

# Smart Jamming Attacks in Wireless Networks During a Transmission Cycle: Stackelberg Game with Hierarchical Learning Solution

Moulay Abdellatif LMATER\*, Majed Haddad†, Abdelillah Karouit† and Abdelkrim Haqiq\*

\*Computer, Networks, Mobility and Modeling laboratory

FST, Hassan 1st University, Settat, Morocco

†LIA/CERI University of Avignon Agroparc,

BP 1228, 84911, Avignon, France

**Abstract**—Due to the broadcast nature of the shared medium, wireless communications become more vulnerable to malicious attacks. In this paper, we tackle the problem of jamming in wireless network when the transmission of the jammer and the transmitter occur with a non-zero cost. We focus on a jammer who keeps track of the re-transmission attempts of the packet until it is dropped. Firstly, we consider a power control problem following a Nash Game model, where all players take action simultaneously. Secondly, we consider a Stackelberg Game model, in which the transmitter is the leader and the jammer is the follower. As the jammer has the ability to sense the transmission power, the transmitter adjusts its transmission power accordingly, knowing that the jammer will do so. We provide the closed-form expressions of the equilibrium strategies where both the transmitter and the jammer have a complete information. Thereafter, we consider a worst case scenario where the transmitter has an incomplete information while the jammer has a complete information. We introduce a Reinforcement Learning method, thus, the transmitter can act autonomously in a dynamic environment without knowing the above Game model. It turns out that despite the jammer ability of sensing the active channel, the transmitter can enhance its efficiency by predicting the jammer reaction according to its own strategy.

**Keywords**—Wireless networks; jamming attacks; game theory; reinforcement learning

## I. INTRODUCTION

Technology and system requirements in the telecommunications domain are changing very rapidly. Over the previous years, since the transition from analogue to digital communications, and from wired to wireless networks, different standards and solutions have been adopted, implemented and modified, often to deal with new and different business requirements. However, in the development of the wireless Next Generation Networks (NGNs) in which the layered architecture is adopted the common challenge of how further improve the resource utilization efficiency and provide better quality-of-service (QoS) is conditioned by the capacity of systems to accommodate changes quickly and with minimum impact on the services already implemented. Furthermore, the flexible topology and the low cost in term of use and setup have motivated the exploration of the wireless NGNs with increasingly higher data rates to meet the rapidly growing demand for wireless access.

Distributed protocols would be required to improve the radio resource utilization and provide high performance for

wireless NGNs. In particular, an integrated design of Medium Access Control (MAC) based on Wireless Random Access (WRA) mechanism may lead to an efficient solution. This is why it is important to design distributed algorithms which can be used by the mobiles to compute the equilibrium strategy and simultaneously achieve the optimal operation points. On the other hand, the basic underlying assumption in legacy WRA protocols is that any concurrent transmission of two or more users causes all transmitted packets to be lost [2]. However, this model does not reflect the actual situation in many practical wireless networks where some information can be received correctly from a simultaneous transmission of several packets. This result is due to the fact that the packet arriving with the highest power has a good chance to be detected accurately, even when other packets are present. The effect of capture on Aloha [9], [10], [11], [18] and on IEEE 802.11 protocol (Carrier Sense Multiple Access-Collision Avoidance (CSMA/CA)) [19], [20], [21] has been studied extensively in the literature and new MAC protocols for channels with capture have been proposed. Furthermore, the full system utilization requires coordination among users which may be impractical given the distributed nature and arbitrary topology changes of wireless collision channels.

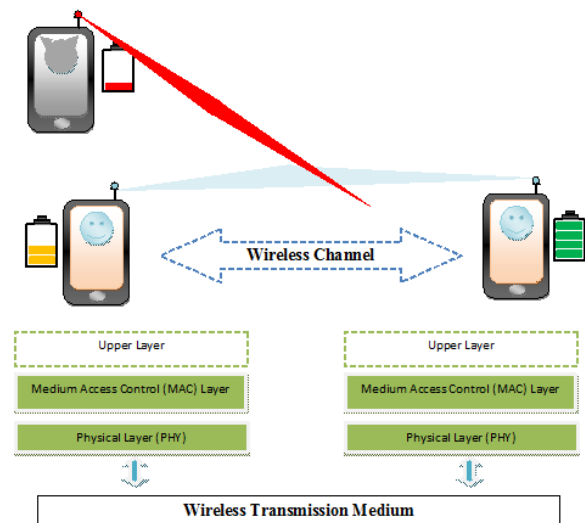


Fig. 1. Layered architecture for wireless networks.

However, while seeking ways to increase the performance of wireless network, there are increasing number of critical security issues that need to be addressed in order to make these wireless NGNs safer [7], [24], [25] (e.g., time-critical services, military operations, etc.). Note that wireless networks are vulnerable to security threats such as distributed denial of service attacks (DoS), spoofing attacks, Sybil attacks, faked sensing attacks and smart jamming attacks [7]. Thus the study of jamming problem in the context of wireless networks is an important challenge since it's easy to destroy communications due to the fact that the jammer can create dynamic and intelligent jamming attacks [23], [5].

The Game theory provides a convenient framework for approaching the power control in wireless based distributed MAC protocols. In fact, given the broadcast nature of the wireless MAC, the users are considered as selfish transmitters [2], and each transmitter seeks to maximize its payoff, while a malicious user tries to degrade the performance of the whole system. In this paper, we consider the IEEE 802.11 MAC CSMA/CA mechanism which is used by a large number of wireless systems, therefore, the problem of jamming can occur during the transmission duty. In addition, the adversary or the jammer has to expend a significant amount of energy to jam the selected frequency bands, also the continuous presence of unusual high interference levels makes these attacks easy to detect. Thus, the main challenge in this paper is to derive the optimal strategy defense against the DoS attacks [16], [3], [4], [8], knowing the fact that the behavior of a malicious user may jam the network by sending abnormal packets to another user to block the channel from doing any things useful (Fig. 1).

