Peer-reviewed version available at Appl. Sci. 2018, 8, 1911; doi:10.3390/app8101911

Article

Location-aware Cooperative Anti-jamming Distributed Channel Selection Approach in UAV Communication Networks

Yifan Xu [1], Guochun Ren [1], Jin Chen [1], Xiaobo Zhang [1], Luliang Jia [1] and Lijun Kong [1]

[1] the College of Communications Engineering, Army Engineering University of PLA, Nanjing 210000, China; yifanxu1995@163.com; guochunren@yeah.net; chenjin99@263.net; xb_zhang2008@126.com; jiallts@163.com; konglijun2018@163.com

Academic Editor: name
Version September 13, 2018 submitted to Preprints

Abstract: This paper investigates the cooperative anti-jamming distributed channel selection problem in UAV communication networks. Considering the existence of malicious jamming and co-channel interference, a location-aware cooperative anti-jamming scheme is designed for the purpose of maximizing the users' utilities. Users in the UAV group cooperate with each other via location information sharing. When the received interference energy is lower than mutual interference threshold, users conduct channel selection strategies independently. Otherwise, users take joint actions with a cooperative anti-jamming pattern under the impact of mutual interference. Aimed at the independent anti-jamming channel selection problem under no mutual interference, a Markov Decision Process framework is introduced, whereas for the cooperative anti-jamming channel selection case under the influence of co-channel mutual interference, a Markov game framework is employed. Furthermore, motivated by reinforcement learning with a "Cooperation-Decision-Feedback-Adjustment" idea, we design a location-aware cooperative anti-jamming distributed channel selection algorithm (LCADCSA) to obtain the optimal anti-jamming channel strategies for the users with a distributed way. In addition, the channel switching cost and cooperation cost, which have great impact on the users' utilities, are introduced. Finally, simulation results show that the proposed algorithm converges to a stable solution with which the UAV group can avoid the malicious jamming as well as co-channel interference effectively.

Keywords: location-aware; cooperative anti-jamming; Markov decision process; Markove game; reinforcement learning

1. Introduction

Unmanned aerial vehicle (UAV) communication networks, as a kind of newly-developing wireless communication networks, have become a hot research issue[1,2]. When important tasks are carried out, how to construct a reliable and robust UAV network is of great significance. In some scenario with strongly competitive characteristics, the destructive effect caused by malicious jamming must be taken into consideration.

In traditional anti-jamming research field, various techniques have been adopted, i.e., power control, Uncoordinated Frequency Hoping (UFH) and Frequency Hopping Spread Spectrum (FHSS)[3]. However, there are some limitations using these techniques: i) Anti-jamming Power control is ineffective under the circumstance of high jamming power. ii) Traditional UFH and FHSS consume a large number of spectrum resources, and they are not able to work well under the dynamic spectrum environment[4–6].

In addition, game theory [7–10], as a strong theoretical tool, is suitable to model the anti-jamming competitive scenario. Specifically, Stackelberg game approach[10], as a kind of hierarchical game, has been widely used in anti-jamming field. For example, in [11], the authors summarized the application of Stackelberg game in anti-jamming dense networks, and introduced several classical

Peer-reviewed version available at Appl. Sci. 2018, 8, 1911; doi:10.3390/app8101911

anti-jamming scenarios and system models. Moreover, a outlook of the application of anti-jamming Stackelberg game was also made. In [12–14], Stackelberg game approaches were adopted for the anti-jamming power control problem, where the user acted as the leader, and the jammer acted as the follower of the game. Utility functions were designed, and Stackelberg Equilibriums (SE) were also obtained via game approaches. Moreover, considering the channel selection problem under malicious jamming environment, a hierarchical anti-jamming channel selection scheme was proposed using a Stackelberg game framework[15]. However, most existing studies under the Stackelberg game framework formulated the interactions between user-side and jammer-side, which brought large deviation in information acquisition. Thus, studies focusing on anti-jamming channel selection under dynamic jamming environment are of great importance.

In fact, the dynamic feature of the channel state brings some challenge to anti-jamming channel selection. In addition, the mobility of UAVs also influences the receiver's signal energy, causing the decline of communication quality[16]. In [17], the author investigated the multi-stage spectrum access problem for Flying Ad-Hoc Network (FANET). Markov decision process (MDP)[18,19], as a decision framework under dynamic channel environment, has been adopted to model the anti-jamming problem. For the purpose of solving the MDP problem, Q-learning[20] methods are usually employed using a "Decision-Feedback-Adjustment" structure to obtain the optimal strategy. For instance, in [4], the author formulated the anti-jamming decision problem as a MDP, and obtained the best anti-jamming scheme via Q-learning. Furthermore, in [21], a deep Q network was built, and the anti-jamming channel selection problem was solved using a deep reinforcement learning method. In addition, in view of the multi-user scenarios in anti-jamming field, MDP problem has been extended to Markov game[22], and several learning algorithms were designed for multi-user scenarios. In [23], a multi-agent learning algorithm was proposed to obtain a stable solution for dynamic spectrum access problem. In [24–26], some multi-user reinforcement learing methods were adopted, where users took actions independently. However, in those methods mentioned in [24–26], users' states are influenced by each other, which leads to unsteady learning environments and poor decision effects.

Taking a overall consideration of the challenges and inspirations brought by above studies, in this paper, we mainly focus on the anti-jamming channel selection problem under dynamic environment, where the channel state and UAVs' locations are time-varying. Moreover, the channel switching cost and cooperation cost are introduced, which have a great impact on the users' utilities. A cooperative anti-jamming mechanism is constructed, in which users can realize information sharing and take actions jointly. Specifically, users in the UAV group sense the location information, and calculate the receiving signal energy as well as estimate whether they are influenced by co-channel interference. For the case where users are not influenced by co-channel interference, a MDP is formulated to model the anti-jamming problem for the UAVs, and a independent Q-learning method is employed to obtain the users' channel selection strategies. For the case where users are indeed influenced by co-channel interference, a Markov game is formulated, and a multi-agent Q-learning method is designed for UAV communication networks. To sum up, the main contributions are summarized as follows:

- A cooperative anti-jamming mechanism is designed for UAV communication networks, where UAVs cooperate via information exchange. Considering the influence of co-channel interference, a MDP and a Markov game are formulated respectively.
- A location-aware cooperative anti-jamming distributed channel selection algorithm (LCADCSA) is designed for the anti-jamming selection problem. Without the influence of co-channel interference, a independent Q-learning method is adopted, while under the influence of co-channel interference, a multi-agent Q-learning method is employed.
- Simulation results exhibit the performance of the proposed LCADCSA, which can avoid the malicious jamming and co-channel interference effectively. Moreover, the influence of channel switching cost and cooperation cost are investigated.

