ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 5, May 2018

Protect Against Blocking the Stochastic Game for Cognitive Radio Networks

S. Udhaya Kumar
Ph.D Computer Science at Joseph Arts & Science college.,Thiruvalluvar University

Abstract: **- Several spectrum management schemes have been proposed in recent years to improve the spectrum use in cognitive radio networks. Few consider the existence of cognitive attackers who can adapt their attack strategy to the environment with a different time spectrum and the secondary consumer strategy In this article, we investigate the security mechanism when secondary users face jamming attack and offer a stochastic game environment to protect against jamming. At each stage of the players, the secondary users observe the availability of the radio spectrum, the quality of the channel and the strategy of attack through the state of the channels detected. Based on this observation, they will decide how many channels are you must reserve for the transmission of control and data messages and how to switch between different channels. By using "mini-learning" training, secondary users can gradually learn the optimal policy that maximizes the expected amount of discounted wages, defined as spectral efficiency. The proposal fixes the anti-jamming policy has shown that it achieves much better results than those achieved of myopic learning, which only maximizes the payment of each phase and a strategy of random defense since successfully assume the dynamics of the environment and the strategic behavior of cognitive aggressors.**

Keywords— Security mechanism, spectrum management, cognitive networks, game theory, reinforcement study.

I. INTRODUCTION

In recent years, cognitive radio technology [1] [2] has been proposed as a promising communication standard to resolve the conflict between scarce resources and the growing demand for wireless services. Through the use of the radio spectrum opportunistically, cognitive radio allows secondary users to understand what part of the spectrum is available, to choose the best channel available to coordinate access to the spectrum with other users and to release the channel when the main the user recovers using the spectrum on the right. Efficient use of spectrum resources has been suggested in the literature for different approaches to spectrum management as a method of spectrum allocation based on price [3] - approaches [10], where the main users, including secondary ones available , Audiovisual-based spectrum, based on the concept and the reflexive model of primary user access [11] - [13]. Although the proposed approach demonstrated that it is capable of improving the use of the spectrum, or to obtain economic benefits from the main users, most of them based on the assumption that consumers seek to maximize the use of the spectrum or in cooperation, it coordinates the same network driver and serves a common purpose or trust when low autonomous users want to maximize their own time. This is not the case when secondary users in a hostile environment where the evil attackers try to harm legitimate users and avoid the effective use of the spectrum. Therefore, the way to guarantee the distribution of the spectrum is fundamental for the widespread use of cognitive radio technology. Bad attackers can initiate different types of attacks at different levels of the cognitive radio network. The authors [14] investigated the attack on the main consumer, where Gnostic intruders mimic the main signal to prevent users from accessing licensed secondary spectrum. It has the position of a protection mechanism that controls the source of the signals that are detected by monitoring the characteristics of the signal and evaluating its location. The work in [15] investigated the false attack sensor data and proposed a theoretical probabilistic analysis in relation to alleviating performance degradation due to sensor error. Other potential problems related to security attacks, such as denial of service in cognitive networks, are discussed in [16] and [17]. However, most of these developments [16] [17] provide only qualitative analysis of the opposition and take into account the real dynamics in the spectral environment and the cognitive capacity of the attackers to adjust their attack strategy. In this paper we focus on the participation in the attack of a cognitive radio network and we recommend a reflective game context for defense planning against blocking, which can accommodate the dynamic capacity spectrum, the quality of the channels and the minor changes

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 5, May 2018

Importance of users and invasive strategies Attack coupling has been studied extensively in wireless networks, and existing solutions in the insert include natural defense surfaces, such as directional antennas [19], and spectral dispersion [20].], ligament layer defenses such gates [22], [23]], [24], [25] and network layer defenses, such as space seizures [26]. However, it is not directly applicable to cognitive networks; spectrum availability continues to change with key users to re-evacuate to license totals. For example, work in [25] suggested the use of error correction codes (n; m) to ensure reliable data communication with high performance. However, this approach requires that whenever there are at least n channels available, which cannot be satisfied if many licensed files occupy primary users. In addition, most projects assume that attackers adopt a firm strategy that will not change over time. However, if the attackers are also equipped with cognitive radio technology, it is very likely that they will adjust their attack strategy based on the dynamics of the environment and the strategy of the secondary users. Therefore, in our work, we are going to model the strategic and potential competition between the cognitive secondary users and the attackers as a reflective game with zero sums. To ensure reliable transmission, we intend to maintain multiple channels to transmit control messages and the control channels must be changed from time to time with the data channels according to the invaders' strategy. We define the availability of the spectrum, the quality of the channel and the observation of the action of the attackers as the state of the game. The action of the secondary users is defined as the number of control channels or data to maintain and how to rotate the control and data channels and its objective is to maximize the performance efficiency RF side defined as the relationship between the expected strong performance total number of active channels used to transmit control and data messages. Using the Minimax-Q learning algorithm, secondary users can obtain the optimal policy with proven convergence. The results of the simulation show that when channel quality is low, secondary users must maintain many data channels and some control channels to improve performance. Since the quality of the channel improves, they should maintain more control channels to guarantee the reliability of the communication. When channel quality increases further, secondary users must be more conservative by maintaining fewer data channels to improve spectrum efficiency Traffic. In states where there are some control or data commands, secondary users must adopt a mixed strategy to avoid serious participation next time. When there is more than one licensed area available, the decision making of the attackers becomes more difficult and the secondary users can take more aggressive measures with more data channels. It also shows that secondary users can achieve higher profits using the stagnant policy of learning Minimax-Q instead of using myopic learning and random strategy.The rest of the paper is organized as follows. In Chapter II we present the system model for the network of secondary users and the defense against blocking. In Section III, we formulate defensive defenses as a stochastic game that defines states, actions, objective functions, and transition rules. In Section IV, we take the optimal policy of the secondary user network using the Minimax-Q learning algorithm. Chapter V presents the results of the simulation, followed by the conclusions of Section VI.

II.SYSTEM MODEL

In this section, we present the assumptions of the model about the network of minor users and the defense against blocking against malicious attackers.

A. Network of secondary users: This paper examines a dynamic spectrum access network where many secondary users equipped with cognitive radio can temporarily access the unused radio spectrum channels that belong to many primary users. There is a secondary base station in the network, which coordinates the use of the spectrum of all secondary users. To avoid harmful collisions or interference with primary users, secondary users must listen to the spectrum before each transmission attempt. We assume that the subnet is a system with slots and at the beginning of each time interval, secondary users must reserve a certain amount of time to detect the presence of a primary user. Several detection techniques are available, such as power detection or feature detection if secondary users know some prior information about the master user's signal. In the exchange of collaborative spectrum, such as a spectrum auction, secondary users can avoid harmful interference by listening to the announcement of the main users about whether to share licensed channels with secondary users. To simplify the analysis, we assume the perfect sense or the shared spectrum in this project. Therefore, the user of the subnet can get everything opportunity to exploit the spectrum with unused license and evacuate the spectrum whenever a primary user recovers the rights to the radio spectrum.

Due to the activity of the main users and the variants of the channel, spectrum availability and quality continue to change. To coordinate the use of the radio spectrum and achieve efficient use of the spectrum, it is necessary to

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 5, May 2018

exchange the necessary control messages between the secondary base station and the secondary users through dedicated control channels1. The control channels serve as a medium that supports high-level network operations such as access control, channel assignment, spectrum transfer, etc. If secondary messages are not received correctly by secondary users or the base station, some network functions will be affected.

B. Defense against interference in radio frequency cognitive networks*:* The radio block is a denial of service (DoS) attack that aims to stop communication in the physical and connection layers of a wireless network. Keeping the wireless spectrum busy, for example, injecting packets continuously into a shared spectrum [22], a docking attacker can prevent legitimate users from accessing an open-spectrum band. Another type of compromise is to inject high insertion force around the area of a victim [18] [26], so the signal-to-noise ratio (SNR) deteriorates and does not obtain correct data.

In a cognitive radio network, malicious attackers can launch an attack to prevent the effective exploitation of spectrum opportunities. In this work, it is assumed that the characteristics of the signal transmitted from the main users and the split secondary users and the attackers also listen to permission zone where the secondary users reach. Attackers block the transmission of secondary users, and do not block unlicensed bands, where the main users are active, either because it can be great penalty attackers if their identities are known by their main users or they do not because the Attackers can approach the main users. In addition, due to the limitation of the number of antennas and / or the total power, we assume that Attackers can block most of the N channels at any time. Subsequently, the aim of the attackers is to cause the greatest damage to the network of secondary users with limited interlocking capacity.

