Proceedings of CECNet 2022
A.J. Tallón-Ballesteros (Ed.)

doi:10.3233/FAIA220523

Anti-Jamming Method of Cognitive Radio Based on Q-Learning

Yichen XIAO[a,1], Haiyu REN[a], Shan WU[a], Lixiang LIU[a], Xiandong MENG[b] and Pengcheng DING[b]
[a] *Science & Technology on Integrated Information System Laboratory, Institute of Software Chinese Academy of Sciences Science, Beijing, China*
[b] *TINNO Communications Co., Ltd, YiBin, Sichuan, China*

Abstract. Due to the exposed nature of wireless links, the communication of wireless networks is vulnerable to jammers. And because the jammer models are usually unknown to communication users, particularly in military confrontation applications, how to ensure maintain communication under different jamming is an active research topic. In this paper, we take the anti-jamming task of cognitive radio as a Markov decision process and propose an anti-jamming method based on Q-learning. The method aim to learn an efficient policy for users to maximize the total channel transmission capacity in different typical jamming scenarios. The simulation results indicate that compared with the traditional anti-jamming methods, the anti-jamming method based on Q-learning can obtain better performance, and more effective against several kinds of typical jamming models.

Keywords. Cognitive Radio, Anti-jamming method, Q-learning algorithm

1. Introduction

The history of jamming attack can be traced back to the early 20th century. In the Second World War, jammers played an important role, interfering with the radio broadcasts of enemy countries and misleading pilots [1]. On the contrary, the development of anti-jamming technology has a long history.

In traditional anti-jamming methods, direct sequence spread spectrum (DSSS) and frequency hopping (FH) are widely used. A novel multiple parallel frequency-hopping (MPFH) communication system was proposed in [2], which prolongs processing time of jammers and finally nullifies follower jamming of interfering party. [3] proposed an anti-jamming method based on IIR and nonlinear compensation for anti-aliasing and out-of-band jamming in CFHR, which can reduce the implementation complexity when the BER performance is not declined. An index-modulation based joint mode-frequency-hopping (IM-MFH) scheme was proposed in [4], which activates several OAM-modes and carrier frequencies to hop simultaneously, thus resulting in low probability of legitimate signals jammed. In [5], a novel spread-spectrum communication theory which combined with the chaotic system and the anti-jamming applications were proposed.

In recent years, Cognitive Radio (CR) [6] has arisen as a potential solution to solve the spectrum shortage problem, where the sensing ability of cognitive users (CUs) helps

[1] Corresponding Author, Yichen Xiao, E-mail: yichen@iscas.ac.cn.

acquire knowledge of the environment. Thus, CUs could effectively use the 'spectrum holes' in spatial, time and frequency domain to reuse the idle spectrum, in order to avoid the effects of jammers. Game-theoretic modeling was a well-known technique used in CR to express the interaction between CUs and adversary jamming attackers since they have opposite objectives [7-8]. In [7], a game-theoretic learning anti-jamming (GTLAJ) paradigm is proposed, and its framework and challenges were introduced. [8] provided a machine-learning-based anti-jamming technique to avoid a hostile jammer, where both the jamming and anti-jamming processes are formulated based on the Markov game framework. However, these approaches of game theory need prior knowledge such as the jamming pattern, which is unpractical in actual usage scenario. Reinforcement learning can also be used to find optimal anti-jamming strategies for many fields of communication, such as Internet of things and UAV [9-10]. Cognitive radio anti-jamming based on the SARSA algorithm was proposed in [11], which took the spectrum energy efficiency of secondary users as a utility function.

Compared with conservative SARSA, Q-learning algorithm is more efficient. In this work, we consider Q-learning to determine optimal anti-jamming strategy. We build the CR anti-jamming model showing how Q-learning algorithm to choose optimal strategy to resist jamming, and via simulation to verify the anti-jamming performance of Q-learning algorithm, i.e. the improvement of total channel transmission capacity and SINR.

