BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **22**, No 4
Sofia • 2022 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2022-0039

A Model-Free Cognitive Anti-Jamming Strategy Using Adversarial Learning Algorithm

Sudha Y., Sarasvathi V.

Department of Computer Science and Engineering, PESIT-Bangalore South Campus-Bangalore and Affiliated to Visvesvaraya Technological University, Belgavi, Karnataka, India
E-mails: sudhasohan@gmail.com sudha.y@presidencyuniversity.in sarsvathiv@pes.edu

***Abstract*:** *Modern networking systems can benefit from Cognitive Radio* (*CR*) *because it mitigates spectrum scarcity. CR is prone to jamming attacks due to shared communication medium that results in a drop of spectrum usage. Existing solutions to jamming attacks are frequently based on Q-learning and deep Q-learning networks. Such solutions have a reputation for slow convergence and learning, particularly when states and action spaces are continuous. This paper introduces a unique reinforcement learning driven anti-jamming scheme that uses adversarial learning mechanism to counter hostile jammers. A mathematical model is employed in the formulation of jamming and anti-jamming strategies based on deep deterministic policy gradients to improve their policies against each other. An open-AI gym-oriented customized environment is used to evaluate proposed solution concerning power-factor and signal-to-noise-ratio. The simulation outcome shows that the proposed anti-jamming solution allows the transmitter to learn more about the jammer and devise the optimal countermeasures than conventional algorithms.*

***Keywords*:** *Cognitive radio network, Reinforcement learning, Smart jammer, Intelligent anti-jamming, Adversarial learning.*

1. Introduction

The Cognitive Radio Network (CRN) is a type of wireless radio network that has been introduced as a frictionless way to resolve the struggle between a restricted spectrum resource pool and a growing demand for it [1-2]. This technology explores the opportunities for non-interference use of a licensed spectrum bands and allows the Secondary Users (SUs) to select and use the ideal channels, and release those occupied channels whenever required by Primary Users (PUs). In the field of CRN, a spectrum sensing, and its sharing have been the subject of a great research [3]. It is important to note that, though the above approaches could improve spectrum efficiency, they depend upon the premise where the SUs utilize the unused spectrum when the PUs are not using it, and they coordinate with one another to accomplish this objective [4]. This consideration, however, often ignores the possibility that, the SUs are susceptible to malicious activity that can jam the communication channels

of CRNs. For CRN to be fully deployed at large scale, it is crucial to provide a secure mechanism in spectrum sensing and sharing [5]. A variety of work has been carried out on the topic of anti-jamming and jamming schemes as discussed in [6]. The traditional approaches in the context of anti-jamming communication, mostly rely on pure signal processing methods, which presents significant performance limitations [7, 8]. Although in the past recent years jamming issues in CRNs are being addressed by machine learning techniques, game theory method, and Markov decision processes [9, 10]. One of the most common applications of machine learning in robotics is Reinforcement Learning (RL) technology [11]. Recent studies have examined the use of RL technology as an anti-jamming solution in CRN due to its decision-making capabilities. This paper mainly discusses the anti-jamming communication based on RL. Following this, the next sub-section discuss the background and potential research issues of the existing works in relation to the context of the proposed research.

1.1. Related work

The selection of related work is based on user-desired keywords to summarize the study background approximately 35 research papers published between 2015 and 2022 have been found to be relevant with the given keyword, and the top 20 closely related papers (Fig. 1) under consideration are selected.

There are many works on anti-jamming carried out based on game theory. For example, the work carried out by [12] is an anti-jamming strategy based on a Stackelberg game approach in which the transmitter is a leader and the jammer is a follower, where both choose their power strategies and determine their optimal strategies. The authors formulate a hierarchical power control mechanism for the anti-jamming process. A similar approach is applied in [13], where Stackelberg game approaches where are discussed to defend jamming by considering various adversarial characteristics. The problem of jamming in the Unmanned Aerial Vehicle (UAV) network is discussed in [14]. This study analyses the competitive relations between the transmitter and the jammer of UAVs and has been conducted using a Bayesian Stackelberg game. An iterative sub-gradient of Bayesian Stackelberg equilibrium is developed to address jamming situations. Most of the anti-jamming solutions are presented in both the power domain and spectrum domain, but in [15], a multi-domain anti-jamming technique is suggested using the Stackelberg game in the power domain, and channel selection with its state is presented in the spectrum domain. However, in the above approaches, when the jammer strategies switches dynamically, it cannot be tracked and counteracted in real-time. Therefore, an application of honeypot is adopted in [16] to resist intelligent or dynamic jammers in CR. In the presented scheme, the attacker's strategy is passively learned from attacks in the past, and active decoy mechanisms are continuously adapted to prevent attacks on legitimate communications. Based on RL, the authors have developed an interference-aware cooperative anti-jamming scheme in [17]. As a result of using the Q-learning technique, users' optimal anti-jamming channel strategies are determined in a distributed manner by means of cooperation, decision feedback, and adjustment mechanisms. The researchers in [18] have adopted a Markov decision process

framework to develop the anti-jamming system. Anti-jamming strategies against sweep jammers are determined by using a collaborative multi-agent approach. In [19], deep RL is applied with a model-free method that does not require previous environmental knowledge. Initially, deep learning is introduced using the sliding window principle to identify jamming patterns. On the other hand, RL is used to select channels in real-time based on recognized jamming patterns.