Given the limited docking feature, attackers can adopt a strategy of attacks against as many channels of data to reduce the gain of the secondary user transmission network data. On the other hand, if the number of control channels is less than N, the amount of data the channels are larger than N, the attackers can try to point to the control channels to make the attack even stronger. If the secondary user network adopts a channel definition scheme to transmit data and control messages, a cognitive intruder can capture said pattern2, distinguish between data channels and control channels and orient only data or control channels and cause the greatest hurt

Therefore, secondary users must perform channel switching / switching to alleviate potential damage due to a channel definition program. As shown in Figure 1, the channels used to transmit data / control messages in this time position can no longer be data / control channels in the next time period. By entering random data in the channel assignment, the access pattern of the secondary user becomes more unpredictable. Subsequently, the attackers must also strategically change the channels they will attack over time. Therefore, channel hopping is more resistant to interference from a fixed channel assignment. When designing the channel skip mechanism in a cognitive radio network, secondary users should consider the following events.

There is a commitment in the selection of a correct amount of control channels. The secondary operation of the network depends to a great extent on the correct reception of the control messages. Therefore, it is more reliable to transmit double control messages through multiple channels (ie, control channels). However, if the secondary user network maintains too many control channels, the number of channels to which the data messages are transmitted (ie, the Data Channels) will be small and the achievable gain using the licensed spectrum will be unnecessarily low. Therefore, a good option should be able to balance the risk of an incorrect reception of control messages and the gain of data transmission. For the defense mechanism to be more general, we assume that the secondary user network can choose not to transmit anything on certain channels, even when the empty band is available. This is due to the fact that when the secondary base station believes that it has maintained enough data or control channels under a very serious attack to the interference, sharing more channels for the transmission of messages can only lead to wasted energy and it would be better to leave some inactive channels, if the power consumption is worrisome of the secondary user network.

The channel jump mechanism must adapt to the invader's strategy. This is because attackers can also equip themselves with cognitive radio technology and adapt their strategies based on observing the dynamics of the environment and the strategic spectrum of secondary users. Therefore, secondary users cannot assume that attackers will adopt a strong attack strategy. Instead, they must create a stochastic model that captures the dynamic adaptation of the attackers' strategy, as well as variations in the environmental spectrum.

Under previous assumptions about the model of the system and the attack against interference, it is known that secondary users aim to maximize the use of the spectrum with carefully designed channel switching schedules and malicious attackers that want to reduce the use of the strategy of spectrum commitment Thus, they have objectives and dynamic interactions can be correctly configured as a non-cooperation game (zero) 3. How is access to a range of all secondary users coordinated by the secondary base station and users supposed to be malicious

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 5, May 2018

cooperate to cause more harm to secondary users, we can see all secondary users in the network as a player and all attackers as another player. On the other hand, while the possibility of spectrum, the quality of the channel and the users and the strategies of secondary malicious attackers are changing over time, the game does not cooperate, it must be considered a contemplative environment, that is, and the dynamic defense against the network blocking secondary users must be formulated as a work of reflection.

III. THE FORM OF THE GAME IN ANTI-JAMING STATISTICALLY

Before finding details about stochastic game blocking tool, first let's enter the reflective game to have a general idea. A thoughtful game [28] is an extension of Markov Decision making process (MDP) [29] examining the interactive competence between the different actors. In a game of considered G, there is a set of states characterized by S, and a collection of action sets, $A_1,,,,,,,,,A_K$, one for each player in the game. The game is played in a series of stages. Inthe game then switches to a new random state with a transition probability determined by the current status and action of each player. Meanwhile, at each stage each player receives a payment $T: S \times A_1 \times \ldots \times A_k \to PD(S)$ which also depends on the current situation and the selected actions. The game is played continuously for various stages, and each player attempts to maximize the expected sum of his discounted earnings, $R_i: S \times A_1 \times \ldots . \times A_k \to \mathbb{R}$ where ri_{t+j} is the reward we receive j in the future from player i and γ is the discount factor. After introducing the concepts of a stochastic game, we will shape the anti-engagement game defining every element of the game at the beginning of each stage the game is in some state. After the players choose them and execute them

A. States and actions

We observe a spectrum grouping system in which the network of secondary users can temporarily use the unused bandwidths belonging to the main users of L. From the bandwidth of the various graduates the bands may be different, we assume that each group with a license is divided into a set of adjacent channels with the same bandwidth so, there are N_l channels in the main user band and we believe everything will be busy released when the primary user recovers releases the band. Then we can mean the states of the primary user in l at time t as Plt, whose value can be each Pt l = 1, which means that the primary user l is active at time t, or Plt = 0, which means that the primary user will do so do not use the licensed bar at time t and secondary users have access to the channels on the band. According to some empirical studies on the basic model of consumer access [27], the states Pt I can model it using the Markov chain with two states where the probability of the transition is denoted by the network of secondary users will achieve a certain benefit using the spectrum capacity in licensed bands. The benefit can be defined as a function of data bandwidth, packet loss, delay or other appropriate QoS measurement, and is often a growing characteristic channel quality Due to the variations in the channels of each licensed bandwidth, the quality of the channel can change from one time slot to another, so the gain of using a licensed group also changes We assume that the profits of each channel within the same authorized group are identical time t and can take any value from a set of discrete values, ie. glt 2 fq1; q2; ¢ ¢ ¢; qng Why channel quality (with respect to SNR) is often modeled as Markov End Mark Marker (FSMC) [21], The profitability dynamics of the group of licenses of Group 1 can also be expressed by the FCMS. Note that The benefit that can be obtained from the use of the licensed frequency bands also depends on the state of the main user, i. when the primary user is active in the 1st band (Plt = 1), secondary users can not access it group l, and so glt = 0. Then the FSMC state must be able to capture the joint dynamics of both the access of the main users and the quality of the channel that can be denoted by (Plt; glt).

The transition probability of FSMC with states (Plt; glt) can be obtained in the following way. When the first licensed group is not available for two consecutive time slots, the transition only depends on the first access model that we have

$$p\left(P_l^{t+1}=1, g_l^{t+1}=0 \middle| P_l^t=1, g_l^t=0\right)=p_l^{1\to 1}. \quad (1)$$

When the first band becomes available with a gain of qn at time t + 1, we have

$$p\left(P_l^{t+1}=0, g_l^{t+1}=q_n \middle| P_l^t=1, g_l^t=0\right)=\left(1-p_l^{1\to 1}\right)p_{gl}^{0\to n}, \quad (2)$$

where p0 gl! n denotes the probability that the gain of band l is qn at time t + 1, given that Plt = 1 and Pt + 1l = 0. When the 1st band is available for two consecutive periods of time, we have a state probability of transition as

$$p\left(P_l^{t+1}=0, g_l^{t+1}=q_n \middle| P_l^t=0, g_l^t=q_m\right)=\left(1-p_l^{0\to 1}\right)p_{gl}^{m\to n}, \quad (3)$$

where pm gl! n is the probability that the gain passes from qm at time t to qn at time t + 1. Finally,when the first band becomes inaccessible from time t to time t + 1, the probability of transition is

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 5, May 2018

$$p\left(P_l^{t+1}=1, g_l^{t+1}=0 \middle| P_l^t=0, g_l^t=q_m\right)=p_l^{0\to 1}, \quad (4)$$

Since the transition does not depend on the gain glt at time t.In the above, we discussed the dynamics of returning to the main users / calling the license and the profits from using the licensed spectrum. It is clear that this dynamic will affect the decisions of the secondary users on how to assign channels for the transmission of control and data messages. For example, to make more use of spectrum capabilities, secondary users tend to allocate more channels with higher profits, such as data channels and lower profits as control channels.

However, your channel allocation decisions must also depend the observations of the malicious attackers' strategies that can be suspected that were detected by the attackers. In this way, secondary users must keep a record what channels were detected by the attackers, and what kind of messages was transmitted on detected channels. Since channels within the same authorized group are accepted to have the same benefit, what counts for secondary users is just the number and type blocked channels Based on these assumptions, observations from the secondary user network they are designated f1; Ct; Jl; D t g, where J1; Ct and Jl; Dt denotes the number of control channels and data
which are retained in the first band observed during a time interval t, and 12 f1; 2; ¢ ¢ ¢; Lg. Such observation can be obtained when secondary users do not receive confirmation of receipt of the message of the receiver. Secondary users can not understand if an inactive channel is blocked or not, because messages are not being transmitted on these channels. Therefore, the number of empty channels that receive jam is not monitoring secondary users and will not be seen in the status of Stochastic game.

