Jamming for Secrecy: Reinforcement Learning Based Anti-Eavesdropping Visible Light Communication

Xianbin Liu, and Sicong Liu*, *Senior Member, IEEE*
*Dept. of Information and Communication Engineering, Xiamen University, Xiamen, 361005, Fujian, China.
Email: liusc@xmu.edu.cn

ICC 2022 - IEEE International Conference on Communications | 978-1-5386-8347-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICC45855.2022.9838520

***Abstract*—Due to the broadcast nature of visible light communication (VLC), the secrecy protection is a crucial issue. In this paper, aiming at the communication scenario of a point-to-point VLC transmission link, an anti-eavesdropping model utilizing friendly jamming is constructed, which is composed of a light-emitting diode (LED) transmitter, a legitimate receiver equipped with a photodiode (PD), and a potential eavesdropper trying to obtain the undisclosed information transmitted. In addition, the system is equipped with a friendly jammer composed of multiple LEDs. A friendly jamming scheme based on reinforcement learning (RL) is proposed, which adopts different friendly jamming actions with the dynamic change of the environment. The simulation results show that compared with the benchmark state-of-the-art method, the proposed scheme can significantly reduce the bit error rate (BER) of the legitimate receiver and improve the overall performance of the anti-eavesdropping VLC system.**

***Index Terms*—Visible light communication, secrecy protection, reinforcement learning, friendly jamming.**

I. Introduction

Visible light communication (VLC), as the next-generation emerging communication technology [1], benefits from the advantages of low energy consumption, strong ability to effectively avoid signal leakage, and low construction complexity [2], [3], and is widely applied in various areas including indoor positioning [4], underwater communications [5], and mobile health surveillance [6], etc. Utilizing friendly jamming and beamforming, VLC can exploit the physical layer characteristics such as channel randomness to defend against eavesdropping attacks [7]. With the rapid development of high-performance processors and the improvement of the storage capacity of smart eavesdropping devices, the security of legitimate VLC users might not be guaranteed by only using the existing upper-layer network encryption or authentication defense schemes. The transmitted information is very likely to be eavesdropped by malicious attackers within the LED coverage area, especially for the transmission of important and secret information in public places. Therefore, it is necessary to design a safe and effective mechanism to prevent eavesdropping for VLC systems.

In state-of-the-art research, security methods such as password protection, beamforming, and friendly jamming are employed to protect user secrecy [8]–[10]. However, as the computing capability and the deciphering ability of the wiretapper continue to improve, traditional secrecy protection schemes are still prone to risks. A mobile communication secrecy protection mechanism based on visible light characteristics is proposed in [11], assuming that the eavesdropper listens from a specific direction and cannot occur in multiple locations at the same time, and the user rotates the transmitter to a specific receiver range for communication to prevent malicious eavesdropping. The achievable secrecy rate of a multiple-input single-output (MISO) eavesdropping channel with one eavesdropper is improved in [12] through optimal zero-forcing beamforming. In addition, optimized beamforming and friendly jamming schemes reduce the quality of the signal received by the eavesdropper, while causing minimal or no jamming to the target receiver [10], [13].

However, it is difficult for traditional schemes to optimize the friendly jamming strategy in a dynamic and complex communication environment. The status of the complex environment is composed of many aspects, such as the BER of the receiver and the secrecy rate, which can be utilized to improve the security performance of the VLC system. As an important branch of machine learning, the reinforcement learning (RL) technology obtains the optimal strategy of a Markov decision process through dynamic interaction with the environment [14]. In recent years, RL has shown good development prospects in improving communication network security. By adaptively learning defense experience from a dynamic and complex environment, it has been applied to effectively deal with the problem of inaccurate and time-variant information and dynamic selection of security policies in different network security scenarios, such as advanced persistent threat attack defense for cloud storage networks, false perception attack defense for mobile group intelligence perception networks, and anti-hostile jamming for water area communication networks, etc [15]–[18]. To overcome the shortcoming of the traditional physical-layer security schemes, this paper proposes an intelligent RL-based friendly jamming (RL-FJ) scheme for VLC based on RL, which exploits the cooperation of multiple light sources as jammers to prevent eavesdroppers from receiving the secret signals and increase the confidentiality of the VLC transmission process.

The proposed RL-FJ scheme optimizes the jamming strategy of the multiple jammer lights based on the status information of the VLC system such as the bit error rate (BER), the secrecy rate and the energy consumption of the jammers in the VLC process. Since the VLC process can be assumed to be a Markov decision process [19]–[21], it is

This work is supported in part by the National Natural Science Foundation of China under grants 61901403, in part by the Science and Technology Key Project of Fujian Province, China (No. 2019HZ020009), in part by the Youth Innovation Fund of Xiamen under grant 3502Z20206039, and in part by the Natural Science Foundation of Fujian Province of China under grant 2019J05001. (*Corresponding Author: Sicong Liu.* email: liusc@xmu.edu.cn).

