Learning multi-channel Power Allocation against smart jammer in Cognitive Radio Networks

Feten Slimeni*, Bart Scheers†, Vincent Le Nir†, Zied Chtourou‡ and Rabah Attia§
‡VRIT Lab - Military Academy of Tunisia, Nabeul, Tunisia
†CISS Departement - Royal Military Academy (RMA), Brussels, Belgium
§SERCOM Lab - EPT University of Carthage, Marsa, Tunisia
Email: {feten.slimeni, ziedchtourou}@gmail.com, {bart.scheers, vincent.lenir}@rma.ac.be, rabah_attia@yahoo.fr

***Abstract*—We model the power allocation interaction between a cognitive radio and a jammer as a two-player zero-sum game. First, we determine the power allocation strategy for the cognitive radio using a modified version of the Q-learning algorithm against fixed jamming strategies. The learned anti-jamming strategy will be compared to the common waterfilling technique. Then, we consider the power allocation game using Q-learning for both the cognitive radio and the jammer. The learned strategies will be compared to the Nash equilibrium found under the assumption of perfect knowledge. Finally, we consider the real scenario of a jammer with imperfect information.**

Keywords– Cognitive radio, cognitive jammer, Q-learning, multi-channel power allocation, imperfect information

I. Introduction

The Cognitive Radio (CR) technology is a promising solution to the imbalance between scarcity and under-utilization of the spectrum [1]. It may exploit sensing, dynamic spectrum access (DSA) and learning capacities to improve both spectrum exploitation and resilience against possible interferer.

Radio jamming is a challenging attack in cognitive radio networks (CRNs) since (i) it may prevent CRs from detecting an available spectrum band during spectrum sensing by keeping the wireless spectrum busy, (ii) it may inject interference during an ongoing communication, so that the signal to interference plus noise ratio (SINR) deteriorates heavily and no data can be received correctly and (iii) it may corrupt control packets by attacking a common control channel to disrupt the totality of the network.

Smart jammers, which have cognitive features such as spectrum sensing, learning and reconfigurability, have recently attracted research attention in both wireless networks [2]–[5] and cognitive radio networks [6], [7]. Proposed anti-jamming protocols for CRNs include error-correcting codes [8], [9] to overcome the jamming impact on the transmitted signal and cooperation among CRs to deceive the jammers either through multi-tiers proxy [10] or honeypot node [11]. The authors in [12]–[14] use game theory to model the CRN jamming attack and apply reinforcement learning algorithms to learn how to avoid jammed channels. Other than learning anti-jamming channel selection, the authors in [15], [16] propose learning algorithms joining one channel selection and power control as anti-jamming strategy.

The interaction between a cognitive jammer and a CR in terms of multi-channel selection and power allocation is presented in [17] as Colonel Blotto game where the two opponents distribute limited resources over a number of battlefields with the payoff equal to SINR. The equilibrium is derived in terms of mixed (probabilistic) strategy via power randomization. Likewise, the authors in [18] adopt a Bayesian approach in studying the power allocation game between the CR and the jammer. They provide the Cumulative Distribution Functions (CDFs) of the transmission powers that should be adopted by the CR and the jammer at the NE to optimize the number of successful transmissions.

In this paper, we model the power allocation interaction between a CR user and a jammer as a zero-sum game. The action sets for both the CR and the jammer are defined by the vectors of power levels and the utility function is defined by the total transmission capacity of the CR. This paper can be seen as a generalization of [15] to multi-channel model; in which both the CR and the jammer are able to learn multi-channel power allocation. We propose, in section III, an online learning algorithm for the CR to select the optimal multi-channel power allocation based on the Q-learning algorithm. We evaluate the performance of the proposed algorithm in two scenarios; (1) against fixed jamming strategies, the learned anti-jamming power allocation strategy will be compared to the common work in this area which leads to explicit waterfilling solution [19], [20]; (2) against a smart jammer using also Q-learning algorithm, the learned jamming and anti-jamming power allocation strategies will be compared to the Nash equilibrium found under the assumption of perfect knowledge, as explained in section IV. In section V, we study the real scenario when the jammer has imperfect information about the CR user and the channel gain coefficients.

