DRL-Based IRS-Assisted Secure Visible Light Communications

Danya A. Saifaldeen, Bekir S. Ciftler ⓘ, Mohamed M. Abdallah ⓘ, and Khalid A. Qaraqe ⓘ, *Senior Member, IEEE*

Abstract—In this paper, we develop a novel Physical Layer Security (PLS) technique for a Visible Light Communication (VLC) system composed of light fixtures assisted by mirror array sheets serving as Intelligent Reflecting Surfaces (IRS). Our objective is to optimize the Secrecy Capacity (SC) by finding the optimal beamforming (BF) weights equipped at the VLC fixtures and mirror orientations at the mirror array sheet. Due to many design parameters, including the beamforming weights, the mirror orientations, and the mobility of the users, conventional optimization techniques may not be practical to optimize the SC capacity. Therefore, we proposed a Deep Reinforcement Learning (DRL) solution based on Deep Deterministic Policy Gradient (DDPG) algorithm to solve the highly complex SC problem by adjusting the BF weights and mirror orientations. The DDPG-based algorithm provides an optimized solution that can adapt to the large size of design parameters and act fast to the channel variations due to users' mobility. Our results show that considering both mirror array sheet and BF vectors provide the highest SC for the system. Moreover, we show the effect of changing the mirror arrangements of the mirror array sheet on SC. We conclude that for a fixed mirror array sheet size, there exists a specific mirror arrangement (i.e., number of mirrors) that optimizes the SC. After this number, the performance of SC saturates. We also show the trade-off between the training complexity and SC performance considering different mirror arrangements in the mirror array sheet.

Index Terms—Deep reinforcement learning, intelligent reflecting surfaces, physical layer security, secrecy capacity, visible light communications.

I. INTRODUCTION

THE advantages of Visible Light Communication (VLC) have positioned the technology to be an appealing complementary for Radio Frequency (RF) to support sixth-generation (6G) networks. VLC can be used for both illumination, and high-speed communication purposes in a broad, free and unlicensed spectrum [1]. Additionally, VLC is considered a cost and energy-efficient technology that provides a high data rate and reliable communication demanded by users nowadays [2]. Despite all the significant advantages mentioned above, VLC has a broadcasting nature that makes communication vulnerable to security threats. Eavesdropping and jamming attacks are common threats in VLC networks, especially when deploying them in public areas such as hospitals, schools, airports, malls, and similar indoor environments [3].

The basic principle of Physical Layer Security (PLS) is to avoid unauthorized listeners (eavesdropper) from obtaining communication data transmitted between transmitter and receiver by exploiting the channel characteristics and signal processing methods [4]. The performance of the PLS can be measured by the Secrecy Capacity (SC), which is defined as the difference between the capacity of the intended user and the capacity of the eavesdropper. The SC can be maximized by PLS techniques, such as artificial noise-aided security and utilization of beamforming (BF) vectors [4], [5]. Among these techniques, transmit BF has been considered a valuable tool for PLS in both RF and VLC systems. Beamforming is defined as directing the transmitted signal towards an intended receiver device [5]. In VLC systems, BF is utilized by concentrating the light in a specific direction or area towards an intended receiver [6], [7].

In addition to BF, Intelligent Reflecting Surfaces (IRS) have been used to improve the PLS performance in RF systems [8]. IRS is a planar surface consisting of a massive number of reflecting elements that are low in cost, which can use intelligently reconfigure the propagation of signals in a wireless environment [9]. Each element has the ability to adjust the reflection amplitude and the phase of each incident signal independently, thus, achieving a high signal-to-noise ratio at the intended receiver [9]. In VLC systems, mirror array sheets are used as IRS to support the communication links [10]. The mirror array sheet consists of multiple mirrors in an arrangement, and the angular orientations of each mirror are critical to focus the light beam properly. Every individual mirror has two angles, the yaw, and the roll angles. As the number of elements increases in the mirror array sheet, controlling the direction of the beam becomes more precise at the price of higher dimensionality and complexity. The conventional methods are not able to cope with these challenges.

In this paper, we consider a VLC system comprising of a single access point (AP) with multiple fixtures equipped with BF and mirror array sheet as IRS. We formulate an optimization problem using BF vectors and mirror array sheet orientations to maximize the SC of the intended user. The formulated optimization problem is non-convex, high-dimensional, and dynamic due to the

Danya A. Saifaldeen, Bekir S. Ciftler, and Mohamed M. Abdallah are with the Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Doha 34110, Qatar (e-mail: dsaifaldeen@hbku.edu.qa; bciftler@hbku.edu.qa; moabdallah@hbku.edu.qa).

Khalid A. Qaraqe is with the Department of Electrical and Computer Engineering, Texas A&M University at Qatar, Doha 34110, Qatar (e-mail: khalid.qaraqe@qatar.tamu.edu).

mobility of the users and the large number of control parameters in IRS-aided VLC systems with BF. Therefore, the conventional optimization techniques will likely fail in finding the optimal value of all these parameters to maximize the SC. On the other hand, Deep Reinforcement Learning (DRL) based techniques address the problem of high dimensionality and complexity due to their model-free nature [11], [12]. Especially, Deep Deterministic Policy Gradient (DDPG), which is very successful in precisely optimizing the parameters using its continuous action space. In this work, we propose a DDPG-based algorithm for jointly optimizing the BF vectors and the mirror orientations to mitigate the existing security issues. We design and implement a centralized DDPG agent at the VLC AP to jointly control the BF weights at the fixtures and the mirror orientations at the mirror array sheet (i.e., yaw and roll angles of each mirror element), as well as the overall yaw angle of the mirror array sheet. To the best of our knowledge, there is no existing work considering DRL-based IRS-aided PLS for VLC systems. Our major contributions in this work are listed below:

- We develop a VLC system model with multiple fixtures for BF and mirror array sheet as IRS that reflects the effect of the mirror angle orientations on the SC.
- We design and apply a DRL-based algorithm to maximize the SC for an IRS-assisted VLC network equipped with BF capabilities with the existence of an eavesdropper.
- We design the reward and the penalty functions for the DRL-based agent to maximize the SC.
- We analyze the performance and the complexity of the DRL-based algorithm with different mirror arrangements and users' positions. The simulation results demonstrate the proposed algorithm's effectiveness and adaptability in maximizing the SC.

