

Cooperative Jamming for Physical Layer Security Enhancement Using Deep Reinforcement Learning

Sayed Amir Hoseini¹, Faycal Bouhafs¹, Neda Aboutorab², Parastoo Sadeghi² and Frank den Hartog¹

¹School of Systems and Computing

²School of Engineering and Technology

University of New South Wales, Canberra, Australia

Email: {s.a.hoseini, f.bouhafs, n.aboutorab, p.sadeghi, frank.den.hartog}@unsw.edu.au

Abstract—Wireless data communications are always facing the risk of eavesdropping and interception. Conventional protection solutions which are based on encryption may not always be practical as is the case for wireless IoT networks or may soon become ineffective against quantum computers. In this regard, Physical Layer Security (PLS) presents a promising approach to secure wireless communications through the exploitation of the physical properties of the wireless channel. Cooperative Friendly Jamming (CFJ) is among the PLS techniques that have received attention in recent years. However, finding an optimal transmit power allocation that results in the highest secrecy is a complex problem that becomes more difficult to address as the size of the wireless network increases. In this paper, we propose an optimization approach to achieve CFJ in large Wi-Fi networks by using a Reinforcement Learning Algorithm. Obtained results show that our optimization approach offers better secrecy results and becomes more effective as the network size and the density of Wi-Fi access points increase.

Index Terms—Artificial Noise, Secrecy, Physical-Layer Security, SDN, Programmable Networks, Friendly Jamming, Reinforcement Learning, Machine Learning.

I. INTRODUCTION

The pervasive and broadcasting nature of electromagnetic waves in wireless networks poses a significant risk of unauthorized access and interception of sensitive data. Although encryption and authentication are common security measures, Physical Layer Security (PLS) is increasingly being recognized as a promising approach to provide an additional layer of protection to wireless communications [1]. However, until recently PLS was primarily restricted to theoretical study of the problem.

We have shown in [2], [3] that PLS could indeed be realized by exploiting the flexibility offered by Software-Defined Networking (SDN) and Spectrum Programming [4]. In our previous work, we have shown that it is possible to implement a PLS solution for Wi-Fi networks using off-the-shelf equipment [2], [3]. In these works, we implemented an algorithm that associates the legitimate station to the Access Point (AP) that provides the highest secrecy capacity in the presence of an eavesdropper [5]. We extended these works in [6] by incorporating the idea of cooperative friendly jamming (CFJ), where APs produce jamming signals to further degrade the capacity of eavesdropping. However, this work was limited to a single user, a single eavesdropper, and two

APs. The optimization formulation and closed-form equations presented in [6] are not scalable to more realistic scenarios that include multiple APs, legitimate stations, and eavesdroppers.

In this paper, we extend our contribution in [6] to larger wireless networks, where APs can work as both legitimate traffic sources and jammers at the same time. More specifically, we aim to optimize the transmit power of each AP that results in the maximum achievable sum secrecy capacity across legitimate users. For this purpose, we propose to use a Reinforcement Learning (RL) technique within Machine Learning (ML) to find a near-optimal radio configuration. In our proposed RL framework, the state of the network is the location of nodes, which underpins the received power at each legitimate or eavesdropper node according to the assumed wireless propagation path-loss model. The action taken is the transmit power from each AP and the revenue is the sum secrecy capacity among legitimate users. Numerical evaluation of our proposed RL method in numerous network scenarios shows significant sum secrecy capacity improvement across the network.

The rest of the paper is structured as follows. In Section II, the related work is presented briefly. Section III formulates the proposed CFJ, optimization problem, and the reinforcement learning model, followed by simulation results in Section IV. Finally, we conclude the paper and discuss the future work in Section V.

II. RELATED WORK

The application of ML to wireless communications has gained significant interest over the last few years [7]. Several ML-based solutions have been proposed in the literature to help address issues such as spectrum sensing [8], RF coexistence [9], and RF transmitter identification [10]. In the context of jamming, contributions focused only on proposing ML-based solutions to detect jamming attacks and their perpetrators [11]–[13]. Authors in [14] proposed to identify possible ML techniques that could be applied to PLS, but focused on Physical Layer Authentication, Antenna Selection, and Relay Node Selection. In [15], authors proposed an ML-based PLS to secure communications between Intelligent Reflecting Surfaces and Mobile Edge Computing nodes. In [16], authors proposed a federated learning framework to enhance PLS

where part of the model aims to optimize the transmit power to reduce the eavesdropping opportunity.