In summary, the state of the stochastic anti-jamming game at time t is determined by st = fst 1; cm 2; ¢ ¢ ¢; St Lg, where stl = (Plt; glt; Jl; Ct; Jl; Dt) means the state associated with the 1st band. After monitoring the status of each stage, both secondary users and attackers will choose your actions for the current time field. Secondary users can no longer choose before locked channels as control or feeds if they believe the attackers will stay blocked the channels until they found activity in the secondary users. On the other hand, if the attackers believe that secondary users will exit the detected channels, they will Select before the channels not previously linked to interfere; then for the secondary users, who still remain in Previously detected channels may be a better option. When you face such uncertainty for all The strategy of other secondary users and attackers must adopt a random strategy. Secondary users will continue to transmit control or data messages to a part of the previously blocked channels in the event that attackers are more likely to detect previously unrelated channels and start broadcasting some of the previously unattached channels if the attackers are more it will probably continue to block pre-retained channels. Similarly, attackers will attack keep some previously affected channels and start obstructing channels that were not stranded in the past.

In addition, as described in Section II, secondary users may need to change channels make the channel access model more unpredictable for attackers and facilitate the potential Damage caused by traffic jam. Therefore, each time secondary users can change control channels to data or inactive channel and vice versa. If so, when there are Nl channels in each license group l, secondary users will have 3N1 different actions to choose from the first group and QL l = 1 3Nl shares in total. This will complicate decision making by secondary users. Yes make a decision for understanding within a reasonable time, we formulate the set of actions for both players as follows. Keep in mind that more complex modeling of the action will only affect performance, without prejudice to the stochastic framework for the fight against computer hackers.

Mathematically, the actions of the secondary users are defined as at = fat 1; at 2; ¢ ¢ ¢; in LG, in I =(in 1; C1 in 1; D1; in 1; C2; in 1; D2) where the action in 1; C1 (or in 1; D1) means that the secondary network will transmit control messages (or data) in the CI channels (or in l; D1) that are selected identically to the previous ones channels attacked and action on it; C2 (or in l; D2) means that the secondary network will transmit the control (or data) messages on channels C1 (or in l; D1) that are selected identically by jamming previously channels Similarly, the actions of the attacker are defined as J = Fat 1; J; at 2; J; ¢ ¢ ¢; to L; Jg, p in (J1), where J1 (or JJ2) means that the attackers will remain in J1 (or in J2) channels that are selected by previously decoupled (or attacked) channels at the current time T. It can be seen that the previous selection of actions has shaped the uncertainty of the players with respect to each of them the other's strategy in the detected and unwanted channels, and the need to change channels.

A. Transitions and state payments at the stage: With the determined game state and action space, we must analyze the state transition rule. We assume that these players choose their actions in each group independently, then the probability of the transition can be expressed by

$$p(\mathbf{s}^{t+1}|\mathbf{s}^t,\mathbf{a}^t,\mathbf{a}_J^t)=\prod_{l=1}^{L} p(s_l^{t+1}|s_l^t,a_l^t,a_{l,J}^t). \quad (5)$$

Since it is supposed to be the dynamics of the main activity of the users and the variations in the channels to be independent of the actions of the players, the probability of

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 5, May 2018

the transition p (st l + ljst l; at l; l; j) may be greater divided into two parts, i.e

$$p(s_l^{t+1}|s_l^t,a_l^t,a_{l,J}^t)=p(J_{l,C}^{t+1},J_{l,D}^{t+1}|J_{l,C}^t,J_{l,D}^t,a_l^t,a_{l,J}^t)\times p(P_l^{t+1},g_l^{t+1}|P_l^t,g_l^t), \quad (6)$$

where the first term on the right side of (6) represents the probability of the numerical transition from the blocked control and the data channels, and the second term is the transition from state of the main user and state of the channel. Since the second term is derived from (1) - (4), you just have to take the first term for the different cases. Case 1: Pt l = 1. As discussed in Section II, we assume that the attackers will not capture the license
when the main users are active; then when the first group is occupied by the main user in the time interval t, i. Plt = 1, the action of the attackers will be on it; J = (0; 0); and the state variable J_1+ 1 1, C and J1; D t + 1 will be 0. Therefore, when Plt = 1, we have

$$p(s_l^{t+1}|s_l^t,a_l^t,a_{l,J}^t)=p(P_l^{t+1},g_l^{t+1}|P_l^t,g_l^t),\ \text{if } J_{l,C}^{t+1}=0 \text{ and } J_{l,D}^{t+1}=0. \quad (7)$$

Case 2: Ptl = 0. When the first band is available to secondary users according to Jt surveillance 1, C and J1; D t at time t for the state of a sedentary channel in the previous time interval,The secondary network will select the action in l = (d1; in l; C2; in l; D2), and the attackers select action of 1, J = (in 1, J1 in 1, J2). Because the control channels (or data) to block the next interval t + 1 include these control channels (or data) that the secondary network has chosen from the previous two channels not detected and detected when the transition p (Jl; Ct + 1; Jl; Dt + 1jJl; Ct; Jl; Dt; l; J; must consider all possible pairs (nC1; nC2) and (nD1; nD2) where nC1 (or nD1) means number of channels for blocked controls (or data) that were not previously detected, nC2 (or nD2) means number of pre-blocked control channels (or data) with nC1 + nC2 = Jl; Ct + 1, and nD1 + nD2 = J1; D t + 1. Since secondary users choose uniformly on channels C1 (or in l; D1) as control channels (or data) of non-sedated NIJ1; CtJl; Dt channels and attackers Sedated uniformly to 1; channels J1, the probability of controlling channels nC1 and data channels nD1 blocked at the time t can be written by

$$p(n_{C_1},n_{D_1}|J_{l,C}^t,J_{l,D}^t,a_l^t,a_{l,J}^t)=\frac{\binom{a_{l,C_1}^t}{n_{C_1}}\binom{a_{l,D_1}^t}{n_{D_1}}\binom{N_{l,1}^t-a_{l,C_1}^t-a_{l,D_1}^t}{a_{l,J_1}^t-n_{C_1}-n_{D_1}}}{\binom{N_{l,1}^t}{a_{l,J_1}^t}}, \quad (8)$$

l; J1where Nt l, l = Nljl; Ctjl; Dt. Similarly, the transition probability of nC2 and nD2 is expressed as

$$p(n_{C_2},n_{D_2}|J_{l,C}^t,J_{l,D}^t,a_l^t,a_{l,J}^t)=\frac{\binom{a_{l,C_2}^t}{n_{C_2}}\binom{a_{l,D_2}^t}{n_{D_2}}\binom{N_{l,2}^t-a_{l,C_2}^t-a_{l,D_2}^t}{a_{l,J_2}^t-n_{C_2}-n_{D_2}}}{\binom{N_{l,2}^t}{a_{l,J_2}^t}}, \quad (9)$$

Ant l; J2 ¢ where Nt l, 2 = Jl, Ct + Jl, D t denotes the number of channels detected. Then the probability of transition of Jt l, C and Jl, Dt becomes

$$p(J_{l,C}^{t+1},J_{l,D}^{t+1}|J_{l,C}^t,J_{l,D}^t,a_l^t,a_{l,J}^t)=\sum_{n_{C_1}+n_{C_2}=J_{l,C}^{t+1}}\ \sum_{n_{D_1}+n_{D_2}=J_{l,D}^{t+1}} p(n_{C_1},n_{D_1}|J_{l,C}^t,J_{l,D}^t,a_l^t,a_{l,J}^t) \quad (10)$$

$$\times p(n_{C_2},n_{D_2}|J_{l,C}^t,J_{l,D}^t,a_l^t,a_{l,J}^t)$$

Substituting (3) (4) and (10) into (6), we can obtain the probability of state transition.After secondary users and attackers choose their actions, secondary users will transmit the control and data messages on the selected channels, and the attackers will block their selected channels. To coordinate spectrum access and simplify operation, we assume that the same control messages are transmitted on all control channels, and a correct copy of the control information in the time t is sufficient to coordinate the spectrum management in the next time interval t + 1. The gain of a given channel can only be achieved when it is used to transmit data messages and at least one the control channel is not blocked by the attackers. Whereas it costs energy for the secondary users to transmit management and data messages and may be limited by the energy of the Secondary users must achieve the highest gain with limited power. Therefore, payment stage of secondary users can be defined as the expected gain for an active channel. Another explanation from the amortization stage, secondary users want to maximize the effective profit spectrum.
Based on these assumptions, step returns can be expressed by

$$r(\mathbf{s}^t,\mathbf{a}^t,\mathbf{a}_J^t)=T(\mathbf{s}^t,\mathbf{a}^t,\mathbf{a}_J^t)\times\left(1-p^{block}(\mathbf{s}^t,\mathbf{a}^t,\mathbf{a}_J^t)\right), \quad (11)$$

where T (st; at; J) denotes the effective gain of the expected spectrum when not all control channels receive blocked and pblock (st; at; at J) means the probability that all control channels in all L bands are stuck As explained in Section III-A, we assume that the attackers choose the same path in channel J1 Nl above; t1 channels not attacked to obstruct and select on l; Channels J2 of the previous Nl; t2 attacked the channels to detect. So, the probability that a channel will not lock at a time t can can be represented by (1aNt1; Jt1l 1) and (1aNt1; l; 2), respectively. Taking into account the gain of the Glt and if it is assumed that different data are transmitted in different channels, we have the expected