II. System model

We consider a zero-sum game between a CR and a jammer; the CR is trying to maximize its total transmission capacity avoiding jammers, the jammer aims to minimize the utility function of the CR by injecting interference. Both the CR and the jammer are assumed to allocate power over the available M channels, a zero power level in one channel means that this channel is not selected by the player. The actions of the CR and the jammer at time n are given by $\mathbf{p} = (p_1, \cdots, p_k, \cdots, p_M)$ and $\mathbf{j} = (j_1, \cdots, j_k, \cdots, j_M)$.The scenario is given in Fig. 1. The Shannon capacity is defined by

$$f(\mathbf{p}, \mathbf{j}) = \sum_{k=1}^{M} B_k log_2(1 + \frac{|h_k|^2 p_k}{|g_k|^2 j_k + n_k}), \quad (1)$$

where n_k and B_k are the noise variance and the bandwidth of channel k. h_k and g_k are the gain coefficients of channel k for the CR and the jammer respectively. We consider $f(\mathbf{p}, \mathbf{j})$ as the CR's utility function and $-f(\mathbf{p}, \mathbf{j})$ the jammer's one.

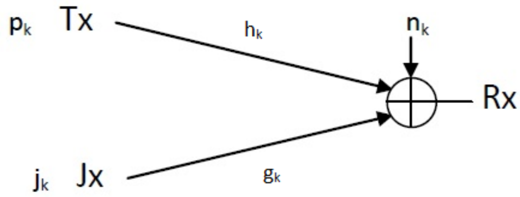

Fig. 1: Scenario of CR jamming attack

III. Q-LEARNING FOR MULTI-CHANNEL POWER ALLOCATION

Since the players can not get a priori information about either the state transition probabilities or the results (rewards) of their actions (decisions) until proceed, reinforcement learning algorithms may be used to solve this game through trial-and-error interactions.

In [21], we have modeled the scenario of fixed jamming strategy as a Markov decision process (MDP) model, and we have proposed a modified Q-learning algorithm (denoted as on-policy synchronous Q-learning: OPSQ-learning) to solve it in terms of non-jammed channel selection. The OPSQ-learning version speeds up the learning period and can be applied during CRN real time communication, it optimizes the Q-values of state/action pairs that the CR goes through until finding an anti-jamming channel selection strategy.

In this paper, we propose the application of this modified Q-learning algorithm to find an anti-jamming technique in terms of multi-channel selection and power allocation to maximize the total transmission capacity, against a jammer having the same cognitive features. We start by considering fixed jamming strategies. Then, we will consider the scenario of a jammer using Q-learning algorithm. The simulation results will be given in section VI.

A. CR using OPSQ-learning against fixed jamming strategies

We consider a fixed jamming strategy, which means that the jammer doesn't change its jamming policy during the game. We generalize our proposed OPSQ-learning algorithm to the multi-channel power allocation scenario, as given by algorithm 1. The CR is using OPSQ-learning to learn the optimal strategy which gives the optimal power allocation that it should choose in each state of the game. We consider that its transmit power p_k in each channel k can be selected from K levels.

We define the state of this game at timeslot n by the pair (Fjx, nb) with Fjx the jammed frequency detected by wideband spectrum sensing (WBSS) and nb the parameter indicating the number of successive occurrence of this jammed frequency. We have opt for mixing spatial and temporal properties in the state space definition to get a Markovian evolution of the environment.

Algorithm 1 Multi-channel anti-jamming power allocation using OPSQ-learning

Set γ and α values.
Initialize the Q matrix Q_0 to zero matrix.
Select a random initial action and observe the initial state s
for n=1, 2, 3, $\cdots$ **do**
  Select an action a verifying $max_x Q_{n-1}(s,x)$
  Observe the subsequent state s'
  Measure the fictive noise N_k in each channel by (2)
  Transmit using the power levels of the chosen action and measures the immediate reward as given by (4)
  Update all Q_n values of the previous state s by doing:
  for i $\in$ {the action set of the CR} **do**
    Observe the subsequent fictive state s_f of taking fictive action i
    Observe the fictive reward r_f as given by (4)
    Update $Q_n(s,i) = (1-\alpha)Q_{n-1}(s,i) + \alpha[r_f + \gamma\ max_x Q_{n-1}(s_f,x)]$
  end for
  $s = s'$
end for

In each timeslot, the CR chooses the action (the power allocation vector $\mathbf{p}$) which corresponds to the maximum Q value in the current state. This action is given by the column having the maximum Q value in the row corresponding to the current state. Considering this greedy strategy instead of doing random exploration, we call the proposed version ON-policy Q-learning since it learns the value of the policy being carried out by the agent, including the exploration steps.