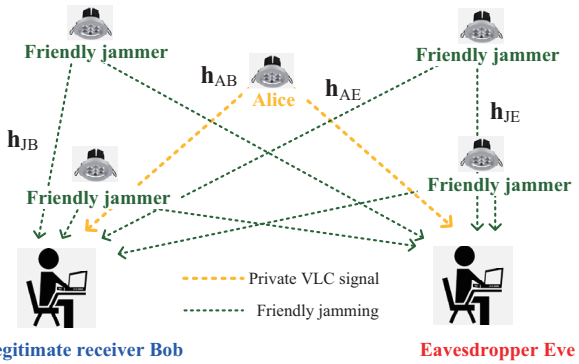


Fig. 1. VLC transmission system from transmitter Alice to legitimate receiver Bob, which is incorporated with four LEDs as friendly jammers in the presence of an eavesdropper Eve.

promising for RL-based methods such as Q-learning [22]–[25] to provide an effective jamming beamformer for the VLC anti-eavesdropping system.

Consequently, the RL-FJ scheme in the framework of RL is proposed in this work, which protects the secret information sent by the transmitter of the VLC system through continuous learning and updating the friendly jamming strategies. The friendly jammer helps the transmitter to ensure the secrecy by employing an intelligently designed jamming beamformer which increases the jamming power received by potential eavesdroppers. In order to achieve the optimal jamming beamformer in the dynamic environment of VLC, the jamming strategy is gradually learnt via a Q-learning process. The results show that after enough learning iterations, the proposed RL-FJ scheme can converge to the optimal jamming beamformer. Compared with the conventional robust jamming scheme, the RL-FJ scheme improves the secrecy rate and the overall utility of the VLC system, and greatly reduces the BER.

The rest of this paper is organized as follows: the VLC model in the presence of eavesdropper is introduced in Section II and the proposed RL-FJ scheme to prevent eavesdropping is presented in Section III. Section IV reports the simulation results and discussions, and conclusions of this work are drawn in Section V.

II. System Model

A. The VLC Transmission Model

The VLC transmission link is modulated by pulse amplitude modulation (PAM) with a DC bias, including an LED at the transmitter and a photodiode (PD) at the receiver. In order to keep the optical signal unipolar for optical radiation, the LED is driven by a current bias represented by I. In the VLC transmission process, the bipolar electrical signal x that conveys information is superimposed on the DC bias I and then converted to a unipolar optical signal with the optical power P_{T} through an electro-optical converter. In order to avoid the nonlinear and clipping distortion problems of electro-optical conversion, the total current of the biased electrical signal $I+x$ must be constrained within a specific range [26]. Therefore, the amplitude of the electrical signal x needs to meet the constraint, i.e., $|x| \leq \alpha I$, where $\alpha \in [0,1]$ is the modulation index. The instantaneous optical power after the electro-optical conversion emitted by the LED is expressed as $P_{\mathrm{T}} = \eta(I+x)$, where η is the electro-optical conversion efficiency coefficient of the LED driver circuits.

The optical signal is then transmitted through the VLC channel and received by the legitimate user, i.e., Bob. Let $P_{\mathrm{R}} = GP_{\mathrm{T}}$ represent the optical power received by the PD at the receiver, where G represents the path gain of the visible light transmission link. The optical power P_{R} received by the PD with a responsivity of R is converted into a corresponding current signal via optical-electro conversion. After removing the DC offset, the electrical signal passes through a signal amplifier with a gain of T to generate the final received information signal denoted as y.

Assuming that the LED radiation pattern in the VLC system is the Lambertian radiation pattern [27]. According to [27] and [28], the visible light path gain G between the transmitter and the receiver is expressed as:

$$G = \begin{cases} \dfrac{n_0^2 A_{\mathrm{P}} (\log\cos\phi_{\frac{1}{2}} - \log 2)}{2\pi d^2 \sin^2(\varphi_{\mathrm{F}}) \log\cos\phi_{\frac{1}{2}}} \cos^{\frac{-\log 2}{\log\cos\phi_{\frac{1}{2}}}}(\phi)\cos(\varphi) & |\varphi| \leq \varphi_{\mathrm{F}}, \\ 0 & |\varphi| > \varphi_{\mathrm{F}}. \end{cases} \tag{1}$$

where ϕ denotes the emission angle of visible light relative to the optical axis of the transmitter, $\phi_{1/2}$ is half the angle of the half luminous intensity of the LED, and φ is the incident angle of visible light with respect to the PD. φ_{F} is the field of view (Fov) of the PD, n_0 is the refractive index of the visible light concentrator, A_{P} is the detection area of the PD, and d is the distance between the transmitter LED and the receiver PD.

