy capacity by tailoring phase shifts to steer reflected signals toward legitimate users while weakening signals received by potential eavesdroppers [13,26]. This has led to the development of RIS-aided wiretap communication models where beamforming strategies are optimized to maximize secrecy rates [27,28]. These models ensure that confidential data remains protected even in highly dynamic network environments.

The adoption of machine learning (ML) algorithms for managing RIS configurations is another significant advancement. Reinforcement learning techniques, such as the DDPG algorithm, have been used to dynamically adjust RIS parameters and optimize beamforming patterns [29,12]. These algorithms are particularly effective in continuous action spaces such as those encountered in RIS systems where precise phase shifts are required to optimize both signal strength and security [30]. DDPG-based frameworks enhance the adaptability of RIS-aided communication systems by enabling real-time optimization in response to changing environmental conditions [31].

Further research has focused on the development of bistatic BackCom systems integrated with RIS. In these systems, the carrier emitter and receiver are spatially separated enabling more flexible network deployments [24]. RIS enhances these setups by introducing controllable propagation paths leading to better signal quality and reduced interference [20]. Resource allocation and beamforming optimization are critical research areas that empower RIS to achieve an effective balance between power distribution and the fulfillment of secrecy constraints [32].

As the number of RIS elements increases in advanced 6G communication systems, the computational complexity of optimization algorithms like DDPG becomes a critical consideration. This paper investigates how the complexity of DDPG scales with the number of RIS elements and demonstrates that the proposed approach remains efficient and scalable, even for large-scale RIS configurations.

Table 1: Comparison of related work

Study	Focus area	Strengths	Limitations	Proposed framework advantage
Pan et al. [15]	RIS for spectral efficiency	High spectral efficiency and reduced fading	No focus on security or BackCom integration	Joint optimization of RIS and BackCom
Jiang et al. [3]	BackCom for IoT	Energy efficiency for IoT devices	Double-fading effects reduce reliability	Mitigates double-fading via RIS
Nguyen et al. [26]	RIS-aided PLS	Improved secrecy capacity using RIS	Assumes perfect CSI and continuous phases	Robust to imperfect CSI and quantized phases
Puspitasari et al. [29]	ML for RIS optimization	Effective real-time optimization	Limited scalability to large-scale RIS	Scalable to large RIS configurations
Liang et al. [20]	Bistatic BackCom with RIS	Enhanced signal propagation and coverage	No integration with security mechanisms	Integrates PLS with RIS and BackCom
Proposed work	Joint RIS, BackCom, and PLS	Comprehensive optimization, real-time learning	–	Dynamic, scalable, and secure solution

Recent studies highlight the effectiveness of Deep Reinforcement Learning (DRL) in optimizing RIS-assisted networks, reference [33] proposed a HAP-based integrated satellite-aerial-terrestrial relay network using a DRL-based Long Short-Term Memory - Double Deep Q-Network (LSTM-DDQN) framework to maximize ergodic rate under Unmanned Aerial Vehicle (UAV) energy constraints. Similarly, reference [34] explored multi-RIS-assisted satellite-UAV-terrestrial networks, addressing dynamic environments and spectrum scarcity with a multi-objective deep deterministic policy gradient (MO-DDPG) algorithm to optimize achievable rate and energy efficiency. These works demonstrate DRL's potential for efficient RIS-enabled integrated networks.

As the role of RIS expands in the realm of 6G networks, several works have established new benchmarks for secrecy rate, spectral efficiency, and power management by integrating RIS, PLS, and advanced ML techniques. This study builds upon these foundations, proposing a novel Joint Backscattering and Communication (JBAC) system that leverages DDPG to dynamically optimize RIS configurations. The framework ensures efficient and secure communication by enhancing legitimate signal paths while suppressing those accessible to eavesdroppers, thereby aligning with the key goals of future 6G-enabled IoT networks.

1.2 Motivation and Contributions

The rapid evolution of wireless communication systems and the advent of 6G networks bring unprecedented opportunities for hyper-connected environments, primarily driven by the proliferation of IoT devices. However, these advancements also pose significant challenges: ensuring secure, efficient, and reliable

communication under stringent resource constraints, addressing the limitations of imperfect CSI, and managing quantized RIS phase shifts.

BackCom emerges as a promising solution for energy-efficient data transfer in IoT networks, yet it suffers from double-fading effects, limited communication range, and security vulnerabilities. Moreover, the dense connectivity envisioned for 6G networks amplifies the risks of eavesdropping, necessitating advanced security strategies that align with resource-efficient operation.

RIS offer a paradigm shift by transforming wireless environments with their ability to manipulate electromagnetic waves. While traditional research treats RIS as a passive signal reflector, this work explores its potential as an active optimization component. By strategically configuring RIS, the limitations of backscatter communication can be mitigated, and both communication performance and security can be enhanced. Furthermore, the integration of advanced reinforcement learning techniques, such as the DDPG algorithm, provides a pathway for real-time adaptability in dynamic wireless environments.

This study offers the following distinctive contributions:

1. **Novel RIS-Enabled Joint Backscatter and Communication (JBAC) System with Imperfect CSI Management:** We present a groundbreaking framework that seamlessly integrates RIS into IoT networks, revolutionizing both backscatter and direct communication modes. Unlike traditional approaches, our model tackles the challenge of imperfect CSI through an adaptive, learning-based strategy, ensuring robust system performance even in the face of channel estimation inaccuracies. By reimagining RIS as an active optimization component rather than a passive reflector, the framework dynamically optimizes signal paths, unlocking new levels of efficiency and reliability for next-generation IoT communication.
2. **Integration of PLS with Quantized RIS Phase Shifts:** While existing PLS frameworks rely on idealized RIS setups, our study incorporates quantized RIS phase shifts a practical constraint in hardware implementations. We demonstrate how RIS selectively enhances signals toward legitimate users while degrading signals directed toward eavesdroppers, effectively mitigating eavesdropping risks even under quantization constraints.
3. **Dynamic Optimization via DDPG Algorithm in Continuous Action Spaces:** To overcome challenges posed by continuous and high-dimensional action spaces, this work leverages the DDPG algorithm for real-time optimization of RIS configurations and power distribution. This approach not only adapts to environmental dynamics but also ensures the joint optimization of beamforming and RIS phase shifts, outperforming conventional optimization techniques.
4. **Addressing Backscatter-Specific Limitations: Double-Fading Effects and Coverage Range:** Our innovative approach revolutionizes backscatter communication by overcoming critical challenges like double fading effects and limited coverage. By seamlessly integrating RIS-assisted signal propagation our framework unlocks new possibilities leveraging RIS to establish additional propagation paths. This not only enhances spectral efficiency but also effectively mitigates fading delivering robust performance in dense IoT environments and paving the way for next-generation connectivity solutions
5. **Benchmarking for Secure 6G Communication Protocols:** By combining RIS functionality with advanced algorithmic methods and addressing practical constraints like imperfect CSI and quantized RIS phases this study establishes new benchmarks for secure and efficient 6G communication protocols. The results provide actionable insights for designing future ready IoT networks that prioritize both operational efficiency and data security.

2 System Model and Problem Formulation

2.1 System Model

We propose an integrated framework that utilizes BackCom with PLS in an RIS-aided environment as shown in Fig. 1. This model incorporates elements from both the bistatic BackCom system and the RIS-assisted wiretap communication system, providing a comprehensive approach to secure communication.