Compare this paper to our previous works[27,28], which studied anti-jamming channel selection in wireless communication networks, and to our previous works [29], the main differences are: i)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2018 doi:10.20944/preprints201809.0227.v1

Peer-reviewed version available at Appl. Sci. 2018, 8, 1911; doi:10.3390/app8101911

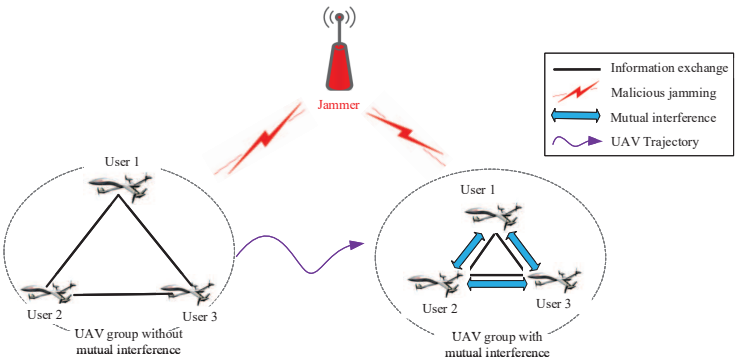


Figure 1. Cooperative anti-jamming UAV communication networks.

Work [27] investigated the multi-agent learning method for anti-jamming problem, and work [28] considered the single reinforcement learning in fading environment. However, both these two works did not take the mobility of UAVs into consideration. Whereas in our paper, the anti-jamming channel selection approach in UAV communication networks is investigated, while taking the mobility of UAVs into consideration, which causes the variation of co-channel interference. Moreover, channel switching cost and cooperation cost are introduced, which influence the users' utilities. ii) In [29], we focused on anti-jamming power control problem in UAV communication networks, whereas in this paper, the cooperative anti-jamming channel selection scheme is designed, and a cooperative anti-jamming algorithm based on multi-agent reinforcement learning is derived, which obtains strategies by interacting with the environment.

The rest of this paper is shown as follows. In Section 2, the system model and problem formulation are investigated. In Section 3, the location-aware cooperative anti-jamming mechanism in UAV Group is designed. In Section 4, the proposed location-aware cooperative anti-jamming distributed channel selection algorithm (LCADCAS) is shown. In Section 5, simulations and discussions are conducted. In the end, we make conclusion in Section 6.

2. System Model and Problem Formulation

The system model is shown in Fig.1. Assume that there are N users (a transmitter-receiver UAV formation is treated as one user) and one jammer in the system scenario. UAVs are under the threat of malicious jammer. In the UAV group, the locations of UAVs are time-varying, and UAVs cooperate with each other via information exchange. Denote the user set as $\mathcal{N}=\{1,...,n,...,N\}$. The available channel set for user is $\mathcal{M}=\{1,...,m,...,M\}$.

Consider two different case of UAV transmission: i) When users are close to each other, and transmitting in the same channel, high received signal energy from other users made them influenced by co-channel interference. ii) When users are far away from each other, the received signal energy from other users is somehow low, which means the users are not influenced by co-channel interference. Mutual interference threshold τ_0 is used to measure the influence of co-channel interference, that is: When received interference energy is lower than τ_0, the UAV communication network is not influenced by co-channel interference, and vice versa.

Assume that channel strategy a_n means user n chooses channel $c_n, c_n \in \mathcal{M}$, to transmit, a_{-n} is the channel strategy combination of all users except user n, a_j is the jamming channel. Users transmit with CSMA pattern, then the throughput of user n is expressed as:

$$Tr_n\left(a_n, a_{-n}, a_j\right) = \left(1-f\left(a_n, a_j\right)\right) \frac{1}{I_n(c_n)} \log_2\left(1+\frac{P_n d_n^{-\alpha}}{N_{c_n}}\right), \quad (1)$$

Peer-reviewed version available at Appl. Sci. 2018, 8, 1911; doi:10.3390/app8101911

where d_n denotes the distance between the transmitter and the receiver of user n, P_n represents the user n's transmission power. α is the path-loss exponent, and N_{c_n} represents the channel noise power. Moreover, $I_n(c_n)$ is the congestion degree of channel c_n, which is expressed as:

$$I_n(c_n) = \begin{cases} 1 + \sum\limits_{x \in \mathcal{N}/n} f(a_n, a_x), & P_x d_{x,n}^{-\alpha} \geq \tau_0, \\ \dots & \\ 1, & P_x d_{x,n}^{-\alpha} < \tau_0. \end{cases} \tag{2}$$

where P_x is the transmission power of user $x, x \in \mathcal{N}/n$, $d_{x,n}$ denotes the interference distance from user x to user n, then $P_x d_{x,n}^{-\alpha}$ can be viewed as the received signal energy from user x to user n. $f(a_n, a_x)$ is a indicator function, which depicts the channel occupation of user n's selected channel, shown as:

$$f(x,y) = \begin{cases} 1, x = y, \\ 0, x \neq y. \end{cases} \tag{3}$$

As shown in Eq. (1), $Tr_n(a_n, a_{-n}, a_j)$ depicts the user n's throughput under the threat of malicious jamming and co-channel interference, and in Eq. (2), the congestion degree $I_n(c_n)$ reflects the number of users who are influenced by co-channel interference.

Consider the channel switching of user, we introduce the channel switching cost unit W_s to evaluate the performance loss. Moreover, if UAVs cooperative with each other to share more information and take actions jointly, a cooperation cost unit W_c is also brought in. Then, as a tradeoff between throughput and its cost, the utility of user n in one time slot is defined as:

$$u_n(a_n, a_{-n}, a_j) = Tr_n(a_n, a_{-n}, a_j) - W_s \delta_s - W_c \delta_c, \tag{4}$$

where δ_s and δ_c are indicator functions for channel switching and cooperation. $\delta_s = 1$ indicates that channel switching occurs at the beginning of current slot, whereas $\delta_s = 0$ means that the user keep its channel strategy. $\delta_c = 1$ indicates that users are cooperation with each other and take joint channel actions, whereas $\delta_c = 0$ means users choose channels independently. The optimization object of user n is:

$$a_n = \arg\max_{a_n \in \mathcal{M}} u_n(a_n, a_{-n}, a_j). \tag{5}$$

Every user in the UAV group wants to employ an optimal anti-jamming channel selection strategy for the purpose of maximizing the network's throughput. However, due to the dynamic feature of the jamming channel and the time-varying locations of UAVs, solving the optimization problem is challenging. Therefore, in the next section, we combine MDP, Markov game and Q-learning to investigate and solve the anti-jamming channel selection problem in UAV communication networks.

3. Location-aware Cooperative Anti-jamming Mechanism in UAV Group

In this part, the location-aware cooperative anti-jamming mechanism in UAV group is designed and analyzed. According to the location sharing information of UAVs, the process of the designed cooperative anti-jamming mechanism is shown in Fig. 2.

