IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 21 January 2025, accepted 31 January 2025, date of publication 5 February 2025, date of current version 13 February 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3539088

RESEARCH ARTICLE

Terahertz-Based IRS-Assisted Secure Symbiotic Radio Communication: A DRL Approach

MUHAMMAD SHAHWAR[1], MANZOOR AHMED[2,3], TOUSEEF HUSSAIN[4], SAJED AHMAD[1], WALI ULLAH KHAN[5], (Member, IEEE), MUHAMMAD SHERAZ[6], AND TEONG CHEE CHUAH[6]

[1]College of Computer Science and Technology, Qingdao University, Qingdao 266071, China
[2]School of Computer and Information Science, Hubei Engineering University, Xiaogan 432000, China
[3]Institute for AI Industrial Technology Research, Hubei Engineering University, Xiaogan 432000, China
[4]School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
[5]Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, 4365 Esch-sur-Alzette, Luxembourg
[6]Faculty of Engineering, Multimedia University, Cyberjaya, Selangor 63100, Malaysia

Corresponding authors: Manzoor Ahmed (manzoor.achakzai@gmail.com) and Teong Chee Chuah (tcchuah@mmu.edu.my)

This work was supported in part by Multimedia University under the Research Fellow Grant MMUI/240021, and in part by Telekom Research and Development Sdn Bhd under Grant RDTC/241149.

ABSTRACT Developing wireless communication technologies is essential to satisfy the requirements of new applications and the increasing proliferation of interconnected devices. This research presents a resilient terahertz (THz)-based secure transmission framework for an active intelligent reflecting surface (IRS)-enabled symbiotic radio (SR) system in the presence of multiple eavesdroppers (Eves). The IRS facilitates secure transmission for the primary transmitter (PT) by intelligently adjusting the phase shifts of the signals from the PT, while simultaneously transmitting its own data to an Internet of Things (IoT) device. In light of the existence of numerous eves and unpredictable channels in real-world situations, we concurrently optimize the active beamforming of the PT and the phase shifts of the IRS to enhance the secrecy of IRS-assisted secure relay networks while adhering to quality-of-service standards and secure communication rates. To address this intricate non-convex stochastic optimization issue, we propose a secure beamforming technique named DDPG-SR, utilizing an effective deep reinforcement learning (DRL)-based deep deterministic policy gradient (DDPG) scheme to determine the optimal beamforming approach against Eves. This method seeks to establish an optimal beamforming strategy to counteract Eves under dynamic environmental circumstances. Comprehensive simulation experiments confirm the effectiveness of our proposed solution, showcasing enhanced performance relative to conventional IRS methods, IRS backscattering-based anti-evesdropping techniques, and other benchmark tactics for secrecy performance.

INDEX TERMS B5G, 6G, deep reinforcement learning, deep deterministic policy gradient, joint-beamforming, non-orthogonal multiple access, physical layer security, secure wireless communication, secrecy rate optimization.

I. INTRODUCTION

As the proliferation of Internet of Things (IoT) devices accelerates, the significance of symbiotic radio communications systems (SRCS) is poised to rise dramatically in the forthcoming sixth-generation (6G) era [1]. The anticipated exponential growth in connected devices as we near the 6G era demands innovative solutions to tackle challenges related to spectrum constraints, energy consumption, and connectivity. One of the key advantages of SRCS in this landscape is their ability to optimize spectrum efficiency. By enabling the coexistence of multiple communication systems within the same frequency bands, SRCS effectively minimizes interference while maximizing the utilization of available spectrum resources [2]. Moreover, SRCS supports energy saving, which is essential for the long-term functioning

The associate editor coordinating the review of this manuscript and approving it for publication was Barbara Masini.

of billions of IoT devices. These systems allow secondary devices to harvest energy from main transmissions, reducing the need for additional power supplies and extending battery life. This capacity aligns seamlessly with 6G's energy efficiency objectives, which call for energy-harvesting methods to enable the sustainable operation of many low-power IoT devices.

Concerning network coverage, SRCS improves connectivity by relaying signals through secondary devices, thereby increasing the reach of primary networks. This capability is critical for 6G's goal of enabling ubiquitous access, even in isolated or underdeveloped regions. By leveraging existing infrastructure and enabling resource sharing, SRCS lessens the need for costly deployments, thereby lowering operational expenses and accelerating the implementation of 6G networks [3]. SRCS also enhances the security of 6G networks by incorporating advanced techniques such as secure beamforming and cooperative jamming, which complicate unauthorized interception of communications. This improved security is crucial for 6G networks, as they will handle sensitive information across critical applications, including healthcare, the metaverse, and autonomous vehicles [4]. Furthermore, SRCS exhibits remarkable adaptability, capable of evolving alongside technological advancements to meet the diverse and evolving requirements of 6G applications. Their dynamic resource allocation features enable them to adjust to fluctuating network conditions and user requirements, facilitating massive machine-type (mMTC) and ultra-reliable low-latency communications (URLLC).

Addressing spectrum scarcity and energy efficiency challenges, symbiotic radio (SR) has emerged as an innovative communication paradigm that enables primary and secondary systems to coexist in a mutually beneficial manner [5]. With increasing demands on wireless networks, it offers a sustainable solution for optimizing spectrum utilization. Ambient backscatter communications (AmBC) and cognitive radio (CR) inspired the SR concept. The primary transmitter (PT) of an SR network also acts as a source of radio frequency (RF) signals for a backscatter device, also known as the secondary transmitter (ST). Useful information from the primary transmitter PT is contained in the signal that the ST backscatters to the primary receiver (PR). Overcoming the shortcomings of traditional CR, the ST serves as a multi-path reflector of the PT signal rather than generating interference.

Moreover, enhancing wireless communication systems has become increasingly vital, leading to the development of intelligent reflecting surfaces (IRS) [6]. This technology utilizes reconfigurable surfaces to manipulate signal paths, improving coverage, energy efficiency, and overall network capacity. In the context of SRCS, IRSs play a crucial role in tackling the issues that emanate from IoT device proliferation and the upcoming 6G era [7]. IRS technology leverages its reconfigurability to improve communication quality and conserve resources in congested networks. Specifically, IRSs can control the wireless signals received from the sources by changing the phase of reflective components, thereby achieving optimized propagation channels. This capability is instrumental in improving the spectrum efficiency and security, which are increasingly critical due to the growing density of connected devices in 6G networks. By integrating IRSs into SRCS, networks can achieve enhanced signal strength, improved security, and expanded coverage without requiring additional powerful transmitters. IRS can amplify and redirect the signals to the intended receivers, thus enhancing the coverage of existing infrastructure. This not only addresses challenges of poor signal penetration but also enables the rapidly growing IoT devices to sustain sufficient clear communication channels.

In addition, THz frequencies are poised to revolutionize 6G wireless communication by enabling ultra-fast data transfer, ultra-low latency, and massive bandwidth for emerging applications [8]. Deep reinforcement learning (DRL) is increasingly being leveraged in THz-based IRS-assisted secure relay networks to improve security in next-generation wireless communication systems [9]. By integrating DRL with IRS operating in the THz frequency range, these networks can dynamically adapt to changing network conditions, optimizing resource allocation, beamforming, and transmission strategies in real-time [10]. This adaptability enhances the ability to mitigate security threats such as eavesdropping and interference, particularly in the high-frequency THz bands where signal propagation is more susceptible to blockage and attenuation. DRL empowers the network to intelligently learn and update security policies, ensuring robust communication and data protection in a highly dynamic and complex environment.

A. SECURE IRS-ASSISTED SR COMMUNICATION

IRS-assisted SR communication represents a significant advancement in wireless communication technologies, particularly in the context of the upcoming 6G networks [11]. This innovative approach integrates IRS with SR systems to enhance spectrum efficiency and energy utilization while facilitating robust communication between primary and secondary users [12]. The integration of IRS into SR systems enhances this capability by providing a programmable environment that can adaptively manage signal propagation. IRS can reflect and manipulate incoming signals, thereby improving the quality of communication for both primary and secondary users [13].

The authors in [14] propose an innovative framework that integrates IRS to enhance hybrid SR networks based on simultaneous wireless information and power transfer (SWIPT). The study develops a practical end-to-end system model and explores a multi-objective optimization approach to maximize system efficiency and performance. This comprehensive research addresses key challenges in balancing spectrum and energy efficiency, providing valuable insights for advancing cognitive communication technologies. A novel IRS-enabled SR system, in which IRS enhances communication between the PT and the PR while simultaneously transmitting its information to the PR through phase shift variations is considered in [15]. The author's primary objective was to collaboratively enhance the active transmit beamforming at the PT and the passive reflecting beamforming at the IRS to reduce the PT's transmit power, while adhering to the signal-to-noise ratio limitations of both primary and IRS transmissions. They formulated an optimization problem linking the IRS phase shifts with both the channel state information (CSI) and transmitted message, analyzing optimal CSI configurations to derive insights into cooperative beamforming strategies between the PT and the IRS.

The research work in [16] explores the use of IRS to enhance secure SR multicast communications. The proposed system involves a PT transmitting confidential information to multiple PUs while the IRS aids this transmission and concurrently delivers its own information to an SU. To ensure signal security and efficiency in the presence of eavesdroppers (Eves), the study optimizes active beamforming at the PT and reflection coefficients at the IRS through an alternating optimization approach. Techniques like successive convex approximation (SCA), semidefinite relaxation (SDR), and sequential rank-one constraint relaxation (SROCR) were employed to optimize the security of the communication system.