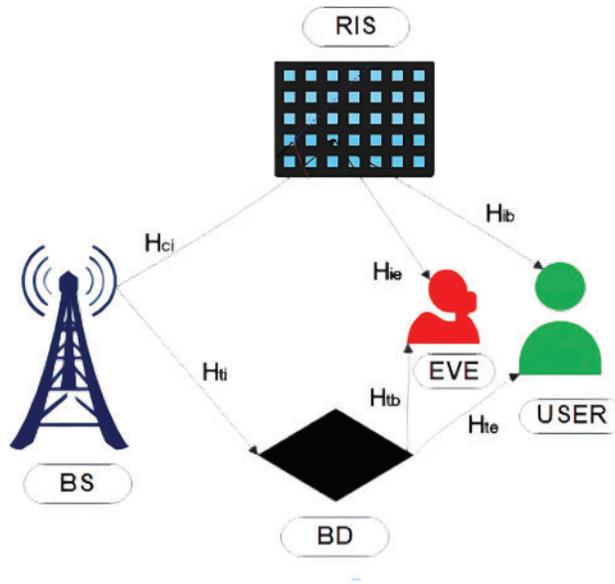


Figure 1: System model of RIS-aided IoT network for 6G communication, enhancing both backscatter and direct modes

The system involves several components. A base station (BS) equipped along M antennas functions as both a Carrier Emitter (CE) for BackCom and a transmitter for direct communication. The setup involves two single-antenna devices: a Backscatter Device (BD) and a legitimate user referred to as Bob. An Eavesdropper (Eve), equipped with M' antennas, aims to intercept communications. Additionally, an RIS with N reflecting elements is employed to enhance signal propagation through adjustable phase shifts.

In this system, the Carrier Emitter (CE) broadcasts a continuous wave signal $s(t)$ with power P and a beamforming vector $\mathbf{w} \in \mathbb{C}^{L \times 1}$. The power of the transmitted signal is subject to the constraint:

$$P = \|\mathbf{w}\|^2, \quad (1)$$

where \mathbf{w} represents the beamforming vector.

The Base Station (BS) also performs direct communication with Bob by transmitting a signal s , which satisfies the power condition $\mathbb{E}\{|s|^2\} = 1$. The beamforming vector \mathbf{w} is constrained by:

$$\|\mathbf{w}\|^2 \leq P_{\max}. \quad (2)$$

The communication channels between different components are represented as follows: the link from the CE/BS to the RIS is denoted as $\mathbf{H}_{CI} \in \mathbb{C}^{N \times L}$ and $\mathbf{H}_{TI} \in \mathbb{C}^{N \times M}$; the channel from the CE/BS to Bob and Eve are denoted as $\mathbf{h}_{CB} \in \mathbb{C}^{M \times 1}$ and $\mathbf{H}_{CE} \in \mathbb{C}^{M \times M'}$, respectively. Similarly, the RIS-to-receiver channels

are represented by $\mathbf{h}_{IB} \in \mathbb{C}^{N \times 1}$ for Bob and $\mathbf{H}_{IE} \in \mathbb{C}^{N \times M'}$ for Eve. Additionally, the direct channel from the Backscatter Device (BD) to Bob is given by $\mathbf{h}_{TB} \in \mathbb{C}$.

The RIS reflects incoming signals by applying phase shifts. These phase shifts are described by the matrix:

$$\Phi = \text{diag} \left(e^{j\Psi_1}, e^{j\Psi_2}, \dots, e^{j\Psi_N} \right), \quad (3)$$

where $\Psi_i \in [0, 2\pi]$ represents the phase shift of the i th reflecting element. These phase shifts influence both the backscatter communication and direct transmission, adding a degree of control to the signal propagation.

The received signal at Bob incorporates contributions from both direct communication and backscatter communication. It is given by:

$$y_B = \left(\mathbf{h}_{IB}^H \Phi \mathbf{H}_{TI} + \mathbf{h}_{TB}^H \right) \mathbf{w}s + n_B, \quad (4)$$

where $n_B \sim \mathcal{CN}(0, \sigma_B^2)$ is the additive noise at Bob. Similarly, the signal received by Eve is expressed as:

$$y_E = \left(\mathbf{H}_{IE}^H \Phi \mathbf{H}_{TI} + \mathbf{H}_{TE}^H \right) \mathbf{w}s + n_E, \quad (5)$$

where $n_E \sim \mathcal{CN}(0, \sigma_E^2 \mathbf{I})$ represents the additive noise at Eve.

The RIS enhances communication quality and secrecy by intelligently adjusting the phase shifts to strengthen the desired signals toward the legitimate user and reduce the signals leaking to the eavesdropper. This dual capability ensures efficient communication while mitigating the risk of eavesdropping.

CSI Acquisition in the Proposed System

Acquiring accurate CSI is essential for optimizing RIS configurations and ensuring secure communication. However, achieving perfect CSI in backscatter communication systems presents significant challenges due to double-fading effects, weaker received signals, and environmental dynamics. To address these issues, we adopt the following CSI acquisition strategies:

1. **Direct Links:** CSI for direct links (e.g., BS-to-Bob and BS-to-Eve) is obtained through traditional pilot-based channel estimation. The BS transmits known pilot sequences to Bob, who estimates the channel response and provides feedback. For Eve, the system assumes partial CSI knowledge via statistical models or worst-case approximations to ensure robust optimization.
2. **Backscatter Links:** - *Active Probing:* The BS sends probing signals that the BD modulates. The received composite signal at the BS is used to estimate the backscatter channel through advanced signal processing techniques such as sparse recovery. - *RIS-Assisted Refinement:* RIS configurations are adjusted iteratively to enhance the backscatter signal during estimation, enabling more accurate CSI acquisition.
3. **RIS-Specific Channels:** - *Cascaded Channel Estimation:* The BS transmits pilots while the RIS adjusts its phase shifts sequentially. The BS estimates the combined BS-to-RIS and RIS-to-receiver channels using a cascaded model:

$$\hat{\mathbf{H}}_{\text{cascade}} = \mathbf{H}_{BS-RIS} \Phi \mathbf{H}_{RIS-User} + \mathbf{E},$$

where \mathbf{E} captures the estimation error. - *Decomposition via Sparse Recovery:* Techniques like compressive sensing separate the individual components of the cascaded channel.

4. **Handling CSI Imperfections:** To account for imperfect CSI, the system incorporates the following strategies:

- Robust optimization techniques consider probabilistic CSI error bounds, optimizing beamforming vectors and RIS phase shifts for worst-case scenarios. The error model is represented as:

$$\hat{\mathbf{H}} = \mathbf{H} + \mathbf{E}, \quad \|\mathbf{E}\| \leq \varepsilon.$$

- The DDPG algorithm dynamically adapts to CSI imperfections, leveraging real-time learning to adjust configurations based on observed system performance.
- A trade-off is maintained between the pilot overhead for CSI acquisition and system performance, ensuring a balance between accuracy and efficiency.

