An Intelligent Anti-Jamming Mechanism against Rule-based Jammer in Cognitive Radio Network

Sudha Y, Dr. Sarasvathi V
Visvesvaraya Technological University
PESIT-Bangalore South Campus and affiliated to VTU, Belgavi, India

***Abstract*—Cognitive Radio Network (CRN) has become a promising technology to overcome the problem of insufficient spectrum utilization. However, the CRN is susceptible to the well-known jamming attack, which reduces its spectrum utilization efficiency. Existing jamming identification schemes and their countermeasure typically require prior statistical information about the communication channel and jamming pattern. This is quite an impractical assumption in the real context. The prime research problem is that the existing schemes are mainly associated with higher computational costs and communication overhead. Hence, the proposed manuscript presents a non-device-centric and efficient anti-jamming mechanism in the form of higher spectrum utilization driven by reinforcement learning techniques to address this above-stated problem. The proposed anti-jamming mechanism is modeled in two phases of implementation. First, the design of the customized environment is introduced as a single wideband cognitive-communication channel where a jammer signal sweeps transversely in the entire band of interest. Secondly, an intelligent agent is designed based on a model-free off-policy algorithm that operates over the same spectrum band. The agent uses its frequency-band knowledge discovery capability to learn frequency band selection and preference strategies to detect and avoid jamming signals, maximizing its successful transmission rate. The simulation results show that the proposed anti-jamming mechanism can effectively eliminate interference and is efficient in power usage and Signal to Noise Ratio (SNR) compared to other existing advanced algorithms.**

Keywords*—*Anti-jamming; agent; cognitive radio network; reinforcement learning

I. INTRODUCTION

Cognitive Radio (CR) is a communication system that perceives its environment and autonomously adjusts according to its radio operating parameters. It has been introduced to address the contradiction between the constrained spectrum resource and the growing demand for spectrum [1]. CR dynamically proposes access to the spectrum, thereby making opportunistic and intelligent use of the spectrum to both Primary User (PU) and Secondary User (SU) [2]. However, a significant security concern arises due to Cognitive Radio Networks (CRNs) openness and dynamic nature [3]-[4]. Since many research studies have been presented in literature towards spectrum sensing and accessing techniques, SU is acquisitive for spectrum holes to collaborate with other SUs to achieve their objectives. However, the previous works ignore that the SUs are vulnerable to different security threats, which can interrupt or block the information flow in CRN. The major security threat in CRN is the jamming attacks that severely degrades network performance [5]-[6]. Jammers can restrict or block the communication channels by introducing unremitting signals, thereby reducing the Signal-To-Noise Ratio (SNR), which may also degrade the throughput of the active communication flow and data transmissions [7]. Many anti-jamming solutions and schemes were introduced in the existing literature to mitigate jamming attacks. The existing anti-jamming solution based on frequency hopping offers a better approach against jamming attacks but introduces higher energy costs to the users [8[-[9]. The researchers have extensively adopted the game theory approach with Direct-Sequence Spread Spectrum (DSSS) technology to counter the impact caused by jamming attacks in CRN [10]-[11]. Although these schemes could deal with jamming effectively, these schemes require prior information about the jamming strategies and communication model, which is quite an impractical assumption considering the real scenario. With the upsurge of Artificial Intelligence (AI) technology, Machine Learning (ML) mechanisms have been extensively utilized in developing anti-jamming models. A class of ML, namely Reinforcement Learning (RL), has received widespread attention to address decision-making problems in recent years. In the context of the anti-jamming solution, RL can be efficiently utilized to explore the characteristics of jamming attacks and build an optimal policy to mitigate the jamming effect. For instance, many researchers applied the Q-learning based RL approach as an anti-jamming solution to choose an appropriate transmission power and optimal frequency hopping channel [12]-[14]. However, the Q-learning-based anti-jamming solution is prone to computational overhead due to the wide expansion in the size of the Q-table in the direction of deriving optimal policy. The researchers also applied a Deep Q-network (DQN) value-based learning mechanism [15]-[16]. DQN is a robust algorithm, and the limitation is, its slow learning rate. It is suitable only for a low-dimensional and discrete action space. Whereas, in the context of CRN anti-jamming, the action space is often both high-dimensional and continuous.

This paper proposes an efficient anti-jamming mechanism using the RL technique. Firstly, a customized environment is designed that mimics the communication scenario and sweep jammer. On the other hand, an agent modeled will operates over the same spectrum band. The training of the agent is carried out using a Deep Deterministic Policy Gradient (DDPG) algorithm. The proposed mechanism does not require any knowledge about the communication model and jamming strategy in the environment. The agent senses the spectrum and takes action in the continuous action space to distinguish whether the current carrier frequency is appropriate based on a

deterministic policy gradient. SNR and low power cost are considered as a basis for the reward for each agent's action. Table I shows some of the short forms and its abbreviations used throughout this paper. The remaining sections of this paper are planned as follows: Section II presents a review of existing works in the context of anti-jamming mechanisms in wireless networks. Section III presents the problem description based on the review analysis. Section IV presents the proposed system design and methodology adopted. Section V presents implementation strategies adopted in the proposed system of anti-jamming. Section VI discusses the outcome and performance analysis, and finally, the overall contribution of this paper is concluded in Section VII.

TABLE I. ABBREVIATION USED

CR	Cognitive Radio
PU	Primary User
SU	Secondary User
CRN	Cognitive Radio Networks
SNR	Signal-To-Noise Ratio
DSSS	Direct-Sequence Spread Spectrum
AI	Artificial Intelligence
ML	Machine Learning
RL	Reinforcement Learning
DDPG	Deep Deterministic Policy Gradient
UAV	Unmanned Aerial Vehicle
BPSK	Binary Phase Shift Keying
SSID	Service Set Identifier
DQN	Deep Q-network

II. RELATED WORK

The existing literatures has extensively studied an anti-jamming problems to enhance the communication and information flow in the severe electromagnetic spectrum of wireless networks. At present, there is less work being carried out towards a rule-based implementation strategy to thwart jamming attacks on cognitive radio networks. The preliminary discussion carried out by Ahmed and Ismail [17] has constructed a rule-based game theory to develop an anti-jamming model. The researcher has developed a Stackelberg framework that uses a rule mechanism to assign an incentive to the defender to protect the channel. Another work carried out by Ye et al. [18] has constructed a rule system using swarm intelligence that emphasizes the allocation of jammer tasks. The study mainly formulates a rule system for making a precise decision for cooperative jamming in the cognitive network.

The work carried out by Singh and Trivedi [19] has integrated game theory with decision-making theory using Markov decision process and zero-sum game. The study has used the reinforcement learning concept to formulate rules that assist in maximizing the gain of anti-jamming. The Q-learning-based concept was adopted by Liu et al. [20], where a rule-based system is constructed. The work carried out by Ibrahim et al. [21] has used game theory to resist jamming attacks in the cognitive radio network. The presented machine learning approach is constructed to develop anti-jamming features, while Markov-Game is used for modeling hamming and anti-jamming processes.

The study of Wang et al. [22] introduces an approach of the dynamic spectrum jamming mitigation scheme based on intelligent algorithms. The presented scheme is adaptive to the dynamic environment and offers effective anti-jamming capability. The work towards optimizing the different jamming parameters such as modulation mode, jamming signal power, and the duty cycle is conducted by Amuru et al. [23]. In this study, the authors have used a multi-arm gambling machine to determine optimal parametric values to guarantee an optimal jamming scheme. The work of Furqan et al. proposed an interference detection scheme based on sparse coding [24], in this study, the sparse coding of the compressed signal is considered, and the convergence mode with machine learning technique is used to distinguish spectrum holes, legitimate primary users, and jammers. Huang et al. [25] introduced channel-hopping technology robust to various jamming attacks. This proposed technology can operate without a pre-assignment role and has a limited time to rendezvous on available channels. Quan et al. [26] presented a multi-pattern frequency hopping scheme to mitigate follower and partial-band jammer. The authors have applied their frequency modes to the data channel to enhance the randomness of the transmission frequency, and the system can effectively suppress jamming. The jamming and jamming mitigation can be regarded as a game process. The rise of advancement in CRN game theory has been extensively studied to mitigate interference attacks. The authors in the study of Wu et al. [27] suggested a scheme of power distribution considering Colonel Blotto's game to resist jamming attacks. Similar works have been carried out by Jia et al. [28]-[29], where Stackelberg's game theory is employed to develop an efficient anti-jamming scheme in the wireless communication channel. The researchers in the work of Wang et al. [30] studied the selection of suitable frequency channels. This study adopts a stochastic game strategy to explore optimal data and control channels to obtain higher throughput under a high-power jamming effect. The work carried out by Hanawal et al. [31] studied joint frequency hopping and transmission adaptation rate to mitigate the reactive-sweep effect. The authors have presented a model that interacts with the user and the jammer as a zero-sum gaming approach, providing a resilient strategy against jamming. The study of Chang et al. [32] presented an anti-jamming model, which adopts a channel hopping mechanism to mitigate the interference effect caused by the jammer. Gao et al. [33] suggested a bi-matrix game model interacting among the user and the jammer. Also, the study derived optimal conditions for Nash Equilibrium under linear constraints. Though the game theory concepts have been successfully employed to model anti-jamming solutions, these methods require prior information such as jamming strategy and communication model, which is quite impractical in real-time scenarios. The adoption of RL technology in anti-jamming modeling is significant against jamming attacks. The RL agent enables the anti-jamming system to get optimal strategy through seamless interaction with the communication environment without depending on the prior information about

the jamming strategies. The work carried out by Gwon et al. [34] suggested a mechanism to cope with the jamming attacks based on the Q-learning to determine the suitable strategy for channel access. Similarly, Liu et al. [35] suggested an anti-jamming scheme oriented on the deep RL technique, enabling users to get optimal decision strategy after exploring different actions through spectrum sensing. The authors in the study of Bi et al. [36] presented a multiuser anti-jamming model by utilizing an advanced version of the Q-learning technique to attain efficiency and optimality in the network resources and communication process. In the work of Liu et al. [37], a consecutive deep RL technique is discussed to deal with dynamic jamming attacks. Table II below shows some of the strengths and limitations of the existing approaches in the above studied literatures.

