sensors

Article

Learning-Based IRS-Assisted Secure Transmission for Mine IoTs

Minghui Min [1,2], Jiayang Xiao [1], Peng Zhang [1], Jinling Song [1,*] and Shiyin Li [1]

[1] School of Information and Control Engineering, China University of Mining and Technology, Xuzhou 221116, China; minmh@cumt.edu.cn (M.M.); xjy2807175506@gmail.com (J.X.)
[2] Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education and School of Cyber Science and Engineering, Wuhan University, Wuhan 430072, China
* Correspondence: jinlingsong@cumt.edu.cn

Abstract: Mine Internet of Things (MIoT) devices in intelligent mines often face substantial signal attenuation due to challenging operating conditions. The openness of wireless communication also makes it susceptible to smart attackers, such as active eavesdroppers. The attackers can disrupt equipment operations, compromise production safety, and exfiltrate sensitive environmental data. To address these challenges, we propose an intelligent reflecting surface (IRS)-assisted secure transmission system for an MIoT device which enhances the security and reliability of wireless communication in challenging mining environments. We develop a joint optimization problem for the IRS phase shifts and transmit power, with the goal of enhancing legitimate transmission while suppressing eavesdropping. To accommodate time-varying channel conditions, we propose a reinforcement learning (RL)-based IRS-assisted secure transmission scheme that enables MIoT device to optimize both the IRS reflecting coefficients and transmit power for optimal transmission policy in dynamic environments. We adopt the deep deterministic policy gradient (DDPG) algorithm to explore the optimal transmission policy in continuous space. This can reduce the discretization error caused by traditional RL methods. The simulation results indicate that our proposed scheme achieves superior system utility compared with both the IRS-free (IF) scheme and the IRS randomly configured (IRC) scheme. These results demonstrate the effectiveness and practical relevance of our contributions, proving that implementing IRS in MIoT wireless communication can enhance safety, security, and efficiency in the mining industry.

Keywords: Internet of things; mining; active eavesdropping; intelligent reflecting surface; reinforcement learning

Citation: Min, M.; Xiao, J.; Zhang, P.; Song, J.; Li, S. Learning-Based IRS-Assisted Secure Transmission for Mine IoTs. *Sensors* **2023**, *23*, 6321. https://doi.org/10.3390/s23146321

Academic Editors: Karim Seddik, Radwa Sultan and Hongliang Zhang

Received: 7 June 2023
Revised: 4 July 2023
Accepted: 6 July 2023
Published: 12 July 2023

1. Introduction

Mine Internet of Things (MIoT) devices are widely applied in intelligent mines to improve safety and mineral production [1]. In the mining industry, IoT networks play a crucial role in controlling mining equipment and gathering essential environmental data, including temperature, humidity, and wind speed, which are instrumental in safeguarding the personal safety of mine workers [2,3]. The accurate, reliable, and durable operation of MIoT devices is essential for the stable and long-term service of intelligent mines. Therefore, MIoT devices must provide high-speed transmission and low energy consumption. However, the wireless transmission characteristics of electromagnetic waves in MIoT often experience severe scattering, substantial interference, and Non-Line-of-Sight (NLoS) propagation, which necessitates innovation in new transmission technology [4].

Furthermore, despite significant advancements in wireless communication technology in recent decades, most MIoT networks, particularly those deployed in open pit mines, remain susceptible to physical layer threats. The open nature of wireless channels exposes these MIoT devices to vulnerabilities such as jamming and eavesdropping, highlighting the need for enhanced security measures [5]. Malicious devices connected to the system can wiretap confidential information, which can lead to data leakages, such as mineral production schedules, the distribution of mineral resources, and safety aspects of the operations. The intruder can use the stolen data to commit fraud and extortion for illegal profit or pose security threats, such as negative impacts on production and deliberately creating catastrophes. In this case, MIoT devices must be able to withstand smart attacks, particularly active eavesdropping, which involves simultaneous eavesdropping and jamming to increase the MIoT device's transmit power and intercept more data [6].

As an emerging technology, intelligent reflecting surfaces (IRS) have attracted extensive research interest. The low cost of IRS makes them a highly suitable technology for wide adoption in MIoT communication. IRS contain metamaterial designed to reflect the incident waves from the source towards the destination [7,8]. With properly adjusted elements, IRS can construct an artificial Line-of-Sight (LoS) link and significantly improve transmission performance in NLoS scenes. Moreover, adding the nonreflected signal and the IRS-reflected signal at the eavesdropper can produce destructive interference, effectively suppressing eavesdropping activity [9]. In this paper, IRS establish a favorable propagation environment, increasing the access point (AP)'s received signal power and decreasing the eavesdropper's received signal power, thus increasing the secrecy rate of the MIoT system in the presence of active eavesdropping.

Due to the complex and random time-varying channel characteristics in MIoT, acquiring the optimal transmission scheme using traditional techniques is typically not feasible [10]. The wiretap policy is also challenging to estimate, making it harder to find the optimum secure transmission policy. Motivated by the advances in model-free deep reinforcement learning (DRL), we model the secure transmission procedure as a Markov Decision Process (MDP). The increasing computational capability of IoT devices, such as the Qualcomm Snap-dragon 800 [11], makes it possible to apply DRL techniques in practical mining IoT communication systems.

In this paper, we propose an innovative secure transmission scheme that leverages IRS and the deep deterministic policy gradient (DDPG) algorithm to enhance the secrecy rate of the system in the presence of an active eavesdropper, specifically in a dynamic MIoT environment. In the proposed scheme, RL is utilized to adapt the time-varying channel characteristics and make the optimal choice without knowing the specific transmission model and attack model. The DDPG-based scheme can select policies in a continuous space while avoiding discretization errors. This enables the MIoT device to jointly optimize the IRS phase shifts and the MIoT device's transmit power in a mine environment. Strategically adjusting the phase shifts and transmission power of IRS, as well as leveraging the utilization of reflected signals, is helpful to enhance the effectiveness of legitimate transmission and ensure a safe mine environment.

According to our simulation results, the proposed DDPG-based IRS-assisted secure transmission (DIST) scheme achieves higher utility than the IRS randomly configured (IRC) scheme and the IRS-free (IF) scheme. By changing the number of IRS elements, we also assess the system utility of both the proposed DIST scheme and the IRC scheme. The main contributions of this paper can be outlined as follows:

- We construct a joint optimization problem of the MIoT device's transmit power and IRS reflecting beamforming to maximize the system's utility. We present an IRS-assisted secure transmission scheme against active eavesdropping in MIoT.
- A DRL-based intelligent beamforming and power control framework is presented to achieve the optimal IRS phase shifts and MIoT device's transmit power. We formulate the control of the IRS elements as an MDP and employ the DDPG algorithm to achieve real-time and continuous phase control based on the dynamic MIoT environment.
- Simulation results demonstrate our proposed DRL-based IRS-assisted secure transmission scheme's performance suppresses eavesdropping and enhances legitimate transmission compared with the IRC and IF schemes.

The subsequent sections of this paper are organized as follows: Section 2 discusses the related works. Section 3 introduces the proposed system model, channel model, and problem formulation. In Section 4, we introduce our proposed DIST scheme. Section 5 provides simulation results, and Section 6 concludes the paper.

Notations: In this paper, we present matrices and vectors with boldface. $(.)^T$ and $(.)^H$ denote the transpose and conjugate transpose operations, respectively. $diag(.)$ denotes a diagonal matrix, and j is the imaginary unit. $\mathbb{E}[.]$ denotes the expectation operation. $|.|$ is the absolute value of a scalar. $\mathbb{C}^{M \times N}$ denotes a complex-valued matrix with a size of $M \times N$.

2. Related Works

Numerous methods have been proposed to improve physical layer security (PLS) performance, including artificial noise (AN) [12], physical layer authentication (PLA) [13], and beamforming [14]. However, these methods have limitations, such as extra power consumption for AN, computing resource requirements for PLA, and limited security guarantees for beamforming. For mining scenario, the authors in [15] investigate PLS in an underground mine environment using an amplify-and-forward relay-aided system with multiple eavesdroppers. The authors employ a block coordinate descent algorithm to design the precoding and jamming matrix at both the source and the relay, similar to other traditional PLS techniques, rather than during the propagation process. Recently, the use of IRS has gained significant attention to address PLS issues in the propagation process. Several studies have explored the use of IRS in secure communication systems in [16–19]. A genetic algorithm (GA) is introduced in [16] to optimize the phase shift of an IRS in a multiple-input multiple-output (MIMO) system, with the goal of improving security performance in the presence of an eavesdropper. To reduce the overhead of computing resources, a low-complexity algorithm is studied in [17] based on fractional programming (FP) and manifold optimization (MO) to circumvent the nonconvex optimization problem and obtain near-optimal IRS phase shifts. However, the optimization technique in both [16,17] rely on a specific transmission model and lack robustness. Moreover, a more practical system model comprising multiple eavesdroppers and imperfect channel state information (CSI) are studied in [18,19]. The interuser interference (IUI) among each mobile user (MU) is studied in [19]. Additionally, none of the existing IRS-assisted PLS approaches consider an active eavesdropper scenario where jamming attacks interfere with the legitimate transmission and raise the transmit power.