2.2 Problem Formulation

The optimization problem for the RIS-aided 6G communication system seeks to maximize the secrecy rate R_S , defined as the difference between the achievable rates at Bob (legitimate user) and Eve (eavesdropper):

$$R_S = [R_B(\mathbf{w}, \Phi) - R_E(\mathbf{w}, \Phi)]^+. \quad (6)$$

The goal is to increase R_S by jointly enhancing the beamforming vector \mathbf{w} and RIS phase shift matrix Φ :

$$\max_{\mathbf{w}, \Phi} R_S(\mathbf{w}, \Phi), \quad (7)$$

subject to the following practical constraints:

Constraints 1. Transmit Power Constraint: Overall power of the beamforming vector must not exceed the set maximum limit:

$$\|\mathbf{w}\|^2 \leq P_{\max}. \quad (8)$$

2. RIS Phase Shift Constraints: Each RIS element's phase shift must be within the feasible range:

$$\Psi_i \in [0, 2\pi], \quad \forall i \in \{1, 2, \dots, N\}, \quad (9)$$

and for quantized phase shifts:

$$\Psi_i \in \{0, \Delta, 2\Delta, \dots, 2\pi - \Delta\}, \quad \Delta = \frac{2\pi}{Q}. \quad (10)$$

3. Imperfect CSI: Robustness to CSI estimation errors is incorporated as:

$$\Pr(\|\mathbf{h} - \hat{\mathbf{h}}\| \leq \varepsilon) \geq \delta, \quad (11)$$

where $\hat{\mathbf{h}}$ is the estimated channel, ε the error tolerance, and δ the confidence level.

4. Energy Efficiency Constraint: The system ensures minimum energy efficiency:

$$\eta = \frac{R_B(\mathbf{w}, \Phi)}{\|\mathbf{w}\|^2} \geq \eta_{\min}, \quad (12)$$

where η_{\min} is the threshold.

5. Multi-Objective Trade-Off: A balance between secrecy and communication quality is captured as:

$$\lambda_1 R_S(\mathbf{w}, \Phi) + \lambda_2 R_B(\mathbf{w}, \Phi), \quad (13)$$

where λ_1 and λ_2 control the relative importance of secrecy and quality.

Problem Statement: Incorporating these constraints, the extended optimization problem is expressed as:

$$\max_{\mathbf{w}, \Phi} \lambda_1 \cdot [R_B(\mathbf{w}, \Phi) - R_E(\mathbf{w}, \Phi)]^+ + \lambda_2 \cdot R_B(\mathbf{w}, \Phi), \quad (14)$$

subject to the transmit power constraint (8), RIS phase shift constraints (9) and (10), imperfect CSI robustness (11), and energy efficiency requirements (12).

This optimization framework integrates practical considerations, ensuring efficient and secure RIS-aided operation under real-world conditions while addressing performance and hardware limitations.

3 DRL Based Methodology for Secure Communication

DRL combines principles from reinforcement learning and deep learning to tackle complex decision-making challenges. By exploiting the analytical potential of deep neural networks, DRL empowers agents to navigate and adapt within intricate and dynamic environments, enabling effective decision-making based on vast and diverse datasets. This approach offers a robust framework for addressing tasks requiring adaptive and data-driven solutions in secure communication systems [35].

The foundation of DRL lies in training agents to determine optimal decisions by learning from iterative process experiences. Actors interact with their surroundings by observing its state, executing actions and receiving responses through rewards which guide their learning process. The goal is to formulate a policy that ensures the maximum total reward accumulated over time [36]. What distinguishes DRL from traditional reinforcement learning approaches is its utilization of deep learning. This integration enables the direct processing of raw sensory inputs, such as visual content or sensory information, excluding relying on physically engineered attributes. Deep neural networks act as powerful estimators, either for policies that identify optimal actions in given situations or for value functions that estimate expected rewards derived from current states and actions [37]. DRL offers notable benefits: Well-suited for managing intricate environments with extensive, high-dimensional state spaces, it is ideal for real-world use cases.

Additionally, agents using DRL leverage the generalization ability of neural networks to apply learned knowledge effectively in novel and unseen scenarios.

The capability for end-to-end learning from raw data to actions simplifies the overall learning process.

It has ability to adapt and scale according to problem size and complexity enables its application across a wide range of tasks, including mastering complex games and controlling autonomous vehicles.

Its ability to adapt and scale according to problem size and complexity enables its application across a wide range of tasks, including mastering complex games and controlling autonomous vehicles.

Nonetheless, challenges exist within DRL. DRL often requires substantial amounts of training data and can be susceptible to overfitting. It may also encounter challenges related to stability and reliability during the learning process, particularly in scenarios where rewards are rare or come with a delay. Despite these challenges, DRL marks a transformative advancement in artificial intelligence, providing a robust toolkit for addressing diverse decision-making problems in complex and ever-changing scenarios.

3.1 MDP

A Markov Decision Process (MDP) serves as a mathematical framework to model decision-making in situations where outcomes are influenced by both the decision-maker's actions and elements of randomness. The key elements of an MDP, along with their mathematical formulations, are described as follows:

- **State Space (S):** The state space S defines the collection of all possible states within the environment. Each state $s \in S$ represents the condition or configuration of the environment at a specific moment in time.
- **Action Set (A):** The action set A includes all the actions that the agent is able to take. For a particular state s the available actions are denoted as $a \in A(s)$. The action set may vary depending on the state as indicated by the notation $A(s)$.
- **Transition Dynamics (P):** The transition dynamics are characterized by the probability function $P(s' | s, a)$, which specifies the likelihood of transitioning to state s' from state s after performing action a . This function captures both deterministic and stochastic elements of the system dynamics:

$$P(s' | s, a) = \Pr(S_{t+1} = s' | S_t = s, A_t = a), \quad (15)$$

where s' represents the state at time $t + 1$, s is the state at time t , and a is the action taken.

- **Reward Function (R):** The reward function $R(s, a, s')$ or $R(s, a)$ defines the immediate reward obtained after transitioning from state s to state s' as a result of action a . It reflects the benefit or cost linked to the transition:

$$R(s, a, s') = \mathbb{E}[R_{t+1} | S_t = s, A_t = a, S_{t+1} = s'], \quad (16)$$

where R_{t+1} is the reward received at time $t + 1$.

- **Discount Factor (γ):** The discount factor $\gamma \in [0, 1]$ determines the weight given to future rewards. A higher value of γ (close to 1) means the agent prioritizes long-term rewards, while a lower value favors immediate rewards. The discount factor ensures convergence of the total reward and accounts for the uncertainty in future outcomes.

MDP is formally represented by (S, A, P, R, γ) . The agent goal is to determine an optimal policy π^* that maximizes the expected sum of discounted future rewards. A policy π maps states to actions and is denoted as:

$$\pi(a | s) = \Pr(A_t = a | S_t = s).$$

The objective is to find the optimal policy π^* that maximizes the anticipated cumulative reward G_t beginning from a given state:

$$\pi^* = \arg \max_{\pi} \mathbb{E}[G_t | S_t = s, \pi], \quad (17)$$

where the total discounted reward G_t from time t onward is defined as:

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}. \quad (18)$$

Solving an MDP involves identifying the optimal policy π^* that optimizes the expected cumulative rewards. Common approaches for solving MDPs, such as Q-Learning, Policy Iteration and Value Iteration, work by iteratively improving the policy or updating state and action values using the MDP's defined components.