TABLE II. COMPARISON OF EXISTING APPROACHES

Approach	Authors	Strength	Limitation
Game theory	[17] [21] [27][28] [29][30] [33]	Good for modeling jamming attack.	Need apriori information of the attack. Complications towards designing complex networks.
Machine learning	[20][23] [24][34] [36][37] [38][39] [40]	Higher accuracy of detection.	Training dependencies, Higher resource involved in the detection, instantaneous attack response is limited.
Swarm Intelligence	[18]	Easier modeling of attack.	Highly iterative, leading to computational complexity.
Adaptive Approach	[22]	Higher scalability.	Needs well-defined attack environment.
Channel-hopping	[25][32]	Effective for spectrum utilization, no dependency of apriori attack information.	Drains more resources from cognitive radio nodes.
Frequency hopping	[26][31]	Effectively suppress jamming.	Not benchmarked with other frequency modes.

The use of the RL technique is also found in the application of Unmanned Aerial Vehicle (UAV) systems. In the study of Gao et al. [38], the RL approach of DQN is implemented to derive optimal policy against jamming attacks. The researchers in Han et al. [39] have introduced a 2D anti-interference model where the SINR of the user's signal is improvised based on the spread spectrum and user mobility. A DQN is used to achieve an optimal strategy for the anti-jamming system. The adoption of RL is carried out by Chen et al. [40] to build an anti-jamming system in the application of wireless body area network In this work, an RL-driven power control mechanism is developed where Q-learning with transfer-learning technique is used to attain optimality in the policy and learning rate. In the work of Lu et al. [41], the anti-jamming technique is presented for UAV-based cellular networks, where a deep RL technique is implemented to determine the best relay strategy. In addition, an approach of transfer learning is used to provide additional support to the anti-jamming system to defend jammers without depending on any form of statistical knowledge about the jamming pattern and the communication model. In addition, various other studies offer a similar form of solutions [42]-[50]. These studies have mainly used game theory and another optimization-based approach. The adoption of game theory offers a good modeling aspect towards the set of actions of the cognitive radio nodes or access points. It also offers the inclusion of static and dynamic scenarios to be modeled in an anti-jamming framework. However, the approaches used are quite dependent on apriori information of the attack or based on a limited set of attack characteristics. This situation in modeling doesn't assist if one node starts exhibiting differential malicious behavior in the cognitive radio network. Hence, they potentially suffer from a lack of strategy towards prevention approach.

III. RESEARCH GAP

From the prior section, it has been seen that not many studies have used rule-based anti-jamming mechanisms. The study carried out by Ibrahim et al. [21], and other researchers have mainly used game-based logic which has its pitfalls. Unfortunately, such studies can only be modeled considering a limited set of actions of jamming, which is then subjected to the machine learning approach. The inclusion of network parameters, especially frequency and bandwidth, are less modeled in such studies, which could offer a prime indicator. Further, other models discussed in the prior section are mainly inclined towards the progressive exploration of jamming points, and hence, the delay could rise. This section discusses some significant open issues explored based on the literature discussed above.

1) No Standard technique: After reviewing different anti-jamming schemes in existing literature, it has been analyzed that there is currently no universal anti-jamming mechanism that can handle both static and dynamic jamming attacks. It has also been analyzed that designing an efficient technique for detecting and mitigating jamming attacks is quite more challenging than executing and implementing jammers in the communication channel of wireless network systems.

2) Lack of effectiveness and efficiency: The existing anti-jamming approaches are ineffective or have limited scope when the jammer covers the entire frequency spectrum. As a result, the wireless communication channel does not recover effectively to provide its services. Another significant issue is how to achieve efficiency in the anti-jamming mechanisms. The frequency hopping-oriented anti-jamming technique can withstand narrowband jamming attacks but at the cost of compromised and degraded spectral efficiency. Similarly, the jamming mitigation based on retransmission protocol can recover communication channels, but it also degrades efficiency and affects the overall communication performances. Most of the schemes in the literature do not consider a trade-off between communication efficiency and the effectiveness of jamming mitigation techniques.

3) Impractical assumption: The existing solutions against jamming attacks based on the model-based analysis are impractical in the real-time implementation scenario. The model-based analysis like game theory and cross-layer

optimization often adopts prior knowledge regarding global channel information and jamming strategy, which is an unrealistic assumption. Also, these approaches are subjected to high computational complexity in the modeling. The jamming mitigation solutions in the existing literature also lack novelty in the design and modeling. It has been found that a similar kind of design consideration and modeling strategy is adopted in most of the studies, which needs optimization in their solution design to meet the requirement of the real-time networking scenario.

4) Curse of high-dimensionality: Recently, reinforcement learning has been extensively studied to design adaptive anti-jamming schemes. However, existing schemes based on Q-learning suffer from the huge dimensionality problem in a complex environment because to achieve optimal policy, the amplitude of actions needs to increase. As a result, the size of the Q-table of Q-learning also increases, which causes slow learning and restricts application scenarios.

5) Lack of suitable environment: Most of the RL-based solutions in the existing studies lacks modeling of a suitable networking environment to assess the RL agent. To validate the scope of the RL agent-based solution, the researchers must design and implement a suitable environment.

Therefore, the prominent research gap identified in existing studies are that, there are few available standard methodologies to deal with jamming issues in complex network system. Further, the formulation of the presented solution is carried out using impractical assumptions that doesn't work on ground reality. Hence, there is a big research gap between the demands to be met in resisting the jamming due to the ineffectiveness of existing solutions. Therefore, the problem statement is stated as "it is challenging to develop a lightweight, robust, and cost-effective computational solution towards mitigating jamming attack in communication environment". Hence, the motivating factor towards adopting the proposed research scheme is due to the rise of security-critical applications for securing the real-world's wireless communication technologies such as Bluetooth, Wi-Fi, and cellular technologies like 3G, 4G, and 5G that demands an intelligent design in the anti-jamming techniques by considering the constraints associated with the network. In this regard, the above factor motivates for an efficient and intelligent jamming scheme that can effectively balance communication efficiency and anti-jamming capabilities. The next section discusses the proposed system to address above discussed research problem.

IV. Proposed System

This section will discuss the proposed system of intelligent anti-jamming mechanisms based on the RL technique. The prime aim of the proposed study is to present an effective and efficient strategy to counter the jammer in the cognitive wireless networking system. Therefore, a model-free and off-policy scheme is employed to design the anti-jamming mechanism. The proposed mechanism does not require or depend on the prior information of jammer signature and channel models. The off-policy algorithm in RL agent design offers a better scope of accurate prediction of the jamming frequency and evades it in advance.

Another significant contribution of the current research study is modeling a customized environment synchronized with Open AI-Gym functions. This is the novel contribution of the current study, where the proposed customized environment imitates the scenario of the communication channel and jammer under consideration. The proposed anti-jamming mechanism operates in this environment over the same set of channels and uses its ability to learn sub-band selection strategies to avoid jamming signals. The objective of the proposed work is to offer a better solution that can meet the requirement of a real-time scenario. The schematic diagram and workflow of the proposed model are shown in "Fig. 1".

A. Scope of the Study

The transmission method considered in the present study is the Binary Phase Shift Keying (BPSK) modulation. The BPSK is popularly used in many modern communication systems like Wi-Fi, Bluetooth, and even car locking systems. Jamming of these systems can have a devastating effect on the company's overall image, which is offering the service. Such jamming might also result in the theft of valuable physical or digital assets. Since the communication protocols like Wi-Fi have a band defined for itself and agreed between the gateway and the end node, the system will always know when a particular channel is busy.

The system which is being proposed will help the communication system to shift the frequencies dynamically between the available channels and the band. It is assumed here that the band which can be observed in the simulation is always free since there is always an understanding between the various transceivers present in the system, and every other transceiver knows the channel and the band being used by another transceiver. Hence, when a foreign attacker element like jammer is introduced, a particular frequency that is known to be free by all the transceivers will get blocked.