The CR transmits with the power levels given by the chosen vector $\mathbf{p}$, observes the new state s', cooperates with the receiver node to measure the fictive noise N_k corresponding to the normalized interference and noise in each channel

$$N_k = \frac{n_k + |g_k|^2 j_k}{|h_k|^2}, \tag{2}$$

and calculates the reward r defined by (4)

$$f_k(\mathbf{p}, \mathbf{j}) = B_k log_2(1 + \frac{p_k}{N_k}). \tag{3}$$

$$r = \sum_k f_k(\mathbf{p}, \mathbf{j}) \tag{4}$$

Having the value of N_k, $\forall k$, the CR is able to do synchronous update all the Q values in the row corresponding to the current state s (defined by the previous actions of the CR and the jammer), instead of the asynchronous update of only one cell in the Q matrix associated to his action. This is the reason of denoting the proposed algorithm as synchronous Q-learning. This means that for each possible action $a = \mathbf{p}$ even the non taken ones, the CR updates the value $Q(s, a)$ as follows

$$Q_n(s,a) = (1-\alpha)Q_{n-1}(s,a) + \alpha[r + \gamma\ max_x Q_{n-1}(s',x)] \tag{5}$$

where γ is the discount factor that controls how much effect future rewards have on the optimal decisions. Small values of γ emphasizing near-term gain and larger values giving significant weight to later rewards. α is a learning rate that controls how quickly new estimates are blended into old estimates.

B. Both the CR and the jammer using Q-learning

We consider that the transmit powers p_k, j_k in each channel k can be selected from K levels. The CR is using the OPSQ-learning as given by algorithm 1 against a jammer who is using Q-learning as given by algorithm 2.

Since the jammer interacts with the CR and changes its jamming strategy, we define the state of this interactive game at each timeslot n by the pair (**p**, **j**) of actions taken by the CR and the jammer at the previous timeslot.

In each timeslot, the CR chooses the action (the power allocation vector **p**) which corresponds to the maximum Q value in the current state and the jammer chooses the action (the power allocation vector **j**) which corresponds to the minimum Q value in the current state. The two players transmit with the corresponding powers in each channel, observe the new state s' and calculate the immediate reward r. Having the value of N_k, the CR updates all the Q values in the row corresponding to the current state s.

In this work, we consider that the jammer can measure the SINR value resulting from its action by observing the acknowledgment packets exchanged between the transmitter-receiver pair [22]. Then, it is able to calculate the immediate reward and the Q value related to the taken actions p_k and j_k. But even having the capacity of WBSS, the jammer can not get the required information to update other Q values.

Algorithm 2 Multi-channel jamming power allocation using Q-learning

```
Set γ and α values.
Initialize the Q matrix Q_0 to zero matrix.
Select a random initial action and observe the initial state s
for n=1, 2, 3, ··· do
    Select an action a verifying min_x Q_{n-1}(s, x)
    Observe the immediate reward r
    Observe the subsequent state s'
    Update Q_n(s, a) = (1 − α)Q_{n-1}(s, a) + α[r + γ min_x Q_{n-1}(s', x)]
    s = s'
end for
```

IV. Nash game with perfect knowledge

We start by presenting the common waterfilling technique, before presenting the Nash game with perfect knowledge.

A. Waterfilling solution

The classic work in the problem of multi-channel power allocation leads to the common waterfilling solution. For example, each CR in [23] applies a waterfilling scheme for allocating power over the available channels and the greedy asynchronous distributed interference avoidance algorithm (GADIA) to solve the mutual interference problem. It is also exploited in [19], [20] to slove multi-channel power allocation in the presence of a jammer in wireless communication networks.

The optimal power allocation to maximize the transmission capacity is given by the expression of the waterfilling solution

$$p_k^* = (\frac{1}{\lambda} - N_k)^+ \tag{6}$$

where $\frac{1}{\lambda}$ is known as the water level, which can be found by bisection and should satisfy

$$\sum_k (\frac{1}{\lambda} - N_k)^+ = P \tag{7}$$

where N_k is the fictive noise power on each channel, as defined by (2) and $(x)^+ = \max(0, x)$.

B. Nash game power allocation

We consider here a sequential game in which both the CR and the jammer make decisions but sequentially and with the assumption of perfect knowledge for the jammer (the jammer knows all what it needs to make its decision, such as the channels gain coefficients and the CR's power allocation). In game theory, a game is said to be sequential if the players choose their actions in a consecutive way and the latter player requires information about the former.

The CR is able through WBSS and cooperation with the receiver to get the fictive noise power N_k in each channel, which is sufficient to implement the waterfilling solution and to adjust its power allocation in each iteration of the game.

