# DRL-Based IRS-Assisted Secure Hybrid Visible Light and mmWave Communications

**DANYA A. SAIFALDEEN** [1], **ABDULLATIF M. AL-BASEER** [1] **(Member, IEEE),**
**BEKIR S. CIFTLER** [2] **(Member, IEEE), MOHAMED M. ABDALLAH** [1] **(Senior Member, IEEE),**
**AND KHALID A. QARAQE** [3] **(Senior Member, IEEE)**

[1]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar

[2]Division of Data Science and Artificial Intelligence, College of Computing and Information Technology, University of Doha for Science and Technology, Doha, Qatar

[3]Department of Electrical and Computer Engineering, Texas A&M University at Qatar, Doha, Qatar

CORRESPONDING AUTHOR: D. A. SAIFALDEEN (e-mail: dsaifaldeen@hbku.edu.qa)

**ABSTRACT** This paper explores a new advancement in physical layer security (PLS) techniques, focusing on the integration of Intelligent Reflecting Surfaces (IRS). This work centers on developing an intelligent hybrid system combining communication lines using millimeter wave (mmWave) and Visible light communication (VLC). The system comprises four VLC access points with light fixtures, reinforced by a mirror array sheet, and a mmWave access point with antennas, supported by a reflecting unit sheet. Within the system, both sheets function as IRS. The aim is to enhance the secrecy capacity (SC) of the system by optimizing the beamforming weights at the VLC fixtures, the beamforming weights at the mmWave AP, the mirror array configurations, and the phase shift vector while meeting specific power constraints. Given the numerous variables and the dynamic nature of user mobility, traditional optimization techniques may be inadequate for improving SC. To address this complexity optimally, we propose a deep reinforcement learning (DRL) approach based on the deep deterministic policy gradient (DDPG) technique. The DDPG algorithm can adapt to channel variations due to user movement and high-dimensional factors. Furthermore, it intelligently selects the optimal technique to improve SC, whether VLC or RF. Simulation results confirm the efficacy of our approach in enhancing the SC for the authorized receiver, particularly in mmWave connections.

**INDEX TERMS** Visible light communications, millimeter wave, deep learning, machine learning, deep reinforcement learning, intelligent reflecting surfaces, secrecy capacity, physical layer security, deep deterministic policy gradient.

## I. INTRODUCTION

THE RAPID growth of cellular systems, connected devices, and the Internet of Things (IoT) has led to an enormous increase in consumers' need for high-speed and reliable networks. This higher demand arises from the need to access various applications, including gaming, social networking, and online video streaming. Nevertheless, the limited availability of licensed spectrum resources and the existing technology significantly constrain conventional radio frequency (RF) communications in light of the increasing demands. Consequently, researchers are currently investigating novel paradigms that operate at higher frequencies in order to address these constraints and meet the needs of future generations. These include a combination of visible light communications (VLC) and millimeter wave (mmWave) technologies.

On the one hand, several advantageous features make VLC an appealing complement to RF technology and an

essential technology for sixth-generation (6G) networks. This technology is fascinating because it offers a large, unrestricted spectrum that is free to use and can be utilized for high-speed communications and illumination [1]. Additionally, VLC provides the dependable connection that customers require while also being an energy-efficient and cost-effective technology to install [2].

On the other hand, mmWave technologies, characterized by frequencies ranging from 30 to 300 GHz, offer a considerable amount of available bandwidth and hold significant promise for facilitating the transmission of multi-Gbps data rates. This capability can potentially increase the capacity of mobile networks and improve both the spectral and energy efficiency in wireless environments [3]. In dense interior situations, both VLC and mmWaves exhibit comparable features because they can transmit significant data rates. Moreover, both technologies include complementary attributes that enhance the resilience of the communication system against obstacles and threats. Hence, incorporating a mmWave link as a supplementary means of communication, capable of transmitting signals when the VLC link is unavailable or insufficiently secure for transmission, becomes advantageous.

In recent literature, the adoption of physical layer security (PLS) technology has gained increasing prominence, primarily focused on strengthening the security framework of wireless communication networks and effectively countering potential vulnerabilities. The primary objective of PLS is to mitigate the risk of unauthorized individuals (i.e., eavesdroppers) intercepting confidential data transmitted between the sender and the receiver. Implementing PLS methodologies involves utilizing channel features from the physical layer and applying signal processing techniques [4]. The measure utilized in assessing communication security is known as secrecy capacity (SC). It is defined as the difference between the capacity of the legitimate connection, which goes from the legitimate transmitter to the legitimate receiver, and the capacity of the eavesdropper link, which goes from the legitimate transmitter to the eavesdropper [5]. PLS approaches such as beamforming vectors and artificial noise-aided security are powerful tools used to tackle privacy issues and boost the secrecy capacity of the communication systems of VLC and RF networks [4], [6]. Beamforming in RF is defined as targeting the broadcast signal at a particular receiver [6]. However, in VLC systems, beamforming is defined as directing light towards a certain region or location to transmit it to the authorized receiver [7], [8].

To further improve the performance of PLS and SC in RF and VLC, several studies proposed the incorporation of IRS into communication systems [9]. IRS is a flat surface made up of several passive, inexpensive reflecting elements that reconfigure the propagation of the signal toward a specific direction in any wireless environment [10]. To enhance the signal-to-noise ratio (SNR) at the designated receiver, each reflecting element can autonomously modify both the phase and amplitude of the reflected input signal. The

IRS facilitates avoiding certain wiretapped communication environments, thereby enhancing the overall security and privacy of the system [10].

Integrating the IRS into VLC systems establishes a potential platform to support various technical trends and new applications. This incorporation aids in mitigating blockages, reducing interference, enabling multi-user access transmissions, increasing spectral efficiency, and facilitating complex interactions among network entities. In addition, leveraging IRS into VLC networks will maintain continuous non-line-of-sight (NLoS) connectivity, thereby supporting user mobility in vehicular communications [11]. In terms of PLS, IRS can enhance the SC by maximizing the channel gain of the legitimate receiver, jam the network by sending random multi-path reflections to the eavesdropper to generate fake noise or create precoding vectors at multiple IRS elements to establish secure beamforming at the legitimate receiver [11]. IRS enhances RF signal strength and propagation characteristics by controlling the smart radio propagation environment. IRS for RF is proven frequency-agnostic and, thus, flexible across different RF applications. IRS-aided improvements in PLS in cognitive radio networks highlight their role in enhancing network security and resilience against threats [12].

Therefore, this paper explores a hybrid system comprising VLC and mmWave connections, in which each link operates independently with a designated power range. The VLC network contains a singular access point (AP) that incorporates several VLC fixtures, each equipped with beamforming weights. The VLC network also includes an IRS mirror array sheet capable of considering different angular orientations to direct the light beam to the desired receiver and improve the overall SC [13]. On the other hand, the mmWave link has a singular AP that incorporates multiple transmitting antennas, each equipped with beamforming weights. Additionally, an IRS containing several reflective elements that exhibit various phase shifts is employed to redirect the communication link and control the secrecy levels within the system [14].

However, the IRS for both VLC and RF still faces several challenges. For example, in IRS-assisted VLC, when the reflecting unit sheet and the mirror array sheet encounter extra parameters, the control of the beam's direction becomes more complicated. In addition, the problem formulated in this paper to maximize the SC of the intended user involves using beamforming weights for both VLC and mmWave signals. Further, the orientations of mirror array sheets, the angles of individual mirrors, and the phase shifts are considered while maintaining power limits. The complicated issue at hand exhibits non-convexity, high dimensionality, and dynamism due to the simultaneous control of several parameters inside the system and the mobility of users. As a result, traditional approaches are ineffective in addressing these complex situations, where existing mathematical equations will be complicated to compensate for the large number of parameters that must be optimized to maximize the overall system's secrecy capacity. For this reason, it is necessary to

introduce alternative solutions to ensure the effective use of IRS for both VLC and mmWave.

To this end, we propose a novel approach in this work by employing deep reinforcement learning (DRL) to address the high dimensionality and complexity issues within the hybrid IRS-assisted VLC/RF network to ensure maximum SC efficiency. This decision is based on the model-free nature of DRL, as highlighted in previous works by Ciftler et al. [15], [16]. We use the deep deterministic policy gradient (DDPG) algorithm, which has been recognized as a very accurate DRL technique due to its ability to optimize a wide range of parameters according to its continuous action space. This paper proposes a strategy based on DDPG to optimize the beamforming vectors, mirror orientations, and phase shifts of VLC and mmWaves simultaneously. Particularly due to the lack of prior studies focusing on the use of DRL in PLS for IRS-assisted systems considering both VLC and mmWave technologies, we propose this approach.

The objective is to mitigate the vulnerability to security breaches from eavesdroppers. We develop a single-agent DDPG algorithm at the mmWave and VLC APs. Within the context of VLC, the agent will collaborate on controlling the angles of individual mirror elements (i.e., roll and yaw angles) present on the mirror array sheet. Also, the agent will adjust the yaw angle of the entire mirror array sheet. Besides that, the agent will manage the beamforming weights associated with each light fixture. In the context of mmWave communication, the agent at the IRS will control the beamforming weights of the transmitted signal and the phase shift of the antenna array.

The following significant contributions characterize our work:

- We develop a hybrid network with two links: VLC with multiple fixtures and mmWave with multiple transmitting antennas. Each link is equipped with beamforming capability.
- We consider an IRS with a mirror array sheet to demonstrate the impact of adjusting the mirror angles on the total SC of the VLC connection.
- We develop a reflecting element sheet acting as IRS, which shows its effect on achieving better SC for the mmWave link.
- We create and model a DRL-based approach to achieve optimal SC in a hybrid VLC and mmWave network that is supported by the IRS and has an eavesdropper engaged.
- We formulate the reward and penalty functions for the DDPG-based agent to optimize secrecy capacity while also weighting each communication technology for optimal selection.
- We simulate and study the DRL-based algorithm's complexity and performance under various mirror configurations and user locations. The outcomes effectively show how the suggested algorithm works while demonstrating its flexibility to optimize the SC of the system.