Artificial Intelligence (AI) has introduced a new way to solve PLS problems through RL. Recent studies in [20,21] have considered PLS problems concerning smart attackers conducting jamming, eavesdropping, and spoofing attacks. For instance, prospect theory (PT) in an unmanned aerial vehicle (UAV) transmission system is investigated in [20], where the attacker is considered to be selfish and subjective. To enhance the secrecy performance and the utility of the legitimate UAV, a power allocation approach utilizing deep Q-networks (DQN) is put forth to determine the optimal policy, in cases where the attack and channel models are unknown. RL techniques are studied in [21] to configure IRS beamforming design. The authors first establish the interaction between the base station (BS) and the smart attacker as a non-cooperative game and derive the Nash equilibrium of the game. Then, a DQN-based antismart attacker strategy is proposed to make the BS and IRS intelligent and restrain the attack, thus improving the system's security. However, since the study assumes a static channel, the proposed strategy may be less adaptable to varying channel conditions, despite its focus on the game-theoretic interaction between the base station (BS) and the attacker. To address these limitations, a novel DRL framework is proposed in [22] to enable the prediction of IRS reflection matrices without the need for extensive channel estimation or beamforming train overhead. Additionally, an integrated DRL and extremum-seeking control (ESC) is studied in [23] to control the IRS and make the system more adaptive to the dynamic channel state without subchannel CSI.

The implementation of IRS and RL in the mining industry is a relatively unexplored research area. Machine learning is applied in a mining system to remove the operator from hazardous environments without compromising task execution [24]. Ref. [25] is the first work implementing IRS in a coal mine. In this study, IRS are placed at the inflection points of the nonlinear routes (i.e., zigzag tunnels) to improve wireless communication quality. Although an approximation-based algorithm is utilized to address the optimization problem, the complex and dynamic nature of the channel state is ignored. Thus, the proposed method in [25] may not be practical in most mining scenarios. Furthermore, neither [24] nor [25] use RL or IRS to solve the PLS problems and enhance secure transmission in a mine environment. The mainly related work is summarized in Table 1.

Table 1. Summary of related literature.

Classification	Related Works	Key Contributions
Traditional techniques used to achieve PLS	[12–15]	- Artificial noise, physical layer authentication, and beamforming techniques. - PLS in underground mine environments with relay-aided systems.
Implementation of IRS in PLS	[16–19]	- Genetic algorithms for IRS phase shift optimization. - Low-complexity algorithms for IRS optimization. - Researches on practical system models.
Implementation of IRS and RL to achieve PLS	[20–23]	- Power allocation using deep Q-networks for UAV systems. - IRS beamforming design using RL techniques. - DRL-based adaptive IRS control.
Implementation of IRS and machine learning in mining systems	[24,25]	- Machine learning applications in mining systems. - Implementation of IRS in coal mines.

3. System Model and Problem Formulation

3.1. System Model

Considering a single-input single-output (SISO) uplink system, as shown in Figure 1, one MIoT device, equipped with a single antenna, establishes communication with a single-antenna AP. Simultaneously, we introduce a single-antenna active eavesdropper with the intention of intercepting the transmission. The MIoT device collects data, such as temperature and gas density, and transmits the data to the AP, which is located $d_{M,A}$ meters away. To establish a dependable communication environment, a passive IRS is deployed at a distance of $d_{M,I}$ meters from the MIoT device, with $N = N_y \times N_z$ reflecting elements. All elements are configured through a wireless IRS controller that receives the control signal from the MIoT device. The IRS reflect the signal to enhance the transmission from the MIoT device to the AP and suppress the wiretap signal at the eavesdropper, thereby obtaining the maximum secrecy rate. The data are then updated to a cloud server and used by the remote control on the ground for digital management in the mine IoT applications.

Upon receiving the control signal, the micro IRS controller sets the bias voltage to apply the phase shift on each IRS reflecting element. The phase shifts configuration can be modeled as $\Theta = diag\left(\beta_1 e^{j\theta_1}, \beta_2 e^{j\theta_2}, \ldots, \beta_N e^{j\theta_N}\right)$, where $\beta_N \in [0,1]$ and $\theta_N \in [0,2\pi)$ are the amplitude reflection coefficient and phase shift of the n-th IRS element, respectively. For simplicity, we set $\beta_N = 1$ for N reflecting elements.

The transmission policy in an IRS-assisted secure transmission system relies on precisely acquiring CSI. In our proposed system model, the legitimate channel state is obtained by the pilot-based channel estimation [26]. We also assume the CSI of the wiretap channel to be perfectly known to the MIoT device. This is because the eavesdropper is considered an active user in the system but is not trusted by the legitimate receiver [9].

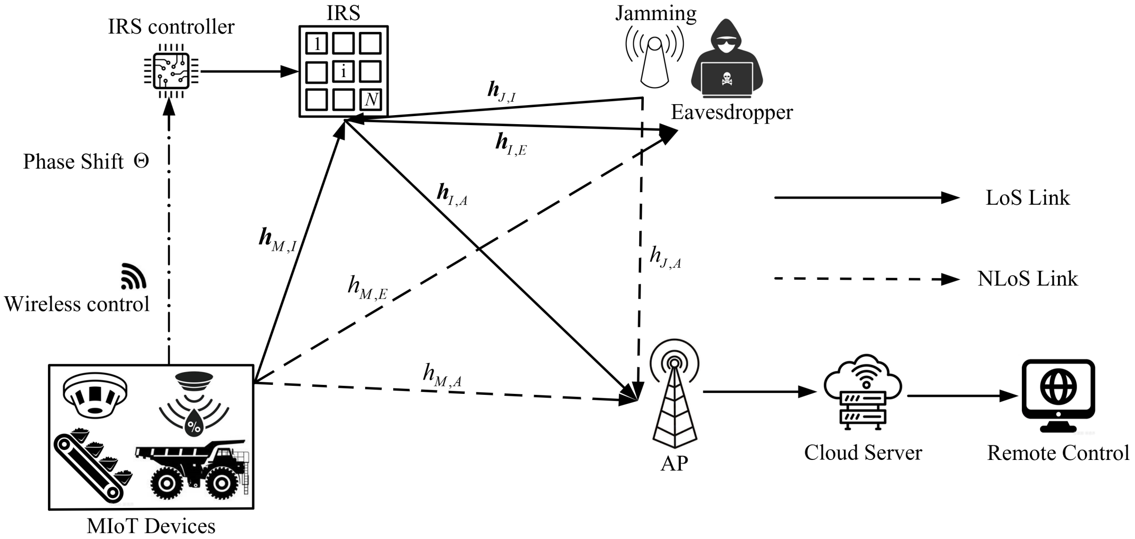


Figure 1. Illustration of the IRS-assisted secure transmission system in MIoT. The MIoT devices choose the transmit power and send phase shift control messages to the IRS controller. At the same time, the active eavesdropper performs jamming to increase the MIoT devices' transmit power for a higher wiretap rate.

3.2. Channel Model

The channel path losses from the MIoT device to the AP, from the MIoT to the eavesdropper, and from the jamming to the AP are denoted by $h_{M,A}$, $h_{M,E}$, and $h_{J,A}$. The channel path losses above are all regarded as Rayleigh fading, which means that the Line-of-Sight signal between the transmitter and receiver is blocked and can be expressed as [27]:

$$\begin{aligned} h_{M,A} &= \sqrt{PL_{M,A}}\widetilde{h}_{M,A} \\ h_{M,E} &= \sqrt{PL_{M,E}}\widetilde{h}_{M,E} \\ h_{J,A} &= \sqrt{PL_{J,A}}\widetilde{h}_{J,A} \end{aligned} \tag{1}$$

where PL is the path loss. $\widetilde{h}$ contains independent and identically distributed (i.i.d) circularly symmetric complex Gaussian distribution with zero mean and unit variance, $\widetilde{h} \sim \mathcal{CN}(0,1)$.