It is well known that the Game theoretical approach is an appropriate concept to dealing with the competitive situation. Compared to the approaches used in previous works [12], [13], [14], [15], [17], etc. we are interested here in the impact of a smart jammer on the transmitter power levels during the period that starts at the first attempt of a packet transmission until the next packet transmission first attempt, due to the fact that when re-transmissions are used, the jammers cause the effective network activity factor (and hence the interference among the Receiver Sides (RSs) to be doubled [24]. In particular, we consider a scenario where a single transmitter (player 1) and a single jammer (player 2) coexist. The case of several transmitters/jammers is a subject of future research. Namely, the strategies of both the jammer and the transmitter are their transmission power levels during the packet transmission cycle. Since each packet transmission attempt incurs a cost in term of power, we consider that the Game objective utilities of both players are functions of the Signal to Interference plus Noise Ratio (SINR) value and the transmission cost. Under this anti-jamming Game based on power control problem, we propose two Game formulations, Nash Game where all players act simultaneously and Stackelberg Game where the transmitter is considered as leader (i.e. first to determine its transmit power) while jammer is considered as follower. At first, we derive the Nash Equilibrium (NE) expression, thereafter, we prove the existence of the Stackelberg Equilibrium (SE) and by using the Simulated Annealing Algorithm we sort out the SE measurement. From the comparison of the two schemes, we deduce that the transmitter can efficiently enhance the system performance. The main limitation with regard to the proposed

power control-based anti-jamming problem is that there may be information loss for unknown jamming patterns. Thus, we consider a worst case scenario where the transmitter has an incomplete information while the jammer has a complete information. We introduce a Reinforcement Learning method, thus, the transmitter can act autonomously in a dynamic environment without knowing neither the estimating jamming patterns and parameters nor the above Game model.

The rest of this paper is outlined as follows. We briefly describe the related work in Section II. Then, we introduce the system model and the Game formulation in Section III. In Sections IV and V, we analyze the system in the presence of a regular and a smart jammer. In Section VI we propose a hierarchical learning solution. Simulation results are provided in Section VII. Finally, we conclude the paper and give some perspectives for future research.

## II. RELATED WORK

Designing mechanisms that can be able to detect wireless network jamming as well as avoid it has been widely studied under several works. In [26], authors investigate the anti-jamming problem with discrete power strategies, they formulate a Stackelberg Game to model the competitive interactions between the user and jammer. Then, they analyzed the asymptotic convergence by proposing a hierarchical power control algorithm (HPCA). In [27], a smart jammer can quickly learn the transmission strategies of the legitimate transmitters, and then he would adjust his strategy to damage the legitimate transmission. Meanwhile, the transmitters are aware of the existence of the smart jammer. The difference from [28] is that they consider relay nodes which help the source counteract a smart jammer. Furthermore, in [29] reinforcement learning can be applied to determine transmission powers against a jammer in a dynamic environment without knowing the underlying Game model. In [1], authors propose an anti-jamming Bayesian Stackelberg Game with incomplete information. In all the previous works on anti-jamming, the authors consider the problem transmitter-jammer during only one transmission attempt.

In this paper, we study the power control problem during a packet transmission cycle in the presence of a smart jammer, which has energy-efficiency and keeps track of the re-transmission attempts of the packet until that it is dropped. We suppose that the power level set is continuous and we consider a non-zero Game by introducing a transmission power cost.

## III. SYSTEM MODEL

Let a mobile use IEEE 802.11 CSMA/CA standard which is the most widely known standard in wireless networks. We assume that a transmission fails with probability that depends on the SINR. If a transmission fails then it is attempted again after some back-off time. After a certain number of attempts  $K$  the packet is dropped. Let's assume that the power is controlled. Hence, the power of the mobile user used at the  $i$ th transmission attempt can be denoted by  $T_i \in [0, \bar{T}]$ . Assume that:

$$T_i = p_0 x^{i-1} \quad (1)$$

where  $p_0 \geq 0$  is the initial transmission power and  $x > 1$  is the power multiplier factor for each re-transmission attempt. In

this paper, we examine a scenario with one transmitter, which has its own traffic to send, and one jammer, which does not have its own traffic and simply wants to jam the transmitter attempts. As the mobile user spreads its signal over a common frequency band and treats interference as noise, thus, the signal to interference plus noise ratio at the  $i$ th transmission attempt  $SINR_i$  at the receiver side is given by

$$SINR_i = \frac{\alpha T_i}{N + \beta J_i} \quad (2)$$

where  $N$  is the background noise level on the channel,  $J_i \in [0, J^{max}]$  is the jammer power at the  $i$ th transmission attempt,  $\alpha > 0$  and  $\beta > 0$  denote the fading channel gain of the mobile user and the jammer, respectively.

Since a jammer chooses which transmission or retransmission to jam, we assume that it jams all packets that are in the back-off stage  $k \geq K_2$ , where  $K_2$  is an integer, that means that the competition starts from the back-off stage  $K_2$ . Since the quick detection of the start of a packet is becoming very harder for the jammer and this is due to the large bandwidths and the widely spread signals, we assume the worst situation in which the jammer can jam the communication from the first transmission attempt despite the fact that it arrives at a completely unpredictable time and frequency.

On the other hand, Let's define a cycle as the period that starts from the first attempt of a packet transmission to the first attempt of the next packet transmission. During a cycle, we consider a Game in which the two mobiles are players.

Moreover, we consider that each transmission occurs a certain cost and let  $C > 0$  and  $D > 0$  be the transmission costs per unit power of the mobile user and jammer respectively. We assume that players have perfect knowledge of the environment state and costs constraint at the beginning of each cycle.

