Enhancing Secure Based on IRS Backscatter Communication

Yuanyuan Miao

Department of Electronic and Electrical Engineering

University of Sheffield

Supervisors: Prof. Jie Zhang, Dr. Salam Khamas

This thesis is submitted for the approval of the

Master of Philosophy

August 2025

This dissertation is dedicated to my parents. They gave me love, support and the courage to start over when facing difficulties.

Acknowledgements

I would like to express the deepest gratitude to my supervisor, Prof. Jie Zhang, for his support and guidances during my research. His insightful feedbacks and academic experience guided me in the direction of progress. His intelligence and integrity are the qualities I strive to emulate in my academic pursuits.

I am also very grateful to Dr. Sai Xu, Dr. Yu Shao and Dr. Jiliang Zhang for their guidance and help. I can always learn a lot from discussions with them, and their research enthusiasm also infects me.

Furthermore, I woulk like to thank all my colleagues in our group who gave me supports and encouragement.

Abstract

The development of wireless communication technology, where a large amount of private information is transmitted through open wireless channels, increases the risk of information leakage. Eavesdroppers may obtain sensitive information by capturing and analyzing transmitted signals, posing a threat to the confidentiality and integrity of communication. Therefore, preventing interference from eavesdropping signals is crucial for ensuring secure communication. This paper proposed a scheme to counter eavesdropper interference.

This paper proposed to combat active eavesdropping using intelligent reflecting surface (IRS) backscatter techniques. IRS is a revolutionary technology that significantly enhances wireless communication performance. By altering the amplitude and phase of incident signals, IRS can reconstruct the entire wireless channel environment, leading to improved communication efficiency and reliability. Backscatter technology relies on reflection to utilize existing radio waves in the environment for communication, without the need for active signal transmission, resulting in extremely low energy consumption. Due to the low power and difficulty in detecting backscattered signals, this technology has strong concealment and helps improve communication security. The combination of backscatter and IRS effectively enhances the security of wireless communication and reduces interference from eavesdroppers.

This paper provided a specific communication environment for the proposed scheme. The source (Alice) sends the confidential information to the intended user (Bob), while the eavesdropper (Willie) transmits a jamming signal to interrupt the transmission for more data interception. To enhance the secrecy, an IRS is deployed and connected with Alice through fiber to transform the jamming signal into the confidential signal by employing backscatter techniques. Based on the considered model, an optimization problem is formulated to maximize the signal-to-interference-plus-noise ratio (SINR) at Bob under the constraints of the transmit power at Alice, the reflection vector at the IRS, and the allowable maximum the SINR at Willie. To address the optimization problem, an alternate optimization algorithm is developed. The simulation results verify the achievable secrecy gain of the proposed scheme. The proposed scheme is effective in combating active eavesdropping. Furthermore, the deployment of large-scale IRS significantly enhances the secrecy rate.

Furthermore, this paper proposed a deep reinforcement learning (DRL) algorithm to solve the optimization problem. Specifically, this paper adopted the deep deterministic policy gradient(DDPG) algorithm to jointly optimize the transmit power at Alice and the reflection vector at the IRS. During the optimization process, this paper adjusted the state, actions, and rewards of the intelligent agent based on the system model. The simulation results show that the DRL based algorithm can significantly improve the communication security of the IRS-backscatter system.

List of Publications

Published

[1]. Miao Y, Shao Y, Zhang J, "IRS Backscatter-Based Secrecy Enhancement against Active Eavesdropping," *Electronics*, vol. 13, no. 265, Jan. 2024.

Table of contents

List of figures

List of tables

Abbreviations

IRS Intelligent reflecting surface

DDPG Deep Deterministic Policy Gradient

SISO Single-input Single-output

MISO Multiple-input Single-output

MIMO Multiple-input Multiple-output

AN Artificial Noise

CSI Channel State Information

TPC Transmit Power Control

RF Radio Frequency

NOMA Non-Orthogonal Multiple Access

IoT Internet of Things

DRL Deep Reinforcement Learning

SNR Signal-to-Noise Ratio

SINR Signal-to-Interference-plus-Noise Ratio

AO Alternation Optimization

SDP Semi-Definite Programs

AP Access Point

BS Base Station

BCD Block Coordinate Descent

MM Minimization Maximization

CW Continuous Wave

AmBC Ambient Backscatter Communication

BD Backscatter Devices

RL Reinforcement Learning

MDP Markov Decision Process

AC Actor-Critic

QCQP Quadratically Constrained Quadratic Programming

SVD Singular Value Decomposition

MRT Maximum Ratio Transmission

MSBE Mean-square Bellman Error

Chapter 1

Introduction

1.1 Research Background

1.1.1 IRS-assisted Communication System

As wireless communication continues to evolve, intelligent reflecting surfaces (IRS) are emerging as a key technology in the field due to their programmability, low power consumption, and cost-effectiveness. IRS is a plane composed of a large number of low-cost passive reflective elements. IRS can reconfigure the wireless propagation environment by intelligently adjusting and controlling the amplitude and phase shift of the reflecting elements [1]. Each reflection element can independently adjust the phase and amplitude of the reflected signal. By coordinating these units, IRS can intelligently control the reflection direction and characteristics of signals, thereby achieving dynamic optimization of the wireless communication environment [2]. Due to its passive reflection of incident signals in the environment, IRS excels in energy conservation and low power consumption, making it suitable for applications with high power consumption requirements [3]. IRS can be flexibly deployed in various environments, allowing it to expand coverage or enhance communication capabilities based on the specific needs of the communication system [4]. In brief, IRS can be applied in a wide range of wireless communication scenarios to enhance the communication performance in various ways.

In IRS assisted communication systems, IRS can enhance the required signal while suppressing interference signals by controlling the reflected phase. In [5], an IRS-assisted single-input single-output (SISO) broadcast channel is studied. By using iterative gradient descent method to calculate and find the optimal transmission rate, further maximizing the effective capacity. In [6], a multiple-input single-output (MISO) downlink multiuser communication system is considered, where the phase shifts of a single IRS and the transmit beamforming vector at the access point are optimized using a two-timescale beamforming optimization algorithm. An IRS-enhanced MISO system with reflection pattern modulation is proposed, in which the IRS is capable of configuring its reflection state to enhance the received signal power through analog beamforming while simultaneously conveying its own information via reflection modulation [7]. A single IRS assisted multiple-input multiple-output (MIMO) cognitive radio wireless communication system is proposed in [8], which maximizes the weighted total rate sum of secondary users at the receiving end by jointly optimizing the transmission precoding of the secondary user transmitter and the reflection phase of the IRS.

1.1.2 IRS-assisted Secure Communication

The primary wireless communication technologies employed to counter eavesdropping attacks include cooperative relaying, artificial noise (AN) injecting, and multi-antenna beamforming. However, these approaches often suffer from high energy consumption and optimization difficulties caused by the high correlation between legitimate and illegitimate links. IRS can effectively compensate for the shortcomings brought by the above-mentioned secure communication technologies, especially by optimizing the IRS reflection phase to enhance legitimate link signals and suppress illegal link signals in low-power situations, while reducing the difficulty of system optimization and improving the security rate of the entire communication system.

Cui et al. investigated how to maximize the security rate of a communication system when the transmission power is fixed and the reflection parameters set at the IRS are constrained [9]. Wang et al. considered using IRS to assist a MISO system for secure communication in the case of unknown eavesdropper's channel state information (CSI), by jointly optimizing the beamforming and IRS reflection phase at the transmitting end to minimize the transmission power at the transmitting end [10]. A MIMO wireless communication system assisted by an IRS is proposed, which maximizes the secure communication rate of the overall link by jointly optimizing the transmit power control (TPC), artificial noise covariance matrix, and IRS reflection phase at the transmitter [11]. These studies demonstrate that IRS can effectively improve the secure communication performance of various wireless communication systems.

1.1.3 Backscatter Secure Communication

In backscatter communication systems, a backscatter transmitter modulates and reflects received radio frequency (RF) signals to transmit data, rather than generating RF signals independently [12]. By utilizing the principle of backscattering, data transmission can be achieved by reflecting radio waves in the environment without the need for active signal transmission. This technology has unique advantages in ensuring communication security.

Because backscatter communication devices do not actively transmit signals and the reflected signal power is extremely low, it becomes challenging for eavesdroppers to detect meaningful signals. This significantly enhances the concealment of the communication.

The researchers of [13] examined a multi-label backscatter communication system in the presence of eavesdroppers, taking into account a realistic backscatter communication scenario where channel correlation might exist between the forward and backscatter links. A new optimization framework is provided to improve the physical layer security of non-orthogonal multiple access (NOMA) environment backscatter communication system [3]. This system model considers the simultaneous operation of NOMA internet of things (IoT) users and backscatter nodes in the presence of multiple eavesdroppers. The model aims to enhance link security by optimizing the reflection coefficient of the backscatter nodes and the transmission power of the base station.

1.1.4 IRS-assisted Communication System Based on DRL

Deep reinforcement learning (DRL) is a technique that utilizes deep learning and reinforcement learning algorithms to optimize decisions. Introducing DRL into IRS assisted communication systems can automatically optimize the control strategy of intelligent reflectors, improving the performance and security of the communication system.

Feng et al. proposed a DRL-based framework to address the non-convex optimization challenge caused by optimizing the passive phase shift of each IRS component to maximize the downlink received signal-to-noise ratio (SNR) [14]. In [15], an IRS assisted NOMA communication system was considered, using DDPG to jointly optimize the beamforming at the transmitter and the IRS reflection phase to maximize the communication rate at the receiver. Yang et al. considered an IRS-assisted communication system aimed at maximizing the secrecy rate for multiple legitimate users in the presence of multiple eavesdroppers within realistic time-varying channels. To address this, a novel DRL-based secure beamforming approach was proposed, enabling the system to achieve optimal beamforming policies against eavesdroppers in dynamic environments [16]. This study highlights that DRL offers significant advantages in solving non-convex optimization problems in complex and dynamic communication systems.

1.2 Research Motivations

IRS is a cost-effective passive equipment with low structural complexity, and is not or rarely equipped with RF link. The advantages of IRS can be introduced into secure communication. The joint optimization of transmission power and IRS reflection coefficient can improve the covert communication performance of the system when the eavesdropper's link statistics are available [17].

The combination of IRS and backscatter technology has been applied in various communication scenarios. Idrees et al. designed a noval scheme to improve the error rate performance of ambient backscattering by utilizing an IRS positioned in close proximity [18]. Backscatter technology can dynamically change the reflection path and modulation characteristics of signals, especially when combined with IRS, making the channel environment extremely complex and difficult to predict. In [19], a method of using IRS based backscatter communication system to resist interference is proposed. Through the modulation of IRS, the undesired signal is backscattered to enhance the user's receiving power, and the undesired signal is used as a resource to combat its own adverse effects

DRL can effectively solve highly complex problems involving IRSs, such as the joint optimization of phase shifts and other related parameters [20]. The study in [21], employed a DRL-based framework to jointly optimize IRS and reader beamforming in IRS-assisted ambient backscatter communication systems.

Motivated by the above research, we proposed a noval scheme that combines the IRS and backscatter to improve the performance of secure communication. We refer to this approach as Backcom-IRS. To maximize the signal-to-interference-plus-noise ratio (SINR) at the legitimate user, we jointly optimize the IRS reflection coefficient and the source beamforming vector. Due to the complexity of the optimization problem, we used two optimization methods: alternating optimization and DDPG algorithm.

1.3 Research Problem Statement, Aim and Objectives

1.3.1 Research Problem Statement

With the rapid development of wireless communication technologies and the widespread transmission of sensitive information through open wireless channels, communication systems face growing risks of active eavesdropping and jamming attacks. Traditional physical layer security technologies such as cooperative relaying, artificial noise injection, and multi-antenna beamforming are often limited by high energy consumption and complex optimization requirements.

IRS have emerged as a promising solution in this background to enhance wireless security by intelligently reconfiguring the wireless propagation environment. Simultaneously, backscatter communication offers low-energy, low-complexity communication that is inherently difficult to detect.

The challenge lies in how to effectively integrate IRS with backscatter communication to enhance physical-layer security, especially under active jamming by eavesdroppers. This involves a non-convex joint optimization problem of IRS reflection coefficients and transmit beamforming under power and secrecy constraints, which cannot be efficiently solved by conventional methods.

1.3.2 Research Aim and Objectives

The aim of this research is to design and evaluate a secure communication system that integrates IRS and backscatter communication to maximize the secrecy rate under active eavesdropping attacks, while maintaining system efficiency.

To achieve the research aim, the research pursues the following objectives.

1. Develop a theoretical model involving a legitimate user, a transmitter, an IRS and an active eavesdropper for a backscatter IRS-assisted communication system.

2. Mathematically formulate the secrecy rate maximization problem under active jamming, considering power constraints and signal-to-interference-plus-noise ratio thresholds for both legitimate and illegitimate receivers. Define the problem as a function of IRS phase shifts and transmitter beamforming vectors.

3. Design an Alternating Optimization algorithm to jointly optimize the source beamforming vector and IRS reflection coefficients. Conduct numerical simulations to validate the effectiveness of the proposed backscatter IRS-assisted schemes and compare performance against baseline methods

4. Develop a Deep Deterministic Policy Gradient algorithm to enable the system to autonomously learn optimal configurations for improved secrecy performance in dynamic wireless environments. Conduct extensive numerical simulations to validate the flexibility and effectiveness and of this algorithm.

1.4 Research Contributions

This thesis proposed a secure communication scheme that integrates backscatter technology with intelligent reflecting surfaces. The research focuses on optimizing the secure communication system by employing an alternation optimization (AO) algorithm and the DDPG algorithm.

Firstly, this thesis established a backcom-IRS secure communication system. The active eavesdropper continuously transmit interference signals to the system. The IRS leverages interference signals to enhance the desired signal through backscattering. This thesis maximized the SINR of the legitimate user by designing the beamforming vector of the source and the reflection coefficient of the IRS, while limiting the SINR of the eavesdropper. The optimization problem of the system model designed in this thesis involves coupling of optimization variables, making the optimization problem non convex and difficult to solve.

This thesis employed an alternating optimization algorithm to solve the non-convex optimization problem. The optimization problem is transformed into two convex subproblems, which are then optimized alternatively. Both of these subproblems are positive semi-definite programs (SDPs) that can be efficiently solved using existing convex optimization solvers.