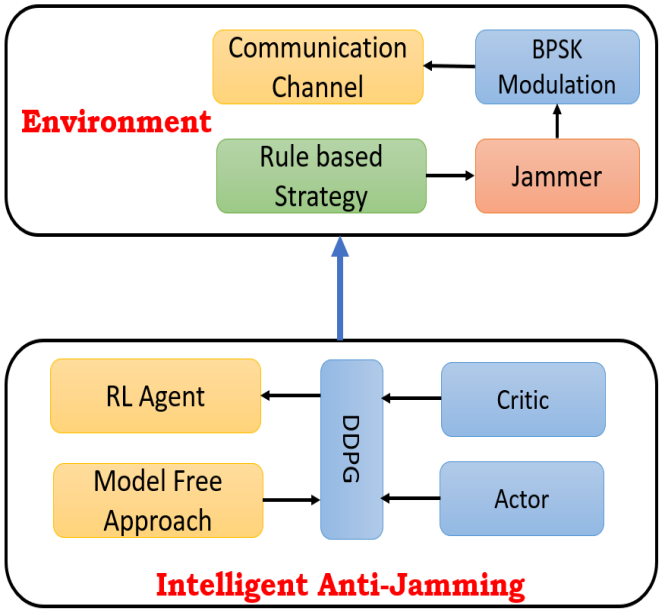

Fig. 1. The Schematic Architecture of the Proposed Anti-Jamming System.

B. System Design

The development of the proposed system is carried out in two implementation phases 1) Environment modeling and 2) agent-based intelligent anti-jamming. In the first phase of implementation, the study presents a design of environment considering a single communication channel with a transmitter, receiver, and rule-based jammer. In particular, the study considers the case scenario of a sweep jamming attack that uses a digital modulation scheme, viz. BPSK, in which the carrier frequency varies according to the modulating digital signal. The phase transition of the BPSK digital modulation is characterized by the π radian, and it has the lowest spectral efficiency at least 1 bit/s/Hz. The signal $s_k(t)$ can be expressed as follows:

$$s_k(t) = f(t) \quad (1)$$

The expression (1) represents a signal concerning the function of time i.e., *f*(t), which is meant for constructing the signal in the simulation model. The variable *t* denotes time in the range [0, ∞] and $s_k(t)$ in the range [0, 1]. A signal s(t) modulated using BPSK during k time interval can be represented as follows:

$$s_k(t) = \sqrt{2R}\cos\left(2\pi f_0 t + d_k\frac{\pi}{2}\right), (k-1)T \le t < kT \quad (2)$$

where R refers to the mean signaling power, the term $d_k \in \{+1, -1\}$ regulates the data bit, f_0 denotes base frequency, and T denotes symbol period. Furthermore, the communication frequency is the unit of frequency between the sender (transmitter module) and the recipient (receiver module). The unit of the frequency used to create interference is referred to as the jamming frequency. The proposed study considers single-channel continuous frequency band and jamming frequencies in the simulation environment. RL techniques involve environment and agent mechanisms that require feedback (reward/penalty) from the environment. In the present work, the SNR obtained at the recipient is considered as the basis of the award. Later sections discuss the jamming model and the relationship between the environment and the RL agent.

C. Sweep Jamming

The jammer forwards a forged signal with a power (P_J) on the target frequency channel to intentionally intrude the information flow or the active communication process. In some cases, the jammer also introduces jamming by reducing the SNR of the data signal received by the recipient with less interference power. The current research work considers the Sweep jamming scenario in the communication channel.

A sweep jammer sequentially blocks N_J adjacent channels with power P_J / N_J from N communication channels in each time slot. In this jamming attack scenario, a narrowband frequency of the jammer's power is recurrently swept over a comparatively wideband frequency with a sweeping rate such that there is enough time to complete its jamming function at any given frequency. Before the jammer terminates, it comes back to that frequency again. The communication channel (C) selected by the jammer (J) at instance (k) can be represented as $C_j^k \in \{1,2,3,\cdots N\}$. The study considers a simple scenario, where the set of jamming actions (J) at a different position such that $C^K \rightarrow [C_j^K]_{1\le j\le J}$. It is also believed that the jammer can jam all communication channels if the transmitter or receiver resides within the proximity of the jammer.

D. Reinforcement Learning

RL is an explicit type of ML technique that is concerned with the function of the agent and the environment. The agent is a programmed AI function that interacts with the unknown environment, learns the behavior, and decides to solve the task. The agent operates in the environment and gets rewards for each operation. The agent's goal is to maximize the expected cumulative reward by selecting the best action for its current situation. “Fig. 2”, depicts a typical relationship between the RL agent and the environment.

The environment refers to the program or task that the agent needs to perform. It takes the current state and operation of the agent as input and returns the Reward/Punishment and the next state as output. The agent is the programmed entity that refers to the learner and decision-maker that executes an action in the environment. The agent gets observation and reward and sends an action to the environment. An action can be described as a set of possible moves that the agent can perform in the environment to solve a given problem. Observation is the state of the current situation that is returned by the environment. A reward/punishment is the feedback return given to the agent for each specific move or action. An agent is designed based on the policy to decide the agent's next move according to the current state of observation.

The RL entails a programmed decision-maker algorithm as an agent learning the behavior through repeated trial and error interaction with the task scenario, i.e., environment. Moreover, the success and failure rate of the agent is decided based on its reward, which indicates how better the action was. Generally, RL is implemented using episodes, a set of time steps during which the agent learns the optimal strategy towards deciding actions to achieve the cumulative reward. During these time steps, the state in the environment may vary, which may also impact the actions learned by the agent. However, the agent attempts to maximize its reward by choosing an optimal action according to the current state of observation and, at the same time, automatically updating the corresponding strategy.

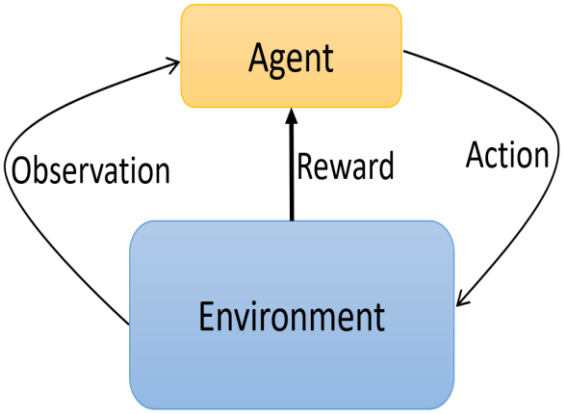

Fig. 2. Relationship between Environment and RL Agent.

V. Methodology and Algorithms

This section discusses the implementation strategies adopted in the environment modeling and development of the anti-jamming system using the RL mechanism.

A. Environment Modeling

The environment consists of the communication model and the jammer. The communication model is designed by considering Wi-Fi communication channels. However, with changing the base frequency and band frequency, any of the following technologies can be considered 5G, 3G, 4G, and Bluetooth. As mentioned, the jammer used here is a sweep jammer that will keep increasing and decreasing the frequency of jamming over time, which means that jamming frequency keeps sweeping the communication spectrum. Fig. 3 shows the modeling of the environment, which mimics the communication scenario. As shown in "Fig. 3", the current research work considers a scenario of CRN communication as an environment that consists of two users (sender and recipient), including one transmitter (T_x), and receiver (R_x), and a jammer (J). The transmitter is accountable for the signal modulation, and the receiver is accountable for demodulating and receiving the signal from the transmitter.

The signal (S) can be jammed by producing noise or interference in the same frequency band. The study considers the base signal using BPSK modulation in the simulation process, as shown in "Fig. 4". The block of a random integer is meant to model the wireless channel's noise before subjecting it to the BPSK block.

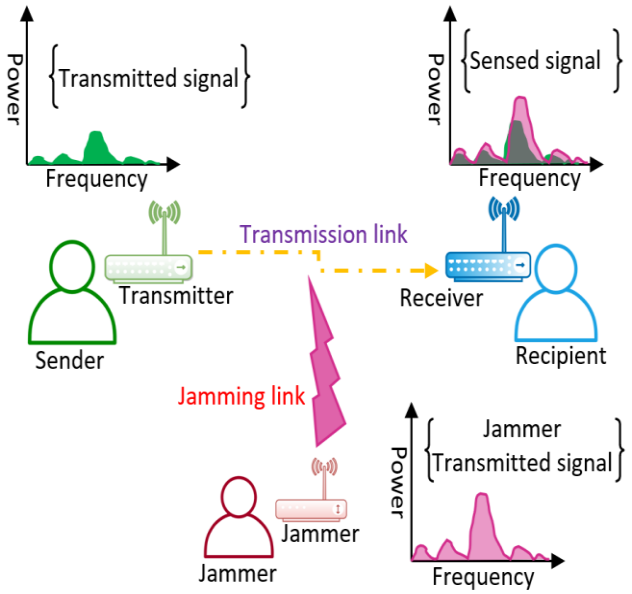

Fig. 3. Environment Model.

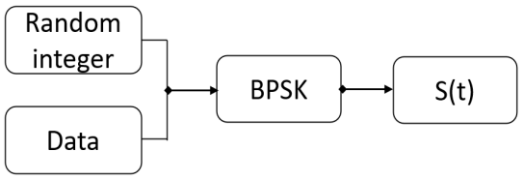

Fig. 4. Signal Construction Process.

B. Design Consideration

The proposed study in the environment modeling considers the processing of signals in continuous time instead of the processing of signals in discrete time-slots. This is because the current work focuses on overcoming jamming at the physical layer. The jammer works continuously, and the proposed anti-jamming overcomes the interference effect caused by jammer (J) on the entire signal of interest instead of the data. Under this consideration, both J and sender (T_x) share the same continuous-time signal. The T_x opts transmission channel frequency (f_s) from the pre-determined set of frequencies such that $f = \{f_1, f_2, f_3 \cdots f_N\}$ of the communication (WiFi) band to forward or carry out the transmission of the data packets to the recipient (R_x) with power P_s. In the case of jamming, the jammer chooses the random transmission channel frequency (f_J) of the same communication band, attempting to impede transmission between T_x and $R_x$with power P_J.