Let  $S_t = \{(p_0, x) | 0 < p_0 \leq p_0^{max}; 1 \leq x \leq x^{max}\}$  the feasible set of the power multiplier and the initial transmission power of the mobile user and  $S_j = \{(J_1, J_2, \dots, J_K) | J_i \geq 0; J_i \leq J^{max}\}$  the feasible set of the jammer power vector. We consider the following power control problem where  $(T, J)$  is to be determined, where  $J = (J_1, J_2, \dots, J_K)$  and  $T = (p_0, x)$ .

The mobile user objective is to achieve the maximum  $\sum_{i=1}^K SINR_i$  with the minimum cost. Intuitively, from (1) and (2), the utility function of the mobile user during a cycle denoted as  $U(T, J)$  is given by:

$$U(T, J) = \sum_{i=1}^K \left( \frac{\alpha p_0 x^{i-1}}{N + \beta J_i} - C p_0 x^{i-1} \right) \quad (3)$$

The jammer objective is to achieve the minimum  $\sum_{i=1}^K SINR_i$  with the minimum cost. From (1) and (2), the utility function of the jammer during a cycle denoted as  $V(T, J)$  is given by:

$$V(T, J) = \sum_{i=1}^K \left( -\frac{\alpha p_0 x^{i-1}}{N + \beta J_i} - D J_i \right) \quad (4)$$

#### IV. NASH GAME

In this section, we assume the presence of a regular jammer, and we consider a Game  $G_n = (\{\text{Transmitter, Regular jammer}\}, \{T, J\}, \{U, V\})$ . Since the regular jammer does not have the capability to sense the ongoing transmission power, all players take actions simultaneously. We focus on finding a Nash equilibrium in which neither the transmitter nor the jammer can increase its utility function by unilaterally changing its strategy. we define the Nash Equilibrium by the following formulation:

$$\begin{aligned} T^{NE} &= \text{Arg max}_{T \in S_t} U(T, J^{NE}) \\ J^{NE} &= \text{Arg max}_{J \in S_j} V(T^{NE}, J) \end{aligned} \quad (5)$$

*Theorem 1:* Let a jammer without the intelligence of learning the transmitter strategy. There exists a NE  $(T^{NE}, J^{NE})$  in the Game, in addition,

$$C > \alpha/N \left\{ \begin{array}{l} T^{NE} = (0, 1) \\ J^{NE} = 0 \end{array} \right. \quad (6)$$

$$C < \alpha/(N + \beta J^{max}) \left\{ \begin{array}{l} T^{NE} = (p_0^{max}, x^{max}) \\ J^{NE} = (\text{Min}(J^{max}, \\ \text{Max}(0, \frac{1}{\beta} (\sqrt{\frac{p_0^{max} \alpha \beta}{D}} x^{max i-1} \\ - N))))_{i \in [1, K]} \end{array} \right. \quad (7)$$

$$ow \left\{ \begin{array}{l} T^{NE} = (\frac{\alpha D}{\beta C^2}, 1) \\ J^{NE} = (\frac{1}{\beta} (\frac{\alpha}{C} - N))_{i \in [1, K]} \end{array} \right. \quad (8)$$

*Proof:* By (3) we have:

$$\frac{\partial U(T, J)}{\partial x} = p_0 \sum_{i=2}^K (i-1) \left( \frac{\alpha}{N + \beta J_i} - C \right) x^{i-2} \quad (9)$$

$$\frac{\partial U(T, J)}{\partial p_0} = \sum_{i=1}^K \left( \frac{\alpha}{N + \beta J_i} - C \right) x^{i-1} \quad (10)$$

The first order partial derivative of  $V(T, J)$  with respect to  $J_i$  for  $i \in [1, K]$ , is

$$\frac{\partial V(T, J)}{\partial J_i} = \frac{\alpha \beta p_0 x^{i-1}}{(N + \beta J_i)^2} - D \quad (11)$$

The second order partial derivatives of the jammer objective function are:

$$\frac{\partial^2 V(T, J)}{\partial J_i \partial J_j} = \left\{ \begin{array}{ll} -\frac{2\alpha \beta^2 p_0 x^{i-1}}{(N + \beta J_i)^3} & i = j \\ 0 & ow \end{array} \right. \quad (12)$$

Therefore, the Hessian matrix of  $V(T, J)$  with respect to the vector  $J$  is negative and  $V(T, J)$  is strictly concave in  $J$ . Thus we consider the following cases:

- $C > \alpha/N$ :  
As  $\frac{\partial U}{\partial T} < 0 \forall T \in S_t$ , thus  $x^{NE} = 1$  and  $p_0^{NE} = 0$  yielding  $T^{NE} = (0, 1)$ . By using the concavity of  $V$  in  $J$  and setting  $\frac{\partial V(T, J)}{\partial J_i}$  to zero, we have  $J_i^0 = \frac{1}{\beta} (\sqrt{\frac{p_0 \alpha \beta}{D}} x^{i-1} - N)$ . Since  $0 \leq J_i \leq J^{max}$ , let  $J' =$

$(\text{Min}(J^{\max}, \text{Max}(0, J^0_i)))_{i \in [1, K]}$ . According to Fig. 1, we have  $\forall J \in S_j : V(T^{NE}, J) \leq V(T^{NE}, J')$ . Thus  $J^{NE} = J' = 0$ .