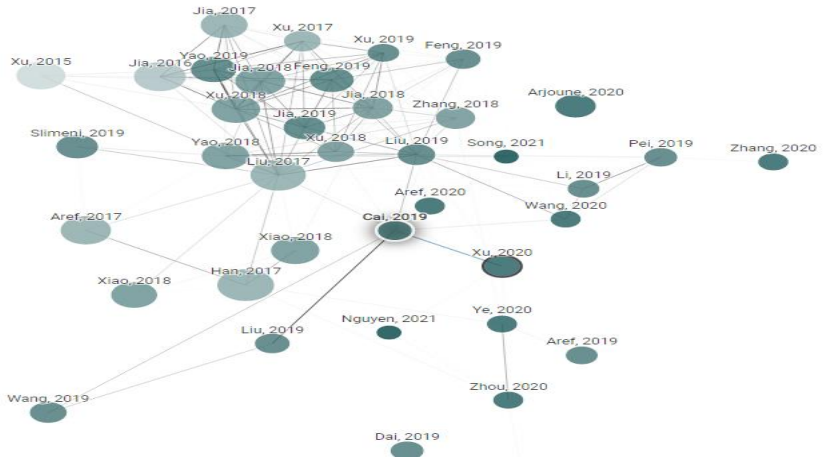

Fig. 1. Citation network of research works related to CRN, jamming, RL and anti-jamming

According to [20], an anti-jamming process can be formulated as a dynamic game with at least one pure strategy under Nash equilibrium. The solution is found by jamming-free channel selection by stochastic learning. The authors in [21] formulate the anti-jamming process as a Stackelberg game, in which the user acts as a leader, and the disruptor chooses a power strategy and then acts as a follower. As an anti-jamming process, a joint optimization algorithm has been developed using the Q-learning algorithm. Similarly. [22] describes a Bayesian three-layer Stackelberg game in which the primary user serves as the leader, the relay user serves as the sub-leader, and the disruptor serves as the followers. In order to obtain the Stackelberg equilibrium, a multi-user hierarchical iterative algorithm is presented using the backward induction method. The power and frequency selection process are optimized using deep-learning methods in [23], when it comes to anti-jamming communication. As reported in [24], the authors have developed a jamming pattern recognition method based on spectrum waterfall containing noisy signals and deep learning to recognize jamming signals in a short time. For detecting radar jamming, Convolutional Neural Networks (CNNs) have been applied in [25] where radar jamming is considered. The anti-jamming as Markov Decision Process (MDP) is presented in [26]. According to authors, an intelligent anti-jamming solution could utilize double Deep Q-learning Network (DQN), an application of deep RL to learn a policy, which maximizes the success rate of transmitters in transmitting. The method presented in [27] reduces jammer's sense probability by hiding anti-jamming techniques. In order to determine the hiding effect of the user’s actions, a deep RL algorithm-based action correlation is used. The aim of [28] is to obtain an optimal

multi-agent anti-jamming algorithm to make online channel selection, so that the sensor nodes can avoid internal mutual interference and effectively tackle external malicious jamming. An efficient communication policy is proposed in [29] that includes channel access and transmission power for making sense of different jamming scenarios. In [30], authors present an adversary modeling and analysis system that uses reward function design to model and analyse adversaries under realistic communication scenarios. In the articles [31, 32], authors have introduced a customized environment developed using libraries from Open-AI gym's benchmarked tool to assess RL's performance. Furthermore, the authors used a deep deterministic policy gradient to design a solution that have learned how to avoid jamming by learning frequency band selection strategies. However, there are several approaches to anti-jamming in CR and wireless networking. But there are still some important issues that need to be resolved. The next sub-section highlights potential issues discovered based on a review of the related work.

1.2. Research problem

There are varieties of solutions that claim to mitigate potential jamming attacks based on simulation statistics in the existing literature on anti-jamming schemes. However, the scope of the existing techniques may not be sufficient to cope with a situation in which the jammer changes strategies dynamically and precariously. Creating an efficient technology is far more challenging than implementing a jammer in the communication channel of a wireless network system. A better solution can be achieved with appropriate exploitation of advanced learning methodologies to cope with jammers' dynamic activities. Following are some potential research issues that can serve as an important direction for researchers in developing better solutions.

- Anti-jamming approaches in literatures have a restricted application when jamming is conducted across the entire frequency spectrum. In addition to being able to withstand narrowband jamming attacks, frequency hopping-based solutions is also subjected to significant drops in spectral efficiency.
- The previous works using game theory approach to anti-jamming is also not suitable for coping with the dynamic strategy of jammers. Similar to this, the schemes based on retransmission protocol can regain communication channels, but at the cost of deteriorate performance.
- There is an unrealistic assumption that game-theory based approaches adopt prior knowledge regarding jamming strategy and channel information. The existing studies frequently adopt a similar design strategy, which results in the lack of novelty in the work as a whole.
- The anti-jamming solutions using Q-learning are not suitable for continuous state and action space. As the size of the Q-table increases, learning process becomes slow. While solution based on DQN are subjected to slow convergence and are unstable when introduced dynamic and complex scenario.
- Most of the RL-based anti-jamming solution in literature are not benchmarked with suitable environment, which must be customized and designed with the packages provided by Open-AI gym tool kit.

- Moreover, the existing work lacks a trade-off between communication efficiency and anti-jamming robustness.

In order to address these research issues, the next sub-section discusses the proposed solution, system design and methodology adopted.

1.3. Proposed solution

The proposed research work is an extension of our previous work [31] where a DDGP based RL agent is trained and used as an anti-jamming solution against sweep jammer. The prime aim of this work is to enhance the capability and efficiency of our intelligent ant-jamming scheme to exploit better policy towards learning pattern of smart jammer and cope up with the scenario where the jammer changes its strategy dynamically and uncertainly. Inspired by the framework presented in the literature, we address a more general channel model introduced in [24] where, the study proposes two different jamming attackers, namely a Feed-forward Neural Network (FNN) attacker and a Deep Reinforcement Learning (DRL) attacker, to perform the jamming attacks on a user performing dynamic multichannel access using a DRL agent itself [12]. Since the rise of security-critical applications securing a real-world wireless communication technologies such as Bluetooth, Wi-Fi, and cellular technologies like 3G, 4G, and 5G demands an intelligent design in the anti-jamming techniques by considering the constraints associated with the network. In this regard, the proposed system presents an efficient and intelligent jamming scheme that can achieve an adequate balance between communication efficiency and anti-jamming capabilities.