In addition, all contributions mentioned above were limited to theoretical models. Our works in [2], [3] were the first implementation of PLS in wireless networks. In these works, we presented an implementation of a PLS solution where a legitimate station is associated with the AP that offers the highest secrecy in the presence of an eavesdropper. The choice of the AP is based on the algorithm we proposed in [5] and the implementation exploits the spectrum programmability concept presented in [4]. In [6], we extended our previous work by proposing a friendly jamming method that improves the secrecy of wireless communication in the presence of an eavesdropper. In [6], we used an optimization model that determines the transmit power that a jamming AP should use to improve the secrecy capacity of the connection. However, this model could only be applied to a scenario that involves two APs and one legitimate user and could not scale to larger deployment scenarios. In this current study, we improve upon this model by proposing a more comprehensive wireless network solution that can enhance PLS through CFJ. We employ Reinforcement Learning [17] to determine the best power allocation to APs. Most existing works on power optimization, such as [18], [19], considered a limited point-to-point scenario (i.e., Alice-Bob communication). To the best of our knowledge, there is no previous work on joint optimization of user association and power allocation for a scenario with multiple APs and jammers. Therefore, the results in our work aim to show the superiority of the proposed model to existing systems in the real world or those proposed in the literature. Furthermore, as far as we know, this is the first research to propose and solve this model using reinforcement learning.

III. SYSTEM MODEL

A. Proposed Friendly Jamming Formulation

Let us assume a network that includes N APs (AP_1, AP_2, \dots, AP_N), K users as u_1, u_2, \dots, u_K , and J eavesdroppers as e_1, e_2, \dots, e_J , which are located in the area that is serviced by a cooperative wireless network. Each user u_k is associated with AP_{α_k} , where $A = [\alpha_1, \dots, \alpha_K]$ is a vector that denotes the association of users to APs and $P^t = [p_1^t, \dots, p_N^t]$ is a vector that represents the transmit power of each AP. APs transmit signals continuously. If there is an associated user to receive data they act as a normal AP and otherwise as a jammer. A simple network with $N = 4$, $K = 2$ and $J = 2$ is illustrated in Figure 1.

We consider the Friis transmission formula [20] for the wireless propagation model where for a transmit power of p^t , the received power at distance d is calculated by

$$p^r = p^t G_t G_r \left(\frac{\lambda}{4\pi}\right)^2 \left(\frac{1}{d}\right)^\gamma, \quad (1)$$

where G_t and G_r are antenna gain for the transmitter and receiver respectively, λ is the wavelength of the radio frequency and γ is the path loss exponent. The received power at u_k and e_j from AP_n are denoted by $p_{n,k}^{ru}$ and $p_{n,j}^{re}$, respectively.

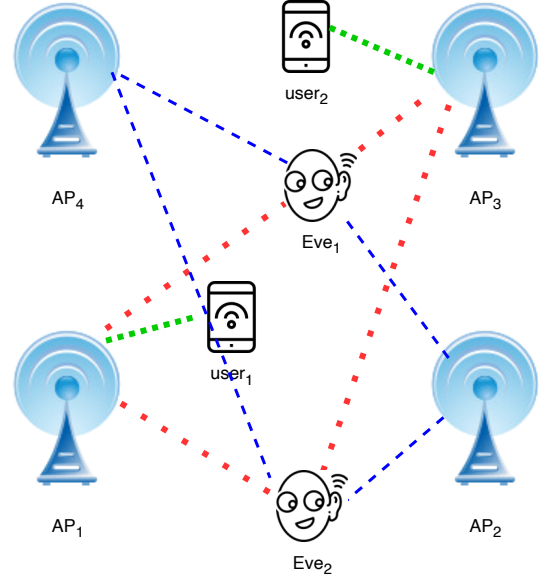


Figure 1: In a cooperative wireless network, user₁ and user₂ receive downlink traffic from AP₁ and AP₃, respectively (shown by green dotted lines). However, Eve₁ and Eve₂ are eavesdropping on their traffic (indicated by red dotted lines). Additionally, AP₂ and AP₄ are idle and act as jammers. It is important to note that all APs and users communicate in the same frequency band. As a result, AP₁'s signal can also be considered as jamming signal for any eavesdropper wiretapping AP₃'s traffic, and AP₂'s signal can also be considered as jamming signal for any eavesdropper wiretapping AP₁'s traffic.