The remaining of the paper is organized as follows. In Section 2, we introduction the system model about cognitive radio network and types of jammers. In Section 3, we obtain Q-learning algorithm for anti-jamming. In Section 4 the simulation results are presented, followed by conclusion in Section 5.

2. System Model

Figure 1 shows a model of cognitive radio communication with jamming. There are a cognitive user (CU) and J jammers. The licensed frequency band is divided into K channels. In each time slot, both CU and jammer are allowed to access only one channel. The dotted line in Figure 1 represents the possible selected channel, and the solid line represents the actually selected channel. Assuming that the channel sensing part is known, the CU can observer the channel situation in real time .The jammer can only jam one channel in each time slot.

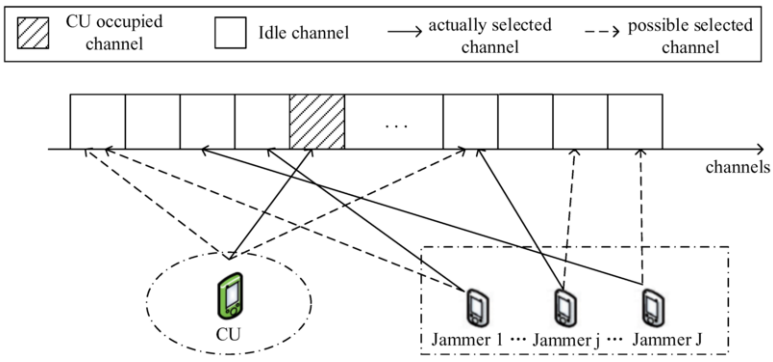

Figure 1. Cognitive Radio Communication with Jamming

The process of interaction between CU and jammers can be constructed into a reinforcement learning model. The jammer tries to block normal communication. If the channel accessed by CU is jammed, it will carry out anti-jamming strategy and select a new channel for data transmission. While avoiding jamming, CU tries to maximize its total transmission capacity. In the paper, the total transmission capacity of the channel is taken as the utility function of CU, which is denoted as

$$u^t = \sum_{k=1}^{K} B_k \log_2 \left(1 + \frac{|h_k|^2 \cdot p_c}{|g_k|^2 \cdot p_j \cdot f(x^t = y^t) + \sigma^2} \right) \tag{1}$$

where σ^2 is the noise variance, and B_k is the bandwidth of channel k. x^t and y^t are the selected channel in slot t of the CU and the jammer respectively. h_k and g_k are the gain of channel k for the CU and the jammer respectively. p_c and p_j are the transmission power of CU and jamming power of jammers respectively. $f(x)$ is an indicator function that equals 1 if x is true, 0 otherwise.

To analyze the impact of jamming attacks, we consider the following four popular kinds of jammers in military confrontation applications, and the jamming models are shown in **Figure 2.**

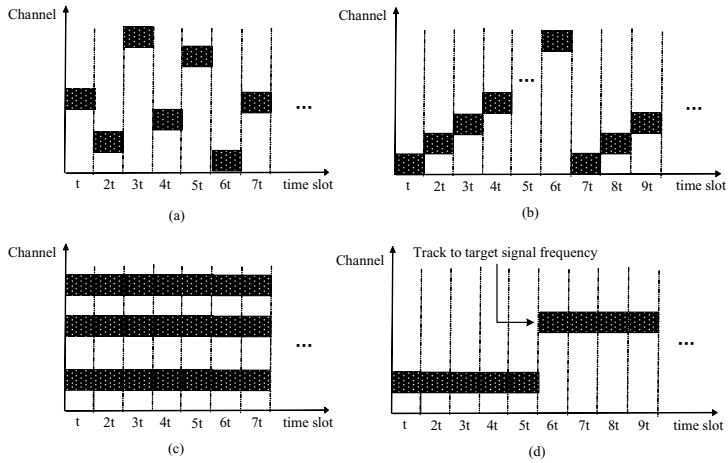

Figure 2. Random jamming model (a), Sweep jamming model (b), Comb spectrum jamming model (c) and Tracking jamming model (d).

