applied sciences

MDPI

Article

Interference-Aware Cooperative Anti-Jamming Distributed Channel Selection in UAV Communication Networks

Yifan Xu, Guochun Ren *, Jin Chen, Xiaobo Zhang, Luliang Jia and Lijun Kong

College of Communications Engineering, Army Engineering University of PLA, Nanjing 210000, China; yifanxu1995@163.com (Y.X.); chenjin99@263.net (J.C.); xb_zhang2008@126.com (X.Z.); jiallts@163.com (L.J.); konglijun2018@163.com (L.K.)

* Correspondence: guochunren@yeah.net

Received: 11 September 2018; Accepted: 11 October 2018; Published: 14 October 2018

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, we design an interference-aware cooperative anti-jamming scheme for the purpose of maximizing users' utilities. Moreover, the channel switching cost and cooperation cost are introduced, which have a great impact on users' utilities. Users in the UAV group sense the co-channel interference signal energy to judge whether they are influenced by co-channel interference. When the received co-channel interference signal energy is lower than the co-channel interference threshold, users conduct channel selection strategies independently. Otherwise, users cooperate with each other and take joint actions with a cooperative anti-jamming pattern under the impact of co-channel interference. Aiming at the independent anti-jamming channel selection problem under no co-channel 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 Q-learning with a "cooperation-decision-feedback-adjustment" idea, we design an interference-aware cooperative anti-jamming distributed channel selection algorithm (ICADCSA) to obtain the optimal anti-jamming channel strategies for users in a distributed way. In addition, a discussion on the quick decision for UAVs is conducted. Finally, simulation results show that the proposed algorithm converges to a stable solution with which the UAV group can avoid malicious jamming, as well as co-channel interference effectively and can realize a quick decision in high mobility UAV communication networks.

Keywords: interference-aware; cooperative anti-jamming; Markov decision process; Markov game; Q-learning

1. Introduction

Unmanned aerial vehicle (UAV) communication networks, as a kind of newly-developing wireless communication network, 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 cases, the existence of a malicious jammer will have a great impact on communications among UAVs, which motivates us to investigate the anti-jamming approach in UAV communication networks.

In the traditional anti-jamming research field, various techniques have been adopted, i.e., power control, uncoordinated frequency hopping (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 amount of spectrum resources, and they are not able to work well under the dynamic spectrum environment [4–6]. As an example of the dynamic spectrum environment, the primary-secondary access scenario in cognitive radio networks has been investigated in [7]. In detail, the primary-secondary access scenario means the spectrum is owned by primary users and can be used by secondary users when it is idle. The spectrum holes, which can be accessed by secondary users, are time-varying, and the environment with spectrum holes is a kind of dynamic spectrum environment. Moreover, in [8], the authors investigated the distributed channel selection problem in an opportunistic spectrum access (OSA) system, in which the channel states are also time-varying. In [9], the authors studied the problem of distributed channel selection for interference mitigation in a time-varying environment, where the channels undergo block fading. This is also a kind of dynamic spectrum environment.

In addition, game theory [9–12], as a strong theoretical tool, is suitable to model the anti-jamming competitive scenario. Specifically, the Stackelberg game approach [12], as a kind of hierarchical game, has been widely used in the anti-jamming field. For example, in [13], the authors summarized the application of the Stackelberg game in anti-jamming dense networks and introduced several classical anti-jamming scenarios and system models. Moreover, an outlook on the application of the anti-jamming Stackelberg game was also made. In [14–16], the 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. Considering the channel selection problem under a malicious jamming environment, the authors proposed a hierarchical anti-jamming channel selection scheme using a Stackelberg game framework [17]. However, most existing studies under the Stackelberg game framework formulated the interactions between the user-side and jammer-side, which brought large deviation in information acquisition. In a word, studies focusing on anti-jamming channel selection under a dynamic jamming environment are of great importance. As an example of a dynamic jamming environment, in [14], the authors investigated the anti-jamming power strategies under the threat of a smart jammer. The jammer can adjust its jamming power adaptively according to the user's transmission power. As the jamming power strategies are not fixed, we can call it the "dynamic jamming power environment". Moreover, in [17], the authors investigated the anti-jamming channel selection problem, in which the channel selection strategies can be adjusted adaptively according to the users' channel strategies, and we can call it the "dynamic jamming channel environment". If the strategies of the jammer are not fixed, we view it as a kind of dynamic jamming environment.

In fact, the dynamic feature of the channel state, which means the channel states are time-varying, brings some challenges to anti-jamming channel selection. For example, the wireless canonical network, in which the channel states are time-varying and there is no information exchange among users, has been investigated in [18]. In addition, the mobility of UAVs also influences the receiver's signal energy, causing the decline of communication quality [19]. In [20], the authors investigated the multi-stage spectrum access problem for a flying ad-hoc network (FANET). The Markov decision process (MDP) [21,22], as a decision framework under the dynamic channel environment, has been adopted widely. For the purpose of solving the MDP problem, Q-learning [23] methods are usually employed using a "decision-feedback-adjustment" structure to obtain the optimal strategy. For instance, in [21–23], the authors investigated the Markov decision process (MDP), Q-learning and the applications of MDP and Q-learning, but did not study the anti-jamming scenarios. In [4], the author formulated the anti-jamming decision problem as an MDP and obtained the best anti-jamming scheme via Q-learning. However, this work did not take the co-channel interference into consideration. In [24], the author dealt with the channel selection problem for cognitive networks via cooperative Q-learning, and in [25], a dynamic spectrum anti-jamming method in a fading environment was investigated. Furthermore, in [26], a deep Q network was built, and the anti-jamming channel selection problem was solved using a deep Q-learning method. These three works did not consider the multi-user case, which also ignored the existence of co-channel interference. In view of the multi-user scenarios in the anti-jamming field, the MDP problem has been extended to the Markov game [27], and several learning algorithms were designed for multi-user scenarios. Nevertheless, [27] did not concern the UAV communication networks where malicious jamming and co-channel interference existed simultaneously. In [18], a multi-agent learning algorithm was proposed to obtain a stable solution for the dynamic spectrum access problem, while it did not concern the attack of a malicious jammer. In [28–30], some multi-user Q-learning methods were adopted, where users took actions independently. However, in those methods mentioned in [28–30], users' states were influenced by each other, which led to unsteady learning environments and poor decision effects. In [31], the author studied the collaborative multi-agent reinforcement learning anti-jamming approach, with no concern about the feature of mobility in UAV communication networks.