The following sections of the paper are structured as follows: Section II comprehensively surveys the existing literature on the studied topics. Section III thoroughly explains the proposed system model for the IRS-assisted secure mmWave and VLC systems. In Section IV, we will present and analyze the DDPG-based method for optimizing the SC while simultaneously controlling all system parameters. The conclusive outcomes are displayed in Section V to demonstrate the performance of the proposed method and validate its efficacy. Finally, we summarize our findings and discuss the next research in Section VI.

## II. RELATED WORK

This section provides a concise overview of previous studies conducted on hybrid systems that combine mmWave and VLC, as well as PLS in RF, mmWave, and VLC systems. Furthermore, it emphasizes the advantages of incorporating IRS and DRL-based methodologies into the suggested systems.

PLS has been studied in many works as an effective tool to secure communication systems, regardless of the channel's classification. A Gaussian MIMO and MISO wiretap channel were studied in [17] and [18], respectively, where the authors used an achievable scheme based on Gaussian signaling and then explored a tight upper bound that met the rate achieved. It has been demonstrated that, for both MIMO and MISO channels, beamforming is the optimal method for ensuring secure communication. In addition, research in [19] explored a new configuration using a multi-input, single-output, multi-eavesdropper (MISOME) channel. Similarly, they concluded that the beamforming approach is capable of obtaining maximum capacity.

The concept of creating a hybrid indoor VLC/mmWave system is introduced in [20], highlighting the advantages and disadvantages of the two technologies. They specifically developed an AP that facilitates mmWave and VLC connections. When the user is available, they assume that VLC serves the downlink and mmWave serves the uplink; otherwise, mmWave is utilized for both lines. The researchers tested the proposed system under various scenarios and showed effectiveness.

A hybrid RF and VLC system is examined in [21] and [22], where a LoS wideband RF communication system in the 60 GHz area is assumed, referred to as mmWaves. Average and outage throughput for every user is used to evaluate the hybrid indoor system's performance. The users are assigned to VLC or RF systems depending on the communication channel's situation. In addition, this work proved that introducing an RF system to VLC improves the user rate performance while satisfying the spectrum and power requirements to enhance the VLC and mmWave system's average and outage performance.

For a hybrid RF and VLC relaying wireless environment, researchers in [4], [7], [23], [24], [25] have concluded that it is a significant advance to use PLS for securing various communication systems from attacks. Through the

use of zero-forcing beamforming and power allocation minimization, Marzban et al. have researched how to achieve the maximum SC of a hybrid RF/VLC network while reducing the amount of electrical power used by the system [7]. The intended objective of Al-Khori et al. work in [23] was to optimize the achievable SC by developing beamforming vectors for both RF and VLC links while reducing the power needed to achieve this SC in the presence of eavesdroppers in various locations. To maximize both the secrecy capacity and outage probability for both cooperative and non-cooperative power saving for the RF/VLC relaying scheme, Al-Khori et al. introduced an approach based on zero-forcing beamforming in another research study [24]. Furthermore, the authors employed beamforming vectors and a cooperative joint relay-jammer selection method to mitigate eavesdropping attempts and optimize the possible SC in an RF/VLC relaying network [4], [25].

The significant advance in using IRS and PLS to secure various communication systems from attacks has positioned the technique to be very promising for research. In an RF system model consisting of a single-antenna eavesdropper, a single-antenna user, and a multi-antenna AP, IRS was incorporated into the system model to secure it [26], [27]. The author's goal in both studies was to increase the legitimate connection's secrecy rate by maximizing the phase elements of the IRS and the transmitter's beamforming vector. Both research findings indicated that the secrecy rate was increased after using the alternating optimization (AO) approach to achieve this aim [26], [27]. Additionally in [28], Chen et al. maximized the secrecy rate by optimizing beamforming vectors at the base station and the reflecting coefficients at the IRS. The maximization problem was solved using the AO approach and the path-following algorithm while considering the reflection coefficients' real-time limitations. In [29], the authors use IRS to improve the PLS of a MIMO non-orthogonal multiple access (NOMA) network that is wiretapped. The particle swarm optimization (PSO) algorithm maximized the secrecy sum-rate. The simulation results confirm that the proposed method's efficiency closely approximates the channel's secrecy capacity.

The authors in [30] used closed-form expressions to find the optimal IRS coefficient that maximizes the SC of a MIMO wiretapped system. The system includes a multi-antenna transmitter, a multi-antenna legitimate user, and a multi-antenna eavesdropper. Their numerical findings demonstrate that the intelligent reflecting surface substantially enhances secrecy rates in MIMO wiretap channels.

In a VLC SIMO indoor scenario in [31], the authors used an IRS mirrors array sheet and the PLS technique to prevent attacks from the eavesdropper and enhance the system's security. Optimizing the channel gain of a legitimate user while decreasing the number of eavesdroppers was the goal to increase the secrecy rate. To achieve this, a PSO-based technique was used to determine the ideal combination of mirror orientations [31]. Aboagye et al. suggested adding an IRS sheet to optimize the attainable rate to prevent the

blockage impact within the VLC transmission system [32]. The authors considered randomly positioned blocks between the direct line connecting the transmitter and the receiver. They developed a sine-cosine algorithm to determine the ideal orientation of IRS components such that obstacles are avoided and the achievable rate is maximized [32]. Recently, IRS has been merged to work with mmWave networks to increase their robustness against blockages in dense environments. In [14], authors established a system consisting of mmWave and multiple IRSs to assist the downlink communication between a base station (BS) and a receiver operating with a single antenna. The authors aimed to maximize the power of the received signal by adjusting the IRSs' phase shift coefficients and the BS's transmit precoding vector simultaneously. The simulation findings confirmed that, in single and multi-IRS scenarios, the quality of the received signal increases four times as the number of reflecting components increases. Furthermore, by creating reliable virtual line-of-sight (LoS) mmWave paths, IRSs improved the system's resilience against blockages.

In real-time, complex, and untraceable wireless communication scenarios, finding optimal solutions for many problems is challenging. Recently, researchers have used deep reinforcement learning techniques in this context. The authors of [16] and [33] employed Deep Reinforcement Learning techniques to tackle security issues in both VLC and RF wireless systems. Yang et al. suggested a DRL-based method to enhance the system's secrecy rate while maintaining service quality in an IRS-aided RF system with numerous legitimate users and eavesdroppers. The proposed algorithm optimizes both beamforming matrices of the base station and the IRS, respectively [16].

In IRS-aided systems, authors used Reinforcement Learning (RL) techniques to mitigate jamming interference in [34], [35]. Both works concluded that re-configuring the reflecting components' surface enhanced the security of the proposed systems. Xiao et al. suggested a VLC beamforming model based on RL for a VLC wiretap system to attain the optimum policy against eavesdropper threats. A VLC beamforming control system based on DRL was also presented to address high-dimensionality, continuous action, and state space [33]. The outcomes improved the overall secrecy rate of the VLC system while ensuring the effectiveness of the suggested DRL scheme. Taking the literature into consideration, no existing work has suggested a DRL-based method with a phase shift sheet and a mirror array sheet serving as IRS to maximize the secrecy capacity of a hybrid VLC/mmWave system.

Throughout this work, we develop and simulate a DRL-based single agent to enhance the secrecy capacity of the VLC/mmWave system. We achieve this by optimizing the beamforming weights for both links, the yaw angle of the mirror sheet, the orientation of the mirror, and the phase shift. We presume that the locations of both the intended receiver and the eavesdropper are known.
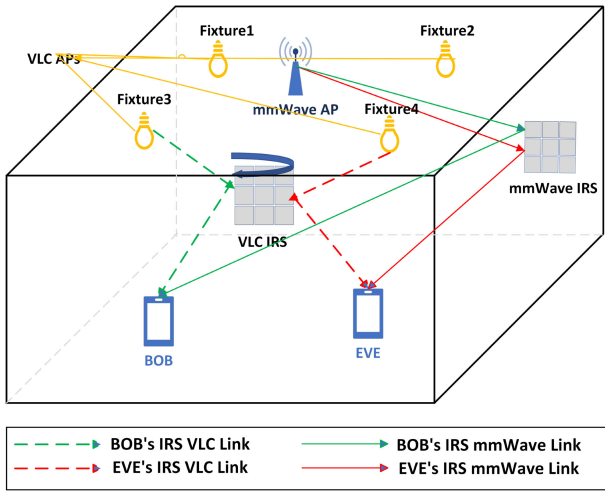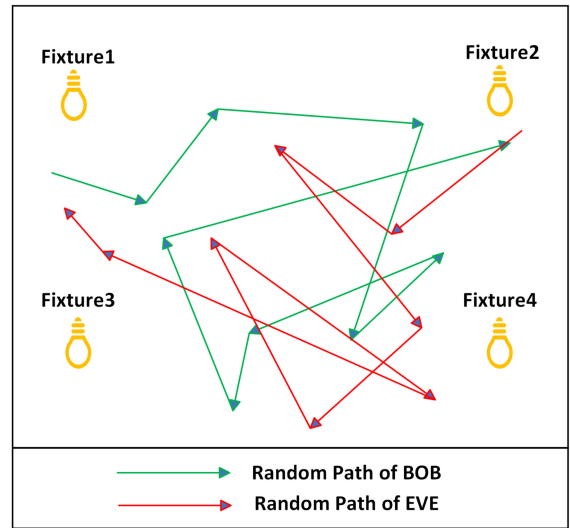
FIGURE 1. Model of IRS-Assisted VLC and mmWave system.



FIGURE 2. Paths of users during training phase.



FIGURE 3. Top view of the VLC system. The rotation of the mirror array sheet is represented by $\gamma$.