This paper is structured as follows. In Section II, we explore the literature to study the existing research work on the proposed system. Then, we provide and discuss the designed model for the IRS-assisted secure VLC system in Section III. Subsequently, we explain the DRL-based VLC BF and mirror arrangement algorithm for SC maximization in Section IV. In Section V, we investigate the performance of the proposed algorithm by providing numerical and simulation results. Finally, in Section VI, we present our future work and concluding remarks.

II. RELATED WORK

In this section, we provide a brief survey of PLS in RF and VLC systems. The focus on PLS has recently gained attention to secure various communication systems such as RF and VLC to mitigate eavesdropping and jamming attacks. In [13]–[15], A. Mostafa et al. used different techniques to apply PLS in VLC systems, such as zero-forcing BF, artificial noise generation, friendly jamming, and robust BF. The aforementioned methods were used to obtain the secrecy rate for single-input single-output (SISO) and multiple-input single-output (MISO) scenarios in VLC systems.

With a random location of eavesdroppers, S. Cho et al. analyzed the SC of a VLC system by developing an analogous approach to model eavesdroppers' location [7]. This approach

was used to develop a BF method that optimizes secrecy performance measures when the transmitter has knowledge of the eavesdropper's intensity measure [7].

PLS of hybrid RF/VLC relaying network scenario has been studied in [6], [16]–[19]. M. Marzban et al. have developed a minimum power allocation algorithm and a zero-forcing BF algorithm to maximize the SC while minimizing the consumed electrical power in the hybrid RF/VLC network [6]. In [16], the objective was to maximize the achievable SC by designing RF and VLC-based BF vectors while minimizing the power that satisfies the required SC with different eavesdropper's location. Also, J. Al-Khori et al. developed zero-forcing BF to maximize the SC and outage probability for non-cooperative and cooperative power saving for this relaying network model [17]. In [18] and [19], a friendly jamming technique was used as an eavesdropping-resilient solution to maximize the achievable SC in RF/VLC relaying network. To achieve this objective, a joint relay-jammer selection algorithm and BF vectors for RF and VLC models were used in the above works, respectively.

IRS has recently received significant attention from researchers to be used along with PLS to enhance the communication systems' security performance. In [20] and [21], an IRS-aided secure wireless communication system consisting of the multi-antenna access point (AP), a single-antenna user, and a single-antenna eavesdropper was considered. Their objective was to maximize the secrecy rate of the legitimate communication channel by optimizing the transmit BF vector and the phase elements at the IRS. Authors in both works obtained improved secrecy results by applying the alternating optimization (AO) technique. In addition, J. Chen et al. used the AO technique based on the path-following algorithm to optimize the beamformers at the BS and the reflecting coefficients at the IRS to maximize the secrecy rate while considering real-time constraints on the reflection coefficients [22].

To address the security problems that may arise in VLC indoor systems, a PLS technique was proposed with the help of IRS mirror array sheet for SISO VLC system to mitigate eavesdropper's attacks in [23]. The main objective of this work was to maximize the secrecy rate by finding the optimal combination of mirrors' orientations that maximizes the channel gain of a legitimate user and reduces the channel gain of an eavesdropper utilizing a Particle Swarm Optimization (PSO)-based method [23].

Recently, S. Aboagye et al. proposed a practical IRS-aided VLC system model considering randomly oriented receiver and randomly deployed blockers in the direct link between the transmitter and the receiver [24]. Their main contribution was to formulate an optimization problem based on a sine-cosine algorithm to configure the orientation of the IRS elements such that the achievable rate is maximized [24].

In the context of finding optimal solutions to real-time, complicated, and intractable problems in wireless communication systems, authors in [12] and [25] used Deep Reinforcement Learning methods to address the IRS effect in RF and VLC systems, respectively. For IRS-aided RF systems, H. Yang et al. used a DRL-based approach to optimize the BF matrix at the base station, and the reflecting BF matrix at the IRS in a multiple legitimate users and multiple eavesdroppers' environment [12].
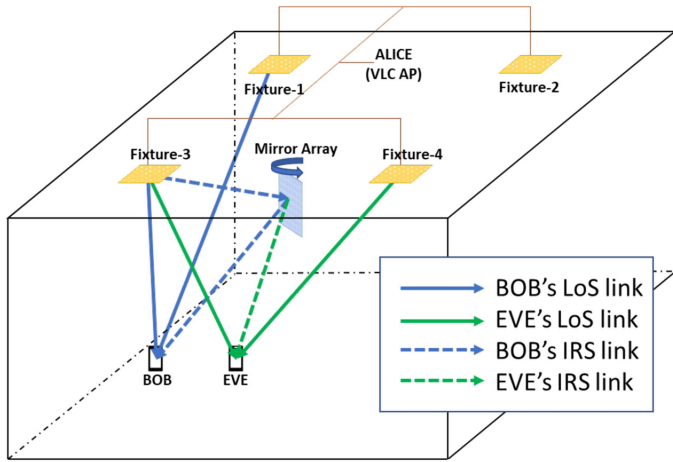
Fig. 1.　DRL-based IRS-Assisted Secure VLC.

Fig. 2.　Training phase for VLC.

The results proved that DRL significantly enhanced the maximization of the system's secrecy rate while guaranteeing the quality of service (QoS) requirements. Reinforcement Learning was also used in [26], [27] to mitigate jamming interference while using IRS. Results showed that adjusting the surface reflecting elements improved the IRS-assisted rate and increased the system's protection level. Liang Xiao et al. proposed a DRL-based intelligent BF framework along with a DRL-based MISO VLC BF algorithm to secure the communication system from eavesdropping attack and to solve the high continuous action space and high dimensional state space [25]. Their simulation results verified the efficiency of the proposed DRL scheme to increase the utility and the secrecy rate of the VLC system.

Considering the existing literature, we believe that there is no other work used DRL with mirror array sheets as IRS to optimize the secrecy capacity of a VLC systems. In this work, we design and propose a DRL-based agent to optimize the BF weights, the mirror array sheet yaw angle, and the mirror orientations to maximize the secrecy capacity of the VLC system given the location of the intended receiver and the eavesdropper.