Next, this thesis applied the DDPG algorithm to optimize the communication system designed in this research. DRL autonomously learns the optimal reflection coefficient settings for the IRS, maximizing the communication system's performance across various environments and conditions. The strategy produced by the DRL agent dynamically adjusts the reflection unit configurations of the IRS, enabling intelligent control of wireless signals.

1.5 Structure of the Thesis

The content of this thesis is organised as follows.

Chapter 2: Literature review

This chapter first reviews IRS-assisted secure communication. The advantages of low power consumption, low cost, and flexible deployment of IRS effectively improve the security performance of wireless communication systems. Then, the secure communication combining IRS and backscatter is introduced. This section primarily discusses the positive impact of backscatter technology in the field of communication security, particularly when it is combined with IRS. Finally, the relevant concepts of DRL are reviewed, along with the key components utilized in solving DRL problems, including reward, value function and Bellman equation.

Chapter 3: The System model and optimization problem formulation

In this chapter, an IRS-assisted secure communication model is designed to maximize the security performance of the entire communication system by optimizing the beamforming vector at the source and the reflection coefficient of the IRS.

Chapter 4: Alternating optimization algorithm to solve optimization problem This chapter proposes using alternation optimization algorithms to solve the system optimization problem. The AO algorithm decomposes the complex optimization problem into several smaller sub problems, and then alternately optimizes each sub problem. This algorithm can effectively solve the non convex optimization problem caused by the coupling of optimization variables in this system.

Chapter 5: DDPG algrithm to solve optimization problem

This section transforms the optimization problem into an reinforcement learning problem and solves it using the DDPG algorithm. DDPG adopts a deterministic strategy, which enables more stable training under a deterministic strategy. DDPG is particularly adept at handling reinforcement learning problems in continuous action spaces. It combines the advantages of strategy gradient and Q-learning

Chapter 6: Conclusions and Future work

This chapter provides a summary of the work presented in this thesis and outlines potential directions for future research.

Chapter 2

Literature Review

Overview

In this chapter, the review begins with IRS-assisted secure communication. Following that, the integration of IRS and backscatter for secure communication is discussed. Finally, the chapter delves into the key concepts of DRL algorithm.

2.1 The IRS-aided Secure Communication

With the rapid advancement of wireless communication in recent years, the demand for more efficient and advanced wireless communication technologies has intensified. IRS have emerged as a key technology in this field due to their low power consumption, low hardware complexity, and high degree of flexibility. IRS can significantly improve the performance of wireless communication by controlling the incident signal.

Figure 2.1 shows the architecture of IRS. IRS consists of three sub layers and one intelligent controller. The outermost side is the reflecting element layer, and a large number of reflecting elements made of digital metamaterials are printed on the two-dimensional medium plane with a certain law. These reflecting elements will independently adjust the phase and amplitude of the incident signal and then reflect it. The middle layer is a copper board, which is mainly used to avoid signal energy leakage. The innermost layer is the

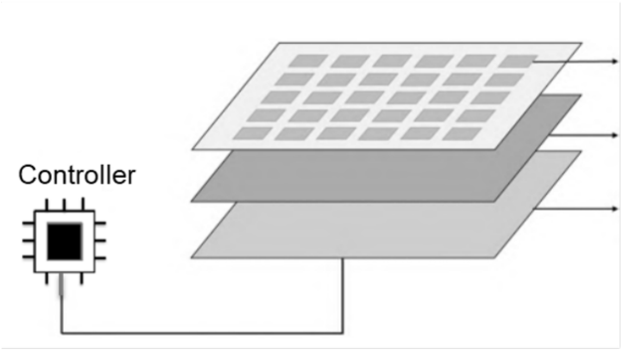

Fig. 2.1 IRS hardware architecture

control circuit board, which is responsible for adjusting the amplitude or phase shift of each reflecting element and is controlled by the controller.

IRS can dynamically adjust the path of reflected signals, which makes it difficult for eavesdroppers to obtain effective signals. The IRS reflected signal can be used to cancel the signal received by the eavesdropper from the base station or to enhance the signal received by the user. IRS increases the data rate of legitimate receivers and reduces the data rate of eavesdroppers. This can improve the difference between the two rates and effectively improve the system safety rate.

In [9], a communication system was designed where a multi antenna access point (AP) sends confidential messages to a single antenna user with a single antenna eavesdropper attempting to intercept them. The eavesdropper's channel in this system is not only stronger, but also highly correlated with legitimate communication channels. They proposed a strategy to maximize confidentiality by jointly optimizing the transmit beamforming at the AP and the reflected beamforming at the IRS and the proposed design significantly increases the secrecy rate compared to scenarios without using the IRS.

The objective of maximizing secrecy in IRS assisted multi antenna system presented a non convex optimization challenge in [22], which requires simultaneously satisfying the transmission power constraint at the source and the unity mode constraint imposed on the IRS phase shift. The solutions of the emission covariance of the source and the phase shift matrix of the IRS are implemented in closed form and semi closed form, respectively. In [2], IRS aided multi-antenna physical layer security is proposed. Chu et al. have designed a secure transmission power allocation and surface reflection phase shift scheme to minimize the transmission power under the confidentiality constraint of legitimate users.

A downlink MISO broadcasting system with multiple eavesdroppers was designed in [23], where the base station (BS) sends independent data streams to multiple legitimate receivers. By jointly optimizing the beamformer at BS and the reflection coefficient at IRS, while considering both continuous and discrete reflection coefficient constraints, the minimum confidentiality rate is maximized.

In order to maximize the secrecy rate of the IRS assisted secure wireless system in [24], the beamformer and IRS phase shift at the transmitter were jointly optimized. Two effective algorithms were developed based on block coordinate descent (BCD) and minimization maximization (MM) techniques to solve the non convex optimization problems of small-scale and large-scale IRS, respectively.

Considering an IRS assisted system, Dong et al. proposed an alternating optimization algorithm to jointly optimize the transmission covariance of the transmitter and the phase shift coefficient of the IRS [25]. The IRS assisted design achieved higher secrecy rate than other benchmark schemes.

Tang et al. designed an IRS-assisted secure transmission that leverages "directional reflection" to enhance physical layer security [26]. Directional reflection enhances the required signal while weakening the signal strength of any eavesdropper located in other directions, significantly improving confidentiality performance.

IRS can control a partial of reflection units to inject artificial noise or interference signals in the direction of eavesdroppers, effectively increasing the decoding difficulty of eavesdroppers and enhancing the physical layer security of the system.

In [27], a scenario has been proposed where multiple potential eavesdroppers with multiple antennas have incomplete known CSI. Under the constraints of transmission power and legal user service quality, they jointly solve the transmitted information beam, artificial noise and reflection coefficient to maximize the system's security rate.

Peng et al. studied the IRS assisted secure communication system in the presence of hardware damage to the IRS and transceiver [28]. In order to maximize the weighted minimum approximate traversal secrecy rate, the beamforming vector at the base station and the phase shift of the reflective elements at the RIS were jointly optimized.

2.2 The Combination of IRS and Backscatter in Secure Communication

Backscatter technology utilizes existing radio waves in the environment to reflect signals for communication. The characteristics of backscatter technology provide advantages for the physical layer security of communication systems [29]. Backscatter communication is generally categorized into three types: monostatic backscatter communication, bistatic backscatter communication, and ambient backscatter communication [30].

Backscatter technology communicates by reflecting radio waves in the environment, with extremely low signal power that is difficult for eavesdroppers to detect. This low-power characteristic poses a huge challenge for eavesdroppers when attempting to intercept and decode signals.

A signal strength balance scheduling scheme has been proposed in [31], which intentionally generates conflicting signals and uses them for communication between readers and tags to prevent eavesdroppers from distinguishing between required signal data and noise signals.

Due to the simple structure of backscatter tags, it is difficult to apply complex security techniques to backscatter networks [32, 33]. A noise injection approach aimed at enhancing security in backscatter networks is developed, the reader injects noise into the continuous wave (CW) signal by generating the CW signal and noise signal simultaneously [29] or by producing random CW signals [34]. These techniques protect the tag's signal by making it difficult for attackers to decode the signal without prior knowledge of the noise or random CW signals, thereby increasing the overall security of the backscatter communication system.

You et al. proposed a new protocol to streamline the communication process and enhance the security of data transmission between multiple tags and the reader [35]. By optimizing the number of tags and training symbols, maximizing the data rate gap between the reader and eavesdropper can improve security performance.

Combining traditional encryption techniques, backscatter communication can add an additional layer of encryption protection on the basis of physical layer security, ensuring the integrity and confidentiality of data during transmission. Ambient Backscatter Communication (AmBC) allows backscatter devices (BDs) to communicate by reflecting ambient RF signals while simultaneously harvesting energy. In [36], a novel physical layer key generation scheme designed for ultra-low-power AmBC systems was proposed.

The traditional backscatter communication security scheme is mainly based on lightweight symmetric cryptography [37]. Key generation typically requires two devices to send channel probing signals to each other in order to measure highly correlated channel characteristics at both ends [38].

Backscatter technology can dynamically change the reflection path and modulation characteristics of signals, especially when combined with IRS. The channel environment becomes extremely complex and difficult to predict, which increases the difficulty of eavesdropping and improves the security performance of the system.

2.3 Optimization Algorithms for Solving Nonconvex Problems in IRS Assisted Communication Systems

2.3.1 Traditional Optimization Algorithms

In IRS assisted wireless communication systems, non convex optimization problems are commonly present, such as joint optimization of transmit beamforming and IRS phase control matrix to maximize system rate [39] or minimize power [40]. These types of problems often become highly non convex due to coupling variables, unit mode constraints, and nonlinear objective functions. Traditional optimization algorithms can be used to solve these non convex optimization problems.

The alternating optimization algorithm decomposes the joint optimization problem into sub problems and iteratively optimizes them. This algorithm has a clear structure and is easy to implement. It is commonly used in optimizing the transmitter beamforming and

IRS reflection matrix. Semipositive definite relaxation transforms non convex unit module constraints into a semi positive definite matrix problem, which can be optimized using mature convex optimization tools such as CVX. This is commonly used to optimize the IRS reflection matrix. An alternative optimization algorithm and semidefinite programming relaxation for designing safe transmit power allocation and surface reflection phase shift are proposed in IRS aided secure transmission [2]. In [41], a suboptimal algorithman based on alternative optimization algorithm successive convex approximation, semidefinite relaxation, and manifold optimization was proposed to solve the non-convex problem. An alternative optimization algorithm utilizing successive convex approximation and semidefinite relaxation was proposed to solve trajectory and transmit power optimization problem for IRS assisted communication in [42]

Fractional Programming aims to maximize processing speed or energy efficiency in the form of class ratios. Fractional programming can transform non convex ratios into iterative convex problems. The typical algorithms for fractional programming include Dinkelbach algorithm and Quadratic Transform. This algorithm is often used for IRS assisted downlink rate or energy efficiency optimization. In [43], fractional Programming is proposed to find the optimal solution to the power allocation subproblem. Dinkelbach method was emploed to jointly select the best set of antennas and optimizes their beamforming [44].

Projected gradient method (PGM) is used for unit mode constrained optimization problems. This method makes gradient descent to the phase variable and projects it back to the unit mode set. A PGM-based algorithm was proposed to solve optimal IRS pattern matrix design focusing on the path angle estimation critical in mmWave communication [45].

2.3.2 The Deep Reinforcement Learning Algorithm

Although the optimization problem of IRS assisted systems can be solved by traditional optimization algorithms, many researchers choose DRL to solve non convex optimization problems in communication systems when the environment is dynamic and computational complexity is high [14, 15, 46]. DRL agent can solve a problem with a large dimensional state space and action space by learning to make an optimal decision through interacting with the environment, such as CSI, transmit power, interference [47]. Therefore, in existing studies [46, 48], model free DRL algorithms are envisioned as optimization solutions for dynamic uncertain environments.

The DDPG algorithm is a reinforcement learning (RL) algorithm based on deterministic policies. The DDPG algorithm is a DRL algorithm used to solve continuous state spaces and continuous action spaces, utilizing policy gradient theory and Actor-Critic model [49]. To solve the problem of maximizing the received signal-to-noise ratio in single user IRS assisted MISO systems, an algorithm based on DDPG was proposed and the results showed that the DDPG algotithm could achieve a better performance, as compared to the semidefinite relaxtion method [14]. In [20], a DDPG based algorithm was proposed to maximize the ergodic sum rate in a multiuser IRS assisted MISO system by jointly optimizing the transmit beamforming at the base station and the reflect beamforming at the IRS. These works verified that the DRL algorithms can solve optimization problems for various wireless communication systems effectively.

Reinforcement learning is the process of an intelligent agent continuously interacting with its environment and gaining experience, in order to improve its decisions based on the experience gained, and ultimately learn an optimal strategy. RL can learn and adjust strategies in dynamic and uncertain environments, adapting to changes in the environment.

DRL is a self-learning artificial intelligence algorithm based on Markov decision process (MDP). By using MDP, the optimization objective function can be considered as the reward, and the optimization variables can be used as actions, allowing the agent to continuously interact with the set environment and ultimately obtain the optimal strategy. This algorithm directly hands over the optimization problem to the agent, allowing it to learn in MDP and obtain an approximate optimal solution.

In the learning process, the agent is not told what actions to take, but constantly tries which actions can generate the maximum total reward. In most cases of RL, the actions taken by the agent not only affect the reward, but also affect the next state, which in turn affects the reward for each subsequent decision. Intelligent agents must try various actions

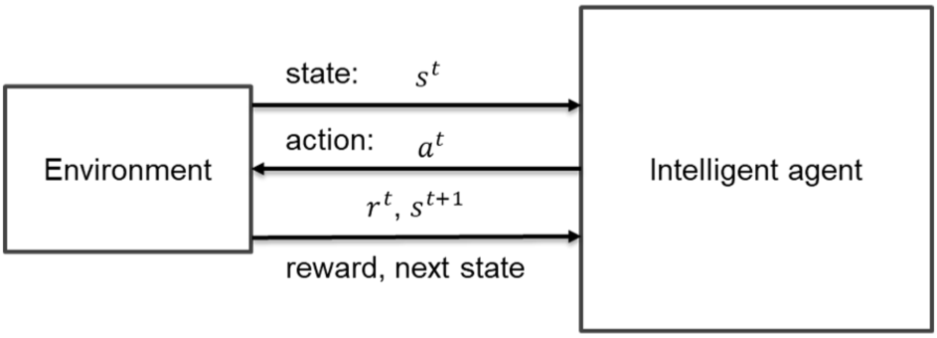

Fig. 2.2 MDP model

and gradually lean towards actions that appear to receive significant rewards. In a task, each action must be attempted multiple times to obtain a reliable estimate of its expected reward.