The details are shown as follows. When users are transmitting, they share location information with each other. After that, each user makes a judgement according to the received interference energy. If users are influenced by co-channel interference, a Markov game is formulated to model the cooperative anti-jamming problem. Every user has to avoid the jamming channel, as well as avoiding the co-channel interference channel for the purpose of realizing higher throughput. If users are in a specific location where they are not influenced by co-channel interference, a Markov decision process is able to formulate the coming optimization problem. Each user makes anti-jamming decision independently via single Q-learning approach.

Peer-reviewed version available at Appl. Sci. 2018, 8, 1911; doi:10.3390/app8101911

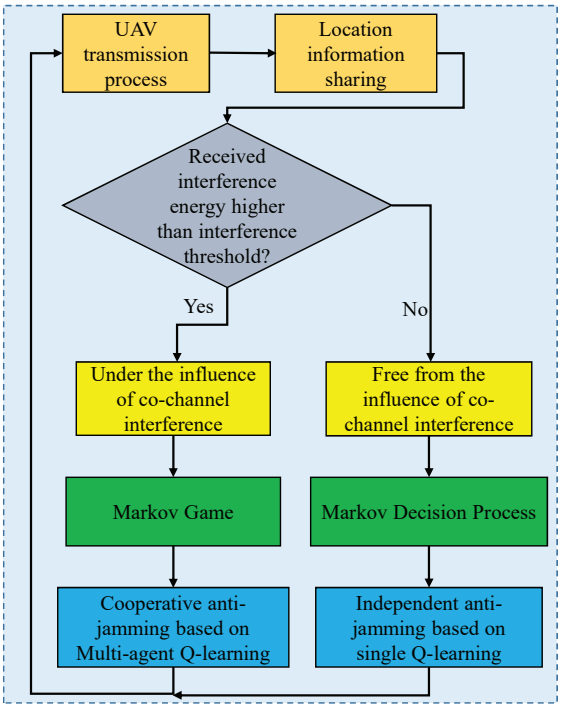


Figure 2. Location-aware cooperative anti-jamming mechanism in UAV group.

3.1. Markov Decision Process

As mentioned above, when users are not influenced by co-channel interference, the anti-jamming channel selection problem can be formulated as a Markov decision process, and each user's strategy is independent to others'.

Definition 1. *When users are free from the influence of co-channel mutual interference, the Markov decision process of user n can be express as* (S_n, A_n, R_n, T_n)*, where:*

- S_n *is the discrete set of user n's environment.*$s_n(t) = \left(f_n(t), f_j(t)\right), s_n(t) \in S_n$ *is the environment state of user n in time t.* $f_n(t)$ *and* $f_j(t)$ *represent user n's transmission channel and jamming channel respectively. In this case, user n's state is not influenced by other users.*
- A_n *is the channel strategy set of user n,* $a_n(t) \in A_n$ *denotes the channel selection strategy under the state of t moment, similarly, user n's strategy is not influenced by others.*
- *The reward function of user n is* R_n*, which satisfies* $S_n \times A_n \to R_n$*. Specifically, for every state* $s_n(t)$*, user can obtain a reward with action* $a_n(t)$*.*
- *The state transition function* T_n*, which satisfies* $S_n \times A_n \to T_n$*. Moreover, it also meets with Markov property, shown as:*

$$\begin{aligned} &P\left[s_n(t+1) \,|s_n(t), a_n(t), ..., s_n(0), a_n(0)\right] \\ &= P\left[s_n(t+1) \,|s_n(t), a_n(t)\right], \;\; a_n(t) \in A_n, s_n(t) \in S_n. \end{aligned} \tag{6}$$

For each user in the UAV group, the corresponding Markov decision process can be solved using single Q-learning method. Optimal anti-jamming selection strategies can be derived as well.

3.2. Markov Game

When users are under the influence of co-channel interference, the anti-jamming channel selection problem can be formulated as a Markov game, each user's strategy is related to other user's strategy.

Thus, all users in the group take joint actions to fight against malicious jammer, and avoid co-channel interference as much as possible.

Definition 2. *When users are influenced by co-channel interference, the anti-jamming channel selection problem can be formulated as a Markov game, which can be expressed as* $\mathcal{G}=\{\mathcal{S},\mathcal{A},\mathcal{T},\mathcal{R}_1,...,\mathcal{R}_N\}$. *The details are shown as follows:*

- $\mathcal{S}$ *is the discrete state set. In cooperative anti-jamming issue,* $s(t)=\left(f_1(t),...,f_N(t),f_j(t)\right),s(t)\in\mathcal{S}$ *represents all users' states and the jammer's state. Users' states are correlative.*
- *Denote* A_n *as the channel selection set of user* n*, and* $\mathcal{A}$ *is the joint action set of all users in the UAV group. The action space is* $\mathcal{A}=A_1\times\cdots\times A_N$.
- $\mathcal{T}$ *is the state transition function, and the state space is* $\mathcal{S}\times\mathcal{A}\times\mathcal{S}$*, which satisfies* $\sum_{s\in\mathcal{S}}\mathcal{T}(s,\mathbf{a},s')=1$. *Specifically,* $\mathbf{a}$ *is the joint channel selection strategy,* s *is the current state.* s' *is the coming state after all users take joint action* $\mathbf{a}$ *under state* s*. The state transition function* $\mathcal{T}$ *satisfies Markov property as well.*
- $\mathcal{R}_1,...,\mathcal{R}_N$ *are the reward functions of each user, and they satisfy* $\mathcal{S}\times\mathcal{A}\to\mathcal{R}_n,n\in\mathcal{N}$*. For UAVs in the group, no matter what joint actions are being taken, each one can obtain an immediate reward.*

Aiming at two different state in the UAV group, the anti-jamming channel selection problem are formulated as Markov decision process and Markov game respectively. For Markov decision process, single Q-learning approach is used to obtain each user's optimal channel selection strategy. Whereas for Markov game, multi-agent learning method is adopted for the purpose of acquiring the joint channel selection strategies for all users.

3.3. Single Q-learning

Single Q-learning method is suitable for the case where UAV group is not influenced by co-channel mutual interference. In traditional single-Q learning algorithm, every user maintains and updates its independent Q table Q^n, for user n, the updating process of Q function is shown as:

$$Q_{t+1}^n(s_n,a_t^n)=(1-\lambda_n)Q_t^n(s_n,a_t^n)+\lambda_n\left[r_t^n+\gamma_n V_n\left(s'_n\right)\right], \tag{7}$$

where λ_n is the learning rate of user n, γ_n represents the discount factor for Q table update. r_t^n is the immediate reward of user n under environment s_n, also can be viewed as the normalized utility, which is:

$$r_t^n=\left(1-f\left(a_t^n,a_j^n\right)\right)\frac{1}{I_t^n(c_n)}-w_s\delta_s-w_c\delta_c, \tag{8}$$

where w_s and w_c are normalized switching cost unit and normalized cooperation cost respectively. $V_n(s'_n)$ is the value function of user n, in single Q-learning, $V_n(s'_n)$ can be expressed as:

$$V_n\left(s'_n\right)=\max\left\{Q_t^n\left(s'_n,a\right)\right\}. \tag{9}$$

The defined value function $V_n(s'_n)$ can be viewed as finding the highest benefit in user n's "memory" under state s'_n.