III. SYSTEM MODEL AND PROBLEM FORMULATION

In this work, we consider an indoor environment with K VLC fixtures and a mirror array sheet as IRS at the center of the room, as shown in Fig. 1. The room dimensions are $x_r \times y_r \times z_r$. The system has two users: the intended receiver (i.e., Bob) and the eavesdropper (i.e., Eve). The goal of Eve is to wiretap the secret message of Bob transmitted by Alice using the broadcast nature of the VLC. The movement of the users is following a random waypoint model [28] with uniformly distributed speeds between 0 and 1 (m/s). At the beginning, each user has a random destination. Once the user reaches this destination, a new destination is randomly created within the room, as shown in Fig. 2.

There are four VLC fixtures of planar source with uniformly emitting radiance over the area. All fixtures are placed at the ceiling of the room, where each one is acting as a single VLC access point (AP) with BF capability. The BF is enabled with a
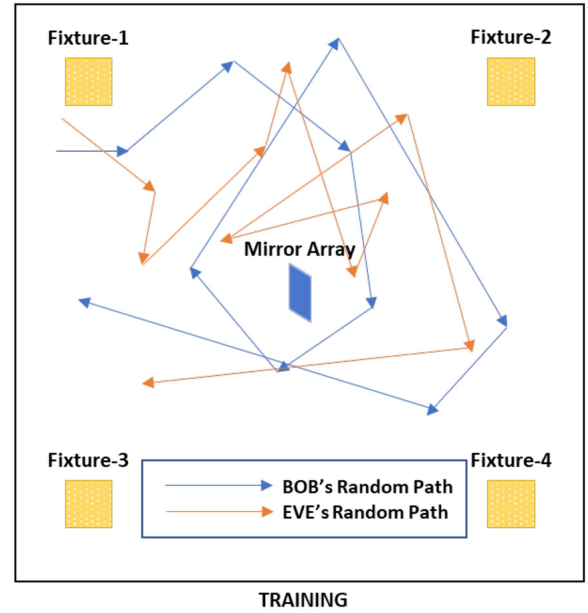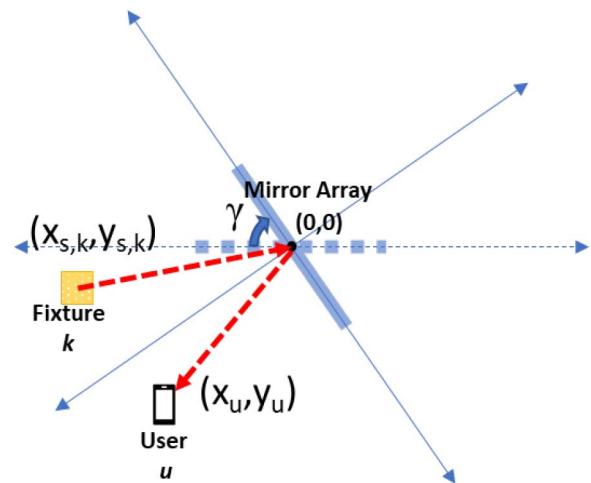

Fig. 3.　Top view of the system. The rotation of the mirror array sheet is represented by γ.

basic power controller, where total power is kept constant and the BF weights defines the share of power at fixtures. The mirror array sheet consists of $N_m \times N_n$ identical rectangular mirrors. The size of each mirror is $b_m \times l_m$, and each mirror can be rotated independently. Additionally, the mirror array sheet can rotate around its vertical axis to optimize its orientation and increase users' capacity. The rotation angle of each individual mirror (i, j) is represented in Fig. 3, where $\beta_{i,j}$ is the yaw angle, and $\alpha_{i,j}$ is the roll angle, while the rotation of the mirror array sheet is given by the yaw angle γ.

The received signal y^l at user l can be expressed as summation of the signals from all the fixtures as follows

$$y^l = \left(\sum_{k=1}^{K} w_k h_{l,k}^{\mathrm{LoS}} + \sum_{k=1}^{K} w_k h_{l,k}^{\mathrm{IRS}} \right) s + n^l, \qquad (1)$$

where, w_k is the BF weight for kth fixture, and $h_{l,k}^{\mathrm{LoS}}$ and $h_{l,k}^{\mathrm{IRS}}$ are the LoS link channel gain and the IRS link channel gain for the kth fixture indicator, respectively. The VLC noise of user $l \in \{\mathrm{Bob}, \mathrm{Eve}\}$ is represented as $n^l \sim \mathcal{N}(0, \sigma^2)$ due to shot and thermal noise, which is assumed to be Additive White Gaussian Noise (AWGN) with zero mean and $\sigma^2 = N_0 B$ variance, where $N_0 \approx 10^{-22} \mathrm{A}^2/\mathrm{Hz}$ [29]. The user might be within the coverage area of one or more fixtures depending on its position within the room.

The channel gain of the LoS link for user l is expressed as

$$h_{l,k}^{\mathrm{LoS}} = \int_{-b_s/2}^{b_s/2} \int_{-l_s/2}^{l_s/2} \frac{\eta(L_a + 1)\sigma\varpi T}{2\pi d_{l,k}^2(x,y)} g(\psi_{l,k}^{\mathrm{LoS,in}}(x,y))$$
$$\cos^{L_a}(\psi_{l,k}^{\mathrm{LoS,ir}}(x,y)) \cos(\psi_{l,k}^{\mathrm{LoS,in}}(x,y))dxdy, \quad (2)$$

where η is the conversion efficiency of the LEDs, $L_a = -\ln 2/\ln(\cos(\phi_{1/2}))$ denotes the order of Lambertian emission for the semi-angle at half illumination of LEDs ($\phi_{1/2}$). The physical area of the detector in a Photodetector (PD) is represented by σ, ϖ is the responsivity of a PD, T is the trans-impedance amplifier gain. The angle of incidence and irradiance between the fixture k and the user l is presented with $\psi_{l,k}^{\mathrm{LoS,in}}(x,y)$ and $\psi_{l,k}^{\mathrm{LoS,ir}}(x,y)$, respectively. For the sake of simplicity, we assume that all the receivers are facing upward vertically, hence $\psi_{l,k}^{\mathrm{LoS,in}}(x,y) = \psi_{l,k}^{\mathrm{LoS,ir}}(x,y)$. The function $g(\cdot)$ gives the optical concentrator gain as follows

$$g(\psi_{l,k}^{\mathrm{LoS,in}}(x,y)) = \begin{cases} a^2/\sin^2 \varphi_c, & \text{if} \psi_{l,k}^{\mathrm{LoS,in}}(x,y) \leq \varphi_c \\ 0, & \text{if} \psi_{l,k}^{\mathrm{LoS,in}}(x,y) > \varphi_c, \end{cases}$$
$$(3)$$