B. The VLC wiretapping Model

An anti-eavesdropping system incorporated with friendly jammers for indoor VLC, is illustrated in Fig. 1, including an LED transmitter (i.e., Alice), a legitimate receiver (i.e., Bob) with a PD to communicate with Alice, and a potential eavesdropper (i.e., Eve) which attempts to intercept the secret information sent from Alice to Bob. To combat against eavesdropping, the system is equipped with N_J LEDs that play the role of friendly jammers.

In the VLC wiretapping channel, the signals received by Bob and Eve can be expressed as

$$y_{\mathrm{B}} = h_{\mathrm{AB}}x + \mathbf{h}_{\mathrm{JB}}^{\mathrm{T}}\mathbf{j} + n_{\mathrm{B}}, \tag{2a}$$

$$y_{\mathrm{E}} = h_{\mathrm{AE}}x + \mathbf{h}_{\mathrm{JE}}^{\mathrm{T}}\mathbf{j} + n_{\mathrm{E}}, \tag{2b}$$

where n_{B} and n_{E} are the additive white Gaussian noise (AWGN) vectors with the variance of σ^2 imposed on the receivers of Bob and Eve, respectively. x is the transmitted information signal, and $\mathbf{j} \in \mathbb{R}^{N_J}$ is jamming signal vector transmitted by the N_J LEDs as jammers. $\mathbf{h}_{\mathrm{JB}} \in \mathbb{R}^{N_J}$ and $\mathbf{h}_{\mathrm{JE}} \in \mathbb{R}^{N_J}$ are the channel gain vectors from the N_J LED jammers to Bob and Eve, respectively, where $\mathbf{h}_{\mathrm{JB}} = RT\eta[G_{1\mathrm{B}}, G_{2\mathrm{B}}, \cdots, G_{\mathrm{N_J B}}]^{\mathrm{T}}$ and $\mathbf{h}_{\mathrm{JE}} = RT\eta[G_{1\mathrm{E}}, G_{2\mathrm{E}}, \cdots, G_{\mathrm{N_J E}}]^{\mathrm{T}}$ with $G_{i\mathrm{B}}$ and $G_{i\mathrm{E}}$ being the path gains from the i-th LED jammer to Bob and Eve, respectively, i = 1, 2, ..., $\mathrm{N_J}$. h_{AB} and h_{AE} are the channel gains from Alice to Bob and Eve, respectively, where $h_{\mathrm{AB}} = RT\eta G_{\mathrm{AB}}$ and $h_{\mathrm{AE}} = RT\eta G_{\mathrm{AE}}$. Due to the limited dynamic range of

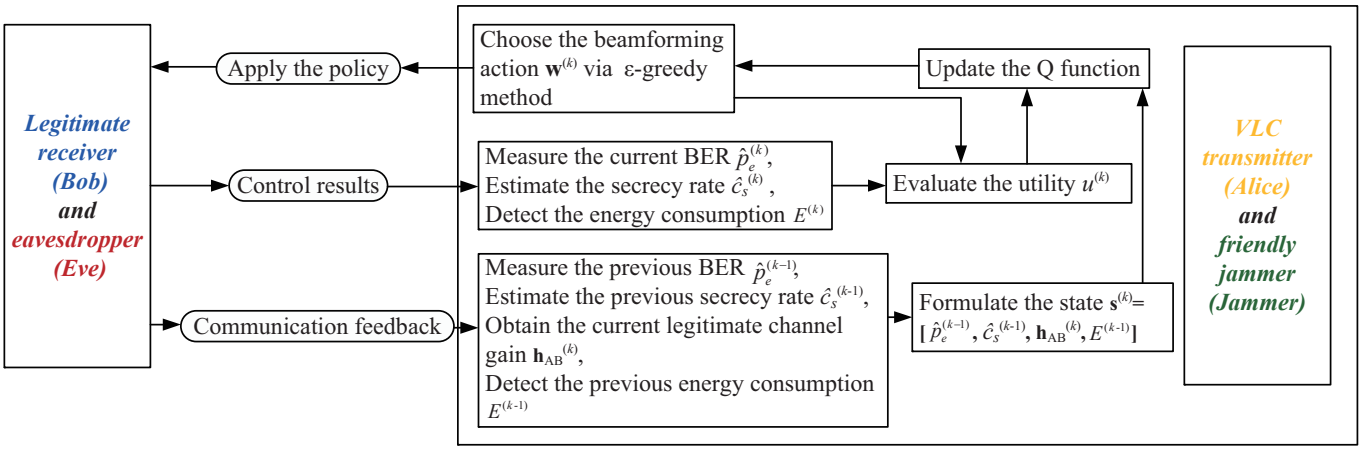


Fig. 2. Proposed friendly jamming control scheme based on RL for secure VLC transmission against eavesdropping.

the LEDs as mentioned previously, the information signal and jamming signals must meet specific amplitude constraints, i.e., $|x| \leq \alpha I$ and $|\mathbf{j}| \prec \alpha I \mathbb{1}$, where $\mathbb{1}$ represents a vector of all ones and $\prec$ denotes that each element in the left vector is smaller than the corresponding value on the right. That is, the absolute value of each element of the jamming signal vector $\mathbf{j}$ is smaller than αI.