This paper presents an adversarial machine learning approach to launch jamming attacks on wireless communications and introduces a defence strategy. The proposed research study presents a mechanism of adversarial learning, i.e., whenever one neural network works against another neural network such that, both networks are improvising their policy against each other as an adversarial learning method. The jammer used in the proposed scheme is a smart jammer designed on application of RL algorithm. The study does not consider the power consumption of jammer or any other parameter because here the focus is on cognitive network. During training process, the cognitive radio (anti-jamming agent) has been trained with the smart jammer agent but while testing it is tested against a sweep jammer/jamming. Since, it has already seen the better policy, so it will naturally perform better against sweep jammer.

The following are the significant contribution of the proposed study:

- This study extends our previous work where an anti-jamming solution has been proposed against sweep jammer. In this study both anti-jamming and jamming process are proposed based on the RL algorithm with an application of adversarial learning.
- The modelling of smart jammer and intelligent anti-jamming adopts DDPG algorithm where the agents can adopt the dynamics of the sub-channels and the strategies of each other for deriving better policy against each other.
- The study considers the customized environment build on Open-AI Gym function and imitates scenario of CRN with jammer and SUs.

2. Methodology

This section discuss about the methodology adopted in the formulation of RL driven anti-jamming task, RL driven jammer and its evaluation on the customized environment. Further, an adversarial learning process is discussed to avoid a jamming in the network. The proposed anti-jamming solution is a model free method as its action is not dependent on the prior information about the jammer and communication. In the proposed study both agents, i.e., smart jammer and anti-jammer adopts off-policy DDGP algorithm [32] as proposed in [31], and offers a better scope of determining policies against each other. In this way, the anti-jamming algorithm exposed to complex and more diverse jamming scenario for learning better strategy to defend potential jamming attacks. Another significant contribution of the proposed system is the adoption of customized environment proposed in our previous work [31] suitable for assessing genuine performance of the agent or learning algorithm.

2.1. Problem formulation

The anti-jamming task can be considered as Markov game [26], which is derived from the MDP in a multi-agent scenario where both jamming and anti-jamming process are driven by RL algorithm. Markov game process can be numerically expressed as $\mathcal{M} = \{S, A_1, A_2, A_3, \cdots, A_N, t, r_1, r_2, r_3, \cdots, r_N\}$, where S is the set of states, A_n denotes the set of actions taken by agent for $n = 1, 2, 3, \cdots, N$, t is the state transition model, and r_n denotes reward given to agent for its each action for $n = 1, 2, 3, \cdots, N$. With respect to [31], state refers to frequency spectrum sensed and observed by the agent. The state s can be numerically expressed as, $s \in S = \{a, f_{j_x}\}$, where $a = \{a_1, a_2, a_3, \cdots, a_N\}$ denoting adversarial action profile and, f_{j_x} is jamming signal. The agent gets reward for its each action where the reward factor depends on the quality of channel selection and power cost factor of a channel switching process.

2.2. Environment

Modelling a suitable environment is crucial to utilize the features of an RL algorithm. The environment refers to the simulation of a given task or problem where the agent interacts with an environment and perceives the states and tries to solve the problem. The environment considered in the current work is quite similar to our previous work [31] developed using functions of benchmarked toolkit namely Open AI Gym, where RL driven learning algorithm as an anti-jamming against the rule-based jammer was proposed. But in the current work, instead of rule-based jammer, the smart or intelligent jammer is considered. The schematic architecture of the environment is as shown in Fig. 2. The environment mimics the scenario of CRN with jammer, and a user having transmitter and receiver module for the communication.

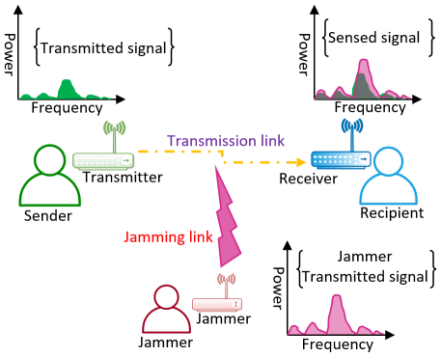

Fig. 2. Environment for CRN communication scenario

As shown in Fig. 2, the proposed environment assumes that there is a frequency synchronization between all SUs for the sake of simplicity. Every SUs explores the available sub-band channels and starts the transmission once a spectrum hole is detected. The environment that mimics the CRN, also has PUs, SUs and a smart Jammer (J) in the network, where each Channel (C) has different capacity with Bandwidth (B). The environmental modelling also considers that the jamming attack is launched only by J by producing noise or interference, which deteriorates the transmission between the SUs in the communication channel of same band. The SUs are equipped with Transmitter (T_x) as a sender, and Receiver (R_x) as recipient for coordination. The study also assumes that there are no other sources of noise or interference such as effect of the multipath fading. For simulation the communication model in environment is designed by considering a Wi-Fi communication channels with BPSK as modulation technique.