- *Random jammer*, which randomly jams a channel in each time slot.
- *Sweep jammer*, which periodically scans the target frequency band. In a scanning period, the jamming frequency moves from the lowest frequency to the highest.
- *Comb spectrum jammer*, which superimposes multiple narrowband jamming signals within a certain bandwidth.
- *Tracking jammer*, which the jamming frequency follows the change of the target signal.

3. Anti-jamming Method based on Q-learning

Q-learning algorithm is a method to solve reinforcement learning control problem by using time sequence difference. In this article, we take the anti-jamming task of cognitive radio as a Markov decision process and propose an anti-jamming method based on Q-learning.

- State

In the interaction between CU and jammers, without knowing the jamming model and wireless channel model, CU can use Q-learning algorithm to determine its optimal strategy, that is, which idle channel is selected for signal transmission when the channel is jammed. The action selection of CU is based on the current state s^t. The state of CU is composed of the state of jammer and base station, which can be represented by *SINR*,

$$SINR = \frac{\left|h_k\right|^2 \cdot p_c}{\sigma^2 + \sum_{j=1}^{J} p_j \cdot \left|g_k\right|^2 \cdot f(x^t = y^t)} \tag{2}$$

- Action

In each time slot, CU and jammers take their own actions respectively. The idle channels among K channels defined as the possible actions x. In time slot t, CU selects channel $x(t) \in \{1,2,\ldots,K\}$ to determine the anti-jamming strategy of the current state, that is, to determine which channel is used to transmit data of the given power p_c.

Considering that at the beginning stage, CU don't know enough about the environmental information, and Q-value table of CU has not been updated a lot. So it is necessary to apply greedy strategy in exploring the results of different actions to update the Q-value table faster. In each time slot, CU has ε probability of exploration and (1-ε) probability to select the action (i.e. channel) maximizing the Q value in the current state. In the previous iterations, CU has a high probability of exploration in order to understand the jamming strategy of the jammer as soon as possible, so ε is bigger. After a period of learning, the Q-value table has been greatly updated and reduced ε. ε is denoted as

$$\varepsilon = 1/\log(n+1) \tag{3}$$

where n is the number of iteration rounds.

- Reward

The utility function in (1) is taken as the real-time reward function of reinforcement learning algorithm. Through the interaction between CU and jammers, CU changes the anti-jamming strategy in real time to maximize the cumulative reward.

- Update

Set function $V_\pi(s)$ represents the expected reward that can be obtained if we continue to carry out strategy π after reaching a certain state s, which is given by

$$V_\pi(s) = E_\pi(G_0 \left| S_0 = s\right.) = E_\pi(\sum_{t=0}^{\infty} \gamma^t R_{t+1} \left| S_0 = s\right.) \tag{4}$$

where π is the mapping from state to action and γ is the discount factor that controls the impact of future rewards on the optimal decision.

In each time slot t, the goal of CU is to select an action x that maximizes the expected accumulated future reward $V_\pi(s)$. The Q-Learning algorithm can finding the optimal decision in each state based on iterative evaluations of possible actions for Q-table. So, CU updates Q-table by Bellman formula as follow,

$$Q(s^t, x^t) = (1-\alpha)Q(s^t, x^t) + \alpha(u^t(s^t, x^t) + \gamma \max_x Q(s^{t+1}, x^t)) \tag{5}$$

The smaller γ, the more attention is paid to immediate income; the higher γ, the greater the weight of future rewards. α is the learning rate, the larger α, the fewer iterations, and vice versa. In our work, we set $\gamma = 0.7$.

The proposed method is summarized in Method 1.