To make it convenient for the proposed RL-FJ scheme to change and determine the friendly jamming strategy for the N_J jammer LEDs, the jamming signal can be further represented by a random jamming signal controlled by beamforming, i.e. $\mathbf{j} = \mathbf{w}j$, where $\mathbf{w} = [w_1, w_2, \cdots, w_{N_J}]^{\mathrm{T}}$ is the friendly jamming beamformer that meets the condition $|\mathbf{w}| \prec \mathbb{1}$, and the j is a zero-mean random jamming signal satisfying the condition $|j| \leq \alpha I$. Therefore, Equations (2a) and (2b) can be rewritten as

$$y_{\mathrm{B}} = h_{\mathrm{AB}} x + \mathbf{h}_{\mathrm{JB}}^{\mathrm{T}} \mathbf{w} j + n_{\mathrm{B}}, \tag{3a}$$

$$y_{\mathrm{E}} = h_{\mathrm{AE}} x + \mathbf{h}_{\mathrm{JE}}^{\mathrm{T}} \mathbf{w} j + n_{\mathrm{E}}. \tag{3b}$$

III. RL-BASED FRIENDLY JAMMING CONTROL SCHEME AGAINST EAVESDROPPING

We first analyze the theoretical secrecy rate of the VLC transmission system in the presence of an eavesdropper. According to the average power constraint of the VLC wiretapping channel in [10], the maximum secrecy rate c_s of a point-to-point VLC transmission link incorporated with friendly jammers in the presence of a wiretapper is given by

$$c_s = \frac{1}{2} \log \left(1 + \frac{2\, h_{\mathrm{AB}}^2 \alpha^2 I^2}{\pi e \sigma^2} \right) - \min \left(\log \frac{h_{\mathrm{AE}}}{|\mathbf{h}_{\mathrm{JE}}^{\mathrm{T}} \mathbf{w}|} + \frac{|\mathbf{h}_{\mathrm{JE}}^{\mathrm{T}} \mathbf{w}|}{h_{\mathrm{AE}}} \log \sqrt{e}, \frac{h_{\mathrm{AE}}}{|\mathbf{h}_{\mathrm{JE}}^{\mathrm{T}} \mathbf{w}|} \log \sqrt{e} \right) \tag{4}$$

Therefore, the jamming beamformer $\mathbf{w}^*$ that maximizes the achievable secrecy rate can be derived by maximizing the function in (4) subject to the constraint of zero-forcing the jamming power on the legitimate user Bob, which is given by

$$\mathbf{w}^* = \arg\max_{\mathbf{w}} c_s = \arg\max_{\mathbf{w}} \mathbf{h}_{\mathrm{JE}}^{\mathrm{T}} \mathbf{w}, \quad \text{s.t.} \quad \mathbf{h}_{\mathrm{JB}}^{\mathrm{T}} \mathbf{w} = 0, \quad |\mathbf{w}| \preceq \mathbb{1}, \tag{5}$$

where $\mathbf{h}_{\mathrm{JB}}^{\mathrm{T}} \mathbf{w} = 0$ is a suboptimal constraint of forcing the jamming power imposed on Bob to be zero, in order to make the legitimate user free from friendly jamming. It is noteworthy that the purpose of solving the optimization problem given in (5) is to find a jamming beamformer $\mathbf{w}$ that maximizes the jamming power on the eavesdropper Eve, while the jamming power received by Bob is nulled, thus inevitably reducing the degrees-of-freedom of the solution space of the jamming beamformer strategy. Since the feasible solution space of the jamming beamformer $\mathbf{w}$ is significantly reduced to the null space of $\mathbf{h}_{\mathrm{JB}}$, the jamming strategy cannot be fully explored so the optimal secrecy rate of the problem might not be achieved. Moreover, for the purpose of reliable and effective VLC transmission, the quality of the signal received by the legitimate user Bob is also a very important metric. Thus the BER performance of Bob should also be considered while determining the optimal solution. However, the BER performance cannot be simply modeled in a linear optimization problem like in (5), but it should be improved via an online iterative learning process. Therefore, it is necessary to explore a more effective method to solve the above problems, so that the system performance can reach the overall optimal.

To this end, the RL-FJ scheme is devised in this paper, which aims to improve the overall performance of the VLC system and find a method that can converge to the theoretical optimal solution of the non-convex problem given in (5). The optimal solution to be find should balance the secrecy rate of VLC system and the receive quality such as good BER performance of the legitimate user Bob. In the dynamic communication environment, after the execution of an action, the state of the VLC system is independent of the previous system states and actions, but only depends on the current state and action. Therefore, the selection of the communication policy of the friendly jammers can be regarded as a Markov decision process, i.e., to achieve the optimal friendly jamming scheme through repeated interactive experiments with the environment.