The authors in [17] investigate how IRS can enhance secure broadcasting in millimeter wave (mmWave) SR networks. This study aims to maximize the minimum secrecy rate among users by co-designing the hybrid precoder at the base station (BS) and the passive beamformer at the IRS, while ensuring quality-of-service (QoS) for both the users and IoT devices. To tackle this secrecy optimization problem, the authors employed a SCA-based iterative algorithm. Similar to [16], the authors in [18] enhanced the security of a multicast communication system. This research explores a robust transmission scheme designed for an active IRS-enabled SR system, which is particularly critical in scenarios involving multiple Eves. They introduced a framework that optimizes both transmit and reflection beamforming to minimize overall system power consumption while ensuring secure multicasting of primary signals and secondary information delivery. Given the challenge of imperfect CSI with Eves, the study applies advanced mathematical techniques, including the S-Procedure and Bernstein-Type Inequality, to address worst-case and statistical CSI error models. This contribution is crucial as it advances the understanding of secure transmission under real-world conditions, where perfect CSI cannot be assumed.

In [19], the authors focused on tackling high-dimensional resource optimizations for low-earth orbit (LEO) satellites, as it is difficult to utilize machine learning (ML)-based and non-ML-based algorithms to tackle high-dimensional resource optimizations within a short duration. To overcome this difficulty, a viable method is to form alliances between numerous LEO satellites and GTs to collaboratively address resource optimization. Motivated by the mutualism inherent in symbiotic radios, this article proposes a novel collaborative learning scheme for implementing intelligent non-terrestrial networks, in which each GT with powerful computing capability operates a ground-tier learning agent to assist LEO satellites in tackling GT resource allocation tasks. Each LEO satellite also functions as a space-tier learning agent, and the LEO satellite's learning model can be transferred to its successor satellite to serve as a starting point for further updating the model.

DRL has recently gained significant attention as an effective approach for optimizing complex communication systems. In the context of wireless networks, the integration of DRL techniques with IRS has shown great potential for enhancing both resource allocation and security [20]. The authors in [21] present an advancement in physical layer security (PLS) by integrating IRS with a hybrid system using mmWave and visible light communication (VLC). The study utilizes a DRL approach, specifically the deep deterministic policy gradient (DDPG) technique, to optimize beamforming and IRS configurations while addressing power constraints and user mobility. However, this study does not incorporate SR networks. The study [22] explores the use of IRS in SR networks, aiming to improve the secrecy energy efficiency (SEE) in a network with potential Eves. The study proposes a robust resource allocation and beamforming design for the SR network, considering uncertainties in CSI and imperfect successive interference cancellation (SIC). The paper presents a DRL approach using proximal policy optimization (PPO) to jointly optimize the active beamforming of the PT and the passive beamforming of the IRS. This approach addresses the complex, non-convex optimization problem and provides a solution that improves security while ensuring the quality of service for both the PT and ST.

B. SECURE THz-BASED IRS COMMUNICATION

The integration of IRS into THz communication systems presents a promising avenue for enhancing security in wireless communications. As the demand for high-speed data transmission increases, particularly in the context of 6G networks, ensuring the confidentiality and integrity of transmitted information becomes paramount [23]. IRS technology, characterized by its ability to manipulate electromagnetic waves, can significantly bolster the security of THz communication systems against eavesdropping and interference.

The authors of [24] address the challenge of securing confidential communication in THz systems with multiple IRS, focusing on potential eavesdropping on both BS-IRS and IRS-user links. For BS-IRS eavesdropping, a zero-forcing beamforming strategy and phase shift solutions for IRSs are applied, while for IRS-user link eavesdropping, an iterative approach manages non-convex optimization within a specified leakage threshold. In contrast to this work, our approach uniquely integrates DRL with THz-based IRS-assisted secure symbiotic radio communication, focusing on dynamic security optimization in highly adaptive THz environments. The study [25] presents a deep learning-based approach for securing THz wireless communication in 6G networks using IRS. The authors propose a framework that integrates IRS with neural network architectures to optimize secure beamforming and mitigate eavesdropping. However, this study does not consider SR-based communication, which is a key component of the proposed model in our work.

The paper [26] examines the design of a beamformer and IRS to improve the security of downlink THz communication in a multi-input single-output (MISO) wiretap channel. The authors propose a low-complexity successive design approach, where they first optimize the IRS phase shifters and then derive the optimal beamformer in closed form. They also present a joint design approach, where the IRS phase shifters and beamformer are alternately optimized until convergence, which achieves high security performance. In contrast, our proposed DDPG-SR approach presented in this work employs DRL to jointly optimize the IRS and beamforming strategies in real-time while also incorporating SR communication, thereby offering a more flexible and adaptive solution for enhancing security.

The research work [27] addresses the problem of securing IRS-assisted mmWave and THz systems from Eves. The authors solve this problem by jointly optimizing transmit beamforming at the base station (BS) and discrete phase shifts at the IRS to maximize secrecy rate. Unlike our proposed DRL-based approach, they employed conventional optimization schemes, i.e., closed-form solutions, semidefinite programming (SDP), and block coordinate descent (BCD) methods. The authors also did not consider the SR environment for security optimization. To improve the performance of 6G-THz communications in indoor environments, the authors of [28] proposed a secure IRS-aided hybrid beamforming design. The authors optimize the placement of multiple IRSs to maximize the coverage area. Under the optimized IRS placement, they proposed a 3D hybrid beamforming scheme at the BS and phase adjustment at the IRSs, which are jointly performed via a low-complexity codebook-based 3D beam scanning approach. However, their approach does not incorporate DRL for optimization, nor does it consider a SR communication system, both of which are key elements in the proposed model of this work.

While existing work has explored the integration of IRS and DRL techniques to enhance various aspects of wireless communication systems, such as resource allocation, security, and performance optimization, IRS-assisted secure THz-based SR communication remains underexplored. The current literature lacks a comprehensive study on leveraging DRL approaches to jointly optimize the active and passive beamforming in an IRS-aided THz-based SR network, particularly in the presence of multiple users and Eves. This research gap presents an opportunity to develop novel solutions that can improve the security and efficiency of THz-based SR communication by leveraging the combined capabilities of IRS and DRL techniques. Our proposed research aims to address this gap and contribute to the advancement of secure THz-based SR communication systems.

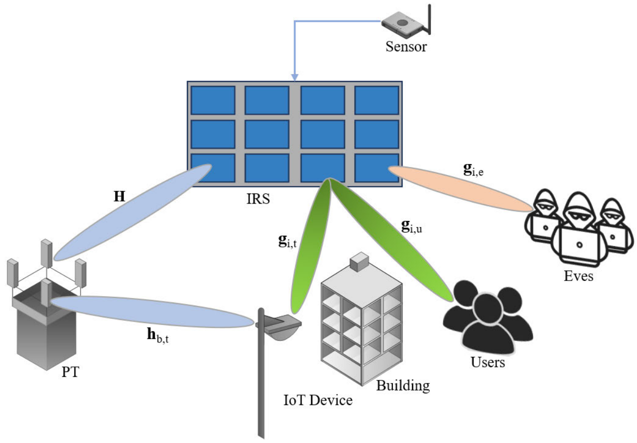

FIGURE 1. An IRS-enabled secure SR communication system.

C. MAIN CONTRIBUTIONS

- We introduce a novel approach that utilizes THz frequencies within an IRS-assisted SR communication framework to enhance security. By integrating THz-enabled IRS, our system achieves dynamic control over signal propagation, significantly strengthening PLS in backscatter and wireless communications.
- We introduce a DDPG-SR technique based on the DDPG framework, tailored to optimize the phase shift coefficient matrix within THz-based IRS-assisted SR networks. This optimization plays a vital role in enhancing PLS, particularly in the presence of multiple users and eavesdroppers.
- We have demonstrated that our proposed DDPG-SR technique outperforms other optimization methods, such as alternating optimization (AO), iterative optimization (IO), and deep Q-network (DQN), in terms of secrecy performance specifically within THz-enabled IRS-assisted SR communication systems. The simulation results underscore the effectiveness of the DDPG-SR approach, highlighting its superior ability to enhance security in these advanced communication environments.

The remaining paper is structured as follows: Section II details the system model, including the signal transmission and channel models, which establish the foundation for our problem formulation. Section III presents the problem formulation and our proposed DDPG-SR approach to optimize beamforming and phase-shift coefficients. Following this, Section IV describes the simulation setup, parameters, and benchmarks used for evaluation. Finally, in Section V, we provide our conclusions, summarizing our findings and suggesting future research directions.

II. SYSTEM MODEL

A. SYSTEM MODEL

In our system model as shown in Figure 1, we consider the IRS enhanced THz-based secure SR communication system in which we have a PT integrated with N_B antennas and N_{RF} radio frequency (RF) chains, IRS integrated with N_I reflecting elements, one single antenna IoT device, U single antenna PRs (Users), and E single antenna eves. In order to organize the calculation and evaluation, we define the set of antennas at the PT, the set of reflecting elements at IRS, the set of PRs, and the set of eves as $N_B = \{1, \ldots, B\}$, $N_I = \{1, \ldots, I\}$, $U = \{1, \ldots, U\}$, and $E = \{1, \ldots, E\}$, respectively. By deploying IRS in the SR communication system, we assume that THz-based primary transmission is blocked from the PT to PRs and Eves due to obstacles. However, the PT utilizes the IRS to send confidential information to the PRs through a THz-based reflected link. At the same time, in the secondary transmission, IRS transmits its own environmental data to the IoT device by using the primary signal. In our considered system, it is assumed that the EVs are situated in close proximity to the PRs to capture the confidential transmitted information. Additionally, we assume there is no link between the Eves and the IoT device because the Eves are near the PRs and far from the IoT device. The Eves' primary mission is to capture the information sent to the PRs, making them unable to intercept the information received by the IoT device.