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 5, May 2018

benefit when using them band l as [in 1; D1 (1aNt1; Jt1l 1) + a l; D2 (1aNt1; J2 l; 2)] GLT. Then we can express T (st; at; at J) as

$$T(\mathbf{s}^t,\mathbf{a}^t,\mathbf{a}_J^t)=\frac{\sum_{l=1}^{L}\left[a_{l,D_1}^t\left(1-\frac{a_{l,J_1}^t}{N_{l,1}^t}\right)+a_{l,D_2}^t\left(1-\frac{a_{l,J_2}^t}{N_{l,2}^t}\right)\right]g_l^t}{\sum_{l=1}^{L}\left(a_{l,C_1}^t+a_{l,D_1}^t+a_{l,C_2}^t+a_{l,D_2}^t\right)}, \quad (12)$$

where the denominator denotes the total number of control and data channels. Thus,(12)reflects spectral efficiency. Only when all control channels in each licensed group l are detected can the secondary network be used be blocked. Therefore, the probability of blocking p block (st; at; at J) can be expressed as

$$p^{block}(\mathbf{s}^t,\mathbf{a}^t,\mathbf{a}_J^t)=\prod_{l=1}^{L}\frac{\binom{a_{l,C_1}^t}{a_{l,C_1}^t}\binom{N_{l,1}^t-a_{l,C_1}^t}{a_{l,J_1}^t-a_{l,C_1}^t}}{\binom{N_{l,1}^t}{a_{l,J_1}^t}}\times\frac{\binom{a_{l,C_2}^t}{a_{l,C_2}^t}\binom{N_{l,2}^t-a_{l,C_2}^t}{a_{l,J_2}^t-a_{l,C_2}^t}}{\binom{N_{l,2}^t}{a_{l,J_2}^t}} \quad (13)$$

$$=\prod_{l=1}^{L}\frac{\binom{N_{l,1}^t-a_{l,C_1}^t}{a_{l,J_1}^t-a_{l,C_1}^t}}{\binom{N_{l,1}^t}{a_{l,J_1}^t}}\times\frac{\binom{N_{l,2}^t-a_{l,C_2}^t}{a_{l,J_2}^t-a_{l,C_2}^t}}{\binom{N_{l,2}^t}{a_{l,J_2}^t}},$$

where the first (or second) term in the product represents the probability of all control channels uniformly selected from the previous channels not detected (or detected) in the 1st band blocked at time t.By replacing (12) and (13) again in (11), we can get the gradual reward for the secondary users, and the payment of the attackers is negative from (11).

IV. DEVELOPMENT OF OPTICAL POLICIES OF THE STOCHASTY GAME:

Based on the stochastic anti-interference wording in the previous section of this section, we are discussing how to come up with the optimal strategy, i. optimal protection policy secondary users. In general, secondary users have a long series of data to transmit, and the energy of attackers can afford to block the secondary network for a long time, since the number of the channels hooked in each stage will not exceed N. "In this way, we can assume that anti-jamming It is for an infinite number of stages. In addition, secondary users see remuneration in different ways stages in a different way, for example, delayed messages usually have a lower value for applications that are slow, and the recent compensation must weigh more than the remuneration that will be received in the future. So, the purpose of secondary users is to derive an optimal policy that will increase the expected amount of Discount wages

$$\max \quad E\{\sum_{t=0}^{\infty}\gamma^t r(\mathbf{s}^t,\mathbf{a}^t,\mathbf{a}_J^t)\}, \quad (14)$$

where γ is the discount factor of the secondary user. Stochastic game policy it refers to a distribution of the probability that the action will be established in any state. So, the politics of the secondary the network is marked with ...: S! PD (A), and the attacker policy can be denoted by ... J: S! P D (AJ), where st 2 S, in 2 A and J 2 AJ. Given the current status if defend the policy ... t (or the policy of silencing ... Jt) at time t is independent of the states and actions in all it is said that past times, politics ... (or ... J) are Markov. If the policy is more independent of time, me. ... t = ... t0, since st = st0 policy is said to be immobile. It is known [30] that every stochastic game does not have an empty set of optimal policies and, at least, one of them is stationary. Since the game between the secondary network and the attackers is a a zero-sum game, the balance of each scene is the unique balance of the mini-jack and so on The optimal policy will be unique for each player. To solve the optimal policy we can use the minimax-Q training method [30].

Here the function Q Q (st; at; at J) of step t is defined as the expected discounted profit when secondary users take action, attackers take action on J, both follow theirs stationary policies after. Since the Q function is an estimate of the total amount expected The discounted wages that evolve over time to maximize the effectiveness of the worst case. at any stage, secondary users should treat Q (st; at; at J) as a payment of game matrix, where to 2 A and to J 2 AJ. Given the gain Q (st, at, j) of the game, secondary users can find minimax balance and update the Q value with the value of the game [30]. Therefore, the value of a state in the game against jamming

$$V(\mathbf{s}^t)=\max_{\pi(\mathbf{a}^t)}\min_{\pi_J(\mathbf{a}_J^t)}\sum_{\mathbf{a}^t\in A}Q(\mathbf{s}^t,\mathbf{a}^t,\mathbf{a}_J^t)\pi(\mathbf{a}^t), \quad (15)$$

where *Q*(**s***t; **a**t; **a**t J*) is updated by

$$Q(\mathbf{s}^t,\mathbf{a}^t,\mathbf{a}_J^t)=r(\mathbf{s}^t,\mathbf{a}^t,\mathbf{a}_J^t)+\gamma\sum_{\mathbf{s}^{t+1}}p(\mathbf{s}^{t+1}\mid\mathbf{s}^t,\mathbf{a}^t,\mathbf{a}_J^t)V(\mathbf{s}^{t+1}). \quad (16)$$

it is worthwhile to use the actual updating of the accrued con [31] [32]

$$Q(\mathbf{s}^t,\mathbf{a}^t,\mathbf{a}_J^t)=(1-\alpha^t)Q(\mathbf{s}^t,\mathbf{a}^t,\mathbf{a}_J^t)+\alpha^t\left[r(\mathbf{s}^t,\mathbf{a}^t,\mathbf{a}_J^t)+\gamma V(\mathbf{s}^{t+1})\right], \quad (17)$$

where "adjustment" means the degree of learning that decays with time by adjustment + 1 = "adjustment, with

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 5, May 2018

0 <" <1 and V (st + 1) it is obtained by (15). In the modified update in (17), the current state status V (st + 1) is used as approximately the expected decrease in future remuneration, which will be improved during the year iteration value; and the estimate of Q (st; at; J) is updated by mixing the previous Q value with a an adjustment of the new assessment to a degree of learning that slowly decays over time. It shows [31], the training approach of miniaturization Q approaches the true values Q and V and, therefore, the optimal policy, while each action tries in each state infinitely many times. Next, summarize the Q minimum training for secondary users to obtain the optimal policy in Table I. Since no secondary user (or attacker) will transmit (or block) a licensed band when the main user is active when the state of the main users differs from one country to another, The respective places for players in these countries are also different. Therefore, the action space it depends on the state. At the beginning of each stage t, secondary users check if they have previously observed state: if not, they will add ST to the observation history for each state of blackboard and initialization of the variables used in the learning algorithm Q, V and policy ... (st; a). If it already exists