The proposed system constructs a condition for a jamming attack towards the ongoing communication. According to this condition, the power variable P_j associated with the channel frequency of jammer f_j is required to be potentially higher than the power variable P_s associated with transmission channel frequency f_s that is received at the receiver end. Moreover, the study in the phase of environment modeling also considers the bandwidth (b) factor, which is equal for both f_s and f_J such that:

$$b_s = b_j \tag{3}$$

$$b \in f_s \rightarrow b_s \tag{4}$$

$$b \in f_j \rightarrow b_j \tag{5}$$

According to the above expressions (3)-(5), it states that bandwidth b_s should be similar to the bandwidth required for jamming b_j as per expression (3) which bears a predefined set of frequency f_s as per expression (4). At the same time, bandwidth b also carries jamming frequency b_j as per expression (5). Hence, expression (3)-(5) will mean that bandwidth consideration in the proposed model bears characteristic of both a predefined set of frequency as well as jamming frequency. This consideration makes the system model difficult to assess the network parameters. This problem becomes more difficult in cognitive radio networks as several secondary users are compared to primary uses over a wide spectrum band. Apart from this, it will also become a challenging aspect to distinguish between the busy channel and jammed channel. This research challenge is further solved as follows: The proposed study considers that if the T_x forwards data in a transmission medium which is also designated by the J, then the SNR of the signal at R_x is corrupted severely. Considering above discussed scenario and notions, the SNR at R_x can be numerically expressed as follows:

$$\text{SNR} = \frac{P_s h_s}{P_J h_J \mathbf{F}(f_s = f_j)} \tag{6}$$

In the above equation (6), the term h_s refers to channel gain from T_x to R_x and the term h_J indicates channel gain from J and R_x. Here, F(x) refers to characteristics function that defines the

probability distribution of the received signal based on the condition numerically expressed as follows:

$$\begin{cases} 1 \text{ if } f_J = f_j : \text{True} \\ 0 \text{ Otherwise} \end{cases} \quad (7)$$

The above expression (7) represents that if input argument (x) of this function is true, the value of this function is equal to 1; otherwise, 0. The study also considers demodulation cutoff value, where data transmission fails when the SNR is lower than this cutoff value. Since at the beginning, the agent (A_g) may not be able to recognize the channel or transmission medium chosen by the J. In this regard, the A_g has to sense the frequencies of channel f_s continually and store the sensing results in the form of a vector. In the proposed modeling, this vector is considered as environment state (S_t) for the corresponding action (A_t) carried out by the A_g (an anti-jamming mechanism). The A_t carried out by A_g is decided based on observed S_t of environment and A_g gets immediate feedback as a reward R_w which is characterized according to the selection of channel and channel switching power cost. In the proposed study, the output of A_g is an un-jamming of transmission medium or communication channel based on its A_t towards channel selection strategies. Therefore, the A_g tries to attain successful communication with a minimal channel switching power cost. Considering this factor R_w for A_g towards its corresponding A_t can be numerically expressed as shown in equation 8:

$$R_w = R_{SNR}(A_t) - c(A_t) \quad (8)$$

In the above equation 8, the first term $R_{SNR}(A_t)$ refers to R_w achieved A_g $\forall$ successful transmission. The success of transmission is considered when the SNR of the signal at R_x surpasses demodulation cutoff value, and in this case, the R_w will be equal to 1; otherwise, it would mean that the transmission has failed, and then the value of R_w will be -1, as expressed in equation 9 as follows:

$$A_g \leftarrow R_w = \begin{cases} 1 \text{ if } SNR_{R_x} \geq SNR_{cutoff} \\ -1 \text{ Otherwise} \end{cases} \quad (9)$$

In equation 9, the term SNR_{R_x} refers to the SNR of the signal at R_x (received signal) considered as a basis for computingR_w. The proposed study considers a control channel that conveys signals in transient to R_x is secure, which the jammer cannot influence. The second term c(At) in the above equation 9 refers to the communication cost factor associated with A_t regarding the channel switching. The T_x and R_x in environment performs transmission of information on the fixed channel with stable power. But during the channel switching process the A_g tries to succeed by taking into account the low power cost. Otherwise, it will be penalized. Hence, the proposed study considers the c(At) term in the computation of R_w for each A_t.

C. Algorithmic Principle

The prime challenge is creating a difference between busy and jammed channels using an intelligent anti-jamming mechanism. The complete idea is implemented on the cognitive radio node and not on the access point. The prime justification behind this is that cognitive radio nodes create a bridge of communication between the base station and access point; therefore, they are more prone to jamming attacks. If the access point experiences a jamming attack, it can easily be rectified by switching over to the next access point, but this is not the case with cognitive radio nodes. If the jamming attack intrudes on cognitive radio nodes, then it will directly affect the primary users. The modeling of the environment contains three significant functions such that i) __init__ function, ii) reset function, and iii) step function.

1) init__ function: This function creates and initializes the environment, as demonstrated in Fig. 5. The environment is built to scan the available channels over the given Service Set Identifier (SSID) tries to block it in this function. Since the proposed study considered a WiFi communication channel, the terms SSID used here are just for recognizing which signal is being blocked. The SSID here is not a real one, but it is simulated. Also, the jamming frequency is set to the first available channel.

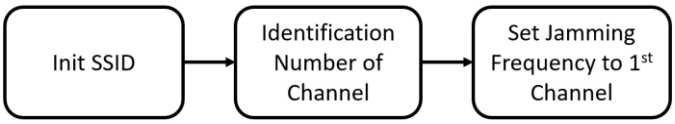

Fig. 5. Process of Environment Initialization.

2) Reset function: The reset function sets the environment back to its original setting. The reset function sets the jamming frequency back to the first available frequency in the channel, the reward is reset to zero, the signal is turned on, and communication channels are observed. The process of the resetting environment is shown in "Fig. 6".

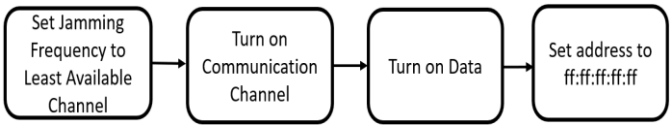

Fig. 6. Process of Environment Setting.

3) Step function: This is the main function of the environment. In this function, the packet of data is tried to be transferred. If the jamming frequency is equal to the base frequency, then the data either gets corrupted or does not pass. Based on this, the environment either gives a positive or negative reward. The process flow of the step function is shown in "Fig. 7".

- Agent (A_g): The agent itself is an anti-jamming system that deals with the decision-making process. Initially, the agent interacts with the environment, senses the frequency spectrum, and instructs the transmitter to select a jamming-free communication channel.
- State (S_t): State refers to the perceived observation of the environment. Here, the state is the agent's frequency spectrum explored and sensed.

- Action (A_t): action refers to the strategic decision taken by the Agent. In the proposed study, the action changes communication band frequency and channel selection.
- Reward (R_w): Reward is the agent's feedback for the action taken for the observation in the environment. The term Reward is the higher SNR and low communication cost in the proposed system.

The proposed intelligent anti-jamming system is based on the function of the RL agent and its interaction with the environment. The environment responds to the agent via SNR so that the agent can inevitably distinguish whether the currently used carrier frequency is appropriate. Concurrently, the output strategy/policy can be further optimized, thereby circumventing surplus and redundant losses from frequency hopping. In the current work, the modeling of the agent is carried out based on the DDPG RL technique, which offers better interaction with the environment and automatically constructs a policy regarding observation and action state. However, the DDPG method involves two functions parameterized with a neural network, namely the Critic and the actor, to derive the best strategy or optimal policy. The actor's job is to specify a decision strategy towards deciding the optimal action for each state of observation. The Critic is responsible for estimating the value functions characterizing how well the actor takes action. The critic function generates the Q-value, representing the expected cumulative long-term reward that the agent receives from the environment for its action towards the current state of observation.

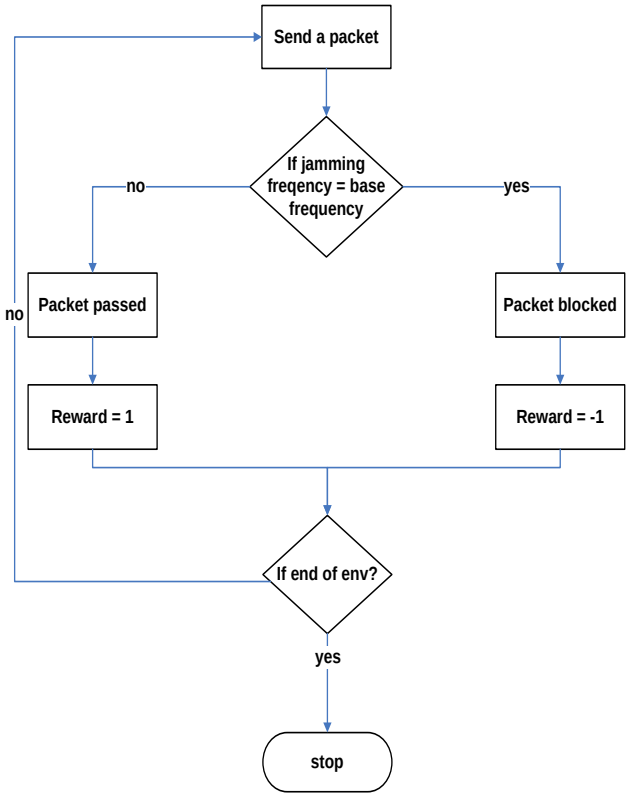

Fig. 7. Process of a Step Function.