where a is the refractive index, and φ_c is half of the receiver field of view. The IRS link is the indirect channel between the fixture and the user via the reflections of the mirror array sheet as shown in Fig. 1. The channel gain of the IRS link for fixture k and user l considering the system model in Fig. 1 is expressed as

$$h_{l,k}^{\mathrm{IRS}}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma) = \sum_{i=1,j=1}^{N_m, N_n} \eta\varpi\sigma T E_{i,j}^{l,k}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma)$$
$$\times g(\theta_{\mathrm{R}_{i,j}}^{\mathrm{P}_{l,k}}) \cos(\theta_{\mathrm{R}_{i,j}}^{\mathrm{P}_{l,k}}), \quad (4)$$

where the irradiance at user l by mirror i, j is represented as $E_{i,j}^{l,k}(\boldsymbol{\alpha}, \boldsymbol{\beta})$. The mirror array sheet center is defined as the origin of the Cartesian coordinate system. The closed-form definition of $E_{i,j}^{l,k}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ can be found via geometrical analysis [23] as follows

$$E_{i,j}^{l,k}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma) = \frac{(L_a + 1)\rho}{2\pi} \int_{-h_m/2}^{h_m/2} \int_{-b_m/2}^{b_m/2} \cos^{L_a}(\theta_{\mathrm{R}_{i,j}}^{\mathbf{I}})$$
$$\times \frac{\mathbf{a}_3^T(\mathbf{P}_{l,k} - \mathbf{R}_{i,j})\widehat{\mathbf{N}}_{i,j}^T(\mathbf{P}_{l,k} - \mathbf{R}_{i,j})}{\|\mathbf{P}_{l,k} - \mathbf{R}_{i,j}\|_2^4}$$
$$\times \mathbb{1}\left(\mathbf{a}_1^T\mathbf{S}_k - \frac{b_s}{2} \leq \mathbf{a}_1^T\mathbf{I} \leq \mathbf{a}_1^T\mathbf{S}_k + \frac{b_s}{2},\right.$$

$$\left. \mathbf{a}_2^T\mathbf{S}_k - \frac{l_s}{2} \leq \mathbf{a}_2^T\mathbf{I} \leq \mathbf{a}_2^T\mathbf{S}_k + \frac{l_s}{2}\right) dx''dz''$$
$$(5)$$

where the mirror reflection efficiency is denoted by ρ, and $\mathbf{a}_i, i \in \{1, 2, 3\}$ is the ith column of 3×3 identity matrix. The functions $\|\cdot\|_2$ and $\mathbb{1}(\cdot)$ represents l_2-norm and the binary indicator, respectively. The binary indicator function determines whether the reflected image of the user l is within the field of view of fixture k using i, jth mirror of the mirror array sheet. The position of the user l with respect to the mirror center is defined as $\mathbf{P}_{l,k}$. The space coordinates of the center of VLC AP k is defined as $\mathbf{S}_k$ in (8).

$$\mathbf{P}_{l,k} = \begin{bmatrix} x_l^* - (x_{s,k}^* + \frac{b_m}{2} + (j-1)b_m) \\ y_l^* \\ h_l - (z_{s,k} + \frac{h_m}{2} + (j-1)h_m) \end{bmatrix} \quad (6)$$

$$\mathbf{R}_{\mathbf{i,j}} = \mathbb{R}_{i,j}^r[x''0z'']^T, \quad (7)$$

$$\mathbf{S}_k = \begin{bmatrix} -(x_{s,k}^* + \frac{b_m}{2} + (j-1)b_m) \\ y_{s,k}^* \\ -(z_{s,k} + \frac{h_m}{2} + (j-1)h_m) \end{bmatrix} \quad (8)$$

$$\mathbf{I} = \begin{bmatrix} \mathbf{a}_1^T\left(\mathbf{R}_{\mathbf{i,j}} + \frac{\mathbf{a}_3^T(\mathbf{S}_k - \mathbf{R}_{i,j})}{\mathbf{a}_3^T\widehat{\mathbf{R}_{\mathbf{i,j}}}}\widehat{\mathbf{R}_{\mathbf{i,j}}}\right) \\ \mathbf{a}_2^T\left(\mathbf{R}_{\mathbf{i,j}} + \frac{\mathbf{a}_3^T(\mathbf{S}_k - \mathbf{R}_{i,j})}{\mathbf{a}_3^T\widehat{\mathbf{R}_{\mathbf{i,j}}}}\widehat{\mathbf{R}_{\mathbf{i,j}}}\right) \\ \mathbf{a}_3^T\mathbf{S}_k \end{bmatrix} \quad (9)$$

where the Cartesian coordinates of user l are represented with x_l and y_l, and the position of kth fixture is represented with $x_{s,k}$ and $y_{s,k}$ with respect to the origin. The rotation of mirror i, j is represented with $\mathbb{R}_{i,j}^r$. When the mirror array sheet rotates around its vertical axis with γ angle, the distance of the user and the light fixture is constant, however, their cartesian coordinates changes as follows

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

$$y_l^* = \sqrt{x_l^2 + y_l^2} \sin\left(\gamma + \tan^{-1}(\tfrac{y_l}{x_l})\right) \quad (11)$$

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

$$y_{s,k}^* = \sqrt{x_{s,k}^2 + y_{s,k}^2} \sin\left(\gamma + \tan^{-1}(\tfrac{y_{s,k}}{x_{s,k}})\right). \quad (13)$$

The binary indicator function in (5) indicates if the reflected image of the user l is within the field of view of fixture k through the mirror i, j. The $N_{i,j}$ is the normal vector of the mirror i, j given the rotation angles $\alpha_{i,j}$ and $\beta_{i,j}$. The normal vector is denoted as follows

$$\widehat{\mathbf{N}}_{i,j} = \begin{bmatrix} \cos(\alpha_{i,j})\sin(\beta_{i,j}) \\ \cos(\alpha_{i,j})\cos(\beta_{i,j}) \\ \sin(\alpha_{i,j}) \end{bmatrix}. \quad (14)$$