More specifically, as shown in Fig. 2, the jammers select the best jamming beamformer according to the current VLC

Algorithm 1: RL-based friendly jamming scheme (RL-FJ).

Initialize greedy factor β and learning rate λ
$Q(\mathbf{s}, \mathbf{w}) = 0, \forall \mathbf{s} \in \Lambda, \mathbf{w} \in W$
for $k = 1, 2, 3, ...$ **do**
  Obtain the BER feedback from Bob at previous slot $\hat{p}_e^{(k-1)}$
  Estimate the secrecy rate at previous slot $\hat{c}_s^{(k-1)}$
  Obtain the current VLC channel gain $\mathbf{h}_{\mathrm{AB}}^{(k)}$
  Obtain the energy consumption $E^{(k-1)}$ for friendly jamming at previous time slot
  Formulate the current status $\mathbf{s}^{(k)} = \left[\hat{p}_e^{(k-1)}, \hat{c}_s^{(k-1)}, \mathbf{h}_{\mathrm{AB}}^{(k)}, E^{(k-1)}\right]$
  Choose the action of friendly jamming beamforming vector $\mathbf{w}^{(k)}$ via ε-greedy method given in (6)
  The jammers employ the selected action $\mathbf{w}^{(k)}$ to perform friendly jamming beamforming
  Obtain the BER feedback from Bob at current time slot $\hat{p}_e^{(k)}$
  Estimate the secrecy rate at current time slot $\hat{c}_s^{(k)}$
  Obtain the energy consumption $E^{(k)}$ for jamming
  Derive the system utility $u^{(k)}$ using (7)
  Update the Q-function by $Q(\mathbf{s}^{(k)}, \mathbf{w}^{(k)}) \leftarrow (1-\lambda)Q(\mathbf{s}^{(k)}, \mathbf{w}^{(k)}) + \lambda(u^{(k)} + \beta \max_{\mathbf{w}} Q(\mathbf{s}^{(k+1)}, \mathbf{w}))$
end

system state and the policy of Q-function, and prevents the potential eavesdropper Eve from wiretapping the legitimate signal through imposing friendly jamming on Eve. Through the interaction and feedback from the VLC system environment, the previous secrecy rate $\hat{c}_s^{(k-1)}$ is estimated according to the prior geometric information of the VLC transmission environment and statistical channel model information. The previous BER of the legitimate receiver represented by $\hat{p}_e^{(k-1)}$ is obtained from the feedback information from Bob. The previous energy consumption of the jammers denoted as $E^{(k-1)}$ is obtained, and the current legitimate channel gain from Alice to Bob $\mathbf{h}_{\mathrm{AB}}^{(k)}$ is estimated. Then, the current system state can be formulated by $\mathbf{s}^{(k)} = [\hat{p}_e^{(k-1)}, \hat{c}_s^{(k-1)}, \mathbf{h}_{\mathrm{AB}}^{(k)}, E^{(k-1)}] \in \Lambda$, where Λ is the state space containing the set of all possible communication system states.

The proposed RL-FJ scheme is summarized in **Algorithm 1**. The friendly jammers select the best jamming beamformer $\mathbf{w}^{(k)}$ according to the obtained and estimated communication system state $\mathbf{s}^{(k)}$ and the Q-function, where $\mathbf{w}^{(k)} = [w_1^{(k)}, w_2^{(k)}, \ldots, w_{N_\mathrm{J}}^{(k)}]^\mathrm{T} \in \mathbf{W}$, $|w_i| \leq 1$, i = 1, 2, ..., N_J, with $\mathbf{W}$ being the action space containing all possible jamming beamformer vectors. In order to facilitate the establishment of the Q-value table, each element w_i in the jamming beamforming vector is quantified as a discrete value of $2L_x + 1$ equally spaced levels during the entire RL decision-making process. Namely, we have $w_i^{(k)} \in \{l/L_x | -L_x \leq l \leq L_x\}$, where the number of the quantization level L_x can be properly configured to balance the learning accuracy and complexity. In the process of jamming beamformer selection, in order to balance between exploration and exploitation, the ε-greedy method is applied. The jammers will select the jamming beamformer $\mathbf{w}^{(k)}$ that maximizes the Q-value with a high probability $1-\varepsilon$, while other jamming beamformer are randomly selected with a low probability ε to avoid being stuck in local optima, which is represented as

$$\Pr\left(\mathbf{w}^{(k)} = \hat{\mathbf{w}}\right) = \begin{cases} 1-\varepsilon, & \hat{\mathbf{w}} = \arg\max\limits_{\mathbf{w}'} Q\left(\mathbf{s}^{(k)}, \mathbf{w}'\right) \\ \frac{\varepsilon}{|\mathbf{W}|-1}, & \text{o.w.} \end{cases} \tag{6}$$

The selected friendly jamming beamformer is then applied to the anti-eavesdropping VLC system. According to the information obtained from the feedback of Bob and the environment, the secrecy rate of the system at the current time slot $\hat{c}_s^{(k)}$ is estimated, and the energy consumption of the jammers at the current time slot $E^{(k)}$ is obtained. Then, the utility of the anti-eavesdropping VLC system at the current time slot can be obtained by

$$u^{(k)} = \hat{c}_s^{(k)} - \delta_1 \hat{p}_e^{(k)} - \delta_2 E^{(k)}, \tag{7}$$

where δ_1 is the synergy factor that balances the contribution of secrecy rate and BER to the utility, and δ_2 is the synergy factor that balances the influence of secrecy rate and energy consumption on the utility.