1. At state $\mathbf{s}^t$, $t = 0, 1, \cdots$
◇ if state $\mathbf{s}^t$ has not been observed previously, add $\mathbf{s}^t$ to $\mathbf{s}_{hist}$,
• generate action set $\mathcal{A}(\mathbf{s}^t)$, and $\mathcal{A}_J(\mathbf{s}^t)$ of the attackers;
• initialize $Q(\mathbf{s}^t, \mathbf{a}, \mathbf{a}_J) \leftarrow 1$, for all $\mathbf{a} \in \mathcal{A}(\mathbf{s}^t)$, $\mathbf{a}_J \in \mathcal{A}_J(\mathbf{s}^t)$;
• initialize $V(\mathbf{s}^t) \leftarrow 1$;
• initialize $\pi(\mathbf{s}^t, \mathbf{a}) \leftarrow 1/\lvert\mathcal{A}(\mathbf{s}^t)\rvert$, for all $\mathbf{a} \in \mathcal{A}(\mathbf{s}^t)$;
◇ otherwise, use previously generated $\mathcal{A}(\mathbf{s}^t)$, $\mathcal{A}_J(\mathbf{s}^t)$, $Q(\mathbf{s}^t, \mathbf{a}, \mathbf{a}_J)$, $V(\mathbf{s}^t)$, and $\pi(\mathbf{s}^t)$;
2. Choose an action $\mathbf{a}^t$ at time t:
◇ with probability p_{exp}, return an action uniformly at random;
◇ otherwise, return action $\mathbf{a}^t$ with probability $\pi(\mathbf{s}^t, \mathbf{a})$ under current state $\mathbf{s}^t$.
3. Learn:
Assume the attackers take action $\mathbf{a}_J^t$, after receiving reward $r(\mathbf{s}^t, \mathbf{a}^t, \mathbf{a}_J^t)$ for moving from state $\mathbf{s}^t$ to $\mathbf{s}^{t+1}$ by taking action $\mathbf{a}^t$
◇ Update Q-function $Q(\mathbf{s}^t, \mathbf{a}^t, \mathbf{a}_J^t)$ according to (17);
◇ Update the optimal strategy $\pi^*(\mathbf{s}^t, \mathbf{a})$ by $\pi^*(\mathbf{s}^t) \leftarrow \arg\max_{\pi(\mathbf{s}^t)} \min_{\pi_J(\mathbf{s}^t)} \sum_{\mathbf{a}} \pi(\mathbf{s}^t, \mathbf{a}) Q(\mathbf{s}^t, \mathbf{a}, \mathbf{a}_J)$;
◇ Update $V(\mathbf{s}^t) \leftarrow \min_{\pi_J(\mathbf{s}^t)} \sum_{\mathbf{a}} \pi^*(\mathbf{s}^t, \mathbf{a}) Q(\mathbf{s}^t, \mathbf{a}, \mathbf{a}_J)$;
◇ Update $\alpha^{t+1} \leftarrow \alpha^t * \mu$;
◇ Go to step 1 until converge.

TABLE I: Learning Minimax-Q for Stochastic Anti-Blocking

in history, secondary users simply call the corresponding sets of actions and function values. Secondary users will choose to continue acting: with a certain probability pexp, they choose explore the entire action space A (st) and return the same action. With a probability of 1 pexp, they decide to take measures in the sense that they are written according to the current ... (st). Once the attackers assume action of J, the secondary users receive the prize and the game goes to the next state st + the secondary users update the values of the function Q and V, update the policy ... (st) of the state and spoil training speed The value of the iteration will continue until ... (ST) approaches the optimal policy, and we will show the approximation of the minimax-Q training in the results of the simulation.

Note that to obtain the value of the state V (st), secondary users have to decide a balance of matrix play where the payment is Q (st, a, aJ) for all 2 A (st) and aJ 2 AJ (st).Suppose that the attackers represent the order of a player whose strategy is denoted by a vector ... J (st) and secondary users form the column whose strategy is denoted by vector ... (st). Then the value the game can be expressed through

$$\max_{\pi(\mathbf{s}^t)} \min_{\pi_J(\mathbf{s}^t)} \pi_J(\mathbf{s}^t)^T Q(\mathbf{s}^t, \mathbf{a}, \mathbf{a}_J) \pi(\mathbf{s}^t), \tag{18}$$

that can not be solved directly. Assuming that the secondary user strategy ... (st) is fixed, then the problem in (18) becomes

$$\min_{\pi_J(\mathbf{s}^t)} \pi_J(\mathbf{s}^t)^T Q(\mathbf{s}^t, \mathbf{a}, \mathbf{a}_J) \pi(\mathbf{s}^t). \tag{19}$$

Since Q (st; a; j) ... (st) is just a vector and ... J (st) is a probability distribution, is equivalent to min [Q (st; aJ) ... (st)] i, i. find the minimum element of Q (st; aJ) ... (st). So, the problem in (18) is simplified

$$\max_{\pi(\mathbf{s}^t)} \min_i \left[Q(\mathbf{s}^t, \mathbf{a}, \mathbf{a}_J) \pi(\mathbf{s}^t) \right]_i. \tag{20}$$

Determine z = mini [Q (st; aJ) ... (st)] i, we have [Q (st; a; st)] i = z. Therefore, the initial problem (18) occurs

$$\begin{aligned} \max_{\pi(\mathbf{S}^t)} \quad & z \\ \text{s.t.} \quad & [Q(\mathbf{s}^t, \mathbf{a}, \mathbf{a}_J) \pi(\mathbf{s}^t)]_i \geqslant z, \\ & \pi(\mathbf{s}^t) \geqslant \mathbf{0}, \\ & \mathbf{1}^T \pi(\mathbf{s}^t) = 1, \end{aligned} \tag{21}$$

where ... (st)> 0 means that every probability element in (st) must be non-negative. Through treatment the target z as well as the variable (21) can refer to the following

$$\begin{aligned} \max_{\pi'} \quad & \mathbf{0}_{aug}^T \pi' \\ \text{s.t.} \quad & Q' \pi' \leqslant \mathbf{0}, \\ & \pi(\mathbf{s}^t) \geqslant \mathbf{0}, \\ & \mathbf{1}_{aug}^T \pi' = 1, \end{aligned} \tag{22}$$

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 5, May 2018

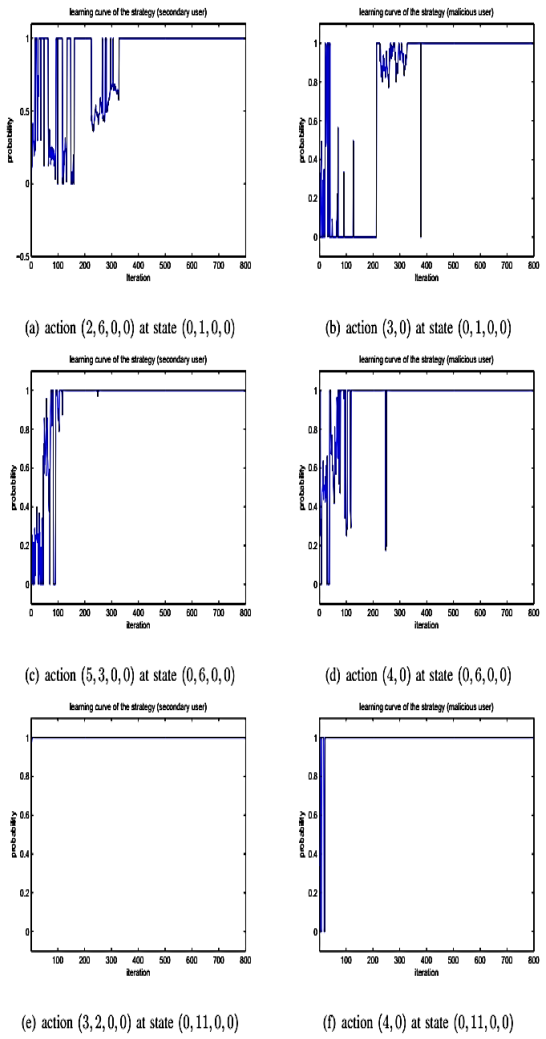
where 0 = [... (zst)], Q0 = ([O 1] ¡[Q (st; aJ) 0]), 1T aug = [1T0] and 0T aug = [0T1]. Issue (22) is a linear program so that secondary users can easily get the value of the game z from optimizer .. 0.

V. RESULTS OF SIMULATION

In this section, we perform simulations to evaluate the performance of the secondary user network under attack First we demonstrate the convergence of the minimax-Q training algorithm and analyze the strategy of secondary users and attackers for several typical states. Then, 17 We compare the achievable performance when secondary users adopt different strategies. About For illustrative purposes, we focus on examples with only one or two licensed bands to provide more insight; however, such policies can be monitored when there are more licensed bands.