RL usually requires a large number of interactive samples, powerful computing resources, especially deep reinforcement learning, which has high training time and computational costs. The training process may experience instability, which may lead to the strategy falling into local optima or oscillations. Despite these drawbacks, RL has enormous potential in solving complex decision problems. With the continuous development of algorithms and the improvement of computing resources, reinforcement learning will play a more important role in wireless communication.

The MDP model is shown in the following figure 2.2.

For a finite MDP, at each time step t, the agent perceives the current environmental state $s^t \in \mathrm{S}$, and based on the state selection action $a^t \in \mathrm{A}(s)$. The environment will provide corresponding feedback based on the actions taken by the intelligent agent, $r^t \in \mathrm{R}$ quantify the quality of actions with rewards. Afterwards, the intelligent agent enters the next state s^{t+1}.

In finite MDP, the set of states S, the set of actions A and reward set R are all finite element sets. Define a probability function p to describe the finite MDP at each time step, at a given moment t,

$$p(s^{'},r|s,a) \doteq \Pr\{s^t = s^{'}, r^t = r|s^{t-1} = s, a^{t-1} = a\}. \tag{2.1}$$

This represents the probability that the intelligent agent will perform action a in state s, then receive reward r and enter the next state $s^{'}$ at the next moment. The probability function p provides a probability distribution for each state s and action a,

$$\sum_{s^{\prime}\in \mathrm{S}}\sum_{r\in \mathrm{R}} p(s^{'},r|s,a) = 1. \tag{2.2}$$

The probability function p provides a complete representation of the dynamic changes in the environment in which the entire agent is located. The probability of each current of state s^t and action a^t is determined only by its previous state s^{t-1} and action a^{t-1}, and is independent of earlier states and actions.

Reward, Value Function, and Bellman Equation

In RL, the ultimate aim of the agent is to maximize the total reward, which means that the agent does not need to maximize the current reward, but rather the long-term cumulative reward. Assuming that starting from time t, the reward received by the intelligent agent in the subsequent sequence is $r^t, r^{t+1}, r^{t+2}, \ldots$, the total reward G^t that the intelligent agent can receive at time t is

$$G^t \doteq r^t + \gamma r^{t+1} + \gamma^2 r^{t+2} + \cdots = \sum_{k=0}^{\infty} \gamma^k r^{t+k} \tag{2.3}$$

γ represents the reward discount coefficient, $0 \leq \gamma \leq 1$. The size of γ determines the extent to which future rewards affect the current cumulative rewards.

As γ increases, it means that in this task, the impact of future decisions on current decisions also becomes increasingly significant. Therefore, the rewards in a task with consecutive time steps are interrelated. When $\gamma = 0$, in this task, future decisions have no impact on current decisions. 2.3 can be transformed into the following equation,

$$\begin{aligned} G^t &\doteq r^t + \gamma r^{t+1} + \gamma^2 r^{t+2} + \cdots \\ &= r^t + \gamma(r^{t+1} + \gamma r^{t+1} + \cdots) \\ &= r^t + \gamma(G^{t+1}) \end{aligned} \tag{2.4}$$

During the task process, the value function can be used to evaluate whether the strategy adopted by the agent is good or bad. The strategy is to map the state to the probability equation for selecting each possible action. Assuming that the intelligent agent follows a strategy π at time step t, $\pi(a|s)$ represents the probability of the agent taking action $a^t = a$ in state $s^t = s$. v_π represents the state value function and q_π denotes the state-action value function.

The state value function of state s is represented as $v_\pi(s)$, which is the expected total reward that the agent can ultimately receive from state s and continues to follow the policy thereafter. The state value function is defined as

$$v_\pi(s) \doteq \mathbb{E}_\pi\left[G^t|s^t = s\right] = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k r^{t+k}|s^t = s\right]. \tag{2.5}$$

The state-action value function of state s performing action a is represented as $q_\pi(s,a)$. It is the expected cumulative reward that the agent can ultimately obtain by taking action a from state s and continuing to follow policy π thereafter. The state-action value function is defined as

$$q_\pi(s,a) \doteq \mathbb{E}_\pi\left[G^t|s^t = s, a^t = a\right] = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k r^{t+k}|s^t = s, a^t = a\right]. \tag{2.6}$$

When the agent consistently follows the policy π and is able to experience each state sufficiently multiple times, the expected total reward that the agent can receive will eventually converge to the value function $v_\pi(s)$ of that state. Similarly, when an intelligent agent consistently follows a strategy and is able to experience each state and take all possible actions a sufficient number of times, the expected total reward that the agent can receive will eventually converge to the value function $q_\pi(s,a)$ of that state action. This method of estimating the value function is also known as Monte Carlo methods.

The value function in reinforcement learning satisfies a recursive relationship. For any policy

pi and any state of the state value function, there exists the following recursive relationship

$$\begin{aligned}
v_\pi(s) &\doteq \mathbb{E}_\pi\left[G^t|s^t=s\right] \\
&= \mathbb{E}_\pi\left[r^t+\gamma(G^{t+1})|s^t=s\right] \\
&= \sum_a \pi(a|s)\sum_s\sum_r p(s^{'},r|s,a)[r+\gamma\mathbb{E}_\pi[G^{t+1}|s^{t+1}=s^{'}]] \\
&= \sum_a \pi(a|s)\sum_{s^{'},r} p(s^{'},r|s,a)[r+\gamma v_\pi(s^{'})]. \qquad (2.7)
\end{aligned}$$

The action $a \in \mathrm{A}(s)$, reward $r \in \mathrm{R}$ and next state $s^{'} \in \mathrm{S}$. The above equation is the Bellman equation for state s. The Bellman equation for state-action is denoted as

$$q_\pi(s,a) = \sum_{s^{'},r} p(s^{'},r|s,a)[r+\gamma v_\pi(s^{'})]. \qquad (2.8)$$

For a given state s, whether a policy is superior or equivalent to another policy depends on whether its state value function is greater than or equal to the state value function of the other policy. In a task, there will be at least one policy that is superior to or equal to the other policies, and this policy is the optimal policy π_*. There may be one or more optimal policies, but they are all referred to as optimal policies.

The state value function and state-action value function corresponding to the optimal policy π_* are respectively referred to as

$$v_*(s) \doteq \max_\pi v_\pi(s), \qquad (2.9)$$

and

$$q_*(s,a) \doteq \max_\pi q_\pi(s,a). \qquad (2.10)$$

2.9 and 2.10 have the following relationship

$$q_*(s,a) = \mathbb{E}[r^t + \gamma v_*(s^{t+1})|s^t = s, a^t = a]. \tag{2.11}$$

v_* is the state value function of the optimal strategy π_*, so it also satisfies the recursive relationship given by the Bellman equation. Since it is the optimal state value function, its numerical value is equivalent to making the optimal action on state s.

$$\begin{aligned} v_*(s) &= \max_{a \in \mathrm{A}(s)} q_{\pi_*}(s,a) \\ &= \max_{a \in \mathrm{A}(s)} \mathbb{E}_{\pi_*}[G^t|s^t = s, a^t = a] \\ &= \max_{a \in \mathrm{A}(s)} \mathbb{E}_{\pi_*}[r^t + \gamma(G^{t+1})|s^t = s, a^t = a] \\ &= \max_{a \in \mathrm{A}(s)} \mathbb{E}_{\pi_*}[r^t + \gamma v_*(s^{t+1})|s^t = s, a^t = a] \\ &= \max_{a \in \mathrm{A}(s)} \sum_{s',r} p(s',r|s,a)[r + \gamma v_*(s')]. \end{aligned} \tag{2.12}$$

The above equation is the Bellman optimal equation for v_*. The Bellman optimal equation for q_* is expressed as

$$q_*(s,a) = \mathbb{E}[r^{t+1} + \gamma \max_{a'} q_*(s^{t+1},a')|s^t = s, a^t = a] \tag{2.13}$$

$$= \sum_{s',r} p(s',r|s,a)[r + \gamma \max_{a'} q_*(s',a')]. \tag{2.14}$$

In a task, each time step an agent takes an action based on its current state, the environment rewards the agent with a certain numerical value to evaluate the quality of the action. In order to numerically describe the quality of actions, a state value function and a state-action value function are provided, and the goal of the agent is to maximize the value function. The Bellman equation describes the relationship between the state value function and the state-action value function at the current and next moments. The Bellman optimal equation is used to indicate how the agent can find the optimal policy based on the current environment.

Policy Gradient and Actor-Critic

In practical applications, the tasks that usually need to be solved are high-dimensional state spaces and action spaces, or both state spaces and action spaces are continuous, so it is almost impossible for agents to traverse all states and actions and find the optimal strategy based on the Bellman optimal equation.

The policy gradient method directly models the strategy and optimizes it through gradient ascent to maximize the expected cumulative reward. The policy gradient method is typically suitable for handling continuous action spaces.

In the policy gradient, the policy π is usually modeled as a parameterized equation with a parameter of θ, which is represented by $\pi(a|s;\theta)$. Let

$$J(\theta) = \sum_{s\in\mathrm{S}} d_\pi(s) v_\pi(s) = \sum_{s\in\mathrm{S}} d_\pi(s) \sum_{a\in\mathrm{A}} \pi(a|s;\theta) q_\pi(s,a). \tag{2.15}$$

$d_\pi(s)$ represents the stationary state distribution of the Markov Chain for the strategy. When $J(\theta)$ increases, it means that the state value function increases. Therefore, using gradient ascent to increase the parameter θ in the direction of increasing $J(\theta)$, we can find the optimal state value function in state s and then find the optimal policy.

Firstly, calculate the gradient of the state value function

$$\begin{aligned}
\bigtriangledown_\theta v_\pi(s) &= \bigtriangledown_\theta \left[\sum_a \pi(a|s;\theta) q_\pi(s,a)\right] \\
&= \sum_a \left[\bigtriangledown_\theta \pi(a|s;\theta) q_\pi(s,a) + \pi(a|s;\theta) \bigtriangledown_\theta \sum_{s',r} p(s',r|s,a)(r + v_\pi(s'))\right] \\
&= \sum_a \left[\bigtriangledown_\theta \pi(a|s;\theta) q_\pi(s,a) + \pi(a|s;\theta) \sum_{s'} p(s'|s,a) \bigtriangledown_\theta v_\pi(s')\right].
\end{aligned} \tag{2.16}$$

Derived from the recursive relationship

$$\bigtriangledown_\theta v_\pi(s') = \sum_{a'} \left[\bigtriangledown_\theta \pi(a'|s';\theta) q_\pi(s',a') + \pi(a'|s';\theta) \sum_{s''} p(s''|s',a') \bigtriangledown_\theta v_\pi(s'')\right]. \tag{2.17}$$

Therefore

$$\nabla_\theta v_\pi(s) = \sum_{x\in \mathrm{S}} \sum_{k=0}^{\infty} \Pr(s \to x, k, \pi) \sum_a \nabla_\theta \pi(a|x;\theta) q_\pi(x,a). \tag{2.18}$$

$\Pr(s \to x, k, \pi)$ represents the probability of transitioning from state s to state x after k time steps under policy π.

Initial state from s_0

$$\begin{aligned}
\nabla_\theta J(\theta) &= \nabla_\theta v_\pi(s_0) \\
&= \sum_s \Big(\sum_{k=0}^{\infty} \Pr(s_0 \to s, k, \pi)\Big) \sum_a \nabla_\theta \pi(a|s;\theta) q_\pi(s,a) \\
&= \sum_s \eta(s) \sum_a \nabla_\theta \pi(a|s;\theta) q_\pi(s,a) \\
&= \left[\sum_s \eta(s)\right] \sum_s \frac{\eta(s)}{\sum_s \eta(s)} \sum_a \nabla_\theta \pi(a|s;\theta) q_\pi(s,a) \\
&\propto \sum_s \frac{\eta(s)}{\sum_s \eta(s)} \sum_a \nabla_\theta \pi(a|s;\theta) q_\pi(s,a) \\
&= \sum_s d_\pi(s) \sum_a \nabla_\theta \pi(a|s;\theta) q_\pi(s,a). \qquad (2.19)
\end{aligned}$$

$\eta(s) = \sum_{k=0}^{\infty} \Pr(s_0 \to s, k, \pi)$, $d_\pi(s) = \frac{\eta(s)}{\sum_s \eta(s)}$ represents the stationary state distribution of the Markov chain for the policy in state s.

$$\nabla_\theta J(\theta) \propto \sum_s d_\pi(s) \sum_a \nabla_\theta \pi(a|s;\theta) q_\pi(s,a), \tag{2.20}$$

this expression is the policy gradient theory. This expression can be further transformed into

$$\begin{aligned}
\nabla_\theta J(\theta) &\propto \sum_s d_\pi(s) \sum_a \nabla_\theta \pi(a|s;\theta) q_\pi(s,a) &(2.21)\\
&= \sum_s d_\pi(s) \sum_a \pi(a|s;\theta) q_\pi(s,a) \frac{\nabla_\theta \pi(a|s;\theta)}{\pi(a|s;\theta)} &(2.22)\\
&= \mathbb{E}_\pi \left[q_\pi(s,a) \nabla_\theta \ln \pi(a|s;\theta) \right]. &(2.23)
\end{aligned}$$

From the above equation, it can be intuitively seen that the value function can help update the policy, because the larger $q_\pi(s,a)$ means the greater the benefit that this action can obtain, and the larger $\bigtriangledown_\theta J(\theta$, the faster the parameter θ changes in the direction of this policy.

The policy gradient theory only considers the updating of policies, and the direction of policy updates depends on the value function. The actor-critic (AC) algorithm refers to the model of policy gradient theory and adds a value function update section, allowing the value function to guide the policy on how to update at each step, enabling the policy to achieve single step updates.

The AC algorithm includes a critic part and an actor part. Actor generates actions based on policy, and critic evaluates actor's actions. The Critic part can be a parameterized state value equation $v_\pi(s;\omega)$ or a parameterized state-action value function $q_\pi(s,a;\omega)$, and both can be fitted using a neural network with parameter ω. Critic is used to update the parameter ω and guide the Actor section on how to update the policy. Actor is a parameterized policy equation $\pi(a|s;\theta)$, and it can be fitted using a neural network with parameter θ. Actor will update the policy based on the size of the value function in Critic.