3.2 Proposed DDPG Model

We apply the Markov Decision Process (MDP) framework to tackle the optimization hurdles in ensuring secure wireless communication systems, including both direct and backscatter communication. MDPs are well-suited for scenarios where current decisions—such as selecting the beamforming vector and RIS phase

shifts—impact future states and rewards. This approach considers long-term consequences for the system's performance and secrecy, unlike straightforward optimization approaches that prioritize immediate results.

To ensure scalability in large-scale RIS deployments, we utilize DDPG, a RL-based framework that excels in managing large action spaces, such as those present in RIS systems with a high number of reflecting elements. DDPG efficiently adapts the system to dynamically fine-tune the phase shift of RIS and beamforming vectors, balancing the trade-off between communication quality and secrecy. It also leverages actor-critic networks, where the actor determines the optimal actions based on the current state, while the critic evaluates these actions by estimating their Q-values, representing the expected rewards.

This reinforcement learning-based approach enables the system to learn and adjust configurations in real-time, making it computationally feasible for large-scale environments, especially an increase in the number of RIS elements. Moreover, parallelization and distributed learning techniques can further enhance the scalability and computational efficiency, allowing the system to handle complex, large-scale RIS deployments in real-world scenarios.

Actor Network: In our secure communication framework, the actor network defines the policy by linking the current environmental state \mathbf{s}_t to a group of actions \mathbf{a}_t , which includes the RIS phase shifts and beamforming vector elements. The policy is given as:

$$\mathbf{a}_t = \pi(\mathbf{s}_t | \theta^\pi), \quad (19)$$

where \mathbf{a}_t is the action vector at time t , \mathbf{s}_t represents the current environmental state information along with the current vector for beamforming and θ^π represents the actor network variables.

The action vector \mathbf{a}_t is structured such that the first N elements correspond to the RIS phase shifts $\Phi_t = \text{diag}(e^{j\Psi_1}, \dots, e^{j\Psi_N})$. These Phase shifts are normalized to fall within the specified range. $[-\pi, \pi]$ to ensure valid values:

$$\Psi_i \in [-\pi, \pi], \quad \forall i \in \{1, 2, \dots, N\}. \quad (20)$$

The remaining elements of \mathbf{a}_t correspond actual and virtual elements associated with the vector components of beamforming \mathbf{w}_t , which must satisfy the transmit power constraint:

$$\|\mathbf{w}_t\|^2 \leq P_{\max}. \quad (21)$$

To ensure compliance, the beamforming vector is normalized as:

$$\mathbf{w}_t \leftarrow \frac{\mathbf{w}_t}{\max(1, \|\mathbf{w}_t\|/\sqrt{P_{\max}})}. \quad (22)$$

By leveraging this approach, the agent can flexibly fine-tune the RIS phase adjustments and beamforming configuration according to the environment, optimizing transmission efficiency and ensuring secure communication.

Critic Network: The role of the critic network is to evaluate the actions generated by the actor network by estimating their Q-values. A Q-value signifies the expected cumulative reward for executing an action \mathbf{a}_t in the state \mathbf{s}_t and continuing to follow the policy π . The Q-value function is given as:

$$Q(\mathbf{s}_t, \mathbf{a}_t | \theta^Q) = \mathbb{E} \left[r_t + \gamma Q(\mathbf{s}_{t+1}, \pi(\mathbf{s}_{t+1} | \theta^\pi) | \theta^{Q'}) \mid \mathbf{s}_t, \mathbf{a}_t \right], \quad (23)$$

where θ^Q defines the parameters of the critic network, while $\theta^{Q'}$ represents the targeted critic network characteristics. The target critic network improves training stability by providing a consistent reference for updating Q-values.

The state s_t consists of the CSI $\mathbf{H}_{TI}, \mathbf{h}_{IB}, \mathbf{H}_{IE}, \mathbf{H}_{CE}, \mathbf{h}_{TB}$, along with the current beamforming vector \mathbf{w}_t and RIS phase shifts Φ_t . The action \mathbf{a}_t encompasses the phase adjustments of RIS and beamforming vector elements.

Reward Function: The immediate reward at time t , denoted by r_t , indicates the system, s secrecy performance, RIS phase stability, and energy efficiency. It is defined as:

$$r_t = \alpha \cdot [R_B(\mathbf{w}_t, \Phi_t) - R_E(\mathbf{w}_t, \Phi_t)]^+ - \beta \cdot \|\Phi_t\|_{\text{variation}}^2 + \eta \cdot \frac{R_B(\mathbf{w}_t, \Phi_t)}{P_{\text{used}}}, \quad (24)$$

where

- $R_B(\mathbf{w}_t, \Phi_t)$ and $R_E(\mathbf{w}_t, \Phi_t)$ are the achievable rates at Bob (legitimate user) and Eve (eavesdropper), respectively.
- $\|\Phi_t\|_{\text{variation}}^2$ penalizes large variations in RIS phase shifts, ensuring stable configuration updates.
- $P_{\text{used}} = \|\mathbf{w}_t\|^2$ denotes the total power used.
- α, β, η are weighting parameters that balance the trade-offs between secrecy, stability, and efficiency.

Training Process: The DDPG algorithm iteratively trains the actor and critic networks using experience replay and target networks to ensure stable learning.

Transitions (s_t, a_t, r_t, s_{t+1}) are stored in a replay buffer \mathcal{B} . During training, a minibatch of transitions is sampled from \mathcal{B} to break correlations between consecutive steps and improve learning stability.

The critic network is optimized by minimizing the following loss function:

$$L = \frac{1}{N} \sum_i (y_i - Q(s_i, a_i | \theta^Q))^2, \quad (25)$$

where $y_i = r_i + \gamma Q'(s_{i+1}, \pi'(s_{i+1} | \theta^{\pi'}) | \theta^{Q'})$ represents the target Q-value, calculated with the help of the target networks.

The actor network is updated using the policy gradient:

$$\nabla_{\theta^\pi} J \approx \frac{1}{N} \sum_i \nabla_a Q(s, a | \theta^Q) \Big|_{s=s_i, a=\pi(s_i)} \nabla_{\theta^\pi} \pi(s | \theta^\pi) \Big|_{s_i}. \quad (26)$$

Target networks are updated slowly to ensure stability:

$$\theta^{\pi'} \leftarrow \tau \theta^\pi + (1 - \tau) \theta^{\pi'}, \quad (27)$$

$$\theta^{Q'} \leftarrow \tau \theta^Q + (1 - \tau) \theta^{Q'}, \quad (28)$$

where $\tau \in (0, 1)$ is the soft update coefficient.

3.3 Computational Complexity Analysis

The computational cost of the DDPG algorithm in RIS optimization is dictated by the architecture of the actor and critic networks, the scale of the state and action spaces (which are proportional to the number of RIS elements N), and the minibatch size B during the training process.

The actor network maps the current state s_t , which includes the channel state information (CSI) and RIS configuration states, to an action a_t , representing the adjustments in RIS phases (N elements) and vector components of beamforming. The critic network evaluates the Q-value of the selected action, providing feedback for policy improvement.

The total complexity for one training iteration can be expressed as:

$$\mathcal{O}(B \cdot l \cdot h^2) + \mathcal{O}(B \cdot N), \tag{29}$$

where l is the number of layers in the neural networks, h is the number of neurons per layer, B is the minibatch size, and N is the number of RIS reflective elements.