Each user in the UAV group adopts independent Q-learning via a "Decision-Feedback-Adjustment" way, and each user can converge to a optimal channel selection strategy.

3.4. Multi-agent Q-learning

Aimed at the case where UAVs are influenced by co-channel interference, an cooperative anti-jamming channel selection algorithm based on multi-agent Q-learning is designed. In the proposed

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2018 doi:10.20944/preprints201809.0227.v1

Peer-reviewed version available at Appl. Sci. 2018, 8, 1911; doi:10.3390/app8101911

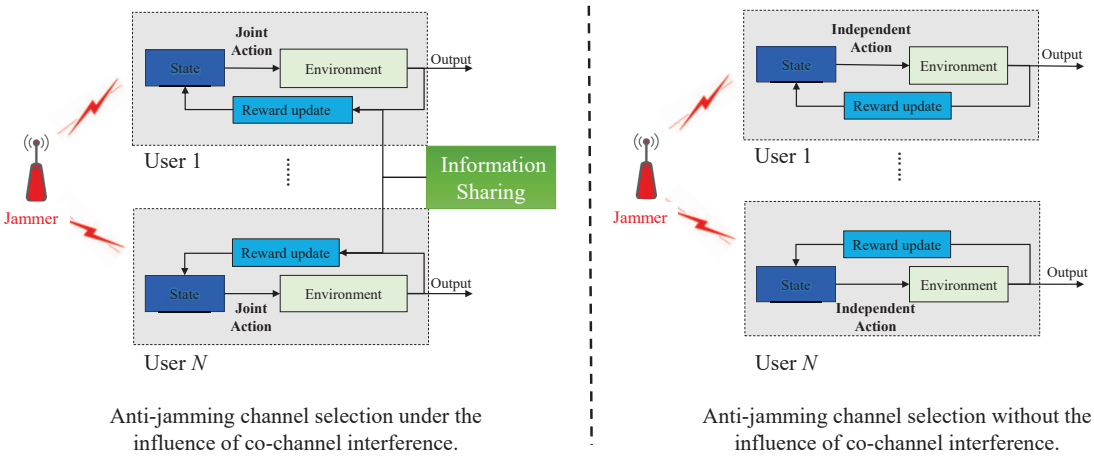


Figure 3. Anti-jamming distributed channel selection framework under different cases.

multi-agent Q-learning, each user maintains and updates a Q table $\tilde{Q}^n$ which is based on joint action $\mathbf{a}$. Similar to single Q-learning, the Q function updates using the following rule:

$$\widetilde{Q}_{t+1}^n\left(s,\mathbf{a}_t\right)=\left(1-\widetilde{\lambda}_n\right)\widetilde{Q}_t^n\left(s,\mathbf{a}_t\right)+\widetilde{\lambda}_n\left[\widetilde{r}_t^n+\widetilde{\gamma}_n\widetilde{V}_n\left(s'\right)\right], \tag{10}$$

where $\widetilde{\lambda}_n$ is user n's learning rate under joint action, $\widetilde{\gamma}_n$ is the discount factor correspondingly. $\widetilde{r}_t^n$ denotes the user n's immediate reward when taking joint action $\mathbf{a}$ under state s. Moreover, $\widetilde{r}_t^n$ represents the normalized throughput under joint action, which can also be shown as:

$$\widetilde{r}_t^n=\left(1-f\left(a_t^n,a_j^n\right)\right)\frac{1}{I_t^n\left(c_n\right)}-w_s\delta_s-w_c\delta_c. \tag{11}$$

$\widetilde{V}_n\left(s'\right)$ is user n's value function in multi-agent Q learning, which is:

$$\widetilde{V}_n\left(s'\right)=\widetilde{Q}_t^n\left(s',\mathbf{a}^*\right), \tag{12}$$

where $\mathbf{a}^*$ represents the best joint action when all users' total benefit reaches maximum. $\mathbf{a}^*$ can be expressed using the following equation:

$$\mathbf{a}^*=\arg\max\sum_{n=1}^{N}\widetilde{Q}_t^n\left(s',\mathbf{a}\right). \tag{13}$$

Without loss of generality, either in single Q-learning or in multi-agent Q-learning, *ε-greedy* policy is introduced for the purpose of avoiding local optimum. Moreover, it is obviously that cooperation cost unit $\delta_c=0$ in single Q-learning, and that $\delta_c=1$ in multi-agent Q-learning. As in single Q-learning, users take actions dependently, while in multi-agent Q-learning, users cooperates with each other to avoid mutual interference.

4. Location-aware Cooperative Anti-jamming Distributed Channel Selection Algorithm

In this section, the location-aware cooperative anti-jamming distributed channel selection algorithm is designed.

As shown in Fig. 3, it depicts the anti-jamming distributed channel selection framework under different cases. In the left part of Fig. 3, the anti-jamming distributed channel selection framework under the influence of co-channel interference is designed. Users in the UAV group adopt a "Joint action-Feedback-Adjustment" idea, and realize cooperative anti-jamming using multi-agent learning.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2018 doi:10.20944/preprints201809.0227.v1

Peer-reviewed version available at Appl. Sci. 2018, 8, 1911; doi:10.3390/app8101911

Algorithm 1: *Location-aware Cooperative Anti-jamming Distributed Channel Selection Algorithm*

Initialization:

Initialize the starting time, ending time and relative learning parameters of the simulation .
Initialize every user n's joint action Q table $\widetilde{Q}^n$ and single Q table Q^n.
Set the initial locations and states of all users.

Repeat Interations:

Each user observes current environment state, and then makes a judgement about the co-channel interference according to shared location information.
If users are under the influence of co-channel interference, go to multi-agent Q-learning.
Otherwise, go to single Q-learning.

Multi-agent Q-learning:

(1) Each user observes and chooses one transmission channel, using the following rules:
- Randomly choose a joint action combination **a** with probability ε.
- Choosing the best joint action $\mathbf{a}^*$ according to Eq. (13), with probability $1-\varepsilon$.

(2) Each user calculates its immediate reward r_t^n via joint action, and then transfers the environment state.
(3) The Q table $\widetilde{Q}^n$ is updated according to Eq. (10).

Single Q-learning:

(1) Each user observes and chooses one transmission channel, using the following rules:
- Randomly choose a independent action a^n with probability ε.
- Choosing the best action a^{n*} with probability $1-\varepsilon$, which realizes the highest Q value in current state.