In this work, we aim to use DRL-based agents to optimize the BF weights and the mirror orientations to maximize the SC of the VLC link. The SC of a receiver (Bob) considering an

eavesdropper (Eve) with known location is expressed as follows

$$C^{\text{sec}}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma, \boldsymbol{w}) = \frac{1}{2} \log \frac{6A^2 \sum_{k=1}^{K} w_k(h_{B,k}^{\text{LoS}} + h_{B,k}^{\text{IRS}}) + 3\pi e\sigma^2}{\pi eA^2 \sum_{k=1}^{K} w_k(h_{E,k}^{\text{LoS}} + h_{E,k}^{\text{IRS}}) + 3\pi e\sigma^2}$$

(15)

where $h_{B,k}^{\text{LoS}}$ and $h_{B,k}^{\text{IRS}}$ are the LoS channel gain and IRS channel gain for Bob from kth fixture, respectively, and $h_{E,k}^{\text{LoS}}$ and $h_{E,k}^{\text{IRS}}$ are the LoS channel gain and IRS channel gain for Eve from kth fixture respectively.

The optimization problem for maximizing the SC by adjusting the BF weights vector and the mirror orientations is denoted as follows:

$$\max_{\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma, \boldsymbol{w}} C^{\text{sec}}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \gamma, \boldsymbol{w})$$

$$\text{subject to(s.t.)} -\frac{\pi}{2} \leq \alpha_{i,j} \leq \frac{\pi}{2},$$

$$-\frac{\pi}{2} \leq \beta_{i,j} \leq \frac{\pi}{2},$$

$$-\frac{\pi}{2} \leq \gamma \leq \frac{\pi}{2},$$

$$0 \leq w_k \leq 1,$$

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

(16)

The optimization problem given in (16) is a non-convex problem with $N_m \times N_n$ mirror elements and K fixtures. There are $K + 2 \times N_m \times N_n + 1$ (i.e., BF vector, mirror yaw angles, mirror roll angles, and mirror array sheet yaw angle) parameters to be optimized, and the conventional methods cannot solve the problem in proper time. Additionally, the mobility of the users increases the complexity of the problem. Hence in the following section, we propose a novel DDPG-based solution to optimize the BF weights and mirror array sheet orientations within a reasonable time.

IV. DRL-BASED IRS-ASSISTED SECRECY CAPACITY MAXIMIZATION

In this section, we explain the DRL-based IRS-assisted SC maximization for VLC networks. In the proposed algorithm, DRL-based agent controls the BF vector for VLC fixtures, the mirror array sheet yaw angle, and the yaw and the roll angles of individual mirrors. The agent has the information of the users' locations (i.e., the positions of Bob and Eve) and it defines its policy based on this information. Hence, the state space ($\mathbf{s}_t$) of the DRL-agent is defined as follows

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

(17)

where $X_B = [x_B, y_B, z_B]$ represents the location of Bob and $X_B = [x_B, y_B, z_B]$ is the location of Eve.

The agent's actions consist of the BF vector for VLC fixtures, the mirror array sheet yaw angle, and the yaw and the roll angles of every mirror. We define the action space with a policy function with correlated noise utilizing Ornstein-Uhlenbeck (OU)

Algorithm 1: Training of DDPG-Based IRS-Assisted SC Maximization for VLC Networks.

1: **Initialization:** Set $t = 0$ and initialize replay buffer of agents $\mathcal{D}$.
2: Initialize the randomized weights of actor network θ^μ and critic network θ^Q.
3: Initialize the target networks using actor and critic networks: $\theta^{\mu'} \leftarrow \theta^\mu$ and $\theta^{Q'} \leftarrow \theta^Q$.
4: **for** $t = 1$ to ∞ **do**
5:     Observe state $\mathbf{s}_t$ (positions of Bob and Eve) and determine an action (BF vector, mirror array sheet yaw angle, and mirrors' yaw and roll angles) for agent $\mathbf{a}_t = \mu(\mathbf{s}_t|\theta^\mu) + \mathbf{n}_t$
6:     Execute all actions $\mathbf{a}_t$.
7:     Receive the reward r_t, and observe next state s_{t+1}, store transition $(\mathbf{s}_t, \mathbf{a}_t, r_t, \mathbf{s}_{t+1})$ in D.
8:     Randomly sample mini-batch transitions from $\mathcal{D}$: $B = \{(\mathbf{s}_i, \mathbf{a}_i, r_i, \mathbf{s}_{i+1})\}$.
9:     Compute the targets for actor and critic networks: $\tilde{Q}(\mathbf{s}_i, \mathbf{a}_i|\theta^Q) = r_i + \gamma_L Q(\mathbf{s}_{i+1}, \mu(\mathbf{s}_i|\theta^{\mu'})|\theta^{Q'})$
10:     Update the θ^Q in critic network by minimizing the loss: $L = \frac{1}{|B|} \sum_{i=1}^{|B|} (\tilde{Q}(\mathbf{s}_i, \mathbf{a}_i|\theta^{Q'}) - Q(\mathbf{s}_i, \mathbf{a}_i|\theta^Q))^2$
11:     Update the θ^μ in actor network according to the sampled policy gradient: $\nabla_{\theta_\mu} \boldsymbol{J} \approx \frac{1}{|B|} \sum_{i=1}^{|B|} \nabla_a Q(\mathbf{s}_i, \mathbf{a}_i|\theta^Q) \nabla_{\theta_\mu} \mu(\mathbf{s}_i|\theta^\mu)$
12:     Update the target networks: $\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'}$ $\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'}$
13: **end for**

---

process as follows

$$\boldsymbol{a}_t = \boldsymbol{\mu}(\mathbf{s}_t|\theta^\mu) + \boldsymbol{n}_t$$

$$= [\boldsymbol{a}_t^{\text{BF}} a_t^\gamma \boldsymbol{a}_t^{\boldsymbol{\alpha}} \boldsymbol{a}_t^{\boldsymbol{\beta}}],$$

(18)

where $\boldsymbol{\mu}$ represents the policy function of the DRL agent, θ is the parameters of the policy function (i.e., neural network weights), and $\boldsymbol{n}$ is the correlated action noise based on OU process [30]. The trade-off between the exploration and the exploitation in our DRL agent is determined by OU process parameters which are provided in Section V. The action of the BF vector is defined as $\boldsymbol{a}_t^{BF}$, with a vector of length K. The mirror array sheet yaw angle action is a scalar value represented by a_t^γ, while actions for the yaw and the roll angles of each individual mirror are defined as $\boldsymbol{a}_t^{\boldsymbol{\alpha}}$ and $\boldsymbol{a}_t^{\boldsymbol{\beta}}$, respectively. The size of each vector is equal to the number of mirrors ($N_m \times N_n$).