The jammer will allocate its jamming power to minimize the total capacity. Mathematically, this is expressed as the following minimizing problem

$$\begin{aligned} \underset{\mathbf{j}}{\text{minimize}} \quad & f(\mathbf{p}, \mathbf{j}) \\ \text{subject to} \quad & \sum_{k=1}^{M} j_k \leq J \end{aligned} \tag{8}$$

We can write the Lagrangian as

$$L(\mathbf{j}, \mu) = -f(\mathbf{p}, \mathbf{j}) - \mu(\sum_{k=1}^{M} j_k - J) \tag{9}$$

Since L is separable in j_k, we can separately minimize each term,

$$\frac{\partial L}{\partial j_k} = \frac{|g_k|^2 |h_k|^2 p_k}{(|h_k|^2 p_k + |g_k|^2 j_k + n_k)(|g_k|^2 j_k + n_k)} - \mu \tag{10}$$

After solving the resulting second order equation in j_k, we get

$$j_k = \left(\frac{1}{2} \sqrt{(\frac{|h_k|^2 p_k}{|g_k|^2})^2 + 4\frac{|h_k|^2 p_k}{|g_k|^2 \mu}} - \frac{|h_k|^2 p_k}{2|g_k|^2} - \frac{n_k}{|g_k|^2} \right)^+ \tag{11}$$

where the KKT multiplier μ is the solution of

$$\sum_{k=1}^{M} j_k \leq J \tag{12}$$

and can be found by bisection. Concerning the other parameters in its power expression (11), the jammer is theorized to have all relevant information with which to make its decision: the channel gain coefficients, the CR's power allocation and the channel noise levels at each iteration. Note that the jammer's strategy, given by equation (11), is not a waterfilling strategy.

V. NASH GAME WITH IMPERFECT INFORMATION

In real scenario, the jammer doesn't have the required information either to apply the Q-learning algorithm (as described in section III) or to play the Nash game (as described in section IV). According to the expression (11), the jammer needs to estimate the CR's power allocation p_k and make assumptions about the parameters n_k, h_k and g_k.

A trivial solution for the jammer would be to make the assumption of flat fading channels, otherwise he has to estimate the different channel coefficients which seems complicated. Let $h_k = h$ and $g_k = g$, $\forall k$. He may consider $g = 1$, which corresponds to the scenario of being near the receiver node. Also, he may neglect the noise $\mathbf{n}$. Furthermore, the jammer may consider that what he detects through spectrum sensing is equal to transmission power $\mathbf{p}$ multiplied by the channel gain $\mathbf{h}$.

VI. SIMULATION AND DISCUSSION

A. OPSQ-learning against fixed jamming strategies

In this scenario, we consider fixed jamming strategies and we will compare the anti-jamming strategy of the CR applying the proposed OPSQ-learning to the waterfilling strategy.

Let's consider $M = 3$ channels. The action set of a sweeping jammer is defined by $A_{\mathbf{j}} = \{(J,0,0),(0,J,0),(0,0,J)\}$ with J as the total jamming power.

To apply the OPSQ-learning algorithm, we consider four power levels for the CR: $\{0, P, \frac{P}{2}, \frac{P}{3}\}$ with P as its maximum power, so the CR may use one/two or the three available channels and its action set is,

$$A_{\mathbf{p}} = \{(P,0,0),(0,P,0),(0,0,P),\\ (\frac{P}{2},\frac{P}{2},0),(\frac{P}{2},0,\frac{P}{2}),(0,\frac{P}{2},\frac{P}{2}),\\ (\frac{P}{3},\frac{P}{3},\frac{P}{3})$$

We consider $P = J = 10$ as the total CR's and jammer's power, $B = (1,1,1)$ as the channels' bandwidth, the discount factor $gamma = 0.95$ and the learning rate $\alpha = 0.1$.

1) Comparison between OPSQ-learning and waterfilling in flat fading channels scenario:

As first case, we consider flat fading channels for both the CR and the jammer with equal channel gain coefficients $\mathbf{g} = \mathbf{h} = (1,1,1)$ and we consider also the same noise level in all the channels $\mathbf{n} = (1,1,1)$. Fig. 3 gives the CR's actions resulting from the application of the learning algorithm against the sweeping jammer. The action indexes are varying from 1 to 7 as given in the action set $A_{\mathbf{p}}$ of the CR. We present in Fig. 2 the average payoff fluctuations during learning, the payoff function is the total transmission capacity of the CR. According to these figures, after some collisions (in timeslots 1 and 4) and some successful transmissions during about 12 timeslots, the CR learns to follow the optimal strategy as given by Table I. For each action of the sweeping jammer, we mention the index of the CR optimal action and the corresponding power allocation found at the convergence of the OPSQ-learning algorithm. As given in this table, the power allocation resulting from the waterfilling strategy equals the power allocation learned using the proposed algorithm.