Scalability with N : The complexity grows linearly with N because the action space dimension is proportional to N . The phase shifts of each RIS element must be optimized, increasing the input size of the networks and the number of optimization variables. Despite this linear scaling, the use of parallelized computations in GPUs and efficient sampling in DDPG ensures scalability for practical deployments with RIS sizes of up to 200–300 elements.

Practical Considerations: Compact neural network designs with fewer layers and neurons (l, h) reduce overhead while maintaining performance. Feature engineering or clustering techniques can also group RIS elements to simplify control in large-scale scenarios.

The scalability of the DDPG algorithm makes it suitable for RIS-aided 6G networks, even as the number of reflective elements increases.

3.4 Adaptive Learning for Enhanced Communication

To enhance security DDPG approach is used by the system for communication that iteratively updates the actor, critic networks influenced by the feedback from the environment, enabling continuous improvement of the policy.

The parameters of critic network's θ^Q improved by minimizing loss function L , which measures the error in the target Q-values and predicted Q-values computed from rewards and the target critic network's output:

$$L = \frac{1}{N} \sum_i (y_i - Q(s_i, a_i | \theta^Q))^2. \tag{30}$$

Here, y_i is the Q-value of target, defined as the reward plus the discounted Q-value of the subsequent state-action combination as estimated by the target critic network.

The actor network updates its parameters θ^π to select actions that yield larger Q-values as evaluated by the network of critic. The actor's objective function J gradient is approximated through:

$$\begin{aligned} \nabla_{\theta^\pi} J \approx \frac{1}{N} \sum_i \nabla_{\mathbf{a}} Q(\mathbf{s}, \mathbf{a} | \theta^Q) \Big|_{\substack{\mathbf{s}=\mathbf{s}_i \\ \mathbf{a}=\pi(\mathbf{s}_i)}} \\ \cdot \nabla_{\theta^\pi} \pi(\mathbf{s} | \theta^\pi) \Big|_{\mathbf{s}_i}. \end{aligned} \tag{31}$$

This iterative training process aligns the actor's policy with the objective of maximizing rewards, improving both the efficiency and security of wireless communication.

3.5 State Space, Action

In the proposed framework for secure data transmission utilizing RIS-assisted backscatter technology, the Markov Decision Process (MDP) is modeled to optimize both transmission efficiency and security. The components of the MDP, namely the state space, action space, and reward mechanism, are described as follows:

State Space: The state at time t , denoted as s_t , encapsulates the real-time conditions of the environment, including channel state information (CSI) for all relevant communication links. It is formally represented as:

$$s_t = \{\mathbf{H}_{CI}, \mathbf{H}_{TI}, \mathbf{h}_{CB}, \mathbf{h}_{IB}, \mathbf{H}_{IE}, \mathbf{H}_{CE}, \mathbf{h}_{TB}\}, \quad (32)$$

where $\mathbf{H}_{CI} \in \mathbb{C}^{N \times L}$ represents the CSI from the Base Station (BS) to the RIS, $\mathbf{H}_{TI} \in \mathbb{C}^{N \times M}$ denotes the CSI from the transmitter to the RIS, and $\mathbf{h}_{CB} \in \mathbb{C}^{M \times 1}$ represents the CSI for the direct link between the BS and Bob (the legitimate user). The CSI for the RIS to Bob link is denoted by $\mathbf{h}_{IB} \in \mathbb{C}^{N \times 1}$, while $\mathbf{H}_{IE} \in \mathbb{C}^{N \times M'}$ represents the RIS to Eve (eavesdropper) link. Additionally, $\mathbf{H}_{CE} \in \mathbb{C}^{M \times M'}$ describes the direct channel from the BS to Eve, and $\mathbf{h}_{TB} \in \mathbb{C}$ captures the direct link from the Backscatter Device (BD) to Bob. These parameters collectively define the current state of the network, enabling informed decision-making for secure communication.

Action Space: The action a_t is the decision variable at time t and consists of two components: the beamforming vector \mathbf{w}_t and the RIS phase shift matrix Φ_t . The action is determined as:

$$a_t = (\mathbf{w}_t, \Phi_t) = \pi(s_t | \theta_\pi) + \mathcal{N}_t, \quad (33)$$

where $\pi(s_t | \theta_\pi)$ represents the actor network's policy function, mapping the current state from s_t to action a_t and \mathcal{N}_t is the introduced during training to promote diverse actions.

The action space is subject to the following constraints. The beamforming vector must satisfy the transmit power constraint:

$$\|\mathbf{w}_t\|^2 \leq P_{\max}, \quad (34)$$

where P_{\max} denotes the maximum allowable transmit power. The RIS phase shifts are constrained as:

$$\Psi_i \in [0, 2\pi], \quad \forall i \in \{1, 2, \dots, N\}, \quad (35)$$

where Ψ_i represents the phase shift of the i -th RIS element. These constraints ensure that the actions taken are both physically feasible and compliant with the system's hardware limitations.

4 Results

In Fig. 2, the "Episode Rewards Over Time" graph illustrates the progression of the DDPG algorithm in optimizing RIS configurations for secure communication in a 6G IoT network (Algorithm 1). Initially, during the exploration phase (0 to 250 episodes), rewards remain low as the algorithm evaluates diverse strategies. This is followed by a significant increase in rewards (250 to 750 episodes), signaling the identification and exploitation of effective actions for improving security. Finally, the rewards stabilize (750 to 2000 episodes), indicating policy convergence and consistent performance. The results highlight the DDPG algorithm, s ability to dynamically adapt and optimize RIS configurations, reinforcing its utility in enhancing secure communication within the 6G IoT environment.

Algorithm 1: DDPG for joint beamforming and RIS phase shift optimization in secure communication

Data: Input: Channel matrices \mathbf{H}_{TI} , \mathbf{h}_{IB} , \mathbf{H}_{IE} , \mathbf{h}_{TB} , \mathbf{H}_{CE} , maximum power P_{\max} , phase shift constraints $\Psi_i \in [0, 2\pi]$

Data: Parameters: Learning rates α_π , α_Q , discount factor γ , soft update coefficient τ , replay buffer size N , exploration noise \mathcal{N}_t

Initialization:

Initialize actor network $\pi(s|\theta^\pi)$ and critic network $Q(s, a|\theta^Q)$ with random weights;

Initialize target networks: $\pi'(s|\theta^{\pi'}) \leftarrow \pi(s|\theta^\pi)$, $Q'(s, a|\theta^{Q'}) \leftarrow Q(s, a|\theta^Q)$;

Initialize replay buffer \mathcal{B} ;

foreach episode do

Initialize exploration noise \mathcal{N}_t (e.g., Ornstein-Uhlenbeck or Gaussian noise);

Receive initial state s_1 (e.g., channel states and initial beamforming vector and RIS phase shifts);

foreach step in episode do

Action Selection: Select action $a_t = (\mathbf{w}_t, \Phi_t)$:

$$a_t = \pi(s_t|\theta^\pi) + \mathcal{N}_t$$

Execute action a_t and observe:

- New state s_{t+1} (e.g., updated CSI, beamforming, and RIS settings)
- Immediate reward:

$$r_t = \alpha \cdot [R_B(\mathbf{w}_t, \Phi_t) - R_E(\mathbf{w}_t, \Phi_t)]^+ - \beta \cdot \|\Phi_t\|_{\text{variation}}^2 + \eta \cdot \frac{R_B(\mathbf{w}_t, \Phi_t)}{\|\mathbf{w}_t\|^2};$$