Therefore, the Shannon capacity [21] of the downlink channel between AP_n and the legitimate user u_k is given as

$$C_{n,k} = W \log(1 + \text{SINR}_{n,k}) \\ = W \log \left(1 + \frac{p_{n,k}^{ru}}{\sum_{\nu=1, \nu \neq n}^N p_{\nu,k}^{ru} + N_{u_k}} \right), \quad (2)$$

where W is the channel bandwidth, $\text{SINR}_{n,k}$ is the Signal-to-Interference-plus-Noise Ratio (SINR) at u_k from AP_n and N_{u_k} is the noise power at u_k . Similarly, the Shannon capacity of the channel between AP_n and the eavesdropper e_j is

$$C_{n,j}^e = W \log(1 + \text{SINR}_{n,j}) \\ = W \log \left(1 + \frac{p_{n,j}^{re}}{\sum_{\nu=1, \nu \neq n}^N p_{\nu,j}^{re} + N_{e_j}} \right), \quad (3)$$

where $\text{SINR}_{n,j}$ is the SINR at e_j from AP_n and N_{e_j} is the noise power at e_j . In this paper, all logs are in base 2 and, therefore, capacities are measured in bits/s. Equation (3) shows all other APs contribute to the jamming and reduce the eavesdropper capacity, regardless of whether the AP is functioning as a jammer or transmitting data to users. We assume there is no collusion among eavesdroppers. In this paper, we consider the wiretap capacity [22] as the PLS performance metric, which is widely studied in the literature, e.g., see [23] for an early work and [24] for more recent results. Also, let us assume the worst-case scenario when the

eavesdropper with the highest capacity is wiretapping. The maximum capacity among all eavesdroppers to wiretap the traffic transmitted by AP_n is defined by

$$C^e(n) = \max(\{C_{n,j}^e : j = 1, \dots, J\}). \quad (4)$$

Assuming the user u_k is associated with AP_{α_k} , the secrecy capacity for the user u_k is given by

$$C^s(u_k, \alpha_k) = [C_{\alpha_k, k} - C^e(\alpha_k)]^+, \quad (5)$$

where $[x]^+$ denotes $\max\{x, 0\}$. In this paper, we aim to find the joint optimal user association and power allocation to the APs that maximize the sum of the secrecy capacity of all legitimate users, which the optimization problem is formulated as follows

$$\begin{aligned} \Psi : \quad & \max_{P^t, A} \left(\sum_{k=1}^K C^s(u_k, \alpha_k) \right), \\ \text{s.t.} \quad & 0 \leq p_n^t \leq p_{max}, \\ & \alpha_k \in \{1, 2, \dots, N\}, \end{aligned} \quad (6)$$

where $A := [\alpha_1, \dots, \alpha_K]$. Solving this joint optimization problem is a difficult task due to the non-linear and non-convex nature of secrecy capacity expression [25] where the user association optimization adds extra complexity. In this paper, we propose to decouple the original problem into two subproblems. First, we employ the AP selection algorithm from [2], [5] where each user u_k is associated with AP_{α_k} that provides the maximum secrecy capacity $C^s(u_k)$ considering eavesdropper locations and a uniform power allocation to APs (i.e., $p_n^t = p_{max}$). That is, the association index α_k for user u_k is obtained by

$$\alpha_k = \arg \max_{n \in \{1, 2, \dots, N\}} (C_{n, k} - C^e(n)). \quad (7)$$

Then, we use this user association $A = [\alpha_1, \dots, \alpha_K]$ and formulate the optimization of the power allocation of APs by

$$\begin{aligned} \Phi = \arg \max_{P^t = [p_1^t, \dots, p_N^t]} & \left(\sum_{k=1}^K C^s(u_k, \alpha_k) \right), \\ \text{s.t.} \quad & 0 \leq p_n^t \leq p_{max}, \end{aligned} \quad (8)$$

which is still a non-convex problem. To solve this, We employ Deep Reinforcement Learning that maximizes the sum secrecy capacity. This approach lets us extend this solution in our future works to consider user mobility or channel selection.