## III. SYSTEM MODEL AND PROBLEM FORMULATION

### A. SYSTEM MODEL

This research investigates using two communication channels, mmWave and VLC, in an indoor environment. The system consists of $K$ VLC fixtures and a sheet of mirror arrays that function as an IRS positioned in the middle of the room. Furthermore, a mmWave AP with $V$ transmitting antennas exists, and a sheet with $I$ reflecting elements for the IRS. This configuration is shown in Fig. 1. The dimensions of the room are denoted by $x_r$, $y_r$, and $z_r$.

The system comprises two users: the intended recipient, referred to as Bob, and the eavesdropper, identified as Eve, and both are represented by the symbol $l$. Under the assumption of NLoS communications, Eve attempts to intercept Bob's confidential message, which is transmitted by Alice (the source) through VLC and mmWave connections in a broadcast manner. This work only utilizes NLoS links because they are the most realistic and reliable. In dynamic environments where objects or people may move and obstruct the direct path between the transmitter and receiver, NLoS connections provide a more robust solution. Using reflections from various IRSs helps maintain connectivity despite environmental changes, and NLoS connections can also extend the system's coverage area. Besides that, we only utilize the NLoS link to reduce the complex structure of the proposed system.

The users' movements follow a uniformly distributed random waypoint model [36], and their speed ranges between 0 and 1 meters per second. Fig. 2 shows that every user starts their journey with a randomly allocated destination. Upon reaching that point, the system creates a new destination within the room using a random process.

At user $l$, the received signal $y^l$ is denoted as:

$$y^l = y^l_{vlc} + y^l_{mmw} \qquad (1)$$

### B. VLC LINK

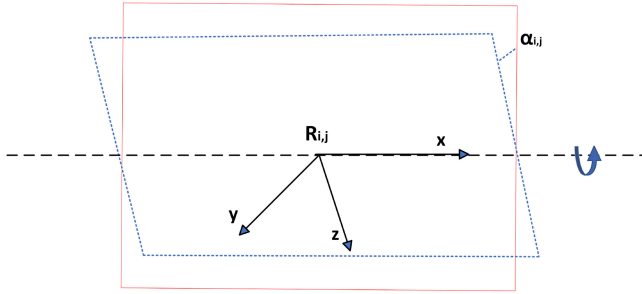This section comprehensively describes all the attributes and features of the VLC link. We assume the presence of four VLC fixtures with planar sources that emit light uniformly throughout the entire area. The fixtures are placed on the room's ceiling and serve as VLC APs equipped with beamforming capabilities.

An IRS sheet is located in the middle of the room with a total of $N_m \times N_n$ mirrors. Each mirror on the array sheet has dimensions of $b_m \times l_m$. Individual mirrors possess independent rotational capabilities to maximize orientation and increase user capacity, while the mirror array sheet may revolve along its vertical axis. The mirror array sheet's rotation is determined by the yaw angle $\gamma$ as shown in Fig. 3. Each mirror's rotation angle $(i, j)$ is shown in Fig. 4, where $\beta_{i,j}$ represents the yaw angle. The symbol $\alpha_{i,j}$ denotes the roll angle of each mirror located at position $(i, j)$ as shown in Fig. 5.

$$y^l_{vlc} = \left( \left( \sum_{k=1}^{K} w_k h^{vlc}_{l,k} \right) \sqrt{P^l_{vlc}} s + n^l_{vlc} \right) \qquad (2)$$

**FIGURE 4.** The rotation of angle $\beta$ in $i$th row and $j$th column.



**FIGURE 5.** The rotation of angle $\alpha$ in $i$th row and $j$th column.

The VLC link is determined by equation (2), where $w_k$ represents the beamforming weight for the $k$th fixture and $h_{l,k}^{\mathrm{vlc}}$ specifies the channel gain of the IRS link corresponding to the $k$th fixture. Our objective is to enhance the power of the VLC connection, denoted as $P_{vlc}^l$, in order to enhance the SC. The variable $s$ represents the signal being communicated. User $l$, where $l$ might be either Bob or Eve, experiences the noise denoted as $n^l$, which follows a normal distribution with mean 0 and variance $\sigma^2$. The noise in VLC is classified as Additive White Gaussian Noise (AWGN), characterized by a variance of $\sigma^2 = N_0 B$ and a mean of zero [37]. Based on the user's position within the room, there is a strong probability that they are within the coverage area of one or more VLC fixtures or mmWave antennas.

The mirror array sheet seen in Fig. 1 generates reflections that indirectly communicate between the fixture and the receiver. This communication line is known as the IRS link.

Given the system model depicted in Fig. 1, the channel gain of the connection between the IRS and fixture $k$ and user $l$ may be calculated using the following formula:

$$h_{l,k}^{\mathrm{vlc}}(\alpha, \beta, \gamma) = \sum_{i=1,j=1}^{N_m, N_n} \eta \varpi \sigma T E_{i,j}^{l,k}(\alpha, \beta, \gamma)$$
$$\times g\left(\theta_{\mathrm{R}_{i,j}}^{\mathrm{P}_{l,k}}\right) \cos\left(\theta_{\mathrm{R}_{i,j}}^{\mathrm{P}_{l,k}}\right), \quad (3)$$

where $E^{l,k}i,j(\alpha, \beta)$ is the irradiance representation at user $l$ through mirror $i, j$. The mirror array sheet is positioned at the origin of the Cartesian coordinate system. Geometrical analysis, as detailed in [31], can be used to derive the closed-form definition of $E^{l,k}i,j(\alpha, \beta)$ as:

$$E_{i,j}^{l,k}(\alpha, \beta, \gamma) = \frac{(L_a + 1)\rho}{2\pi} \int_{-h_m/2}^{h_m/2} \int_{-b_m/2}^{b_m/2} \cos^{L_a}\left(\theta_{\mathrm{R}_{i,j}}^{\mathbf{I}}\right)$$
$$\times \frac{\mathbf{a}_3^T(\mathbf{P}_{l,k} - \mathbf{R}_{i,j})\hat{\mathbf{N}}_{i,j}^T(\mathbf{P}_{l,k} - \mathbf{R}_{i,j})}{\|\mathbf{P}_{l,k} - \mathbf{R}_{i,j}\|_2^4}$$
$$\times \mathbb{1}\left(\mathbf{a}_1^T\mathbf{S}_k - \frac{b_s}{2} \leq \mathbf{a}_1^T\mathbf{I} \leq \mathbf{a}_1^T\mathbf{S}_k + \frac{b_s}{2}, \right.$$
$$\left. \mathbf{a}_2^T\mathbf{S}_k - \frac{l_s}{2} \leq \mathbf{a}_2^T\mathbf{I} \leq \mathbf{a}_2^T\mathbf{S}_k + \frac{l_s}{2}\right)dx''dz'' \quad (4)$$

where the $i$th column of the $3 \times 3$ identity matrix is represented by $\mathbf{a}_i$, and $\rho$ represents the mirror reflection efficiency. The $l_2$-norm and binary indicator functions are denoted by $|\cdot|_2$ and $\mathbb{1}(\cdot)$, respectively. To determine if user $l$ has an image reflection inside fixture $k$'s field of view, the binary indicator function is applied to the mirror at position $i, j$ on the mirror array sheet. The coordinates of the center of VLC AP $k$ in the room are specified as $\mathbf{S}_k$ in equation (7), while the location of user $l$ compared to the center of the VLC mirror is denoted as $\mathbf{P}_{l,k}$.

$$\mathbf{P}_{l,k} = \begin{bmatrix} x_l^* - \left(x_{s,k}^* + \frac{b_m}{2} + (j-1)b_m\right) \\ y_l^* \\ h_l - \left(z_{s,k} + \frac{h_m}{2} + (j-1)h_m\right) \end{bmatrix} \quad (5)$$

$$\mathbf{R}_{i,j} = \mathbb{R}_{i,j}^r[x'' \ 0 \ z'']^T, \quad (6)$$

$$\mathbf{S}_k = \begin{bmatrix} -\left(x_{s,k}^* + \frac{b_m}{2} + (j-1)b_m\right) \\ y_{s,k}^* \\ -\left(z_{s,k} + \frac{h_m}{2} + (j-1)h_m\right) \end{bmatrix} \quad (7)$$

$$\mathbf{I} = \begin{bmatrix} \mathbf{a}_1^T\left(\mathbf{R}_{i,j} + \frac{\mathbf{a}_3^T(\mathbf{S}_k - \mathbf{R}_{i,j})}{\mathbf{a}_3^T\widehat{\mathbf{R}_{i,j}}}\widehat{\mathbf{R}_{i,j}}\right) \\ \mathbf{a}_2^T\left(\mathbf{R}_{i,j} + \frac{\mathbf{a}_3^T(\mathbf{S}_k - \mathbf{R}_{i,j})}{\mathbf{a}_3^T\widehat{\mathbf{R}_{i,j}}}\widehat{\mathbf{R}_{i,j}}\right) \\ \mathbf{a}_3^T\mathbf{S}_k \end{bmatrix} \quad (8)$$

where the position of user $l$ in Cartesian coordinates is represented by $x_l$ and $y_l$, while $x_{s,k}$ and $y_{s,k}$ indicate the $k$th fixture's location with respect to the origin. The mirror $i, j$ is rotating, and it is described by $\mathbb{R}_{i,j}^r$. On the other hand, the rotation of the mirror array sheet at an angle $\gamma$ around its vertical axis changes their coordinates in Cartesian space as illustrated in the equations below; however, the user's distance from the light fixture stays the same.

$$x_l^* = \sqrt{x_l^2 + y_l^2}\cos\left(\gamma + \tan^{-1}\left(\frac{y_l}{x_l}\right)\right) \quad (9)$$

$$y_l^* = \sqrt{x_l^2 + y_l^2}\sin\left(\gamma + \tan^{-1}\left(\frac{y_l}{x_l}\right)\right) \quad (10)$$

$$x_{s,k}^* = \sqrt{x_{s,k}^2 + y_{s,k}^2}\cos\left(\gamma + \tan^{-1}\left(\frac{y_{s,k}}{x_{s,k}}\right)\right) \quad (11)$$

$$y_{s,k}^* = \sqrt{x_{s,k}^2 + y_{s,k}^2} \sin\left(\gamma + \tan^{-1}\left(\frac{y_{s,k}}{x_{s,k}}\right)\right). \quad (12)$$

where the function represented by the binary indicator in equation (4) assesses if user $l$'s reflected image is visible to fixture $k$ through mirror $i,j$. The mirror at $i,j$ had a normal vector denoted by $N_{i,j}$ and is determined based on the rotation angles $\alpha_{i,j}$ and $\beta_{i,j}$.