(2) Each user calculates its own immediate reward r_t^n, and then transfers the environment state.
(3) The Q table Q^n is updated according to Eq. (7).

End

Jump out the repeat process when the algorithm reaches the maximal iterations.

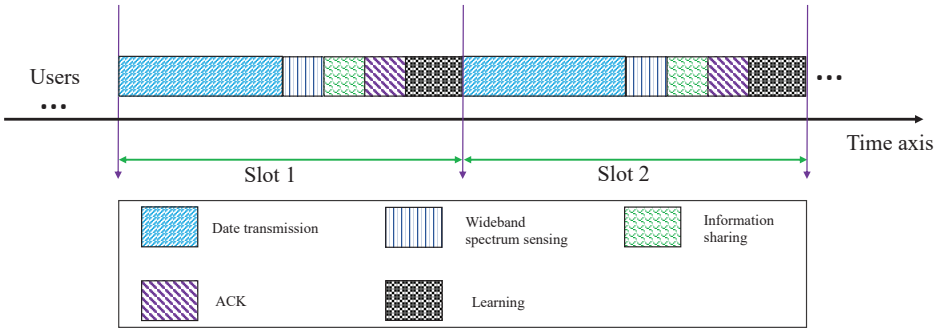


Figure 4. Anti-jamming transmission slot structure for UAVs.

In this framework, users ought to share their strategies so that they can take joint actions. In the right part of Fig. 3, the anti-jamming framework under the case that users are not influenced by co-channel interference is shown. In this framework, users adopt the single Q-learning mechanism which uses a "Independent action-Feedback-Adjustment" idea. After receiving the immediate reward, each user adjusts its strategy independently. All users in the UAV group ought to sense location information to judge whether they are influenced by co-channel interference. The details of location-aware cooperative anti-jamming distributed channel selection algorithm are shown in Algorithm 1.

In addition, the transmission slot structure of UAVs are shown in Fig. 4. Each user chooses a channel to transmit packet firstly. And then, a process of wideband spectrum sensing (WBSS) is conducted for the purpose of acquiring the currently available channel state. Later, users start information sharing, and make judgement about whether they are influenced by co-channel interference. Then, users receive ACKs from they receivers to confirm whether the data packets are sent successfully. During the last process of the user's slot, each user starts learning to update their strategies in the next slot.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2018 doi:10.20944/preprints201809.0227.v1

Peer-reviewed version available at Appl. Sci. 2018, 8, 1911; doi:10.3390/app8101911

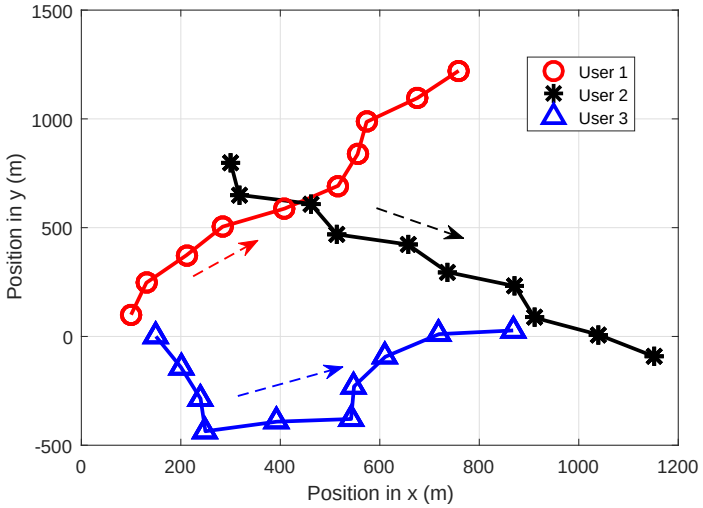


Figure 5. The trajectory setting for UAVs.

5. Simulation Results And Discussions

5.1. Simulation Setting

In the simulation part, a UAV communication network which consists of three users and one jammer is investigated. The available channel number for users to access is 4. The jammer send sweeping jamming signal to the available channels, and the jamming signal stays at one channel for about 2.28ms. The transmission time in each user's slot is $T_{tr} = 0.98$ms, and the time for WBSS, information sharing, ACK and learning are totally $T_{wbss} + T_{is} + T_{ack} + T_{le} = 0.2$ms in each slot.

Other simulation settings are shown as follows. Assume the transmission power of each user is 0.1W, the initial locations of three users are (100m,100m), (300m,800m) and (150m,0m) respectively. The trajectories of UAVs are shown in Fig. 5, and the flying time is divided into 10 epoches. UAVs move 150m per epoch, and the duration time of each epoch is set to be 3s. Furthermore, the pass-loss factor $\alpha = 2$, co-channel interference threshold is 6.25×10^{-7}W. The total simulation time is equal to the flying time(approximately 30s). Motivated by [30], $\lambda_1 = ... = \lambda_n = \tilde{\lambda}_1 = ... = \tilde{\lambda}_n = 0.8$, $\gamma_1 = ... = \gamma_n = \tilde{\gamma}_1 = ... = \tilde{\gamma}_n = 0.6$, $\varepsilon = 0.1$.

Moreover, Fig. 6 depicts the interference distance of UAVs. In detail, the flying process is divided into four stages. During flying time 0s to 6s (the first stage), the distance between user 1 and user 3 is less than 400m, and they will influenced by co-channel interference as the received signal energy is higher than threshold. During 6s to 15s (the second stage), user 1 and user 3 are influenced by co-channel interference. From 15s to 21s (the third stage), all users are keep relative far from each other, so there exists no co-channel interference, while from 21s to 30s (the fourth stage), user 2 and user 3 are influenced by co-channel interference.

5.2. Cumulative Normalized Utility of Users

In this part, the user's performance analysis is mainly investigated. As is mentioned in Algorithm1, when users are influenced by co-channel interference, the proposed location-aware cooperative anti-jamming distributed channel selection algorithm (LCADCSA) is based on multi-agent Q-learning. When users are not influenced by co-channel interference, the proposed LCADCSA algorithm is based

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2018 doi:10.20944/preprints201809.0227.v1

Peer-reviewed version available at Appl. Sci. 2018, 8, 1911; doi:10.3390/app8101911

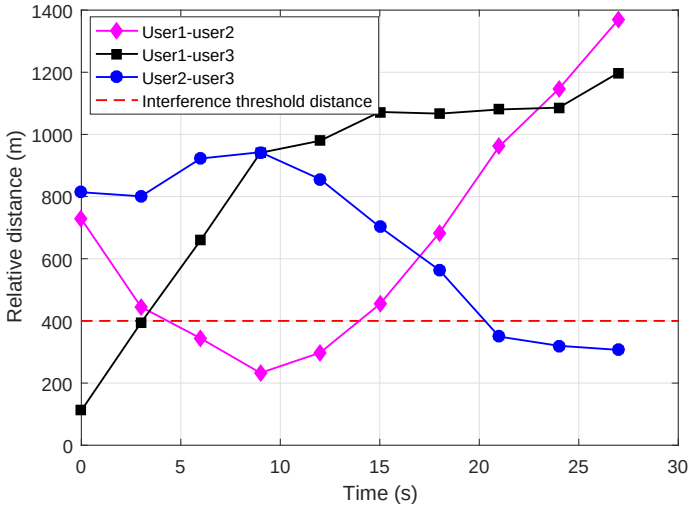


Figure 6. Relative distance variation setting.