Store transition (s_t, a_t, r_t, s_{t+1}) in \mathcal{B} ;

Minibatch Sampling: Sample N transitions (s_i, a_i, r_i, s_{i+1}) from \mathcal{B} ;

Compute target Q-value for each transition:

$$y_i = r_i + \gamma Q'(s_{i+1}, \pi'(s_{i+1}|\theta^{\pi'})|\theta^{Q'}).$$

Critic Update: Update critic network θ^Q by minimizing the loss:

$$L = \frac{1}{N} \sum_i (y_i - Q(s_i, a_i|\theta^Q))^2.$$

Actor Update: Update actor network θ^π using policy gradient:

$$\nabla_{\theta^\pi} J \approx \frac{1}{N} \sum_i \nabla_a Q(s, a|\theta^Q) \Big|_{s=s_i, a=\pi(s_i)} \nabla_{\theta^\pi} \pi(s|\theta^\pi) \Big|_{s_i}.$$

Target Network Update: Softly update target networks:

$$\theta^{\pi'} \leftarrow \tau \theta^\pi + (1 - \tau) \theta^{\pi'};$$

$$\theta^{Q'} \leftarrow \tau \theta^Q + (1 - \tau) \theta^{Q'}.$$

Ensure Action Constraints:

- *Beamforming Power Constraint:* Normalize \mathbf{w}_t :

$$\mathbf{w}_t \leftarrow \frac{\mathbf{w}_t}{\max(1, \|\mathbf{w}_t\|/\sqrt{P_{\max}})}$$

- *RIS Phase Shift Range:* Clamp RIS phases:

$$\Psi_i \leftarrow \text{mod}(\Psi_i, 2\pi), \quad \forall i \in \{1, 2, \dots, N\}.$$

end

Decay exploration noise \mathcal{N}_t for next episode;

end

In Fig. 3, the ‘‘Secrecy Rate vs. Transmit Power’’ graph compares the proposed DDPG-based approach with the Alternate Optimization method and a system excluded RIS. The results demonstrate that the DDPG approach consistently delivers the high secrecy rate, with a steeper improvement as transmit power increases, highlighting its efficiency in optimizing RIS configurations for secure communication. In contrast,

the AO method delivers moderate performance, while the system without RIS performs poorly due to its inability to enhance signal security. This demonstrates the DDPG approach's superiority in achieving higher secrecy rates, particularly at higher transmit powers, reinforcing its applicability for secure and efficient 6G IoT networks.

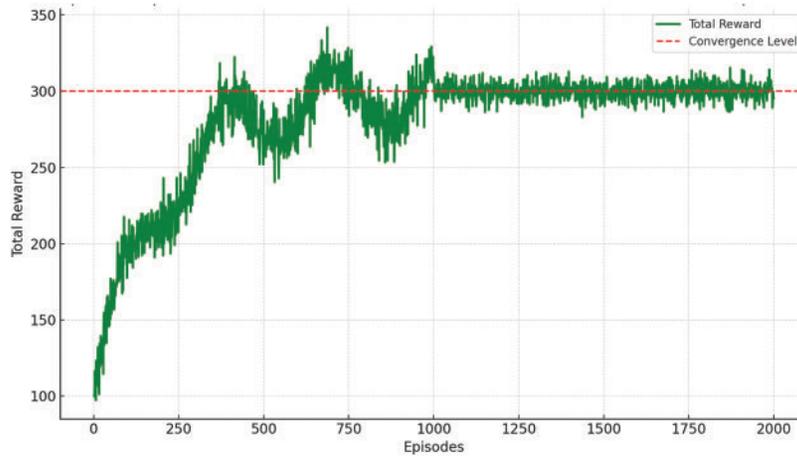


Figure 2: Episode vs. rewards

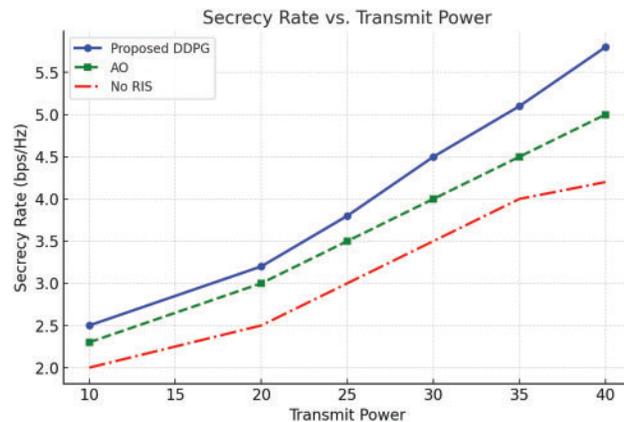


Figure 3: Secrecy rate vs. transmit power

In Fig. 4, the impact of different learning rates on the DDPG algorithm, its performance in optimizing RIS configurations for secure communication is illustrated. A learning rate of 0.0001 achieves the highest average reward (17 to 18) with stable convergence, showcasing its effectiveness in balancing exploration and exploitation. Conversely, a learning rate of 0.001 stabilizes at a moderate reward level (12 to 14), while 0.00001 exhibits the lowest rewards (9 to 10) and higher variability, indicating slower and less reliable learning. These results emphasize the critical role of learning rate selection in enhancing the stability and efficiency of the DDPG algorithm for RIS-aided secure communication in 6G IoT networks. Further exploration of learning rates could refine these insights, enabling more robust algorithm design for secure communication systems.

Fig. 5 evaluates the scalability of the proposed DDPG algorithm by analyzing the secrecy rate dependent on the number of reflective elements of RIS. The results indicate a consistent increase in secrecy rate with

more RIS elements, demonstrating the method, s effectiveness in optimizing larger RIS configurations. The DDPG algorithm outperforms the Alternate Optimization (AO) and Deep Q-Network (DQN) methods, showcasing superior handling of the expanded action space introduced by larger RIS arrays. As the number of elements increases, the DDPG framework dynamically adjusts RIS phase shifts and beamforming parameters, achieving higher secrecy rates and enhanced communication security. Notably, even with 120 RIS elements, the algorithm maintains computational efficiency, emphasizing its practicality for large-scale 6G IoT deployments. These findings highlight the DDPG algorithm’s scalability and its suitability for real-world RIS-aided secure communication systems.