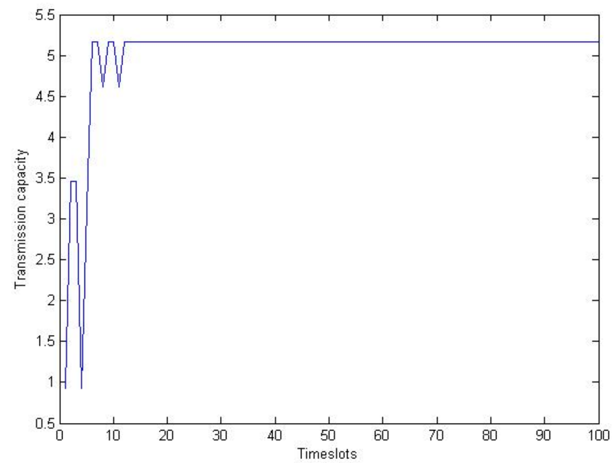

Fig. 2: The transmission capacity over flat fading channels in the presence of sweeping jammer

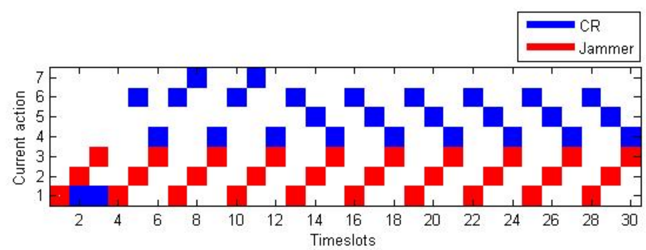

Fig. 3: The learned anti-jamming strategy against sweeping jammer over flat fading channels

Jx	index	1	2	3
	power	(10,0,0)	(0,10,0)	(0,0,10)
OPSQ-learning	index	6	5	4
	power	(0, 5, 5)	(5, 0, 5)	(5, 5, 0)
waterfilling		(0, 5, 5)	(5, 0, 5)	(5, 5, 0)
Capacity		5.1699	5.1699	5.1699

TABLE I: CR using OPSQ-learning/waterfilling against sweeping jammer over flat fading channels

2) Comparison between OPSQ-learning and waterfilling in selective channels scenario:

In this scenario, we consider selective channels for both the CR and the jammer with the channel gain coefficients $\mathbf{g} = \mathbf{h} = (2,1,3)$ and we consider the noise vector $\mathbf{n} = (2,3,1)$. We have chosen these values to make channel 3 better than channel 1 which is better than channel 2, for both the CR and the jammer.

Fig. 4 gives the CR's actions applying the OPSQ-learning algorithm. After some collisions (in timeslots 1 and 4) and some successful transmissions during about 6 timeslots, the CR learns to follow the optimal strategy as given by Table II. Considering the same parameters, the waterfilling solution against each action of the jammer results in a power allocation close to the solution found by using the OPSQ-learning algorithm; the two solutions avoid almost the same channels but differs slightly in the allocated power levels and the payoff values. This difference is due to the number of possibilities (i.e. the power levels) which is infinite for the waterfilling strategy and finite for the proposed algorithm.

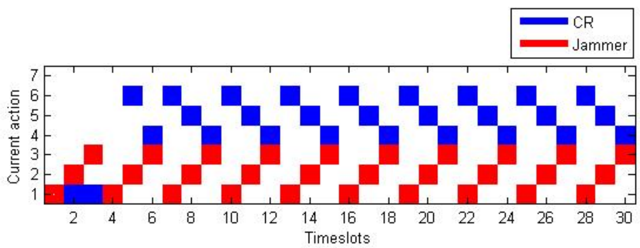

Fig. 4: The learned anti-jamming strategy against sweeping jammer over selective channels

Jx	index	1	2	3
	power	(10,0,0)	(0,10,0)	(0,0,10)
OPSQ-learning	index	6	5	4
	power	(0, 5, 5)	(5, 0, 5)	(5, 5, 0)
Capacity (OPSQ)		6.9386	8.983	4.8745
waterfilling		(0, 3.5556, 6.4444)	(4.8056, 0, 5.1944)	(6.25, 3.75, 0)
Capacity (waterfilling)		7.0104	8.9849	4.9248