The core principle of DDPG is to combine and get the benefits of both deterministic policy gradient mechanism and Q-learning function. The key aspect of the deterministic policy gradient mechanism is that it can handle continuous actions space while minimizing learning time. The mathematical representation of the DDPG is as shown in "Fig. 8".

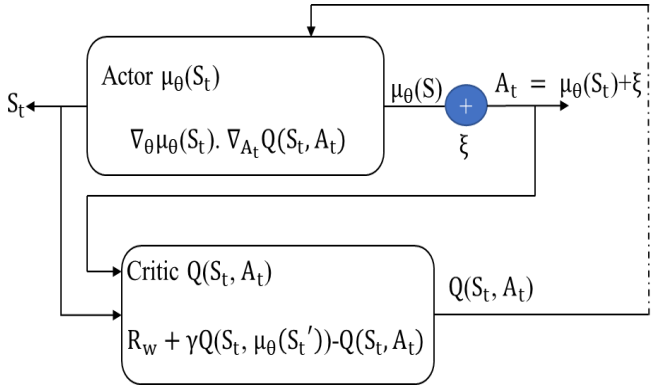

Fig. 8. Schematic Representation of DDPG.

Above Fig. 8 shows the building block of the DDPG with actor $\mu_\theta(S_t)$ and critic $Q(S_t, A_t)$ representation. However, to estimate the optimal policy and Q-value function, a DDPG algorithm maintains two networks i) current network and ii) target network. The current network consists of two function approximators such as current actor $\mu_\theta(S_t)$ and Critic $Q_\theta(S_t, A_t)$. The target network consists of two similar function approximators such as target actor $\mu_\theta'(S_t)$ and target critic $Q_\theta'(S_t, A_t)$, both the current network and target network have similar configuration, function, and parameterization. The target network is the copy of the main network (current) that helps to maintain stability in the training process of the main network.

1. Algorithmic Principle for Agent Modelling

Init Param
Param = [E, BS, BL, ξ]
Agent = f (Current, Target, Param)

The modeling of agent considers three major procedures such as i) building actor-network, ii) building critic network and iii) initialization of target actor-critic network. In the agent modeling, the algorithm also considers parameters (Param) such as ξ random process for initial state observation, batch size (BS) buffer length (BL), and E episode.

Procedure-1: Building Current Network

Input layer (S_t)
denselayer1 (size =100, connect=inputlayer, activation=relu)
denselayer2 (size =100, connect= denselayer1, activation=relu,)
denselayer3 (size =100, connect= denselayer2, activation=relu)
Actor = f_1(actor_Network, S_t)

The above-discussed procedure is mentioned for actor configuration with input layers and multiple dense layers connected in a feed-forward manner. A $f_1(x)$ refers to the deterministic policy actor function used to construct an actor model with input arguments of actor_network, which includes all configured layers, and observation state S_t from the environment and returns the corresponding action that is expected to maximize the long-term cumulative reward.

Procedure-2: Building Critic_Network

Config → StatePath
inputlayer (S_t, 'state')
denselayer (size = 100, connect = input, activation=relu)
Config → ActionPath
inputlayer (A_t, 'action')
denselayer (size = 100, connect = input, activation=relu)
Config → State_Action
Netlayers (dim= (1,2))
Add activation_function: relu
denselayer (size = 100, connect = state_action)
Add activation_function: relu
Critic $= f_2$*(StatePath, ActionPath, State_Action path)*

The above-discussed procedure is mentioned for Critic configuration with state path, action path, and state action path. The Critic takes S_t and A_t as inputs and returns the Q-value representing the expected long-term reward. The $f_2(x)$ refers to the value function used to construct Critic network, which maps input value consisting of S_t and A_t to an expected and cumulative output value as reward based on the Q-value

Procedure-3: Building Target network

Actor_target $= f_1$*(Actor,* S_t*)*
Critic_target $= f_2$ (Critic, StatePath, ActionPath, State_Action path)

To achieve stability in the training phase of the current network, the algorithm initializes the target network with the same configuration as the current network. Therefore, the proposed RL agent adopts a four-function estimator (i.e., two from the current network and two from the target network) to perform the anti-jamming operation by sensing the spectrum without relying on any kind of statistical information about the jamming pattern and communication mode. The proposed anti-jamming mechanism has the advantage that it is suitable for continuous action space and overcomes the limitation associated with Q-learning and model-based jamming mitigation approaches. Thus, it can meet the requirement of a real-time deployment scenario. The algorithm for anti-jamming based on the proposed off policy and model-free Agent is discussed as follows:

2. Algorithmic Principle for Anti-jamming

Input: $E_P, A_t, T, \gamma, \mu_\theta, \mu_\theta', Q_\theta, Q_\theta', B_S, A_g$
Output: Unjammed signal
Start
1. Init random θ for the current network: [μ, Q]
2. Init random θ' for target network: [μ',Q']
3. For episode $= 1: E_P$ do
4. Init ξ for A_t exploration
5. Sense frequency spectrum
6. get initial state S_t
7. For i $= 1: T$ do
8. select action $A_t = \mu(S)$
9. select channel $f_s \in C_i$ and (P_s) according to current policy
10. Execute: $\mathbf{A_t} \leftarrow$ initiate communication
11. Sense new spectrum $\mathbf{S_{t+1}}$
12. Compute $\mathbf{R_W}$
13. Using equation 8
14. store experience: $[S_t, A_t, R_w, S_{t+1}]$ into experience pool ψ
15. Check $|\psi| \geq \mathbf{B_S}$ do
16. random selection of B_S from ψ
17. For each experience in B_S
18. Compute target actor and critic value
19. Using equation 14
20. Update current Critic by minimizing the loss
Using equation 15
21. Update the current actor using policy gradient
Using equation 16
22. Update target actor and target critic
Using equation 17 & 18
End

The above-mentioned algorithmic steps describe the working principle of the agent to perform un-jamming based on the interaction with the environment where it senses the spectrum and takes the action by exploring and learning the channel switching pattern. Concurrently, the agents get rewarded and observe the next state. The proposed algorithm takes a set of inputs such as number of episodes (E), number of iterations (T), Set of Action (A_t), discount factor (γ), current-actor (μ_θ), target-actor (μ_θ'), current Critic (Q_θ), target critic (Q_θ') and batch size (B_S). In the beginning steps, the algorithm randomly initializes the weighting parameters (θ and θ') for both current and target network, respectively (Step:1-2). In the next step, the algorithm initializes ξ a random process for A_t exploration based on the initial S_t perceived by sensing the spectrum frequency (Step:3-6). In these steps, the A_g interacts with the environment (the communication scenario and jammer). In the communication scenario, the study considers channel link such that $C_i \in \{1,2,3,\cdots N\}$. The A_g senses the C_i and for each C, the proposed algorithm considers value 1 if C is in a good state, otherwise, 0. Therefore, the A_g gets feedback (R_w) $+1$ if a C with a good state is selected, otherwise R_w will be -1. When A_g interacts with the environment, it gets the initial S_t such that $S_t = \{1,2,3,\cdots N\}$, and performs current A_t. Here, N denotes the number of channels (C). According to policy (i.e., current A_t) the channel $f_s \in C_i$ with transmitter power P_s is selected to launch the communication (Step:8-10). The proposed algorithm also considers that the A_g keeps sensing the spectrum and gets set of states such that $S_t \in M$, where M is the n-dimension vector the holds most recent observations S_t space. At every instance t, the recent state S_t will be added to the M, and the previous S_t (S_{t-M}) will be ignored (Step:11). The recent S_t at next instance, $t+1$ can be represented as follows:

$$S_{t+1} \leftarrow \{S_t, S_{t-1}, S_{t-2}, \cdots, S_{t-(N-1)}\} \quad (10)$$

In the next step of the algorithm, a set of $A_t \in K$, where K is the n-dimension vector space that holds distinct $A_t$sets such that $K \in \{1,2,3,\cdots N\}$, where each genuine A_t in the K denotes the channel index. Hence, in this regard, when an A_t is selected, the A_g will guide the transmitter to access the corresponding C and subsequently A_g gets the R_w which exposes the state-selected C (Step 12). It is to be noted that the A_g is allowed to access a single channel to sense the spectrum and learn the channel selection strategy in each iteration. According to numerical expression 8, attaining higher R_w meant getting a higher possibility of successful transmission. Therefore, in this regard, the core objective of the A_g to get an optimal policy or strategy π, that maps the S_t to the K, towards maximizing the long-term cumulative R_w for the decisions choosing anti-jamming action towards accessing un-jammed channel such that:

$$R_{w_K} = \sum_{i=0}^{\infty} \gamma^i R_{w+i+1} \quad (11)$$

The agent mechanism meets this objective by computing the ideal action-value such that:

$$Q^*(S_t, A_t) = \max_{\pi} \chi\left[R_{w_K} \middle| S_t = M, A_t = K, \pi\right] \quad (12)$$