The objective function of the maximization problem defined in (16) aims to achieve the highest SC possible as in (15) by applying the optimal policy given the constraints such as BF weights and angle limits. Hence the reward function is defined as the SC

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

(19)

where the penalty function P_t is added to mitigate the problem of having zero SC, and it is given as

$$P_t = \begin{cases} 0, & \text{if } C_t^{\text{sec}} > 0 \\ \Gamma, & \text{if } C_t^{\text{sec}} = 0 \end{cases}, \qquad (20)$$

where the penalty value is Γ.

A. DDPG Algorithm

The DDPG-Based training to maximize the SC of the IRS-assisted VLC network is provided by Algorithm 1. The capacity of the replay buffer is M, and it is empty at the beginning of the training. We generate the actor and the critic network weights randomly in Step 3 as the final initialization step. For every iteration t, Steps from 5 to 12 of the algorithm are repeated during the whole training process. At Step 5, we observe the environment state (i.e., the positions of Bob and Eve) and use the policy function to determine the action set. We apply the actions in Step 6. Following that, we compute the reward in Step 7, and observe the next step. In Step 8, a mini-batch of transitions is randomly sampled from the replay buffer. We compute the targets for the actor and critic networks in Steps 9, and the critic network is updated in Steps 10. Then we update the actor-network with the policy gradient method using the mini-batch in Step 11. Finally, we update the target networks to ensure stability in Step 12.

B. Complexity Discussion

The DDPG agent is based on four neural networks. The first two are the actor and the critic networks, which are used for acting on the states and providing feedback results from the action. The rest are the target actor and the target critic networks, which are included to increase the agent's stability while training. In each network, there are two hidden layers, and in each hidden layer there are hidden nodes equal to the number of mirrors in the mirror array sheet. The actor network has 4 inputs, (i.e., the locations of Bob and Eve on XY-plane), and the output size is $4 + 2 \times N_m \times N_n + 1$ (i.e., BF vector, mirror yaw angle, mirror roll angle, and mirror array sheet yaw angle). In our numerical results section, we show the effect of changing the total number of mirrors (i.e, mirror arrangements) in terms of SC and time complexity.

V. NUMERICAL RESULTS

In this section, we present the numerical results to show the SC performance of our proposed system model and benchmark it against PSO-based algorithm [23]. First, we will simulate the algorithm under three various link scenarios The first scenario only uses light fixtures with BF capabilities (BF only). The second scenario only uses the mirror array sheets as IRS, where LoS link from light fixtures does not exist (mirror array sheets only). Whereas the last scenario combines both BF and mirror array sheets links.

Our simulation setup comprises a square room with 10 meters width and length. In this room, we consider a square-shaped mirror array sheet, with each side having a length of 0.5 meters

TABLE I
SIMULATION PARAMETERS

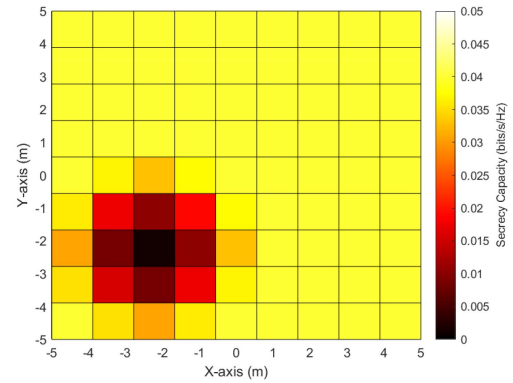| Parameter	Value
P_{V}^{max}	2 W
σ_{V}^2	-100 dBm/MHz
W_{V}	2 MHz
Ψ_{fov}	$45°$
$\Psi_{1/2}$	$70°$
A_{pd}	10^{-4}
λ	0.4
H_f	1
n_c	1.5
κ	1
ν	1.6
d_0	1 Meter
Γ	100
γ_L	0.9
τ	0.05

Fig. 4.   SC for considering the BF-only case (i.e., without IRS).

and is hanging 2 meters above the ground. We consider the origin (0,0) to be the coordinates of the center of the room and the mirror array sheet that is horizontally located. At the ceiling, there are four VLC fixtures located at the four quarters of the room, thus $\boldsymbol{x}_s = [-1, 1, -1, 1]$ and $\boldsymbol{y}_s = [-1, -1, 1, 1]$, to make IRS efficient for optimizing secrecy capacity. All the fixtures are located 3 meters above the ground.

The simulation parameters for the system and the DDPG agent are provided in Table I. We use i7-11800H CPU, 32 GB RAM, and RTX 3080 Mobile (with Thermal Design Power of 130 W) at maximum performance for training the networks. The system is simulated for 1000 epochs with 100 iterations in each epoch for training. The trained model was tested in 1000 Monte Carlo experiments with 100 iterations. During the training phase, users are freely moving within the room at a random speed uniformly distributed between 1 to 2 m/s. Both users are following a random waypoint model as shown in Fig. 2, which is widely used for mobility in indoor scenarios [28].

The 3D mesh graphs represented in Figs. 4, 5 and 6 demonstrate the Secrecy Capacity of Bob in the case of using 2×2 mirror arrangement for three different scenarios: BF only, mirror array sheets only, and BF with mirror array sheets, respectively. In all figures, we simulate the system's performance when the position of Bob is fixed at $(x_B = -2.8, y_B = -2.8)$, while Eve
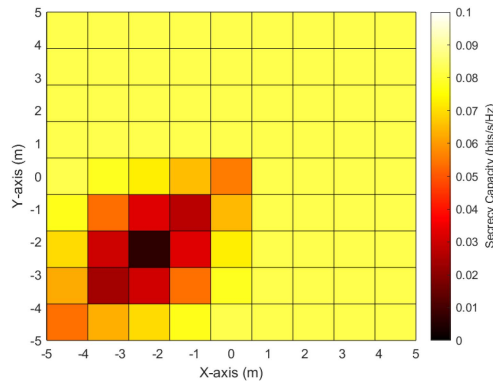
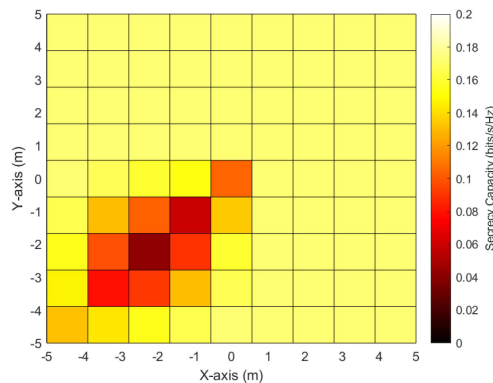Fig. 5. SC for considering IRS with blocked LoS from fixtures.