$$\widehat{\mathbf{N}}_{i,j} = \begin{bmatrix} \cos(\alpha_{i,j})\sin(\beta_{i,j}) \\ \cos(\alpha_{i,j})\cos(\beta_{i,j}) \\ \sin(\alpha_{i,j}) \end{bmatrix}. \quad (13)$$

### C. MMWAVE LINK

Our proposed system incorporates an IRS-assisted mmWave downlink to assist the existing IRS-assisted VLC link during secure data transmission, as shown in Figure 1. An IRS with $I$ reflecting units is employed to transmit information from the BS to a single-antenna receiver $L$. The mmWave link is represented as follows:

$$y_{mmw}^l = \left(\left(\sum_{i=1}^{I}\sum_{v=1}^{V} w_v f_{i,v}^{\text{mmw}}\theta_i z_{l,i}\right)\sqrt{P_{mmw}^l}s + n_{mmw}^l\right) \quad (14)$$

where $w_v$ is the beamforming weight for $v$th antenna, and $f_{v,i}^{\text{mmw}}$ is the mmW link channel gain from each antenna $v$ to the $i$th mmWave IRS element. $\theta_i$ is the phase shift matrix of the $i$th mmWave IRS element, $z_{l,i}$ is the channel gain between the $i$th mmWave IRS element and user $l$. $P_{mmw}^l$ is the power associated with the mmWave link, where we aim to maximize it to maximize the SC. The transmitted signal is $s$, and the mmWave noise of user $l \in \{\text{Bob, Eve}\}$ is expressed as $n^l \sim \mathcal{N}(0, \sigma^2)$, where $N_0 \approx 10^{-22}\text{A}^2/\text{Hz}$. It is worth mentioning that AWGN with zero mean and variance of $\sigma^2 = N_0 B$ is assumed [37].

In Fig.1, the mmWave link represents an indirect communication between the transmitting antenna and the end receivers, involving reflections from the reflecting unit's sheet. The channel gain from antenna $v$ to mmWave IRS element $i$ and from mmWave IRS element $i$ to user $l$ can be expressed based on the system model described previously.

$$f_{i,v}^{\text{mmw}} = \sqrt{10^{PL_{dB}(d_{i,v})/10}} \times a_{iv} \quad (15)$$

where, $a_{iv}$ in the antenna response vector between the mmWave IRS element and user $l$, which is given by:

$$a_{i,v} = \left(1, e^{-j\frac{2\pi d}{\lambda}\sin\vartheta_{i,v}\cos\phi_{i,v}}, \ldots, e^{-j\frac{2\pi d}{\lambda}\sin\vartheta_{i,v}(N_x-1)\cos\phi_{i,v}}\right) \quad (16)$$

$$\otimes\left(1, e^{-j\frac{2\pi d}{\lambda}\sin\vartheta_{i,v}\cos\phi_{i,v}}, \ldots, e^{-j\frac{2\pi d}{\lambda}\sin\vartheta_{i,v}(N_y-1)\cos\phi_{i,v}}\right) \quad (17)$$

where $I_x$ and $I_y$ are the numbers of the reflecting elements in the $x \times y$ plane, $d$ is the spacing between mmWave IRS elements, and $\lambda$ is the wavelength. The azimuth departure angle is represented by $\vartheta_{i,v}$, and $\phi_{i,v}$ is the elevation angle of departure.

$$\vartheta_{i,v} = \sin^{-1}\frac{h_{i,v}^{\text{mmw}}}{d_{i,v}} \quad (18)$$

$$\phi_{i,v} = \cos^{-1}\left(\frac{y_i - y_u}{\sqrt{(x_i - x_v)^2 + (y_i - y_v)^2}}\right) \quad (19)$$

The channel gain matrix from $z_{l,i}$, the mmWave IRS element, to the mmW access point follows the same model.

### D. OBJECTIVE FUNCTION

In this work, the main objective is to employ agents based on DRL to maximize the secrecy capacity of the mmWave and VLC links at the intended user while maintaining power constraints for the overall system. The objective of the VLC link will be achieved by optimizing the beamforming weights of the VLC link and the mirror angles of the IRS sheet and the individual mirrors. For the mmWave link, optimization will be done for the beamforming weights and phase shifts of the antenna arrays.

The SC of the VLC link, taking into account the presence of an eavesdropper whose position is known, can be mathematically formulated as:

$$C_{vlc}^{\text{sec}}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma, \boldsymbol{w_k}) = \frac{1}{2}\log\frac{6A^2\sum_{k=1}^{K}w_k h_{Bob,k}^{\text{vlc}} + 3\pi e\sigma^2}{\pi eA^2\sum_{k=1}^{K}w_k h_{Eve,k}^{\text{vlc}} + 3\pi e\sigma^2}. \quad (20)$$

From the above equation, $h_{Bob,k}^{\text{vlc}}$ represents the IRS channel gain of the connection link between Bob and the $k$th fixture. Similarly, the value of $h_{Eve,k}^{\text{vlc}}$ describes the IRS channel gain of the connection link between Eve and $k$th fixture.

For the mmWave link, we model the SC of Bob, taking into account the presence of an eavesdropper whose location is known, as follows:

$$C_{mmW}^{\text{sec}}(\boldsymbol{\theta}, \boldsymbol{w_v}) = \log_2\left(1 + \frac{|w_v f_{i,v}^{\text{mmw}}\theta_i z_{l,i}|^2}{\sigma^2}\right) \quad (21)$$

By modifying the beamforming weight vectors, and the mirror orientations or phase shifts, and weighting the total power $P_T$ based on the selected system (i.e., VLC or mmWave), the optimization problem for enhancing the SC of both VLC and mmWave connections is given as follows:

$$\max_{\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma, \boldsymbol{w_k}, \theta_i, \boldsymbol{w_v}, P_{vlc}, P_{mmw}} C_{vlc}^{\text{sec}} + C_{mmW}^{\text{sec}}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma, \boldsymbol{w_k}, \theta_i, \boldsymbol{w_v}) \quad (22)$$

subject to (s.t.)

$$P_{vlc}^l + P_{mmw}^l \leq P_T \quad (23)$$

$$-\frac{\pi}{2} \leq \alpha_{i,j} \leq \frac{\pi}{2}, \quad (24)$$

$$-\frac{\pi}{2} \leq \beta_{i,j} \leq \frac{\pi}{2}, \quad (25)$$

$$-\frac{\pi}{2} \leq \gamma \leq \frac{\pi}{2}, \quad (26)$$

$$0 \leq \theta_i \leq 2\pi, \quad (27)$$

$$0 \leq w_v \leq 1, \quad (28)$$

$$\sum_{v=1}^{V} w_v = 1. \quad (29)$$

$$0 \leq w_k \leq 1, \tag{30}$$

$$\sum_{k=1}^{K} w_k = 1. \tag{31}$$

Setting the decision variables $P_{vlc}$ and $P_{mmw}$ determines the power allocation for VLC and mmWave. This aids in directing the system to utilize the optimal technology that maximizes the SC, either VLC or mmWave. The sum power of VLC and mmWave links should be less than the total power of the system ($P_T$).

The yaw angle of the mirror is $\alpha$, and the roll angle of the mirror is $\beta$, whereas the whole mirror array sheet's yaw angle is $\gamma$. Limiting all mirror angles to a range of $-\frac{\pi}{2}$ to $\frac{\pi}{2}$ is practical and intuitive for controlling the orientation of reflective surfaces. It also ensures a consistent representation, where negative angles correspond to reflections in one direction and positive angles correspond to reflections in the opposite direction.

Theta $\theta_i$ is the phase angle of the mmWave link, which is considered to be between 0 and $2\pi$. Restricting the phase angle to a specific range eliminates redundancy and ensures a concise representation of the phase shift. It is important to recall that in this context, we only optimize the phases of the mmWave IRS, where there is only a single receiver.

$w_k$ and $w_v$ are the beamforming weight vectors for VLC and mmWave links, respectively. The constraints set both weights between 0 and 1. The normalization equations expressed in Eqs. (29) and (31) provide a standardized and consistent approach to power control and distribution in both the VLC system and the mmWave antenna arrays.

The optimization problem defined in (22) requires optimizing all listed parameters while maintaining the constraint listed in (31). This is a non-convex problem, and its complexity will be examined in the next section. In addition, the mobility of both users, Eve and Bob, puts the problem at a higher level of complexity to be solved with conventional methods. Consequently, we propose a novel solution based on DDPG to address the optimization issue presented in (22) within a proper timing frame. We will discuss the details of this solution in the following section.

## IV. DRL-BASED IRS-ASSISTED SECRECY CAPACITY MAXIMIZATION

For the created hybrid mmWave/VLC network, this section discusses the details of the DRL-based IRS-assisted SC maximizing algorithm.

The DRL-based agent controls the beamforming weight vector of the fixtures, the yaw angle of the mirror array sheet, and the roll and yaw angles of each mirror for the VLC connection. Meanwhile, for the mmWave link, the agent controls the signal's beamforming weight and the phase shift vector. In addition, the agent controls the power ratio, which allocates a certain percentage of power to either the VLC or mmWave link concerning the location of Eve and Bob. The policy is set based on the locations of Bob and Eve, and in this system, it is assumed to be known by the agent.