- $C < \alpha / (N + \beta J^{\max})$ :  
As  $\frac{\partial U}{\partial T} > 0 \forall T \in S_t$ , then,  $p_0^{NE} = p_0^{\max}$  and  $x^{NE} = x^{\max}$ . By using the concavity of V in J and setting  $\frac{\partial V(T, J)}{\partial J_i}$  to zero, we have  $J^0_i = \frac{1}{\beta} (\sqrt{\frac{p_0^{\max} \alpha \beta}{D} x^{\max i-1}} - N)$ . Since  $0 \leq J_i \leq J^{\max}$ , let  $J' = (\text{Min}(J^{\max}, \text{Max}(0, J^0_i)))_{i \in [1, K]}$ . According to Fig. 1, we have  $\forall J \in S_j : V(T^{NE}, J) \leq V(T^{NE}, J')$ . Thus  $J^{NE} = J'$ .

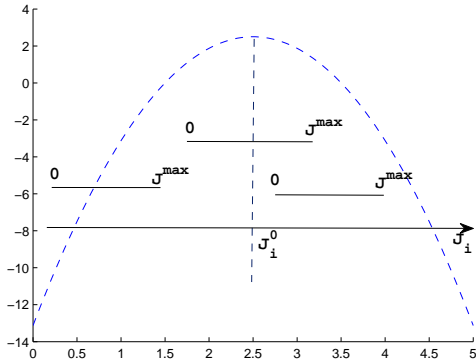


Fig. 2. The assumption of  $V_i(T^{NE}, J_i)$  with respect to  $J_i$ .

- $\alpha / (N + \beta J^{\max}) \leq C \leq \alpha / N$ :  
let  $J' = (\frac{1}{\beta} (\frac{\alpha}{C} - N))_{i \in [1, K]}$ , then  $U(T, J') = 0 \forall T \in S_t$ . By using the concavity of V in J and setting  $\frac{\partial V(T, J)}{\partial J_i}$  to zero, we have  $J^0_i = \frac{1}{\beta} (\sqrt{\frac{p_0 \alpha \beta}{D} x^{i-1}} - N)$ . In order to have  $J^0_i = J'_i$  for  $i \in [1, K]$  we must have  $x = 1$  and  $p_0 = \frac{\alpha D}{\beta C^2}$ , without loss of generality we assume that  $\frac{\alpha D}{\beta C^2} \leq p_0^{\max}$ . As result, we get,  $T^{NE} = (\frac{\alpha D}{\beta C^2}, 1)$  and  $J^{NE} = (\frac{1}{\beta} (\frac{\alpha}{C} - N))_{i \in [1, K]}$ . That means:  
 $\forall T \in S_t : U(T, J^{NE}) = U(T^{NE}, J^{NE}) = 0$ .  
 $\forall J \in S_j : V(T^{NE}, J) \leq V(T^{NE}, J^{NE})$ .

## V. STACKELBERG GAME

We assume the presence of a smart jammer. Since this kind of jammer has the capability to sense the ongoing transmission power, we model this problem as a Stackelberg Game denoted as:  $G_s = (\{\text{Transmitter, Regular jammer}\}, \{T, J\}, \{U, V\})$ , where the leader is the transmitter and the follower is the jammer. Thus, the leader fixes its optimal strategy based on the reaction of the follower, then the follower optimizes its own utility according to the leader strategy, namely, we define the Stackelberg Equilibrium by the following formulation:

$$\begin{aligned} T^{SE} &= \text{Arg max}_{T \in S_t} U(T, \text{Arg max}_{J \in S_j} V(T, J)) \\ J^{SE} &= \text{Arg max}_{J \in S_j} V(T^{SE}, J) \end{aligned} \quad (13)$$

## A. Jammer's Optimal Reaction

Assume that the two players have a complete information about the environment.

*Lemma 1:* Let T be a given strategy of the transmitter. There exists a unique  $J^*(T)$  such that  $J^*(T) = \text{Arg max}_j V(T, j)$ . In addition, the optimal jammer reaction is given by:

$$J_{i \in [1, K]}^*(T) = \begin{cases} 0 & E_1 \\ J^{\max} & E_2 \\ \frac{1}{\beta} (\sqrt{\frac{p_0 \alpha \beta}{D} x^{i-1}} - N) & \text{ow} \end{cases} \quad (14)$$

The conditions are given by:

- $E_1 : x^{i-1} < \frac{N^2 D}{p_0 \alpha \beta}$
- $E_2 : x^{i-1} > \frac{(N + \beta J^{\max})^2 D}{p_0 \alpha \beta}$

*Proof:* According to (4),  $V(T, \cdot)$  is a continuous function on the compact set  $S_j$  and it can achieve its maximum value at some point  $J \in S_j$ . Since the first order partial derivative of the jammer objective function with respect to  $J_i, \forall i \in [1, K]$  is:

$$\frac{\partial V(T, J)}{\partial J_i} = \frac{\alpha \beta p_0 x^{i-1}}{(N + \beta J_i)^2} - D \quad (15)$$

and the second order partial derivatives of the jammer objective function are:

$$\frac{\partial^2 V(T, J)}{\partial J_i \partial J_j} = \begin{cases} -\frac{2\alpha \beta^2 p_0 x^{i-1}}{(N + \beta J_i)^3} & i = j \\ 0 & \text{ow} \end{cases} \quad (16)$$

Therefore, the Hessian matrix of  $V(T, J)$  with respect to the vector J is negative and  $V(T, J)$  is strictly concave in J, [30]. Thus there exists a unique solution  $J^*(T)$  such that  $J^*(T) = \text{Arg max}_{J \in S_j} V(T, J)$ .