In the iterative learning process, the Q-value of each jamming action obtained by the proposed RL-FJ algorithm is represented by $Q(\mathbf{s}, \mathbf{w})$, which represents the long-term discounted benefit of the jammers performing a certain jamming action in a certain system state. The friendly jammers update the next state $\mathbf{s}^{(k+1)}$ according to the iterative Bellman equation as given by

$$\begin{aligned} Q\left(\mathbf{s}^{(k)}, \mathbf{w}^{(k)}\right) \leftarrow (1-\lambda) Q\left(\mathbf{s}^{(k)}, \mathbf{w}^{(k)}\right) + \\ \lambda\left(u^{(k)} + \beta \max_{\mathbf{w}} Q\left(\mathbf{s}^{(k+1)}, \mathbf{w}\right)\right) \end{aligned} \tag{8}$$

where the learning rate $\lambda \in [0, 1]$ represents the weight of the current Q-value when updating the Q-function. To be specific, the agent with a learning rate of zero will not learn anything, and a learning rate of one means that the jammer only considers updating the latest Q-value without considering the past Q-value. The greedy factor $\beta \in [0, 1]$ represents the uncertainty of the learning algorithm for future benefits. In detail, a greedy factor of zero means that the jammer only focuses on the current reward, while a greedy factor of one means that the jammer will consider the reward for a long time in the future.

IV. SIMULATION RESULTS

The performance of the proposed RL-FJ scheme for VLC anti-eavesdropping transmission is evaluated through simulation experiments, which is conducted in a typical indoor VLC scenario, as shown in Fig. 3. The size of the room is $5 \times 5 \times 3$ m^3, where there exist 5 down-facing light fixtures attached to the ceiling. Each light fixture contains 4 identical LEDs and each LED radiates an average optical power of 1 W.

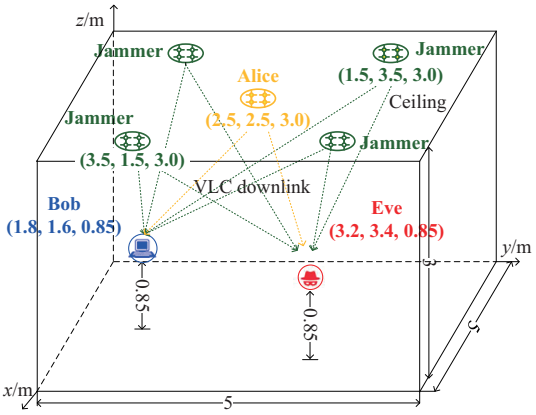


Fig. 3. Simulation experimental environment setup for the VLC transmission system with friendly jammers in the presence of an eavesdropper.

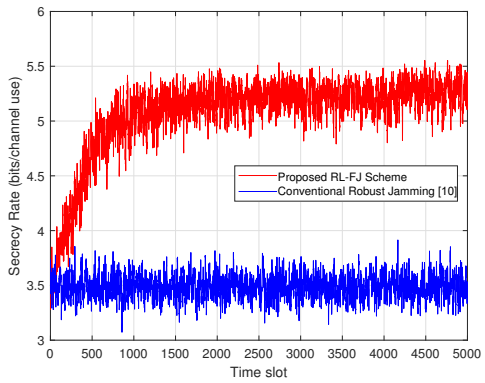
The light in the center is the transmitter Alice that modulates and transmits the information signal, and the 4 lights around Alice form a group of intelligent friendly jammers. All the LEDs employ 4PAM modulation, and the modulation index α is 10%. The legitimate receiver Bob and the eavesdropper Eve are located on a table 0.85m above the ground, both equipped with a PD with a Fov angle φ_{F} of 60° and a detection area A_{P} of 1 cm^2. The semi-angle of half luminous intensity is $\phi_{1/2} = 60^{\circ}$ and the optical concentrator refractive index is n_0 = 1.5. The PD responsivity is R = 0.54 A/W and the average noise power is σ^2 = -98.82 dBm.

In the experimental simulations, we set the learning rate as $\lambda = 0.5$, greedy factor as $\beta = 0.5$ and synergy factor as $\delta_1 = 3.5$, as $\delta_2 = 0.4$. To better explore the solution space, the value of ε adopted in the ε-greedy method is decreased linearly from 1.0 downto 0.1 in the first 400 time slots of the learning process, and subsequently fixed to 0.1 to maintain the stability of the learnt strategy of the VLC system.