B. SIGNAL TRANSMISSION MODEL

The PT transmits signal streams to the users while allowing the IRS to send its own data to the IoT device. The IRS utilizes binary phase shift keying (BPSK) modulation to encode its information symbol $c \in \{-1, 1\}$ onto the incoming signals from the PT. We deploy a commensal SR system in which a ST (IRS) leverages the existing transmission infrastructure and signals of a PT to communicate its own data. In this setup, the primary and secondary systems coexist, with the primary system being largely unaffected by the presence of the secondary system. Typically, the primary system operates at a much higher symbol rate compared to the symbol rate of the secondary system [29]. The PT can communicate a large number of primary symbols L ($L \gg 1$), within one secondary symbol period. During the IRS symbol transmission in the time period P_c, we assume that the knowledge of channel remains unchanged. During this process, the PT sends L primary symbols, i.e., $P_c = L.P_s$ where P_s denotes the time period of the primary symbol. Figure 2 illustrates the frame patterns of the primary and secondary transmission symbols. We consider the processes of the PT and ST to be perfectly aligned. Because perfect alignment is an idealized assumption, its benefits lie in enabling theoretical tractability, providing a foundation for system design, and serving as a benchmark for assessing the performance impact of practical misalignment. In future work, imperfections in alignment could be modeled to explore their effects on system performance and devise robust countermeasures.

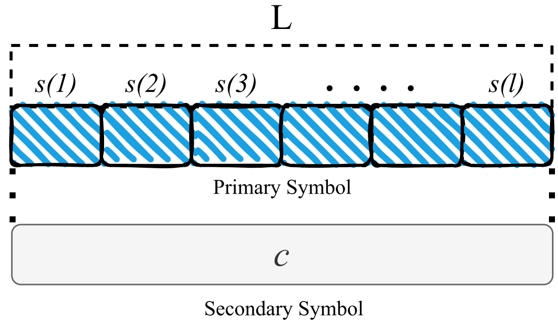

FIGURE 2. Transmission symbols for the SR system.

Let $s(l)$ denote the primary symbol transmitted by the PT, where $s(l) \sim \mathcal{CN}(0, 1)$, which follows a complex normal distribution. Here, $l \in \{1, \ldots, L\}$ which encompasses a set of all primary symbols. To save on power and costs, the PT uses a partially connected hybrid beamforming structure [30], [31]. The PT has N_{RF} RF chains connected to N_B antennas, with each RF chain linked to $\frac{N_B}{N_{RF}}$ antennas through analog phase shifters. Let $\boldsymbol{w} = \mathbf{F}_a\mathbf{f}_d$ represent a beamforming vector used by the PT to transmit its data, where $\mathbf{F}_a \in \mathcal{C}^{N_B \times N_{RF}}$ and $\mathbf{f}_d \in \mathcal{C}^{N_{RF} \times 1}$ are analog and digital beamformers, respectively. Thus, the final transmitted signal by the PT is $\boldsymbol{w}s(l)$.

A perfect CSI is assumed by the PT with the existing channel estimation methods, which allows for optimal resource allocation, precise beamforming, and maximized secrecy performance by effectively suppressing interference and enhancing the legitimate user's channel. This assumption simplifies the analysis and provides a theoretical benchmark for system performance. However, in practical scenarios, channel estimation errors and imperfect CSI can arise due to noise, feedback delays, or hardware limitations. As shown in Figure 1, the channel coefficient matrices/vectors from the PT to the IRS, from the PT to the IoT device, from the IRS to the U-th PR, from the IRS to the E-th eve, and from the IRS to the IoT device are denoted by $\boldsymbol{H} \in \mathcal{C}^{N_I \times N_B}$, $\boldsymbol{h}_{b,t} \in \mathcal{C}^{N_T \times N_B}$, $\boldsymbol{g}_{i,u} \in \mathcal{C}^{1 \times N_I}$, $\boldsymbol{g}_{i,e} \in \mathcal{C}^{1 \times N_I}$, and $\boldsymbol{g}_{i,t} \in \mathcal{C}^{N_T \times N_I}$, respectively, where $b \in N_B$, $u \in U$, $t \in N_T$, $i \in N_I$, and $e \in E$. By using a non-orthogonal multiple access (NOMA) approach with each beam the PT sends the superimposed signal to the user via IRS. By employing NOMA, it is possible for each beam to accommodate numerous users. Additionally, the interference between users is minimized by implementing the SIC approach for users with weaker channel gains [32].

The l-th signal received at the U-th PR within a specific secondary symbol can be formulated as

$$y_u(l) = G_t G_r \boldsymbol{g}_{i,u}(c\Phi)\boldsymbol{H}\boldsymbol{w}s(l) + n_u(l), \tag{1}$$

where $n_u(l) \sim \mathcal{CN}(0, \sigma_U^2)$ is the additive white Gaussian noise (AWGN) at the U-th PR. Similarly, the l-th signal received at the e-th Eve can be formulated as

$$y_e(l) = G_t G_r \boldsymbol{g}_{i,e}(c\Phi)\boldsymbol{H}\boldsymbol{w}s(l) + n_e(l), \tag{2}$$

where $n_e(l) \sim \mathcal{CN}(0, \sigma_T^2)$ is the AWGN at the E-th eve. The received signals at the IoT device can be expressed as

$$y_t(l) = [\boldsymbol{h}_{b,t} + \boldsymbol{G}_{i,t}(c\Phi)\boldsymbol{H}]G_tG_r\boldsymbol{w}s(l) + n_t(l), \quad (3)$$

where $n_t(l) \sim \mathcal{CN}(0, \sigma_T^2)$ is the AWGN at the IoT device, G_t and G_r denote the gains of the transmitting and receiving antennas, respectively. $\Phi = diag(\alpha_1 e^{j\theta 1}, \alpha_2 e^{j\theta 2}, \ldots, \alpha_I e^{j\theta_I})$ denotes the IRS reflection coefficient matrix, whereas $\theta_I \in [0, 2\pi]$ and $\alpha_I \in [0, 1]$ represent the phase shift coefficient and the amplitude reflection coefficient on the combined transmitted signal, respectively.

C. CHANNEL MODEL

In this setup, all the channels operate within a shared narrow frequency band. The reason for focusing on narrowband systems stems from the increasing demand for low-power applications like smart metering, industrial automation, remote manufacturing, and smart factories. These applications prioritize simplicity and energy efficiency, making narrowband communication an ideal choice [33]. Additionally, THz signals are quickly absorbed by water when traveling through the air, causing the scattered components to be much weaker than the line-of-sight (LoS) component [34]. Therefore, we can disregard the scattered components in the channels, and $\boldsymbol{H}$ represents the THz MIMO channel between the PT and the IRS, can be represented as

$$\boldsymbol{H} = q(f, d)\hat{\boldsymbol{H}}, \quad (4)$$

where $q(f, d) = \frac{c}{4\pi f d} e^{-\frac{1}{2}\tau(f)d}$ denotes the path loss, which includes both the free-space path loss and molecular absorption loss in THz communications, where c represents the speed of light, f is the operational frequency, $\tau(f)$ indicates the absorption value of water molecules in the air, and d represents the distance between the PT and the IRS. $\hat{\boldsymbol{H}} = a(\phi)b(\theta)$, where $a(\phi)$ and $b(\theta)$ are the transmitter's and the receiver's array steering vectors, respectively, which can be formulated as

$$a(\phi) = \frac{1}{\sqrt{N_{\mathrm{B}}}}\left[1, e^{j\pi\phi}, e^{j2\pi\phi}, \cdots, e^{j(N_{\mathrm{B}}-1)\pi\phi}\right], \quad (5)$$

$$b(\theta) = \frac{1}{\sqrt{N_{\mathrm{I}}}}\left[1, e^{j\pi\theta}, e^{j2\pi\theta}, \cdots, e^{j(N_{\mathrm{I}}-1)\pi\theta}\right]. \quad (6)$$

Here, $\phi = 2d_0 f \sin(\varphi_t)/c$ and $\theta = 2d_0 f \sin(\varphi_r)/c$, where d_0 is the array spacing, and $\varphi_i \in [-\frac{\pi}{2}, \frac{\pi}{2}], i = [t, r]$, is the angle of arrival (AOA) and the angle of departure (AOD). Similarly, $\boldsymbol{h}_{b,t}$ and $\boldsymbol{g}_{i,t}$ can be formulated as

$$\boldsymbol{h}_{b,t} = q(f, d_{bt})\hat{\boldsymbol{h}}_{b,t}, \quad (7)$$

$$\boldsymbol{g}_{i,t} = q(f, d_{it})\hat{\boldsymbol{g}}_{i,t}, \quad (8)$$

where d_{bt} and d_{it} are the distances between PT to IoT and IRS to IoT device respectively. Angles for $\hat{\boldsymbol{h}}_{b,t}$ and $\hat{\boldsymbol{g}}_{i,t}$ are formulated by using (5) and (6). Similarly $\boldsymbol{g}_{i,u}$ can be formulated as

$$\boldsymbol{g}_{i,u} = q(f, d_{iu})\hat{\boldsymbol{g}}_{i,u}, \quad (9)$$

where $\hat{\boldsymbol{g}}_{i,u} = \frac{1}{\sqrt{N_{\mathrm{I}}}}\left[1, e^{j\pi\phi}, e^{j2\pi\phi}, \cdots, e^{j(N_{\mathrm{I}}-1)\pi\phi}\right]$, and $q(f, d_{iu}) = \frac{c}{4\pi f d_{iu}} e^{-\frac{1}{2}\tau(f)d_{iu}}$, where d_{iu} denotes the distance between the IRS and U-th primary user.

Similarly, $\boldsymbol{g}_{i,e}$ can be formulated as

$$\boldsymbol{g}_{i,e} = q(f, d_{ie})\hat{\boldsymbol{g}}_{i,e}, \quad (10)$$

where $\hat{\boldsymbol{g}}_{i,e} = \frac{1}{\sqrt{N_{\mathrm{I}}}}\left[1, e^{j\pi\phi}, e^{j2\pi\phi}, \cdots, e^{j(N_{\mathrm{I}}-1)\pi\phi}\right]$, and $q(f, d_{ie}) = \frac{c}{4\pi f d_{ie}} e^{-\frac{1}{2}\tau(f)d_{ie}}$, where d_{ie} is the distance between the IRS and E-th Evesdropper.