On the other hand, by resolving the following equation  $\frac{\partial V(T, J)}{\partial J_i} = 0$ , we have  $J^0_i = \frac{1}{\beta} (\sqrt{\frac{p_0 \alpha \beta}{D} x^{i-1}} - N)$ . Since  $0 \leq J_i \leq J^{\max}$ . 1) If  $J^0_i > J^{\max}$ , yielding  $E_2$ , then  $V(T, J)$  increases in  $S_j$  and thus  $J_{i \in [1, K]}^*(T) = J^{\max}$ . 2) If  $J^0_i < 0$ , yielding  $E_1$ , then  $V(T, J)$  decreases in  $S_j$  yielding  $J_{i \in [1, K]}^*(T) = 0$ . 3) If  $0 \leq J^0_i \leq J^{\max}$ , yielding  $E_3$ , therefore,  $J_{i \in [1, K]}^*(T) = J^0_i$ . Thus, we deduce the property of the optimal jammer strategy given the strategy of the transmitter given in lemma 1. ■

## B. Stackelberg Equilibrium

Let's now focus on analyzing the transmitter objective function given the reaction of the jammer.

*Theorem 2:* There exists  $T^{SE} \in S_t$  such that  $(T^{SE}, J^*(T^{SE}))$  is a Stackelberg Equilibrium of the Game.

*Proof:* To do so, we begin by proving the continuity of  $J^*$  on  $S_t$ . It's obvious that  $J^*$  is continuous in  $S_t \setminus \bigcup_{i=1}^K \{S a_i, S b_i\}$ , where for each  $i \in [1, K]$ ,  $S a_i$  is the set of couple  $(p_0, x) \in S_t$  such that  $p_0 x^{i-1} = \frac{N^2 D}{\alpha \beta}$ , and  $S b_i$  is the set of couple

$(p_0, x) \in S_t$  such that  $p_0 x^{i-1} = \frac{(N+\beta J^{max})^2 D}{\alpha \beta}$ . Let now  $t_{a_i} \in Sa_i$  and  $t_{b_i} \in Sb_i$ , it's clear from lemma 2 that  $\lim_{T \rightarrow t_{a_i}^-} J_i^*(T) = \lim_{T \rightarrow t_{a_i}^+} J_i^*(T) = 0$ ;  $\lim_{T \rightarrow t_{b_i}^-} J_i^*(T) = \lim_{T \rightarrow t_{b_i}^+} J_i^*(T) = J^{max}$ . The continuity of  $J^*$  on all  $S_t$  is proved. Since  $U(T, J)$  is a continuous function on  $S_t \times S_j$ , thus  $U(T, J^*(T))$  is continuous in  $T$ .

Since the set  $S_t$  is compact,  $U(T, J^*(T))$  achieves its maximum at some point  $T^{SE} \in S_t$ . This prove the existence of  $T^{SE} \in S_t$  such that  $(T^{SE}, J^*(T^{SE}))$  is a Stackelberg Equilibrium of the Game. ■

**U(T, J\*(T)) is not a concave function::** Despite we proved the existence of a SE  $(T^{SE}, J^*(T^{SE}))$ , calculating the SE is a challenging due to the non-concavity of the function  $U(T, J^*(T))$ . We use an example to show that there exists  $(T_1, T_2, t) \in S_t^2 \times ]0, 1[$  where:  $U(t.T_1 + (1-t)T_2, J^*(t.T_1 + (1-t)T_2)) < t.U(T_1, J^*(T_1)) + (1-t).U(T_2, J^*(T_2))$

Let  $N = 0.2; E = 1; C = 0.1; p_{0_1} = p_{0_2} = 2; k = 10; \alpha = \beta = 0.5; x_1 = 1.05, x_2 = 1.1, t = 0.63$ .

In this example we have,  $U(t.x_1 + (1-t)T_2, J^*(t.T_1 + (1-t)T_2)) = 13.7463$ ,  $U(T_1, J^*(T_1)) = 13.3062$ ,  $U(T_2, J^*(T_2)) = 14.5018$ .

Hence,  $U(t.T_1 + (1-t)T_2, J^*(t.T_1 + (1-t)T_2)) - t.U(T_1, J^*(T_1)) + (1-t).U(T_2, J^*(T_2)) = -0.0023$ . Thus  $U(T, J^*(T))$  is not a concave function on the set  $S_t$ . This results proves the complexity of finding a closed form of the global optimum, that's why we propose a simulated annealing technique as shown in Algorithm 2 in order to approximate the global optimum of our given function  $U(T, J^*(T))$ .

---

**Algorithm 1** Calculate  $T^{SE} = Arg \max_{T \in S_t} U(T, J^*(T))$

---

**Require:**  $T \in S_t$   
Initialize the system parameters.  
Initialize G with a large value.  
T0=[0,1];  
**while** ( $G \neq 0$ ) **do**  
    **while** (Accepted states number is below a threshold level)  
    **do**  
        Pick a random neighbor,  $T_{new} \leftarrow neighbour(T)$   
         $\delta T = neighbour(T) - T_{new}$   
         $\delta U = U(T_{new}, J^*(T_{new})) - U(T, J^*(T))$   
        **if**  $\delta U > 0$  **then**  
             $T \leftarrow T_{new}$   
        **else**  
             $T \leftarrow T_{new} + \delta T \cdot \exp(\delta U/G)$   
        **end if**  
    **end while**  
     $G \leftarrow G - 1$   
**end while**

---

## VI. ANTI-JAMMING WITH REINFORCEMENT LEARNING

Reinforcement Learning (RL) is considered as a method in which the player takes action in a current time step and receives the corresponding reward in the next time step to evaluate its previous action [6]. RL is capable of solving more complex problems, specially, as the player does not require knowledge about the environment reaction and the

reward function. However, the player learns just from previous experiences by interacting with the environment.

Through the above Game model, where both the transmitter and the jammer have a complete information of each other (i.e., channel gain and transmission cost), the SE strategies are derived. However, in view of the fact that Neither the jammer physical location nor its transmission cost is known by the transmitter due to the assumption that firstly, the jammer can change its physical location in a completely unforeseen time; secondly, the value of the jammer's transmission cost is not shared over the channel. Consequently, we introduce a reinforcement learning technique, especially the Q-learning method, so that the transmitter can act autonomously in a dynamic environment without knowing the above Game model.