2.4 Gap Analysis

Many researches have been developed in the field of IRS assisted secure communication, but there are still several gaps in the current literature that have not been addressed. This section aims to analyze and determine the research gaps filled by this paper.

While IRS and backscatter communication have individually demonstrated potential for enhancing physical layer security, few studies explore their joint integration in a unified framework. Most existing works focus on either IRS assisted systems without considering the passive nature of backscatter, or backscatter systems without leveraging the reconfigurability and beamforming capabilities of IRS. This gap overlooks the synergistic benefits of combining IRS's intelligent beam forming with the covert and low-power transmission of backscatter.

Lots of the existing research addressed passive eavesdropping, where the attacker only listens to the channel. In more aggressive and realistic scenarios, active eavesdroppers send jamming signals to degrade legal user's received sigal and increase the risk of information leakage. This research gap is to design a secure IRS backscatter communication system against active jamming attack.

Many studies have used traditional optimization algorithms, such as alternating optimization and semidefinite programming, to effectively solve the coupled variables in non convex optimization problems for IRS assisted systems. Most optimization schemes for IRS systems are model-based static optimization schemes. These methods are struggled to adapt to real-time dynamic scenarios under unknown or rapidly changing conditions. Meanwhile, DRL has significant advantages in these scenario. This research gap indicates the need for DRL to autonomously adjust IRS configuration to maximize secrecy in real world environments.

Chapter 3

The System model and optimization problem formulation

Overview

This chapter first establishes an IRS-assisted MISO secure communication system. Then analyze the optimization problem of maximizing the security performance of the system.

3.1 The System model

Consider an IRS-based backscatter wireless communication system countermeasure against active eavesdropping, as shown in Figure 3.1. A source (Alice), an IRS, a legitimate user (Bob), and an eavesdropper (Willie) constitute the communication system of this thesis.

In this system, Alice continuously transmits information to all directions. Willie is sending interference signals to prevent Bob from receiving the required signals to eavesdrop. Generally, the stronger the interference signals from an eavesdropper like Willie, the lower the communication system security is, which needs the IRS to enhance this system's security. The IRS, as a transmitter, aims to convert all received signals into desired signals through backscattering. The signal processed by IRS uses the interference signal from Willie to ensure the communication safety of legal user Bob. The number of antennas equipped by Alice is N, while Willie and Bob are equipped with one antenna each. IRS has L elements. It

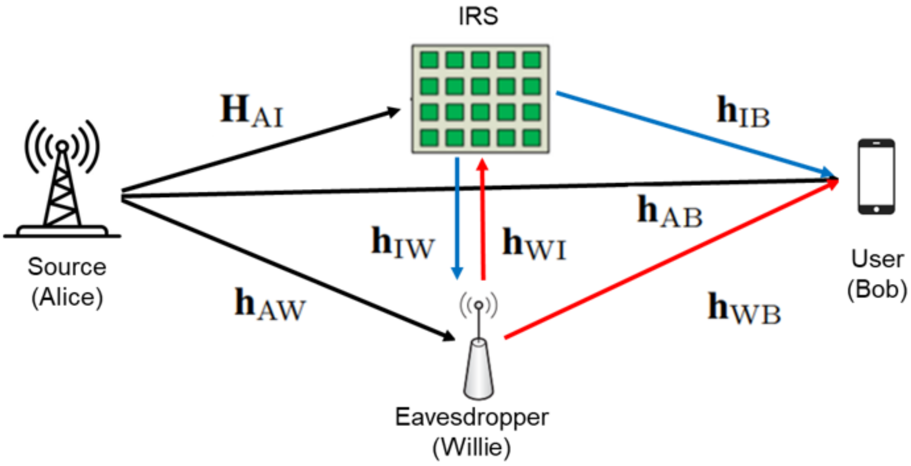
IRS
$\mathbf{H}_{\mathrm{AI}}$
$\mathbf{h}_{\mathrm{IB}}$
$\mathbf{h}_{\mathrm{AB}}$
$\mathbf{h}_{\mathrm{IW}}$
$\mathbf{h}_{\mathrm{WI}}$
Source
(Alice)
$\mathbf{h}_{\mathrm{AW}}$
$\mathbf{h}_{\mathrm{WB}}$
User
(Bob)
Eavesdropper
(Willie)

Fig. 3.1 IRS-aided backscatter wireless communication system under active eavesdropper's attack.

is assumed that all channels in the system experience quasi-static flat fading. Additionally, we consider that the channel state information of all the involved channels in the system is precisely and accurately known in order to determine the limit of the security rate. We assume the frequency is non-selective and constant in each fading block.

The channel gains from Alice to Bob, from the IRS to Bob, from Willie to Bob, from Willie to the IRS, from Alice to the IRS, from the IRS to Willie, and from Alice to Willie are represented by the following notations: $\mathbf{h}_{\mathrm{AB}} \in \mathbb{C}^{N\times 1}$, $\mathbf{h}_{\mathrm{IB}} \in \mathbb{C}^{L\times 1}$, $\mathbf{h}_{\mathrm{WB}} \in \mathbb{C}^{1\times 1}$, $\mathbf{h}_{\mathrm{WI}} \in \mathbb{C}^{L\times 1}$, $\mathbf{H}_{\mathrm{AI}} \in \mathbb{C}^{L\times N}$, $\mathbf{h}_{\mathrm{IW}} \in \mathbb{C}^{1\times L}$, and $\mathbf{h}_{\mathrm{AW}} \in \mathbb{C}^{1\times N}$. The beamforming vector used by Alice to transmit the desired signal s is denoted by $\mathbf{w} \in \mathbb{C}^{N\times 1}$, while the beamforming vector used by Willie to transmit the interference signal a is represented by $\mathbf{v} \in \mathbb{C}^{1\times 1}$.

The signals s and a represent the transmitted signals from Alice, the legitimate transmitter, and Willie, the eavesdropper, respectively. In addition, $\mathbb{E}[|s|^2] = 1$ and $\mathbb{E}[|a|^2] = 1$. n_{B} and n_{W} represent the Gaussian white noise at Bob and Willie, respectively. These noises have zero mean and variances of σ_{B}^2 and σ_{W}^2, respectively. The amplitude and phase shift incurred by the l-th reflective element of the IRS are represented by $\mathbf{\Theta} = \mathrm{diag}(\beta_1 e^{j\alpha_1}, \beta_2 e^{j\alpha_2}, \cdots, \beta_L e^{j\alpha_L})$ with $\alpha_l = [0, 2\pi]$, $l \in \mathscr{L} = \{1, 2, \cdots, L\}$ and $\beta_l = [0, 1]$. In this model, we do not consider the interaction of IRS-neighboring reflective units and assume that each IRS reflective unit reflects the signal independently. Due to strong path loss, we overlook signals that were reflected multiple times by the IRS. After the above design and construction of the whole system, the received signals at Bob and Willie can be respectively formulated as

$$y_{\mathrm{B}} = \mathbf{h}_{\mathrm{AB}}^{H}\mathbf{w}s + \mathbf{h}_{\mathrm{WB}}^{H}\mathbf{v}a + \mathbf{h}_{\mathrm{IB}}^{H}\mathbf{\Theta}(\mathbf{h}_{\mathrm{AI}}\mathbf{w}s + \mathbf{h}_{\mathrm{WI}}\mathbf{v}a) + n_{\mathrm{B}}, \tag{3.1}$$

$$y_{\mathrm{W}} = \mathbf{h}_{\mathrm{AW}}^{H}\mathbf{w}s + \mathbf{h}_{\mathrm{IW}}^{H}\mathbf{\Theta}(\mathbf{H}_{\mathrm{AI}}\mathbf{w}s + \mathbf{h}_{\mathrm{WI}}\mathbf{v}a) + n_{\mathrm{W}}. \tag{3.2}$$

Equation (3.1) mathematically describe the received signal at the legal user Bob includes direct signal from Alice, reflected signal via the IRS, jamming signal directly from Willie and additive white Gaussian noise at Bob. The received signal at Bob affected by Alice's beamforming and IRS reflection control, which helps to calculate Bob's SINR. In order to improve the secrecy of the system, It is necessary to maximize the SINR of Bob.

Equation (3.2) mathematically describe the received signal at the eavesdropper Willie includes direct signal from Alice, reflected signal from the IRS and additive white Gaussian noise at Willie. It forms the basis for calculating Willie's SINR, which must be minimized or kept under a threshold to ensure secure communication.

3.2 optimization problem formulation

In this system, the backscattering of signals through IRS does not need to distinguish the source signal s and the interference signal a. The goal of this system is to maximize the SINR at Bob γ while we constrained Willie's SINR ξ by setting the maximum value. The transmit powers at Alice and Willie are represented as P_s and P_a, respectively. Define $\boldsymbol{\theta} = [\beta_1 e^{j\theta_1}, \beta_2 e^{j\theta_2}, \cdots, \beta_L e^{j\theta_L}]^H$. The problem of this system is expressed as

$$\begin{aligned}
\max_{\boldsymbol{\theta},\mathbf{w}} \quad & \gamma = \frac{\left|\mathbf{h}_{\mathrm{AB}}^{H}\mathbf{w} + \mathbf{h}_{\mathrm{IB}}^{H}\mathbf{\Theta}(\mathbf{H}_{\mathrm{AI}}\mathbf{w} + \mathbf{h}_{\mathrm{WI}}\mathbf{v})\right|^2}{\sigma_{\mathrm{B}}^2 + \left|\mathbf{h}_{\mathrm{WB}}^{H}\mathbf{v}\right|^2}, \\
s.t. \quad & \mathrm{Tr}(\mathbf{w}\mathbf{w}^H) \leq P_s, \\
& \theta_l \leq 1, \forall l \in \mathscr{L}, \\
& \xi \leq \varepsilon.
\end{aligned} \tag{3.3}$$

The SINR at Bob is primarily dependent on two key variables: the beamforming vector $\mathbf{w}$ used by Alice and the reflection coefficient vector $\boldsymbol{\theta}$ utilized by the IRS. The SINR at Willie is formulated as

$$\xi = \frac{\left|\mathbf{h}_{\mathrm{AW}}^{H}\mathbf{w} + \mathbf{h}_{\mathrm{IW}}^{H}\boldsymbol{\Theta}(\mathbf{H}_{\mathrm{AI}}\mathbf{w} + \mathbf{h}_{\mathrm{WI}}\mathbf{v})\right|^{2}}{\sigma_{\mathrm{W}}^{2}}. \tag{3.4}$$

The goal of this system model is to jointly optimize the transmission power of the transmitter and the reflection coefficient of the IRS to maximize the SINR of the legitimate user Bob, while considering the constrain that the signal-to-interference-plus-noise ratio of the eavesdropper is below the threshold.

To further elaborate the proposed system model, the network configuration and simulation settings are defined as follows: the source Alice is equipped with $N = 4$ antennas, while the legitimate user Bob and the eavesdropper Willie each have a single antenna. The IRS comprises $L = 40$ passive reflecting elements, strategically placed near Bob to enhance the desired signal and suppress jamming. The channel links are modeled using quasi-static Rician fading with a path loss exponent of 3, and additive white Gaussian noise with a noise variance of $\sigma^2 = 10^{-5}$. The transmit power at Alice is P_s = 9 dBW.

The optimization problem in this model is non-convex. To solving this optimization problem involves transforming problem 3.3 into a quadratically constrained quadratic programming (QCQP) problem. Subsequently, finding a sub-optimal solution to problem 3.3 involves employing AO optimization methods. The AO method iteratively optimizes the source beamforming and IRS configuration under SINR constraints. In addition, the DDPG based DRL approach was proposed to solve the non-convex optimization problem. The DRL based method adaptively learns optimal IRS policies based on observed environmental states, which effectively enhancing the secrecy rate in dynamic settings.

Chapter 4

Alternating Optimization Algorithm and System Security Performance Analysis

Overview

In this chapter, the AO algorithm is employed to solve the optimization problem. Subsequently, simulations are conducted based on the system model proposed in Chapter 3, followed by the analysis of the simulation results.

4.1 Alternation Optimization

Alternation optimization is a classical approach for solving optimization problems involving multiple sets of coupled variables, especially when optimization problem is non-convex. Solving the problem of all variables simultaneously is computationally challenging. On the other hand, The AO method fixes one set of variables and optimizes another set of variables, alternating between them.

We develop an alternating optimization algorithm to solve the optimization problem. Due to the coupling between the variables $\boldsymbol{\theta}$ and $\mathbf{w}$, directly solving the non-convex optimization problem can be challenging. Specifically, we address this non-convex problem by iteratively solving two sub-problems: sub-problem 1, which focuses on optimizing the beamforming vector $\mathbf{w}$ with a fixed reflection coefficient vector $\boldsymbol{\theta}$, and sub-problem 2, which focuses on optimizing $\boldsymbol{\theta}$ with a fixed $\mathbf{w}$.

Before optimizing these two sub-problems respectively, we need to convert the objective function equivalently. Since $\boldsymbol{\Theta} = \text{diag}\left\{\boldsymbol{\theta}^H\right\}$, that expression can be converted to

$$\mathbf{h}_{\text{AB}}^H\mathbf{w} + \mathbf{h}_{\text{IB}}^H\boldsymbol{\Theta}(\mathbf{H}_{\text{AI}}\mathbf{w} + \mathbf{h}_{\text{WI}}\mathbf{v}) = \mathbf{h}_{\text{AB}}^H\mathbf{w} + \boldsymbol{\theta}^H\boldsymbol{\Phi}\mathbf{w} + \boldsymbol{\theta}^H\mathbf{a}. \tag{4.1}$$

$\boldsymbol{\Phi} = \text{diag}\left\{\mathbf{h}_{\text{IB}}^H\right\}\mathbf{H}_{\text{AI}}$ and $\boldsymbol{a} = \text{diag}\left\{\mathbf{h}_{\text{IB}}^H\right\}\mathbf{h}_{\text{WI}}\mathbf{v}$. This equation simplifies the total received signal at Bob, which originally includes direct transmission from Alice and reflected signal via IRS into a more compact, vectorized form suitable for optimization. This transformation prepares the optimization model for alternationg optimization.