TABLE II: CR using OPSQ-learning/waterfilling against sweeping jammer over selective channels

3) OPSQ-learning against other fixed jamming strategies:

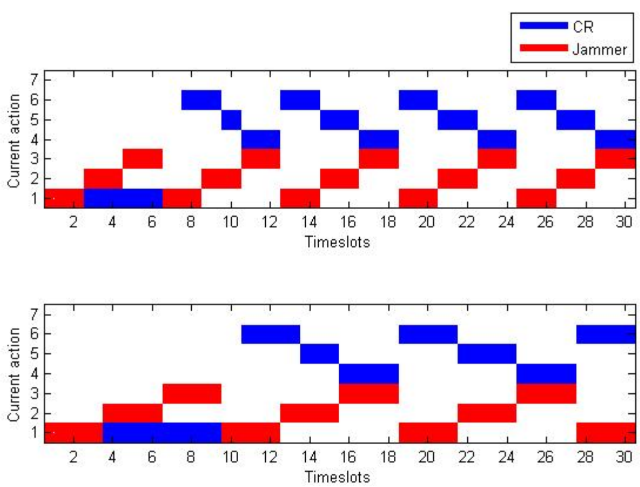

Fig. 5: The learned anti-jamming strategies against sweeping jammer attacking the same channel for 2 TSs and 3 TSs

Since we have considered the time in the definition of the state set, the CR succeeds to learn anti-jamming power allocations against a sweeping jammer staying in the same channel for two and three timeslots, as given by Figure 5.

B. OPSQ-learning against a jammer using Q-learning

In this scenario, we consider the same previous simulation parameters. Here, the CR applies the OPSQ-learning as given by algorithm 1 and the jammer uses the Q-learning algorithm 2. We consider that the jammer has the same action set as the CR: $A_{\mathbf{j}} = A_{\mathbf{p}}$. We will compare the strategies learned by the CR and the jammer to the optimal strategies found at the convergence of the Nash game under perfect knowledge.

Fig. 6 gives the payoff fluctuations during learning. After about 13000 timeslots, the CR's and the jammer's actions (power allocations) are no longer fluctuating and the transmission capacity reaches the fixed value $C = 2.4859$. The jammer's final power allocation is $\mathbf{j} = (5, 0, 5)$ and the CR's final power allocation is $\mathbf{p} = (\frac{10}{3}, \frac{10}{3}, \frac{10}{3})$.

In the Nash game, the CR uses the waterfilling expression (6) and proceeds by bisection until reaching the value of λ corresponding to the allocation of the total power. The jammer proceeds by bisection and calculates the sum of the allocated powers to all the channels using the expression (11) until reaching the value of μ corresponding to the allocation of the total jamming power. At the Nash equilibrium (NE) of the described game, we get (after 37 iterations) the jammer's power allocation $\mathbf{j} = (4.0370, 1.5370, 4.4259)$ and the CR's power allocation $\mathbf{p} = (3.3333, 3.3333, 3.3333)$ with the transmission capacity $C = 2.384$, as given by Fig. 7. This result is close to the result found using OPSQ-learning for the CR and Q-learning for the jammer.

C. The CR with perfect knowledge against a jammer with imperfect knowledge

We consider the Nash game in two scenarios; (1) the jammer has perfect knowledge, (2) the jammer does the assumption detailed before. We present in Table III a comparison between the NEs of the two scenarios.

These results corresponds to the channel coefficients $|h| = (1.9821, 0.9848, 3.3178)$ and $|g| = (0.533, 0.0985, 1.1683)$. Hence, the jammer having perfect knowledge avoids bad channels (e.g. channel two since it has low gain coefficient) even if it used by the CR. This allows the jammer to attack the other channels with higher powers which reduces the total channel capacity of the CR. Without perfect knowledge, the jammer occupies all the channels with almost the same power level which results in a limited payoff gain for the CR (i.e. limited loss in the effectiveness of the jamming attack).