B. Proposed Reinforcement Learning Model

Since the wireless network is controlled by software, APs, and stations' information are available for the controller's software that can be used to control the network globally [4]. Therefore, the reinforcement learning model for our proposed system is defined as follows:

- **State (S)** is defined based on the observed information of the environment that includes the location of users, eavesdroppers, and APs. Thus, we define the state as $\mathbf{S} = \{L_{AP}, L_u, L_e\}$ where $L_{AP} = [l_{AP1}, l_{AP2}, \dots, l_{APN}]$

is a vector that denotes the location of APs, $L_u = [l_{u1}, l_{u2}, \dots, l_{uK}]$ typifies the location of user stations, and $L_e = [l_{e1}, l_{e2}, \dots, l_{eJ}]$ represents the location of eavesdroppers.

- **Action (a)** is defined as the vector of transmit power of APs: $P^t = [p_1^t, \dots, p_N^t]$.
- **Revenue (R)** is the combination of rewards and penalties after taking action \mathbf{a} at state \mathbf{S} . In our model, we only consider a reward for the sum of positive secrecy capacities among users:

$$R(\mathbf{S}, \mathbf{a}) = \sum_{k=1}^K C^s(u_k, \alpha_k). \quad (9)$$

This paper utilizes the Soft Actor-Critic (SAC) algorithm [17], which is an off-policy approach resulting in a stochastic policy for optimal actions. The SAC algorithm's main advantage is the ability to balance exploration and exploitation, which the model can adjust during training to prevent the policy from converging too early to an undesirable local optimum [26].

IV. PERFORMANCE EVALUATION

We evaluated the performance of the proposed RL-based CFJ model using simulations with Matlab. We simulated three Wi-Fi systems: a) a normal Wi-Fi implementation where the user is associated with the AP with the highest SINR regardless of the eavesdroppers' location and eavesdropping capacity; b) the smart AP implementation where the user is associated with the AP that provides the highest secrecy capacity according to (7) (but without any power optimization); and c) The RL-Based CFJ implementation where each AP can work as a source of legitimate traffic or jamming to increase the secrecy of communication according to (7) and (8) where its transmit power is optimized. In more detail, each AP is sending either data frames or jamming signals continuously. If APs associate with more than one user, time division multiplexing will be used. However, instantaneous capacity is considered for secrecy. As shown in (3), each AP can contribute to the friendly jamming effectiveness against eavesdropping on other AP's traffic. In the normal Wi-Fi (a) and smart AP (b) implementation, the transmit power of every AP is set to 1 watt, and the idle APs (not associated with any legitimate user) are neither transmitting data traffic nor jamming. The RL-based (c) implementation firstly employs (7) to select the AP for each user assuming the transmit powers are maximum across APs and then optimizes transmit powers to achieve the maximum sum secrecy capacity. The optimal transmit power is limited to a maximum of 1 Watt.

The map considered for the simulations has dimensions of $50m \times 50m$, where all coordinates are expressed in meters (m). Users and eavesdroppers were randomly located, and the APs were placed sufficiently far from each other to cover the entire map area. The communication between users and APs is conducted over Wi-Fi operating at $f_0 = 2.4$ GHz. The noise power at all nodes was set to $N = N_m = N_e = -85$ dBm =