The distance-dependent path loss PL is modeled as

$$PL = PL_0 - 10\xi \log_{10} \frac{d}{d_0} \tag{2}$$

where $PL_0 = -30$ dB is the reference channel path loss for the reference distance $d_0 = 1$ m, ξ is the path loss exponent, and d is the distance from the transmitter to the receiver.

The channel path loss from the MIoT device to the IRS, from the IRS to the AP, from the IRS to the eavesdropper, and from the jamming to the IRS are denoted by $\mathbf{h}_{M,I} \in \mathbb{C}^{N\times1}$, $\mathbf{h}_{I,A} \in \mathbb{C}^{N\times1}$, $\mathbf{h}_{I,E} \in \mathbb{C}^{N\times1}$, and $\mathbf{h}_{J,I} \in \mathbb{C}^{N\times1}$. The channel path losses above are all assumed to be small-scale Rician fading, which suggests the LoS link coexists with NLoS link, and the channel path loss can be expressed as [7,23]

$$
\begin{aligned}
\mathbf{h}_{M,I} &= \sqrt{PL_{M,I}}\left(\sqrt{\frac{K_{M,I}}{K_{M,I}+1}}\overline{\mathbf{h}}_{M,I} + \sqrt{\frac{1}{K_{M,I}+1}}\widetilde{\mathbf{h}}_{M,I}\right) \\
\mathbf{h}_{I,A} &= \sqrt{PL_{I,A}}\left(\sqrt{\frac{K_{I,A}}{K_{I,A}+1}}\overline{\mathbf{h}}_{I,A} + \sqrt{\frac{1}{K_{I,A}+1}}\widetilde{\mathbf{h}}_{I,A}\right) \\
\mathbf{h}_{I,E} &= \sqrt{PL_{I,E}}\left(\sqrt{\frac{K_{I,E}}{K_{I,E}+1}}\overline{\mathbf{h}}_{I,E} + \sqrt{\frac{1}{K_{I,E}+1}}\widetilde{\mathbf{h}}_{I,E}\right) \\
\mathbf{h}_{J,I} &= \sqrt{PL_{J,I}}\left(\sqrt{\frac{K_{J,I}}{K_{J,I}+1}}\overline{\mathbf{h}}_{J,I} + \sqrt{\frac{1}{K_{J,I}+1}}\widetilde{\mathbf{h}}_{J,I}\right)
\end{aligned}
\tag{3}
$$

where K is the Rician-K factor and denotes the proportion of power between the LoS link and the NLoS link. $\widetilde{\mathbf{h}}$ is the random components caused by multipath effect with i.i.d and $\mathcal{CN}(0,1)$ distributed elements. The deterministic component $\overline{\mathbf{h}}$ is position-dependent and can be expressed as [28]

$$\overline{\mathbf{h}} = \mathbf{h}^{A}\mathbf{h}^{D} \tag{4}$$

where the superscripts "A" and "D" stand for "Arrival" and "Departure", respectively.

Without loss of generality, we place the IRS on the yOz plane. So, the component $\mathbf{h}^{A(D)}$ in Equation (4) can be expressed as

$$\mathbf{h}^{A(D)} = \mathbf{h}^{A(D)}_{y-axis}\mathbf{h}^{A(D)}_{z-axis} \tag{5}$$

where

$$\mathbf{h}^{A(D)}_{z-axis} = \left[1, e^{-j\frac{2\pi}{\lambda_c}d\cos\psi}, \dots, e^{-j\frac{2\pi}{\lambda_c}d\cos\vartheta(N_z-1)}\right]^T \tag{6}$$

$$\mathbf{h}^{A(D)}_{y-axis} = \left[1, e^{-j\frac{2\pi}{\lambda_c}d\cos\vartheta\sin\psi}, \dots, e^{-j\frac{2\pi}{\lambda_c}d\cos\vartheta\sin\psi\left(N_y-1\right)}\right] \tag{7}$$

where λ_c is the carrier wavelength, and d is the distance between two adjacent IRS elements. Furthermore, ϑ represents the azimuth angle and ψ represents the elevation angle. The LoS component $\overline{\mathbf{h}}$ is solely dependent on ϑ and ψ, meaning that once the locations of each unit are obtained, $\overline{\mathbf{h}}$ is fully determined.

For the proposed system model, the MIoT device sends message m with zero mean and unit variance to the AP with transmission power p, where $\mathbb{E}\left\{|m|^2\right\} = 1$, $p \in [P_{\min}, P_{\max}]$, $P_{\min}$, and $P_{\max}$ is the minimum, and the maximum values of the MIoT device transmit power, respectively.

3.3. Problem Formulation

The received signal y_A at the AP and y_E [9] at the eavesdropper can be denoted as

$$y_A = \left(\mathbf{h}^{H}_{I,A}\Theta\mathbf{h}_{M,I} + h_{M,A}\right)\sqrt{p}m + n_k \tag{8}$$

$$y_E = \left(\mathbf{h}^{H}_{I,E}\Theta\mathbf{h}_{M,I} + h_{M,E}\right)\sqrt{p}m + n_k \tag{9}$$

where $n_k \sim \mathcal{CN}\left(0,\sigma^2\right)$ denotes the complex additive white Gaussian noise (AWGN).

The active eavesdropper aims to wiretap more data by increasing the jamming power p_J. Therefore, we assume that the active eavesdropper has no self-interference. That is to say, we ignore the LoS channel between the eavesdropper and the jamming device and only consider the IRS-reflected jamming signals. Thus, the received jamming signal J_A at the AP and J_E at the eavesdropper can be expressed as

$$J_A = \left(\mathbf{h}_{I,A}^H \Theta \mathbf{h}_{J,I} + h_{J,A}\right)\sqrt{p_J} + n_k \tag{10}$$

$$J_E = \left(\mathbf{h}_{I,E}^H \Theta \mathbf{h}_{J,I}\right)\sqrt{p_J} + n_k \tag{11}$$

Then, we can calculate the signal-to-interference-and-noise ratio (SINR) ρ_A at the AP and ρ_E at the eavesdropper [29], and they can be expressed as

$$\rho_A = \frac{\left|\left(\mathbf{h}_{I,A}^H \Theta \mathbf{h}_{M,I} + h_{M,A}\right)\sqrt{p}\right|^2}{\left|\left(\mathbf{h}_{I,A}^H \Theta \mathbf{h}_{J,I} + h_{J,A}\right)\sqrt{p_J}\right|^2 + \sigma^2} \tag{12}$$

$$\rho_E = \frac{\left|\left(\mathbf{h}_{I,E}^H \Theta \mathbf{h}_{M,I} + h_{M,E}\right)\sqrt{p}\right|^2}{\left|\left(\mathbf{h}_{I,E}^H \Theta \mathbf{h}_{J,I}\right)\sqrt{p_J}\right|^2 + \sigma^2} \tag{13}$$

We evaluate the eavesdropping policy according to the AP's received jamming power $\widetilde{p}_J$, which can be denoted as

$$\widetilde{p}_J = \left|\left(\mathbf{h}_{I,A}^H \Theta \mathbf{h}_{J,I} + h_{J,A}\right)\sqrt{p_J}\right|^2 \tag{14}$$

The achievable rates at AP R_A and eavesdropper R_E in bps/Hz can be denoted as [6,19]

$$R_A = \log_2\left(1 + \rho_A\right) \tag{15}$$

$$R_E = \log_2\left(1 + \rho_E\right) \tag{16}$$

Thus, the achievable secrecy rate R_{sec} [19,30] can be denoted as

$$R_{sec} = R_A - R_E \tag{17}$$

To achieve the maximum secrecy rate R_{sec}, there is a trade-off in configuring the IRS reflecting coefficient matrix Θ. On the one hand, we synchronize the phase of the reflected channel $\mathbf{h}_{I,A}^H \Theta \mathbf{h}_{M,I}$ with the direct channel $h_{M,A}$ to strengthen AP's received signal and thus maximize R_A. On the other hand, we reverse synchronize the phase of the reflect channel $\mathbf{h}_{I,E}^H \Theta \mathbf{h}_{M,I}$ with the direct channel $h_{M,E}$ to weaken the eavesdropper's received signal and decrease R_E.