---

**Algorithm 1** Training of DDPG-Based IRS-Assisted Hybrid VLC and mmWave Network to Maximize SC

---

1: **Initialization:** Set $t = 0$ and initialize a replay buffer of agents $\mathcal{D}$.
2: Initialize random weights of actor-network $\theta^{\mu}$ and critic network $\theta^{Q}$.
3: Initialize the target networks using actor and critic networks: $\theta^{\mu'} \leftarrow \theta^{\mu}$ and $\theta^{Q'} \leftarrow \theta^{Q}$.
4: Initialize convergence flag isConverged $\leftarrow$ False and convergence threshold $\epsilon$, $t_{lim}$.
5: **while** not isConverged or $t < t_{lim}$ **do**
6:     Observe state $\mathbf{s}_t$ (positions of Bob and Eve) and determine an action (VLC beamforming vector, mmWave beamforming vector, mirror array sheet yaw angle, mirrors' yaw and roll angles) for agent $\boldsymbol{a}_t = \mu(\mathbf{s}_t|\theta^{\mu}) + \boldsymbol{n}_t$
7:     Execute all actions $\boldsymbol{a}_t$.
8:     Receive the reward $r_t$, and observe next state $s_{t+1}$, store transition $(\mathbf{s}_t, \boldsymbol{a}_t, r_t, \mathbf{s}_{t+1})$ in $D$.
9:     Randomly sample mini-batch transitions from $\mathcal{D}$: $B = \{(\mathbf{s}_i, \boldsymbol{a}_i, r_i, \mathbf{s}_{i+1})\}$.
10:     Compute the targets for actor and critic networks: $\tilde{Q}(\mathbf{s}_i, \boldsymbol{a}_i|\theta^{Q}) = r_i + \gamma_L Q(\mathbf{s}_{i+1}, \mu(\mathbf{s}_i|\theta^{\mu'})|\theta^{Q'})$
11:     Update the $\theta^{Q}$ in the critic network by minimizing the loss: $L = \frac{1}{|B|} \sum_{i=1}^{|B|} \left( \tilde{Q}(\mathbf{s}_i, \boldsymbol{a}_i|\theta^{Q'}) - Q(\mathbf{s}_i, \boldsymbol{a}_i|\theta^{Q}) \right)^2$
12:     Update the $\theta^{\mu}$ in actor-network according to the sampled policy gradient: $\nabla_{\theta^{\mu}} J \approx \frac{1}{|B|} \sum_{i=1}^{|B|} \nabla_a Q(\mathbf{s}_i, \boldsymbol{a}_i|\theta^{Q}) \nabla_{\theta^{\mu}} \mu(\mathbf{s}_i|\theta^{\mu})$
13:     Update the target networks: $\theta^{Q'} \leftarrow \tau\theta^{Q} + (1-\tau)\theta^{Q'}$ $\theta^{\mu'} \leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu'}$
14:     Check for convergence (e.g., based on changes in $\theta^{\mu}$ or loss $L$): isConverged $\leftarrow$ True if convergence criterion (e.g., $|L_{new} - L_{old}| < \epsilon$) is satisfied
15:     $t \leftarrow t + 1$
16: **end while**

---

As a result, we can define the agent's DRL state space ($\mathbf{s}_t$) as stated below.

$$\mathbf{s}_t = [\mathbf{x}_B, \mathbf{x}_E]^T, \tag{32}$$

$X_B = [x_B, y_B, z_B]$ is where Bob is located and $X_B = [x_B, y_B, z_B]$ represent the coordinates of the eavesdropper. After learning, the agent's actions for the VLC link consist of the beamforming vector of the light fixtures, the yaw angle of the mirror array sheet, and each mirror's yaw and roll angles. Meanwhile, for the mmWave, the actions consist of the beamforming weights and phase shift angles. As discussed earlier, the power ratio is another action to determine the percentage of the power allocated for VLC and RF.

By using the Ornstein-Uhlenbeck (OU) process, the action space is characterized as a policy function with associated noise, and the equation is demonstrated below:

$$\boldsymbol{a}_t = \boldsymbol{\mu}(\mathbf{s}_t|\theta^{\mu}) + \boldsymbol{n}_t$$
$$= \left[ \boldsymbol{a}_t^{\text{BF}_{\text{VLC}}} \; a_t^{\gamma} \; \boldsymbol{a}_t^{\alpha} \; \boldsymbol{a}_t^{\beta} \; \boldsymbol{a}_t^{\text{BF}_{\text{MMW}}} \; PR \right], \tag{33}$$

where $\boldsymbol{\mu}$ represents the DRL agent's policy function, and $\boldsymbol{n}$ represents the noise of the correlated action based on the OU

process. The parameters of the DRL agent, characterized by neural network weights, are defined by $\theta$ [38]. Apart from $\mathbf{n}$, the OU process parameters given in Section V have an impact on the DRL agent in terms of the trade-off between exploration and exploitation.

The action vector of the VLC beamforming weights is stated as $\boldsymbol{a}_t^{BF_{VLC}}$, with a vector length of $K$. $a_t^{\gamma}$ is a scalar value representing the action of the mirror array sheet angle. The actions representing the yaw angle of the mirror are defined as $\boldsymbol{a}_t^{\alpha}$. In contrast, the actions of the roll angles are represented as vector $\boldsymbol{a}_t^{\beta}$. The size of each vector is $(N_m \times N_n)$, equivalent to the number of mirrors. On the other hand, the mmWave beamforming weights action vector is defined as $\boldsymbol{a}_t^{BF_{MMW}}$, and the power ratio action is symbolized as $PR$.

For the proposed system, our objective function aims to maximize the problem stated in (22), where the goal is to achieve the optimal (highest) SC as in (20). The optimum policy function can be used to accomplish this, given all stated limitations, such as the system's total power, the beamforming weights of both links and the angles' limits. As a result, the reward function in this algorithm is defined as the SC, which is represented as follows:

$$r_t = C_t^{sec} + P_t, \tag{34}$$

where the penalty function, $P_t$, is introduced to ensure a non-zero SC and is expressed as

$$P_t = \begin{cases} 0, & \text{if } C_t^{sec} > 0 \\ \Gamma, & \text{if } C_t^{sec} = 0, \end{cases} \tag{35}$$

where $\Gamma$ represents the penalty value.

### A. DDPG ALGORITHM

DDPG, a model-free, off-policy reinforcement learning algorithm, is designed for environments with continuous action spaces. It combines the benefits of deep neural networks with the principles of deterministic policy gradients, enabling efficient learning and robust performance. Unlike stochastic policies, DDPG learns deterministic policies. This allows for easier optimization and updates with gradient ascent. In particular, we employ the actor-critic method, combining the advantages of policy-based and value-based reinforcement learning approaches. Figure 6 demonstrates the flowchart of the DDPG algorithm.

Algorithm 1 illustrates the comprehensive algorithm based on DDPG that is suggested to optimize the SC of an IRS-assisted hybrid VLC/mmWave network.

At the start of training, we initialize an empty reply buffer with a capacity of $M$ to store transitions containing state, action, reward, and next-state tuples. In Step 2, the algorithm initializes the actor and critic networks with corresponding random weights. The actor-network learns the policy directly, mapping states to actions, whereas the critic network evaluates the actions the actor suggests by estimating the Q-value function. Step 3 generates the target
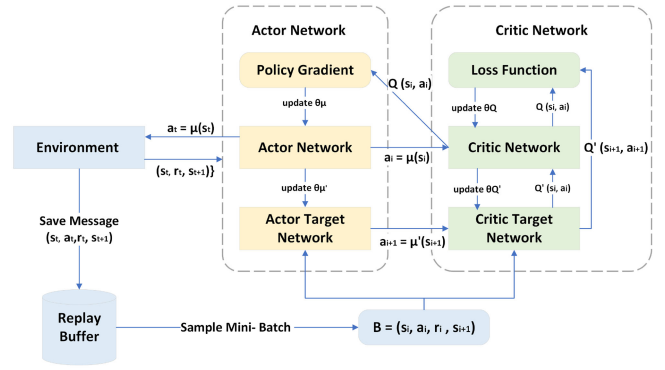


**FIGURE 6.** Overview flowchart of DDPG algorithm with actor-critic method.

networks using the actor and critic networks with the same weights. The algorithm determines a convergence threshold using a predefined threshold and sets the iteration limit as a final initialization step.

During the training process, the algorithm's steps 5 through 15 are repeated for each iteration $t$. In Step 6, the policy function decides the action set once the environment state (i.e., Bob and Eve's locations) is seen. Action is executed within the environment in Step 7, resulting in reward and observation for the next state, which will be stored in the transitions as stated in Step 8.

Step 9 involves a random sampling for a mini-batch of transitions from the replay buffer. Using the Bellman equation, the target values of the critic and actor networks are computed in Step 10. Steps 11 update the critic network by minimizing the loss function that quantifies the difference between predicted and target Q-values. In this case, mean squared error (MSE) loss is used.

Following that, in Step 12, the mini-batch assists in updating the actor-network by utilizing the policy gradient approach to maximize the expected return. The critic network approximates the gradient using the sampled actions and Q-values. In the last step, the target networks are updated in Step 13 to guarantee stability. In steps 14 and 15, the convergence criterion is checked, and the iteration (episode) is increased by 1, respectively.

### B. COMPLEXITY DISCUSSION

The DDPG agent utilizes a framework consisting of four neural networks. The initial two networks include the actor and critic components, respectively. These networks execute actions based on the current state and provide feedback outcomes about the performed action. The remaining components consist of the target actor and target critic networks, which are included to improve the agent's stability during the training process. Each network comprises four hidden layers, with the number of hidden nodes in each layer corresponding to the number of VLC IRS mirrors and the number of mmWave IRS reflecting elements. The actor-network is characterized by four inputs, which correspond to the positions of Bob and Eve on the XY plane. The

**TABLE 1.** Simulation Parameters.

| Parameter | Value |
|---|---|
| $P_V^{max}$ | 2 W |
| $\sigma_V^2$ | $-100$ dBm/MHz |
| $W_V$ | 2 MHz |
| $\Psi_{fov}$ | $45°$ |
| $\Psi_{1/2}$ | $70°$ |
| $A_{pd}$ | $10^{-4}$ |
| $\lambda$ | 0.4 |
| $H_f$ | 1 |
| $n_c$ | 1.5 |
| $\kappa$ | 1 |
| $\nu$ | 1.6 |
| $d_0$ | 1 Meter |
| $\Gamma$ | 100 |
| $\gamma_L$ | 0.9 |
| $\tau$ | 0.05 |

expression $4 + 4 \times N_m \times N_n + 1$ determines the output size of the network, with $N_m$ and $N_n$ representing the number of mirrors. This output contains many components, including the VLC beamforming weight vector, mmWave beamforming weight vector, mirror yaw angle, mirror roll angle, mirror sheet yaw angle, phase shift, and power ratio. The numerical findings section discusses the impact of varying the number of mirrors on SC, power ratio, and average rate.