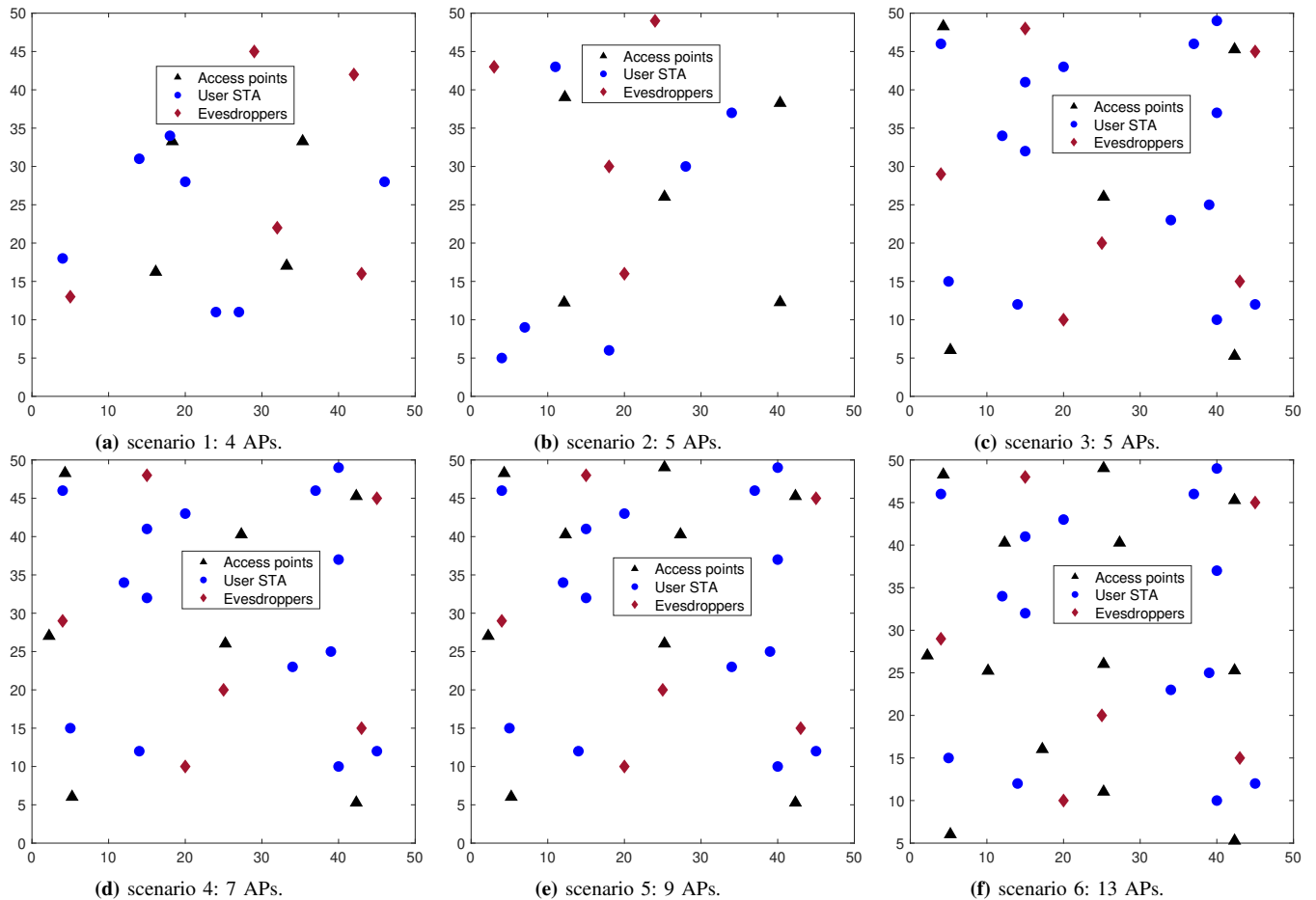


Figure 2: These figures display the locations of users, eavesdroppers, and APs in various scenarios. Scenarios 3, 4, 5, and 6 involve identical numbers and locations of users and eavesdroppers, with only new APs being introduced.

3.16×10^{-12} Watts. We assumed a path-loss exponent of $\gamma = 2$ for the entire map. To simulate the environment and train the RL agent, we used Matlab’s Reinforcement Learning Toolbox. The action, as defined in Section III-B, was implemented as a continuous space. The map was divided into $1m \times 1m$ cells to discretize the observation space and limit possible states. The deep learning networks for the SAC agent’s critic and actors comprised nine layers of depth and 32 to 256 hidden units, depending on the number of APs.

In order to investigate the impact of the user node and AP location on our system, we conducted simulations of six different scenarios, as presented in Figure 2. Scenarios 3 to 6 featured the same number of users and eavesdroppers, all located in the same positions. However, the number of APs varied, increasing from 5 in scenario 3 to 7, 9, and 13 in scenarios 4, 5, and 6, respectively. This allowed us to explore the effects of more densely deployed APs on our system. For each scenario, we simulated the three implementations of normal Wi-Fi, smart AP and RL-based CFJ as described above. For the results, we calculated a) *sum secrecy capacity* $\sum_{k=1}^K \max(C^{s}(u_k, \alpha_k)$, b) *sum Eve capacity* where the capacity of each eavesdropper is the highest eavesdropping capacity on K users according to (4)

and c) the *secrecy ratio*, which we define as the percentage of users that can achieve positive secrecy capacity. The capacities are shown for $W = 1$ Hz or equivalently bit per second per hertz (bps/Hz). The antenna gains at the transmitter, G_t , and the receiver, G_r , are assumed to be unity.