Then, the MIoT system's utility function [6] is defined as follows:

$$U(\boldsymbol{\theta}, p) = \omega_1 R_{sec} - \omega_2 p \tag{18}$$

where $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_n, \ldots, \theta_N]$, $\forall n \in \{1, 2, \ldots, N\}$; weights ω_1 and ω_2 denote the coefficients. The coefficients ω_1 and ω_2 represent the weight of the achievable secrecy rate and the transmit power, which are set for balancing the influence factors of the utility function.

We aim to optimize the IRS phase shifts $\boldsymbol{\theta}$ and the MIoT device's transmit power p to maximize the utility. The following formulation represents the optimization problem:

$$\begin{aligned} \max_{\boldsymbol{\theta}, p} \quad & U(\boldsymbol{\theta}, p) \\ s.t. \quad & \begin{cases} \theta_n \in [0, 2\pi) \\ p \in [p_{\min}, p_{\max}] \end{cases} \end{aligned} \tag{19}$$

However, it is difficult to solve the formulated problem, as its objective function is nonconvex concerning either $\boldsymbol{\theta}$ or p. Additionally, the complex time-varying channel fading makes it impossible to obtain an optimal solution for long-term system utility using traditional optimization techniques.

4. Proposed DIST Scheme

4.1. Main Elements of DIST

In previous sections, we discussed the challenges in MIoT wireless communication. To address these issues, we propose a model-free RL approach. More specifically, we introduce a DDPG-based IRS-assisted secure transmission (DIST) scheme to efficiently search the policy space and improve the secure transmission performance while remaining independent of any specific system model or wiretap policy [31]. The DIST scheme is designed to be applicable to a wide range of MIoT systems, making it a valuable contribution to the field. By considering the IRS-assisted MIoT device's transmission system as the dynamic environment and the MIoT device itself as the learning agent, our method is able to adapt to various situations and effectively address the security concerns in MIoT wireless communication. In the following specifications, we outline the main components of the framework employed by the DIST scheme.

State space: At time slot k, the MIoT device observes the environment and formulates the state $\mathbf{s}^{(k)}$, which is modeled as follows:

$$\mathbf{s}^{(k)} = \left[\mathbf{h}^{(k)}, \widetilde{p_J}^{(k-1)}\right] \in \Lambda \tag{20}$$

where Λ is the state space. $\mathbf{h}^{(k)} = \left\{\mathbf{h}_{M,I}^{(k)}, h_{M,A}^{(k)}, h_{M,E}^{(k)}, \mathbf{h}_{I,A}^{(k)}, \mathbf{h}_{J,I}^{(k)}, \mathbf{h}_{I,E}^{(k)}, h_{J,A}^{(k)}\right\}$, $\mathbf{h}^{(k)}$ are the channel path loss at time slot k. $\widetilde{p_J}^{(k-1)}$ is the AP's received jamming power at time slot $k-1$.

Action space: We denote $\mathbf{A}$ as the action space. According to the observed state $\mathbf{s}^{(k)}$ at time slot k, the MIoT device designs the IRS phase shifts $\boldsymbol{\theta}^{(k)}$ and chooses the transmit power $p^{(k)}$. Then, the phase shifts control signal is sent to the IRS controller. Hence, the secure transmission policy $\mathbf{a}^{(k)} \in \mathbf{A}$ can be formulated by

$$\mathbf{a}^{(k)} = \left[\boldsymbol{\theta}^{(k)}, p^{(k)}\right] \tag{21}$$

Reward function: In the proposed DIST scheme, the reward function evaluates the secure transmission policy according to the current state. In the presented paper, we aim to achieve the maximum long-term utility of the system, as addressed in Equation (19). Thus, the reward function is denoted as follows:

$$r^{(k)}(\mathbf{s}, \mathbf{a}) = U^{(k)} \tag{22}$$

4.2. Main Process of DIST

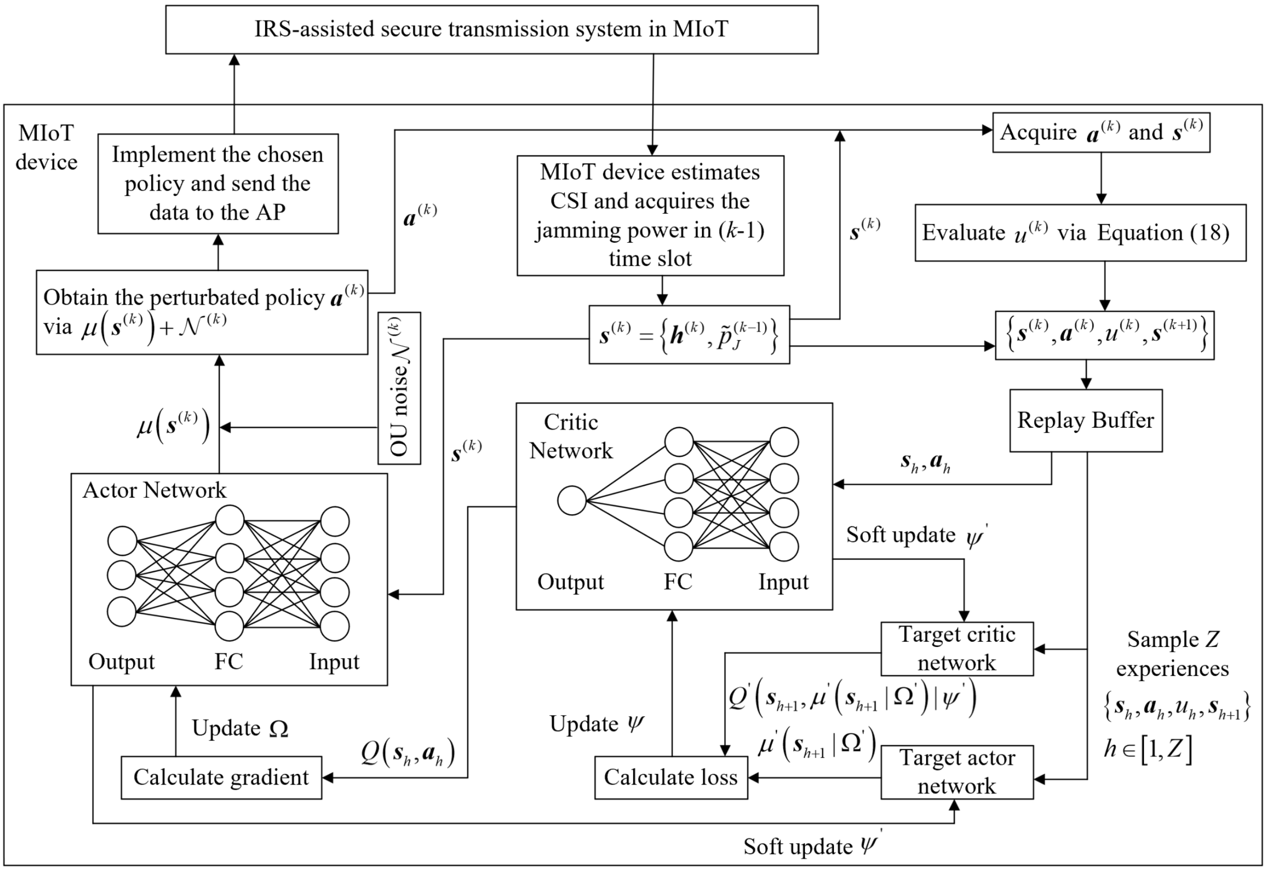
Our proposed DIST scheme contains a critic network and an actor network, denoted as $Q(\mathbf{s}, \mathbf{a}|\Psi)$ and $\mu(\mathbf{s}|\Omega)$ with parameters Ψ and Ω, respectively. The actor network is responsible for choosing the secure transmission policy, while the critic network assesses the policy selected by the actor network. Moreover, a target critic network $Q'(\mathbf{s}, \mathbf{a}|\Psi')$ and a target actor network $\mu'(\mathbf{s}|\Omega')$ are designed to promote convergence.

At the beginning of each episode, the MIoT device sets a random phase shift on each element. The MIoT device observes the environment and acquires the global CSI and the AP's received jamming power. Then, the MIoT device formulates the initial state $\mathbf{s}$ and inputs it into the actor network to generate corresponding transmission policy $\mathbf{a}$.