Taking an overall consideration of the challenges and inspirations brought by the above studies, in this paper, we mainly focus on the anti-jamming channel selection problem under the dynamic environment, where the channel state and UAVs' locations are time-varying. We also investigate two different threats to UAVs: malicious jamming and co-channel interference. The malicious jamming refers to the jamming attack launched by enemies or the opponent, and if the jamming signal occupies a channel, users are not able to communicate in this channel anymore. The co-channel interference refers to the interference signal caused by other users when they are communicating in the same channel. If there exists co-channel interference, users can still communicate via a CSMA or TDMA pattern. Malicious jamming has the purpose of destroying or degrading the communications of users subjectively, while co-channel interference does not have subjectivity.

Motivated by [14], we define the user's utility as the trade-off between its throughput and cost. Considering the influence of co-channel interference and malicious jamming simultaneously, we firstly design the users' throughput. While considering the cost of channel switching and the cooperation among users, which have a great impact on the users' utilities, we introduce the channel switching cost unit and cooperation cost unit. Moreover, in our paper, 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 available channel state and co-channel interference signal energy and make a judgment whether they are influenced by co-channel interference. For the case where users are not influenced by co-channel interference, an MDP is formulated to model the anti-jamming problem for the users, and an 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 obtain joint channel selections. Via using the learning experience, users can realize a quick anti-jamming channel selection decision in 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 joint Q table sharing. Considering the influence of co-channel interference, an MDP and a Markov game are formulated, respectively.
- An interference-aware cooperative anti-jamming distributed channel selection algorithm (ICADCSA) is designed for the anti-jamming selection problem. Without the influence of co-channel interference, an 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 ICADCSA, which can avoid the malicious jamming and co-channel interference effectively. Moreover, the influence of channel switching cost and cooperation cost are investigated.

Comparing this paper to our previous works [25,31], which studied the anti-jamming channel selection in wireless communication networks, and to our previous work [32], we summarize the main differences as: (i) The work in [31] investigated the multi-agent learning method for anti-jamming problem, and [25] considered the single reinforcement learning in a fading environment. However, neither of them took the mobility of UAVs into consideration; whereas in our paper, the anti-jamming channel selection approach in UAV communication networks is investigated, taking the mobility of

UAVs into consideration, which causes the variation of the co-channel interference level. Moreover, channel switching cost and cooperation cost are introduced, which influence the users' utilities. (ii) In [32], we focused on the 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 Q-learning is derived, which obtains strategies by interacting with the environment.

The rest of this paper is as follows. In Section 2, the system model and problem formulation are investigated. In Section 3, the interference-aware cooperative anti-jamming mechanism in the UAV group is designed. In Section 4, the proposed interference-aware cooperative anti-jamming distributed channel selection algorithm (ICADCAS) is shown, and the complexity is analyzed. Moreover, a discussion of the quick decision for UAVs is also presented. In Section 5, simulations and discussions are conducted. In the end, we make a conclusion and investigate future work in Section 6.

2. System Model and Problem Formulation

The system model is shown in Figure 1. We construct the UAV communication network, which is a kind of dynamic UAV canonical network. Moreover, we assume that there are N users and one malicious jammer in the system scenario. A user in the UAV canonical networks is a collection of multiple UAVs with intra-communications, and there is a heading UAV managing the whole cluster, transmitting command and control information or sharing some important messages. For simplicity, we assume that UAVs in the same cluster keep relatively static and denote one UAV cluster as one user. Examples of users in traditional canonical networks are given in [18,33]. Under these assumptions, we think when different users are close to each other, the existence of co-channel interference is a matter that influences the intra-communications of users. Moreover, in our system scenario, UAVs are under the threat of a malicious jammer and co-channel interference simultaneously. The locations of users in the UAV group are time-varying, and users can cooperate with each other via information exchange. The users' set is denoted as $\mathcal{N} = \{1, ..., n, ..., N\}$, and the available channel set for the user is $\mathcal{M} = \{1, ..., m, ..., M\}$.

We consider two different cases of users' transmissions: (i) When users are close to each other, and transmitting in the same channel, high received signal energy from other users causes them to be influenced by co-channel interference. (ii) When users are far away from each other, the received signal energy from other users is somehow low; thus, users are not influenced by co-channel interference. The mutual interference threshold τ_0 is used to measure the influence of co-channel interference, that is: when the received co-channel interference signal energy is lower than τ_0, the UAV communication network is not influenced by co-channel interference, and vice versa. The co-channel interference threshold τ_0 is defined as the received signal energy threshold, which is transmitted from other users:

$$\tau_0 = p_0 d_{th}^{-\alpha}, \tag{1}$$

where p_0 is the transmission power of users, d_{th} is the co-channel interference distance threshold among users and α is the path-loss factor.