The results are visualized as in Figure 3. It can be observed that the RL-based CFJ outperforms the normal Wi-Fi and smart AP implementations in all scenarios, achieving a higher sum secrecy capacity. Please note that the revenue function in (9) only rewards the sum secrecy capacity according to (6) and there is no explicit penalty for eavesdropper’s capacity in this equation. Nevertheless, such RL method also manages to achieve a lower eavesdropping capacity for most scenarios. To demonstrate the distribution of secrecy capacity improvement among users, we presented another group of bar charts depicting the percentage of secrecy ratio in Figure 3. The results clearly show that the proposed RL-based model outperforms the other two implementations. Moreover, the smart AP implementation also shows better performance than normal Wi-Fi as the latter does not take into account the PLS in any form.

The second scenario results are distinct in the sense that we can see for all the implementations, the secrecy ratio is 100%.

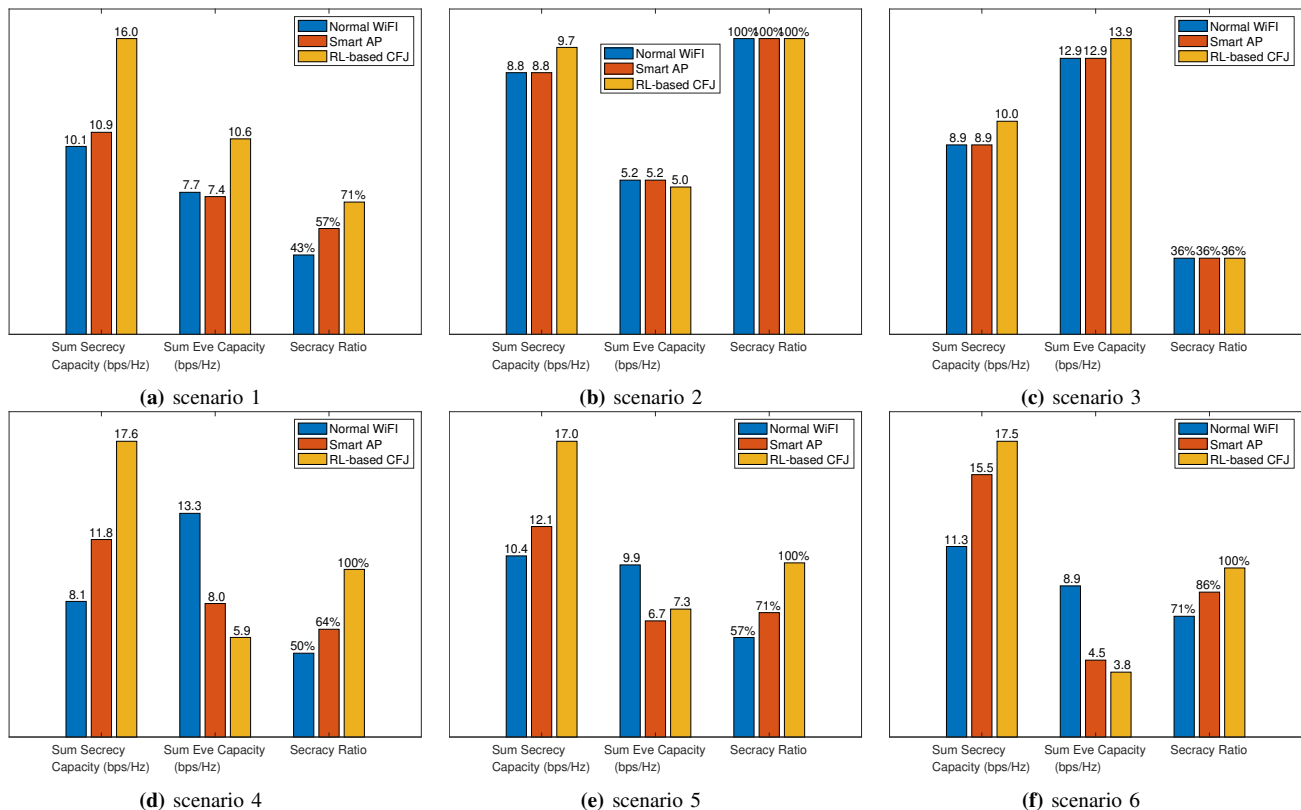


Figure 3: Results for different scenarios for 3 different implementations. RL-based CFJ can generally outperform other implementations. Scenarios 3, 4, 5, and 6 involve identical numbers and locations of users and eavesdroppers, with only new APs being introduced. Therefore, secrecy can be improved by deploying more APs. Although scenario 5 is expected to perform as well as or better than scenario 4 for RL-based CFJ, the results show a slight downgrade. This is because the AP selection algorithm is run separately from power optimization and before it.

This is due to the eavesdroppers' location, where almost all users are closer than the eavesdroppers to at least one AP. This leads to positive secrecy for all users in all implementations where the smart AP selection as presented in (7) results in the same selection as normal Wi-Fi (where the closest AP is simply selected).

In the third scenario, there is an increase in both the number of users and eavesdroppers, leading to a significant increase in eavesdropping capacity and a decrease in secrecy ratio. RL-Based power optimization also cannot improve the secrecy capacity notably. However, by adding two more APs to scenario 3 and keeping the same number and location for users and eavesdroppers, it can be observed in scenario 4 that there is a significant improvement in the results. The proposed RL-based implementation remarkably outperforms the other two approaches.