We assume that the transmitter can choose its power multiplier and its initial transmission power from  $M$  and  $N$  levels respectively. Let  $P$  and  $A^{MN}$  denote the power action taken by the transmitter and the set of power action respectively. Meanwhile, the state observed by the transmitter is denoted by  $s_t^n$ . In each transmission cycle, the transmitter and the jammer take actions sequentially, we denote by  $J$  the jammer power action. At the beginning of the n-th transmission cycle, the transmitter first takes action and the decision making of its power action  $P_n$  is based on the transmission state in the previous transmission cycle, i.e.,  $s_t^n = (J_{n-1})$ . sequentially, based on the observed state  $s_j^n = (P_n)$ , the jammer chooses its optimal power  $J_n$  given by (14). The received utility value of the transmitter is denoted by  $u_n$ . Let now describe the anti-jamming power control strategy based on Q-learning. Let  $\alpha_t$  and  $\beta_t$  denote the learning rate and the discount factor of the transmitter. The Q-function with the power action  $P$  in the state  $s_t$  is denoted by  $Q(s_t, P)$ . The maximum Q value in the state  $s_t$  is denoted by  $V(s_t)$ . We define the update rule of the Q-function in the n-th transmission cycle as follows:

$$Q(s_t^n, P_n) \leftarrow Q(s_t^n, P_n) + \alpha_t(u_n + \beta_t V_{n+1} - Q(s_t^n, P_n)) \quad (17)$$

$$V(s_t^n) \leftarrow \max_{\forall P \in A^{MN}} Q(s_t^n, P) \quad (18)$$

As a well-known reinforcement learning method, Q-learning should try to balance between exploration and exploitation according to  $\epsilon$ -greedy policy where the transmitter chooses with a high probability  $1-\epsilon$  the power action that maximizes the Q value in the state  $s_t$  while other power actions are taken with an equal low probability  $\frac{\epsilon}{MN-1}$ . Thus, the probability of power action  $x$  taken by the transmitter is given by the following formulation:

$$Pr(P = x) = \begin{cases} 1 - \epsilon & x = \underset{\forall P \in A^{MN}}{\operatorname{argmax}} Q(s_t^n, P) \\ \frac{\epsilon}{MN-1} & \text{ow} \end{cases} \quad (19)$$

Anti-Jamming Strategy of the transmitter with Q-learning is shown in detail as Algorithm 2.

## VII. SIMULATION RESULTS

In this section, numerical results are performed to evaluate the performance of the proposed power control problem during

**Algorithm 2** Anti-Jamming Strategy of the transmitter with Q-learning

---

**Require:**  $P \in S_t$   
 Set the system parameters:  $\beta_t, \alpha_t, \epsilon$ , episode  
 Set  $s_t, A^{MN}$   
 Initialize  $Q(s_t, P), V(s_t)$  as zero  $\forall s_t, P \in A^{MN}$   
**while** (*episode*  $\neq 0$ ) **do**  
   Set the starting state  $s_t^1$   
   For  $n = 1, 2, 3, \dots$   
   Observe the current state  $s_t^n$   
   Pick a random power  $P_n$  from  $s_t^n$  according to Equation (19).  
   **if**  $P_n = \operatorname{argmax}_{P \in A^{MN}} Q(s_t^n, P)$  **then**  
     Decrease  $\epsilon$   
   **end if**  
   Observe the next state  $s_t^{n+1}$  and  $u_n$   
   Update  $Q(s_t^n, P_n)$  by Equation (17).  
   Update  $V(s_t^n)$  by Equation (18).  
   Break if convergence: small deviation on  $Q$   
   *episode*  $\leftarrow$  *episode*  $- 1$   
**end while**

---

a cycle in both scenarios: 1) Transmitter against smart jammer (in which the jammer has the intelligence to quickly learn the transmission power of the transmitter and adjust its own transmission power). 2) Transmitter against regular jammer (in which both players play the Game simultaneously in a non-cooperative manner).

Among all the system variables, only fading channel gains of the transmitter and the jammer, may vary significantly due to the fact that the players can change their physical locations. Thus, we investigate the relations of the utilities of all players in equilibrium with respect to  $\alpha$  and  $\beta$ . Let  $N = 1, D = 0.2, C = 0.2$  and  $K = 10$ .

Fig. 3 shows the impact on the Utility function with respect to  $\alpha$  at NE and SE. We observe that, as  $\alpha$  increases, transmitter's SE utility increases while jammers' SE utility decreases; this phenomenon is due to the fact that the larger  $\alpha$  became, the closer the transmitter became from the receiver. In addition, we depict in Fig. 4 the Utility function at NE and SE of both players with respect to  $\beta$ . As we can remark, the transmitter's utility at the SE decreases with  $\beta$ , while the jammer's utility increases with it; this is due to the fact that the larger  $\beta$  became, the closer the jammer became from the receiver. Moreover, in both Fig. 3 and 4, the transmitter at the NE has a lower utility than that at the SE, because at the latter the transmitter knows the existence of a jammer and utilizes its transmit power more efficiently. Similarly, a jammer obtains a higher utility at the SE than that at the NE, due to its ability to learn and adjust its own power according to the ongoing transmission power. This results proves that despite the jammer ability of sensing the active channel, the transmitter can enhance its efficiency by predicting the jammer reaction according to its own strategy.