We assume that channel strategy $a_n(t)$ means user n chooses channel $c_n, c_n \in \mathcal{M}$, to transmit in time slot t, $a_{-n}(t)$ is the channel strategy combination of all users except user n and $a_j(t)$ is the jamming channel. Note that in [31], the authors considered that if there existed co-channel interference in one channel, users were not able to transmit in this channel; while in our paper, we assume that if there exists co-channel interference, users can still transmit to their receivers via accessing the channel with a relatively fair pattern, which is more realistic. Then, the throughput of user n in slot t is expressed as:

$$Tr_n\left(a_n(t), a_{-n}(t), a_j(t)\right) = \left(1 - f\left(a_n(t), a_j(t)\right)\right) \frac{1}{I_n(c_n)} \log_2\left(1 + \frac{P_n d_n^{-\alpha}}{N_{c_n}}\right), \tag{2}$$

where d_n denotes the distance between the transmitter and the receiver of user n, P_n represents the user n's transmission power and N_{c_n} represents the channel noise power. Moreover, $I_n\left(c_n,t\right)$ is the congestion degree of channel c_n in the current slot, which is expressed as:

$$I_n\left(c_n,t\right)=\begin{cases}1+\sum\limits_{x\in\mathcal{N}/n} f\left(a_n\left(t\right),a_x\left(t\right)\right), & P_x d_{x,n}^{-\alpha}\left(t\right)\geq\tau_0,\\ \dots & \\ 1, & P_x d_{x,n}^{-\alpha}\left(t\right)<\tau_0.\end{cases} \tag{3}$$

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

$$f\left(x,y\right)=\begin{cases}1, & x=y,\\ 0, & x\neq y.\end{cases} \tag{4}$$

As is shown in Equation (2), $Tr_n\left(a_n\left(t\right),a_{-n}\left(t\right),a_j\left(t\right)\right)$ depicts the user n's throughput under the threat of malicious jamming and co-channel interference, and in Equation (3), the congestion degree $I_n\left(c_n,t\right)$ reflects the number of users who are influenced by co-channel interference.

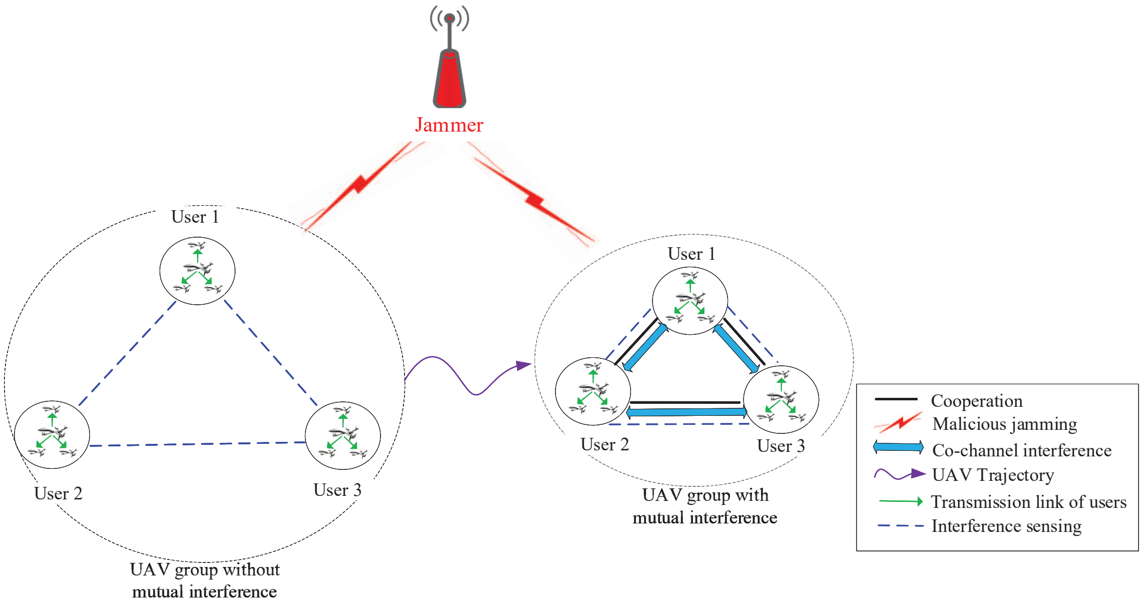


Figure 1. Cooperative anti-jamming UAV communication networks.

Different from [25,31], this work consider the channel switching of users and introduces the channel switching cost unit W_s to evaluate the performance loss. Moreover, if users cooperate with each other and take actions jointly, a cooperation cost unit W_c is also brought in. The introduction of channel switching cost makes it meaningful to investigate the variation of channel switching number, whereas the introduction of cooperation cost makes it interesting to decide whether to cooperate. Then, as a tradeoff between throughput and its cost, the utility of user n in one time slot is defined as:

$$u_n\left(a_n\left(t\right),a_{-n}\left(t\right),a_j\left(t\right)\right)=Tr_n\left(a_n\left(t\right),a_{-n}\left(t\right),a_j\left(t\right)\right)-W_s\delta_s\left(a_n\left(t\right),a_n\left(t-1\right)\right)-W_c\delta_c\left(I_n\left(c_n,t\right)\right), \tag{5}$$

where δ_s and δ_c are indicator functions for channel switching and cooperation and can be expressed as:

$$\begin{aligned}
&\delta_s\left(a_n(t), a_n(t-1)\right)=\begin{cases}1, & a_n(t)\neq a_n(t-1),\\ 0, & a_n(t)=a_n(t-1).\end{cases}\\
&\delta_c\left(I_n(c_n,t)\right)=\begin{cases}1, & I_n(c_n,t)>1,\\ 0, & otherwise.\end{cases}
\end{aligned} \tag{6}$$

$\delta_s = 1$ indicates that channel switching occurs at the beginning of the current slot, whereas $\delta_s = 0$ means that the user keeps its channel strategy. $\delta_c = 1$ indicates that the users are influenced by co-channel interference, cooperate with each other and take joint channel actions, whereas $\delta_c = 0$ means users choose channels independently. The optimization object of user n is:

$$a_n(t) = \underset{a_n(t)\in\mathcal{M}}{\arg\max}\, u_n\left(a_n(t), a_{-n}(t), a_j(t)\right). \tag{7}$$

Every user in the UAV group wants to employ an optimal anti-jamming channel selection strategy for the purpose of maximizing its utility. 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, the Markov game and Q-learning to investigate and solve the anti-jamming channel selection problem in UAV communication networks.