Method 1 : Anti-jamming method based on Q-Learning

Initialize: α, γ, ε
Initialize: $Q(s,x) = 0 \quad \forall s, x$
For t =1, 2 …
  Choose an initial state at random $s_0 = SINR^0$;
  While do
    Choose x^t via epsilon-greedy
    Generate a random number r
      If ε<r:
        $x^t = \arg\max Q(s, x)$
      Else:
        Exploration, randomly choose a action;
      End if
    Use Channel x^t to send signals with power p_c;
    Obtain next state s^{t+1};
    Obtain u^t
    Update $Q(s, a)$:
      $Q(s^t, x^t) = (1-\alpha)Q(s^t, x^t) + \alpha(u^t(s^t, x^t) + \gamma \max_x Q(s^{t+1}, x^t))$
    Update $s = s^{k+1}$;
  End while
End for

4. Simulation and Results

The simulation is based on MATLAB to build a common scenario of cognitive radio communication with jamming. Table 1 shows the simulation parameters of the scenario. And we make the following settings to simplify the scenario:

- The channel gain obeys a uniform distribution over (a, b) and ignores the impact of distance;
- CU and jammers have the same numbers of channels and time slots, and the time slots are strictly aligned;
- The basic unit of time is time slot, and the basic unit of frequency is channel (apart from channel capacity).

Table 1. Simulation parameter settings

	Parameter	Value
Parameters of CU	number of channels	K=256

	number of slots	T=512
	channel gain	$h_k \sim U(0, 1)$
	bandwidth	B=1MHz
	power	p_c=5
	background noise power	σ^2=1
Major parameters of jammers	channel gain	$h_j \sim U(0, 1)$
	power(expect comb spectrum jammers)	p_j=8
	power(comb spectrum jammers)	$p_j=8/(K/\Delta f)$
	frequency interval (comb spectrum jammers)	Δf =16 channels
	sweep rate (sweep jammers)	1 channel per slot
	random rate (random jammers)	1 channel per slot
	response time (tracking jammers)	T_j=16 slots

In addition, the traditional anti-jamming methods, i.e. the DSSS method and the FH method, are used as the simulation comparison group. The spreading gain of the DSSS method is 4 and the frequency hop rate of the FH method is 1 hop/slot.

For jamming scenarios, we employ four typical jammers: Random jammers, Comb spectrum jammers, Sweep jammers and Tracking jammers, corresponding to scenario 1, scenario 2, scenario 3 and scenario 4, respectively.

From **Figure 3** to **Figure 6** show the performance when dealing with the four typical jammers respectively. And after the ant-jamming methods are stable, the results are shown in Table 2.

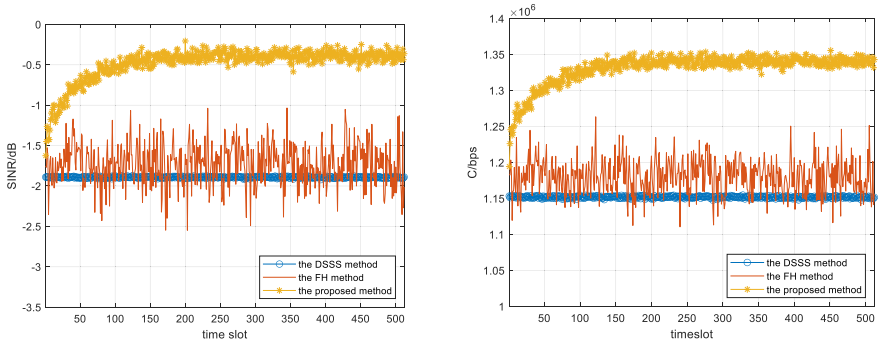

Figure 3. Performance against random jammers.

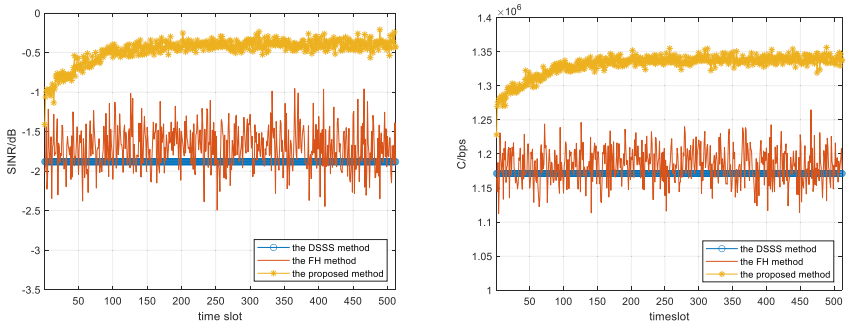

Figure 4. Performance against comb spectrum jammers.