## V. PERFORMANCE EVALUATION

This section presents a thorough performance evaluation of the proposed DRL-based approach. We begin with a detailed description of the experimental setup and proceed to an in-depth analysis of the numerical results obtained.

### A. EXPERIMENTAL SETUP

In this simulation, we examine an enclosed environment characterized by a room measuring 10 meters by 10 meters. In this context, we consider the presence of a mmWave IRS sheet and a square-shaped VLC IRS mirror sheet, both positioned 2 meters above the ground. The dimensions of the mirror sheet are each half a meter in length and width, while the number of IRS-assisted mmWave is varied in our simulations. Specifically, we use 8, 16, 32, and 64.

Regarding the VLC connection, the origin $(0, 0)$ is established as the reference point corresponding to both the center of the horizontally positioned mirror array sheet and the center of the room. The strategic placement of four light fixtures in the room's corners enhances the implementation of VLC IRS to maximize the capacity for maintaining confidentiality. The vectors $x_s = [-1, 1, -1, 1]$ and $y_s = [-1, -1, 1, 1]$ represent the location of the fixtures, and each fixture is situated at a height of three meters from the floor. On the other hand, the mmWave MISO antenna sits at the central location of the room, marked by the Cartesian coordinates $x, y, z = [0, 0, 3]$. As shown in Fig. 1, the mmWave IRS reflecting sheet is placed on the right wall of the room.

Table 1 presents the DDPG agent and the system's simulation parameters. Throughout the process of network training, the computational resources utilized include an i7-11800H Central Processing Unit (CPU), 32 gigabytes of Random Access Memory (RAM), and an RTX 3080 Mobile Graphics Processing Unit (GPU) with a thermal design power of 130 watts.

During this DDPG simulation, we conducted 1000 episodes to train the system simulation, each consisting of 1000 iterations. The buffer size is assumed to be 1000000, and the batch size is set to 10. We set the learning rate for the actor-critic models at 0.0001 and 0.0002, respectively. The target networks update at a rate of $\tau = 0.0001$. The discount factor used for future rewards is set to 0.25.

The action noise is classified as AWGN, characterized by a variance of $\sigma^2 = N_0 B$ and a mean of zero. The Gaussian noise standard deviation has an exact value of 0.01.

Throughout the training phase, users exhibit unrestricted movement inside the room, with their speed being randomly dispersed between 1 and 2 m/s according to a uniform distribution. Both users follow the random waypoint model, as seen in Fig. 2, which is a commonly employed approach for simulating mobility in indoor settings [36].
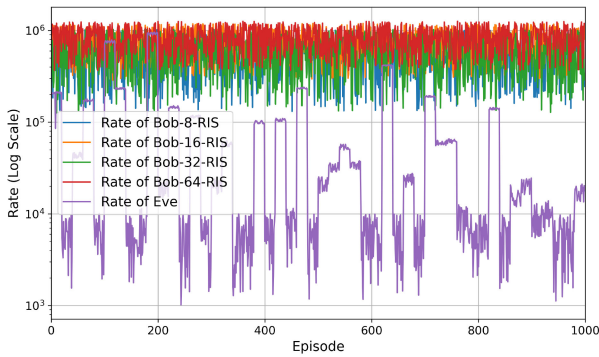
### B. NUMERICAL RESULTS

We simulate the SC of the system in three different cases. First, we demonstrate the use of only the VLC link; second, we use only the mmWave link; and in the third case, we integrate both links within the system. For all cases described in this work, we execute the simulation for 1000 episodes, and we consider the NLoS links only between transmitter and receiver.
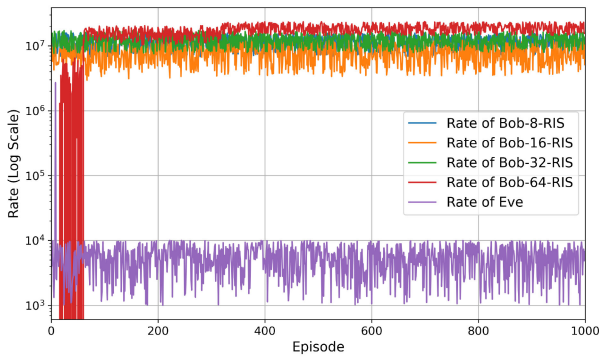
We first examine the system's secrecy capacity when deploying the VLC link only. In this scenario, the mmWave connection is inactive, and the system's communication is solely carried out via the VLC connection. Fig. 7(a) demonstrates the presence of fluctuations in the VLC link's rate. These fluctuations occur because the link is susceptible to interruption by obstacles or bodies, resulting in message delivery failures.

Furthermore, Fig. 7(b) illustrates the system's SC performance while only utilizing the mmWave link. In this scenario, we disregard the VLC link, resulting in improved and more consistent system performance. This is due to the constant availability of the mmWave link, which ensures that the system's SC level is maintained with fewer breaches. The eavesdropper exhibits a very low and constant rate compared to Bob, as seen in Fig. 7(b). In addition, increasing the number of mmWave IRS elements enhanced the secrecy rate of Bob's link within the system.

Fig. 11 illustrates the third scenario, which integrates both connections, mmWave and VLC, to improve the proposed system's secrecy capacity. In this scenario, we investigate the impact of setting the VLC IRS mirrors to a fixed value of 49 ($7 \times 7$ mirrors) while adjusting the mmWave IRS element number from 8, 16, 32 to 64. Compared to VLC and RF-only scenarios, we can improve the secrecy rate of the hybrid system. Eve has a very low and fluctuating rate,

(a) SC of the system using VLC link only
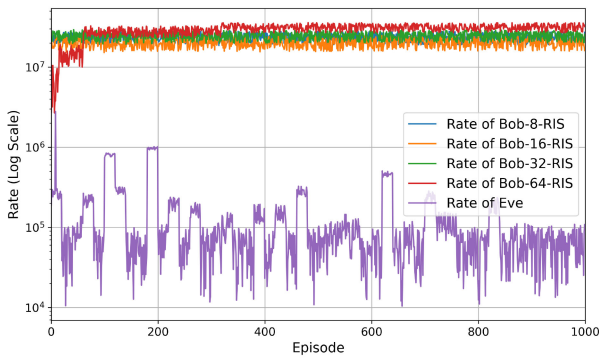


(b) SC of the system using mmWave link only



**FIGURE 7.** Comparison of the system's SC using log scale using (a) VLC link only, (b) mmWave link only, and (c) VLC and mmWave links.



**FIGURE 8.** Power Ratio vs Episodes for Different mmWave IRS elements and 49 VLC IRS mirrors.



**FIGURE 9.** Average Rates for Bob and Eve for Different mmWave IRS elements and 49 VLC IRS mirrors.



**FIGURE 10.** Summed Rate of Bob and Eve for Different mmWave IRS elements.

whereas Bob's rate performance is improving as we increase the number of reflective elements at the IRS.