***A. Convergence and Strategic Analysis*:**

1) *Protection against jamming in a licensed band:* First we investigate the case when there is only one a licensed bandwidth available for secondary users, that is. L = 1. The license has eight channels group, among which attackers can choose at most four channels to block at any time. Benefit of the use of each channel in the group of licenses glt can take any value of f1; 6; 11g and the transition the probability of gain of any qj a qi is pj g! l 1 = pj g! l 2 = 0: 4, pj g! 13 = 0: 2, for j = 1; 2; 3 too and for j = 0 when the primary user becomes inactive. Transition probabilities for the access of the primary user is given by p1 l! 1 = 0: 5 and p0 1! 1 = 0: 5. The duration of the time interval is 2 ms. First, we investigate the strategy of secondary users and attackers in these countries when the main the user is inactive and there are no channels that were detected successfully in the previous stage. I remember that the stochastic furtive game state with L = 1 is st = fP1t; g1t; J1t C; J1t; Dg,where Jt 1; C and J1t; D represents the number of control channels and blocked data observed by (0; 1; 0; 0); (0; 6; 0; 0); and (0; 11; 0; 0). We show the learning curve of the secondary user strategy in these countries in the left column of Figures 2 and 3



(a) action $(2,6,0,0)$ at state $(0,1,0,0)$ (b) action $(3,0)$ at state $(0,1,0,0)$

(c) action $(5,3,0,0)$ at state $(0,6,0,0)$ (d) action $(4,0)$ at state $(0,6,0,0)$

(e) action $(3,2,0,0)$ at state $(0,11,0,0)$ (f) action $(4,0)$ at state $(0,11,0,0)$

Fig. 2: Learning curve of the secondary users (left column) and the attackers (right column).

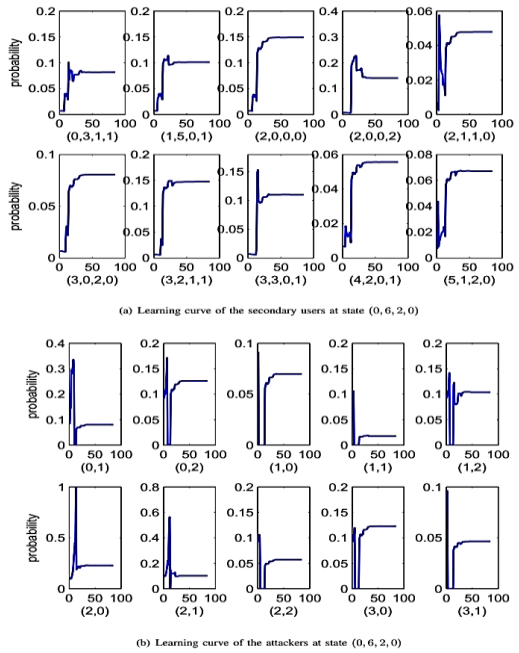
The training curve of the attackers' strategy in the right column. From Figure 2 we see that we use mini-Q-learning, secondary user strategies and attackers converge less than 400 time intervals (0.8 seconds), and the optimal strategy for each player is a clean strategy. Stage (1, C1, 1, D1, 1, C2, 1, D2), and the second user of the 1st band is denoted by the action of the aggressors is in (1, J1, 1, J2). Then, in Figures 2 (a) and 2 (b) for the state (0; 1; 0; 0), we see that the optimal strategy of the secondary users is finally approaching (2; 6; 0; 0), which means that secondary users

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 5, May 2018

select both 2 channels and control channels and 6 channels as data channels; and the optimal attack strategy approaches (3; 0); jam This is because the gain of each channel in this state is only 1 and the secondary users choose to reserve multiple channels of transmission of data messages and multiple channels of control messages, in the hope of obtaining more profits, but with a greater risk of blocking all control channels.

When the gain increases to 6 per channel as shown in Figures 2 (c) and 2 (d), the secondary users be more cautious when booking 5 control channels and 3 data channels, as well as attackers Be aggressive by attacking as many channels as you can. That is because the benefit of each channel is higher, and secondary users want to ensure a certain profit through insurance when renting a blocking control channel. When the gain is further increased to 11 (Fig 2 (e) and 2 (e)), secondary users become even more conservative, having only 2 data channels
and 3 control channels. This is because the destination of the secondary users is defined as Effective spectrum gain as in (12), and leaving more channels as empty can increase Finale Then we see how the strategy of the players will change when there are some variables of the country for example, some control or information channels are detected by the attackers in the previous stage. We choose only two states to illustrate the state (0; 6; 2; 0) and the state (0; 6; 0; 2)with strategy in state (0; 6; 0; 0).



(a) Learning curve of the secondary users at state $(0, 6, 2, 0)$

(b) Learning curve of the attackers at state $(0, 6, 2, 0)$

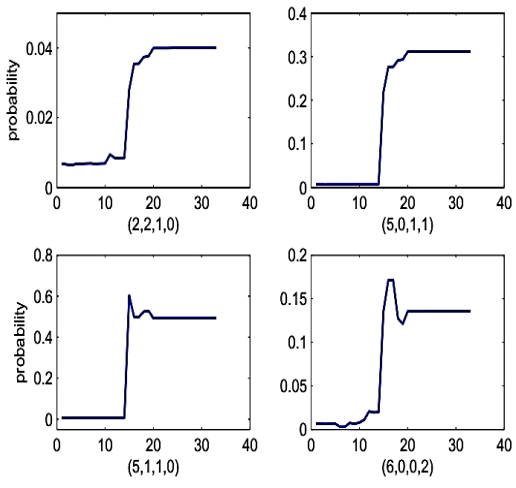
Fig. 3: Learning curve of the secondary users and the attackers at state $(0, 6, 2, 0)$.

Figure 3 shows the training curve for secondary users and attackers in the state(0; 6; 2; 0) where 2 control channels are blocked in the previous step. We see, Approximate strategies within 50 time intervals (0.1 seconds) and the optimal rules of both players in that state They are mixed strategies. As in the previous stage, the attackers successfully blocked 2 control channels; most other unwanted channels probably have data channels. That way attackers tend to detect previously non-sedentary channels with a relatively high probability because shown by actions (1; 0); (twenty); (3; 0); (2; 1) in Figure 3 (b), whose general probability is very high in the beginning Subsequently, secondary users tend to reserve most of the pre-locked channels as data channels, as shown in those actions where: D2> 1 is more likely greater than 0: 9; and reserve only some of the previous undetected channels, such as data channels, as shown by the actions where up to D1 6 3 with a total probability greater than 0: 8. Also, from Attackers will attack less than 3 channels of previous, undetected, secondary channels users retain only a maximum of 3 control channels to ensure reliable communication. The attackers it usually obstructs less than 4 channels. If you choose to mute 4 channels, secondary users face them the high probabilities of an attack will leave more empty channels. This can increase in return reward expected for secondary users and, therefore, attackers of up to 3 channels. The strategies of both players in state (0; 6; 0; 2) are shown in Figure 4.

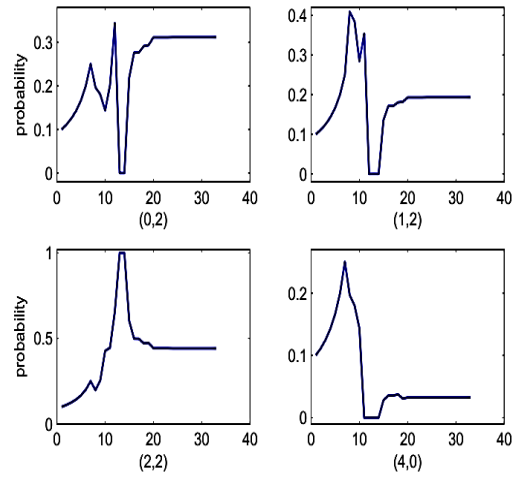
Since there are 2 data channels successfully caught in the previous stage, secondary users tend to book less than 1 channel which were previously blocked as data channels to avoid the "second detected", as can be seen in the actions (5; 0; 1; 1) and (5; 1; 1; 0) with a general probability of more than 0: 7. Since the attackers will probably attack previously unsecured channels, secondary users reserve most of the unwanted channels as control channels to provide reliability, again, as shown by the actions (5; 0; 1; 1)and (5; 1; 1; 0) where 5 undetected channels are selected as control channels. In response to Secondary users strategy, attackers will continue to attack previously detected channels as (0; 2); (1; 2); (2; 2) with a total probability of more than 0:94; where J2 = 2. Compare Figure 4 and Figure 3,