In communication scenario, jammer J executes interfering signals continuously such that, simultaneously an intelligent anti-jammer mechanism directs user (SUs) to choose optimal channel by overcoming the interference effect on the communication channel of interest. Here, both smart jammer and user share the same continuous-time signal. The transmission channel frequencies as $f = \{f_1, f_2, f_3, \cdots, f_N\}$ are consider from a pool of frequencies in a communication band to communicate with the other user recipient (SU) with power factor, associated with a channel. In the case of jamming attack, the jammer J chose the channel frequency (f_J) of a same band randomly to obstruct the transmission between users. The success of jamming attack depends on the factor that the power of jamming signal must be greater than power of the received user signals, which is numerically expressed as follows:

$$P_\mathrm{J}(f_\mathrm{J}) > P_\mathrm{u}(f_\mathrm{u}), \tag{1}$$

where, P_J denotes power factor of jammer frequency f_J, and P_u is the power of user frequency at receiver side. Along with this, the study also considers additional factor namely, Bandwidth (B) of the channel, identical for both f_u and f_J such that:

$$\mathrm{B}(f_\mathrm{u}) \cong \mathrm{B}(f_\mathrm{j}). \tag{2}$$

Under the above interpretation, if a user communicates or transmits data to another user in the same communication channel chosen by the J, then the SNR of the

transmission signal of user on receiver side is despoiled severely, and numerical expressed as follows:

$$\text{SNR} = \frac{P_{\text{u}} \cdot g_{\text{s}}}{P_{\text{J}} g_{\text{J}} \boldsymbol{F}(B(f_{\text{u}}) \cong B(f_{\text{j}}))}. \tag{3}$$

Here, g_{s} and g_{j} refers to channel gain of user signal and channel gain of jammer signal (i.e., from sender to receiver), and $\boldsymbol{F}(B(f_{\text{u}}) \cong B(f_{\text{j}}))$ is a function that characterize the probability distribution of the received signal based on certain condition as follows:

$$\begin{cases} 1 & \text{if } B(f_{\text{u}}) \cong B(f_{\text{j}}) \text{ is True,} \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

The DDPG agents for jamming and anti-jamming task, gets rewards for their action in the environment respectively, which depends on the factor of channel selection and its associated power cost. The agent for jamming task tries to maximize its reward by jamming signal and the agent for anti-jamming task tries to attain successful communication with a minimal power cost associated with channel switching. The mathematical equation for deciding reward factor we use the same equation mentioned in [31] as follows:

$$R_{\text{w1}} = R_{\text{SNR}}(A_{\text{t}}) - c(A_{\text{t}}), \tag{5}$$

where, $R_{\text{SNR}}(A_{\text{t}})$ refers to reward given to agent action (anti-jamming) for all successful transmission and c is communication cost factor associated with action concerning channel switching process. Under this interpretation the reward for a jammer agent (J) will be

$$R_{\text{w2}} = -1 \times R_{\text{w1}}. \tag{6}$$

Using above equations, we get

$$R_{\text{w}} = \begin{cases} 1 \text{ if } & \text{SNR}_{\text{R}_{\text{U}}} \geq \text{SNR}_{\text{cutoff}}, \\ -1 & \text{otherwise,} \end{cases} \tag{7}$$

where cutoff is demodulation threshold, when the SNR at receiver ($\text{SNR}_{\text{R}_{\text{U}}}$) is lower than this threshold then the data transmission fails, therefore the agent for jamming process will get positive reward while agent for anti-jamming process will get negative reward.

2.3. Adversarial learning

Based on the exploration of environment, the agent driven anti-jamming algorithm sense the state and gets the observation about communication channel, and the jammer. The objective of the anti-jamming algorithm is to guide SUs to carefully choose a sub-channel to maximize its spectrum utilization by avoiding the jam. On the other hand, the intelligent jammer aims to forbid the SU from effective channel utilization by a strategic jamming approach. The objectives of the two agents, namely the intelligent jammer and the intelligent anti-jamming, are opposite to each other. Therefore, the dynamic interaction between them is well formulated as an adversarial learning or learning action policy against each other, where the gain of one agent is the loss of another agent.

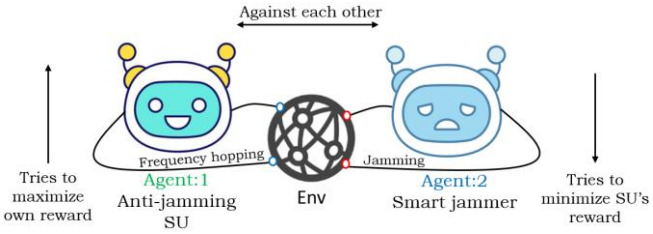

Fig. 3. Environment for CRN communication scenario

As shown in Fig. 3, the interaction between both agents is non-cooperative. The Agent-1 learns action policy against Agent-2 and tries to maximize its reward by providing jamming free channel. On the other hand, the Agent-2 tries to minimize the reward value for Agent-1 by building a new action policies towards jamming the signal. Here both agents are intelligent and will exhibit rational behaviour to maximize their own rewards according to their individual actions. The key entities for learning algorithms are highlighted as follows:

- Agents: There are two agents viz i) the SU (anti-jamming), and ii) smart jammer (J).
- States: It is the representation of environment and in the proposed case study sub-bands are the state.
- Action: It is an optimal decision taken by both agents. In the proposed study, the action of SU is, changing the sub-band frequency and selection of a jamming free channel, while action of jammer is to jam the frequency band.
- Reward: It is the feedback to the agent for corresponding actions. In our case, the reward can be positive and negative depending on the achievement of the agents.

Algorithm 1. Intelligent Anti-Jamming

Input: E(Episodes), I(Iteration),α(Learning rate),γ(Discount), B(Batch size)

Output: Selection of jamming free signal

Start

Step 1. $\forall\ s \in S,\ a^{u} \in A(s)$ *and* $a^{j} \in A(s)$

Step 2. Initialize θ randomly for current network $[\mu,\ Q]$

Step 3. Initialize θ randomly for target network $[\mu',\ Q']$

Step 4. **While** episode → 1:E **do**

Step 5. Initialize ξ for action a^{u} and get initial state s

Step 6. **For** i = 1: I **do**

Step 7. Choose action a^{u} → $\mu(s)$

Step 8. Perform frequency hopping // perceive communication channel

Step 9. Select channel $i \in C$ and P_{u} according to the current policy

Step 10. Action a^{u} → initiate transmission

Step 11. Get reward $R(s, a^{u})$ using Eq. (4)