D. SECRECY RATE

Considering that the PT transmits $s(l)$ to the U-th PR through the formulated channel. With a constant c, the instantaneous signal-to-noise ratio (SNR) for decoding $s(l)$ at the U-th PR can be represented as

$$\gamma_U = \frac{\left|\alpha G_t G_r \boldsymbol{g}_{i,u}(c\Phi)\boldsymbol{H}\boldsymbol{w}\right|^2}{\sigma_U^2 + \left|\left|\boldsymbol{g}_{i,u} c\Phi\right|\right|^2}, \quad (11)$$

where α is the path correction variable. The main focus of the Eves is to steal the primary information transmitted by the PT to U-th PR through the IRS. Therefore, the SNR for the E-th Eve to decode $s(l)$ can be expressed as

$$\gamma_E = \frac{\left|\alpha G_t G_r \boldsymbol{g}_{i,e}(c\Phi)\boldsymbol{H}\boldsymbol{w}\right|^2}{\sigma_E^2 + \left|\left|\boldsymbol{g}_{i,e} c\Phi\right|\right|^2}. \quad (12)$$

As indicated in [35], the maximum likelihood detection technique enables the IoT device to jointly decode $s(l)$ and c by utilizing its strong computational capabilities for reliable communications. Upon decoding $s(l)$ at the IoT device, the SIC approach is employed to eliminate the term $\boldsymbol{H}_{b,t}$ of 3. Thus, after successfully receiving the signal at IoT, the SNR to decode the c can be expressed as

$$\gamma_T = \frac{L\left|\alpha G_t G_r \boldsymbol{g}_{i,t}\Phi\boldsymbol{H}\boldsymbol{w}\right|^2}{\sigma_T^2 + \left|\left|\boldsymbol{g}_{i,t}\Phi\right|\right|^2}. \quad (13)$$

Therefore, it is possible to calculate the secrecy rate that the U-th primary user achieved as

$$R_{sec} = \log_2(1 + \gamma_U) - \log_2(1 + \gamma_E). \quad (14)$$

E. POWER CONSUMPTION MODEL

The overall power consumption for the system with the active IRS, represented as P_T^{IRS}, is expressed as

$$P_T^{IRS} = P_{tr} + P_{ac}^{IRS}, \quad (15)$$

where $P_{tr} = ||\boldsymbol{w}||^2$ is the maximum transmitting power by the PT, and P_{ac}^{IRS} presents the consumption of power by active IRS. In [36], authors calculated the power consumption of IRS with N_I reflecting elements as

$$P_{ac}^{IRS} = N_I P_c + N_I P_d + P_{ac}^{Out}, \quad (16)$$

where P_{c} denotes the power consumption of controlling each reflecting element, P_{d} denotes power consumption by the direct current biasing, and $P_{\mathrm{ac}}^{\mathrm{Out}}$ is the output power of active IRS and can be calculated as

$$P_{ac}^{Out} = ||\Phi \boldsymbol{H} \boldsymbol{w}||^2 + ||\Phi||^2. \tag{17}$$

We have noticed that the power consumption of the system depends on how we set up the transmit and reflection beamforming. In essence, the design of these two types of beamforming critically influences the overall energy efficiency of the system.

III. DEFINING THE PROBLEM SPACE AND DRL SOLUTION PROPOSAL

A. PROBLEM FORMULATION

The primary objective of our work is to optimize the secrecy rate while ensuring a strong SNR at the PR through the use of physical layer security (PLS) techniques. In this approach, the IRS not only reflects incoming signals but also transmits its own data to an IoT device by leveraging the SR technology. To accomplish this, we aim to optimize two key components: the active beamforming vector $\boldsymbol{w}$ at the PT and the reflection coefficient matrix $\boldsymbol{\Phi}$ at the IRS. The IRS plays a crucial role in enhancing signal reception for users while simultaneously enabling secure communication. The ultimate goal is to establish a secure wireless communication link between the PT and the PRs while maintaining a minimum secrecy rate (R_{sec}) and improving signal reception at the PRs. From a computational perspective, this optimization problem can be formulated as

$$\begin{aligned}
P(1):\ \max_{\boldsymbol{w},\boldsymbol{\Phi}}\ & \sum_{u=1}^{U} R_{sec}^{u} && \text{(18a)}\\
\text{s.t. } & \gamma_U \geq \gamma_p, && \text{(18b)}\\
& \gamma_E \leq \gamma_e, && \text{(18c)}\\
& \gamma_T \geq \gamma_s, && \text{(18d)}\\
& R_{sec} \geq R_{sec}^{min}, && \text{(18e)}\\
& \boldsymbol{w} \leq P_{tr}, && \text{(18f)}\\
& 0 < \theta_I \leq 2\pi, \forall I \in N_I, && \text{(18g)}
\end{aligned}$$

where P_{tr} denotes the transmit power of the PT, the terms γ_p and γ_s represent the decoding SNRs necessary for the primary and secondary signals, respectively, while γ_e denotes the SNR threshold for the Eves. In our system, several constraints are critical to ensuring secure and reliable communication. To optimize the secrecy rate, (18a) serves an objective function, which focuses on adjusting the beamforming vector along with the amplitude and phase shift of the IRS. Equation (18b) presents the SNR requirement for accurate decoding of the $s(l)$ signal at users, while (18c) ensures secure transmission by regulating the maximum SNR achievable at Eves. Similarly, (18d) guarantees that the primary signals can be correctly decoded by the IoT device, enabling proper implementation of the SIC technique. Equation (18e) refers to the minimum secure rate requirement. The maximum power constraint at the PT is adhered to, while the phase of the IRS are managed within specified ranges to achieve the desired system performance in (18f) and (18g), respectively. The optimization problem in 18 poses significant challenges due to its non-convex nature with respect to both $\boldsymbol{w}$ and $\boldsymbol{\Phi}$, as well as the coupling between these optimization variables. This complexity complicates the determination of an optimal solution. Furthermore, we aim to design a robust beamforming structure that maximizes the system's worst-case achievable secrecy rate while satisfying worst-case constraints.

B. DRL ALGORITHM DESIGN AND SOLUTION STRATEGY

Our research seeks to boost the secrecy rate for primary users of wireless networks by optimizing the beamforming of the signal transmitted by the PT and the IRS reflection coefficients. This non-convex optimization problem poses considerable computational complexity. Additionally, the system exhibits time-varying CSI, including fluctuations in path loss, fading, and interference levels. To address this non-convex optimization problem, we reformulate it as a markov decision process (MDP), enabling the application of a DRL approach. Specifically, we utilize the DDPG algorithm to optimize secure and efficient beamforming while navigating the complexities of transmit power management and the intricate relationship between amplitude and phase shifts configurations within the defined constraints. DDPG is particularly well-suited for this application due to its ability to handle high-dimensional continuous action spaces, such as optimizing the IRS phase shift coefficients. Compared to algorithms like asynchronous advantage actor-critic (A3C) or proximal policy optimization (PPO), which are effective for discrete or hybrid action spaces, DDPG's deterministic policy and off-policy training offer faster convergence and better sample efficiency in continuous control tasks. Our method accounts for various dynamic factors that influence the performance of IRS-enabled secure communication systems in real-world scenarios, including application-specific requirements, variations in channel quality, and the capabilities of PRs.

C. MDP FRAMEWORK

This section commences with an elucidation of the essential components of the MDP. Research has demonstrated that the DDPG algorithm is particularly successful in addressing continuous control challenges [10], [37]. A central controller functioning as the DRL agent is incorporated into the PT. Specifically within the context of the DDPG algorithm, the agent-environment interaction is formalized as a MDP. This MDP is defined by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$, where $\mathcal{S}$ denotes the state space, $\mathcal{A}$ represents the action space, $\mathcal{P}$ specifies the state transition probabilities, $\mathcal{R}$ is the reward function, and γ is the discount factor. The DDPG agent navigates this MDP by traversing the continuous state space, selecting actions from the continuous action space, and receiving rewards based on the current state $\mathcal{S}$ and chosen action $\mathcal{A}$. The agent aims to learn a deterministic policy that optimizes the expected cumulative discounted reward. The core components of an MDP—state, action, and reward—are illustrated below.

- **State Space:** The agent observes the current state $s \in \mathcal{S}$ at each time step t of the environment in which controller collects the channel condition and information, including transmission rate and required secrecy rate. Finally, the state with respect to time s_t is formulated as

$$s_t = \{\boldsymbol{H}, \boldsymbol{h}_{b,t}, \boldsymbol{g}_{i,t}, \boldsymbol{g}_{i,u}, \boldsymbol{g}_{i,e}, R_U, R_T, R_{sec}^{min}\}, \tag{19}$$

  where R_U and R_T are the transmission rates of PRs and IoT device respectively, while R_{sec}^{min} is the required threshold for secrecy. The state serves as the input to the agent and based on the observed state, the agent uses its policy to select an action from the action space.
- **Action Space:** The action space encompasses all potential actions available to the agent. According to our model, the agent optimizes the considered variables, namely the active beamforming vector $\boldsymbol{w}$ and reflection coefficient matrix $\boldsymbol{\Phi}$ at each time step. For the DDPG, the action space is continuous, and the agent learns a deterministic policy to choose precise values for its actions. The selected action is then applied to the environment. Therefore, the action space $a_t \in \mathcal{A}$ can be formulated as

$$a_t = \{\boldsymbol{w}, \Phi\}. \tag{20}$$

- **Reward:** The feedback signal that the agent gets from the environment after acting is the reward. The reward function is structured to optimize security while reducing energy usage, in accordance with the system's operational limitations. There are two main parts of the reward's function, penalty and reward, which is formulated as

$$\text{Penalty} = \begin{cases} 0, & \text{if } R_{sec} \geq R_{sec}^{min}, \\ 1, & \text{if } R_{sec} < R_{sec}^{min}. \end{cases} \tag{21}$$

  Our optimization target is defined by the reward function, aiming to maximize the system secrecy rate of each user while satisfying their QoS criteria. Thus the reward function $r_t \in \mathcal{R}$ can be formulated as

$$r_t = \mu_1 R_{sec}(t) - \mu_2 P_t, \tag{22}$$

  where P_t is the utilization of power at time t while $\mu_1 and \mu_2$ are the weight balancing factors.