3. Interference-Aware Cooperative Anti-Jamming Mechanism in the UAV Group

In this part, the interference-aware cooperative anti-jamming mechanism in the UAV group is designed and analyzed. According to the wideband spectrum sensing and co-channel interference sensing of users, the process of the designed cooperative anti-jamming mechanism is shown in Figure 2.

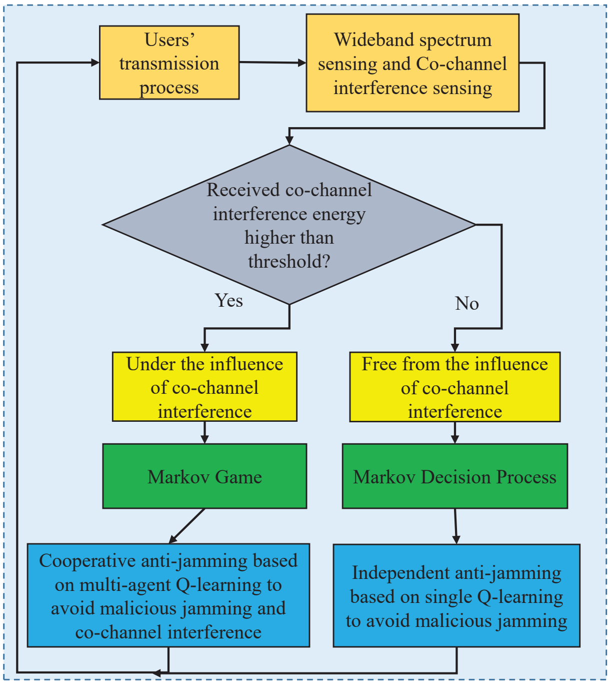

Users' transmission process
Wideband spectrum sensing and Co-channel interference sensing
Received co-channel interference energy higher than threshold?
Yes
No
Under the influence of co-channel interference
Free from the influence of co-channel interference
Markov Game
Markov Decision Process
Cooperative anti-jamming based on multi-agent Q-learning to avoid malicious jamming and co-channel interference
Independent anti-jamming based on single Q-learning to avoid malicious jamming

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

The details are shown as follows. After users accomplish transmissions with their receivers, they conduct the step of wideband spectrum sensing (WBSS) and co-channel interference sensing. In detail, wide spectrum sensing is used to obtain the current channel state, whereas the co-channel interference sensing is to judge whether other users are flying closer so that users are influenced by co-channel interference. If users are influenced by co-channel interference, a Markov game is formulated to model the cooperative anti-jamming problem, and every user has to avoid the jamming channel, as well as avoid the co-channel interference channel for the purpose of realizing higher utility. If users judge that they are not influenced by co-channel interference, a Markov decision process is able to formulate the coming optimization problem, and each user makes the anti-jamming decision independently via the single Q-learning approach.

3.1. Markov Decision Process

As was 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 (MDP), and each user's strategy is independent. Motivated by [21,22], we define MDP as:

Definition 1. *When users are free from the influence of co-channel 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.* $s_n(t) = (f_n(t), f_j(t)), s_n(t) \in S_n$ *is the environment state of user n at 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 n can obtain a reward with action* $a_n(t)$*.*
- *The state transition function* T_n *satisfies* $S_n \times A_n \to T_n$*. Moreover, it also meets the Markov property, shown as:*

$$\begin{aligned} & P\left[s_n(t+1)\,|s_n(t), a_n(t), \ldots, s_n(0), a_n(0)\right] \\ & = P\left[s_n(t+1)\,|s_n(t), a_n(t)\right], \;\; a_n(t) \in A_n, s_n(t) \in S_n. \end{aligned} \tag{8}$$

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

3.2. Single Q-Learning

Motivated by [22,24,25], we think the single Q-learning method is suitable for the case where the UAV group is not influenced by co-channel mutual interference. In the 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:

$$Q_{t+1}^n(s_n(t), a_n(t)) = (1-\lambda_n)\, Q_t^n(s_n(t), a_n(t)) + \lambda_n\left[r_n(t) + \gamma_n V_n({s'}_n)\right], \tag{9}$$

where λ_n is the learning rate of user n and γ_n represents the discount factor for Q table update. $r_n(t)$ is the immediate reward of user n while taking action $a_n(t)$ under environment $s_n(t)$, which also can be viewed as the normalized utility, which is:

$$r_n(t) = \left(1 - f\left(a_n(t), a_j(t)\right)\right) \frac{1}{I_n(c_n, t)} - w_s \delta_s\left(a_n(t), a_n(t-1)\right) - w_c \delta_c\left(I_n(c_n, t)\right), \tag{10}$$

where w_s and w_c are the normalized switching cost unit and normalized cooperation cost, respectively. The relationship between W_s and w_s, W_c and w_c is:

$$\begin{aligned} w_s &= \frac{W_s}{\log_2\left(1+\frac{P_n d_n^{-\alpha}}{N c_n}\right)}, \\ w_c &= \frac{W_c}{\log_2\left(1+\frac{P_n d_n^{-\alpha}}{N c_n}\right)}. \end{aligned} \tag{11}$$

$V_n\left(s'_n\right)$ is the value function of user n; in single Q-learning, $V_n\left(s'_n\right)$ can be expressed as:

$$V_n\left(s'_n\right) = \max\left\{Q_t^n\left(s'_n, a\right)\right\}. \tag{12}$$

The defined value function $V_n\left(s'_n\right)$ 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" and can converge to an optimal channel selection strategy [34]. As the channel switching cost is introduced, it makes sense for users to decrease their channel switching number.

3.3. 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 users' strategies. Thus, all users in the group take joint actions to fight against the malicious jammer and avoid co-channel interference as much as possible. Inspired by [27], we define the Markov game as:

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 the 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* $\sum\limits_{s \in \mathcal{S}} \mathcal{T}\left(s, \mathbf{a}, s'\right) = 1$. *Specifically,* $\mathbf{a}$ *is the joint channel selection strategy, and* 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 the 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 states in the UAV group, the anti-jamming channel selection problems are formulated as a Markov decision process and a Markov game, respectively. For the Markov decision process, the single Q-learning approach is used to obtain each user's optimal channel selection strategies; whereas for the Markov game, the multi-agent learning method is adopted for the purpose of acquiring the joint channel selection strategies for all users.