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 5, May 2018



(a) Learning curve of the secondary users at state $(0,6,0,2)$

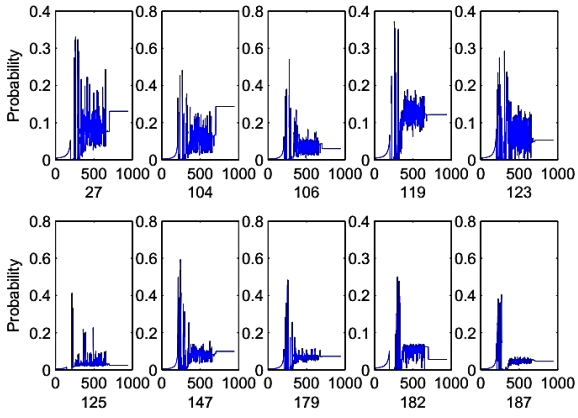


(b) Learning curve of the attackers at state $(0,6,0,2)$

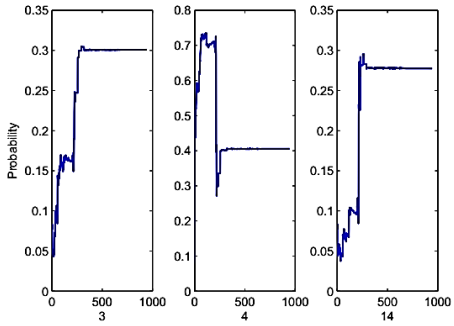
Fig. 4: Learning curve of the secondary users and the attackers at state $(0,6,0,2)$.

we find that when attackers successfully stutter some data channels, more information about the secondary user strategy (when finding data channels) revealed that the damage of the interference attack would be heavier and that the secondary users reserve more channels for usage control, resulting in a reduced payment.

2) *Anti-blocking protection in two licensed bands*: we are now discussing the average strategy users and attackers when there are two licensed bands available, that is. L = 2. There are four channels in each bandwidth and channel gain in each band it still has a value of f1; 6; 11g, p the same transition probability as in the case of a tape. Transition probability for The user's main access to the first bar is p1 1! 1 = p0 1! 1 = 0: 5, while the transition probability for the second band is p12! 1 = p0 2! 1 = 0: 2, which means that the probability that the second band is is higher than the first group. Attackers can detect up to four channels each weather To compare the case with a single bar, we first study the strategy of both players in the state ((0; 6; 0; 0); (0; 6; 0; 0)), where both bands are available, with gain g1t = g2t = 6, and without control or data channels They are blocked in the previous stage. We show the training curve of the two players in Figure 5,



(a) Learning curve of the secondary users



(b) Learning curve of the attackers

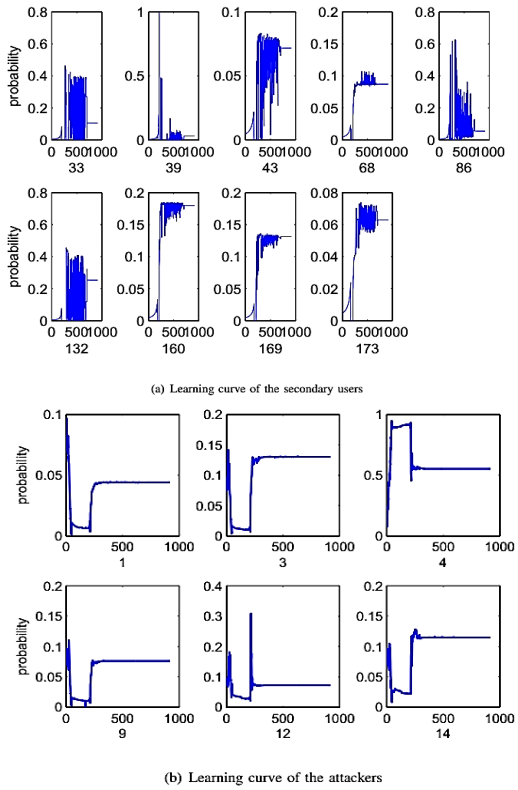
Fig. 5: Learning curve at state $((0,6,0,0),(0,6,0,0))$ when $L=2$.

where the number below each section shows the action index shown in this section. We see this the secondary user's strategy approaches the optimal policy within 800 time intervals (1.6 s), while The strategy of the attackers converges within 400 time intervals (0.8 seconds). Under the optimal policy, the secondary (0, 2, 0, 0), (1, 3, 0, 0)) indices such as 104, action ((2,) indexed as 27, action ((2, 0,

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 5, May 2018

0, 0), (1, 3, 0, 0) indexed as 119 and action ((2, 2, 0, 0), 0)) indexed as 147; the attackers act mainly ((0, 0), (3, 0)) by indexing 3, action ((0, 0), (4, 0)) indexed as 4, and action ((4; 0); (0; 0)) indexed as 14. Since the presence of the second band is higher, the attackers tend to detect the channels in the second band (with a general probability of 0: 7 of Action 3 and 4). But there is still a possibility that they will attack the first group by pointing out that The strategy of the attackers is accidental. Compared with the equivalent state (0; 6; 0; 0) in the case of a tape, where the policy of secondary users is (5; 3; 0; 0), the policy of secondary users in the case of two lanesThen we study the strategy in state (0, 1; 0; 0), (0; 6; 0; 0)) where g1t = 1 and g2t = the curves are shown in Figure 6.



(a) Learning curve of the secondary users

(b) Learning curve of the attackers

Fig. 6: Learning curve at state $((0,1,0,0),(0,6,0,0))$ when $L=2$.

Since the second band has a higher profit and is more will probably be available in the next slot, the attackers tend to muffle the second tape, as can be seen from probability of action ((0; 0); (3; 0)) indexed as 3; and action ((0; 0); to attackers' strategy, secondary users tend to reserve more control channels in the first since it is less likely to attack more data channels in the second group as it does higher gain for each channel as shown by the probability of action (2; 1; 0; 0); (1; 1; 0; 0) such as 132 and action ((0; 0; 0)); (0; 4; 0; 0)) indexed as 160;

A. Comparison of different strategies:

We also compare the efficiency of secondary users when they adopt the optimal policy obtained from the Q-minimum training with other policies to evaluate the proposed stochastic dazzle game and the learning algorithm. We assume that the attackers use their optimal stationary device a policy that is trained against secondary users who accept the minimum-Q training. Than we consider the following three scenarios with different strategies for secondary users. Secondary users accept the stationary policy derived from the Minimax-Q Training (denoted from "optimal"). Secondary users adopt a stationary policy derived from myopic learning. With myopic, we mean they are more interested in immediate remuneration than in future salaries. In the matter myopic policy, we assume that secondary users ignore the effect of their current action future payouts, so is the extreme case when $\gamma = 0$ (denoted as "myopic"). Secondary users adopt a fixed strategy that makes action equal to the action space A (st) for each ST (labeled "fixed").In Figure 7 we compare the accumulated average salary

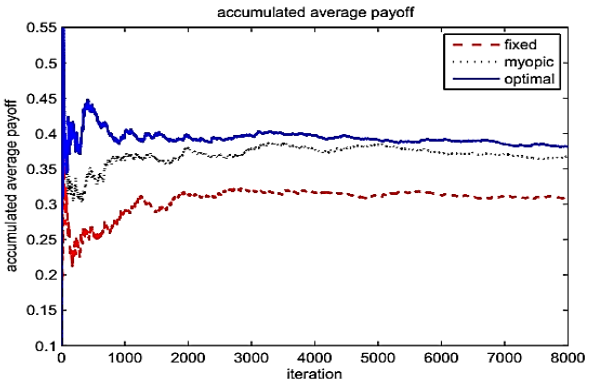


Fig. 7: Average payoff of different strategies.

or each iteration t0 calculated by R "

$$\bar{r}(t') = \frac{1}{t'} \sum_{t=1}^{t'} r(\mathbf{s}(t), \mathbf{a}(t), \mathbf{a}_J(t)). \qquad (23)$$

We see that, since the "optimal" strategy and the "myopic" strategy maximize the worst case while the "fixed" strategy only chooses a single action, strategy, the first two strategies have higher average salaries than the "fixed" strategy. Further, as shown in Figure 8, since the "optimal" strategy also takes into account the future amortization in optimization the strategy of the current stage, the "optimal

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 5, May 2018

strategy" achieves the highest amount of salaries at a discount(15% more than the myopic strategy and 42% more than the "fixed" strategy).Therefore, when secondary users are faced with a group of smart attackers who can adapt their strategy to the dynamics of the environment and the strategy of the enemy, by adopting training in the "mini-card" Q in The stochastic modeling of the anti-glare game achieves the best performance.