Discount Factor and Policy: The agent's goal is to maximize the overall discounted reward, which comes from:

$$R_t = \sum_{k=0}^{T} \gamma^k r_{t+k}, \tag{23}$$

where $\gamma \in [0, 1]$ serves as the discount factor, indicating the weight given to future rewards.

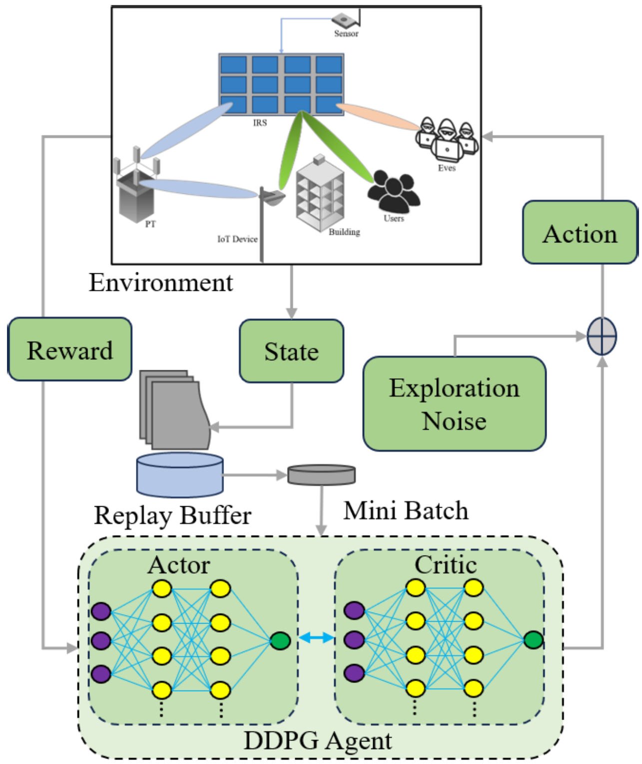

FIGURE 3. DDPG training model.

The policy in the DRL framework dictates how the agent chooses actions given a state. In DDPG, the policy is deterministic, meaning the policy is a function $\pi(s_t)$ that maps states to actions. Thus, the policy function is represented as

$$\pi(s_t) = a_t, \tag{24}$$

D. LEARNING AND TRAINING PROCESS OF THE DDPG AGENT

Figure 3 illustrates the integration of the DDPG algorithm within the proposed system to optimize the IRS phase shift matrix for enhanced secrecy performance. It provides an overview of the key components of the DDPG framework, including the actor and critic networks, the replay buffer, and the target networks, all of which work collaboratively to learn optimal phase shift strategies. By interacting with the THz-based IRS-assisted SR environment, the DDPG algorithm refines its policy through continuous feedback, aiming to maximize the secrecy rate. This process aligns with the system model by ensuring adaptive optimization in dynamic environments, directly contributing to improved PLS in the presence of multiple users and potential Eves. Moreover, we will discuss how to train the DDPG agent to find the optimal values for the active beamforming vector $\boldsymbol{w}$ and the reflection coefficient matrix $\boldsymbol{\Phi}$, while adhering to the MDP framework's constraints on secrecy rates and power consumption.

1) ACTOR-CRITIC ARCHITECTURE

The DDPG algorithm employs an Actor-Critic framework, with the actor responsible for selecting actions and the critic assessing the quality of these actions by estimating the value function, as illustrated in Fig. 3. The two networks used are:

- **Actor Network:** This network outputs continuous actions given the current state s_t. In our case, the actor network generates the active beamforming vector $\boldsymbol{w}$ and reflection coefficient matrix $\boldsymbol{\Phi}$ at each time step t.
- **Critic Network:** This network estimates the action-value function $Q(s_t, a_t)$, representing the expected cumulative reward for choosing action a_t in state s_t and then adhering to the current policy π.

The parameters for the actor and critic networks are θ^μ and θ^Q, respectively. The training process iterates over a series of updates to these parameters to improve the policy $\pi_{\theta^\mu}(s)$ and the value function $Q_{\theta^Q}(s, a)$.

2) EXPLORATION STRATEGY: NOISE ADDITION

To encourage exploration in the continuous action space, we apply noise to the actor's action during training. Specifically, we incorporate Ornstein-Uhlenbeck noise, which is suitable for problems involving inertia, like our beamforming and reflection coefficient optimization. The modified action with noise is expressed as:

$$a_t = \pi(s_t|\theta^\mu) + \mathcal{N}_t, \tag{25}$$

where $\mathcal{N}_t$ represents the added noise at time step t. This ensures the agent explores a variety of actions and prevents premature convergence to sub-optimal policies.

3) EXPERIENCE REPLAY BUFFER

The experience replay buffer is used to store the agent's interactions as tuples (s_t, a_t, r_t, s_{t+1}). During training, the agent draws mini-batches from this buffer to reduce correlations between sequential experiences, thus enhancing learning stability. The use of a replay buffer helps the agent learn from past experiences and avoids over fitting to recent interactions. The replay buffer is denoted as $\mathcal{D}$, and training samples are drawn randomly from this buffer:

$$(s_t, a_t, r_t, s_{t+1}) \sim \mathcal{D}. \tag{26}$$

4) TRAINING PROCEDURE

The training of the DDPG agent involves updating both the actor and critic networks using the samples drawn from the experience replay buffer. The training procedure is detailed as follows:

- **Critic Network Update:** The critic network is trained by minimizing the difference between the predicted Q-value and the target Q-value. The target Q-value is calculated using the Bellman equation, which is a fundamental principle in reinforcement learning that defines the relationship between the value of a state and the values of its successor states

$$y_t = r_t + \gamma Q'(s_{t+1}, \pi'(s_{t+1}|\theta^{\mu'})|\theta^{Q'}), \tag{27}$$

where Q' and π' are the target critic and actor networks, and $\theta^{\mu'}$, $\theta^{Q'}$ represent the target network parameters. These target networks are updated using a soft update mechanism:

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}, \tag{28}$$
$$\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}, \tag{29}$$

where τ is the soft update parameter (typically a small value like 0.001). The critic loss function is:

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

where N is the batch size.

Algorithm 1 DDPG Approach

Initialization: Given initialized No. of episodes E, and No. of steps S

Initialize the $Q(s_t, a_t)$, $Q_{\theta^Q}(s, a)$, θ^μ and θ^Q

for `Episode` $= 0, 1, 2, \cdots, E-1$ **do**
  Collect initial state s_t and action a_t
  for `t`$=0, 1, 2, \cdots, T-1$ **do**
    Select action $a_{t+1} = \pi(s_t|\theta^\mu, a_t)$ from the actor network
    Observe new state s_{t+1} and rewards r_{t+1}
    Save the experience (s_t, a_t, r_t, s_{t+1}) in the buffer
    Sample from the minibatch
    Target Q-value computed by the equation (27)
    Loss function computed by the equation (30)
    Update the policy gradient by the equation (31)
    Update target networks by equation (28) and (29)
  end for
end for

- **Actor Network Update:** The actor network is updated by maximizing the expected Q-value. The policy gradient is computed using the deterministic policy gradient theorem:

$$\nabla_{\theta^\mu} J \approx \frac{1}{N}\sum_i \nabla_a Q(s_i, a|\theta^Q)|_{a=\pi(s_i|\theta^\mu)} \nabla_{\theta^\mu}\pi(s_i|\theta^\mu). \tag{31}$$

This gradient is used to update the actor network parameters θ^μ using gradient ascent. Algorithm 1 presents the steps of the DDPG algorithm, detailing the initialization, action selection, and update processes used in each episode and time step to optimize policy and value functions.

E. SECURE BEAMFORMING OPTIMIZATION THROUGH DDPG TRAINING

Secure Beamforming Optimization through DDPG dynamically adjust beamforming parameters, thereby optimizing security in our considered environment. This approach addresses the continuous control requirements of secure beamforming by training an agent to maximize secrecy rates, minimize interference, and meet QoS constraints. The DDPG agent, functioning as the central controller, observes the current state of our communication environment and selects actions from a continuous action space to fine-tune the beamforming vector as shown in Algorithm 2. The agent receives a reward based on the effectiveness of its actions, focusing on improving secrecy rates and reducing power usage while ensuring secure transmission. Through iterative training, the agent learns a beamforming policy that prioritizes long-term rewards, reinforcing strategies that enhance security and efficiency in response to environmental changes like user movement or channel variations. Consequently, DDPG-based training enables real-time, adaptive secure beamforming, yielding a robust and optimized communication framework.

Algorithm 2 DRL-Based Beamforming Optimization for THz-Based IRS-Asisted SR Networks

Input: Channels and secrecy rate.
Output: The optimal $a_t = \{\boldsymbol{w}, \Phi\}$, Q function
Initialization: $\boldsymbol{w}$, Φ, Memory, DRL parameters.

1: **while do**
2: **for** `Episode` $= 0, 1, 2, \cdots, E-1$ **do**
3: Collect all channels for first state s_0
4: **for** `t`$=0, 1, 2, \cdots, T-1$ **do**
5: Update the action $a_t = \{\boldsymbol{w}_t, \Phi_t\} = \pi(s_t|\theta^{\mu})$
6: Execute the DDPG Algorithm 1
7: Modify the DRL parameters
8: Feed them into the agent as s_{t+1}
9: **end for**
10: **end for**
Until: Either converges or reaches the maximum number of iterations.
11: **end while**

IV. SIMULATION RESULTS AND DISCUSSION

This section presents a numerical assessment of the proposed DRL-based hybrid beamforming method for IRS-enhanced wireless THz communication systems with multiple Eves and users. We start by describing the configurations and parameters employed in the THz communication framework, followed by an analysis of the significant results.