According to the observed state $\mathbf{s}^{(k)}$ at time slot k, the MIoT device selects the secure transmission policy $\mathbf{a}^{(k)} = \left[\boldsymbol{\theta}^{(k)}, p^{(k)}\right]$ through the actor network. The actor network links each state to a corresponding transmission policy using function $\mu\left(\mathbf{s}^{(k)}|\Omega^{(k)}\right)$. To enable the MIoT device to explore the environment, we model an Ornstein–Uhlenbeck (OU) process as the exploration noise $\mathcal{N}^{(k)}$, which is known as the OU-noise. The OU-noise is used to improve the exploration efficiency and find the optimal policy with better convergence. Thus, the secure transmission policy $\mathbf{a}^{(k)}$ is given by

$$\mathbf{a}^{(k)} = \mu\left(\mathbf{s}^{(k)}|\Omega^{(k)}\right) + \mathcal{N}^{(k)} \tag{23}$$

The MIoT device then sends the phase shifts control signal to the IRS controller and transmits the data to the AP with the transmit power p. Then, the MIoT device calculates the achievable rate at AP and eavesdropper via Equations (15) and (16). As a result, the MIoT obtains an immediate reward $u^{(k)}$, and the system state $\mathbf{s}^{(k)}$ is updated to a new state $\mathbf{s}^{(k+1)}$, which is denoted as $\mathbf{s}^{(k+1)} = \left[\mathbf{h}^{(k+1)}, \widetilde{p_J}^{(k)}\right]$. Next, the MIoT device stores the transition $\left(\mathbf{s}^{(k)}, \mathbf{a}^{(k)}, u^{(k)}, \mathbf{s}^{(k+1)}\right)$ in the replay buffer, where the oldest experience is systematically discarded in a rolling manner as the buffer reaches its maximum capacity. When the buffer size is larger than the batch size Z, the MIoT device randomly samples Z experiences from the replay buffer for exploring the optimal transmission policy in the dynamic MIoT environment. The detailed structure is shown in Figure 2.



IRS-assisted secure transmission system in MIoT
MIoT device
Implement the chosen policy and send the data to the AP
$a^{(k)}$
MIoT device estimates CSI and acquires the jamming power in (k-1) time slot
$s^{(k)}$
Acquire $a^{(k)}$ and $s^{(k)}$
Evaluate $u^{(k)}$ via Equation (18)
Obtain the perturbated policy $a^{(k)}$ via $\mu\left(s^{(k)}\right)+\mathcal{N}^{(k)}$
OU noise $\mathcal{N}^{(k)}$
$s^{(k)}=\left\{h^{(k)},\tilde{p}_J^{(k-1)}\right\}$
$\left\{s^{(k)},a^{(k)},u^{(k)},s^{(k+1)}\right\}$
$\mu\left(s^{(k)}\right)$
$s^{(k)}$
Critic Network
Replay Buffer
s_h,a_h
Output
FC
Input
Soft update ψ'
Actor Network
Output
FC
Input
Target critic network
Sample Z experiences
$\left\{s_h,a_h,u_h,s_{h+1}\right\}$
$h\in[1,Z]$
$Q'\left(s_{h+1},\mu'\left(s_{h+1}|\Omega'\right)|\psi'\right)$
$\mu'\left(s_{h+1}|\Omega'\right)$
Update Ω
$Q\left(s_h,a_h\right)$
Update ψ
Calculate gradient
Calculate loss
Target actor network
Soft update ψ'

Figure 2. The DDPG-based IRS-assisted secure transmission scheme in MIoT.

We formulate the minibatch $e_h = \{\mathbf{s}_h, \mathbf{a}_h, u_h, \mathbf{s}_{h+1}\}$, $h \in [1, Z]$ and utilize the Adam optimizer to update the critic network's weight Ψ [32], where the loss function is denoted as

$$\Psi = \arg\min_{\Psi} \frac{1}{H} \sum_{h=1}^{H} \left(u_h + \gamma Q'\left(\mathbf{s}_{h+1}, \mu'\left(\mathbf{s}_{h+1}|\Omega'\right)|\Psi'\right) - Q(\mathbf{s}_h, \mathbf{a}_h|\Psi)\right)^2 \tag{24}$$

where the discount factor $\gamma \in [0, 1]$.

The weights of the actor network are updated by leveraging the gradient of the Q-value [32], which can be expressed as follows:

$$\triangledown\Omega \approx \frac{1}{H}\sum_{h=1}^{H} \triangledown_{\mathbf{a}} Q(\mathbf{s}=\mathbf{s}_h, \mathbf{a}=\mu(\mathbf{s}_h)|\Psi) \times \triangledown_{\Omega}\mu(\mathbf{s}=\mathbf{s}_h|\Omega) \tag{25}$$

Lastly, the MIoT device uses the soft update strategy to ensure the target network changes slowly, thus guaranteeing stability. The soft update can be denoted as follows:

$$\begin{aligned} \Psi' &= \tau\Psi + (1-\tau)\Psi' \\ \Omega' &= \tau\Omega + (1-\tau)\Omega' \end{aligned} \tag{26}$$

where the τ represents the learning rate. The more detailed process is illustrated in Algorithm 1.

Algorithm 1 DDPG-based IRS-assisted secure transmission scheme (DIST)

```
Initialize: actor network, critic network, target critic network, target actor network, and replay buffer
for episode e = 1, 2, 3, ..., E do
    Initialize action exploration noise $\mathcal{N}$
    Obtain the channel state information $\mathbf{h}_{M,I}, h_{M,A}, h_{M,E}, \mathbf{h}_{I,A}, \mathbf{h}_{J,I}, \mathbf{h}_{I,E}, h_{J,A}$
    Randomly choose the IRS phase shifts $\boldsymbol{\theta}$
    Evaluate the AP's received jamming power $\widetilde{p}_J$
    Formulate the initial state $s$ according to Equation (20)
    for Time slot k = 1, 2, 3, ..., T do
        Select transmission policy $\mathbf{a}^{(k)}$ with state $\mathbf{s}^{(k)}$ and noise $\mathbf{N}^{(k)}$ based on the current policy.
        Execute transmission policy $\mathbf{a}^{(k)}$ and obtain the reward and utility $U^{(k)} = r^{(k)}(\mathbf{s}, \mathbf{a})$
        Obtain the AP's received jamming power $\widetilde{p_J}^{(k)}$
        Obtain the channel path loss $\mathbf{h}_{M,I}^{(k+1)}, h_{M,A}^{(k+1)}, h_{M,E}^{(k+1)}, \mathbf{h}_{I,A}^{(k+1)}, \mathbf{h}_{J,I}^{(k+1)}, \mathbf{h}_{I,E}^{(k+1)}, h_{J,A}^{(k+1)}$
        Formulate the state $\mathbf{s}^{(k+1)}$
        Store the transition $\left(\mathbf{s}^{(k)}, \mathbf{a}^{(k)}, u^{(k)}, \mathbf{s}^{(k+1)}\right)$ to the replay buffer
        if buffer length > $Z$ then
            Randomly sample a minibatch of $Z$ transitions $(\mathbf{s}_h, \mathbf{a}_h, u_h, \mathbf{s}_{h+1})$
            Update the critic network and actor network via Equations (24) and (25)
            Update the target actor network and target critic network via Equation (26)
        end if
    end for
end for
```

In Table 2, we present the advantages and potential limitations associated with the DIST scheme.

Table 2. Advantages and limitations of the proposed DIST scheme.

Advantages	**Limitations**
Adapt to time-varying and dynamic channel conditions	In practice, MIoT devices can hardly obtain the perfect CSI in a timely manner, causing performance degradation
Reduce transmit power consumption and promote energy efficiency	Computationally intensive due to the application of DRL
Enhance wireless communication in challenging mining environments	Only suitable for single-device scenarios

5. Simulation Setup and Results

In this section, we comprehensively illustrate the performance of our proposed DIST scheme under the presence of an active eavesdropper in mining scenarios. The system topology and coordinate of each unit are shown in Figure 3. The red line, blue line, and black line represent the eavesdropping channel, jamming channel, and legitimate transmission channel, respectively. In real-world mining operations, the positions of devices may vary. The changing positions may affect the value of the system performance. However, it will not impact the advantage trend of the proposed DIST scheme compared with the benchmarks. Simulations are implemented using Pytorch 1.13.1 with Python 3.9. The number of MIoT devices, jamming devices, and active eavesdroppers is set to 1, and they are all equipped with one single antenna. The MIoT device observes and estimates the CSI at each time slot. The IRS is composed of a total of $N = 12$ [6] reflecting elements, specifically $N_y = 2$ elements aligned parallel to the y-axis and $N_z = 6$ elements aligned parallel to the z-axis. The background noise power σ^2 is set to -80 dBm [27]. The MIoT device is specifically configured to operate within a transmit power range, with a minimum power $P_{\min}$ setting of 1 mW and a maximum power $P_{\max}$ setting of 9 mW. The jamming power is randomly generated in the range of 1 mW and 5 mW. The Rician factors $K_{M,I}$, $K_{I,A}$, $K_{I,E}$, and $K_{J,I}$ are assumed to be equal and set to 10 [33]. $\xi_{LoS} = 2.2$ and $\xi_{NLoS} = 3.8$ [34] are the path loss exponents of the LoS link and NLoS link, respectively.