Letting $\hat{\boldsymbol{\theta}} = [\boldsymbol{\theta}^H, 1]^H$ and $\hat{\boldsymbol{\Phi}} = [\boldsymbol{\Phi}^H, \mathbf{h}_{\text{AB}}]^H$, then Equation 4.1 is transformed as:

$$\mathbf{h}_{\text{AB}}^H\mathbf{w} + \boldsymbol{\theta}^H\boldsymbol{\Phi}\mathbf{w} + \boldsymbol{\theta}^H\mathbf{a} = ([\boldsymbol{\theta}^H, 1][\boldsymbol{\Phi}^H, \mathbf{h}_{\text{AB}}]^H)\mathbf{w} + \boldsymbol{\theta}^H\mathbf{a} = \hat{\boldsymbol{\theta}}\hat{\boldsymbol{\Phi}}\mathbf{w} + \boldsymbol{\theta}^H\mathbf{a}. \tag{4.2}$$

This equation takes the expression from Equation 4.1 and reorganizes it in a compact and vectorized form by introducing augmented variables $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\Phi}}$. This transformation is used to unify the received signal expression into a clean linear form: $\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\Phi}}\mathbf{w}$.

Assuming $\hat{\boldsymbol{a}} = [\boldsymbol{a}^H, 0]^H$, convert expression 4.2 to:

$$\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\Phi}}\mathbf{w} + \boldsymbol{\theta}^H\mathbf{a} = \hat{\boldsymbol{\theta}}\hat{\boldsymbol{\Phi}}\mathbf{w} + [\boldsymbol{\theta}^H, 1]^H[\boldsymbol{a}^H, 0]^H = \hat{\boldsymbol{\theta}}\hat{\boldsymbol{\Phi}}\mathbf{w} + \hat{\boldsymbol{\theta}}\hat{\boldsymbol{a}}. \tag{4.3}$$

This expression combines both desired signal and interference into one linear algebraic structure. Letting $\hat{\mathbf{w}} = [\mathbf{w}^H, 1]^H$, and $\check{\boldsymbol{\Phi}} = [\hat{\boldsymbol{\Phi}}, \hat{\boldsymbol{a}}]^H$, expression 4.3 can be transformed into:

$$\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\Phi}}\mathbf{w} + \hat{\boldsymbol{\theta}}\hat{\boldsymbol{a}} = \hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\Phi}}\mathbf{w} + \hat{\boldsymbol{a}}) = \hat{\boldsymbol{\theta}}\check{\boldsymbol{\Phi}}\hat{\mathbf{w}}. \tag{4.4}$$

This equation combines two separate terms from the signal expression 4.3 into a single matrix vector product, which expresses the full received signal as a bilinear form involving both optimization variables.

Therefore, the objective equation can be deduced as

$$\left|\mathbf{h}_{\mathrm{AB}}^{H}\mathbf{w}+\mathbf{h}_{\mathrm{IB}}^{H}\mathbf{\Theta}(\mathbf{H}_{\mathrm{AI}}\mathbf{w}+\mathbf{h}_{\mathrm{WI}}\mathbf{v})\right|^{2}=\hat{\boldsymbol{\theta}}^{H}\check{\mathbf{\Phi}}\hat{\mathbf{w}}\hat{\mathbf{w}}^{H}\check{\mathbf{\Phi}}^{H}\hat{\boldsymbol{\theta}}. \tag{4.5}$$

This equation expresses the power of the total received signal at Bob, as a quadratic form. It converts the objective function into a form suitable for the use of convex solvers (such as CVX) for efficient calculation.

Similarly, for the expression for the SINR at Willie, we transform it in the same way.

$$\mathbf{h}_{\mathrm{AW}}^{H}\mathbf{w}+\mathbf{h}_{\mathrm{IW}}^{H}\mathbf{\Theta}(\mathbf{H}_{\mathrm{AI}}\mathbf{w}+\mathbf{h}_{\mathrm{WI}}\mathbf{v})=\hat{\boldsymbol{\theta}}\check{\boldsymbol{\delta}}\hat{\mathbf{w}}. \tag{4.6}$$

letting $\check{\boldsymbol{\delta}}=[\hat{\boldsymbol{\delta}},\hat{\boldsymbol{b}}]^{H}$, $\hat{\boldsymbol{\delta}}=[\boldsymbol{\delta}^{H},\mathbf{h}_{\mathrm{AW}}]^{H}$, $\hat{\boldsymbol{b}}=[\boldsymbol{b}^{H},0]^{H}$, $\boldsymbol{\delta}=\mathrm{diag}\left\{\mathbf{h}_{\mathrm{IW}}^{H}\right\}\mathbf{H}_{\mathrm{AI}}$ and $\boldsymbol{b}=\mathrm{diag}\left\{\mathbf{h}_{\mathrm{IW}}^{H}\right\}\mathbf{h}_{\mathrm{WI}}\mathbf{v}$. This equation expresses the received signal at Willie in the same compact matrix-vector form used for Bob. This transformation is essential for later deriving the optimization problem.

Define $\hat{\mathbf{W}}=\hat{\mathbf{w}}\hat{\mathbf{w}}^{H}$, $\hat{\mathbf{\Theta}}=\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}^{H}$, $\mathrm{rank}(\hat{\mathbf{W}})=1$, $\mathrm{rank}(\hat{\mathbf{\Theta}})=1$, $\hat{\mathbf{W}}\succeq\mathbf{0}$, $\hat{\mathbf{\Theta}}\succeq\mathbf{0}$. Then optimization problem is reformulated as

$$\begin{aligned}
\max_{\hat{\mathbf{\Theta}},\hat{\mathbf{W}}}\quad & \hat{\boldsymbol{\theta}}^{H}\check{\mathbf{\Phi}}\hat{\mathbf{w}}\hat{\mathbf{w}}^{H}\check{\mathbf{\Phi}}^{H}\hat{\boldsymbol{\theta}}=\mathrm{Tr}(\hat{\mathbf{W}}\check{\mathbf{\Phi}}^{H}\hat{\mathbf{\Theta}}\check{\mathbf{\Phi}}),\\
s.t.\quad & \mathrm{Tr}(\hat{\mathbf{W}})\le P_s+1,\\
& \hat{\boldsymbol{W}}_{N+1,N+1}=1,\\
& \hat{\mathbf{\Theta}}_{l,l}=1,\ l\in\mathscr{L}\ \text{or}\ l=L+1,\\
& \hat{\mathbf{\Theta}}\succeq\mathbf{0},\mathrm{rank}(\hat{\mathbf{\Theta}})=1,\\
& \hat{\mathbf{W}}\succeq\mathbf{0},\mathrm{rank}(\hat{\mathbf{W}})=1,\\
& \xi\le\varepsilon.
\end{aligned} \tag{4.7}$$

This expression reformulated the maximization problem a matrix form using the variable transformations. It maximizes Bob's received signal power in a matrix trace version under physical and network constraints. These constraints include limiting the total transmit power of the beamformer, ensuring that the augmented vector maintains the structure, the IRS phase shift must have a unit modulus, the matrix must be positive semidefinite, the outer product of the matrix variable execution vector and the constraints of the eavesdropper's SINR leakage. Expression 4.7 is the core optimization problem in this system model.

We transform problem 4.7 into its relaxed form by removing the constraints of $\text{rank}(\hat{\mathbf{W}}) = 1$ and $\text{rank}(\hat{\boldsymbol{\Theta}}) = 1$. Subsequently, problem 4.7 is more tractable and can be solved using convex optimization techniques. This relaxation allows for a wider range of solutions. In addition, 4.7 can be expressed as

$$\text{Tr}(\hat{\mathbf{W}}\check{\boldsymbol{\Phi}}^H\hat{\boldsymbol{\Theta}}\check{\boldsymbol{\Phi}}) = \text{vec}(\check{\boldsymbol{\Phi}})^H\left(\hat{\mathbf{W}}^T \otimes \hat{\boldsymbol{\Theta}}\right)\text{vec}\left(\check{\boldsymbol{\Phi}}\right). \tag{4.8}$$

Equation 4.8 rewrites the objective trace as a vector-matrix-vector quadratic form using Kronecker products and vectorized matrices. This form is useful for algorithmic implementation.

(sub-problem 1) When $\hat{\boldsymbol{\Theta}}$ is given,

$$\begin{aligned} \max_{\hat{\mathbf{W}}} \quad & \text{Tr}(\hat{\mathbf{W}}\check{\boldsymbol{\Phi}}^H\hat{\boldsymbol{\Theta}}\check{\boldsymbol{\Phi}}), \\ s.t. \quad & \text{Tr}(\hat{\mathbf{W}}) \le P_s + 1, \\ & \hat{\boldsymbol{W}}_{N+1,N+1} = 1, \\ & \hat{\mathbf{W}} \succeq \mathbf{0}, \\ & \xi \le \varepsilon. \end{aligned} \tag{4.9}$$

(sub-problem 2) When $\hat{\mathbf{W}}$ is given,

$$\begin{aligned} \max_{\hat{\boldsymbol{\Theta}}} \quad & \text{Tr}(\hat{\mathbf{W}}\check{\boldsymbol{\Phi}}^H\hat{\boldsymbol{\Theta}}\check{\boldsymbol{\Phi}}), \\ s.t. \quad & \hat{\boldsymbol{\Theta}}_{l,l} = 1,\ l \in \mathscr{L} \text{ or } l = L+1, \\ & \hat{\boldsymbol{\Theta}} \succeq \mathbf{0}, \\ & \xi \le \varepsilon. \end{aligned} \tag{4.10}$$

Expression 4.9 corresponds to the beamforming optimization at the transmitter Alice under fixed IRS configuration. It maximizes the matrix trace form of Bob's received signal power. The constrains ensure that the transmit power does not exceed the maximum power, homogenization in SDR formulation, positive semidefiniteness of the matrix variable and the SINR at Willie remains below a tolerable threshold. Expression 4.10 addresses the design of the IRS phase shifts while keeping the beamforming vector fixed. The objective is to maximize Bob's signal power. The constraints ensures that each IRS element applies a unit modulus and the SINR at Willie remains below a threshold.

These two subproblems resulting from the relaxation are convex. We can converge to an optimized solution for $\hat{\mathbf{W}}^*$ and $\hat{\mathbf{\Theta}}^*$ through iteratively solving the relaxed sub-problems 1 and 2 alternately. The above problem currently is a convex-positive semi-definite program (SDP), and it can be efficiently solved using existing convex optimization solvers. If $\hat{\mathbf{W}}^*$ and $\hat{\mathbf{\Theta}}^*$ are rank 1, restore $\hat{\mathbf{w}}^*$ and $\hat{\boldsymbol{\theta}}^*$ using singular value decomposition (SVD). When using the SVD for rank reduction, we can choose to keep the first few largest singular values and set the others to zero. This will result in a lower rank approximation, but not usually a complete reduction to rank one. A rank-one solution is a special case and is unlikely to be realized in the general case. In other cases, recover the approximate solution $\mathbf{w}^*$ and $\boldsymbol{\theta}^*$ using the standard Gaussian randomization method. This randomization method provides an approximate solution and is particularly useful when dealing with matrices of higher rank. The quality of the approximation depends on the properties of the original matrices and the size of the random matrix.

The secrecy rate of this system model denotes as R. It is calculated based on the differece between the achievable rate of Bob, and the achievable rate of Willie. The mathematical definition of the secrecy rate is as follows.

$$\mathrm{R} = [\mathrm{R_B} - \mathrm{R_W}]^+ = [\log_2(1+\gamma) - \log_2(1+\xi)]^+ \tag{4.11}$$

$\mathrm{R_B}$ is the data rate at Bob, $\mathrm{R_W}$ is the data rate at Willie, γ is the SINR at Bob and ξ is the SINR at Willie.

According to the above analysis, we recapitulate the overall algorithm for problem 3.3 as Algorithm 1. The objective value of optimizing problem is represented by $\mathrm{R}^{(k)}$ with variables $\hat{\mathbf{\Theta}}^{(k)}$ and $\hat{\mathbf{W}}^{(k)}$ in the k-th iteration, while ε denotes a small threshold set to 0.001. In each iteration of Algorithm 1, after updating either the beamforming vector or the IRS reflection vector, the algorithm calculates the updated SINR at Bob based on the current beamforming vector and the IRS reflection vector, the updated SINR at Willie, then computes R with equation 4.11. At each iteration, $\hat{\mathbf{\Theta}}^{(k-1)}$ is fixed, and $\hat{\mathbf{W}}^{(k)}$ is optimized. Then $\hat{\mathbf{W}}^{(k)}$ is fixed, and $\hat{\mathbf{\Theta}}^{(k)}$ is updated. This loop continues until the R value convergence. R is used to track convergence and record the best secrecy achieved so far. The alternating optimization algorithm has many applications in wireless communication [9, 2, 27]. Algorithm 1 always converges, as the objective value is non-decreasing over iterations and has a finite upper bound.

Algorithm 1 Algorithm for Solving Optimization Problem

```
**Initialization**: Set $k = 0$, $\hat{\boldsymbol{\theta}}^{(0)} = 1_L$. Input variables: $\hat{\mathbf{\Theta}}^{(0)}, \hat{\mathbf{W}}^{(0)}$.
Compute $\hat{\mathbf{\Theta}}^{(0)} = \hat{\boldsymbol{\theta}}^{(0)H}\hat{\boldsymbol{\theta}}^{(0)}$; $\mathrm{R}^{(0)} = f(\hat{\mathbf{W}}^{(0)}, \hat{\mathbf{\Theta}}^{(0)})$, according to 4.8.
**repeat**
    Set $k = k+1$.
    With given $\hat{\mathbf{\Theta}}^{(k-1)}$, optimize the sub-problem 1, $\hat{\mathbf{W}}^{(k)}$ by 4.9.
    With given $\hat{\mathbf{W}}^{(k)}$, optimize the sub-problem 2, $\hat{\mathbf{\Theta}}^{(k)}$ by 4.10.
    Compute $\mathrm{R}^{(k)} = f(\hat{\mathbf{W}}^{(k)}, \hat{\mathbf{\Theta}}^{(k)})$.
**until** $\frac{\mathrm{R}^{(k)} - \mathrm{R}^{(k-1)}}{\mathrm{R}^{(k)}} < \varepsilon$.
Recover rank-one approximate solution output variables $\mathbf{w}^*$ and $\boldsymbol{\theta}^*$.
```

4.2 Numerical Results

In order to evaluate the security of the proposed approach in this paper, numerical simulations were conducted on an IRS-assisted backscatter communication system. We used the MATLAB to generate the simulation. To ensure the reliability of the simulation results, appropriate network circumstances have been carefully considered. The deployment of the network is a multiple antenna base station Alice communicates with a single antenna legitimate user Bob, while an active eavesdropper Willie attempts to intercept the transmission. An IRS is deployed near Bob to assist in enhancing the SINR while suppressing the Willie's SINR. All wireless links are modeled as quasi-static flat-fading channels. The direct and reflected links are assumed to follow Rician or Rayleigh fading, based on their line-of-sight or non-line-of-sight characteristics. Path loss is calculated using the log-distance model. The transmit power at Alice is $P_s = 9$ dBW. The noise variance is fixed at $\sigma^2 = 10^{-5}$. The jamming power from Willie and the number of reflecting elements is varied for evaluate the system robustness and scalability, respectivly. These conditions establish a controlled and real simulation environment, which can accurately evaluate the performance of the proposed system model and verify its effectiveness in the actual wireless circumstances. Additionally, for comparative analysis, this paper also provides several different schemes.