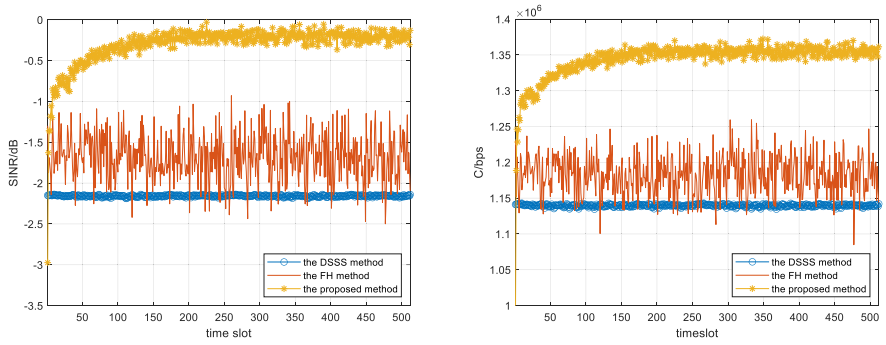

Figure 5. Performance against sweep jammers.

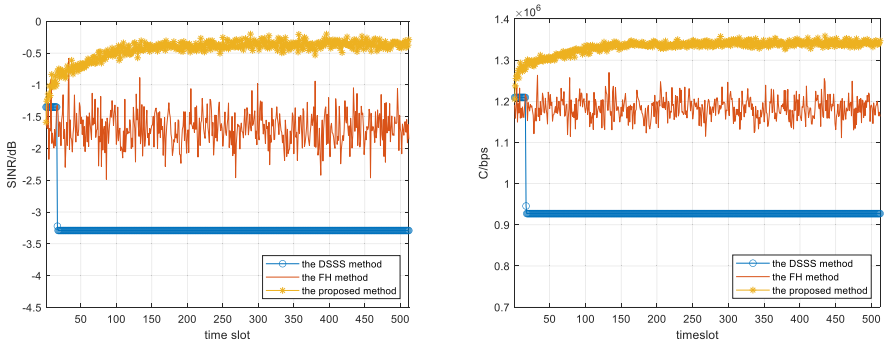

Figure 6. Performance against tracking jammers.

Table 2. Simulation results of the ant-jamming methods

Scenario	Performance	Mean of DSSS	Mean of FH	Mean of Proposed	improved (cp. DSSS)	improved (cp. FH)
Scenario 1	SINR	-1.892dB	-1.735dB	-0.391dB	1.501dB	1.344dB
	total channel capacity	1.152Mbps	1.180Mbps	1.340Mbps	16.37%	13.56%
Scenario 2	SINR	-1.882dB	-1.698dB	-0.394dB	1.488dB	1.304dB
	total channel capacity	1.171Mbps	1.184Mbps	1.339Mbps	14.35%	13.09%
Scenario 3	SINR	-2.157dB	-1.68dB	-0.203dB	1.954dB	1.478dB
	total channel capacity	1.14Mbps	1.182Mbps	1.355Mbps	18.86%	14.64%
Scenario 4	SINR	-1.882dB	-1.698dB	-0.394dB	2.931dB	1.325dB
	total channel capacity	0.927Mbps	1.183Mbps	1.341Mbps	44.66%	13.36%

Simulation and results show that:

- Under the different jamming models, the proposed method takes a certain time for iterative learning to stabilize the performance, but after that, it is more stable and better than the FH method. The SINR can be increased by 1.3-1.5dB, and the total channel capacity can be increased by 13%~15% specifically. That's because the proposed method effectively avoids the interference by learning, while the FH method still has the possibility of collision.
- Under the tracking jammers, the performance of DSSS method deteriorates rapidly since its frequency is fix and tracked by jammers, but the performance

of the proposed method remains good. Compared with the DSSS method, the SINR performance can be improved by 2.9dB, and the total channel capacity can be increased by 44.66%, while the improvement values under other jamming models are 1.4dB~2dB and 14%~19% respectively.