3.4. Multi-Agent Q-Learning

Inspired by [30,31], we aim at the case where UAVs are influenced by co-channel interference and design a cooperative anti-jamming channel selection algorithm based on multi-agent Q-learning. In the proposed multi-agent Q-learning, each user maintains and updates a Q table $\tilde{Q}^n$, which is based on joint action $\mathbf{a}(t)$. Similar to single Q-learning, the Q function updates using the following rule:

$$\widetilde{Q}_{t+1}^n\left(s, \mathbf{a}(t)\right) = \left(1 - \widetilde{\lambda}_n\right)\widetilde{Q}_t^n\left(s, \mathbf{a}(t)\right) + \widetilde{\lambda}_n\left[\widetilde{r}_n(t) + \widetilde{\gamma}_n\widetilde{V}_n\left(s'\right)\right], \tag{13}$$

where $\widetilde{\lambda}_n$ is user n's learning rate under joint action and $\widetilde{\gamma}_n$ is the discount factor correspondingly. $\widetilde{r}_n(t)$ denotes the user n's immediate reward when taking joint action $\mathbf{a}$ under state s. Moreover, $\widetilde{r}_n(t)$ represents the normalized utility under joint action, which can also be shown as:

$$\widetilde{r}_n(t) = \left(1 - f\left(a_n(t), a_j(t)\right)\right)\frac{1}{I_n\left(c_n, t\right)} - w_s\delta_s\left(a_n(t), a_n(t-1)\right) - w_c\delta_c\left(I_n\left(c_n, t\right)\right). \tag{14}$$

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

$$\widetilde{V}_n\left(s'\right) = \widetilde{Q}_t^n\left(s', \mathbf{a}^*\right), \tag{15}$$

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

$$\mathbf{a}^* = \arg\max \sum_{n=1}^{N} \widetilde{Q}_t^n\left(s', \mathbf{a}\right). \tag{16}$$

Without loss of generality, either in single Q-learning or in multi-agent Q-learning, the ε-greedy policy is introduced for the purpose of avoiding the local optimum. Moreover, it is obvious that indicator function $\delta_c = 0$ in single Q-learning and $\delta_c = 1$ in multi-agent Q-learning. As in single Q-learning, users take actions dependently, while in multi-agent Q-learning, users cooperate with each other to avoid co-channel interference. Note that in [34], the author investigated the convergence of multi-agent Q-learning and proved that the multi-agent Q-learning can converge to an optimal joint action, so we will not discuss the convergence analysis in our paper. Moreover, as the cooperation cost is introduced, the utilities of users may decrease in multi-agent Q-learning due to joint Q-table sharing.

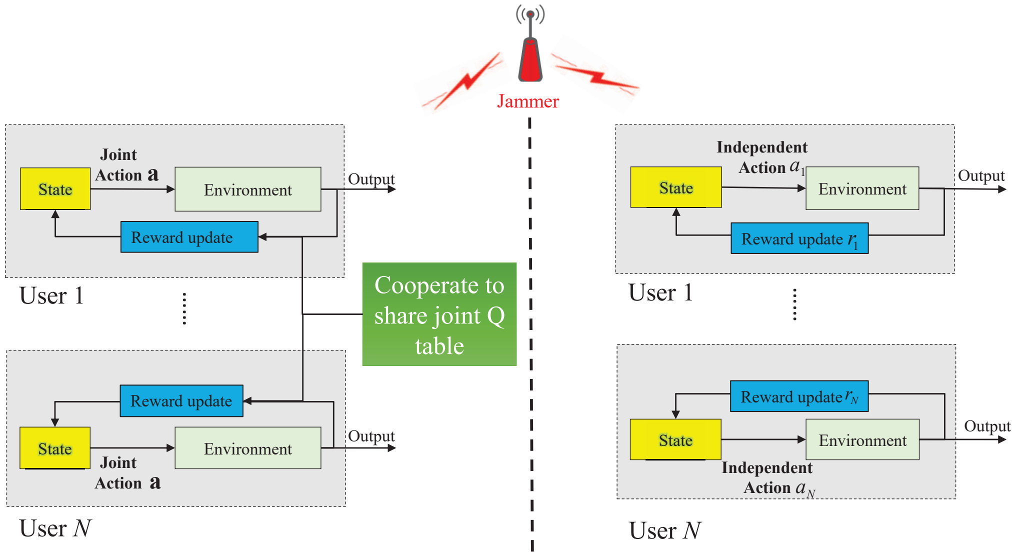
4. Interference-Aware Cooperative Anti-Jamming Distributed Channel Selection Algorithm

In this section, the interference-aware cooperative anti-jamming distributed channel selection algorithm is designed, and then, the complexity analysis is described.

4.1. Algorithm Description

As is shown in Figure 3, the anti-jamming distributed channel selection framework under different cases is depicted. In the left part of Figure 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 Q-learning. In this framework, users ought to cooperate with each other to share the joint Q table so that they can take joint actions. In the right part of Figure 3, the anti-jamming framework under the case where users are not influenced by co-channel interference is shown. In this framework, users adopt the single Q-learning mechanism, which uses an "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 co-channel interference signal energy to judge whether they are influenced by co-channel interference. The details of interference-aware cooperative anti-jamming distributed channel selection algorithm are shown in Algorithm 1.