Let now consider a scenario with power control strategy based on Q-learning. In this simulation, we set  $M = N = 20$  and we set the maximum episode numbers in the learning to 120 in order to ensure the transmitter can learn an optimal

action. The learning rate  $\alpha_t = 0.8$  which indicates how far the current estimate value of  $Q$  is adjusted toward the update target value of  $Q$ . The discount factor of the source  $\beta_t = 0.8$  that indicates the increasing uncertainty about rewards that will be received in the future. We assume a transmitter that does not have a complete knowledge about the dynamic environment, while the jammer has these knowledge. The initialization of the  $\epsilon$  value for greedy algorithm is starting from 0.5 to ensure that the transmitter can try all actions in all states repeatedly. The utility of the transmitter received by the receiver according to the learning episodes are shown in Fig. 5. We can remark that the utility of the transmitter converges towards the solution proved in the above model. This result validate the proposed power control model. Note that, as the transmitter is gradually aware of the dynamic environment with the learning episodes increasing, which indicates a well anti-jamming performance. This is due to the fact that the transmitter chooses a more proper power action after has a well knowledge about the environment.

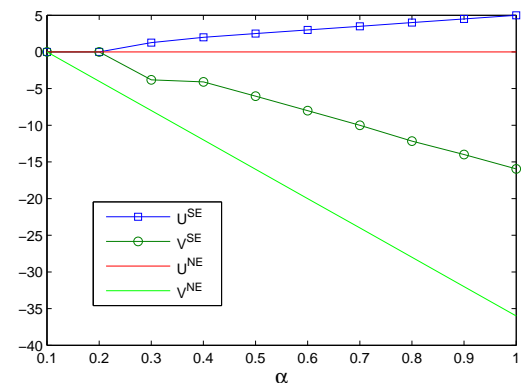


Fig. 3. The impact of  $\alpha$  on the utility function of Jammer/transmitter at NE and SE.  $\beta=0.5$ .

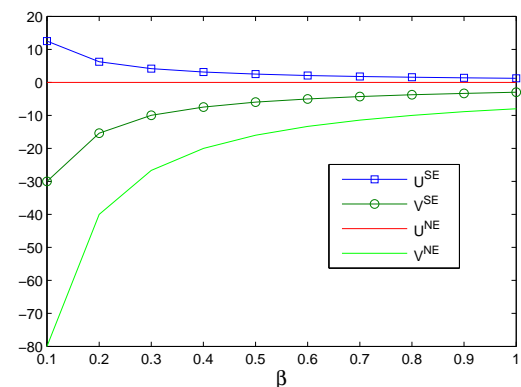


Fig. 4. The impact of  $\beta$  on the utility function of Jammer/Transmitter at NE and SE.  $\alpha=0.5$ .

VIII. CONCLUSION

In this paper we studied denial of service vulnerability in wireless networks in the presence of jamming attacks. We choose a Game theoretical approach which is an abstract concept that indicates how the final outcome of a competitive



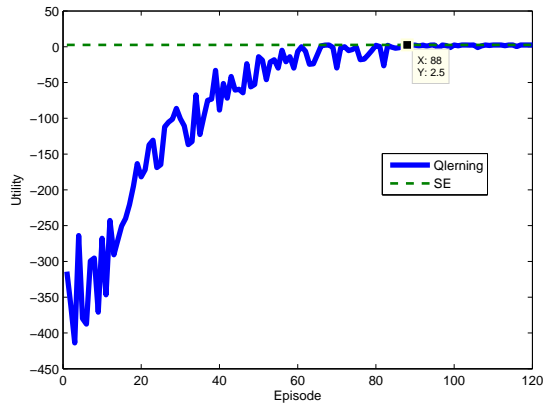


Fig. 5. Utility function of the transmitter, where the transmitter action is chosen based on Q-learning.

situation is dictated by interactions among the players. We considered a jamming during a transmission cycle. We studied the case where all players take action simultaneously and the case where the transmitter is the leader and the jammer is the follower. We proposed a Nash Game in the simultaneous Game and a Stackelberg Game in the hierarchical Game. A closed form of Nash Equilibrium is derived, then, we proved the existence of Stackelberg equilibrium. We sorted out the Stackelberg problem by using a simulated annealing algorithm. Moreover, we studied the relations of the utilities of all players in Nash and Stackelberg equilibrium. In order to validate our Stackelberg model, Q-learning method can be considered to be used by the transmitter to determine their transmission power actions in the presence of a smart jammer in a dynamic environment without knowing the underlying Game model. Simulation results have verified that despite the jammer ability of sensing the active channel, the transmitter can enhance its efficiency by predicting the jammer reaction according to its own strategy. Finally, this work can be extended to the case of several jammers that operate on a single sub-carrier during a single time slot in order to investigate the interaction among jammers who have interest to damage the source node transmission.

#### REFERENCES

- [1] L. Jia, F. Yao, Y. Sun, Y. Niu and Y. Zhu "Bayesian Stackelberg Game for Anti-jamming Transmission with Incomplete Information." IEEE Communications Letters, vol. 20, no. 10, p. 1991-1994, 2016.
- [2] Lmater Moulay Abdellatif, Abdelillah Karout, and Abdelkrim Haqiq. "An efficient pricing mechanism of random access in wireless network with self-interested mobile users." Wireless Networks and Mobile Communications (WINCOM), 2015 International Conference on. IEEE, 2015.
- [3] Guan, Yanpeng, and Xiaohua Ge. "Distributed secure estimation over wireless sensor networks against random multichannel jamming attacks." IEEE Access, vol. 5, p. 10858 - 10870, June 2017.
- [4] Koh, Jing Yang, and Pengfei Zhang. "Localizing Wireless Jamming Attacks with Minimal Network Resources." International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage. Springer, vol. 10658, p. 322-334, 2017.
- [5] Jaitly, Sunakshi, Harshit Malhotra, and Bharat Bhushan. "Security vulnerabilities and countermeasures against jamming attacks in Wireless Sensor Networks: A survey." Computer, Communications and Electronics (Comptelix), 2017 International Conference on. IEEE, 2017.