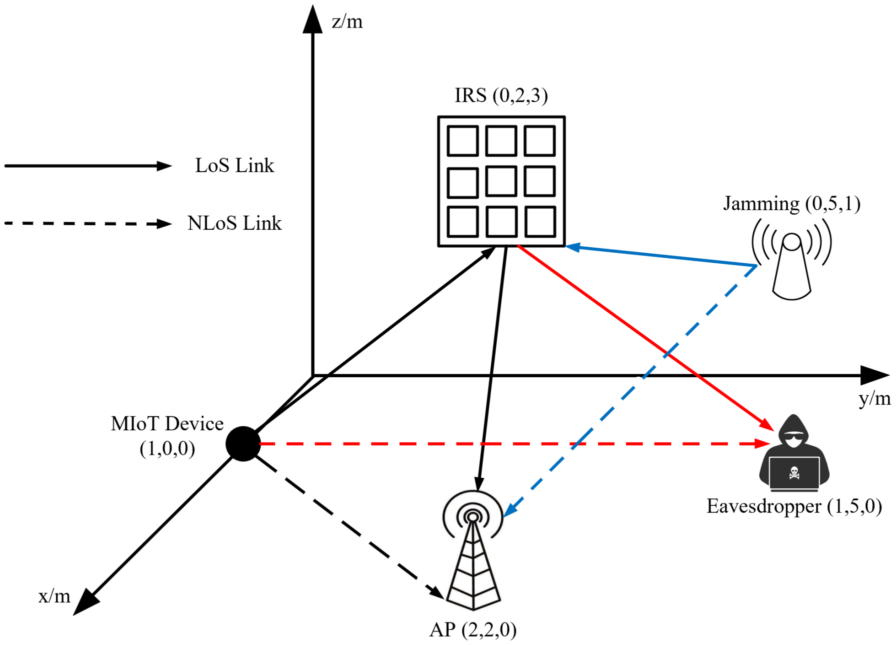


Figure 3. Simulation setting for an IRS-assisted secure transmission system in MIoT.

The learning model in the proposed DIST framework consists of a three-layer deep neural network (DNN). The hidden layer contains 32 neurons. The actor and critic learning rates are set to 5×10^{-7} and 5×10^{-4}, respectively. Moreover, the discount factor is determined to be $\gamma = 0.3$, whereas the soft update parameter is configured to be $\tau = 0.005$. We set the max buffer size to 10,240 and the batch size to 16. Moreover, we set the time slot number in each episode to $T = 256$ and the episode number to $E = 1024$. The parameters ω_1 and ω_2 in Equation (18) are set to 1 and 500, respectively, to balance the secrecy rate gain and power consumption loss. For the settings of the parameters mentioned above, we determined them through multiple experiments conducted by our research team.

Two benchmark schemes are considered, shown as follows:

IRS randomly configured (IRC): In this case, the reflection coefficients of each IRS element are generated randomly. We only use the DDPG algorithm to optimize the transmit power [35].

IRS-free (IF): We consider a classical communication system in MIoT without introducing the IRS. In this case, the MIoT device only chooses the transmit power based on the DDPG algorithm [36].

Figure 4 provides a comprehensive evaluation of the system utility across all schemes. Our proposed DIST scheme converges after 400 episodes and achieves the utility increment from −1.8 to 1.3. Specifically, in episode 600, our proposed DIST scheme achieves 5.5 and 4.25 times higher utility than the IF and IRC schemes, respectively. This proves the remarkable utility increase from applying the IRS in MIoT wireless communication. And it also emphasizes the significance of applying the RL mechanism to solve the IRS beamforming design problem in a secure transmission scene.

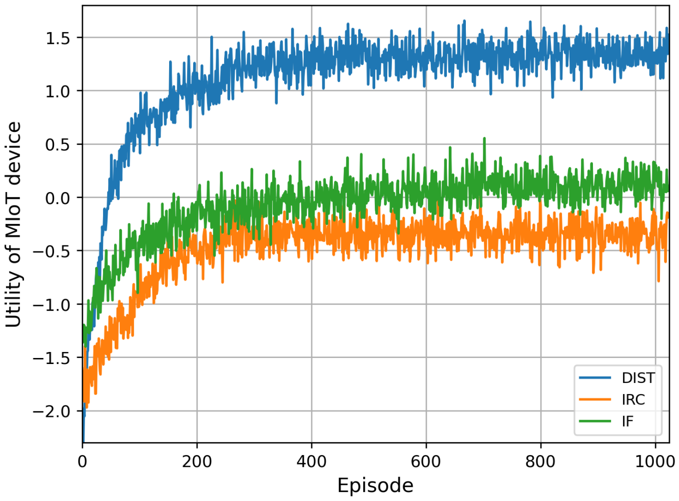
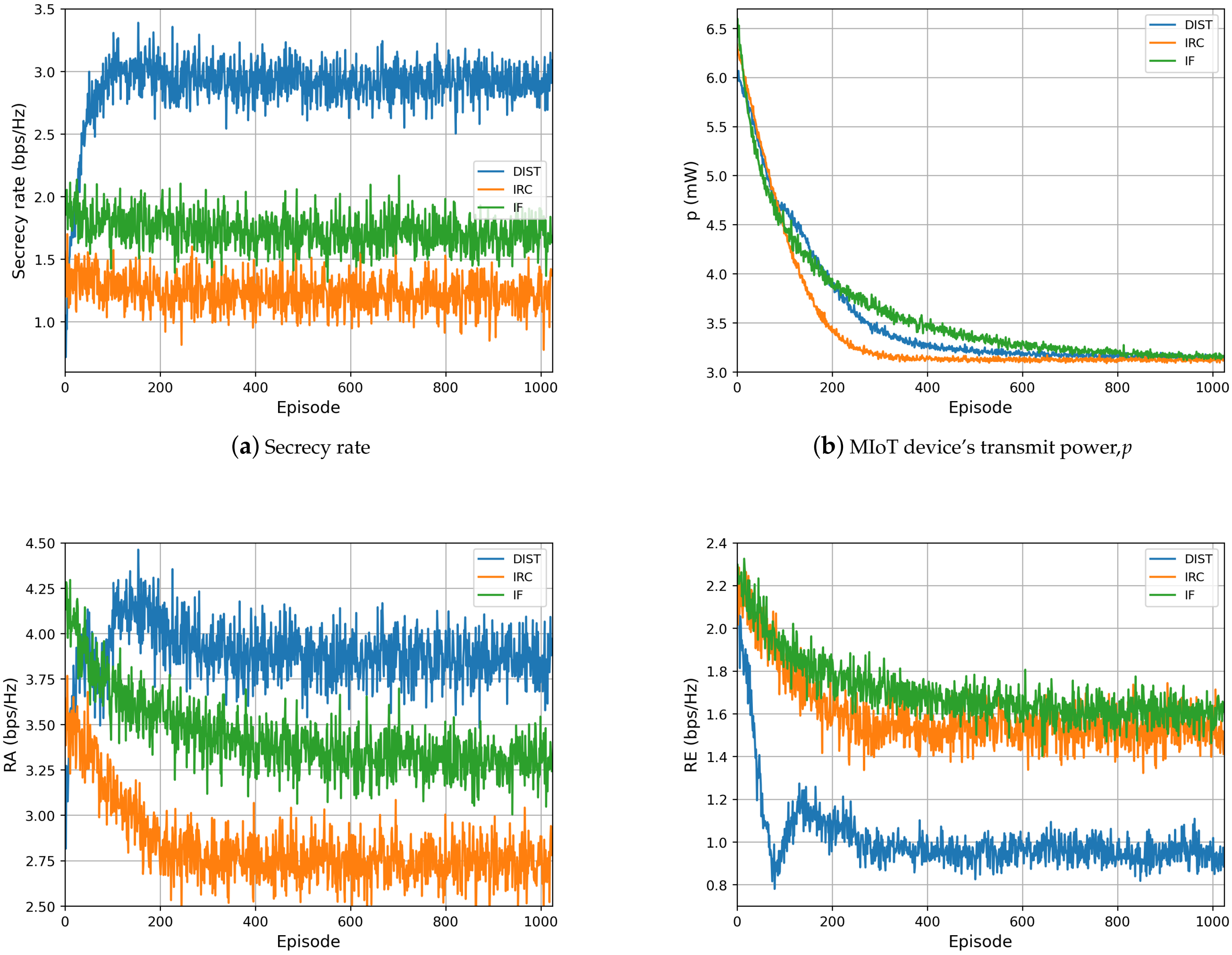


Figure 4. Utility of MIoT device of the DIST scheme compared with the IRC and IF schemes.