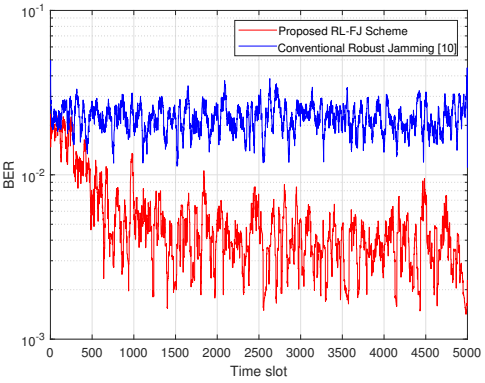
The performance of the proposed RL-FJ scheme is reported in Fig. 4, which is also compared with the conventional robust jamming scheme, in the aspects of secrecy rate, BER of the legitimate receiver, and VLC system utility. It can be noted from Fig. 4 that, the proposed RL-FJ scheme outperforms the conventional robust jamming scheme, achieving a lower BER, higher utility and higher secrecy rate.

More specifically, as shown in Fig. 4(a), the secrecy rate of the proposed RL-FJ scheme increases rapidly in the iterative learning process and converges to 5.25 bits/channel use after 5000 time slots, which is 53.1% higher than that at the initial stage of learning. It is worth noting that at the 5000th time slot, the secrecy rate of the proposed RL-FJ scheme is about 49.6% higher than that of the robust jamming scheme, which verifies the effectiveness of the proposed learning approach towards achieving better security for VLC transmission against wiretapping.

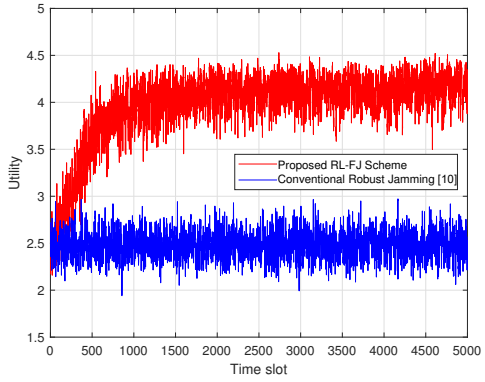
As shown in Fig. 4(b), the BER of the legitimate receiver Bob drops rapidly using the proposed RL-FJ scheme, and finally approaches a mean value of 2.3×10^{-3}, which is much lower than the BER at the beginning of the learning process. It can also be noted from Fig. 4(b) that the BER of the proposed RL-FJ scheme is far much lower than that of



(a) Secrecy rate of the VLC system



(b) BER of the legitimate receiver Bob



(c) Utility of the VLC system

Fig. 4. Performance evaluation of the proposed RL-FJ scheme compared with the conventional robust jamming scheme: (a) secrecy rate, (b) BER of the legitimate receiver, and (c) VLC system utility.

the robust jamming scheme at the 5000th time slot, which shows the superior quality of the VLC transmission using the proposed method.

For the overall utility of the VLC system, as shown in Fig. 4(c), the utility of the proposed RL-FJ scheme increases with the iterations of the learning process and converges to a value of about 4.21 after 5000 time slots, which is 108.4% higher than the initial stage of learning. Besides, at the 5000th time slot, the utility of the RL-FJ scheme significantly exceeds that of the robust jamming scheme by 75.4%, which verifies

the good performance of the proposed scheme in preventing eavesdropping for secure and efficient VLC transmission.

V. Conclusion

In this paper, an intelligent secrecy protection framework for secure VLC transmission has been proposed, which can realize undisclosed arfaoui2018artificialinformation protection by exploiting smart friendly jamming on the eavesdropper via multiple LEDs. To obtain the optimal friendly jamming scheme in the dynamic and complex VLC transmission environment in the presence of wiretapping, a friendly jamming scheme based on RL has been devised to iteratively learn the optimal policy of the friendly jamming beamforming vectors that maximize the system performance rapidly and effectively. Simulation results have shown that compared with the conventional robust jamming scheme, the proposed RL-FJ scheme can significantly increase the secrecy rate and the overall utility of the VLC system, while reducing the BER of the legitimate receiver. The proposed method is also promising to be applied in other different VLC scenarios where there exist eavesdroppers to guarantee the transmission security.

References

[1] T. Komine and M. Nakagawa, "Fundamental analysis for visible-light communication system using LED lights," *IEEE Trans. Consum. Electron.*, vol. 50, no. 1, pp. 100–107, Feb. 2004.

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

[3] F. Yang, J. Gao, S. Liu, and J. Song, "Clipping noise elimination for OFDM systems by compressed sensing with partially aware support," *IEEE Trans. Broadcast.*, vol. 63, no. 1, pp. 103–110, Mar. 2017.

[4] D. Su, X. Liu, and S. Liu, "Three-dimensional indoor visible light localization: A learning-based approach," in *UbiComp '21*, pp. 672–677, New York, NY, Sept. 2021.