From the above numerical expression, the objective of the maximizing cumulative R_w is achieved to increase the χ (expectation) on all S_t accessible by the policy of the R_w achieved after eachA_t, where $Q^*(S_t, A_t)$ refers to the optimal action-value function and π refers to policy rule (i.e., policy mapping sequences to A_t) . Therefore, the optimal policy for a precise A_t in any accessible S_t can be achieved by solving the following numerical equation as follows:

$$\pi^* = \arg\left(\underset{\pi}{\text{Max}}\, Q^*(S_t, A_t)\right) \quad (13)$$

Where $\pi*$ refers to the optimal decision policy regarding precise action in any given state of the environment. Therefore, A_t with maximum $Q^*(S_t, A_t)$ have maximum cumulative R_w and a higher probability of being selected. Thus, the agentA_g can choose the action A_t with maximum action-value to effectively overcome the impact of jamming attacks. Since the action space is continuous, computing max in the action-value function $\underset{\pi}{\text{Max}}\, Q^*(S_t, A_t)$ value is extremely a difficult task because the algorithm is required to be executed every time the A_g interacts environment to perform an A_t which is unacceptable in the real-time scenario. But adopting the DDPG algorithm in the proposed anti-jamming modeling, the agent perceives a value function (by Critic network) and action strategy (by actor-network) synchronously. The Critic uses off-policy and the Bellman equation to learn the value functions ($Q^*(S_t, A_t)$ and exploiting this information, the actor learns the action strategy (π^*) in such a way that it works particularly well for continuous action spaces in the environment. In this process, it is considered that derivative of function ($Q^*(S_t, A_t)$ exists at all points in its domain concerning the A_t argument. This enables to derive an optimal gradient-based learning protocol the policy $\mu(S_t)$, i.e., a function of the actor towards deciding optimal action. In this regard, estimation of the action value can be done by approximating $Q(S_t, A_t, \theta) \approx Q^*(S_t, A_t)$ instead of executing complex and iterative optimization procedures every time to compute $\underset{\pi}{\text{Max}}\, Q^*(S_t, A_t)$. But to achieve stability in the agent learning process, the algorithm maintains two networks, i.e., the current actor-critic network and target actor-critic network. The current network performs a selection of action and evaluation of action-value, and the target network is used to maintain stability in the training process of the current network. The training of the current network requires experience pool ψ to approximate$Q^*(S_t, A_t)$. The experience pool ψ is a set of transitions obtained via interaction between the current actor-network and the environment which is stored in the experience pool $\psi \epsilon \{S_t, A_t, R_w, S_{t+1}\}$ (Step:14). In this next step, the algorithm performs a sampling process by taking a random B_S (minibatches) of transitions from the ψ to update the parameters of current Critic and current actor networks (Step:15-16). The process of sampling B_S splits the transition data into small batches to compute error and update network parameters. To stabilize the performance of the agent mechanism, the algorithm then computes the target actor value and target Critic value for the next S_t. Q-value (Step:18) numerically obtained through Bellman equation as follows:

$$y_k = R_w + \gamma Q_\theta{}'\left((S_t{}', \mu_\theta{}'(S_t{}')\right) \quad (14)$$

In the next step of the algorithm (Step:20), the training of the current Critic network is accomplished by updating its parameter based on the loss function numerically expressed as follows:

$$\mathcal{L}(\theta) = \frac{1}{B_S} \sum_k (y_k - Q_\theta(S_t, A_t))^2 \quad (15)$$

In the above equation (15), MSE (mean-squared-error) is used as a loss function that computes the sum of squared distances between the updated target Critic (action-value) and current Critic (action-value). Further, in the next step of the algorithm (Step:21), an iterative optimization algorithm (gradient descent) is used to update and rejuvenate the weights θ of the current actor towards determining appropriate A_t decision strategy that maximizes the R_w numerically expressed as follows:

$$\nabla_\theta J(\theta) = \frac{1}{B_S} \sum_k \nabla_\theta \mu_\theta(S_k) \times \nabla_{A_t} Q_\theta(S_t, A_t)|_{A_t = \mu_\theta(S_k)} \quad (16)$$

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_S of transitions. Finally, the target actor and target Critic network is updated to provide better stability in the learning process of the current actor and Critic network numerically expressed in equation 17 and 18 (Step:22):

$$\mu_\theta{}' \leftarrow \emptyset\theta + (1 - \emptyset)\mu_\theta{}' \quad (17)$$

$$Q_\theta{}' \leftarrow \emptyset\theta + (1 - \emptyset)Q_\theta{}' \quad (18)$$

Where $\emptyset$ refers to the hyperparameter 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. “Fig. 9”, illustrates the architecture of the proposed anti-jamming mechanism based on the above-discussed algorithm.

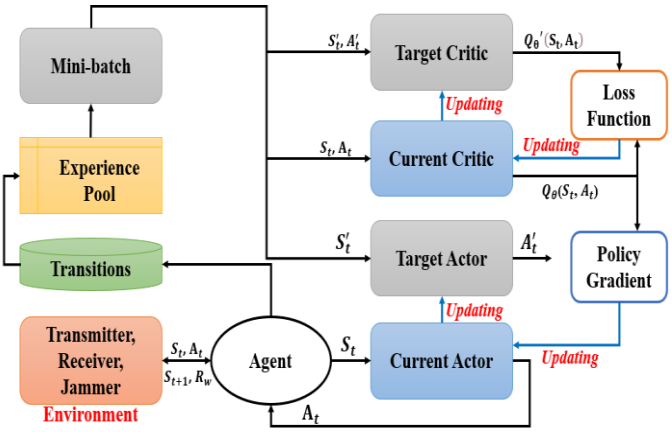

Fig. 9. Illustration of the Proposed Anti-Jamming Mechanism.

VI. Implementation and Result

This section presents the outcome achieved regarding the power cost and SNR. Also, the effectiveness and scope of the proposed RL-driven anti-jamming system are justified based on the comparative analysis with two other RL techniques, viz. i) Q-learning and ii) DQN. The Q-learning and DQN both have a higher frequency of implementation in the existing literature for solving the jamming problem. Therefore, both techniques are considered suitable for comparison with the proposed method.

The design and development of the proposed anti-jamming solution are carried out using Python with an Anaconda development environment configured on Windows 64-bit Intel i7 CPU, 16GB ram, and NVIDIA GTX graphic card. The parameters considered in the simulation process of the proposed system are highlighted in Table III.

The simulation setting considers a pair of transmitter-receiver and jammer operating in a frequency band of a WiFi communication channel, i.e., 2.4GHz. The number of communication channels considered in the proposed case is 11. The bandwidth of the jammer and transmitter is considered equal to 20MHz. The jamming power is considered equal to 30dB. The transmitter signal power is considered equal to 25-45 dB, and the demodulation cutoff value is considered set equal to 10 dB. The carrier or base signal frequency is considered equal to 5 GHz, and the data rate is equal to 2 Mbps. The digital modulation technique adopted is BPSK. The cost associated with the channel switching factor is considered equal to 0.2, the discount factor is considered in the range of [0,1]. Basically, in the proposed simulation setting, it is considered equal to 0.96, and the value of the minibatch size is considered equal to 32.

The performance of the proposed system is analyzed for power factor and SNR. Also, the scope and effectiveness of the proposed system is justified based on the comparative analysis, where the proposed system is compared with two similar existing reinforcement learning techniques, i.e., Q-learning and DQN. "Fig.10", shows the time-frequency analysis of the jammer and transmitter, where the light blue block represents the channel selected by the jammer, and light orange blocks represent the data transmission scenario. The red block represents the jammed signal. In this case, if the user sends data packets through the channel selected by the jammer, the SNR of the signal at the receiving end will degrade, resulting in transmission failure. The problem faced by the agent is how to determine the optimal strategy in an unknown environment.

Since the agent does not know the channel selected by the jammer in the unknown environment, it must sense the communication spectrum and select the channel according to its previous interaction with the environment to guide the transmitter to select an undisturbed channel for transmission without affecting the system utility. Therefore, adoption of the above-mentioned environment assists to understand more about the signal system with respect to the transmission with less iterative way and more progressive way towards mitigating jamming issue in communication. Hence, jamming problem is controlled to a large extent when deployed over the proposed test environment to see a visible beneficial outcome as discussed in the consecutive part of this section. The idea is to offer better balance between computation and communication.

TABLE III. Simulation Parameters

Parameters	Value
WiFi Frequency band	2.4 GHz
Number of communication channels	11
Jamming Model	Sweep Jammer
Jamming power	30 dBm
Transmitter signal power	25-45 dBm
The bandwidth of the Transmitter signal	20MHz
The bandwidth of the Jamming signal	20MHz
Demodulation cut-off	10 dB
Data rate	2 Mbps
Digital modulation technique	BPSK
Channel switching Cost	0.2
Discount factor	0.96
Minibatch size	32

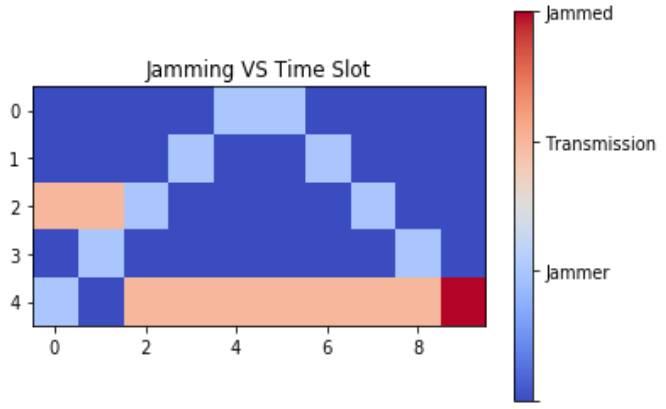

Fig. 10. Illustration of Jammer Process and Transmitter Process.