Figure 5 investigates the R_A, R_E, R_{sec}, and p of all schemes. For the secrecy rate shown in Figure 5a, our proposed DIST scheme outperforms the IF scheme and the IRC scheme by 70.6% and 141.7% in secrecy capacity. We then dig into the detailed performance. Particularly, in Figure 5d, in our proposed DIST scheme, we observe that the eavesdropping rate increases from 0.8 bps/Hz to 1.2 bps/Hz from episode 80 to 160, and then falls to 0.9 bps/Hz. The reason is that the MIoT device explores the environment and chooses the policy aiming to obtain the maximum utility. In this process, the eavesdropping rate may go up a bit, but in Figure 5a,c, the signal transmission rate at AP and the secrecy rate are still rising. Several factors contribute to the continuous rise in system utility in this interval. Among these are the factors mentioned above and the declining transmit power, as shown in Figure 5b.

Additionally, in Figure 5c, the signal transmission rate at AP degrades a little bit from 4.25 bps/Hz to 3.9 bps/Hz after 190 episodes. The reason is that the MIoT device's transmit power converges more slowly than the IRS phase shifts. After 200 episodes, the transmit power is still declining. According to Equations (12) and (15), lower transmit power will lead to a lower AP signal transmission rate when the reflecting coefficients converge to the optimal value.



(**a**) Secrecy rate

(**b**) MIoT device's transmit power,p

(**c**) RA

(**d**) RE

Figure 5. Performance of our proposed DIST scheme compared with the IRC and IF schemes: (**a**) Secrecy rate, R_{sec}. (**b**) MIoT device's transmit power, p. (**c**) Signal transmission rate at AP, R_A. (**d**) Eavesdropping rate, R_E.

As shown in Figure 6, we investigate the performance of our proposed DIST scheme and the IRC scheme by varying the number of IRS elements. The significant improvement of our proposed scheme demonstrated in Figure 6 results from more IRS elements bringing more reflected signals. When the IRS are well-adjusted, the reflected signal can be intelligently combined at AP to provide higher signal strength and deliberately manipulated at the eavesdropper to attenuate its received signal power, thereby diminishing its ability to intercept the transmission.

Moreover, the system utility of the IRC scheme decreases slightly as the number of IRS elements increases. This is because without IRS properly adjusted, the reflected signal with random phase will be added constructively or destructively, generating a stronger or weaker signal. Thus, the more IRS elements used, the larger the range of the SINR. According to Equations (15) and (16), the average R_A and R_E will decrease due to the different slope of function $\log_2(x)$ when the SINR range gets bigger, eventually resulting in performance degradation.

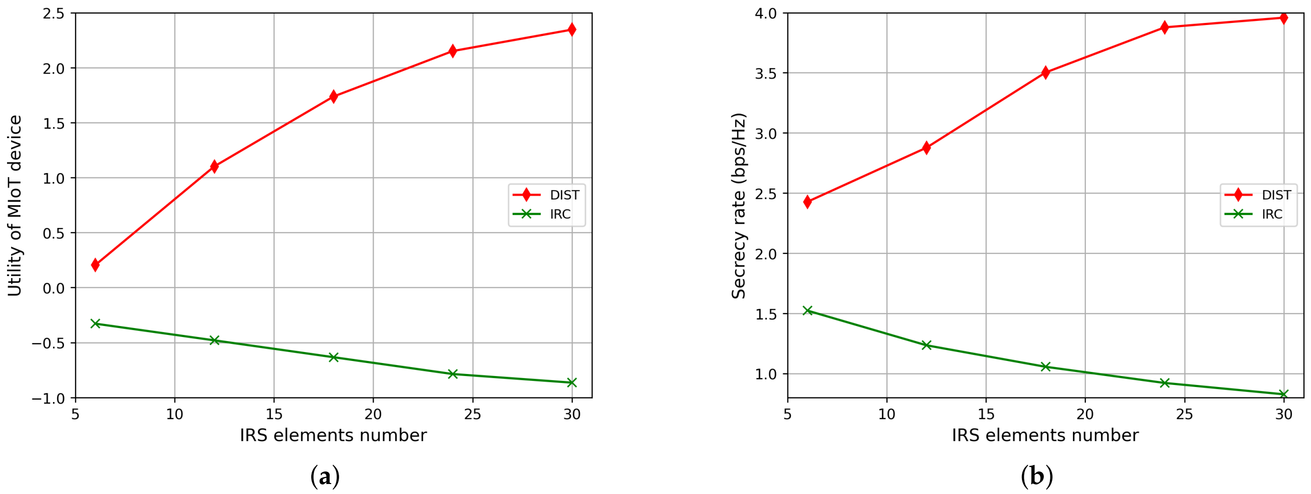


Figure 6. Average performance of the DIST scheme and the IRC scheme in MIoT, with the IRS elements number changing from 6 to 30 and averaged over 1024 episodes against active eavesdroppers: (**a**) Utility of MIoT device. (**b**) Secrecy rate.

6. Conclusions

In this paper, we investigated a secure transmission scheme against an active eavesdropper and formulated the optimization problem to maximize the utility of an MIoT device for a dynamic MIoT communication environment. We proposed a DDPG-based IRS-assisted secure transmission scheme in MIoT that enables the MIoT device to jointly design the optimal IRS phase shifts and transmit power. Simulation results demonstrate the effectiveness of our proposed scheme in enhancing secrecy rates and reducing power consumption. Comparing our DIST scheme with the IF scheme and the IRC scheme, our DIST scheme achieves a substantial performance improvement in utility of 5.5 times and 4.25 times, respectively. These results demonstrate the vital role of IRS in bolstering physical layer security and enhancing transmission performance in the MIoT wireless communication environment. This work can also be applied to handle secure transmission in other NLoS scenarios, such as large-scale underground supermarkets. Our future work will focus on developing a multiagent learning-based method to solve multidevice scenarios, including multiple receivers and eavesdroppers.