	NE under perfect knowledge	NE under imperfect knowledge
p	(3.3184, 3.5958, 3.0858)	(3.3316, 3.4675, 3.2009)
j	(4.9801, 0, 5.0198)	(3.4679, 3.2153, 3.3167)
Capacity	10.8422	11.4391

TABLE III: Knowledge effect on the NE

VII. Conclusion

In this paper, we considered a power allocation game between a CR user and a jammer. First, we proposed a modified version of the Q-learning algorithm (OPSQ-learning) which allows the CR to learn an anti-jamming power allocation

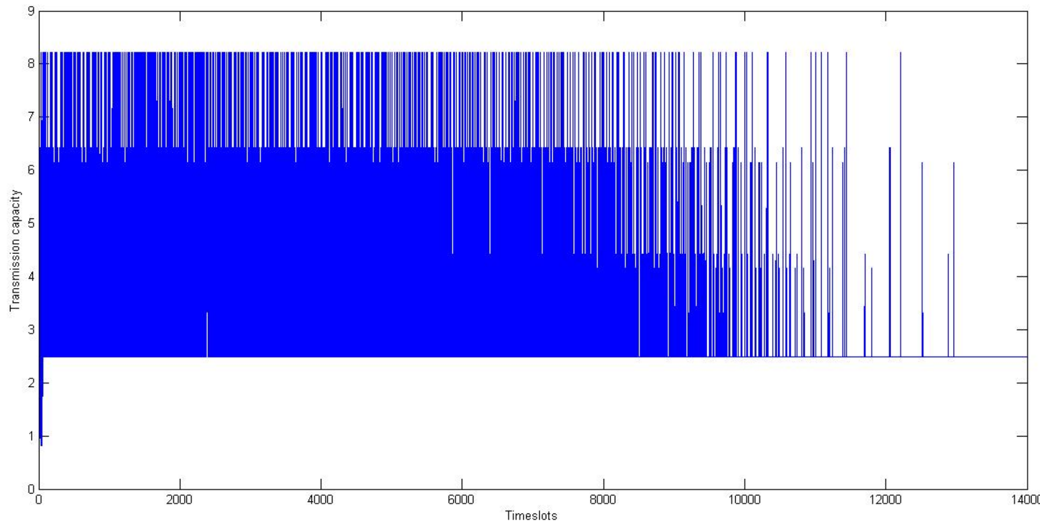

Fig. 6: The transmission capacity over selective channels against a jammer using Q-learning

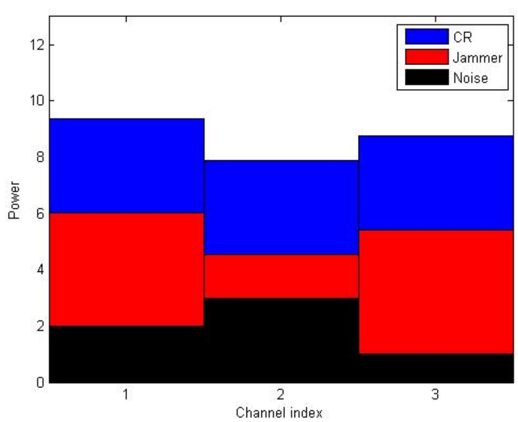

Fig. 7: The jamming and anti-jamming strategies at the NE

strategy. Then, we presented the game under both perfect and imperfect information. Against fixed jamming strategies, the learned solution equals the common explicit waterfilling solution over flat fading channels and it is slightly different over selective channels. Furthermore, we considered a smart jammer using the Q-learning algorithm. The learned jamming and anti-jamming power allocation strategies are almost equal to the optimal Nash equilibrium strategies found under the assumption of perfect knowledge. Finally, we studied the real scenario when the jammer has imperfect information about the CR user and the channel gain coefficients. Under this condition, the jammer occupies all the channels with almost the same power level which results in a limited payoff gain for the CR.

REFERENCES

[1] J. Mitola, "Cognitive radio for flexible mobile multimedia communications," in *Mobile Multimedia Communications, 1999. (MoMuC '99) 1999 IEEE International Workshop on*, 1999, pp. 3–10.

[2] M. K. Hanawal, M. Abdel-Rahman, and M. Krunz, "Game theoretic anti-jamming dynamic frequency hopping and rate adaptation in wireless systems," in *12th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, WiOpt 2014, Hammamet, Tunisia, May 12-16, 2014*, 2014, pp. 247–254.

[3] Y. Li, L. Xiao, J. Liu, and Y. Tang, "Power control stackelberg game in cooperative anti-jamming communications," in *Game Theory for Networks (GAMENETS), 2014 5th International Conference on*, Nov 2014, pp. 1–6.

[4] K. Xu, Q. Wang, and K. Ren, "Joint ufh and power control for effective wireless anti-jamming communication," in *INFOCOM, 2012 Proceedings IEEE*, March 2012, pp. 738–746.