The first scheme considers the transmission scenario of a traditional wireless communication system that does not incorporate an IRS, which is expressed as Without-IRS. By comparing the performance of the IRS-assisted system against this Without-IRS scenario, we can evaluate the added benefits or improvements brought by the IRS. The optimization for this scheme is specifically for the beamformer $\mathbf{w}$. Therefore, it may not take advantage of the additional capabilities that an IRS can offer in terms of enhancing communication links, mitigating interference, or improving the overall system performance.

The second scheme is the wireless communication system assisted by passive reflection IRS, which is denoted by Reflection-IRS. This scheme involves the integration of an IRS into the wireless communication system. In this system, the IRS acts as a passive relay, reflecting signals to enhance communication links. This scheme jointly optimizes the IRS reflection coefficient $\boldsymbol{\theta}$ and the source beamformer $\mathbf{w}$. The optimization process considers both the reflection properties of the IRS and the beamforming at the source. In contrast, the difference in the scheme proposed in this paper is that the IRS in the backscatter system utilizes the interference from the eavesdropper to enhance the receiving power of the legitimate user. This scheme is used to compare the effects of backscatter technology for IRS-assisted communication systems.

The third scheme involves only optimizing the reflection coefficient vector of the IRS using the maximum ratio transmission (MRT), referred to as MRT-IRS. In this approach, the beamforming vector of the source is not involved in optimization. The primary optimization in the MRT-IRS scheme is directed towards the IRS reflection coefficient vector. The goal is to maximize the received signal power at the legal user by adjusting the IRS reflections. By optimizing only the IRS reflection coefficients and not involving the source beamforming vector, it provides a reference for evaluating the impact of IRS reflections alone on the system performance. This scheme simplifies the optimization process and could limit the overall performance compared to schemes that optimize both the source and the IRS.

The last scheme is a relay. By introducing a relay with a set number of antennas and specific transmit power, the performance of this relay-based system can be compared with the IRS-assisted system. This scheme employs a relay with four antennas in place of the IRS, and its position is set to be identical to the IRS in the BackCom-IRS approach for comparison purposes. This positioning ensures a fair and relevant comparison between the two approaches. Unlike the IRS-assisted system, which primarily reflects signals to enhance communication, the relay scheme actively amplifies and forwards signals. The transmit power of this relay is denoted by P_r.

The link from Alice to Willie is modeled as a slow-fading Rayleigh channel. The channel gain from the IRS to Bob follows a Rician distribution with a Rician factor of 3 on a small scale. The path loss model for all channels in the system is denoted by $PL = PL_0 - 10lg(d/d_0)$ dB. The path loss at the reference distance of $d = d_0$ and d_0 = 1 m is denoted by PL_0 = -30 dB. The transmit power at Alice is P_s = 9 dBW. The transmit power of relay is P_r = 0.1 W. The noise variance is $\sigma^2 = 10^{-5}$. The distances from Alice to Bob, Alice to the IRS, Alice to Willie, the IRS to Bob, Willie to Bob, and Willie to the IRS are d_{AB} = 60 m, d_{AI} = 55 m, d_{AW} = 55 m, d_{IB} = 15 m, d_{WB} = 15 m, and d_{WI} = 15 m, respectively. The summary of the parameters is presented in Table 4.1.

Table 4.1 Parameters setting.

Parameter	**Value**
Rician factor	$\kappa = 3$
Path loss	$PL_0 = -30$ dB $d_0 = 1$ m
Transmit power at Alice	$P_s = 9$ dBW
Transmit power of relay	$P_r = 0.1$ W
Noise variance	$\sigma^2 = 10^{-5}$
Distances	$d_{\text{AB}} = 60$ m, $d_{\text{AI}} = 55$ m, $d_{\text{AW}} = 55$ m, $d_{\text{IB}} = 15$ m, $d_{\text{WB}} = 15$ m, $d_{\text{WI}} = 15$ m

Figure 4.1 shows the SINR at user Bob in the BackCom-IRS scheme as a function of the iteration number t under randomly generated observations. It represents a typical result selected from several generated observations. In this case, $P_a = 9$ dBW, $L = 40$. As the iterations progress, the SINR at the legal user Bob tends to stabilize. When the number of iterations reaches ten, the objective function converges to 10^{-3} and the convergence is monotonically increasing.

Figure 4.2 shows that the SINR at Bob changes with the transmission power of the eavesdropper Willie. It can be seen from this figure that the SINR at Bob decreases as P_a increases. Increasing the transmission power at the eavesdropper Willie is detrimental to the user's received information. Higher transmission power of the eavesdropper negatively affects the communication link to user Bob. The proposed scheme outperforms the MRT-IRS scheme, traditional reflection IRS scheme, Relay scheme, and Without-IRS scheme, as demonstrated in the simulation results. This implies that, even under conditions of increased eavesdropper power, the proposed scheme is more effective in maintaining a satisfactory SINR at user Bob. Especially in comparison to the conventional reflection-IRS, the IRS with integrated backscatter exhibits a more substantial difference in the SINR as the eavesdropper's transmit power increases. This implies that backscatter technology can enhance the security of IRS-assisted communication systems. There are near linear patterns in this figure. The linear pattern in the curves arises from interference behavior in the system.

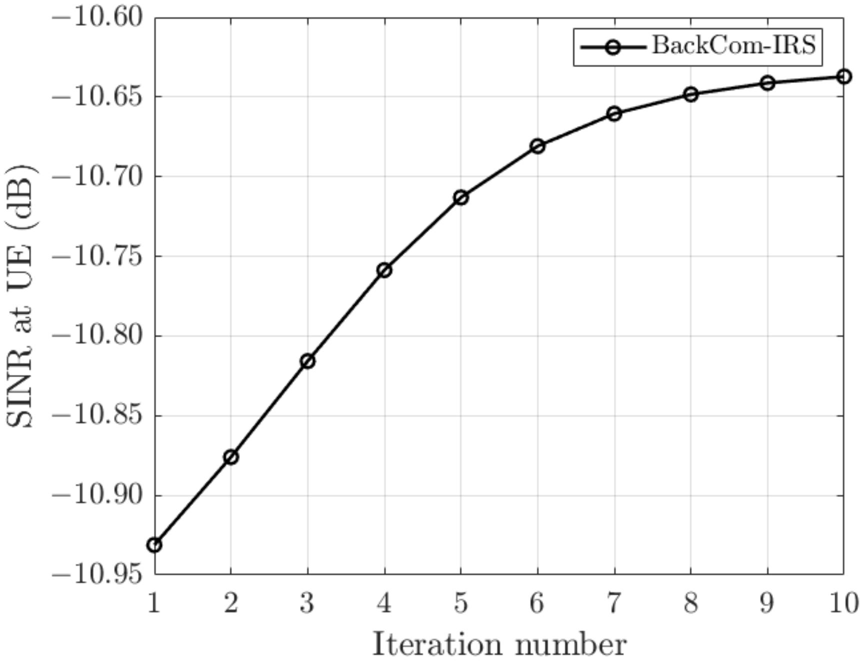

Fig. 4.1 Convergence of the BackCom-IRS scheme in a random observation.

Because the transmit power of Willie increases linearly, the interference experienced by Bob also increases approximately linearly. When there are no adaptive mechanisms used at Bob to suppress this interference in the simulated system, the SINR degradation follows a predictable and nearly linear trend. In [9, 24], as the transmit power at Tx increases, IRS-assisted systems exhibit better secrecy rate gains. In [50], the downlink network IRS-aided system has better performance than full-dupl decode-and-forward relay (FDR)-aided system in the high SNR. In the low-SNR, the IRS transmission experiences severe path loss and performs worse than FDR-aided systems.

Figure 4.3 illustrates how the SINR at Bob is affected by the total number of reflective elements L at the IRS. This figure shows that the SINR at Bob increases as the number of elements L increases. The number of IRS elements does not affect the Without-IRS and relay scheme. At first, the relay scheme and the without-IRS scheme will be better than the scheme proposed in this paper. As the number of IRS components increases, the scheme proposed in this paper will achieve higher performance gains than other schemes. This implies that a larger number of elements in the IRS contributes positively to the security of the wireless

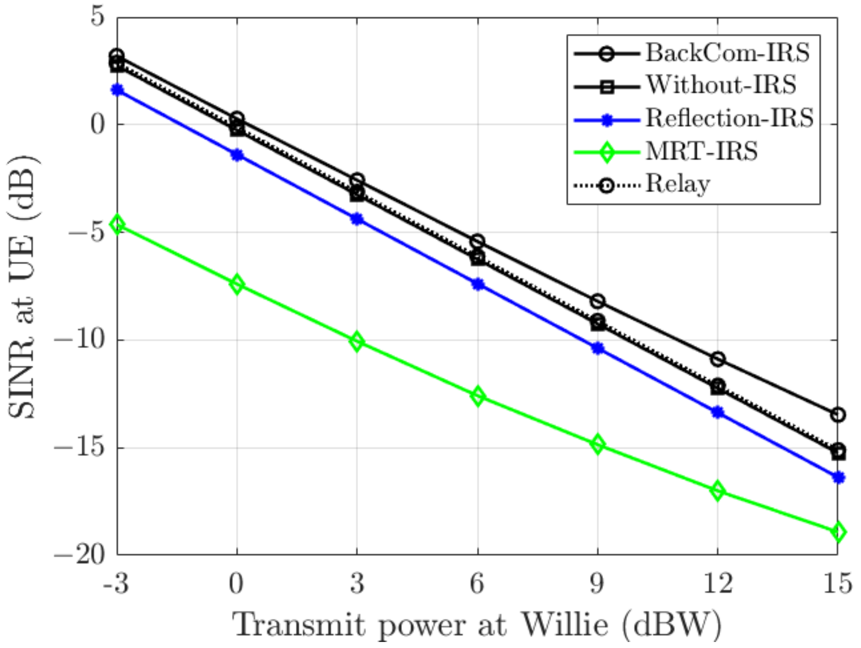

Fig. 4.2 The SINR γ at Bob versus the transmit power P_a at Willie.

communication system. Both [24, 23] indicate that adding more IRS reflective elements is beneficial for improving the secrecy rate in the IRS-assisted system. When the number of reflecting elements is small, the received signal at legal user Bob is dominated by the direct link other than the IRS-assisted link. The performance differences between the proposed scheme and the MRT-IRS scheme increase with the reflecting elements of IRS. This indicates that as the number of reflective elements increases, the joint optimization of the beamforming vector of the source and the reflection coefficient of the IRS becomes more flexible, and the performance gain also becomes higher.

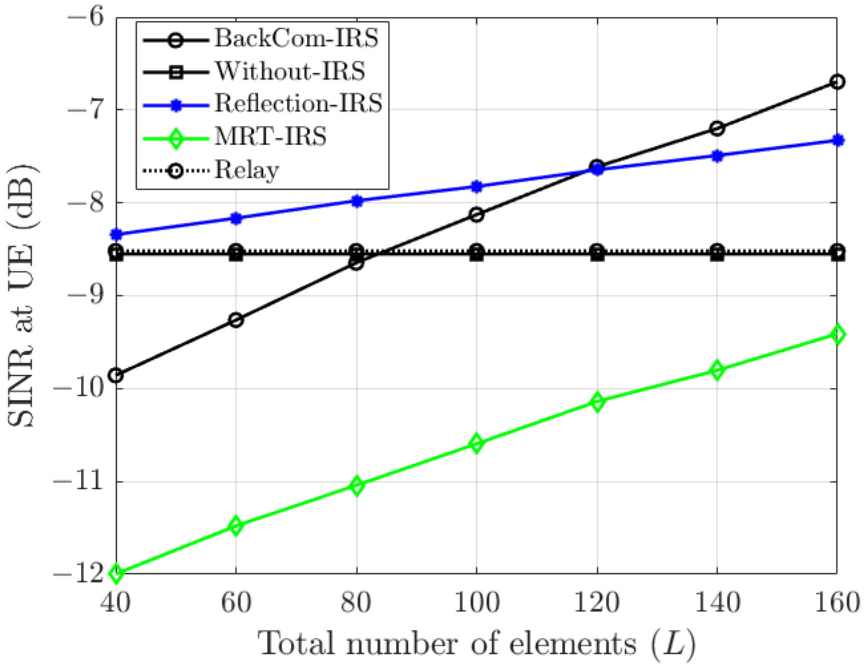

Fig. 4.3 The SINR γ at Bob versus the total number of reflecting elements L at the IRS.

Chapter 5

DDPG Algorithm and System Security Performance Analysis

Overview

In this chapter, the DDPG algorithm is employed to solve the optimization problem. Following that, simulations are conducted based on the system model proposed in Chapter 3, and the simulation results are analyzed.

5.1 DDPG Algorithm

While classical optimization algorithm as alternating optimization are effective for solving the secrecy rate maximization problem, there are several limitations. It usually requires perfect and instantaneous CSI. Every time the environment changes, the optimization process must be rerun from the beginning. It may be computationally intensive, especially in large scale IRS settings. In order to overcome these limitations and achieve real-time, adaptive and decision-making, we introduce DRL. By using DRL, we allow an agent to interact with the wireless environment and learn a strategy that maps the observed state of the system to the optimal action. The agent learns through trial and error and receives reward based on how well the action improves the secrecy rate. Over time, it learns to take actions that maximize long term expected reward, which is secure communication performance in this system.