Author Contributions: Funding acquisition, M.M. and S.L.; conceptualization, M.M. and S.L.; investigation, J.X. and M.M.; methodology, J.S.; project administration, J.S.; software, J.X. and P.Z.; supervision, M.M.; validation, J.S.; writing—original draft, J.X. and P.Z.; writing—review and editing, S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (grant number 62101557 and 61771474), and China Postdoctoral Science Foundation (2022M713378), and the Fundamental Research Funds for the Central Universities (2042022kf0021).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhang, G.; Chen, C.H.; Cao, X.; Zhong, R.Y.; Duan, X.; Li, P. Industrial Internet of Things-enabled monitoring and maintenance mechanism for fully mechanized mining equipment. *Adv. Eng. Inform.* **2022**, *54*, 101782. [CrossRef]
2. Zhang, J. Exploration on coal mining-induced rockburst prediction using Internet of things and deep neural network. *J. Supercomput.* **2022**, *78*, 13988–14008. [CrossRef]
3. Ali, M.H.; Al-Azzawi, W.K.; Jaber, M.; Abd, S.K.; Alkhayyat, A.; Rasool, Z.I. Improving coal mine safety with internet of things (IoT) based Dynamic Sensor Information Control System. *Phys. Chem. Earth* **2022**, *128*, 103225. [CrossRef]
4. Farjow, W.; Raahemifar, K.; Fernando, X. Novel wireless channels characterization model for underground mines. *Appl. Math. Model.* **2015**, *39*, 5997–6007. [CrossRef]
5. Cao, Y.; Duan, L.; Jin, M.; Zhao, N. Cooperative double-IRS aided proactive eavesdropping. *IEEE Trans. Commun.* **2022**, *70*, 6228–6240. [CrossRef]
6. Xiao, L.; Hong, S.; Xu, S.; Yang, H.; Ji, X. IRS-Aided Energy-Efficient Secure WBAN Transmission Based on Deep Reinforcement Learning. *IEEE Trans. Commun.* **2022**, *70*, 4162–4174. [CrossRef]
7. Kumar, C.; Kashyap, S. On the Power Transfer Efficiency and Feasibility of Wireless Energy Transfer Using Double IRS. *IEEE Trans. Veh. Technol.* **2023**, *72*, 6165–6180. [CrossRef]
8. Zhang, S.; Zhang, H.; Di, B.; Tan, Y.; Di Renzo, M.; Han, Z.; Song, L. Intelligent omni-surfaces: Ubiquitous wireless transmission by reflective-refractive metasurfaces. *IEEE Trans. Wirel. Commun.* **2021**, *21*, 219–233. [CrossRef]
9. Cui, M.; Zhang, G.; Zhang, R. Secure wireless communication via intelligent reflecting surface. *IEEE Wirel. Commun. Lett.* **2019**, *8*, 1410–1414. [CrossRef]
10. Basar, E.; Di Renzo, M.; De Rosny, J.; Debbah, M.; Alouini, M.S.; Zhang, R. Wireless communications through reconfigurable intelligent surfaces. *IEEE Access* **2019**, *7*, 116753–116773. [CrossRef]
11. Min, M.; Xiao, L.; Chen, Y.; Cheng, P.; Wu, D.; Zhuang, W. Learning-Based Computation Offloading for IoT Devices with Energy Harvesting. *IEEE Trans. Veh. Technol.* **2019**, *68*, 1930–1941. [CrossRef]
12. Jang, G.; Kim, D.; Lee, I.H.; Jung, H. Cooperative Beamforming with Artificial Noise Injection for Physical-Layer Security. *IEEE Access* **2023**, *11*, 22553–22573. [CrossRef]
13. Alhoraibi, L.; Alghazzawi, D.; Alhebshi, R.; Rabie, O.B.J. Physical Layer Authentication in Wireless Networks-Based Machine Learning Approaches. *Sensors* **2023**, *23*, 1814. [CrossRef] [PubMed]
14. Dong, F.; Wang, W.; Li, X.; Liu, F.; Chen, S.; Hanzo, L. Joint Beamforming Design for Dual-Functional MIMO Radar and Communication Systems Guaranteeing Physical Layer Security. *IEEE Trans. Green Commun. Netw.* **2023**, *7*, 537–549. [CrossRef]
15. Meng, W.; Gu, Y.; Bao, J.; Gan, L.; Huang, T.; Kong, Z. Cooperative Jamming with AF Relay in Power Monitoring and Communication Systems for Mining. *Electronics* **2023**, *12*, 1057. [CrossRef]
16. Ren, H.; Liu, X.; Pan, C.; Peng, Z.; Wang, J. Performance Analysis for RIS-Aided Secure Massive MIMO Systems with Statistical CSI. *IEEE Wirel. Commun. Lett.* **2022**, *12*, 124–128. [CrossRef]
17. Feng, K.; Li, X.; Han, Y.; Jin, S.; Chen, Y. Physical layer security enhancement exploiting intelligent reflecting surface. *IEEE Commun. Lett.* **2020**, *25*, 734–738. [CrossRef]
18. Hong, S.; Pan, C.; Ren, H.; Wang, K.; Chai, K.K.; Nallanathan, A. Robust transmission design for intelligent reflecting surface-aided secure communication systems with imperfect cascaded CSI. *IEEE Trans. Wirel. Commun.* **2020**, *20*, 2487–2501. [CrossRef]
19. Yang, H.; Xiong, Z.; Zhao, J.; Niyato, D.; Xiao, L.; Wu, Q. Deep reinforcement learning-based intelligent reflecting surface for secure wireless communications. *IEEE Trans. Wirel. Commun.* **2020**, *20*, 375–388. [CrossRef]
20. Xiao, L.; Xie, C.; Min, M.; Zhuang, W. User-centric view of unmanned aerial vehicle transmission against smart attacks. *IEEE Trans. Veh. Technol.* **2017**, *64*, 3420–3430. [CrossRef]
21. Li, B.; Shi, T.; Zhao, W.; Wang, N. Reinforcement Learning-Based Intelligent Reflecting Surface Assisted Communications against Smart Attackers. *IEEE Trans. Commun.* **2022**, *70*, 4771–4779. [CrossRef]
22. Taha, A.; Zhang, Y.; Mismar, F.B.; Alkhateeb, A. Deep reinforcement learning for intelligent reflecting surfaces: Towards standalone operation. In Proceedings of the IEEE International Workshop on Signal Processing Advances in Wireless Communications, Atlanta, GA, USA, 26–29 May 2020; IEEE: New York, NY, USA, 2020; pp. 1–5.
23. Wang, W.; Zhang, W. Intelligent reflecting surface configurations for smart radio using deep reinforcement learning. *IEEE J. Sel. Areas Commun.* **2022**, *40*, 2335–2346. [CrossRef]
24. Fidêncio, A.X.; Glasmachers, T.; Naro, D. Application of Reinforcement Learning to a Mining System. In Proceedings of the 2021 IEEE 19th World Symposium on Applied Machine Intelligence and Informatics, Herl'any, Slovakia, 21–23 January 2021; IEEE: New York, NY, USA, 2021; pp. 111–118.
25. Guo, T.; Li, X.; Mei, M.; Yang, Z.; Shi, J.; Wong, K.K.; Zhang, Z. Joint Communication and Sensing Design in Coal Mine Safety Monitoring: 3D Phase Beamforming for RIS-Assisted Wireless Networks. *IEEE Internet Things J.* **2022**, *10*, 11306–11315. [CrossRef]
26. Hu, X.; Zhang, R.; Zhong, C. Semi-passive elements assisted channel estimation for intelligent reflecting surface-aided communications. *IEEE Trans. Wirel. Commun.* **2021**, *21*, 1132–1142. [CrossRef]
27. Feng, K.; Wang, Q.; Li, X.; Wen, C.K. Deep reinforcement learning based intelligent reflecting surface optimization for MISO communication systems. *IEEE Wirel. Commun. Lett.* **2020**, *9*, 745–749. [CrossRef]
28. Dong, R.; Wang, B.; Cao, K. Deep learning driven 3D robust beamforming for secure communication of UAV systems. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 1643–1647. [CrossRef]

29. Cao, Y.; Lv, T.; Ni, W. Intelligent reflecting surface aided multi-user mmWave communications for coverage enhancement. In Proceedings of the IEEE International Symposium on Personal Indoor and Mobile Radio Communications Workshops-PIMRC Workshops, London, UK, 31 August–3 September 2020; IEEE: New York, NY, USA, 2020; pp. 1–6.
30. Han, H.; Cao, Y.; Sheng, M.; Zhao, N.; Liu, J.; Niyato, D. IRS-Aided Secure NOMA Networks Against Internal and External Eavesdropping. *IEEE Trans. Commun.* **2022**, *70*, 7536–7548. [CrossRef]
31. Min, M.; Yang, S.; Zhang, H.; Ding, J.; Ding, G.; Pan, M.; Han, Z. Indoor Semantic Location Privacy Protection with Safe Reinforcement Learning. *IEEE Trans. Cogn. Commun. Netw.* 2023, *accepted*. [CrossRef]
32. Min, M.; Wang, W.; Xiao, L.; Xiao, Y.; Han, Z. Reinforcement learning-based sensitive semantic location privacy protection for VANETs. *China Commun.* **2021**, *18*, 244–260. [CrossRef]
33. Han, Y.; Tang, W.; Jin, S.; Wen, C.K.; Ma, X. Large intelligent surface-assisted wireless communication exploiting statistical CSI. *IEEE Trans. Veh. Technol.* **2019**, *68*, 8238–8242. [CrossRef]
34. Boutin, M.; Benzakour, A.; Despins, C.L.; Affes, S. Radio wave characterization and modeling in underground mine tunnels. *IEEE Trans. Antennas Propag.* **2008**, *56*, 540–549. [CrossRef]
35. Zheng, B.; You, C.; Zhang, R. Intelligent reflecting surface assisted multi-user OFDMA: Channel estimation and training design. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 8315–8329. [CrossRef]
36. Lv, Z.; Xiao, L.; Du, Y.; Niu, G.; Xing, C.; Xu, W. Multi-Agent Reinforcement Learning based UAV Swarm Communications Against Jamming. *IEEE Trans. Wirel. Commun.* 2023, *accepted*. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.