Step 12. Get next state (s+1)

Step 13. Get Tuple $[s, a^{u}, R, s+1]$ into experience pool ψ

Step 14. If $|\psi| \geq$ B **do**

Step 15. Randomly select B from ψ
Step 16. **End if**
Step 17. **For** each experience in B **do**
Step 18. Compute value for μ' and Q' value using Eq. (8)
Step 19. Update $Q_\theta(s, a^u)$ by minimizing the loss $\mathcal{L}(\theta)$ using Eq. (9)
Step 20. Update the $\mu(s)$ based on policy gradient using Eq. (11)
Step 21. Update μ' and Q' using Eq. (13) & (14)
Step 22. **End for**
Step 23. **End for**
Step 24. $s \leftarrow s+1$ and $a^u \leftarrow a^{u+1}$
Step 25. **End while**
Step 26. **If** $a = a^j$ **then**
Step 27. Signal jammed, and Smart Jammer gets reward
Step 28. else
Step 29. Transmission successful, agent rewarded
Step 30. **End if**
End

The entire mechanism of intelligent anti-jamming agent is as discussed in the above Algorithm 1. Since the development of agent is carried out using DDPG, the algorithm takes necessary input as episodes (E), number of iterations (I), learning rate $\alpha \in [0\ 1]$, discount rate $\gamma \in [0\ 1]$, and a batch size B and returns a jamming free transmission with a higher bandwidth utilization after processing all the steps. For all set of states $s \in S$, SU action $a^u \in A(s)$ and jammer action $a^j \in A(s)$, the algorithm initializes weights (θ) randomly for the current network consisting two parameterized function approximator's in DDPG algorithm namely actor (μ) and critic (Q). Similarly, in the next step, the algorithm initializes the weights (θ) randomly for the target network which is copy of current network to maintain consistency in the training process. The actor is responsible for determining optimal policy for an action (a) and the critic generates Q-value that characterize the goodness of action (Steps 1-3). Further, an initialization of random process (ξ) for an agent to take action (a^u) is to observe the initial state(s) and by interacting with the environment. The agent while interacting with the environment performs hopping, and perceives channel information to select the channel (C) for the transmission according to the current policy with user transmission power (P_u). The agent gets positive reward (Equation (4)) if the channel state is in favourable condition or jamming-free otherwise it will be penalized with negative reward value (Steps 4-11). In the next step the agent continues to explore the environment and get the new state (s+1) and takes action accordingly (Step 12). Here, the agent with its frequency hopping strategy always tries to learn better policy against jammer strategy and takes appropriate action to maximize its cumulative reward value. Therefore, for each iteration, i.e., the agents get new state stored in a vector M and similarly, the corresponding actions is also stored into a vector N by denoting the index of channels which are in good state. Since, this study has adopted DDPG algorithm, the algorithm maintains two networks, i.e., the current actor-critic and target actor-critic network.

The training of current network needs an experience pool which is a tuple consisting set of state, corresponding action, reward and new state (Step 13).

In the next step, the algorithm maintains an experience pool ψ to determine an optimal action value denoted as $Q^*(s, a^{\mathrm{u}})$. Then, sampling is done to update current network $[\mu, Q]$ parameters using mini-batch size (B) of transitions from the pool ψ (Steps14, 15). This process will lead to compute error and update $[\mu, Q]$ parameters. To stabilize the performance of the agent and to learn the anti-jamming policy (Step 18), the algorithm computes the target network value $[\mu', Q']$ using Bellman equation numerically expressed as follows:

(8) $$y_k = R + \gamma Q'((s, \mu'(s)).$$

Then, further steps (Step 19) are subjected to updating parameters of current network $[\mu, Q]$ by computing loss function as Mean Square Error (MSE). mathematically given as follows:

(9) $$\mathcal{L}(\theta) = \frac{1}{B}\sum_k (y_k - Q(s, a^{\mathrm{u}}))^2.$$

In the above Equation (9), y_k is the updated action-value of target network and $Q_\theta(s, a^{\mathrm{u}})$ is the updated action-value of critic network. But in case of jamming, the agent as a jammer also adopts the same strategy against anti-jamming agent to jam the SU with maximum action-value given as follows:

(10) $$\mathcal{L}(\theta) = \frac{1}{B}\sum_k \left(y_k - Q(s, a^{\mathrm{j}})\right)^2.$$

Further, in the next step of the algorithm (Step 20), an iterative optimization algorithm (gradient descent) is used to update the weights θ of the current Actor towards determining appropriate A_{t} decision strategy that maximizes the R_{w} numerically expressed as follows:

(11) $$\nabla_\theta J(\theta) = \frac{1}{B}\sum_k \nabla_\theta \mu_\theta(s) \times \nabla_{a^{\mathrm{u}}} Q_\theta(s, a^{\mathrm{u}})|_{a^{\mathrm{u}}=\mu(s)}.$$

In the above numerical expression, the current-Actor network is updated with the average of the sum of gradients in an off-policy manner with B of transitions. Similarly, the updated rule for jammer is given as follows:

(12) $$\nabla_\theta J(\theta) = \frac{1}{B}\sum_k \nabla_\theta \mu_\theta(s) \times \nabla_{a^{\mathrm{j}}} Q_\theta(s, a^{\mathrm{j}})|_{a^{\mathrm{j}}=\mu(s)}.$$

Finally, the target Actor and target Critic network is updated (Step 21) to provide better stability in the learning process of the current Actor and Critic network numerically expressed as follows:

(13) $$\mu' \leftarrow \emptyset\theta + (1 - \emptyset)\mu',$$

(14) $$Q' \leftarrow \emptyset\theta + (1 - \emptyset)Q',$$

where $\emptyset$ refers to the hyperparameter ranging between $[0, 1]$. In this operation, a sliding averaging mechanism is used to update the parameters of the target network once per update of the current network. As a result, the target network slowly tracks the learned networks, thereby significantly improving stability in learning the action values. The Algorithm 2 discuss the procedure adopted for a smart jammer. An anti-jamming agent is designed to guide the SU in choosing a sub-channel carefully, based on the knowledge of a channel, the system, and the attacker, to maximize spectrum utilization and prevent jamming. By using a strategic jamming approach, jammers prevent the SUs from using channels effectively.