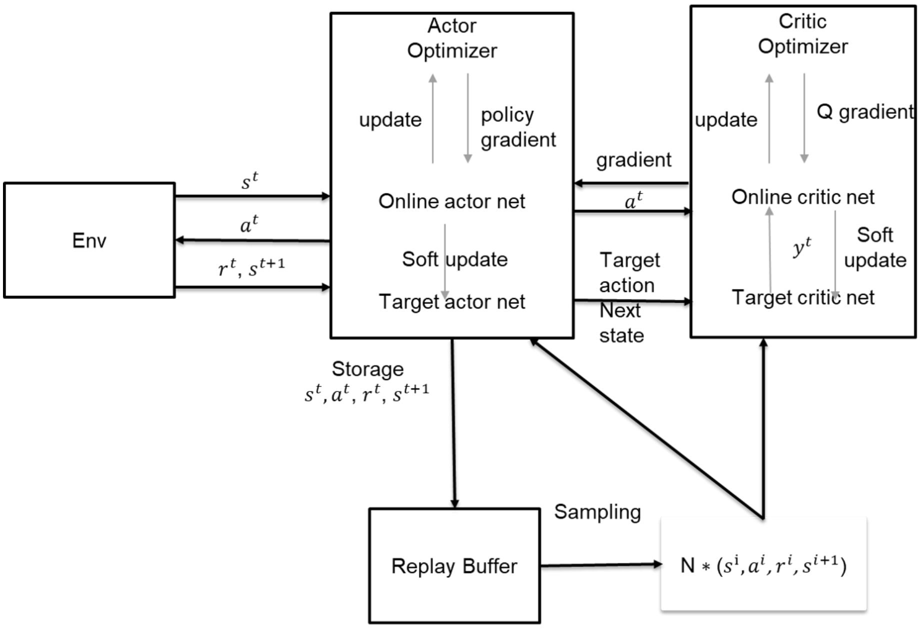

Fig. 5.1 DDPG structure

We used DDPG as a typical DRL algorithm in this work. The learning framework used in this work is mainly based on policy, and the value function is estimated using actor critic method. The agent in DDPG algorithm learns a deterministic policy: a mapping from state to action and also learns a value function (the Q value), which helps improve the policy. It is a hybrid of policy learning and value prediction. The DDPG algorithm framework is shown in the figure 5.1, which includes four neural networks: critic network, action network, target critic network, and target action network.

Firstly, we create an experience replay buffer of size $\mathscr{D}$. At time t, the agent performs the current action a^t in the current state s^t and receives the current reward r^t from the environment. At the same time, the agent enters the next state s^{t+1} and describes the transition characteristics of the agent at time t using tuple (s^t, a^t, r^t, s^{t+1}), which is stored in the replay buffer.

The learning process of DDPG algorithm is described as follows. Randomly select the k-th transition tuple (s^k, a^k, r^k, s^{k+1}) from replay buffer. In the policy critic, take (s^k, a^k) as the input of the critic network and output the current Q Value $q_\pi(s^k, a^k; \theta c)$ is used to critic the current policy. θc represents as a critic network parameter. Take s^{k+1} as the input of the target actor network and output the actor a^{k+1}. Take (s^{k+1}, a^{k+1}) as the input of the target critic network and output the target Q Value $q_\pi(s^{k+1}, a^{k+1}; \theta c)$.

In the policy improvement stage, in order to make the current Q value and target Q value as close as possible, the reward r, the current Q value and target Q value are used to construct a mean square Bellman error (MSBE) function, and the MSBE is minimized to update the critic network parameter θc. The action network parameter θa is updated by taking the derivative of $q_\pi(s^k, a^k; \theta c)$ with respect to the action network parameter using gradient rise. The target critic network parameter and target action network parameter are updated using Soft Update.

The updates on the training critic network is given as follow:

$$\theta_c^{t+1} = \theta_c^t - \mu_c \nabla_{\theta_c^{train}} L(\theta_c^{train}). \tag{5.1}$$

The loss function of the training critic network is

$$L(\theta_c^{train}) = \left(r^t + \gamma q(\theta_c^{target} | s^{t+1}, a') - q(\theta_c^{train} | s^t, a^t)\right)^2 \tag{5.2}$$

μ_c is the learning rate for the update on training critic network. a' is the action output from the target actor network. $\gamma \in (0, 1]$ represents the discount coefficient of Q value.

The update on the training actor network is given as

$$\theta_a^{t+1} = \theta_a^t - \mu_a \nabla_{\theta_a} q(\theta_c^{target} | s^t, a) \nabla_{\theta_a^{train}} \pi(\theta_a^{train} | s^t). \tag{5.3}$$

μ_a is the learning rate for the update on training actor network.

The updates on the target critic network and the target actor network are given as follows, respectively

$$\theta_c^{target} \leftarrow \tau_c \theta_c^{train} + (1 - \tau_c) \theta_c^{target}, \tag{5.4}$$

Algorithm 2 DDPG Based Algorithm for Sovling Optimization Problem

```
Initialization: experience replay memory $\mathscr{M}$ with size $\mathscr{D}$, training actor network parameter $\theta_a^{train}$, target actor network parameter $\theta_a^{train} = \theta_a^{target}$, training critic network parameter $\theta_c^{train}$, target critic network parameter $\theta_c^{train} = \theta_c^{target}$.
Input: the transmit beamforming matrix $\mathbf{W}$, the reflection coefficient $\mathbf{\Theta}$.
for episode $j = 1, 2, ..., J$ Do
    Obtain the initial state $s^0$.
    for time step $t = 1, 2, ..., T$ Do
        Obtain action $a^t = \{\mathbf{W}^t, \mathbf{\Theta}^t\} = \pi(\theta_a^{train})$ from the actor network.
        Obtain the next state $s^{t+1}$ given the action $a^t$, and calculate the reward $r^t$. Store the experience $(s^t, a^t, r^t, s^{t+1},)$ in the replay memory.
        Obtain the $Q$ value function from the critic network. $Q = q(\theta_c^{train}|s^t, a^t)$
        Sample random mini-batches of experiences from replay memory $\mathscr{M}$ with size $N_B$.
        Update the training critic network $\theta_c^{train}$ by 5.1.
        Update the training actor network $\theta_a^{train}$ by 5.3.
        Update the target critic network $\theta_c^{target}$ by 5.4.
        Update the target actor network $\theta_a^{target}$ by 5.5.
        update the state $s^t = s^{t+1}$
    end for
end for
Output: Obtain the optimal action $a^*$ via the actor network.
```

$$\theta_a^{target} \leftarrow \tau_a \theta_a^{train} + (1 - \tau_a)\theta_a^{target}. \tag{5.5}$$

The DDPG based algorithm for optimization problem is shown in Algorithm 2.

5.2 Numerical Results

(State)

The current state includes the channel information, the last time slot beamforming vector at source and reflection coefficients at IRS.

$$s^t = \left[\mathbf{h}_{\mathrm{AB}}, \mathbf{h}_{\mathrm{IB}}, \mathbf{h}_{\mathrm{WB}}, \mathbf{h}_{\mathrm{WI}}, \mathbf{H}_{\mathrm{AI}}, \mathbf{h}_{\mathrm{IW}}, \mathbf{h}_{\mathrm{AW}}, \mathbf{W}^{t-1}, \mathbf{\Theta}^{t-1}\right] \tag{5.6}$$

Due to the fact that neural networks only accept real inputs, it is necessary to convert the imaginary parts of the system's channels into real inputs for the neural network.

(Action)

The action space consists of the beamforming vector at source (Alice) and the updated reflection coefficients at the IRS based on the current state.

$$a^t = \left[\mathbf{W}^t, \boldsymbol{\Theta}^t\right] = \left[\mathbf{w}_1^t, \ldots, \mathbf{w}_N^t, \boldsymbol{\theta}_1^t, \ldots, \boldsymbol{\theta}_L^t\right] \tag{5.7}$$

The agent controls both IRS reflection coefficients and beamforming vector, the size of the actions list is $L+2N$. For the beamforming vector, we represent it as a real-valued vector of size $2N$, including real and imaginary parts.

(Reward)

The objective is to maximize the received SINR. Thus, the received SINR at legal user Bob is used as the reward.

$$r^t = \gamma^t \tag{5.8}$$

γ^t is denoted as expression 3.3.

The hyper-parameters used in the algorithm are shown in Table 5.1. The transmit power at Alice is P_s = 9 dBW. The noise variance is $\sigma^2 = 10^{-5}$. The numbers of reflective elements at the IRS is 60. The actor and critic networks in the DDPG agent consisted of two hidden layers, each with 2048 and 1024 neurons. Although the neural network used in this work is not very deep by deep learning standards, it still qualifies as a deep neural network due to having multiple hidden layers and the ability to learn complex, nonlinear mappings from state to action. It required approximately 6 hours to train this model.

The DDPG agent was able to learn an optimal policy that dynamically adjusts transmission strategies based on the observed environment, without the need to repeatedly solve complex optimization problems. This enabled real-time adaptation in response to changes in channel conditions and jamming power. Although training the DDPG model required an offline phase with multiple episodes of environment interaction, once trained, the agent could

Table 5.1 Hyper-Parameters for the DDPG based algorithm.

Parameter Description	Value
learning rate for training critic network	0.0001
learning rate for training actor network	0.0001
learning rate for target critic network	0.0001
learning rate for target actor network	0.0001
decaying rate for the training networks	0.00001
discounted rate for future reward	0.99
size of experience reply	100000
size of mini-batch	128

quickly determine optimal actions in real time. This is especially useful for fast changing environments where classical optimization would be too slow or computationally demanding. Simulation results showed that the DDPG based approach outperformed traditional baseline methods, especially in dynamic or uncertain environments. The learned policy was able to enhance the SINR at the legitimate receiver while simultaneously minimizing the received signal at the eavesdropper, thus boosting the overall secrecy rate.

Figure 5.2 shows the result of average rewards versus time steps under different transmit powers P_a at the Willie. As the transmit power P_a emitted by the eavesdropper Willie increases, the average reward decreases. The average reward increases as training progresses and it demonstrates that the DDPG agent successfully learns an effective policy over time, gradually improving the secrecy rate through optimized beamforming and IRS phase shifts. The proposed DDPG based scheme can adaptively optimize system parameters to improve secrecy performance, even in the presence of active jamming.

Figure 5.3 shows the result of average rewards versus time steps under different numbers of reflective elements at the IRS. With the increase of elements, the average rewards also increase gradually. As the number of IRS reflection units increases, the convergence of the secure communication system designed in this thesis improves. More IRS elements enable finer grained control over the reflected signal phase, thereby enhancing constructive interference at Bob and destructive interference at Willie. Deploying more IRS elements leads to better physical layer security performance of this system.

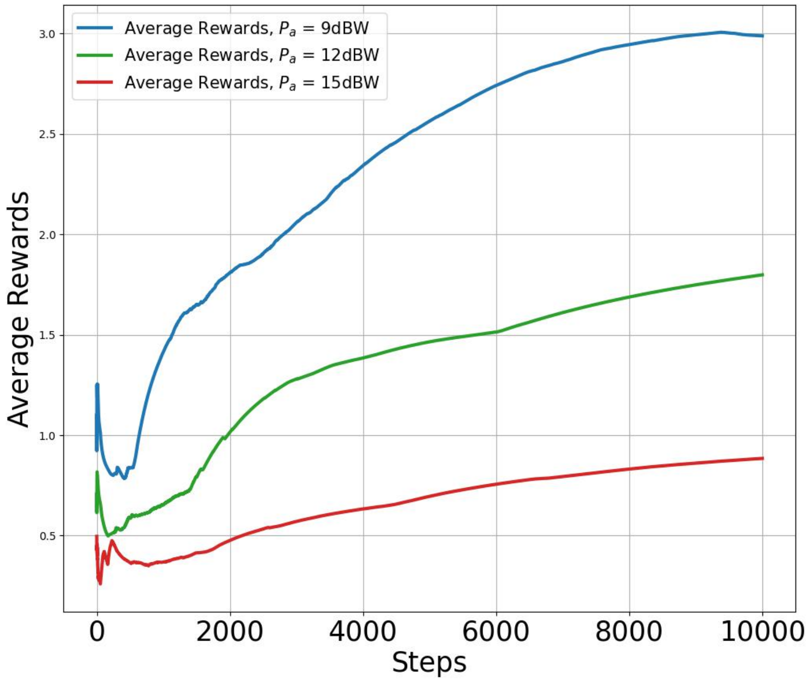

Fig. 5.2 Average reward versus time steps under different transmit powers P_a at the Willie.

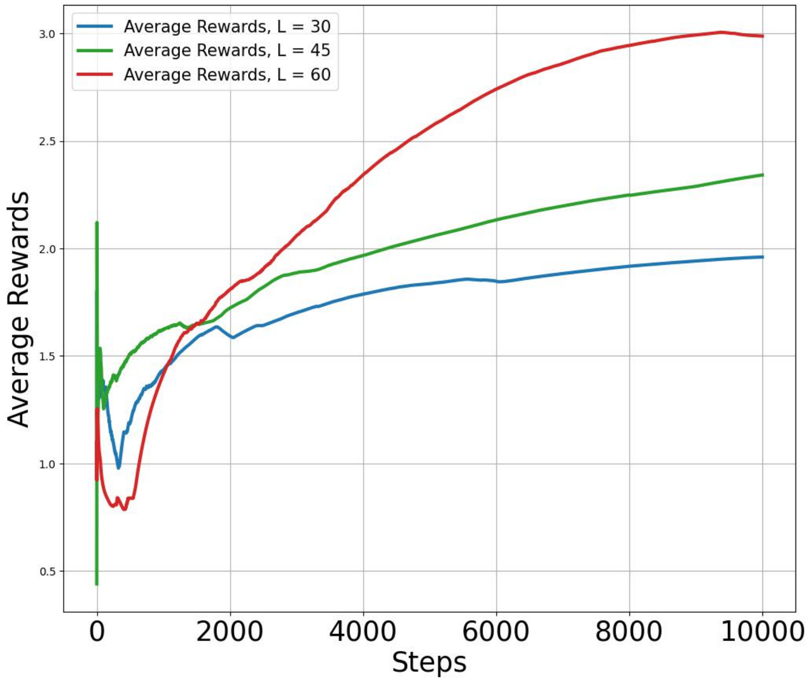

Fig. 5.3 Average reward versus time steps under different numbers of reflecting elements L at the IRS.