From the Section II of this paper, it can be seen that there was previous technique to solve the problem, however, DQN and Q-Learning is the extensively used methodologies that offers better guidelines towards proposed problem mitigation. Therefore, the outcome of the proposed technique is compared with DQN and Q-Learning which represents the frequently adopted solution.

"Fig. 11" shows the performance analysis of the proposed system regarding the power cost factor. The graph trend exhibits that the proposed system performs much better than the other two RL techniques regarding reduction in the fluctuation of the power consumption during data transmission. The performance of the Q-learning technique is worse than DQN based anti-jamming process. This is because state space is so huge that it suffers in finding an optimal policy. The Q-learning depends on the Q-table, which is discrete, and also it uses a simple Bellman algorithm to perform the frequency hopping, which directly depends on Q-function. The agent algorithm tries to choose the frequency which gives the highest reward. In the case of Q-learning, the agent algorithm uses a Q-table to decide the optimal reward; hence the frequency needs to keep changing. Therefore, more frequency hopping leads to more power consumption. The DQN exhibited little fluctuation in the power consumption and has achieved closer performance to the proposed system. The DQN is more efficient than Q-learning, tries to optimize action-value for the present scenario. The Q-learning is discrete and has a limited number of rows and columns in Q-Table, whereas in DQN, one neural network decides the best action and state pair for the given scenario. The proposed system uses the DDPG algorithm suitable to continuous action space and has less randomness than Q-learning and DQN. The proposed system converges to less fluctuation and achieves lower power consumption.

In "Fig. 12", it can be analyzed that both proposed and DQN exhibits similar performance regarding SNR. However, in the case of Q-learning, the graph trend exhibits lower SNR compared to the others. The SNR refers to the overall number of useful packets received over the total number of packets transmitted. In the transmission scenario, two things are possible once the packet is transmitted, i) the packet will not receive at the receiver side, and ii) the packet may corrupt by the time it is received. If the packet is corrupt in practicality, there will be syndrome bits that can be used to correct the corrupted packets. However, the proposed study does not consider the syndrome bits, but the study considers the communication channel whether the packets transferred are intact or not. Thereby, the system has only one control variable, i.e., whether the signal is being jammed or not. The Q-learning algorithm always considers discrete frequencies. Since the discrete hopping frequency of Q-learning is always much higher than sweeping width, the communication gets blocked more often in Q-learning than in a continuous system like DQN and the proposed agent algorithm. Hence, while DQN and proposed show a similar SNR ratio, Q-learning shows a much lower SNR ratio. Even if DQN and proposed show a similar SNR ratio, the proposed method consumes less power compared to DQN. Hence, the proposed algorithm is deemed to be better than the other Q-learning and DQN methods. It could be also noted that in order to develop an anti-jamming mechanism over complex network environment, there is a need to fine tune the proposed scheme. However, interestingly, there are just a need of amending the communication topology as well as constraint information within the model. Hence, no reengineering technique is required to ensure that proposed system does operate over complex network environment. The primary reason behind this claim is that proposed system harnesses learning functions using low epoch to generate the optimal result, whereas the existing approaches do demands higher dependency of resource leading to degradation of signal quality in its outcome.

Further, Table IV highlights that the proposed system offers better improvement of power reduction, higher signal quality, and lower processing time in contrast to frequently used Q-Learning and DQN approach. Similar trend is expected with any other existing approach toward problem solution and hence better consistency is noted.

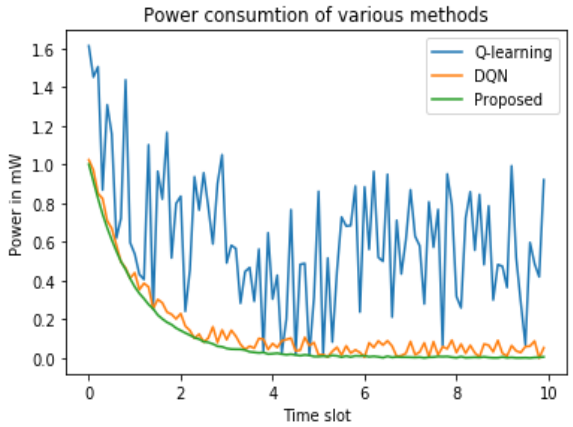

Fig. 11. Analysis of the Power.

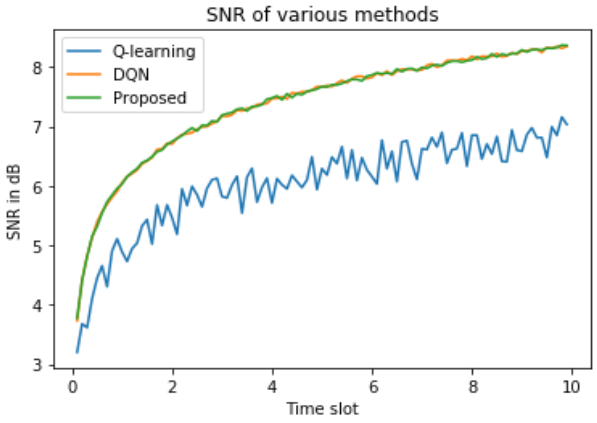

Fig. 12. Analysis of the SNR.

TABLE IV. COMPARISON OF OUTCOMES

Techniques	Improvement in Power Reduction	Improvement in Signal Quality	Processing Time
Q-Learning	87.57%	98.41%	0.26619s
DQN	76.29%	82.63%	0.59851s
Proposed	35.81%	81.59%	0.98791s

VII. Conclusion

The previous works on anti-jamming mechanisms for wireless communication have concentrated on how to prevent jamming attacks but neglected the possibility that jammers could obtain frequency action. However, existing anti-jamming methods can guarantee a solution on a temporary basis. They may fail to ensure efficient performance for the long run, especially where intelligent jammers are deployed. Aiming at this issue, the proposed system has been presented as an efficient and intelligent anti-jamming scheme based on the model-free and off-policy agent mechanism. Overall implementation of the proposed work is carried out in a multi-fold manner, such as 1) Building a customized RL environment for evaluating agent mechanism 2) designing the intelligent anti-jamming agent algorithm to perform the anti-jamming process. The proposed system is more intelligent than the existing works, it has robust environmental applicability, and avoids the additional overhead caused by continuous action space. The simulation results show the proposed system's effectiveness compared to the widely adopted Q-learning and DQN based anti-jamming system. Moreover, the proposed work is limited to modeling the intelligent anti-jamming scheme, and it is not evaluated against intelligent anti-jamming techniques. The current research work will be extended with intelligent jamming scheme, which will be introduced against the proposed anti-jamming solution.

References

[1] Rehmani MH, Dhaou R, editors. The cognitive radio, mobile communications and wireless networks. Springer International Publishing; 2019.

[2] Marinho J, Monteiro E. Cognitive radio: survey on communication protocols, spectrum decision issues, and future research directions. Wireless networks. 2012 Feb;18(2):147-64.

[3] N. Mansoor, A.M. Islam AM, M. Zareei, S. Baharun, T. Wakabayashi, S. Komaki. Cognitive radio ad-hoc network architectures: a survey. Wireless Personal Communications, .vol. 81, Iss.3, pp.1117-42, 2015.

[4] Y. Sudha and V. Sarasvathi, "Evolution of the Security Models in Cognitive Radio Networks: Challenges and Open Issues," 2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT), pp. 1-6, 2020, doi: 10.1109/3ICT51146.2020.9311956.

[5] S.B. Nanthini SB, M. Hemalatha, D. Manivannan, L. Devasena, "Attacks in cognitive radio networks (CRN)-A survey". Indian Journal of science and Technology, vol.1, Iss.7, pp.530, 2014.

[6] S. Bhattacharjee, R. Rajkumari, N. Marchang, "Cognitive radio networks security threats and attacks: a review". International Journal of Computer Applications. 2014; 975:8887.

[7] DK. Jasim, S.B. Sadkhan, "Cognitive Radio Network: Security and Reliability trade-off-Status, Challenges, and Future trend". IEEE 1st Babylon International Conference on Information Technology and Science (BICITS),Apr 28 pp. 149-153, 2021.

[8] M.K. Hanawal, M.J. Abdel-Rahman, M. Krunz, "Joint adaptation of frequency hopping and transmission rate for anti-jamming wireless systems", IEEE Transactions on Mobile Computing, vol.15, Iss.9, pp.2247-59, 2015.

[9] Di Pietro, R. and G. Oligeri, "Jamming mitigation in cognitive radio networks", IEEE Network, vol.27, Iss.3, pp.10-15, 2013.

[10] Xiao L. Spread spectrum-based anti-jamming techniques. InAnti-Jamming Transmissions in Cognitive Radio Networks 2015 (pp. 5-9). Springer, Cham.

[11] K. Ibrahim K, Qureshi IM, Malik AN, Ng SX. Bandwidth-Efficient Frequency Hopping based Anti-Jamming Game for Cognitive Radio assisted Wireless Sensor Networks. In2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring) 2021 Apr 25 (pp. 1-5). IEEE.