on single Q-learning. For better clarification, we use cumulative normalized utility p to show the effective of LCADCSA approach, which is defined as follows:

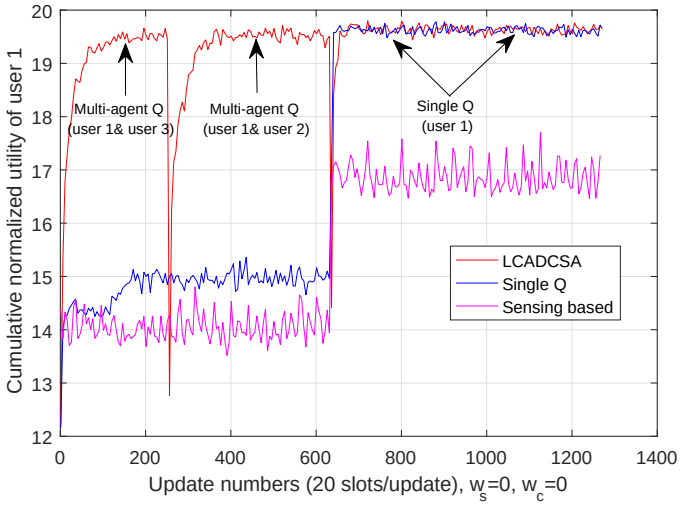
$$U_{cum} = \sum_{i=1}^{PN} \left(\left(1 - f\left(a_t^n, a_j^n\right)\right) \frac{1}{I_t^n\left(c_n\right)} - w_s \delta_s - w_c \delta_c \right). \tag{14}$$

where PN is the number of packet in every update, and PN is set to be 20 in the simulation, which means the cumulative normalized utility updates per 20 slots, and the time of each update is 23.6ms.

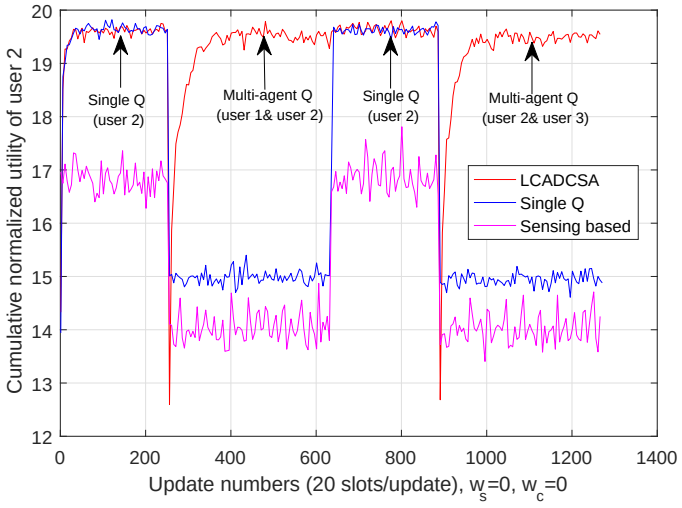
The cumulative normalized utilities of users are shown in Fig. 7(a), Fig. 7(b) and Fig. 7(c) respectively, where the channel switching cost and cooperation cost are set to be 0. As is shown in those three figures, the users' channel selection processes are divided into four stages: In the first stage, user 1 and user 3 cooperate with each other, and adopt multi-agent Q-learning, user 2 employs sing Q-learning. In the second stage, user 1 and user 2 cooperate with each other, and adopt multi-agent Q-learning, while user 3 employs sing Q-learning. In the third stage, as all users are not influenced by co-channel interference, each user adopts sing Q-learning method. In the fourth stage, user 2 and user 3 cooperative via multi-agent Q-learning, whereas user 1 chooses its transmission channel independently via single Q-learning.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2018 doi:10.20944/preprints201809.0227.v1

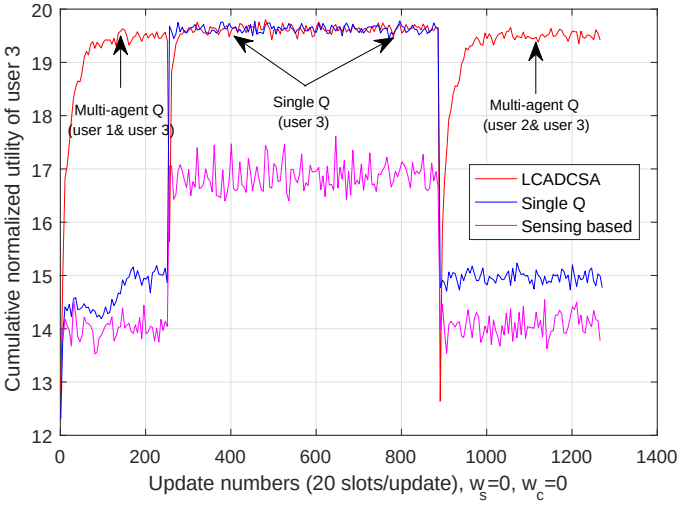
Peer-reviewed version available at Appl. Sci. 2018, 8, 1911; doi:10.3390/app8101911



(a) Cumulative normalized utility of user 1.



(b) Cumulative normalized utility of user 2.



(c) Cumulative normalized utility of user 3.

Figure 7. Coverage performance consider number of UAVs and weighting coefficient

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2018 doi:10.20944/preprints201809.0227.v1

Peer-reviewed version available at Appl. Sci. 2018, 8, 1911; doi:10.3390/app8101911

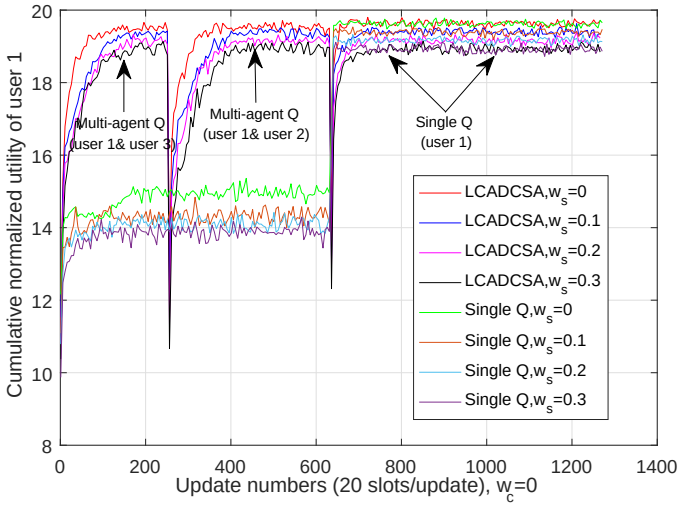


Figure 8. The cumulative normalized utility of user 1 with different channel switching cost.