A. SIMULATION SETUP

Given the considerable signal attenuation in THz propagation and the absorption effects caused by water molecules, our study focuses on scenarios involving short-distance communication. We define the distance from the PT to the IRS as 15 meters, while the distance to the IoT device is set as 12 meters. Furthermore, all users, along with Eves, are positioned 5 meters away from the IRS, with all users and Eves arranged within a circular area of 2 meter radius.

TABLE 1. DRL parameters.

Parameters	Description	Values
γ	Discounted factor of the future reward	0.99
τ	Learning rate of actor and critic network	0.001
λ_c	Decaying rate of training critic network	0.005
λ_a	Decaying rate of training actor network	0.005
B	Replay buffer size	35000
E	No. of episodes	1100
S	No. of steps in each episode	5000
Z	Batch size	32
N_h	No. of hidden layers	2
N_n	No. of neurons in each hidden layer	64
η	Learning rate of model	0.001

1) SYSTEM PARAMETERS

In our system model, we deployed a hybrid beamforming framework where the PT has $N_B = 4$ antennas, with $P_{tr} = 20dbm$ as transmit power, the IRS has $N_I = 32$ reflecting elements, and the numbers of users and Eves are $U = E = 4$. The transmission frequency is designated as $f = 0.3$ THz, utilizing a fixed bandwidth of 12 GHz. Small-scale fading in the channels is modeled using Rician fading with a Rician factor of $\mathcal{K} = 5$.

2) DRL PARAMETERS

Here, we outline the key parameters utilized in the proposed DDPG-based algorithm for training the agent in our study. In this study, the architecture of the actor and critic networks in the DDPG model was carefully designed to optimize performance in a continuous action space. The actor and critic networks each consist of two hidden layers with 64 neurons per layer, using the ReLU activation function. These parameters are crucial for ensuring optimal performance in the hybrid beamforming scenario. The values specified in Table 1 are selected based on preliminary experiments and relevant literature.

B. BENCHMARKS AND COMPARISON

Several optimization techniques have been proposed to enhance security in IRS-assisted symbiotic radio communication systems [24], [38], [39]. To evaluate the effectiveness of our proposed DDPG-SR optimization method and its performance against three approaches: deep Q-network (DQN), two conventional non-DRL techniques—the iterative optimization algorithm (IO) and the alternating optimization (AO) scheme. AO iteratively optimizes one subset of variables while keeping others fixed, facilitating convergence in IRS applications with reduced computational demands. IO further builds upon this by continually refining phase shifts across all variables until an optimal solution

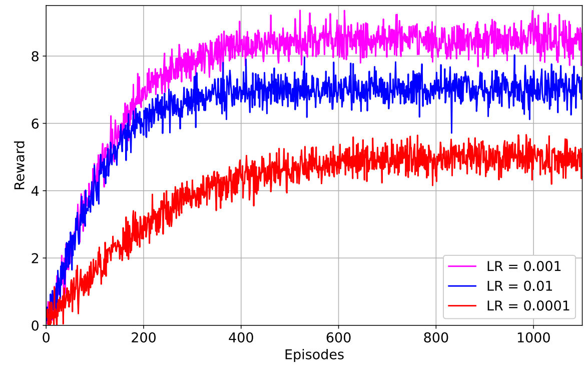

FIGURE 4. Reward progression for different learning rates.

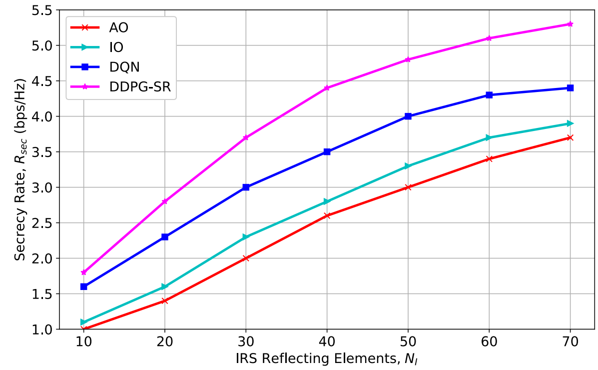

FIGURE 6. Secrecy rate vs. number of IRS elements.

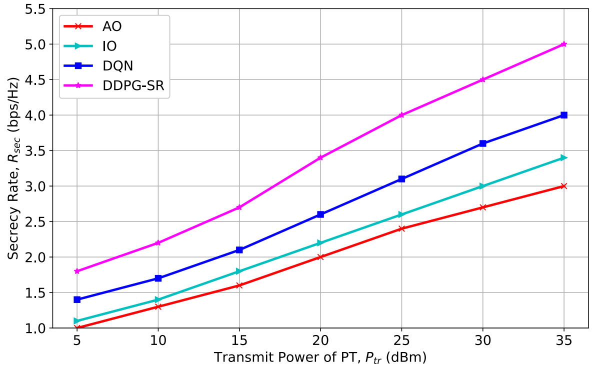

FIGURE 5. Secrecy rate vs. transmit power of PT.

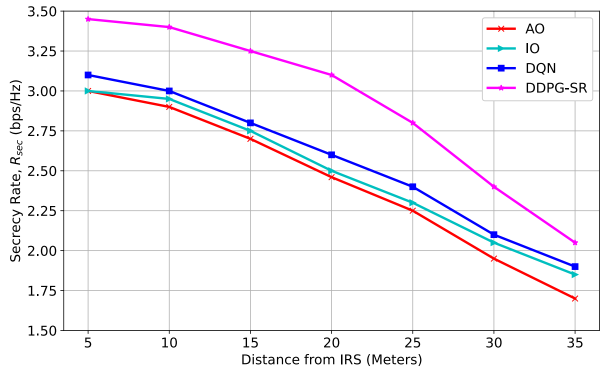

FIGURE 7. Secrecy rate vs. user distance from IRS.

is achieved, offering improved accuracy but at the cost of increased computation.

Recently, DQN has emerged as a powerful reinforcement learning alternative for IRS-based optimization. While DQN is effective at approximating optimal policies in discrete action spaces, its applicability in continuous action spaces is limited. Nevertheless, it provides a valuable benchmark for comparing adaptive optimization strategies. Together, AO, IO, and DQN offer diverse perspectives in optimizing IRS-assisted communication security, forming a foundation for evaluating advanced models like DDPG in multi-user, multi-eavesdropper scenarios. In contrast, our DDPG-SR approach continuously learns and adapts to the environment by leveraging a DRL framework, which allows it to optimize security parameters dynamically and respond to environmental changes more efficiently. The comparative analysis highlights the advantages of the DDPG-SR scheme in terms of secrecy rate, demonstrating its superior capability to maintain secure communication in THz frequencies where conventional techniques may struggle. Our proposed approach not only provides better security performance but also adapts effectively in complex, multi-user THz-based IRS-assisted networks.

The Fig. 4 illustrates the relationship between reward progression and different learning rates (LR) over several epochs, highlighting the impact of LR selection on model performance. Three different LRs $\eta = \{0.01, 0.001, 0.0001\}$ were tested. The LR of 0.001 outperforms both higher (0.01) and lower (0.0001) LRs due to its balanced approach to updating the model's parameters. With a higher LR, like 0.01, the model takes larger steps in the direction of the gradient, which can lead to overshooting the optimal solution, resulting in instability and fluctuating rewards. This makes it difficult for the model to converge to an optimal policy. On the other hand, a very low learning rate, such as 0.0001, allows only minimal updates to the model parameters. While this can lead to a stable, gradual approach, it significantly slows down the learning process. The LR 0.001 facilitates smoother convergence by neither overshooting nor progressing too slowly, enabling the DDPG model to reach higher rewards effectively.

Figure 5 compares the secrecy rate performance of different optimization schemes with the proposed DDPG-SR method across different transmit power levels of the PT. The DDPG-SR method consistently outperforms other schemes, reaching a maximum secrecy rate, showcasing its effectiveness in enhancing secure communication. While AO and IO achieve moderate increases, DQN performs slightly better but still falls short of DDPG-SR. This demonstrates DDPG-SR's superiority in dynamically optimizing security for IRS-assisted THz systems. This comparison highlights DDPG-SR's capability to dynamically adapt phase shifts and optimize security with different power levels of PT in IRS-assisted THz communication systems.

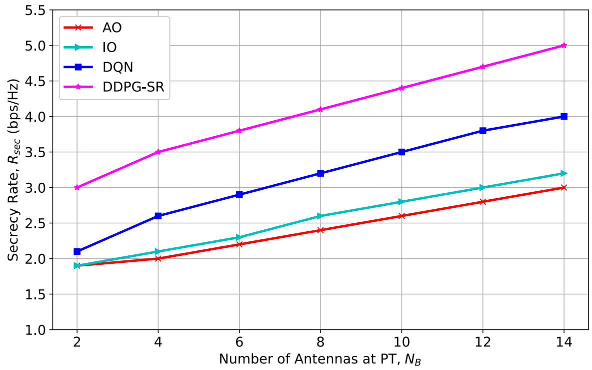

FIGURE 8. Secrecy rate vs. number of PT antennas.

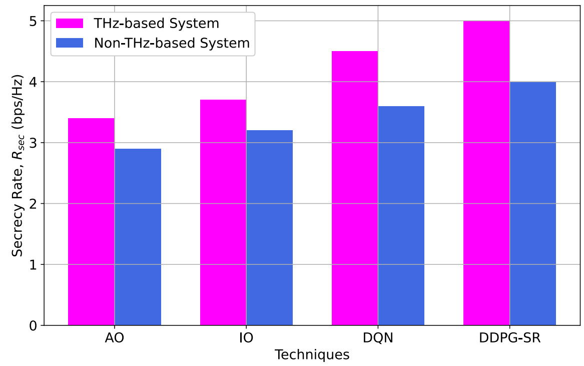

FIGURE 9. Comparison of secrecy rates for different techniques in THz and non-THz systems.