However, by adding two more APs in scenario 5, the sum secrecy capacity is slightly lower and the eavesdropper capacity increases for RL-based results. This was unexpected, given that we anticipated achieving at least the same level of performance as scenario 4. Our investigation revealed this is because some users are now associated with newly added APs. We conclude that since the AP selection algorithm is done separately and first, by assuming uniform power allocation,

and then the RL model is used to find the optimal power vector, the whole process is not optimal. This should be improved by integrating the AP selection in the RL model in future work. Generally speaking, we can conclude that having a more dense AP deployment improves the secrecy of users as can be seen in scenarios 3 to 6.

V. CONCLUSIONS AND FUTURE WORK

In this paper, we presented an optimization solution for the transmit power allocation, necessary to achieve CFJ in large Wi-Fi networks, by leveraging the Soft Actor-Critic RL agent methodology within reinforcement learning. The obtained results show that the proposed solution enhances the secrecy of wireless communication in the presence of multiple eavesdroppers. Furthermore, we found that a higher density of APs generally offers more opportunities to improve the secrecy of communication.

These results demonstrate the effectiveness of ML in addressing the complex problem of transmit power optimization of friendly jamming to achieve PLS. They also showcase the potential of ML in addressing other optimization problems in PLS that could not be solved with conventional theoretical or numerical optimization methods. More specifically, we aim to integrate the AP selection algorithm into the RL model in our

future work, to further optimize the secrecy of the wireless network. Finally, we are also considering the extension of this work to mobile legitimate stations and eavesdroppers, by training the RL model to the dynamic environment and mobility patterns that characterize large Wi-Fi networks.