Furthermore, as can be seen from Fig. 7(a), Fig. 7(b) and Fig. 7(c), either single-Q learning or multi-agent Q-learning can realize high cumulative utilities within 50 update numbers (about 1.18s). For the purpose of evaluate the effective of the LCADCSA algorithm, it is compared to the sensing based algorithm and multi-user single Q-learning. In sensing based algorithm, users select channels that are not jammed by the jammer after sensing current channel states, and in multi-user single Q-learning, each user adopts single Q-learning independently to avoid the jamming channel while ignoring the existence of mutual interference. Simulation results shows that users can also achieve higher cumulative normalized utilities U_{cum} using LCADCSA algorithm when there exists mutual interference between users. The reason is: In the proposed algorithm, users can learn the action of jammer, and can also adjust their channel selection strategy jointly according to their location and interference information. Thus, the users can avoid malicious jamming and co-channel interference simultaneously.

In addition, In Fig. 8 and Fig. 9, we make comparisons of user 1's cumulative normalized utilities with different channel switching cost and cooperation cost. As shown in Fig. 8, with the increase of channel switching cost, user 1's cumulative normalized utility decrease either in LCADCSA algorithm or in multi-user single Q-learning algorithm. As shown in Fig. 9, with the increase of cooperation cost, user 1's cumulative normalized utility decreased a lot in multi-agent Q-learning stages, and the utility keep invariant in single Q-learning stages. The reason is that in multi-agent Q-learning, users cooperate with each other and share their joint Q tables as well as actions, whereas in single Q-learning, users only need to take actions and update their Q tables independently. Moreover, if the cooperation cost is too high, the influence of cooperation is greater than co-channel interference, which makes it unwise to cooperate to avoid co-channel interference.

5.3. Channel Selection Strategies of Users and the Jammer

As an example, Fig. 10 shows the time-frequency diagram after the LCADCSA algorithm converging in the first stage(4800ms-4850ms), where user 1 and user 3 are influenced by co-channel interference. The red square denotes the jamming channel, the blue square, black square and yellow square represent the channel selection of user 1, user 2 and user 3 respectively. The mixed color square means that either more than two users choose the same channel, or users and the jammer choose the same channel in one certain slot. During the first stage, users are under the threat of malicious jammer and co-channel interference. Thus, user 1 and user 3 adopt multi-agent Q-learning and take joint channel selection, whereas user 3 employs single-agent Q-learning as it is not influenced by co-channel

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2018 doi:10.20944/preprints201809.0227.v1

Peer-reviewed version available at Appl. Sci. 2018, 8, 1911; doi:10.3390/app8101911

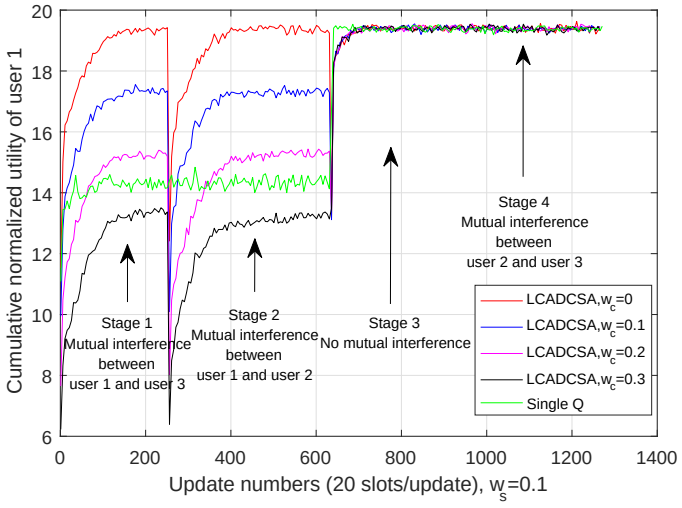


Figure 9. The cumulative normalized utility of user 1 with different cooperation cost.

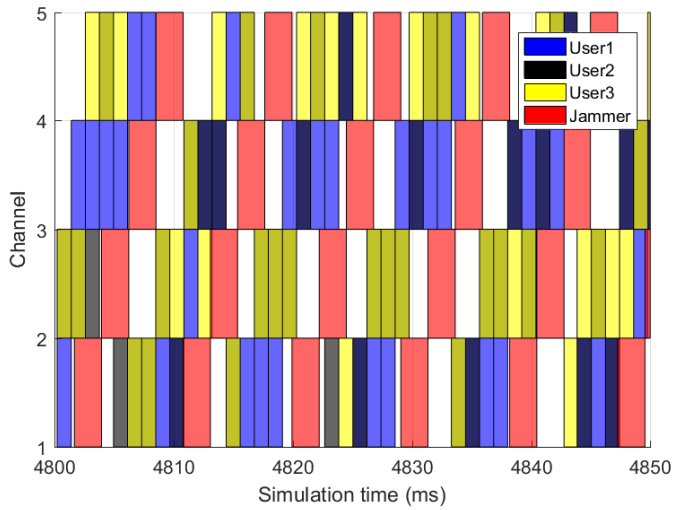


Figure 10. The time-frequency diagram after LCADCSA converging in the first stage.

interference. As is shown in Fig. 10, the users' channel selections avoid the vast majority of jamming channels. Moreover, user 1 and user 3 avoid being influenced by co-channel interference as they select different channel in each time slot. In addition, although there exist some overlapping areas between user 2's channels and other users' channels, the communication of user 2 would not be influenced by co-channel interference as its received co-interference signal energy is lower than threshold. In a word, the time-frequency diagram shows that the proposed LCADCSA algorithm is effective.

6. Conclusion

This paper investigated the anti-jamming channel selection problem in UAV communication networks. Via constructing an cooperative anti-jamming mechanism, users can realize information sharing and then take actions according to the interference level in the network. The channel switching cost and cooperation cost, which had a great impact on the users' utilites, were introduced. For the case where users were not influenced by co-channel interference, a Markov decision process was formulated for independent anti-jamming channel selection, and a single Q-learning method was designed to obtain the independent anti-jamming channel selection strategies. For the case where

users were influenced by co-channel interference, a Markov game was formulated for the interactions between users and malicious jammer, and a multi-agent Q-learning method was adopted to obtain the joint anti-jamming channel selection strategies. Simulation results depicted that the proposed LCADCSA algorithm can avoid malicious jamming and co-channel interference effectively.