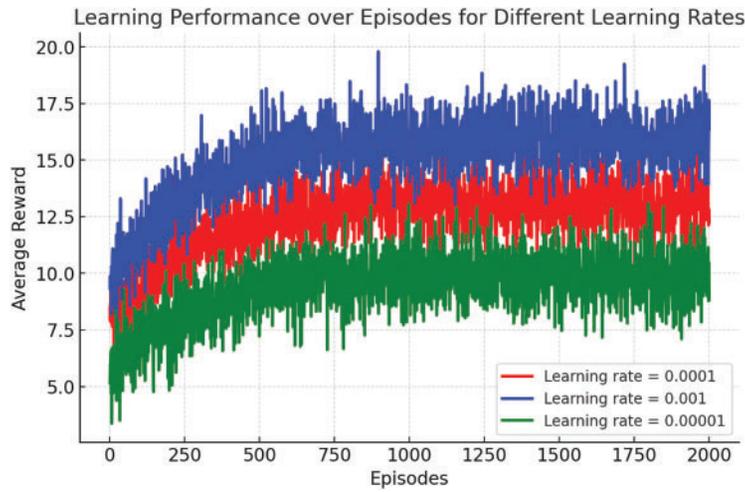


Figure 4: Learning performance for different rates

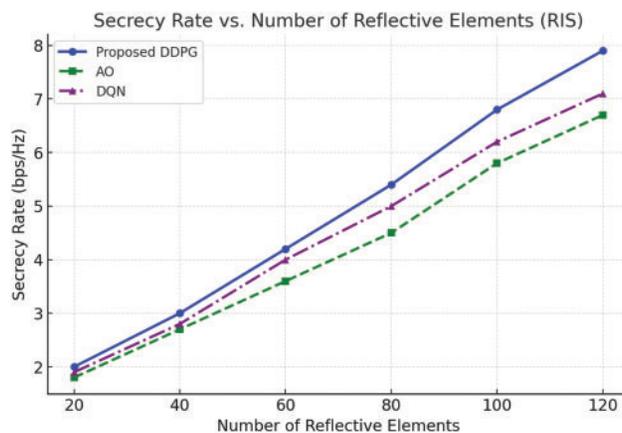


Figure 5: Secrecy rate vs. RIS elements over different techniques

Fig. 6 illustrates the critical role of RIS phase optimization in enhancing secrecy rates as the number of reflective elements increases. With optimized RIS phases the secrecy rate consistently improves, reaching approximately 7 bps/Hz for 100 elements. In contrast, non-optimized RIS phases show slower improvements achieving only around 3.5 bps/Hz. This stark contrast underscores the importance of phase optimization in maximizing system security and performance in RIS-aided networks. As the number of elements grows,

phase optimization becomes indispensable for achieving secure and efficient communication, particularly in advanced 6G IoT applications.

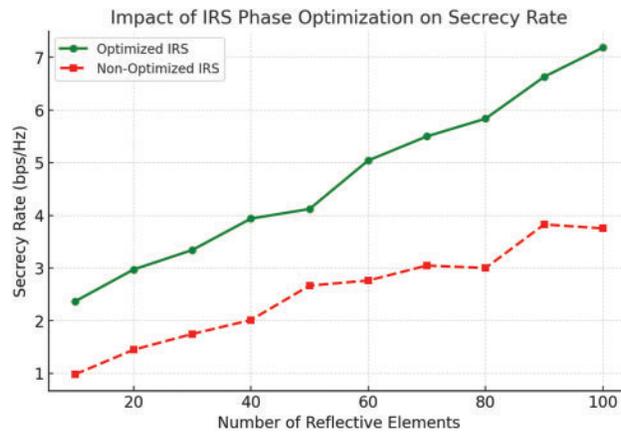


Figure 6: RIS phase optimization vs. non-optimized RIS

5 Conclusion

This study illuminates the transformative potential of RIS in revolutionizing 6G communication for IoT networks. By redefining participant R a passive IS as an active than reflector, we unlock pathways f new or c optimizing communication efficiency and spectral utilization rather ion in both backscatter and direct communication modes. A standout achievement of this research is the innovative integration of PLS with the DDPG algorithm, a powerful combination that enhances security and operational efficacy. Through meticulous strategic configurations of RIS and optimized power distribution, we achieve remarkable improvements in transmission efficiency while effectively addressing critical security challenges. Ultimately, this research sets pioneering benchmarks for the future of 6G communication protocols, laying a robust foundation for the creation of secure, interconnected, and highly efficient wireless systems. This study also addresses the computational complexity of the DDPG algorithm in the context of RIS optimization. By demonstrating that the algorithm's complexity scales linearly with the number of RIS elements, we establish its feasibility for real-world 6G deployments. The results validate that the performance gains outweigh the computational costs, ensuring scalability and efficiency for large-scale RIS-assisted networks. As we advance toward the next generation of networks, the methodologies and insights presented here will serve as a guiding light, shaping the future landscape of secure IoT communications.

Acknowledgement: Not applicable.

Funding Statement: The project was funded by the deanship of scientific research (DSR), King Abdukaziz University, Jeddah, under grant No. (G-1436-611-225). The authors, therefore acknowledge with thanks DSR technical and financial support.

Author Contributions: The authors declare their contributions to the paper as follows: study conception and design: Syed Zain Ul Abideen, Mian Muhammad Kamal; data collection: Mian Muhammad Kamal, Eaman Alharbi, Ashfaq Ahmad Malik; analysis and interpretation of results: Syed Zain Ul Abideen, Wadee Alhalabi, Muhammad Shahid Anwar; draft manuscript preparation: Mian Muhammad Kamal, Muhammad Shahid Anwar, Liaqat Ali. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data used in this study are available upon request, though some may be restricted due to privacy, confidentiality, or ethical concerns.

Ethics Approval: Not applicable.

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

References

1. Wu Q, Zhang R. Towards smart and reconfigurable environment: intelligent reflecting surface aided wireless network. *IEEE Commun Magaz.* 2019;58(1):106–12. doi:10.1109/MCOM.001.1900107.
2. Sheen B, Yang J, Feng X, Chowdhury MMU. A deep learning based modeling of reconfigurable intelligent surface assisted wireless communications for phase shift configuration. *IEEE Open J Commun Soc.* 2021;2:262–72. doi:10.1109/OJCOMS.2021.3050119.
3. Jiang T, Zhang Y, Ma W, Peng M, Peng Y, Feng M, et al. Backscatter communication meets practical battery-free Internet of Things: a survey and outlook. *IEEE Commun Surv Tutor.* 2023;25(3):2021–51. doi:10.1109/COMST.2023.3278239.
4. Sood S. An overview of backscatter communication technique for performing wireless sensing in green communication networks. In: 2023 International Conference on Power Energy, Environment & Intelligent Control (PEEIC); 2023; Greater Noida, India: IEEE. p. 7–11.
5. Tan F, Xu X, Chen H, Li S. Energy-efficient beamforming optimization for MISO communication based on reconfigurable intelligent surface. *Phys Commun.* 2023;57(3):101996. doi:10.1016/j.phycom.2022.101996.
6. Liu Y, Liu X, Mu X, Hou T, Xu J, Di Renzo M, et al. Reconfigurable intelligent surfaces: principles and opportunities. *IEEE Commun Surv Tutor.* 2021;23(3):1546–77. doi:10.1109/COMST.2021.3077737.
7. Huang C, Zappone A, Alexandropoulos GC, Debbah M, Yuen C. Reconfigurable intelligent surfaces for energy efficiency in wireless communication. *IEEE Transact Wireless Commun.* 2019;18(8):4157–70. doi:10.1109/TWC.2019.2922609.
8. Yang B, Cao X, Huang C, Guan YL, Yuen C, Di Renzo M, et al. Spectrum-learning-aided reconfigurable intelligent surfaces for “green” 6G networks. *IEEE Network.* 2021;35(6):20–6. doi:10.1109/MNET.110.2100301.
9. Ahmad S, Khan S, Khan KS, Naem F, Tariq M. Resource allocation for IRS-assisted networks: a deep reinforcement learning approach. *IEEE Commun Stand Magaz.* 2023;7(3):48–55. doi:10.1109/MCOMSTD.0002.2200007.
10. Zou Y, Zhu J, Wang X, Hanzo L. A survey on wireless security: technical challenges, recent advances, and future trends. *Proc IEEE.* 2016;104(9):1727–65. doi:10.1109/JPROC.2016.2558521.
11. Cui M, Zhang G, Zhang R. Secure wireless communication via intelligent reflecting surface. *IEEE Wirel Commun Lett.* 2019;8(5):1410–4. doi:10.1109/LWC.2019.2919685.
12. Naem F, Ali M, Kaddoum G, Huang C, Yuen C. Security and privacy for reconfigurable intelligent surface in 6G: a review of prospective applications and challenges. *IEEE Open J Commun Soc.* 2023;4:1196–217. doi:10.1109/OJCOMS.2023.3273507.
13. Peng Z, Zhang Z, Kong L, Pan C, Li L, Wang J. Deep reinforcement learning for RIS-aided multiuser full-duplex secure communications with hardware impairments. *IEEE Int Things J.* 2022;9(21):21121–35. doi:10.1109/JIOT.2022.3177705.
14. Gong S, Lu X, Hoang DT, Niyato D, Shu L, Kim DI, et al. Toward smart wireless communications via intelligent reflecting surfaces: a contemporary survey. *IEEE Commun Surv Tutor.* 2020;22(4):2283–314. doi:10.1109/COMST.2020.3004197.
15. Pan C, Ren H, Wang K, Kolb JF, Elkashlan M, Chen M, et al. Reconfigurable intelligent surfaces for 6G systems: principles, applications, and research directions. *IEEE Commun Magaz.* 2021;59(6):14–20. doi:10.1109/MCOM.001.2001076.
16. Yuan X, Zhang YJA, Shi Y, Yan W, Liu H. Reconfigurable-intelligent-surface empowered wireless communications: challenges and opportunities. *IEEE Wireless Commun.* 2021;28(2):136–43. doi:10.1109/MWC.001.2000256.