[12] L. Xiao, Y. Li, J. Liu, and Y. Zhao, "Power control with reinforcement learning in cooperative cognitive radio networks against jamming," The Journal of Supercomputing, vol. 71, no. 9, pp. 3237–3257, 2015.

[13] L. Xiao, Y. Li, C. Dai, H. Dai, and H. V. Poor, "Reinforcement learning based noma power allocation in the presence of smart jamming," IEEE Transactions on Vehicular Technology, vol. 67, no. 4, pp. 3377–3389, 2018.

[14] Lmater MA, Haddad M, Karouit A, Haqiq A. Smart Jamming Attacks in Wireless Networks During a Transmission Cycle: Stackelberg Game with Hierarchical Learning Solution. International Journal of Advanced Computer Science and Applications (IJACSA). 2018 Apr 1;9(4).

[15] G. Han, L. Xiao, and H. V. Poor, "Two-dimensional anti-jamming communication based on deep reinforcement learning," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2017, pp. 2087–2091.

[16] Han, Guoan & Xiao, Liang & Poor, H. Vincent. (2017). Two-dimensional anti-jamming communication based on deep reinforcement learning. 2087-2091. 10.1109/ICASSP.2017.7952524.

[17] S. Ahmed, Ismail, "Stackelberg-Based Anti-Jamming Game for Cooperative Cognitive Radio Networks", University of Calgary's Digital Repository, Doctorial Thesis, 2017.

[18] F. Ye, F. Che and H. Tian, "Cognitive cooperative-jamming decision method based on bee colony algorithm," 2017 Progress in Electromagnetics Research Symposium - Fall (PIERS - FALL), 2017, pp. 531-537, doi: 10.1109/PIERS-FALL.2017.8293195.

[19] TS. Singh, A. Trivedi, "Anti-jamming in cognitive radio networks using reinforcement learning algorithms", Ninth International Conference on Wireless and Optical Communications Networks, DOI:10.1109/WOCN.2012.6331885, 2012.

[20] H. Liu, H. Zhang, Y. He, and Y. Sun, "Jamming Strategy Optimization through Dual Q-Learning Model against Adaptive Radar", MDPI, vol.22, Iss.145, 2022, https://doi.org/10.3390/s22010145.

[21] K. Ibrahim, S. X. Ng, I. M. Qureshi, A. N. Malik and S. Muhaidat, "Anti-Jamming Game to Combat Intelligent Jamming for Cognitive Radio Networks," in IEEE Access, vol. 9, pp. 137941-137956, 2021, doi: 10.1109/ACCESS.2021.3117563.

[22] Wang, X., Wang, J., Xu, Y., et al.: 'Dynamic spectrum anti-jamming communications: challenges and opportunities', IEEE Commun. Mag., 2020, 58, (2), pp. 79–85.

[23] Amuru, S., Tekin, C., Schaar, M.V., et al.: 'Jamming bandits–A novel learning method for optimal Jamming', IEEE Trans. Wirel. Commun., 2016, 15, (4), pp. 2792–2808.

[24] Furqan HM, Aygül MA, Nazzal M, Arslan H. Primary user emulation and jamming attack detection in cognitive radio via sparse coding. EURASIP Journal on Wireless Communications and Networking. 2020 Dec;2020(1):1-9.

[25] Huang JF, Chang GY, Huang JX. Anti-jamming rendezvous scheme for cognitive radio networks. IEEE Transactions on Mobile Computing. 2016 May 3;16(3):648-61.

[26] Quan, H., Zhao, H. & Cui, P. Anti-jamming Frequency Hopping System Using Multiple Hopping Patterns. Wireless Pers Commun 81, 1159–1176 (2015).

[27] Y. Wu, B. Wang, K. J. R. Liu, and T. C. Clancy, "Anti-jamming games in multi-channel cognitive radio networks," IEEE J. Sel. Areas Commun., vol. 30, no. 1, pp. 4–15, Jan. 2012.

[28] L. Jia, Y. Xu, Y. Sun, S. Feng, and A. Anpalagan, "Stackelberg game approaches for anti-jamming defence in wireless networks," IEEE Wireless Commun., vol. 25, no. 6, pp. 120–128, Dec. 2018.

[29] L. Jia, F. Yao, Y. Sun, Y. Xu, S. Feng, and A. Anpalagan, "A hierarchical learning solution for anti-jamming stackelberg game with discrete power strategies," IEEE Wireless Commun. Lett., vol. 6, no. 6, pp. 818–821, Dec. 2017.

[30] B. Wang, Y. Wu, K. J. R. Liu, and T. C. Clancy, "An anti-jamming stochastic game for cognitive radio networks," IEEE J. Sel. Areas Commun., vol. 29, no. 4, pp. 877–889, Apr. 2011.

[31] Hanawal MK, Abdel-Rahman MJ, Krunz M. Joint adaptation of frequency hopping and transmission rate for anti-jamming wireless

systems. IEEE Transactions on Mobile Computing. 2015 Oct 19;15(9):2247-59.

[32] G.-Y. Chang, S.-Y. Wang, and Y.-X. Liu, "A jamming-resistant channel hopping scheme for cognitive radio networks," IEEE Transactions on Wireless Communications, vol. 16, no. 10, pp. 6712–6725, 2017.

[33] Gao, Y., Xiao, Y., Wu, M., Xiao, M. and Shao, J., 2018. Game theory-based anti-jamming strategies for frequency hopping wireless communications. IEEE Transactions on Wireless Communications, 17(8), pp.5314-5326.

[34] Y. Gwon, S. Dastangoo, C. Fossa, and H. T. Kung, "Competing mobile network game: Embracing anti-jamming and jamming strategies with reinforcement learning," in Proc. IEEE Conf. Commun. Netw. Secur. (CNS), Oct. 2013, pp. 28–36.

[35] Liu, X., Xu, Y., Jia, L., et al.: 'Anti-jamming communications using spectrum waterfall: a deep: reinforcement learning approach', IEEE Commun. Lett., 2018, 22, (5), pp. 998–1001.

[36] Y. Bi, Y. Wu, and C. Hua, "Deep reinforcement learning based multiuser anti-jamming strategy," in Proc. IEEE Int. Conf. Commun. (ICC), May 2019, pp. 1–6.

[37] S. Liu, Y. Xu, X. Chen, and, "Pattern-aware intelligent anti-jamming communication: A sequential deep reinforcement learning approach," IEEE Access, vol. 7, pp. 169204–169216, 2020.

[38] N. Gao, Z. Qin, X. Jing, Q. Ni, and S. Jin, "Anti-intelligent UAV jamming strategy via deep Q-Networks," IEEE Trans. Commun., vol. 68, no. 1, pp. 569–581, Jan. 2020.

[39] G. Han, L. Xiao, and H. V. Poor, "Two-dimensional anti-jamming communication based on deep reinforcement learning," in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), New Orleans, LA, USA, Mar. 2017, pp. 2087–2091.

[40] G. Chen, Y. Zhan, Y. Chen, L. Xiao, Y. Wang, and N. An, "Reinforcement learning based power control for in-body sensors in WBANs against jamming," IEEE Access, vol. 6, pp. 37403–37412, 2018.

[41] X. Lu, L. Xiao, C. Dai, and H. Dai, "UAV-aided cellular communications with deep reinforcement learning against jamming," IEEE Wireless Commun., vol. 27, no. 4, pp. 48–53, Aug. 2020.

[42] Jiang, S. and Xue, Y., 2011. Providing survivability against jamming attack for multi-radio multi-channel wireless mesh networks. Journal of Network and Computer Applications, 34(2), pp.443-454.

[43] Xiao, L., Li, Q., Chen, T., Cheng, E. and Dai, H., 2015, December. Jamming games in underwater sensor networks with reinforcement learning. In 2015 IEEE Global Communications Conference (GLOBECOM) (pp. 1-6). IEEE.

[44] Xiao, L., Xie, C., Chen, T., Dai, H. and Poor, H.V., 2016. A mobile offloading game against smart attacks. IEEE Access, 4, pp.2281-2291.

[45] Zhang, H., Qi, Y., Wu, J., Fu, L. and He, L., 2016. DoS attack energy management against remote state estimation. IEEE Transactions on Control of Network Systems, 5(1), pp.383-394.

[46] Al Mamoori, S., Nizampatnam, M. and Jaekel, A., 2019. Optimal attack-aware RWA for scheduled lightpath demands. SN Applied Sciences, 1(11), pp.1-12.

[47] Xie, C. and Xiao, L., 2016, October. User-centric view of smart attacks in wireless networks. In 2016 IEEE International Conference on Ubiquitous Wireless Broadband (ICUWB) (pp. 1-6). IEEE.

[48] Tu, S., Waqas, M., Meng, Y., Rehman, S.U., Ahmad, I., Koubaa, A., Halim, Z., Hanif, M., Chang, C.C. and Shi, C., 2020. Mobile fog computing security: A user-oriented smart attack defense strategy based on DQL. Computer Communications, 160, pp.790-798.

[49] Qin, J., Li, M., Wang, J., Shi, L., Kang, Y. and Zheng, W.X., 2020. Optimal Denial-of-Service attack energy management against state estimation over an SINR-based network. Automatica, 119, p.109090.

[50] Zhao, C., Wang, Q., Liu, X., Li, C. and Shi, L., 2021. Reinforcement learning based a non-zero-sum game for secure transmission against smart jamming. Digital Signal Processing, 112, p.103002.