The correlation between secrecy rate and IRS elements is depicted in Figure 6, revealing distinct patterns across benchmark schemes. The DDPG-SR demonstrates superior performance, exhibiting a more pronounced increase in secrecy rate as the number of IRS elements grows. In comparison, the DDPG-SR shows a consistently rise in secrecy rate, while DQN shows moderate improvement and the AO displays the least improvement, generally resulting in lower secrecy rates. Figures 5 and 6 collectively emphasize the efficacy of the DDPG-SR, which maximizes the benefits of both increased transmit power and the number of IRS elements. This underscores its pivotal role in boosting the secrecy performance of THz-based SR networks enabled by IRS technology.

The Figure 7 illustrates the relationship between the secrecy rate and the distance of users from the IRS, comparing the performance of four algorithms: AO, IO, DQN, and DDPG-SR. As the distance from the IRS increases from 5 to 35 meters, the secrecy rate decreases across all methods, showcasing the impact of distance on model performance in secure communication scenarios. The DDPG-SR algorithm achieves the highest secrecy rate throughout, maintaining strong performance even at greater distances due to its effective parameter optimization for secure communication. At THz frequencies, signal attenuation is significantly higher compared to lower frequencies. As the distance between the IRS and the users increases, path loss grows rapidly, leading to a weaker received signal for the legitimate users. This results in a lower achievable rate for the legitimate communication link and a sudden drop in secrecy performance at higher distances compared to other benchmarks. Overall, DDPG-SR stands out as the most effective approach in this context, offering a smoother and higher secrecy rate progression, enabling it to maintain security more effectively over varying distances.

Figure 8 illustrates the impact of increasing the number of antennas at the PT on the secrecy rate in a THz-based IRS-assisted SR communication system, comparing DDPG-SR with benchmarks. As the number of antennas increases from 2 to 14, all techniques demonstrate a positive trend in secrecy rate, indicating that adding more antennas enhances secure communication. The DDPG-SR technique consistently achieves the highest secrecy rate, significantly outperforming the other methods across all antenna counts. The analysis of Figure 5 and Figure 8 shows that both increasing the transmit power and the number of antennas at the PT contribute positively to the secrecy rate in the communication system. Higher transmit power enhances the signal-to-noise ratio, enabling more secure transmissions. Simultaneously, an increased number of antennas facilitates more precise beamforming, directing the signal towards the intended receiver while minimizing the risk of interception by Eves.

Figure 9 shows a comparative analysis of the secrecy rates achieved by various optimization techniques in THz-based and non-THz-based communication systems. Across all techniques, the THz-based system demonstrates superior secrecy rates compared to the non-THz-based system, reflecting the enhanced security capabilities associated with THz frequencies. THz-based communication systems outperform non-THz systems due to several intrinsic advantages of THz frequencies. Their high directionality enables narrow beamforming, which minimizes signal leakage and reduces interception risks. Additionally, the high path loss of THz waves limits the range for potential Eves, while their wide bandwidth supports sophisticated modulation and encryption techniques, further enhancing security. Among the methods evaluated, DDPG-SR achieves the highest secrecy rate in both systems, underscoring its effectiveness in optimizing secure communication channels.

V. CONCLUSION

This work offers a comprehensive study on the security enhancement of a THz-based IRS-assisted SR communication system using DRL. The integration of IRS with THz communication enhances the security strength of the primary communication links, incorporating the proposed model with improvements even in scenarios involving multiple Eves. The proposed algorithm, based on DDPG, enables the dynamic optimization of the active beamforming vector and IRS reflection coefficient matrix, facilitating efficient real-time adaptation of the system to environmental changes and channel dynamics. Simulation results demonstrate that DDPG-SR outperforms IO, AO, and the DQN methods, particularly under dynamic environmental and network conditions. These studies highlight DRL's capability to solve complex problems like beamforming in the THz range. Besides, this work has demonstrated the feasibility of DRL-based optimization for IRS-assisted SR systems while also providing essential insights into the development of robust, high-performance networks capable of meeting the emerging requirements of next-generation IoT applications within the 6G framework and beyond. For future research, this work can be extended by investigating larger-scale networks, exploring diverse system topologies, and incorporating additional environmental factors such as mobility and channel uncertainty to enhance the reliability and adaptability of IRS-assisted SR communication systems. Future work will also focus on extending the proposed framework to account for imperfect CSI and ensure reliable performance under realistic conditions.

REFERENCES

[1] X. Xu, Y.-C. Liang, G. Yang, and L. Zhao, "Reconfigurable intelligent surface empowered symbiotic radio over broadcasting signals," *IEEE Trans. Commun.*, vol. 69, no. 10, pp. 7003–7016, Oct. 2021.

[2] N. Kassri, A. Ennouaary, S. Bah, and H. Baghdadi, "A review on SDR, spectrum sensing, and CR-based IoT in cognitive radio networks," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, pp. 1–22, 2021.

[3] H. F. Alhashimi, M. N. Hindia, K. Dimyati, E. B. Hanafi, N. Safie, F. Qamar, K. Azrin, and Q. N. Nguyen, "A survey on resource management for 6G heterogeneous networks: Current research, future trends, and challenges," *Electronics*, vol. 12, no. 3, p. 647, Jan. 2023.

[4] M. Sheraz, T. C. Chuah, K. Sultan, M. Ahmed, I. E. Lee, and S. C. Tan, "Improving cache-enabled D2D communications using actor–critic networks over licensed and unlicensed spectrum," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 36, no. 10, Dec. 2024, Art. no. 102249.

[5] M. B. Janjua and H. Arslan, "A survey of symbiotic radio: Methodologies, applications, and future directions," *Sensors*, vol. 23, no. 5, p. 2511, Feb. 2023.

[6] X. Lei, M. Wu, F. Zhou, X. Tang, R. Q. Hu, and P. Fan, "Reconfigurable intelligent surface-based symbiotic radio for 6G: Design, challenges, and opportunities," *IEEE Wireless Commun.*, vol. 28, no. 5, pp. 210–216, Oct. 2021.

[7] M. Ahmed, M. Shahwar, F. Khan, W. U. Khan, A. Ihsan, U. S. Khan, F. Xu, and S. Chatzinotas, "NOMA-based backscatter communications: Fundamentals, applications, and advancements," *IEEE Internet Things J.*, vol. 11, no. 11, pp. 19303–19327, Apr. 2024.

[8] W. Jiang, Q. Zhou, J. He, M. A. Habibi, S. Melnyk, M. El-Absi, B. Han, M. D. Renzo, H. D. Schotten, F.-L. Luo, T. S. El-Bawab, M. Juntti, M. Debbah, and V. C. M. Leung, "Terahertz communications and sensing for 6G and beyond: A comprehensive review," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 4, pp. 2326–2381, 4th Quart., 2024.

[9] M. Shehab, M. Elsayed, A. Almohamad, A. Badawy, T. Khattab, N. Zorba, M. Hasna, and D. Trinchero, "Terahertz multiple access: A deep reinforcement learning controlled multihop IRS topology," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 1072–1087, 2024.

[10] T. Hussain, M. Shahwar, S. Z. U. Abdin, Z. Zaheen, N. Jan, and M. Shoaib, "IRS integrated UAV communication to enhance physical layer security: A deep reinforcement learning approach," *Asian Bull. Big Data Manage.*, vol. 4, no. 2, pp. 135–149, May 2024.

[11] X. Peng, Q. Tao, X. Gan, and C. Zhong, "Intelligent reflecting surface enhanced cell-free symbiotic radio systems," *IEEE Internet Things J.*, vol. 10, no. 22, pp. 19545–19557, Apr. 2023.

[12] B. Ji, Y. Liu, T. Li, L. Xing, W. Wang, S. Mumtaz, X. Shang, W. Liu, and C. Han, "Anti-jamming null space projection beamforming based on symbiotic radio," *Comput. Model. Eng. Sci.*, vol. 138, no. 1, pp. 679–689, Jan. 2024.

[13] S. Hu, C. Liu, Z. Wei, Y. Cai, D. W. Kwan Ng, and J. Yuan, "Beamforming design for intelligent reflecting surface-enhanced symbiotic radio systems," in *Proc. IEEE Int. Conf. Commun.*, May 2022, pp. 2651–2657.

[14] R. Ma, J. Tang, X. Yin Zhang, K.-K. Wong, and J. A. Chambers, "RIS-assisted hybrid symbiotic radio network: From practical end-to-end model to multi-objective optimization," *IEEE Trans. Cognit. Commun. Netw.*, vol. 10, no. 5, pp. 1730–1743, Oct. 2024.

[15] H. Zhou, X. Kang, Y.-C. Liang, S. Sun, and X. Shen, "Cooperative beamforming for reconfigurable intelligent surface-assisted symbiotic radios," *IEEE Trans. Veh. Technol.*, vol. 71, no. 11, pp. 11677–11692, Nov. 2022.

[16] C. Zhou, B. Lyu, D. T. Hoang, and S. Gong, "Reconfigurable intelligent surface assisted secure symbiotic radio multicast communications," in *Proc. IEEE 96th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2022, pp. 1–6.

[17] C. Wang, Z. Li, T.-X. Zheng, D. W. K. Ng, and N. Al-Dhahir, "Intelligent reflecting surface-aided secure broadcasting in millimeter wave symbiotic radio networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 11050–11055, Oct. 2021.

[18] B. Lyu, C. Zhou, S. Gong, D. T. Hoang, and Y.-C. Liang, "Robust secure transmission for active RIS enabled symbiotic radio multicast communications," *IEEE Trans. Wireless Commun.*, vol. 22, no. 12, pp. 8766–8780, Apr. 2023.