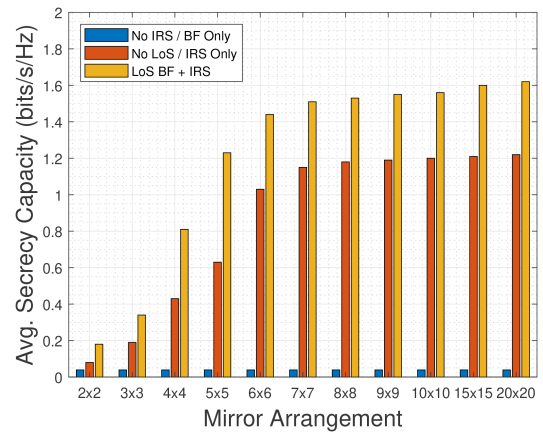Fig. 6. SC for considering when IRS and BF combined.

Fig. 7. Average SC for various mirror arrangements with a constant mirror array sheet size for LoS and NLoS cases.
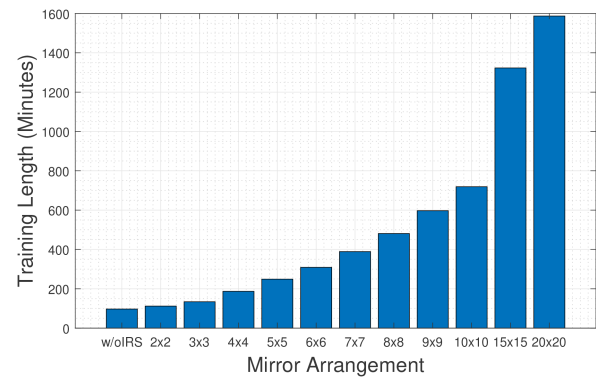

Fig. 8. Training length for various mirror arrangements with constant total size.
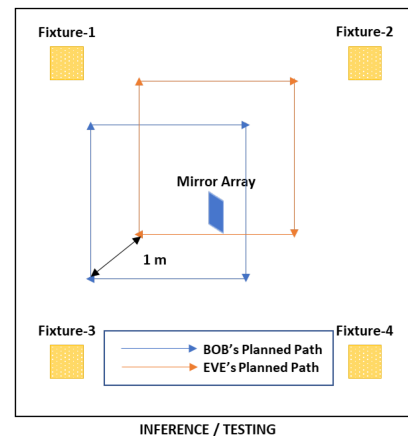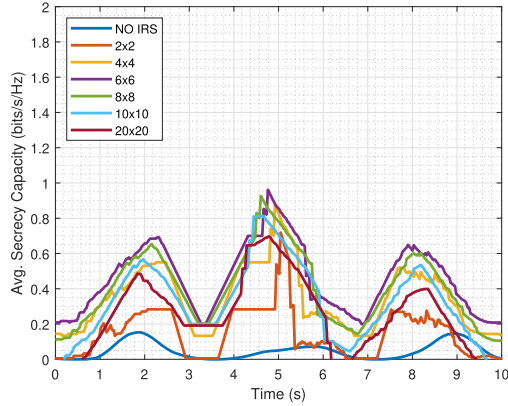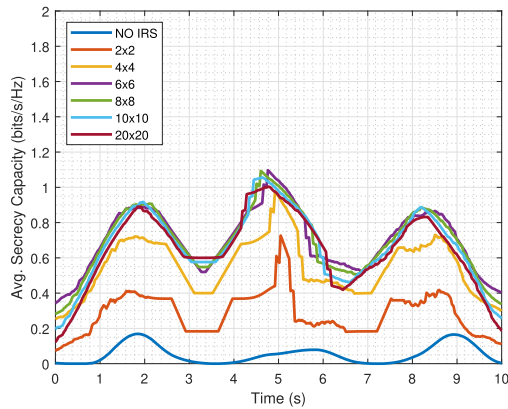

Fig. 9. Testing path is square shaped.

is moving around the room. The location of Eve is represented by the X and Y axes.

The baseline simulation of our system model is represented in Fig. 4. In this case, only light fixtures with BF capabilities exist; therefore, users can only communicate over LoS links. It can be shown from the graph that SC, in this case, cannot exceed 0.05 bits/sec/Hz. We represent the second scenario in Fig. 5 where the mirror array sheet is added as IRS to the system. In this scenario, the LoS link between the users and the fixtures is blocked, so the transmission is only available on the IRS link. The graph shows a significant performance advantage in terms of SC, which reaches up to 0.1 bits/sec/Hz in this case. The third scenario is represented in Fig. 6, where both LoS and IRS links are available, while considering optimized mirror orientations and BF weights to increase the SC. The graph shows a significant performance advantage in terms of SC, which reaches 0.18 bits/sec/Hz in this case. On the other hand, it can be shown from all figures that the SC drops to the lowest value when Eve is at the closest location to Bob (i.e., $x_E = -3, y_E = -3$) and improves significantly when Eve moves away from Bob.

We present the average SC for different mirror arrangements in Fig. 7. In this figure, both users move with the random waypoint model during training and testing. As the number of mirrors increases from 2×2 to 7×7, the average SC increases rapidly, after which it reaches saturation. Using a larger number of mirrors provides more control over the direction of the beam, hence, increases the overall SC of the network. However, this
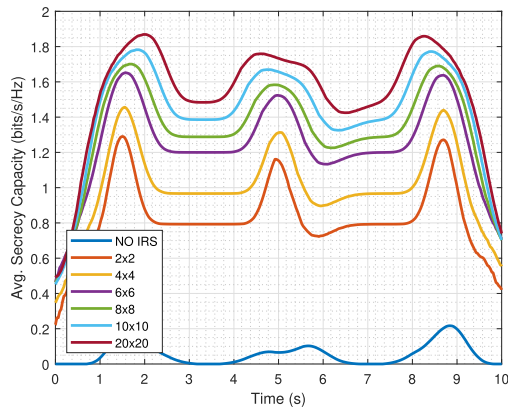
increment requires more training time and larger neural network to map the state to the optimized actions. The time complexity of this system is illustrated in Fig. 8. It shows the training duration for various mirror arrangements simulated with the same number of iterations. As stated in the complexity discussion in subsection IV-B, the number of nodes in the DDPG agent's neural layers

(a) Inference Mode after training with random way point.