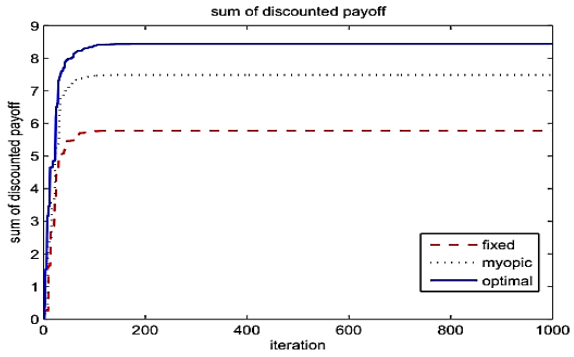


Fig. 8: Sum of discounted payoff of different strategies.

VI. CONCLUSION

In this article, we investigate the design of the anti-block mechanism in the cognitive radio network Looking at half the spectrum, time is different and cognitive attackers can use adaptive strategies, model interactions between secondary users and attackers like a stochastic game with a zero amount. The secondary users adapt their strategy to the way of booking and switch between the control and data channels according to their spectrum observations availability, channel quality and attacker actions. The results of the simulation show that the optimum the policy derived from training the mini-Q in the stochastic game can be much better efficiency in terms of spectrum efficiency compared to the policy of learning about myopia that only increases the maximum reward of each stage without taking into account the dynamics of the environment and cognitive abilities of the attackers and an accidental protection policy. The proposed stochastic game the framework can be summarized to model different protection mechanisms in other layers of the cognitive radio network, because you can also model the various environmental dynamics cognitive aggressors.

REFERENCES

[1] J. Mitola, "Cognitive radio: an integrated agent architecture for software defined radio," Ph.D. dissertation, KTH Royal Institute of Technology, Stockholm, Sweden, 2000.

[2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[3] M. M. Halldorson, J. Y. Halpern, L. Li, and V. S Mirrokni, "On spectrum sharing games," *Proc. ACM on Principles of distributed computing*, pp. 107–114, 2004. 22

[4] O. Ileri, D. Samardzija, and N. B. Mandayam, "Demand responsive pricing and competitive spectrum allocation via a spectrum server," *IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN'05)*, pp. 194–202, Baltimore, Nov.2005.

[5] J. Huang, R. Berry, and M. L. Honig, "Auction-based spectrum sharing," *ACM/Springer Mobile Networks and Apps.*, vol. 11,no. 3, pp. 405–418, June 2006.

[6] C. Kloeck, H. Jaekel, and F. K. Jondral, "Dynamic and local combined pricing, allocation and billing system with coginitive radios," *IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN'05)*, pp. 73–81, Baltimore, Nov.2005.

[7] S. Gandhi, C. Buragohain, L. Cao, H. Zheng, and S. Suri, "A general framework for wireless spectrum auctions," *IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks (DySPAN'07)*, pp. 22–33, Dublin, Apr. 2007.

[8] Z. Ji and K. J. R. Liu, "Belief-assisted pricing for dynamic spectrum allocation in wireless networks with selfish users," in *Proc. IEEE Int'l Conference on Sensor, Mesh, and Ad Hoc Communications and Networks (SECON)*, pp. 119–127, Reston,Sep. 2006.

[9] Z. Ji and K. J. R. Liu, "Multi-stage pricing game for collusion-resistant dynamic spectrum allocation," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 182–191, Jan. 2008.

[10] Y. Wu, B. Wang, K. J. R. Liu, and T. C. Clancy, "A Scalable Collusion-Resistant Multi-Winner Cognitive Spectrum Auction Game," to appear, *IEEE Trans. on*

Communications.

[11] Y. Xing, R. Chandramouli, S. Mangold, and S. Shankar, "Dynamic spectrum access in open spectrum wireless networks,"*IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 626-637, Mar. 2006.

[12] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks:A POMDP framework," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 3, pp. 589-600, Apr. 2007.

[13] B. Wang, Z. Ji, K. J. R. Liu, and C. Clancy, "Primary-prioritized Markov approach for efficient and fair dynamic spectrum allocation," *IEEE Trans. on Wireless Communications*, vol. 8, no. 4, pp. 1854–1865, Apr. 2009.

[14] R. Chen, J. Park, and J. H. Reed, "Defense against primary user emulation attacks in cognitive radio networks," *IEEE Journalon Selected Areas in Communications*, vol. 26, no. 1, pp. 25–37, Jan. 2008.

[15] R. Chen, J. Park, and K. Bian, "Robust Distributed Spectrum Sensing in Cognitive Radio Networks," *IEEE 27th Conferenceon Computer Communications (INFOCOM)*, pp. 31-35, Phoenix, AZ, Apr. 2008.

[16] T. X. Brown and A. Sethi, "Potential cognitive radio denial-of-service vulnerabilities and protection countermeasures: a multidimensional analysis and assessment," *Mobile Networks and Applications*, vol. 13, no. 5, pp. 516–532, Oct. 2008.

[17] T. C. Clancy and N. Goergen, "Security in cognitive radio networks: threats and mitigation," *3rd International Conference onCognitive Radio Oriented Wireless Networks and Communications (CrownCom)*, pp. 1-8, Singapore, May 2008.

[18] A. Wood and J. Stankovic, "Denial of service in sensor networks," *IEEE Computer*, 35(10):54-62, Oct. 2002.
[19] G. Noubir, "On connectivity in ad hoc network under jamming using directional antennas and mobility," *International Conference on Wired /Wireless Internet Communications*, pp. 186–200, 2004.

[20] R. L. Pickholtz, D. L. Schilling, and L. B. Milstein, "Theory of spread spectrum communications-a tutorial," *IEEE Trans.Commun.*, vol. 20, no. 5, pp. 855-884, May 1982.

[21] Q. Zhang and S. A. Kassam, "Finite-state Markov model for Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 47, no.11, pp. 1688-1692, Nov. 1999.

[22] V. Navda, A. Bohra, S. Ganguly, and D. Rubenstein, "Using channel hopping to increase 802.11 resilience to jamming attacks,"*IEEE 26th Conference on Computer Communications (INFOCOM)*, pp. 2526-2530, Anchorage, AK, Apr. 2007.23

[23] R. Gummadi, D. Wetherall, B. Greenstein, and S. Seshan, "Understanding and mitigating the impact of RF interference on 802.11 networks," *ACM SIGCOMM Computer Communication Review*, vol. 37, no. 4, pp. 385-396, 2007.

[24] A. D. Wood, J. A. Stankovic, and G. Zhou, "DEEJAM: defeating energy-efficient jamming in IEEE 802.15.4-based wireless networks," *Proc. 4th IEEE Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, pp. 60-69, San Diego, CA, June 2007.

[25] S. Khattab, D. Mosse, and R. Melhem, "Modeling of the channel-hopping anti-jamming defense in multi-radio wireless networks," *Proc. 5th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services (MobiQuitous)*, pp. 1-10, Dublin, Ireland, July 2008.

[26] W. Xu, T. Wood, W. Trappe, and Y. Zhang, "Channel surfing and spatial retreats: defenses against wireless denial of service,"in *Proc. 3rd ACM Workshop on Wireless Security (WiSe)*, pp. 80-89, Philadelphia, PA, Oct. 2004.

[27] S. Geirhofer, L. Tong, and B. M. Sadler, "Cognitive medium access: constraining interference based on experimental models,"*IEEE Journal on Selected Areas in Communications*, vol. 26, no. 1, pp. 95-105, Jan. 2008.

[28] L.S. Shapley, "Stochastic games," *Proc. Nat. Acad. Sci. USA*, vol. 39, no. 10, pp. 1095–1100, 1953.
[29] J. Filar and K. Vrieze, *Competitive Markov Decision Processes*, Springer-Verlag, 1997.

[30] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," *Proc. 11th International Conference on Machine Learning*, pp. 157–163, 1994.

[31] M. L. Littman and C. Szepesvari, "A generalized reinforcement-learning model: Convergence and applications," *Proc. 13th International Conference on Machine Learning*, pp. 310–318, 1996.

[32] J. Hu and M. P. Wellman, "Multiagent reinforcement learning: Theoretical framework and an algorithm," *Proc. 15thInternational Conference on Machine Learning*, pp. 242-250, 1998.

[33] A. Neyman and S. Sorin, *Stochastic Games and Applications*, Kluwer Academic Press, 2003.

[34] J. F. Mertens and A. Neyman, "Stochastic Games," *International Journal of Game Theory*, vol. 10, pp. 53-66, 1981.

[35] T. S. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall, 1996.

[36] A. Chan, X. Liu, G. Noubir, and B. Thapa, "Control channel jamming: resilience and identification of traitors," *Proc. ISIT*,2007.