Chapter 6

Conclusions and Future Work

IRS reconstructs the channel of wireless communication systems by controlling the amplitude and phase of the incident signal, enhancing the required signal while suppressing interference signals. Due to its high flexibility, low power consumption, and low hardware complexity, IRS can effectively improve the performance of communication systems in various wireless communication scenarios.

The combination of backscatter technology and IRS can effectively utilize the interference signal sent by the active eavesdropper to the system, enhancing the desired signal at the legitimate user. Especially when the interference signal power of the eavesdropper is higher, the energy that IRS can perform backscattering is higher, which significantly improves the safety performance of the system.

The DRL algorithm learns how to take action to maximize total rewards through the interaction between the agent and the environment. Reinforcement learning can learn and adjust strategies in dynamic and uncertain environments, adapt to environmental changes, and is suitable for a variety of complex tasks. DDPG adopts a deterministic policy, which enables more stable training under a deterministic policy. It effectively reduces the correlation of samples and improves the stability of training through experience replay mechanism and target network.

The objectives achieved are summarized as follows.

- We proposed a scheme using the combination of IRS and backscatter techniques to combat active eavesdropping. We provided a specific communication environment for the proposed scheme. The source (Alice) sends the desired signal to the legitimate user (Bob), while the eavesdropper (Willie) sends the interference signal. IRS is deployed in the system to convert the interference signal into the desired signal through the use of backscatter technology. We developed a MISO system, Alice has multiple antennas, while Bob and Willie both have single antenna.

- We jointly optimized the transmit power at Alice and the reflection coefficient at IRS to maximize Bob's SINR under the SINR constraint at Willie. The optimization problem is mathematically defined and transformed into a nonconvex problem, which can be solved by algorithm.

- We proposed an alternating optimization algorithm to solve the non-convex problem. It was reformulated using matrix trace and vectorization identities and solved using SDR within an AO framework. Alternating optimization is more suitable for optimization problems with clear structures that can be decomposed into multiple sub problems, and it is a relatively stable and easy to implement method. The simulation results verify the proposed scheme is effective in combating active eavesdropping. It also have verified that the proposed scheme which combined the IRS and backscattering outperforms other schemes without IRS and traditional IRS schemes in terms of security performance

- We developed a DDPG-based algorithm to solve the optimization problem. DDPG is suitable for RL problems that require handling continuous action spaces, especially in complex and dynamic environments. It has strong learning ability and flexibility, but requires high exploration strategies, training stability, and computing resources. The simulation results show that the DDPG algorithm can significantly improve the secrecy performance of the system compared to AO algorithm.

Building on the current work, several directions can be explored in future research.

- Due to the fact that the model established in this article is based on the assumption of perfect CSI, which is difficult in practical situations, future work will first investigate the estimation of CSI or further study the above issues under imperfect CSI.

- With the development of communication technology, different application scenarios will involve different or even more complex communication systems. Extending the system model to include multiple legitimate users or multiple eavesdroppers would provide insight into more practical and complex security scenarios. Therefore, in the following work, various IRS-aided systems will be studied.

- DDPG uses deterministic strategies, which can easily fall into suboptimal strategies, leading to insufficient exploration and other disadvantage. In the future work, We will study other DRL algorithms to compensate for the shortcomings of the DDPG algorithm. In addition, combining traditional optimization algorithms AO and DRL to leverage their advantages.

- We will consider the practicability of the proposed scheme using a real-world hardware based testbeds to verify the IRS-assisted secure systems.

References

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

[2] Z. Chu, W. Hao, P. Xiao, and J. Shi, “Intelligent reflecting surface aided multi-antenna secure transmission,” *IEEE Wireless Communications Letters*, vol. 9, no. 1, pp. 108–112, 2020.

[3] X. Li, Z. Xie, Z. Chu, V. G. Menon, S. Mumtaz, and J. Zhang, “Exploiting benefits of irs in wireless powered noma networks,” *IEEE Transactions on Green Communications and Networking*, vol. 6, no. 1, pp. 175–186, 2022.

[4] F. Naeem, G. Kaddoum, S. Khan, K. S. Khan, and N. Adam, “Irs-empowered 6g networks: Deployment strategies, performance optimization, and future research directions,” *IEEE Access*, vol. 10, pp. 118 676–118 696, 2022.

[5] W. Aman, M. M. Ur Rahman, S. Ansari, A. A. Nasir, K. Qaraqe, M. A. Imran, and Q. H. Abbasi, “On the effective capacity of irs-assisted wireless communication,” *Physical Communication*, vol. 47, p. 101339, 2021.

[6] M.-M. Zhao, Q. Wu, M.-J. Zhao, and R. Zhang, “Intelligent reflecting surface enhanced wireless networks: Two-timescale beamforming optimization,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 2–17, 2021.

[7] S. Lin, B. Zheng, G. C. Alexandropoulos, M. Wen, M. D. Renzo, and F. Chen, “Reconfigurable intelligent surfaces with reflection pattern modulation: Beamforming design

and performance analysis," *IEEE Transactions on Wireless Communications*, vol. 20, no. 2, pp. 741–754, 2021.

[8] L. Zhang, Y. Wang, W. Tao, Z. Jia, T. Song, and C. Pan, "Intelligent reflecting surface aided mimo cognitive radio systems," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 10, pp. 11 445–11 457, 2020.

[9] 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, 2019.

[10] H.-M. Wang, J. Bai, and L. Dong, "Intelligent reflecting surfaces assisted secure transmission without eavesdropper's csi," *IEEE Signal Processing Letters*, vol. 27, pp. 1300–1304, 2020.

[11] S. Hong, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "Artificial-noise-aided secure mimo wireless communications via intelligent reflecting surface," *IEEE Transactions on Communications*, vol. 68, no. 12, pp. 7851–7866, 2020.

[12] J. Kimionis, A. Bletsas, and J. N. Sahalos, "Bistatic backscatter radio for power-limited sensor networks," in *2013 IEEE Global Communications Conference (GLOBECOM)*, 2013, pp. 353–358.

[13] Y. Zhang, F. Gao, L. Fan, X. Lei, and G. K. Karagiannidis, "Secure communications for multi-tag backscatter systems," *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1146–1149, 2019.

[14] K. Feng, Q. Wang, X. Li, and C.-K. Wen, "Deep reinforcement learning based intelligent reflecting surface optimization for miso communication systems," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 745–749, 2020.

[15] Z. Yang, Y. Liu, Y. Chen, and J. T. Zhou, "Deep reinforcement learning for ris-aided non-orthogonal multiple access downlink networks," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6.

[16] H. Yang, Y. Zhao, Z. Xiong, J. Zhao, D. Niyato, K.-Y. Lam, and Q. Wu, “Deep reinforcement learning based intelligent reflecting surface for secure wireless communications,” in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6.

[17] C. Wu, S. Yan, X. Zhou, R. Chen, and J. Sun, “Intelligent reflecting surface (irs)-aided covert communication with warden’s statistical csi,” *IEEE Wireless Communications Letters*, vol. 10, no. 7, pp. 1449–1453, 2021.

[18] S. Idrees, X. Jia, S. Durrani, and X. Zhou, “Design of intelligent reflecting surface (irs)-boosted ambient backscatter systems,” *IEEE Access*, vol. 10, pp. 65 000–65 010, 2022.

[19] S. Xu, J. Liu, and J. Zhang, “Resisting undesired signal through irs-based backscatter communication system,” *IEEE Communications Letters*, vol. 25, no. 8, pp. 2743–2747, 2021.

[20] C. Huang, R. Mo, and C. Yuen, “Reconfigurable intelligent surface assisted multiuser miso systems exploiting deep reinforcement learning,” *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1839–1850, 2020.

[21] X. Jia and X. Zhou, “Irs-assisted ambient backscatter communications utilizing deep reinforcement learning,” *IEEE Wireless Communications Letters*, vol. 10, no. 11, pp. 2374–2378, 2021.

[22] H. Shen, W. Xu, S. Gong, Z. He, and C. Zhao, “Secrecy rate maximization for intelligent reflecting surface assisted multi-antenna communications,” *IEEE Communications Letters*, vol. 23, no. 9, pp. 1488–1492, 2019.

[23] J. Chen, Y.-C. Liang, Y. Pei, and H. Guo, “Intelligent reflecting surface: A programmable wireless environment for physical layer security,” *IEEE Access*, vol. 7, pp. 82 599–82 612, 2019.

[24] X. Yu, D. Xu, and R. Schober, “Enabling secure wireless communications via intelligent reflecting surfaces,” in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[25] L. Dong and H.-M. Wang, “Enhancing secure mimo transmission via intelligent reflecting surface,” *IEEE Transactions on Wireless Communications*, vol. 19, no. 11, pp. 7543–7556, 2020.

[26] J. Tang, H. Wen, H.-H. Song, and R.-F. Wang, “On the design of irs-assisted directional reflection for physical layer secure transmission,” in *2021 IEEE 21st International Conference on Communication Technology (ICCT)*, 2021, pp. 852–857.

[27] 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, 2020.

[28] Z. Peng, R. Weng, C. Pan, G. Zhou, M. D. Renzo, and A. L. Swindlehurst, “Robust transmission design for ris-assisted secure multiuser communication systems in the presence of hardware impairments,” *IEEE Transactions on Wireless Communications*, vol. 22, no. 11, pp. 7506–7521, 2023.

[29] W. Saad, X. Zhou, Z. Han, and H. V. Poor, “On the physical layer security of backscatter wireless systems,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 6, pp. 3442–3451, 2014.

[30] N. Van Huynh, D. T. Hoang, X. Lu, D. Niyato, P. Wang, and D. I. Kim, “Ambient backscatter communications: A contemporary survey,” *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2889–2922, 2018.

[31] Y. An, H. Park, and W. Lee, “Signal strength balanced scheduling for secure ambient backscatter networks,” in *2023 International Conference on Information Networking (ICOIN)*, 2023, pp. 56–61.

[32] Q. Jing, A. V. Vasilakos, J. Wan, J. Lu, and D. Qiu, “Security of the internet of things: perspectives and challenges,” *Wireless Networks*, vol. 20, no. 8, p. 2481–2501, 2014.

[33] P. Tedeschi, S. Sciancalepore, and R. Di Pietro, “Security in energy harvesting networks: A survey of current solutions and research challenges,” *IEEE Communications Surveys & Tutorials*, vol. 22, no. 4, pp. 2658–2693, 2020.

[34] Q. Yang, H.-M. Wang, Q. Yin, and A. L. Swindlehurst, “Exploiting randomized continuous wave in secure backscatter communications,” *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 3389–3403, 2020g.

[35] J. You, G. Wang, and Z. Zhong, “Physical layer security-enhancing transmission protocol against eavesdropping for ambient backscatter communication system,” in *6th International Conference on Wireless, Mobile and Multi-Media (ICWMMN 2015)*, 2015, pp. 43–47.

[36] J. Li, P. Wang, L. Jiao, Z. Yan, K. Zeng, and Y. Yang, “Security analysis of triangle channel-based physical layer key generation in wireless backscatter communications,” *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 948–964, 2023.

[37] T. Eisenbarth, S. Kumar, C. Paar, A. Poschmann, and L. Uhsadel, “A survey of lightweight-cryptography implementations,” *IEEE Design & Test of Computers*, vol. 24, no. 6, pp. 522–533, 2007.

[38] S. N. Premnath, S. Jana, J. Croft, P. L. Gowda, M. Clark, S. K. Kasera, N. Patwari, and S. V. Krishnamurthy, “Secret key extraction from wireless signal strength in real environments,” *IEEE Transactions on Mobile Computing*, vol. 12, no. 5, pp. 917–930, 2013.

[39] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, “Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design,” *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 14 960–14 973, 2020.

[40] L. Gu and A. Mohajer, “Joint throughput maximization, interference cancellation, and power efficiency for multi-irs-empowered uav communications,” *Signal, Image and Video Processing*, vol. 18, no. 5, pp. 4029–4043, 2024.

[41] D. Xu, X. Yu, Y. Sun, D. W. K. Ng, and R. Schober, “Resource allocation for secure irs-assisted multiuser miso systems,” in *2019 IEEE Globecom Workshops (GC Wkshps)*, 2019, pp. 1–6.

[42] Z. Ji, W. Yang, X. Guan, X. Zhao, G. Li, and Q. Wu, “Trajectory and transmit power optimization for irs-assisted uav communication under malicious jamming,” *IEEE Transactions on Vehicular Technology*, vol. 71, no. 10, pp. 11 262–11 266, 2022.

[43] 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, 2022.

[44] A. Rezaei, A. Khalili, J. Jalali, H. Shafiei, and Q. Wu, “Energy-efficient resource allocation and antenna selection for irs-assisted multicell downlink networks,” *IEEE Wireless Communications Letters*, vol. 11, no. 6, pp. 1229–1233, 2022.

[45] S. Noh, H. Yu, and Y. Sung, “Training signal design for sparse channel estimation in intelligent reflecting surface-assisted millimeter-wave communication,” *IEEE Transactions on Wireless Communications*, vol. 21, no. 4, pp. 2399–2413, 2021.

[46] H. Yang, Z. Xiong, J. Zhao, D. Niyato, Q. Wu, M. Tornatore, and S. Secci, “Intelligent reflecting surface assisted anti-jamming communications based on reinforcement learning,” in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–6.

[47] C. Zhong, M. Cui, G. Zhang, Q. Wu, X. Guan, X. Chu, and H. V. Poor, “Deep reinforcement learning-based optimization for irs-assisted cognitive radio systems,” *IEEE Transactions on Communications*, vol. 70, no. 6, pp. 3849–3864, 2022.

[48] P. D. Thanh, H. T. H. Giang, and I.-P. Hong, “Anti-jamming ris communications using dqn-based algorithm,” *IEEE Access*, vol. 10, pp. 28 422–28 433, 2022.

[49] F. Xu, T. Hussain, M. Ahmed, K. Ali, M. A. Mirza, W. U. Khan, A. Ihsan, and Z. Han, “The state of ai-empowered backscatter communications: A comprehensive survey,” *IEEE Internet of Things Journal*, vol. 10, no. 24, pp. 21 763–21 786, 2023.

[50] Y. Cheng, K. H. Li, Y. Liu, K. C. Teh, and H. Vincent Poor, “Downlink and uplink intelligent reflecting surface aided networks: Noma and oma,” *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 3988–4000, 2021.