5. Conclusion

In this paper, one anti-jamming method of cognitive radio based on Q-learning is proposed. By applying the Q-learning algorithm, without knowing the channel model and the jamming model, proposed method can achieve anti-jamming communication with better performance of cognitive radio. To verify the effectiveness of the proposed method, the anti-jamming performance under four typical jammers is simulated. The result show that the proposed method has higher SINR and higher system transmission capacity compared with the traditional anti-jamming methods. In the future work, we plan to accelerate the convergence speed of the proposed method, and investigate multi CUs scenarios which exists the competition and collaboration among CUs.

References

[1] Mpitziopoulos A, Gavalas D. An effective defensive node against jamming attacks in sensor networks. Security and Communication Networks, 2009, 2(2): 145-163.

[2] Z. Chao, X. Zhu, W. He, Y. Ban and S. Chen, A Frequency-Hopping Communication System Based On Multiple Parallel Hopping, 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2018 Oct 12-14; Chongqing, China:pp. 2163-2169.

[3] Xiao Ma; Xuanhe Yang; Xin Jin; Xiaqing Miao; Shuai Wang. A Digital Coherent Frequency Hopping Anti-jamming Receiver Based on IIR, 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). 2020 Jun 12-14; Chongqing, China: 10.1109/ ITNEC48623.2020.9084695.

[4] X. Meng, R. Tao, L. Jia, Liping Liang, Wenchi Cheng, Hailin Zhang. Index Modulation Based Joint Mode-Frequency Hopping. IEEE Communications Letters, 2021, 25(6), pp.1810-1814.

[5] Wenzhun Huang, Shanwen Zhang, Robert Hao Yan. Novel Spread Spectrum Communication Theory and the Anti-jamming Applications, 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021 Feb 20-22; Coimbatore, India: 10.1109/ICICT50816.2021.9358748.

[6] J. Mitola. Cognitive radio for flexible mobile multimedia communications, 1999 IEEE International Workshop on Mobile Multimedia Communications (MoMuC'99) (Cat. No.99EX384), 1999 Nov 15-17; San Diego, CA, USA: pp. 3-10.

[7] Luliang Jia, Nan Qi, Feihuang Chu, Shengliang Fang, Ximing Wang, Shuli Ma, Shuo Feng. Game-Theoretic Learning Anti-Jamming Approaches in Wireless Networks, IEEE Communications Magazine, 2022, 60(5), pp.60-66.

[8] Khalid Ibrahim, Soon Xin Ng, Ijaz Mansoor Qureshi, Aqdas Naveed Malik, Sami Muhaidat. Anti-Jamming Game to Combat Intelligent Jamming for Cognitive Radio Networks, IEEE Access, 2021, Oct 04; pp.137941-137956

[9] Ximing Wang, Yuhua Xu, Jin Chen, Chunguo Li, Xin Liu, Dianxiong Liu, Yifan Xu. Mean Field Reinforcement Learning Based Anti-Jamming Communications for Ultra-Dense Internet of Things in 6G. 2020 International Conference on Wireless Communications and Signal Processing (WCSP), 2020 Oct 21-23; Nanjing, China: 10.1109/WCSP49889.2020.9299742.

[10] Jinlin Peng, Zixuan Zhang, Qinhao Wu, Bo Zhang. Anti-Jamming Communications in UAV Swarms: A Reinforcement Learning Approach, IEEE Access, 2019, 7, pp.180532-180543.

[11] ZHU Rui, MA Yongtao, NAN Yafei, ZHANG Yunlei. Cognitive Radio Anti-Jamming Decision Algorithm Based on Improved Reinforcement Learning. Journal of Frontiers of Computer Science and Technology, 2019, 13(4): 693-701.