Algorithm 2. Smart Jammer

Step 1. $\forall\ s \in S,\ a^{u} \in A(s)$ *and* $a^{j} \in A(s)$
Step 2. Initialization θ randomly for $[\mu,\ Q]$ and $[\mu', Q']$
Step 3. **While** episode → $1{:}E$ **do**
Step 4. Initialize ξ for action a^{j} and get initial state s
Step 5. **For** i = 1: I **do**
Step 6. Choose action a^{j} → $\mu(s)$
Step 7. Select channel $i \in C$ with P_{j}
Step 8. Action a^{u} → initiate jamming
Step 9. Get reward $R(s, a^{u})$ using Eq. (8)
Step 10. Get next state (s+1)
Step 11. Get Tuple $[s,\ a^{j},\ R,\ s+1]$
Step 12. If $|\psi| \geq B$ **do**
Step 13. Randomly select B from ψ
Step 14. **End if**
Step 15. **For** each experience in B **do**
Step 16. Compute value for μ' and Q' value using Eq. (8)
Step 17. Update $Q_{\theta}(s, a^{j})$ using Eq. (10)
Step 18. Update $\mu(s)$ using Eq. (12)
Step 19. Update μ' and Q' using Eq. (13) & (14)
Step 20. **End for**
Step 21. **End for**
Step 21. s ← $s+1$ and a^{j} ← a^{j+1}
Step 21. **End while**
Step 21. **If** $a^{j} = a$ **then**
Step 21. Signal jammed, and Smart Jammer gets reward
Step 21. else
Step 21. Transmission successful, agent rewarded
Step 21. **End if**
End

3. Results and discussion

The proposed system is designed and executed using Python with Anaconda distribution installed on Windows 11 64-bit Intel Core i7 with 16 GB of RAM. This section presents the results and performance analysis of the proposed system by comparing it with other RL algorithms in terms of power cost and SNR. The study consider, Q-learning, DQN and an anti-jamming learning algorithms (DDPG) as proposed in [31]. The simulation parameters used in system performance are highlighted in Table 1.

Table 1. Simulation Parameters and its values

Parameters	Value
Wi-Fi Frequency band	2.4 GHz
Number of communication channels	11
Jamming Model	Smart Jammer
Jamming power	30 dB.m
Transmitter signal power	25-45 dB.m
Bandwidth of Transmitter signal	20 MHz
Bandwidth of Jamming signal	20 MHz
Demodulation cutoff	10 dB
Data rate	2 Mbps
Digital modulation technique	BPSK
Channel switching Cost	0.2
Discount factor	0.96
Minibatch size	32

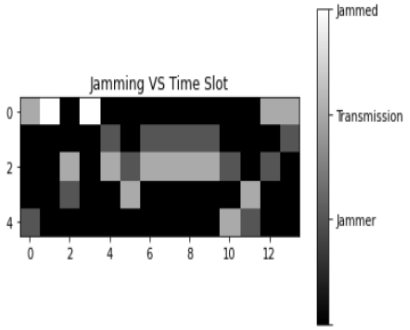

Fig. 4. Illustration of smart jammer and SU hopping with time-frequency analysis

Based on the time-frequency analysis of hopping is depicted in Fig. 4, SUs and smart jammer consistently expect to achieve their goals with their sub-band channels. Jammers continuously anticipate their own sub-band channels to achieve their purpose, so the SU has to anticipate the other available sub-channels to avoid jamming. Grey blocks represent transmission, while dark grey blocks represent the smart jammer's selected channel. Meanwhile, a white block indicates that the signal has been jammed. Using the jammer-selected channel, the user will be unable to transmit data packets because the SNR of the received signal is impaired. In an unknown environment, determining an optimal strategy is a challenge task for the agent. It is necessary that the agent understands the communication spectrum and interacts with the environment in order to help the SU to select the unobstructed transmission channel without interfering the system performance in an unknown environment.

In Fig. 5, the system performance is analysed with respect to the power cost (milli-watt) exhibited by different RL algorithm. Based on the graph trend, it can be observed that the proposed adversarial learning algorithm exhibits lower power expenditure during transmission and is more consistent than other RL algorithms over the progressive time slots. As previously discussed in [31], Q-learning techniques suffer from the issue of dimension disaster since the environmental conditions and action space are continuous, and Q-learning finds an optimal policy using Q-tables

which are discrete by nature. With Q-learning, the learning algorithm use a Q-table to determine the optimal action, so the frequency has to change constantly, consuming more power with more frequency hopping.

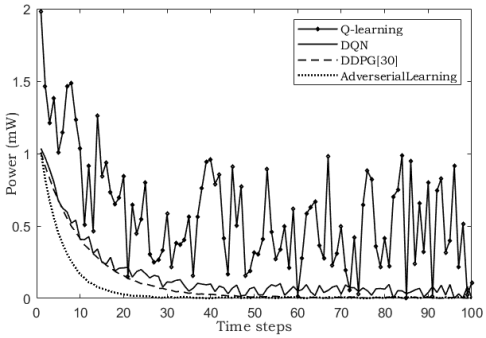

Fig. 5. Analysis of the power cost

The performance of DQN and DDPG appear to be similar, with only small differences. As DQN uses neural networks to optimize action-value for a given task, it is more efficient than Q-learning, but as compared to DDPG, its performance is not as stable. As opposed to Q-learning and DQN, DDPG uses two neural networks called actor and critic that makes it more suitable to continuous action space and less random. In the proposed adversarial learning algorithm, two DDGP algorithms are trained against each other, where one is anti-jamming, and the other is smart jammer, learning from each other to develop better policies. Therefore, the proposed adversarial learning algorithm achieves less fluctuation and uses less power.