In addition, the example of the transmission slot structure of users is shown in Figure 4. The heading UAV of each user firstly chooses a channel to transmit to its receivers. After that, the user obtains the feedback from their receivers to judge whether the transmission is successful. In the following part of the slot, a process of wideband spectrum sensing (WBSS) and co-channel interference sensing is conducted for the purpose of acquiring the currently available channel state and sensing the received co-channel interference signal energy. Later, users compare the signal energy to the threshold and make a judgment whether they are influenced by co-channel interference in the current time slot. If users are influenced by co-channel interference, they start cooperating to share the joint Q table with each other via a reliable control channel and take joint channel selection at the beginning of next slot to avoid co-channel interference, if users are not influenced by co-channel interference, they do not share their Q tables and take actions independently. Then, each user starts multi-agent Q-learning or single Q-learning to update their Q tables. During the last process of the slot, users broadcast the following transmission channel to their receivers.



Anti-jamming channel selection under the influence of co-channel interference.

Anti-jamming channel selection without the influence of co-channel interference.

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

Algorithm 1: Interference-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 Iterations:
 Each user senses and observes the current environment state and then makes a judgment
 about the co-channel interference according to co-channel interference sensing.
 If users are under the influence of co-channel interference, go to multi-agent 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 $\mathbf{a}$ with probability $\varepsilon$.
     • Choosing the best joint action $\mathbf{a}^*$ according to Equation (16), 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 Equation (13).
 Otherwise, go to single Q-learning.
  Single Q-learning:
   (1) Each user observes and chooses one transmission channel, using the following rules:
     • Randomly choose an independent action $a^n$ with probability $\varepsilon$.
     • Choosing the best action $a^{n*}$ with probability $1 - \varepsilon$, which realizes the highest Q value
       in the 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 Equation (9).
End
 Jump out of the repeat process when the algorithm reaches the maximal iterations.
```

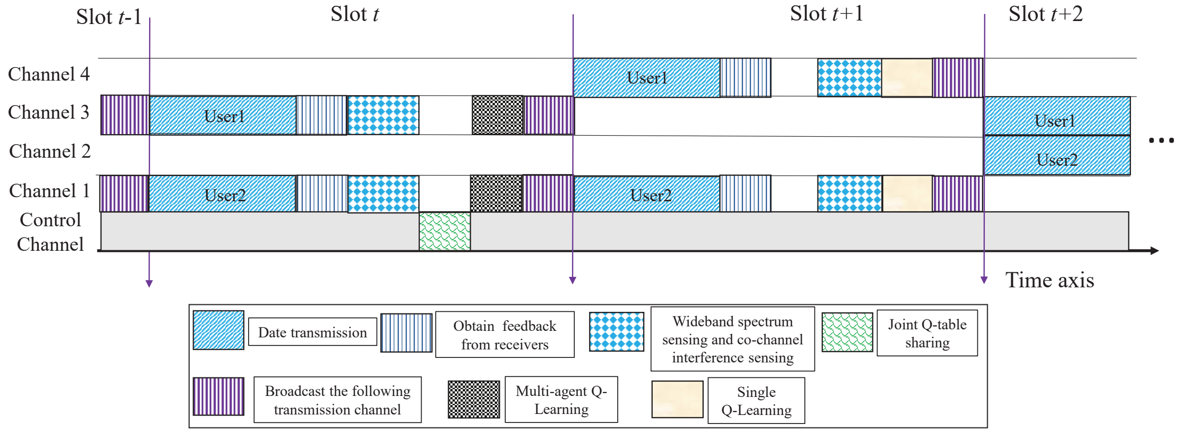


Figure 4. Anti-jamming transmission slot structure for UAVs.

4.2. Complexity Analysis

Inspired by [35], we analyze the complexity cost of the proposed algorithm in the following part. As is shown in Table 1, the complexity is divided into computational complexity, storage size and information sharing size. The details are shown as follows.

Firstly, we consider the computation complexity for users. In one slot, if multi-agent Q-learning is adopted, each user chooses the joint action according to Equation (16); the complexity is $O\left(C_1\right)$, where C_1 is a small constant; and each user updates its joint Q table according to Equation (13); the complexity is $O\left(C_2\right)$, where C_2 is a constant. Thus, the computation complexity for one user in multi-agent Q learning is $O\left(C_1\right)+O\left(C_2\right)$. If single Q-learning is adopted, each user chooses its action independently according to Equation (12); the complexity is $O\left(C_3\right)$; and then updates its single Q table according to Equation (9); the complexity is $O\left(C_4\right)$. C_3 and C_4 are small constants. Assuming there are N users in the UAV group, for user n, the time slot number of multi-agent Q-learning is Tm_n, and the time slot number for single-agent Q-learning is Ts_n, then the computation complexity for the UAV group in the whole time slot is:

$$C_{comp}=\sum_{n=1}^{N}\left\{Ts_n\left[O\left(C_1\right)+O\left(C_2\right)\right]+Tm_n\left[O\left(C_1\right)+O\left(C_2\right)\right]\right\}. \tag{17}$$

Secondly, for every user n, it has to store one single Q-table and $N-1$ joint Q tables. Then, the storage size of the UAV group is $N\left(O\left(S_1\right)+\left(N-1\right)O\left(S_2\right)\right)$, where $O\left(S_1\right)$ and $O\left(S_2\right)$ are the storage sizes of single Q-table and joint Q table, respectively.

Thirdly, information sharing is needed in multi-agent Q learning to share the joint Q-tables. Thus, the information sharing size of the UAV group is $\sum\limits_{n=1}^{N} Tm_n O\left(S_1\right)$. Hence, the total complexity of the UAV group is expressed as:

$$\begin{aligned}C_{total}=&\sum_{n=1}^{N}\left\{Ts_n\left[O\left(C_1\right)+O\left(C_2\right)\right]+Tm_n\left[O\left(C_1\right)+O\left(C_2\right)\right]\right\}+\\&\sum_{n=1}^{N}Tm_nO\left(S_1\right)+N\left(O\left(S_1\right)+\left(N-1\right)O\left(S_2\right)\right).\end{aligned} \tag{18}$$

In a word, we can find that the computational complexity of each user in each time slot is a small constant, and the storage size in each time slot is not too large. Moreover, the data of information sharing in each time slot is also a constant, which means information sharing is realistic for users. In the simulation setting part, the settings for the Q table size also indicate that the data size of information sharing and the storage size are acceptable. Thus, we think the total complexity is affordable for users in the UAV group.

Table 1. Complexity analysis of the UAV group.

Computation	$\sum\limits_{n=1}^{N} \{Ts_n [O(C_1) + O(C_2)] + Tm_n [O(C_1) + O(C_2)]\}$
Storage Size	$N(O(S_1) + (N-1) O(S_2))$
Information Sharing	$\sum\limits_{n=1}^{N} Tm_n O(S_1)$

4.3. A Discussion on the Quick Decision for UAVs

In the designed anti-jamming transmission slot structure, we have illustrated that users make wideband spectrum sensing and co-channel interference sensing in every time slot, which means that users can sense the states of malicious jamming and co-channel interference precisely. Then, users decide to take joint channel selections via multi-agent Q-learning or take independent channel selections via single Q-learning. After the proposed ICADCSA algorithm converges for one certain scenario, users can make quick decisions via previous learning experience and do not need to learn again when they encounter the same scenario next time. Thus, our proposed approach can realize quick anti-jamming channel selection in high mobility UAV communication networks. This discussion will be investigated and analyzed in the following simulation section.

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, was investigated. The available channel number for users to access was four. Moreover, we assumed that there existed one dedicated control channel that was reliable for users to share information. The jammer sent a sweeping jamming signal to the available channels, and the jamming signal stayed at one channel for about 2.28 ms. Referring to [24], the transmission time in each user's slot was set to be $T_{tr} = 0.98$ ms, and the time for obtaining feedback, sensing, information sharing, learning and broadcasting were in total $T_{of} + T_{se} + T_{is} + T_{le} + T_{br} = 0.2$ ms in each slot.

Other simulation settings are shown as follows. We assumed that the transmission power of each user was 0.1 W, and the initial locations of three users were (100 m, 100 m), (300 m, 800 m) and (150 m, 0 m), respectively. The trajectories of the UAVs are shown in Figure 5, and the flying time was divided into 10 epochs (Note that we designed the trajectories of users for the purpose of supplying a simulation environment, but this does not mean that the location information of users was known in advance. The proposed algorithm had good universality and could work well in high mobility UAV communication networks.). UAVs moved 150 m per epoch, and the duration time of each epoch was set to be 3 s. Furthermore, the pass-loss factor $\alpha = 2$, co-channel interference threshold distance was set to be 400 m; hence, the co-channel interference threshold was 6.25×10^{-7} W. We assumed that each user consisted of one heading UAV and one receiver, and the distance between the transmitter and receiver was 20 m. The channel noise power was assumed to be -110 dBm. The total simulation time was equal to the flying time (approximately 30 s). Motivated by [24], $\lambda_1 = ... = \lambda_n = \tilde{\lambda}_1 = ... = \tilde{\lambda}_n = 0.8$, $\gamma_1 = ... = \gamma_n = \tilde{\gamma}_1 = ... = \tilde{\gamma}_n = 0.6$, $\varepsilon = 0.1$.

In multi-agent Q-learning, we assumed that the joint Q table of each user was a matrix with 64 rows and 16 columns, whereas in single Q-learning, the independent Q table of each user was a matrix with 16 rows and four columns. Each user had to maintain two joint Q tables and one single Q table. When users cooperated with each other to share joint Q table, the data of Q table sharing in every slot were 1024 bytes. If $T_{is} = 0.1$ ms, then the required transmission rate for Q table sharing was 9.76 MB/s, which was realistic. Moreover, the normalized switching cost unit w_s and cooperation cost unit w_c varied from zero to 0.3, respectively. Each user's immediate reward can be viewed as the normalized utility, which can be calculated by Equations (10) and (14).