17. Alhammedi A, Shayea I, El-Saleh AA, Azmi MH, Ismail ZH, Kouhalvandi L, et al. Artificial intelligence in 6G wireless networks: opportunities, applications, and challenges. *Int J Intell Syst.* 2024;2024(1):8845070. doi:10.1155/2024/8845070.
18. Li J, Hong Y. Intelligent reflecting surface aided communication systems: performance analysis. In: 2021 IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC); 2021; Helsinki, Finland: IEEE. p. 519–24.
19. Kim SH, Park SY, Choi KW, Lee TJ, Kim DI. Backscatter-aided cooperative transmission in wireless-powered heterogeneous networks. *IEEE Transact Wireless Commun.* 2020;19(11):7309–23. doi:10.1109/TWC.2020.3010544.
20. Liang YC, Zhang Q, Wang J, Long R, Zhou H, Yang G. Backscatter communication assisted by reconfigurable intelligent surfaces. *Proc IEEE.* 2022;110(9):1339–57. doi:10.1109/JPROC.2022.3169622.
21. Zeng M, Bedeer E, Li X, Pham QV, Dobre OA, Fortier P, et al. IRS-empowered wireless communications: state-of-the-art, key techniques, and open issues. In: 6G wireless. UK: CRC Press; 2021. p. 15–38.
22. Wu H, Ren H, Pan C. Movable antenna-enabled RIS-aided integrated sensing and communication. *arXiv:240703228.* 2024.
23. Byun Y, Kim H, Kim S, Shim B. Channel estimation and phase shift control for UAV-carried RIS communication systems. *IEEE Transact Vehic Technol.* 2023;72(10):13695–700. doi:10.1109/TVT.2023.3274871.
24. Niu H, Lin Z, Chu Z, Zhu Z, Xiao P, Nguyen HX, et al. Joint beamforming design for secure RIS-assisted IoT networks. *IEEE Int Things J.* 2022;10(2):1628–41. doi:10.1109/JIOT.2022.3210115.
25. Do DT, Le AT, Ha NDX, Dao NN. Physical layer security for Internet of Things via reconfigurable intelligent surface. *Future Generat Comput Syst.* 2022;126(1):330–9. doi:10.1016/j.future.2021.08.012.
26. Nguyen KK, Masaracchia A, Sharma V, Poor HV, Duong TQ. RIS-assisted UAV communications for IoT with wireless power transfer using deep reinforcement learning. *IEEE J Select Top Signal Process.* 2022;16(5):1086–96. doi:10.1109/JSTSP.2022.3172587.
27. Faisal K, Choi W. Machine learning approaches for reconfigurable intelligent surfaces: a survey. *IEEE Access.* 2022;10(10):27343–67. doi:10.1109/ACCESS.2022.3157651.
28. Zhou H, Erol-Kantarci M, Liu Y, Poor HV. A survey on model-based, heuristic, and machine learning optimization approaches in RIS-aided wireless networks. *IEEE Commun Surv Tutor.* 2024;26(2):781–823.
29. Puspitasari AA, An TT, Alsharif MH, Lee BM. Emerging technologies for 6G communication networks: machine learning approaches. *Sensors.* 2023;23(18):7709. doi:10.3390/s23187709.
30. Pogaku AC, Do DT, Lee BM, Nguyen ND. UAV-assisted RIS for future wireless communications: a survey on optimization and performance analysis. *IEEE Access.* 2022;10(4):16320–36. doi:10.1109/ACCESS.2022.3149054.
31. Basharat S, Hassan SA, Pervaiz H, Mahmood A, Ding Z, Gidlund M. Reconfigurable intelligent surfaces: potentials, applications, and challenges for 6G wireless networks. *IEEE Wireless Commun.* 2021;28(6):184–91. doi:10.1109/MWC.011.2100016.
32. Yang H, Liu S, Xiao L, Zhang Y, Xiong Z, Zhuang W. Learning-based reliable and secure transmission for UAV-RIS-assisted communication systems. *IEEE Trans Wirel Commun.* 2024;23(7):6954–67.
33. Wu M, Guo K, Li X, Lin Z, Wu Y, Tsiftsis TA, et al. Deep reinforcement learning-based energy efficiency optimization for RIS-aided integrated satellite-aerial-terrestrial relay networks. *IEEE Trans Commun.* 2024;72(7):4163–78.
34. Guo K, Wu M, Li X, Song H, Kumar N. Deep reinforcement learning and NOMA-based multi-objective RIS-assisted IS-UAV-TNs: trajectory optimization and beamforming design. *IEEE Transact Intell Transport Syst.* 2023;24(9):10197–210. doi:10.1109/TITS.2023.3267607.
35. François-Lavet V, Henderson P, Islam R, Bellemare MG, Pineau J. An introduction to deep reinforcement learning. *Foundat Trends[®] Mach Learn.* 2018;11(3–4):219–354. doi:10.1561/22000000071.
36. Wang X, Wang S, Liang X, Zhao D, Huang J, Xu X, et al. Deep reinforcement learning: a survey. *IEEE Transact Neural Netw Learn Syst.* 2022;35(4):5064–78. doi:10.1109/TNNLS.2022.3207346.
37. Kaelbling LP, Littman ML, Moore AW. Reinforcement learning: a survey. *J Artif Intell Res.* 1996;4:237–85. doi:10.1613/jair.301.