[5] X. Ma, F. Yang, S. Liu, and J. Song, "Channel estimation for wideband underwater visible light communication: A compressive sensing perspective," *Opt. Express*, vol. 26, no. 1, pp. 311–321, Aug. 2018.

[6] P. Pathak, X. Feng, P. Hu, and P. Mohapatra, "Visible light communication, networking, and sensing: A survey, potential and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2047–2077, Sept. 2015.

[7] F. Yang and J. Gao, "Dimming control scheme with high power and spectrum efficiency for visible light communications," *IEEE Photon. J.*, vol. 9, no. 1, pp. 1–13, Feb. 2017.

[8] F. Mousa, N. Almaadeed, K. Busawon, *et al.*, "Secure MIMO visible light communication system based on user's location and encryption," *J. Lightw. Technol.*, vol. 35, no. 24, pp. 5324–5334, Dec. 2017.

[9] B. Zhang, K. Ren, G. Xing, *et al.*, "SBVLC: Secure barcode-based visible light communication for smartphones," *IEEE. Trans. Mob. Comput.*, vol. 15, no. 2, pp. 432–446, Feb. 2016.

[10] A. Mostafa and L. Lampe, "Securing visible light communications via friendly jamming," in *Globecom Workshops*, pp. 524–529, Austin, TX, Dec. 2014.

[11] L. Sun and Q. Du, "Physical layer security with its applications in 5G networks: A review," *China Commun.*, vol. 14, no. 12, pp. 1–14, Dec. 2017.

[12] 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, Sept. 2015.

[13] H. Zaid, Z. Rezki, A. Chaaban, and M. S. Alouini, "Improved achievable secrecy rate of visible light communication with cooperative jamming," in *GlobalSIP*, pp. 1165–1169, Orlando, FL, Dec. 2015.

[14] L. Xiao, Y. Ding, J. Huang, S. Liu, Y. Tang, and H. Dai, "UAV anti-jamming video transmissions with QoE guarantee: A reinforcement learning-based approach," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 5933–5947, Sept. 2021.

[15] H. Yang, Z. Xiong, J. Zhao, *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.

[16] L. Xiao, G. Sheng, S. Liu, *et al.*, "Deep reinforcement learning-enabled secure visible light communication against eavesdropping," *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 6994–7005, Oct. 2019.

[17] Y. Liu, H. Yu, S. Xie, and Y. Zhang, "Deep reinforcement learning for offloading and resource allocation in vehicle edge computing and networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 11158–11168, Nov. 2019.

[18] C. Li, J. Xia, F. Liu, *et al.*, "Dynamic offloading for multiuser muti-CAP MEC networks: A deep reinforcement learning approach," *IEEE Trans. Veh. Technol.*, vol. 70, no. 3, pp. 2922–2927, Mar. 2021.

[19] N. Mastronarde and M. Schaar, "Fast reinforcement learning for energy-efficient wireless communication," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 6262–6266, Dec. 2011.

[20] S. A. Osia, A. Shahin Shamsabadi, S. Sajadmanesh, *et al.*, "A hybrid deep learning architecture for privacy-preserving mobile analytics," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4505–4518, May. 2020.

[21] C. Jiang, H. Zhang, Y. Ren, *et al.*, "Machine learning paradigms for next-generation wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 98–105, Apr. 2017.

[22] D. Yang, G. Xue, J. Zhang, *et al.*, "Coping with a smart jammer in wireless networks: A stackelberg game approach," *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 4038–4047, Aug. 2013.

[23] X. He, H. Dai, and P. Ning, "Faster learning and adaptation in security games by exploiting information asymmetry," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3429–3443, Jul. 2016.

[24] H. Metwaly Saad, A. Mohamed, and T. Elbatt, "Cooperative Q-learning techniques for distributed online power allocation in femtocell networks," *Wirel. Commun. Mob. Comput.*, vol. 15, no. 15, pp. 1929–1944, Oct. 2015.

[25] Y. Xiao, G. Niu, L. Xiao, Y. Ding, S. Liu, and Y. Fan, "Reinforcement learning based energy-efficient internet-of-things video transmission," *Intelligent and Converged Networks*, vol. 1, no. 3, pp. 258–270, Dec. 2020.

[26] F. Yang, Y. Sun, and J. Gao, "Adaptive LACO-OFDM with variable layer for visible light communication," *IEEE Photon. J.*, vol. 9, no. 6, pp. 1–8, Dec. 2017.

[27] L. Zeng, D. O'Brien, H. Le Minh, *et al.*, "High data rate multiple input multiple output (MIMO) optical wireless communications using white LED lighting," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 9, pp. 1654–1662, Dec. 2009.

[28] J. Kahn and J. Barry, "Wireless infrared communications," *Proc. IEEE*, vol. 85, no. 2, pp. 265–298, Feb. 1997.