Author Contributions: Yifan Xu, Guochun Ren and Jin Chen conceived and designed the model; Yifan Xu, Luliang Jia and Lijun Kong performed the theoretical analysis and simulation; Xiaobo Zhang analyzed the simulation result; Yifan Xu wrote the paper; Guochun Ren, Jin Chen, Xiaobo Zhang, Luliang Jia and Lijun Kong also provided some valuable suggestions for this paper.

Funding: This work was supported by the National Natural Science Foundation of China under Grant No. 61771488, in part by the Natural Science Foundation for Distinguished Young Scholars of Jiangsu Province under Grant No. BK20160034, and in part by the Open Research Foundation of Science and Technology on Communication Networks Laboratory. (*Corresponding author: Guochun Ren*)

Conflicts of Interest: The authors declare no conflict of interest.

References

1. L. Gupta, R. Jain, and G. Vaszkun, "Survey of important issues in UAV communication networks," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1123-1152, Second Quarter. 2016.
2. D. Liu, Y. Xu, J. Wang, *et al.*, "Self-organizing relay selection in UAV communication networks: A matching game perspective," *IEEE Wireless Commun.*, to appear. (available: https://arxiv.org/abs/1805.09257)
3. Y. Zou, J. Zhu, X. Wang, *et al.*, "A survey on wireless security: Technical challenges, recent advances, and future trends," *Proc. IEEE*, vol. 104, no. 9, pp. 1727-1765, Sept. 2016.
4. C. Chen, M. Song, C. Xin, *et al.*, "A game-theoretical anti-jamming scheme for cognitive radio networks," *IEEE Network*, vol. 27, no. 3, pp. 22-27, May. 2013.
5. L. Zhang, Z. Guan, and T. Melodia, "United against the enemy: Anti-jamming based on cross-layer cooperation in wireless networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 8, pp. 5733-5747, Aug. 2016.
6. H. Zhu, C. Fang, Y. Liu, *et al.*, "You can jam but you cannot hide: Defending against jamming attacks for geo-location database driven spectrum sharing," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 10, pp. 2723-2737, Oct. 2016.
7. Z. Han, *et al.*, *Game theory in wireless and communication networks*. Cambridge University Press, 2012.
8. Y. Xu, J. Wang, Q. Wu, *et al.*, "A game-theoretic perspective on self-organizing optimization for cognitive small cells," *IEEE Commun. Mag.*, vol. 53, no. 7, pp. 100-108, Jul. 2015.
9. Q. Wu, Y. Xu, J. Wang, *et al.*, "Distributed channel selection in time-varying radio environment: Interference mitigation game with uncoupled stochastic learning," *IEEE Trans. Veh. Technol.*, vol. 62, no. 9, pp. 4524-4538, Nov. 2013.
10. Y. Sun, J. Wang, F. Sun, *et al.*, "Energy-aware joint user scheduling and power control for two-tier femtocell networks: A hierarchical game approach," *IEEE Syst. J.*, vol. PP, no. 99, pp. 1-12, 2017.
11. L. Jia, Y. Xu, Y. Sun, *et al.*, "Stackelberg game approaches for anti-jamming defence in wireless networks," *IEEE Wireless Commun.*, to appear. (available: https://arxiv.org/abs/1805.12308)
12. D. Yang, G. Xue, J. Zhang, *et al.*, Coping with a smart jammer in wireless networks: A stackelberg game approach, *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 4038-4047, Aug. 2013.
13. Y. Li, L. Xiao, J. Liu, *et al.*, "Power control stackelberg game in cooperative anti-jamming communications," *in Proc. GAMENETS 2014*, pp. 1-6.
14. L. Xiao, T. Chen, J. Liu, *et al.*, "Anti-jamming transmission stackelberg game with observation errors," *IEEE Commun. Lett.*, vol. 19, no. 6, pp. 949-952, June. 2015.
15. F. Yao, L. Jia, Y. Sun, *et al.*, "A hierarchical learning approach to anti-jamming channel selection strategies," *Wireless Netw.*, doi: 10.1007/s11276-017-1551-9.
16. L. Xiao, X. Lu, D. Xu, *et al.*, "UAV relay in VANETs against smart jamming with reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4087-4097, Jan. 2018.
17. J. Chen, Y. Xu, Q. Wu, *et al.*, " Distributed channel selection for multicluster FANET based on real-time trajectory: a Potential game approach," submitted to *IEEE Trans. Veh. Technol.*, 2018.
18. Q. Hu, and W. Yue, *Markov decision processes with their applications.* Springer, US, 2007.

19. M. L. Puterman, *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons, 2009.
20. C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3-4, pp. 279-292, 1992.
21. X. Liu, Y. Xu, L. Jia, *et al.*, "Anti-jamming communications using spectrum waterfall: a deep reinforcement learning approach", *IEEE Commun. Lett.*, vol. 22, no. 5, pp. 998-1001, Mar. 2018.
22. L. Busoniu, R. Babuska, and B. D. Schutter, "A comprehensive survey of multi-agent reinforcement learning," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 38, no.2, pp.156-172, Feb. 2008.
23. Y. Xu, J. Wang, Q. Wu, *et al.*, " Dynamic spectrum access in time-varying environment: Distributed learning beyond expectation optimization," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5305-5318, Dec. 2017.
24. M. A. Aref, S. K. Jayaweera, and S. Machuzak, "Multi-agent reinforcement learning based cognitive anti-jamming," *in Proc. IEEE WCNC 2017*, pp. 1-6.
25. M. A. Aref, and S. K. Jayaweera, "A novel cognitive anti-jamming stochastic game," *in Proc. Cognitive Communications for Aerospace Applications Workshop 2017*, pp.1-4.
26. M. A. Aref, and S. K. Jayaweera, "A cognitive anti-jamming and interference-avoidance stochastic game," in Proc. *IEEE ICCI*CC 2017*, pp. 520-527.
27. F. Yao and L. Jia, "A Collaborative Multi-agent Reinforcement Learning Anti-jamming Algorithm in Wireless Networks," submitted to *IEEE Wireless Commun. Lett*, Aug. 2018.
28. L. Kong, Y. Xu, Y. Zhang, *et al.*, "A Reinforcement Learning Approach for Dynamic Spectrum Anti-jamming in Fading Environment," *in Proc. IEEE ICCT 2018*, pp. 1-7.
29. Y. Xu, G. Ren, J. Chen, *et al.*, "A one-leader multi-follower Bayesian-Stackelberg game for anti-jamming transmission in UAV communication networks," *IEEE Access*, vol. 6, pp. 21697-21709, Apr. 2018.
30. F. Slimeni, Z. Chtourou, B. Scheers, *et al.*, "Cooperative Q-learning based channel selection for cognitive radio networks," *Wireless Netw.*, DOI:10.1007/s11276-018-1737-9, to be published.