REFERENCES

- [1] Y. Liu, H.-H. Chen *et al.*, “Physical layer security for next generation wireless networks: Theories, technologies, and challenges,” *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 347–376, 2016.
- [2] S. A. Hoseini, F. Bouhafs *et al.*, “A practical implementation of physical layer security in wireless networks,” in *2022 IEEE 19th Annual Consumer Communications Networking Conference (CCNC)*, 2022, pp. 1–4.
- [3] —, “Realizing Physical Layer Security with common off-the-shelf WiFi equipment,” in *CCNC 2023 : IEEE Consumer Communications and Networking Conference*, 2023.
- [4] F. Bouhafs, M. Mackay *et al.*, “Wi-5: A programming architecture for unlicensed frequency bands,” *IEEE Communications Magazine*, vol. 56, no. 12, pp. 178–185, 2018.
- [5] F. Bouhafs, F. den Hartog *et al.*, “Realizing physical layer security in large wireless networks using spectrum programmability,” in *2020 IEEE Globecom Workshops*, 2020, pp. 1–6.
- [6] S. A. Hoseini, P. Sadeghi *et al.*, “Network-Controlled Physical-Layer Security: Enhancing Secrecy Through Friendly Jamming,” in *27th IEEE Symposium on Computers and Communications (ISCC 2022)*, 2022.
- [7] T. Erpek, T. O’Shea *et al.*, “Deep learning for wireless communications,” *Development and Analysis of Deep Learning Architectures*, pp. 223–266, 2020.
- [8] D. Janu, K. Singh *et al.*, “Machine learning for cooperative spectrum sensing and sharing: A survey,” *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 1, p. e4352, 2022.
- [9] M. Ahmed, B. Gabr *et al.*, “Machine learning-based module for monitoring lte/wifi coexistence networks dynamics,” in *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*. IEEE, Conference Proceedings, pp. 1–6.
- [10] K. Youssef, L. Bouchard *et al.*, “Machine learning approach to rf transmitter identification,” *IEEE Journal of Radio Frequency Identification*, vol. 2, no. 4, pp. 197–205, 2018.
- [11] Y. Arjoun, F. Salahdine *et al.*, “A novel jamming attacks detection approach based on machine learning for wireless communication,” in *2020 International Conference on Information Networking (ICOIN)*. IEEE, Conference Proceedings, pp. 459–464.
- [12] G. Kasturi, A. Jain *et al.*, “Detection and classification of radio frequency jamming attacks using machine learning,” *J. Wirel. Mob. Networks Ubiquitous Comput. Dependable Appl.*, vol. 11, no. 4, pp. 49–62, 2020.
- [13] O. Puñal, I. Aktaş *et al.*, “Machine learning-based jamming detection for ieee 802.11: Design and experimental evaluation,” in *Proceeding of IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks 2014*. IEEE, Conference Proceedings, pp. 1–10.
- [14] A. K. Kamboj, P. Jindal *et al.*, “Machine learning-based physical layer security: techniques, open challenges, and applications,” *Wireless Networks*, vol. 27, pp. 5351–5383, 2021.
- [15] L. Zhang, S. Lai *et al.*, “Deep reinforcement learning based irs-assisted mobile edge computing under physical-layer security,” *Physical Communication*, vol. 55, p. 101896, 2022.
- [16] T. Wang, Y. Li *et al.*, “Secrecy driven federated learning via cooperative jamming: An approach of latency minimization,” *IEEE Transactions on Emerging Topics in Computing*, vol. 10, no. 4, pp. 1687–1703, 2022.
- [17] T. Haarnoja, A. Zhou *et al.*, “Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor,” in *International conference on machine learning*. PMLR, 2018, pp. 1861–1870.
- [18] R. Eletreby, H. Rahbari *et al.*, “Supporting phy-layer security in multi-link wireless networks using friendly jamming,” in *2015 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2015, pp. 1–6.
- [19] M. Adams and V. K. Bhargava, “Using friendly jamming to improve route security and quality in ad hoc networks,” in *2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2017, pp. 1–6.
- [20] T. S. Rappaport, “Mobile radio propagation: Large-scale path loss,” *Wireless communications: principles and practice*, pp. 107–110, 2002.
- [21] T. M. Cover, *Elements of information theory*. John Wiley & Sons, 2006.
- [22] A. D. Wyner, “The wire-tap channel,” *Bell system technical journal*, vol. 54, no. 8, pp. 1355–1387, 1975.
- [23] J. P. Vilela, M. Bloch *et al.*, “Wireless secrecy regions with friendly jamming,” *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 2, pp. 256–266, 2011.
- [24] F. Yang, J. Wang *et al.*, “Physical-layer security for indoor VLC wiretap systems under multipath reflections,” *IEEE Transactions on Wireless Communications*, vol. 21, no. 12, pp. 11 179–11 192, 2022.
- [25] M. Zhang, R. Xue *et al.*, “Secrecy capacity optimization in coordinated multi-point processing,” in *2013 IEEE International Conference on Communications (ICC)*, 2013, pp. 5845–5849.
- [26] “Soft actor-critic (sac).” [Online]. Available: <https://spinningup.openai.com/en/latest/algorithms/sac.html>