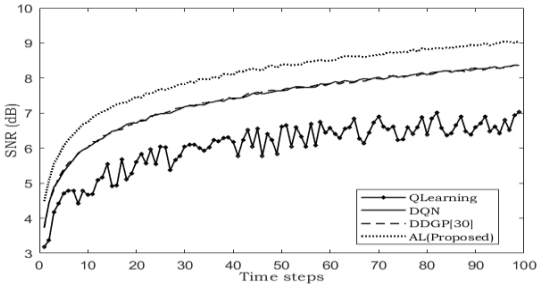

Fig. 6. Analysis of the SNR

A performance analysis of adversarial learning is shown in Fig. 6. The SNR factor determines how many packets are received over how many packets are transmitted. Basically, there are only two outcomes: either transmission is successful or it is not successful due to jamming. Comparing the graph trend with other RL algorithms over progressive time slots, the proposed adversarial learning algorithm has higher SNR. DQN and DDPG both have similar SNR trends, but they outperform Q-learning. The interesting aspect here is the higher SNR in the proposed system. According to the proposed adversarial learning mechanism, two neural networks compete and learn from each other. An anti-jamming agent builds a better policy to beat smart jammer agents, while the jammer tries to beat the anti-jamming agent. In

contrast, the proposed algorithm will build a better policy if the jammer is static, and if the jammer is dynamic, the policy will continue to improve.

4. Conclusion

In this paper, the proposed study has explored the effectiveness of adversarial learning scheme in conjunction with multi-agent environment for defending jamming attacks in CRN. Both jammer and anti-jamming mechanism have been implemented with DDPG a dual neural network system making them intelligent with the self-exploration capabilities. In addition, a customized environment is adopted which is developed using functions of Open-AI Gym for the adequate assessment of learning agent algorithm. The study has devised a non-cooperative frequency-hopping system in which, the agent an intelligent anti-jamming mechanism, and the smart jammer are under adversarial learning and builds a policy against each other to meet their objective. The prime ideology behind devising adversarial learning scheme is to train an intelligent anti-jamming agent in such a way that it exposes to more dynamic situation given by smart jammer and learn multi-faceted recovering strategy to help secondary user to avoid jamming and, hence enhances the bandwidth utilization of user. The effectiveness and scope of the proposed adversarial learning driven intelligent frequency hopping mechanism is confirmed through the simulation results. The outcomes shows that the presented scheme outperforms the other existing reinforcement-learning algorithms and is capable of intelligently selecting jamming free sub-channels, while avoiding effects of potential jammer on the transmission.

References

1. Alias, D. M. Cognitive Radio Networks: A Survey. – In: Proc. of International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET'16), IEEE, 2016, pp. 1981-1986.
2. Saleem, Y., M. Husain Rehmani. Primary Radio User Activity Models for Cognitive Radio Networks: A Survey. – Journal of Network and Computer Applications, Vol. **43**, 2014, pp. 1-16.
3. Salahdine, F., N. Kaabouch. Security Threats, Detection, and Countermeasures for Physical Layer in Cognitive Radio Networks: A Survey. – Physical Communication, Vol. **39**, 2020, 101001.
4. Hu, F., B. Chen, K. Zhu. Full Spectrum Sharing in Cognitive Radio Networks Toward 5G: A Survey. – IEEE Access, Vol. **6**, 2018, pp. 15754-15776.
5. Sudha, Y., V. Sarasvathi. An Intelligent Anti-Jamming Mechanism against Rule-Based Jammer in Cognitive Radio Network. – International Journal of Advanced Computer Science and Applications, Vol. **13**, 2022, No 3.
6. Bouabdellah, M., N. Kaabouch, F. El Bouanani, H. Ben-Azza. Network Layer Attacks and Countermeasures in Cognitive Radio Networks: A Survey. – Journal of Information Security and Applications, Vol. **38**, 2018, pp. 40-49.
7. Jasim, D. K., S. B. Sadkhan. Cognitive Radio Network: Security and Reliability Trade-off-Status, Challenges, and Future Trend. – In: Proc. of 1st Babylon International Conference on Information Technology and Science (BICITS'21), IEEE, 2021, pp. 149-153.
8. Arjoune, Y., N. Kaabouch. A Comprehensive Survey on Spectrum Sensing in Cognitive Radio Networks: Recent Advances, New Challenges, and Future Research Directions. – Sensors, Vol. **19**, 2019, No 1, 126.