To further evaluate the effectiveness of the hybrid system, we simulated several parameters to study its performance and confirm its efficacy. The figures below illustrate the relationship between episodes, power ratio, number of IRS elements, and SC. The same parameters previously specified are applied for the subsequent simulations (49 VLC IRS mirrors and varying mmWave IRS elements). We chose this mirror arrangement about our previous work in [39], where the results of this paper indicate that using an IRS sheet with 49 mirrors is the optimal configuration for maximizing the SC of the VLC network.
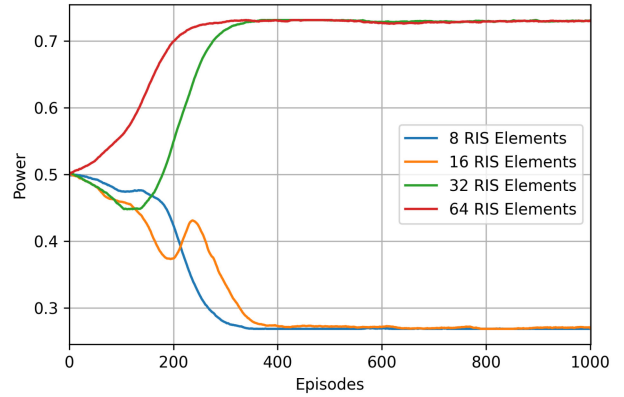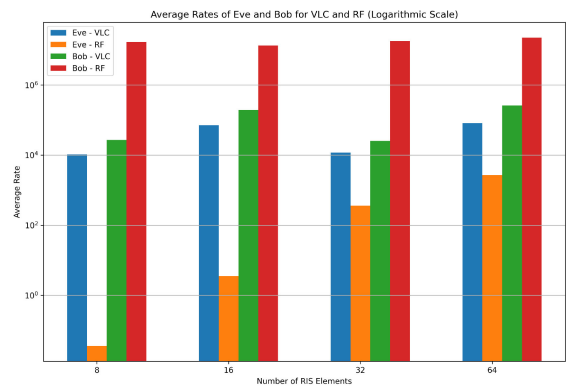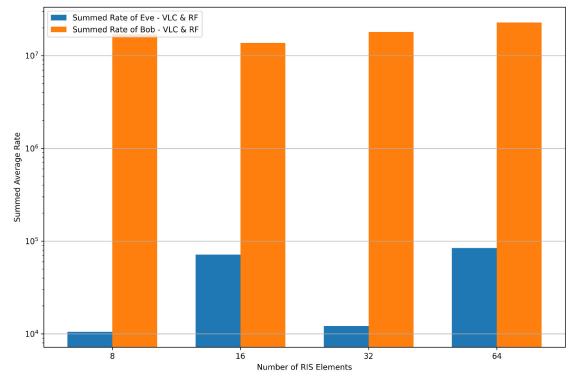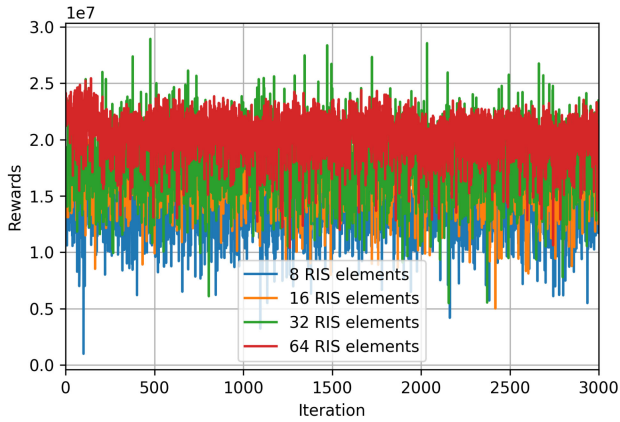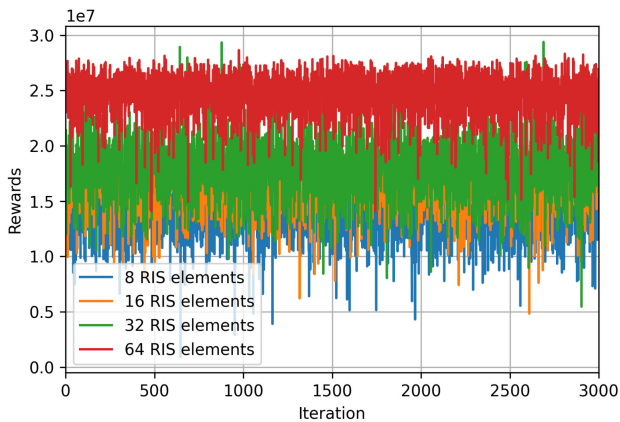
In Fig. 8, we examine the power ratio of the system. The results indicate that employing 8 and 16 mmWave IRS elements resulted in greater power allocation towards the RF mmWave connection for data transmission. However, an IRS with 32 and 64 elements will allocate a larger power ratio for transmission to the VLC link. This is because these configurations receive more mmWave signals, making it advantageous to provide more power to VLC. Fig. 9 illustrates the average rates of Bob and Eve using both links, VLC and mmWave. Empirical evidence demonstrates that

(a) Iteration 100000 to 103000 (100 to 103 Episodes)



(b) Iteration 700000 to 703000 (700 to 703 Episodes)



(c) Iteration 980000 to 983000 (980 to 983 Episodes)

**FIGURE 11.** Convergence of Rewards within Episodes for episodes (a) 100 to 103 Episodes, (b) 700 to 703 Episodes, and (c) 980 to 983 Episodes.

Bob surpasses Eve in both connections, thus guaranteeing the effectiveness of the suggested approach in mitigating eavesdropping attacks.

Fig. 10 illustrates the combined average rates of Eve and Bob, achieved by employing various mirror configurations. The results indicate a significant improvement in Bob's rate performance compared to Eve across all mmWave IRS
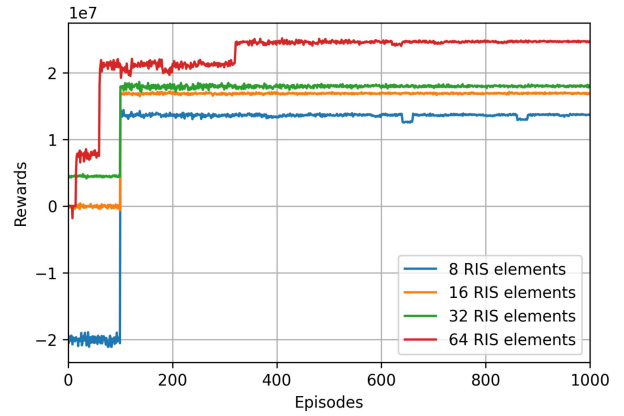


**FIGURE 12.** Rewards vs Episodes for Different mmWave IRS elements and 49 VLC IRS mirrors.

configurations. This further confirms the successful execution of the algorithm in preventing unauthorized interception attempts.

During the training process of the model, we record the outcomes at predefined intervals of the iterations. Figures 11(a), 11(b), and 11(c) display the convergence charts of the rewards (secrecy capacity) for three thousand iterations, specifically from iteration 100000 to 103000, 700000 to 703000, and 980000 to 983000, respectively. Further iterations of model training enhance the convergence of rewards, as demonstrated by the simulation results. This could certainly be seen when specifically comparing Figs. 11(a) and 11(c). Training with more iterations enhances the learning process, improving actions and feedback and leading to more consistent and improved rewards.

Figure 12 analyzes the average reward, which describes the total secrecy capacity of the system. In this outcome, we train the model for 1000 episodes; each episode consists of 1000 iterations. We exclude the first 100 episodes of the training time, as they are disregarded in the learning process. Following that, the model achieves convergence of average reward values for all mirror numbers.

The same figure also demonstrates how changing the number of mmWave reflective elements affects the SC. As shown in Fig. 12, increasing the number of mmWave IRS elements from 8 to 64 provides greater control over the direction of the signal, leading to an enhancement in the system's secrecy capacity.

## VI. CONCLUSION
In this paper, we have introduced a new DRL approach to optimize the SC in a hybrid VLC and mmWave system with the assistance of the IRS. The DDPG-based agent maintained control over the beamforming weights of the light fixtures and mmWave access point, the angles of each mirror, the orientations of the mirrors on the mirror array sheet, and the phase shift of the mmWave IRS. Controlling the listed parameters is critical to maximizing the hybrid system's SC. The simulation results demonstrate that the

proposed method based on DDPG is effective in addressing the obstacles and intricacies of the system. These issues encompass the mobility of users, the high dimensionality of parameters, and the overall complexity of the scenario. During the simulation, we investigated the impact of varying the number of reflective elements in the mmWave IRS sheet on the total SC, power allocation, and average rate. The experimental findings indicated that utilizing 64 elements in the mmWave IRS sheet yielded an optimization of the system's performance with respect to SC, especially in the mmWave link.

## ACKNOWLEDGMENT

## REFERENCES

[1] H. Haas, L. Yin, Y. Wang, and C. Chen, "What is LiFi?" *J. Lightw. Technol.*, vol. 34, no. 6, pp. 1533–1544, Jan. 12, 2016.

[2] A. Jovicic, J. Li, and T. Richardson, "Visible light communication: Opportunities, challenges and the path to market," *IEEE Commun. Mag.*, vol. 51, no. 12, pp. 26–32, Dec. 2013.

[3] S. A. Busari, K. M. S. Huq, S. Mumtaz, L. Dai, and J. Rodriguez, "Millimeter-wave massive MIMO communication for future wireless systems: A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 2, pp. 836–869, 2nd Quart., 2018.

[4] J. Al-Khori, G. Nauryzbayev, M. Abdallah, and M. Hamdi, "Secrecy capacity of hybrid RF/VLC DF relaying networks with jamming," in *Proc. Int. Conf. Comput. Netw. Commun. (ICNC)*, 2019, pp. 67–72.

[5] X. Li, H.-N. Dai, Q. Wang, M. Imran, D. Li, and M. Imran, "Securing Internet of Medical Things with friendly-jamming schemes," *Comput. Commun.*, vol. 160, pp. 431–442, Jul. 2020.

[6] Y. Zou and J. Zhu, *Physical-Layer Security for Cooperative Relay Networks*. Heidelberg, Germany: Springer, 2016.

[7] M. F. Marzban, M. Kashef, M. Abdallah, and M. Khairy, "Beamforming and power allocation for physical-layer security in hybrid RF/VLC wireless networks," in *Proc. 13th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, 2017, pp. 258–263.

[8] S. Cho, G. Chen, and J. Coon, "Securing visible light communication systems by beamforming in the presence of randomly distributed eavesdroppers," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2918–2931, May 2018.

[9] Z. Tang, T. Hou, Y. Liu, J. Zhang, and L. Hanzo, "Physical layer security of intelligent reflective surface aided NOMA networks," *IEEE Trans. Veh. Technol.*, vol. 71, no. 7, pp. 7821–7834, Jul. 2022.

[10] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, Jan. 2020.

[11] H. Abumarshoud, L. Mohjazi, O. A. Dobre, M. Di Renzo, M. A. Imran, and H. Haas, "LiFi through reconfigurable intelligent surfaces: A New Frontier for 6G?" *IEEE Veh. Technol. Mag.*, vol. 17, no. 1, pp. 37–46, Mar. 2022.

[12] 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, Dec. 2023.

[13] A. M. Abdelhady, A. K. S. Salem, O. Amin, B. Shihada, and M.-S. Alouini, "Visible light communications via intelligent reflecting surfaces: Metasurfaces vs mirror arrays," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 1–20, 2021.

[14] P. Wang, J. Fang, X. Yuan, Z. Chen, and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: Joint active and passive precoding design," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14960–14973, Dec. 2020.

[15] B. S. Ciftler, A. Alwarafy, and M. Abdallah, "Distributed DRL-based downlink power allocation for hybrid RF/VLC networks," *IEEE Photon. J.*, vol. 14, no. 3, Jun. 2022, Art. no. 8632510.

[16] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao, and Q. Wu, "Deep reinforcement learning-based intelligent reflecting surface for secure wireless communications," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 375–388, May 2020.

[17] S. Shafiee, N. Liu, and S. Ulukus, "Towards the secrecy capacity of the Gaussian MIMO wire-tap channel: The 2-2-1 channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4033–4039, Sep. 2009.