(b) Taking the planned path during the training mode.

(c) The SC after training with planned path.

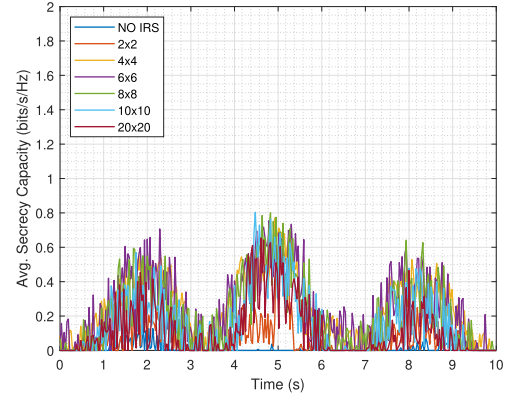Fig. 10. The average SC for the planned path in Fig. 9.

Fig. 11. The SC for the planned path with PSO algorithm [23].

larger number of mirrors increases the SC insignificantly, while the increase in the training complexity is substantial for this particular setting. The best mirror arrangement may differ if the mirror array sheet's total area or the fixtures' position is changed in a different scenario.

To test the adaptability of the proposed algorithm to the location of the users, we simulate the system model with planned paths for users as a realistic scenario. In Fig. 9, we present a planned path for Bob and Eve, assuming Eve is following Bob closely to eavesdrop the VLC channel. Bob moves in a square-shaped path, and Eve follows Bob while maintaining a one-meter gap distance with the speed of 1 m/s. Testing the SC performance of the proposed algorithm with various settings and at different stages is provided in Fig. 10. The initial inference of the trained model is shown in Fig. 10(a). At this stage, the model is only trained with a random waypoint method, and it is using inference only to make decisions without training. In other words, in the inference mode, the DRL-based algorithm takes action without training, hence having significantly less complexity. In inference-only mode, we stop training the model, and we use it to infer the best actions regarding the state (i.e., the location of Bob and Eve). Since the model is trained with the random waypoint method and inference-only mode is used, the sudden change of movement direction at each edge of the squared path model shown in Fig. 9 causes the spikes and dips in the SC shown in Fig. 10(a).

In Fig. 10(b), training mode is enabled, and Bob and Eve are moving along the planned path while learning. Enabling the training mode allows the agent to perform more stable decisions than the inference-only mode since the actions correlate with previous actions. The sudden changes in users' direction still cause some blind points causing steep changes in the SC. However, it is better compared to inference only mode. In Fig. 10(c), the results after training for 10 completed planned paths are provided. Since the model is trained with 10 episodes, the end result is much more smoother compared to (a) and (b), and the BF and the mirror orientations are optimized to maximize the SC. Note that the BF-only (NO IRS) case has low SC values since the beamforming method cannot change the direction of the beams; instead, it only changes the weights of the fixtures.

increases linearly with the number of mirrors, and therefore increases the number of actions.

The trade-off between secrecy performance and time complexity for the DRL-based system is demonstrated by comparing Figs. 7 and 8. It is shown that increasing the number of mirrors requires more time for training and inference, while the increase in the SC is limited. In the context of this trade-off, we show that the best mirror arrangement for this DRL-based IRS-assisted VLC system model is 7×7 with 49 mirrors in total. Using a

In Fig. 11, we present the performance of the PSO-based algorithm in terms of secrecy capacity. The PSO-based algorithm does not learn from experience, hence in every turn the SC is similar to the previous iteration. In each iteration, each particle searches for the maximum SC by updating its value and velocity, however the position of the users change in every iteration as well. Thus, the SC is not stable between iterations. Overall, the performance of the PSO-based algorithm is worse than DRL-based algorithm.

VI. Conclusion

In this work, we proposed a novel DRL-based IRS-assisted SC optimization for the VLC system. The DDPG-based agent controls the BF weights at the light fixtures and the mirror orientations at the mirror array sheet to optimize the SC. We showed that the DDPG-based algorithm can cope with the high complexity and the mobility of users by adapting to new situations. We explored the effect of number of mirrors within the mirror array sheet with fixed total area in terms of the trade-off between the SC and the training complexity. Our results showed that for the given system model, using 7×7 mirror arrangement has the optimal performance and complexity trade-off considering the security of the VLC communication system.

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, 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. Ma *et al.*, "Robust beamforming design for rate splitting multiple access-aided MISO visible light communications," 2021. [Online]. Available: https://arxiv.org/abs/2108.07014

[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.*, 2019, pp. 67–72.

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

[6] 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.*, 2017, pp. 258–263.

[7] 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 2017.

[8] 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, doi: 10.1109/TVT.2022.3168392.

[9] 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.

[10] 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.

[11] 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, doi: 10.1109/JPHOT.2021.3139678.

[12] 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, Jan. 2021.

[13] A. Mostafa and L. Lampe, "Physical-layer security for MISO visible light communication channels," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 9, pp. 1806–1818, Sep. 2015.

[14] A. Mostafa and L. Lampe, "Physical-layer security for indoor visible light communications," in *Proc. IEEE Int. Conf. Commun.*, 2014, pp. 3342–3347.

[15] A. Mostafa and L. Lampe, "Securing visible light communications via friendly jamming," in *Proc. IEEE Globecom Workshops*, 2014, pp. 524–529.

[16] 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.*, 2018, pp. 1–6.

[17] 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.

[18] 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.

[19] 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.*, 2019, pp. 67–72.

[20] 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.

[21] 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.

[22] 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.

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

[24] 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.

[25] 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.

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

[27] 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.

[28] 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.

[29] 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, 2008.

[30] M. Plappert *et al.*, "Parameter space noise for exploration," 2017. [Online]. Available: https://arxiv.org/abs/1706.01905