[19] Y. Cao, S.-Y. Lien, Y.-C. Liang, and D. Niyato, "Toward intelligent non-terrestrial networks through symbiotic radio: A collaborative deep reinforcement learning scheme," *IEEE Netw.*, vol. 39, no. 1, pp. 211–219, Jan. 2025.

[20] M. Tariq, S. Ahmad, M. A. Jan, and H. Song, "Deep learning aided intelligent reflective surfaces for 6G: A survey," *ACM Comput. Surv.*, vol. 57, no. 3, pp. 1–37, Nov. 2024.

[21] D. A. Saifaldeen, A. M. Al-Baseer, B. S. Ciftler, M. M. Abdallah, and K. A. Qaraqe, "DRL-based IRS-assisted secure hybrid visible light and mmWave communications," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 3007–3020, 2024.

[22] B. Li, W. Liu, and W. Xie, "Secure transmission design for RIS-aided symbiotic radio networks: A DRL approach," *Digit. Commun. Netw.*, vol. 10, no. 6, pp. 1566–1575, Dec. 2024.

[23] M. Ahmed, A. Wahid, W. U. Khan, F. Khan, A. Ihsan, Z. Ali, K. M. Rabie, T. Shongwe, and Z. Han, "A survey on RIS advances in terahertz communications: Emerging paradigms and research frontiers," *IEEE Access*, vol. 12, pp. 173867–173901, 2024.

[24] J. Qiao, C. Zhang, A. Dong, J. Bian, and M.-S. Alouini, "Securing intelligent reflecting surface assisted terahertz systems," *IEEE Trans. Veh. Technol.*, vol. 71, no. 8, pp. 8519–8533, Aug. 2022.

[25] A. Kiran, A. Sonker, S. Jadhav, M. M. Jadhav, J. V. N. Ramesh, and E. Muniyandy, "Secure communications with THz reconfigurable intelligent surfaces and deep learning in 6G systems," *Wireless Pers. Commun.*, pp. 1–17, May 2024. [Online]. Available: https://link.springer.com/article/10.1007/s11277-024-11163-7

[26] B. Ning, Z. Chen, W. Chen, and L. Li, "Improving security of THz communication with intelligent reflecting surface," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2019, pp. 1–6.

[27] J. Qiao and M.-S. Alouini, "Secure transmission for intelligent reflecting surface-assisted mmWave and terahertz systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 10, pp. 1743–1747, Oct. 2020.

[28] N. U. Saqib, S. Hou, S. H. Chae, and S.-W. Jeon, "Reconfigurable intelligent surface aided hybrid beamforming: Optimal placement and beamforming design," *IEEE Trans. Wireless Commun.*, vol. 23, no. 9, pp. 12003–12019, Sep. 2024.

[29] R. Long, Y.-C. Liang, H. Guo, G. Yang, and R. Zhang, "Symbiotic radio: A new communication paradigm for passive Internet of Things," *IEEE Internet Things J.*, vol. 7, no. 2, pp. 1350–1363, Feb. 2020.

[30] Z. Wan, Z. Gao, F. Gao, M. D. Renzo, and M.-S. Alouini, "Terahertz massive MIMO with holographic reconfigurable intelligent surfaces," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4732–4750, Jul. 2021.

[31] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. L. H. Nguyen, L. Li, and K. Haneda, "Hybrid beamforming for massive MIMO: A survey," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 134–141, Sep. 2017.

[32] H. M. Al-Obiedollah, K. Cumanan, J. Thiyagalingam, A. G. Burr, Z. Ding, and O. A. Dobre, "Energy efficient beamforming design for MISO non-orthogonal multiple access systems," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4117–4131, Jun. 2019.

[33] Y.-H. You, Y.-A. Jung, S.-C. Choi, and I. Hwang, "Complexity-effective sequential detection of synchronization signal for cellular narrowband IoT communication systems," *IEEE Internet Things J.*, vol. 8, no. 4, pp. 2900–2909, Feb. 2021.

[34] S. Priebe, M. Kannicht, M. Jacob, and T. Kürner, "Ultra broadband indoor channel measurements and calibrated ray tracing propagation modeling at THz frequencies," *J. Commun. Netw.*, vol. 15, no. 6, pp. 547–558, Dec. 2013.

[35] Q. Zhang, Y.-C. Liang, and H. V. Poor, "Reconfigurable intelligent surface assisted MIMO symbiotic radio networks," *IEEE Trans. Commun.*, vol. 69, no. 7, pp. 4832–4846, Jul. 2021.

[36] R. Long, Y.-C. Liang, Y. Pei, and E. G. Larsson, "Active reconfigurable intelligent surface-aided wireless communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 4962–4975, Aug. 2021.

[37] A. Wahid, S. Z. U. Abideen, M. Ahmed, W. U. Khan, M. Sheraz, T. C. Chuah, and Y. L. Lee, "Advanced security measures in coupled phase-shift STAR-RIS networks: A DRL approach," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 36, no. 9, Nov. 2024, Art. no. 102215.

[38] Y. Pan, K. Wang, C. Pan, H. Zhu, and J. Wang, "UAV-assisted and intelligent reflecting surfaces-supported terahertz communications," *IEEE Wireless Commun. Lett.*, vol. 10, no. 6, pp. 1256–1260, Jun. 2021.

[39] J. Xu, Z. Zhu, Z. Chu, H. Niu, P. Xiao, and I. Lee, "Sum secrecy rate maximization for IRS-aided multi-cluster MIMO-NOMA terahertz systems," *IEEE Trans. Inf. Forensics Security*, vol. 18, pp. 4463–4474, 2023.

MUHAMMAD SHAHWAR received the B.E. degree in computer system engineering and the M.S. degree in computer engineering from Pakistan Universities, in 2017 and 2021, respectively. He is currently pursuing the Ph.D. degree in software engineering with the College of Computer Science and Technology, Qingdao University, Qingdao, Shandong, China. His current research interests include wireless communications, physical layer security, reconfigurable intelligent surfaces, terahertz communications, symbiotic radio, machine learning, and 6G technology.

MANZOOR AHMED received the Ph.D. degree in communication and information systems from Beijing University of Posts and Telecommunications, Beijing, China, in 2015. From 1997 to 2000, he was a Lecturer with the Balochistan University of Engineering and Technology, Turbat, Pakistan, and a Telecomm Engineer with government-owned telecommunication service provider National Telecommunication Corporation, Islamabad, Pakistan, from 2000 to 2011. He was a Postdoctoral Researcher with the Electrical Engineering Department, Tsinghua University, Beijing, from 2015 to 2018. He was an Associate Professor with the Department of Computer Science and Technology, Qingdao University, Qingdao, China. He is currently a Professor with the School of Computer and Information Science and the Institute for AI Industrial Technology Research, Hubei Engineering University, Xiaogan, China. He has several research publications in IEEE top journals and conferences. His research interests include resource allocation and offloading in vehicular communications and networking, fog and edge computing, physical-layer security, RIS, backscattering, and UAV communication. He received several awards, including the Distinction Award from the President of Pakistan, the Best Employee Award from NTC, and the Excellent Paper and Best Paper Awards from Intelligent and Converged Networks and the GameNets Conference, in 2023 and 2014, respectively.

TOUSEEF HUSSAIN received the B.E. degree in telecommunications engineering from The Islamia University of Bahawalpur, Pakistan, in 2020, and the M.S. degree in software engineering from Qingdao University, Shandong, China, in 2024. He is currently pursuing the Ph.D. degree in electronic science and technology with the School of Electronic Engineering, Beijing University of Posts and Telecommunication, Beijing, China. His current research interests include intelligent topology shaping in wireless communications, physical layer security, reconfigurable intelligent surfaces, symbiotic radio, machine learning, and 6G technology.

SAJED AHMAD received the bachelor's degree in computer science from the University of Swat, Pakistan, and the master's degree in software engineering from the College of Computer Science and Technology, Qingdao University, Qingdao, China, in June 2024. He specializes in wireless communication, focusing on UAV communication, physical layer security, and optical wireless systems. His research interests include securing networks, optimizing resources, and leveraging emerging technologies, such as RIS and NOMA. Dedicated to advancing 5G, 6G, and the IoT, he aims to enhance the efficiency, reliability, and security of modern communication systems.

WALI ULLAH KHAN (Member, IEEE) received the master's degree in electrical engineering from COMSATS University Islamabad, Pakistan, in 2017, and the Ph.D. degree in information and communication engineering from Shandong University, Qingdao, China, in 2020. He is currently with the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg. He has authored/co-authored more than 100 publications, including international journals, peer-reviewed conferences, and book chapters. His research interests include convex/nonconvex optimizations, non-orthogonal multiple access, reflecting intelligent surfaces, ambient backscatter communications, the Internet of Things, intelligent transportation systems, satellite communications, autonomous aerial vehicles, physical layer security, and applications of machine learning.

MUHAMMAD SHERAZ is currently a Research Fellow with the Faculty of Engineering, Multimedia University, Malaysia. His research interests include information-theoretic aspects of wireless networks, signal processing, mmWave and THz communications, AI, digital twin networks, resource allocation, non-orthogonal multiple access (NOMA), reconfigurable intelligent surfaces (RIS), ultra-reliable low latency communications (URLLC), edge computing, the IoTs, intelligent transportation systems (ITS), and industry 4.0. He has served as a TPC member for several conferences.

TEONG CHEE CHUAH received the B.Eng. and Ph.D. degrees from Newcastle University, U.K., in 1999 and 2002, respectively. He is currently a Professor with the Faculty of Engineering, Multimedia University, Malaysia. His current research interests include wired, wireless, and convergent networks, including network digital twins, mobile network lifecycle management, telecom OSS/BSS, multi-access edge computing, and the application of artificial intelligence and machine learning in 5G/6G networks.