[5] D. Yang, G. Xue, J. Zhang, A. W. Richa, and X. Fang, "Coping with a smart jammer in wireless networks: A stackelberg game approach." *IEEE Transactions on Wireless Communications*, vol. 12, no. 8, pp. 4038–4047, 2013.

[6] C. X. Wednel Cadeau, Xiaohua Li, "Markov model based jamming and anti-jamming performance analysis for cognitive radio networks." Communications and Network, 2014.

[7] W. Conley and A. Miller, "Cognitive jamming game for dynamically countering ad hoc cognitive radio networks," in *Military Communications Conference, MILCOM 2013 - 2013 IEEE*, Nov 2013, pp. 1176–1182.

[8] A. Asterjadhi and M. Zorzi, "JENNA: a jamming evasive network-coding neighbor-discovery algorithm for cognitive radio networks," *IEEE Wireless Communications*, vol. 17, no. 4, pp. 24–32, 2010.

[9] V. Balogun, "Anti-jamming performance of hybrid FEC code in the presence of CRN random jammers," *International Journal of Novel Research in Engineering and Applied Sciences (IJNREAS)*, vol. 1, no. 1, 2014.

[10] W. Wang, S. Bhattacharjee, M. Chatterjee, and K. Kwiat, "Collaborative jamming and collaborative defense in cognitive radio networks," *Pervasive and Mobile Computing*, vol. 9, no. 4, pp. 572–587, 2013.

[11] S. Bhunia, X. Su, S. Sengupta, and F. J. Vázquez-Abad, "Stochastic model for cognitive radio networks under jamming attacks and honeypot-based prevention," in *Distributed Computing and Networking - 15th International Conference (ICDCN '14), pages 438-452, Coimbatore, India, January 4-7, 2014. Proceedings*.

[12] Y. Wu, B. Wang, and K. J. Ray Liu, "Optimal defense against jamming attacks in cognitive radio networks using the markov decision process approach," in *GLOBECOM'10*, 2010, pp. 1–5.

[13] C. Chen, M. Song, C. Xin, and J. Backens, "A game-theoretical anti-

jamming scheme for cognitive radio networks," *IEEE Network*, vol. 27, no. 3, pp. 22–27, 2013.

[14] B. Wang, Y. Wu, K. J. R. Liu, and T. C. Clancy, "An anti-jamming stochastic game for cognitive radio networks," *IEEE Journal on Selected Areas in Communications*, 2011.

[15] T. Chen, J. Liu, L. Xiao, and L. Huang, "Anti-jamming transmissions with learning in heterogenous cognitive radio networks," in *Wireless Communications and Networking Conference Workshops (WCNCW), 2015 IEEE*, March 2015, pp. 293–298.

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

[17] Y. Wu, B. Wang, K. J. R. Liu, and T. C. Clancy, "Anti-jamming games in multi-channel cognitive radio networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 1, pp. 4–15, 2012.

[18] R. El-Bardan, S. Brahma, and P. Varshney, "Power control with jammer location uncertainty: A game theoretic perspective," in *Information Sciences and Systems (CISS), 2014 48th Annual Conference on*, March 2014, pp. 1–6.

[19] E. Altman, K. Avrachenkov, and A. Garnaev, "A jamming game in wireless networks with transmission cost." in *NET-COOP*, ser. Lecture Notes in Computer Science, vol. 4465. Springer, 2007, pp. 1–12.

[20] R. H. Gohary, Y. Huang, Z.-Q. Luo, and J.-S. Pang, "A generalized iterative water-filling algorithm for distributed power control in the presence of a jammer," in *ICASSP'09*, 2009, pp. 2373–2376.

[21] F. Slimeni, B. Scheers, Z. Chtourou, and V. Le Nir, "Jamming mitigation in cognitive radio networks using a modified q-learning algorithm," in *Military Communications and Information Systems (ICMCIS), 2015 International Conference on*, May 2015, pp. 1–7.

[22] Z. Serceki and M. Wilhoyte, "Method for determining packet error rate of wireless LAN stations," Patent US 20 040 076 138A1, 22 April, 2004.

[23] S. Baksi and D. C. Popescu, "Distributed power allocation for rate maximization in cognitive radio networks with horizontal spectrum sharing," in *2015 IEEE Wireless Communications and Networking Conference, WCNC 2015, New Orleans, LA, USA, March 9-12, 2015*, 2015, pp. 932–936.