- [6] Watkins, Christopher JCH, and Peter Dayan. "Q-learning." Machine learning 8.3-4 (1992): vol. 8, no. 3-4, p. 279-292, May 1992.
- [7] Pelechrinis, Konstantinos, Marios Iliofotou, and Srikanth V. Krishnamurthy. "Denial of service attacks in wireless networks: The case of jammers." IEEE Communications surveys and tutorials 13.2, p. 245-257, 2011.
- [8] Dhuria, Shivam, and Monika Sachdeva. "Detection and Prevention of DDoS Attacks in Wireless Sensor Networks." Networking Communication and Data Knowledge Engineering. Springer, Singapore, vol. 3, p. 3-13, 2017.
- [9] Arnbak, J. E. N. S. C., and Wim Van Blitterswijk. "Capacity of slotted ALOHA in Rayleigh-fading channels." IEEE Journal on Selected Areas in Communications 5.2, p. 261-269, 1987.
- [10] Nguyen, Gam D., Anthony Ephremides, and Jeffrey E. Wieselthier. "Comments on 'Capture and retransmission control in mobile radio'." IEEE Journal on Selected Areas in Communications 24.12, p. 2340-2341, 2006.
- [11] LaMaire, Richard O., Arvind Krishna, and Michele Zorzi. "On the randomization of transmitter power levels to increase throughput in multiple access radio systems." Wireless Networks 4.3, p. 263-277, 1998.
- [12] Guan, Yanpeng, and Xiaohua Ge, "Distributed Attack Detection and Secure Estimation of Networked Cyber-Physical Systems Against False Data Injection Attacks and Jamming Attacks". IEEE Transactions on Signal and Information Processing over Networks, vol. 4, no. 1, p. 48-59, March 2018.
- [13] Tayebi, Arash, Stevan Berber, and Akshya Swain. "Performance analysis of chaotic DSSS-CDMA synchronization under jamming attack." Circuits, Systems, and Signal Processing, vol. 35, no. 12, p. 4350-4371, December 2016.
- [14] Kim, Yongchul, Young-Hyun Oh, and Jungho Kang. "Asynchronous Channel-Hopping Scheme under Jamming Attacks." Mobile Information Systems, vol. 98, no. 4, p. 3583-3610, February 2018.
- [15] Wei, Xianglin, et al. "Collaborative mobile jammer tracking in multi-hop wireless network." Future Generation Computer Systems, vol. 78, no. 3, p. 1027-1039, January 2018.
- [16] Li, Li, et al. "Security estimation under Denial-of-Service attack with energy constraint." Neurocomputing, vol. 292, no. 31, p. 111-120, May 2018.
- [17] Peng, Lianghong, et al. "Energy efficient jamming attack schedule against remote state estimation in wireless cyber-physical systems", vol. 272, no. 10, p. 571-583, January 2018.
- [18] Sant, Jeetendra, and Vinod Sharma. "Performance analysis of a slotted-ALOHA protocol on a capture channel with fading." Queueing Systems 34.1-4, p. 1-35, 2000.
- [19] Manshaei, Mohammad Hossein, et al. "Performance analysis of the IEEE 802.11 MAC and physical layer protocol." World of Wireless Mobile and Multimedia Networks, 2005. WoWMoM 2005. Sixth IEEE International Symposium on a. IEEE, 2005.
- [20] Nyandoro, Alfandika, Lavy Libman, and Mahbub Hassan. "Service differentiation in wireless lans based on capture." Global Telecommunications Conference, 2005. GLOBECOM'05. IEEE. Vol. 6. IEEE, 2005.
- [21] Hadzi-Velkov, Zoran, and Boris Spasenovski. "Capture effect in IEEE 802.11 basic service area under influence of Rayleigh fading and near/far effect." Personal, Indoor and Mobile Radio Communications, 2002. The 13th IEEE International Symposium on. Vol. 1. IEEE, 2002.
- [22] X Ge, J Yang, H Gharavi, Y Sun, "Energy Efficiency Challenges of 5G Small Cell Networks" in IEEE Communications Magazine, vol. 55, no. 5, pp. 184 - 191, May 2017.

- [23] Pelechrinis K, Iliofotou M, Krishnamurthy SV, "Denial of Service Attacks in wireless networks: The Case of Jammers" IEEE Communications Surveys and Tutorials , vol. 13, no. 2, pp. 245 - 257, May 2010.
- [24] SD Amuru, HS Dhillon, RM Buehrer, "On Jamming Against wireless networks" IEEE Transactions on Wireless Communications, vol. 16, no. 1, pp. 412 - 428, November 2016.
- [25] C. Kaufman and R. Perlman, M. Speciner, "Network security: private communication in a public world". Prentice Hall Press, 2002.
- [26] L. Jia, et al., "A Hierarchical Learning Solution for Anti-Jamming Stackelberg Game With Discrete Power Strategies" IEEE Wireless Communications Letters , vol. 6, no. 6, pp. 818 - 821, August 2017.
- [27] Y. Li, L. Xiao, J. Liu and Y. Tang "Power control stackelberg Game in cooperative anti-jamming communications". In Int. Conf. Game Theory for Networks (GameNETS), pp. 1-6, Nov. 2014.
- [28] D. Yang, J. Zhang, X. Fang and A. Richa "Optimal transmission power control in the presence of a smart jammer". In : Global Communications Conference (GLOBECOM), 2012 IEEE, p. 5506-5511, December 2012.
- [29] XIAO, Liang, LI, Yan, LIU, Jinliang, et al., "Power control with reinforcement learning in cooperative cognitive radio networks against jamming". The Journal of Supercomputing, vol. 71, no. 9, p. 3237-3257, 2015.
- [30] Boyd, Stephen, and Lieven Vandenberghe. "Convex optimization". Cambridge university press, 2004.