[18] S. Shafiee and S. Ulukus, "Achievable rates in gaussian MISO channels with secrecy constraints," in *Proc. IEEE Int. Symp. Inf. Theory*, 2007, pp. 2466–2470.

[19] A. Khisti and G. Womell, "Secure transmission with multiple antennas: The MIMOME wiretap channel, August 2008," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5515–5532, Dec. 2010.

[20] O. Ulgen, U. Ozmat, and E. Gunaydin, "Hybrid implementation of millimeter wave and visible light communications for 5G networks," in *Proc. IEEE 26th Telecommun. Forum (TELFOR)*, 2018, pp. 1–4.

[21] D. A. Basnayaka and H. Haas, "Hybrid RF and VLC systems: Improving user data rate performance of VLC systems," in *Proc. IEEE 81st Veh. Technol. Conf. (VTC Spring)*, 2015, pp. 1–5.

[22] D. A. Basnayaka and H. Haas, "Design and analysis of a hybrid radio frequency and visible light communication system," *IEEE Trans. Commun.*, vol. 65, no. 10, pp. 4334–4347, Oct. 2017.

[23] J. Al-Khori, G. Nauryzbayev, M. Abdallah, and M. Hamdi, "Physical layer security for hybrid RF/VLC DF relaying systems," in *Proc. IEEE 88th Veh. Technol. Conf. (VTC-Fall)*, 2018, pp. 1–6.

[24] J. Al-Khori, G. Nauryzbayev, M. M. Abdallah, and M. Hamdi, "Secrecy performance of decode-and-forward based hybrid RF/VLC relaying systems," *IEEE Access*, vol. 7, pp. 10844–10856, 2019.

[25] J. Al-khori, G. Nauryzbayev, M. Abdallah, and M. Hamdi, "Joint beamforming design and power minimization for friendly-jamming relaying hybrid RF/VLC systems," *IEEE Photon. J.*, vol. 11, no. 2, Apr. 2019, Art. no. 7902718.

[26] M. Cui, G. Zhang, and R. Zhang, "Secure wireless communication via intelligent reflecting surface," *IEEE Wireless Commun. Lett.*, vol. 8, no. 5, pp. 1410–1414, Oct. 2019.

[27] H. Shen, W. Xu, S. Gong, Z. He, and C. Zhao, "Secrecy rate maximization for intelligent reflecting surface assisted multi-antenna communications," *IEEE Commun. Lett.*, vol. 23, no. 9, pp. 1488–1492, Sep. 2019.

[28] J. Chen, Y.-C. Liang, Y. Pei, and H. Guo, "Intelligent reflecting surface: A programmable wireless environment for physical layer security," *IEEE Access*, vol. 7, pp. 82599–82612, 2019.

[29] Y. Qi and M. Vaezi, "IRS-assisted physical layer security in MIMO-NOMA networks," *IEEE Commun. Lett.*, vol. 27, no. 3, pp. 792–796, Mar. 2023.

[30] M. Kim and D. Park, "Intelligent reflecting surface-aided MIMO secrecy rate maximization," *ICT Exp.*, vol. 8, no. 4, pp. 518–524, 2022.

[31] L. Qian, X. Chi, L. Zhao, and A. Chaaban, "Secure visible light communications via intelligent reflecting surfaces," in *Proc. IEEE Int. Conf. Commun. (ICC)*, 2021, pp. 1–6.

[32] S. Aboagye, T. M. N. Ngatched, O. A. Dobre, and A. R. Ndjiongue, "Intelligent reflecting surface-aided indoor visible light communication systems," *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3913–3917, Dec. 2021.

[33] L. Xiao, G. Sheng, S. Liu, H. Dai, M. Peng, and J. Song, "Deep reinforcement learning-enabled secure visible light communication against eavesdropping," *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 6994–7005, Oct. 2019.

[34] H. Yang et al., "Intelligent reflecting surface assisted anti-jamming communications based on reinforcement learning," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, 2020, pp. 1–6.

[35] H. Yang et al., "Intelligent reflecting surface assisted anti-jamming communications: A fast reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1963–1974, Mar. 2021.

[36] C. Bettstetter, H. Hartenstein, and X. Pérez-Costa, "Stochastic properties of the random waypoint mobility model," *Wireless Netw.*, vol. 10, no. 5, pp. 555–567, 2004.

[37] J. Grubor, S. Randel, K.-D. Langer, and J. W. Walewski, "Broadband information broadcasting using LED-based interior lighting," *J. Lightw. Technol.*, vol. 26, no. 24, pp. 3883–3892, Dec. 2008.

[38] M. Plappert et al., "Parameter space noise for exploration," 2017, *arXiv:1706.01905*.

[39] D. A. Saifaldeen, B. S. Ciftler, M. M. Abdallah, and K. A. Qaraqe, "DRL-based IRS-assisted secure visible light communications," *IEEE Photon. J.*, vol. 14, no. 6, pp. 1–9, Dec. 2022.

**DANYA A. SAIFALDEEN** received the B.S. degree in electrical and computer engineering from Texas A&M University, Doha, Qatar, in 2014, the M.S. degree in wireless communications from the University of Southampton, U.K., in 2016, and the Ph.D. degree in computer science and engineering from Hamad Bin Khalifa University, Doha, in 2024. His current research interests include physical-layer security in visible light communications and mmWave networks.

**ABDULLATIF M. AL-BASEER** (Member, IEEE) received the M.Sc. degree in computer networks from the King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, in 2017 and the Ph.D. degree in computer science and engineering from Hamad Bin Khalifa University, Doha, Qatar, in 2022, where he is a Postdoctoral Research Fellow with the Smart Cities and IoT Lab. He also has six U.S. patents in the area of the wireless network edge. He has authored and co-authored over 30 conference and journal papers in IEEE ICC, IEEE Globecom, IEEE CCNC, IEEE WCNC, and IEEE Transactions. His current research interests include AI for networking, AI for cybersecurity, distributed AI, and edge LLMs.

**BEKIR S. CIFTLER** (Member, IEEE) received the B.S. degree from Middle East Technical University, Ankara, Turkey, in 2011, the M.S. degree from the TOBB University of Economics and Technology, Ankara, in 2013, and the Ph.D. degree in electrical and computer engineering from Florida International University, Miami, FL, USA, in 2017. His current research interests include wireless localization and tracking for Internet of Things, software-defined radios, 5G and mmWave networks, and MIMO systems.

**MOHAMED M. ABDALLAH** (Senior Member, IEEE) received the B.Sc. degree from Cairo University, Giza, Egypt, in 1996, and the M.Sc. and Ph.D. degrees from the University of Maryland, College Park, MD, USA, in 2001 and 2006, respectively. From 2006 to 2016, he held academic and research positions with Cairo University and Texas A&M University at Qatar, Doha, Qatar. He is currently a Founding Faculty Member with the rank of an Associate Professor with the College of Science and Engineering, Hamad Bin Khalifa University, Doha. He has published more than 150 journals and conferences and four book chapters, and co-invented four patents. His current research interests include wireless networks, wireless security, smart grids, optical wireless communication, and blockchain applications for emerging networks. He is a recipient of the Research Fellow Excellence Award with Texas A&M University at Qatar in 2016, the Best Paper Award in multiple IEEE conferences, including IEEE BlackSeaCom 2019, the IEEE First Workshop on *Smart Grid and Renewable Energy* in 2015, and the Nortel Networks Industrial Fellowship from 1999 to 2003. His professional activities include an Associate Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE OPEN ACCESS JOURNAL OF COMMUNICATIONS, the Track Co-Chair of the IEEE VTC Fall 2019 Conference, the Technical Program Chair of the 10th International Conference on Cognitive Radio-Oriented Wireless Networks, and the Technical Program Committee Member of several major IEEE Conferences.

**KHALID A. QARAQE** (Senior Member, IEEE) was born in Bethlehem. He received the B.S. (Hons.) degree in electrical engineering from the University of Technology, Bagdad, Iraq, in 1986, the M.S. degree in electrical engineering from the University of Jordan, Amman, Jordan, in 1989, and the Ph.D. degree in electrical engineering from Texas A&M University at College Station, College Station, TX, USA, in 1997. From 1989 to 2004, he has held a variety positions in many companies and he has over 12 years of experience in the telecommunication industry. He has worked on numerous projects and has experience in product development, design, deployments, testing, and integration. He joined the Department of Electrical and Computer Engineering, Texas A&M University at Qatar, in July 2004, where he is currently a Professor and the Managing Director with the Center for Remote Healthcare Technology. He has published more than 131 journal papers in top IEEE journals and presented 250 papers at prestigious international conferences. He has 20 book chapters published, four books, three patents, and presented several tutorials and talks. His research interests include communication theory and its application to design and performance, analysis of cellular systems and indoor communication systems. Particular interests are in mobile networks, broadband wireless access, cooperative networks, cognitive radio, diversity techniques, index modulation, visible light communication, FSO, tele-health, and noninvasive bio sensors. He has been awarded more than 20 research projects consisting of more than USD 13M from local industries, Qatar, and the Qatar National Research Foundation. He received the Itochu Professorship Award from 2013 to 2015, the Best Researcher Award from QNRF 2013, the Best Paper Award from the IEEE First Workshop on *Smart Grid and Renewable Energy* in 2015, the Best Paper Award from the IEEE Globecom 2014, the Best Poster Award IEEE Dyspan Conference in 2012, the TAMUQ Research Excellence Award in 2010, the Best Paper Award from ComNet in 2010, the Best Paper Award from CROWNCOM in 2009, the Best Paper Award from the ICSPC'0 in 2007, the IEEE SIGNAL PROCESSING MAGAZINE Best Column Award in 2018, the Best Oral Presentation Award from CECNet 2018, the Best Paper Award, Green Communications, Computing, and Technologies Conference in 2016, the Faculty of the Year (Student Award) in 2016 and 2015, and the Best Faculty Award by the Student Body in 2015.