9. K a u r, A., K. K u m a r. A Comprehensive Survey on Machine Learning Approaches for Dynamic Spectrum Access in Cognitive Radio Networks. – Journal of Experimental & Theoretical Artificial Intelligence, Vol. **34**, 2022, No 1, pp. 1-40.
10. G a n e s h B a b u, R., V. A m u d h a. A Survey on Artificial Intelligence Techniques in Cognitive Radio Networks. – In: Emerging Technologies in Data Mining and Information Security. Singapore, Springer, 2019, pp. 99-110.
11. W a n g, Y., Z. Y e, P. W a n, J. Z h a o. A Survey of Dynamic Spectrum Allocation Based on Reinforcement Learning Algorithms in Cognitive Radio Networks. – Artificial Intelligence Review, Vol. **51**, 2019, No 3, pp. 493-506.
12. J i a, L., F. Y a o, Y. S u n, Y. X u, S. F e n g, A. A n p a l a g a n. A Hierarchical Learning Solution for Anti-Jamming Stackelberg Game with Discrete Power Strategies. – IEEE Wirel. Commun. Lett., Vol. **6**, 2017, No 6, pp. 818-821. DOI: doi:org/10.1109/lwc.2017.2747543.
13. J i a, L., Y. X u, Y. S u n, S. F e n g, A. A n p a l a g a n. Stackelberg Game Approaches for Anti-Jamming Defence in Wireless Networks. – IEEE Wirel. Commun., Vol. **25**, 2018, No 6, pp. 120-128.
14. X u, Y., et al. A One-Leader Multi-Follower Bayesian-Stackelberg Game for Anti-Jamming Transmission in UAV Communication Networks. – IEEE Access, Vol. **6**, 2018, pp. 21697-21709.
15. J i a, L., Y. X u, Y. S u n, S. F e n g, L. Y u, A. A n p a l a g a n. A Multi-Domain Anti-Jamming Defense Scheme in Heterogeneous Wireless Networks. – IEEE Access, Vol. **6**, 2018, pp. 40177-40188.
16. B h u n i a, S., E. M i l e s, S. S e n g u p t a, F. V a z q u e z-A b a d. CR-Honeynet: A Cognitive Radio Learning and Decoy-Based Sustenance Mechanism to Avoid Intelligent Jammer. – IEEE Trans. Cogn. Commun. Netw., Vol. **4**, 2018, No 3, pp. 567-581.
17. X u, Y., G. R e n, J. C h e n, X. Z h a n g, L. J i a, L. K o n g. Interference-Aware Cooperative Anti-Jamming Distributed Channel Selection in UAV Communication Networks. – Applied Sciences, Vol. **8**, 2018, No 10, 1911.
18. Y a o, F., L. J i a. A Collaborative Multi-Agent Reinforcement Learning Anti-Jamming Algorithm in Wireless Networks. – IEEE Wireless Communications Letters, Vol. **8**, 2019, No 4, pp. 1024-1027.
19. L i u, S., Y. X u, X. C h e n, X. W a n g, M. W a n g, W. L i, Y. L i, Y. X u. Pattern-Aware Intelligent Anti-Jamming Communication: A Sequential Deep Reinforcement Learning Approach. – IEEE Access, Vol. **7**, 2019, pp. 169204-169216.
20. J i a, L., Y. X u, Y. S u n, S. F e n g, L. Y u, A. A n p a l a g a n. A Game-Theoretic Learning Approach for Anti-Jamming Dynamic Spectrum Access in Dense Wireless Networks. – IEEE Transactions on Vehicular Technology, Vol. **68**, 2018, No 2, pp. 1646-1656.
21. F e n g, Z., G. R e n, J. C h e n, C. C h e n, X. Y a n g, Y. L u o, K. X u. An Anti-Jamming Hierarchical Optimization Approach in Relay Communication System via Stackelberg Game. – Applied Sciences, Vol. **9**, 2019, No 16, 3348.
22. F e n g, Z., G. R e n, J. C h e n, X. Z h a n g, Y. L u o, M. W a n g, Y. X u. Power Control in Relay-Assisted Anti-Jamming Systems: A Bayesian Three-Layer Stackelberg Game Approach. – IEEE Access, Vol. **7**, 2019, pp. 14623-14636.
23. L i, Y., Y. X u, X. W a n g, W. L i, W. B a i. Power and Frequency Selection Optimization in Anti-Jamming Communication: A Deep Reinforcement Learning Approach. – In: Proc. of 5th IEEE International Conference on Computer and Communications (ICCC'19), IEEE, 2019, pp. 815-820.
24. C a i, Y., K. S h i, F. S o n g, Y. X u, X. W a n g, H. L u a n. Jamming Pattern Recognition Using Spectrum Waterfall: A Deep Learning Method. – In: Proc. of 5th IEEE International Conference on Computer and Communications (ICCC'19), IEEE, 2019, pp. 2113-2117.
25. L i u, Q., W. Z h a n g. Deep Learning and Recognition of Radar Jamming Based on CNN. – In: Proc. of 12th International Symposium on Computational Intelligence and Design (ISCID'19), Vol. **1**, IEEE, 2019, pp. 208-212.
26. X u, J., H. L o u, W. Z h a n g, G. S a n g. An Intelligent Anti-Jamming Scheme for Cognitive Radio Based on Deep Reinforcement Learning. – IEEE Access, Vol. **8**, 2020, pp. 202563-202572.
27. W a n g, Y., X. L i u, M. W a n g, Y. Y u. A Hidden Anti-Jamming Method Based on Deep Reinforcement Learning. – arXiv preprint arXiv:2012.12448, 2020.

28. Z h o u, Q., Y. L i, Y. N i u. Intelligent Anti-Jamming Communication for Wireless Sensor Networks: A Multi-Agent Reinforcement Learning Approach. – IEEE Open Journal of the Communications Society, Vol. **2**, 2021, pp. 775-784.
29. N g u y e n, P. K. H., V. H. N g u y e n. A Deep Double Q-Learning Based Scheme for Anti-Jamming Communications. – In: 2020 28th European Signal Processing Conference (EUSIPCO'21), IEEE, 2021, pp. 1566-1570.
30. D o n g, J., S. W u, M. S u l t a n i, V. T a r o k h. Multi-Agent Adversarial Attacks for Multi-Channel Communications. – arXiv preprint arXiv:2201.09149, 2022.
31. S u d h a, Y., V. S a r a s v a t h i. An Intelligent Anti-Jamming Mechanism against Rule-Based Jammer in Cognitive Radio Network. – International Journal of Advanced Computer Science and Applications, Vol. **13**, 2022, No 3.
32. L i l l i c r a p, T. P., J. J. H u n t, A. P r i t z e l, N. H e e s s, T. E r e z, Y. T a s s a, D. S i l v e r, D. W i e r s t r a. Continuous Control with Deep Reinforcement Learning. – arXiv preprint arXiv:1509.02971, 2015.

Received: 31.08.2022; Accepted: 20.10.2022