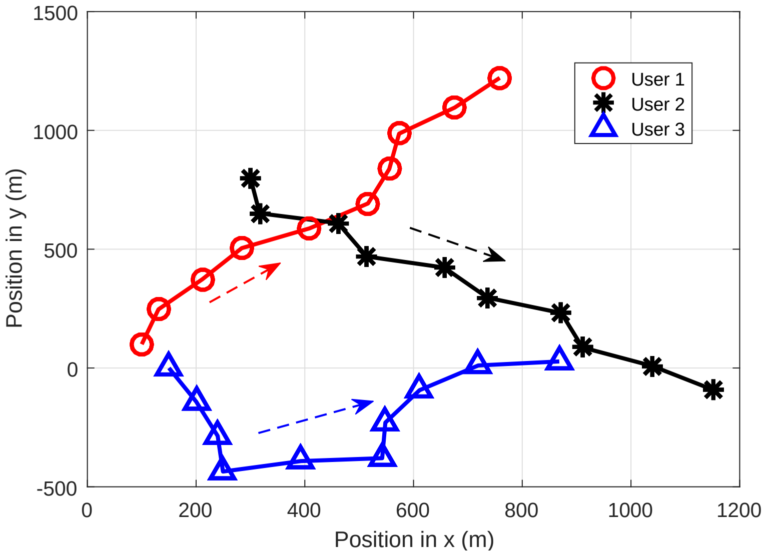


Figure 5. The trajectory setting for UAVs.

In addition, Figure 6 depicts the interference distance of UAVs. In detail, the flying process is divided into four stages. During flying time 0 s to 6 s (the first stage), the distance between User 1 and User 3 was less than 400 m, and they would be influenced by co-channel interference as the received signal energy was higher than threshold. During 6 s to 15 s (the second stage), User 1 and User 3 were influenced by co-channel interference. From 15 s to 21 s (the third stage), all users kept relatively far from each other, so there existed no co-channel interference, while from 21 s to 30 s (the fourth stage), User 2 and User 3 were influenced by co-channel interference.

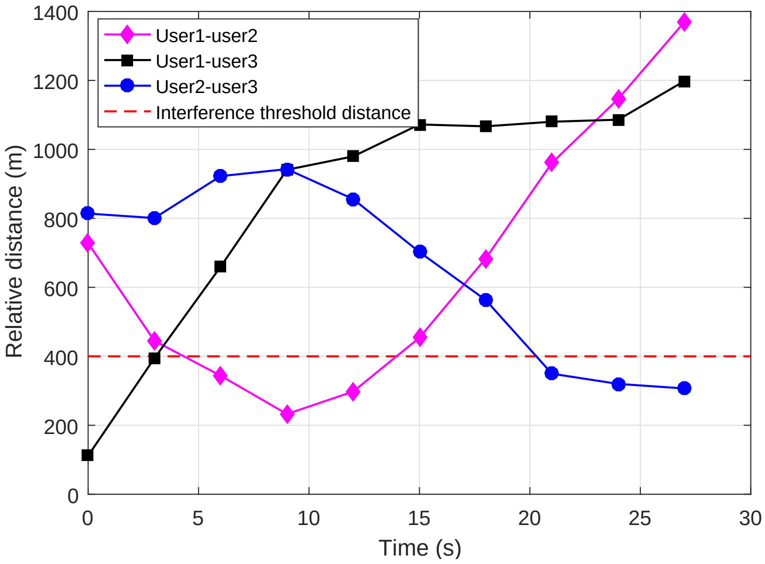


Figure 6. Relative distance variation setting.

5.2. Channel Selection Strategies of Users and the Jammer

As an example, Figure 7 shows the time-frequency diagram after the ICADCSA algorithm converging in the first stage (4800 to 4850 ms), where User 1 and User 3 were influenced by co-channel interference. In Figure 7, the Y-axis represents the channels that users and the jammer can choose (from Channel 1 to 4), while the X-axis represents the simulation time. For better description, we used u1, u2, u3 and J to denote the channel selection of User 1, User 2, User 3 and the jammer, respectively. The red square denotes the jamming channel, and the yellow square, black square and light pink square represent the channel selection of User 1, User 2 and User 3, respectively. The mixed color square illustrates 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 were under the threat of a malicious jammer and co-channel interference. Thus, User 1 and User 3 adopted multi-agent Q-learning and took joint channel selection, whereas User 3 employed single Q-learning as it was not influenced by co-channel interference. As is shown in Figure 7, the users' channel selections avoided the vast majority of jamming channels. Moreover, User 1 and User 3 avoided being influenced by co-channel interference, as they selected different channels in each time slot. In addition, although there existed 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 was lower than the threshold. In a word, the time-frequency diagram shows that the proposed ICADCSA algorithm was effective.

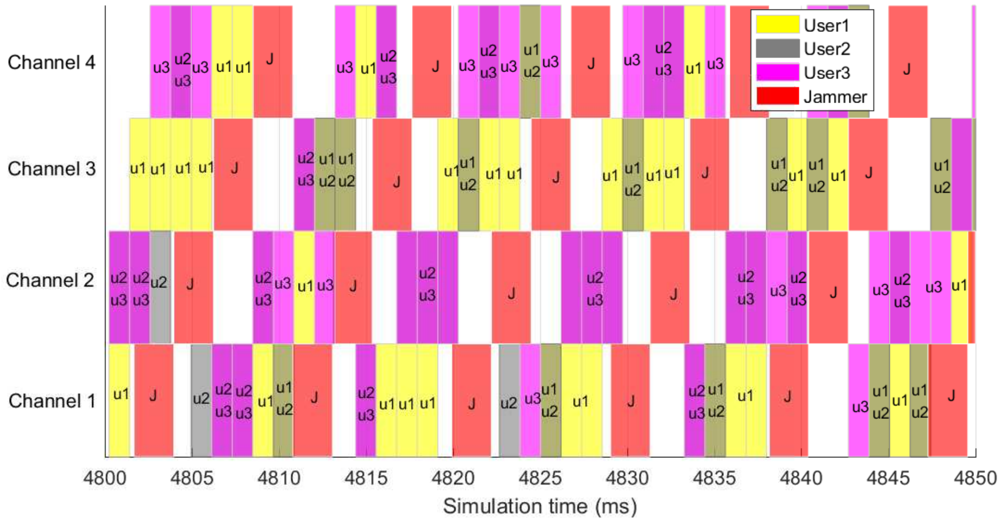


Figure 7. The time-frequency diagram after LCADCSA converging in the first stage.

5.3. Performance Analysis of Users

In this part, the user's performance analysis is mainly investigated. As is mentioned in Algorithm 1, when users were influenced by co-channel interference, the proposed interference-aware cooperative anti-jamming distributed channel selection algorithm (ICADCSA) was based on multi-agent Q-learning. When users are not influenced by co-channel interference, the proposed ICADCSA algorithm was based on single Q-learning. For better clarification, we use cumulative normalized utility U_{cum} to show the effective of ICADCSA approach, which is defined as follows:

$$U_{cum} = \sum_{i=1}^{PN} \left(1 - f\left(a_n\left(t\right), a_j\left(t\right)\right)\right) \frac{1}{I_n\left(c_n, t\right)} - w_s \delta_s\left(a_n\left(t\right), a_n\left(t-1\right)\right) - w_c \delta_c\left(I_n\left(c_n, t\right)\right), \tag{19}$$

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, while the time of each update is 23.6 ms.

5.3.1. Performance Analysis without Cost

We first investigate the cumulative normalized utility without cost. The cumulative normalized utilities of users are shown in Figure 8, Figures 9 and 10, respectively, where the channel switching cost and cooperation cost are set to be zero. As is shown in those three figures, the users' channel selection processes were divided into four stages: In the first stage, User 1 and User 3 cooperated with each other and adopted multi-agent Q-learning; User 2 employed single Q-learning. In the second stage, User 1 and User 2 cooperated with each other and adopted multi-agent Q-learning, while User 3 employed single Q-learning. In the third stage, as all users were not influenced by co-channel interference, each user adopted single Q-learning method. In the fourth stage, User 2 and User 3 cooperated via multi-agent Q-learning, whereas User 1 chose its transmission channel independently via single Q-learning.

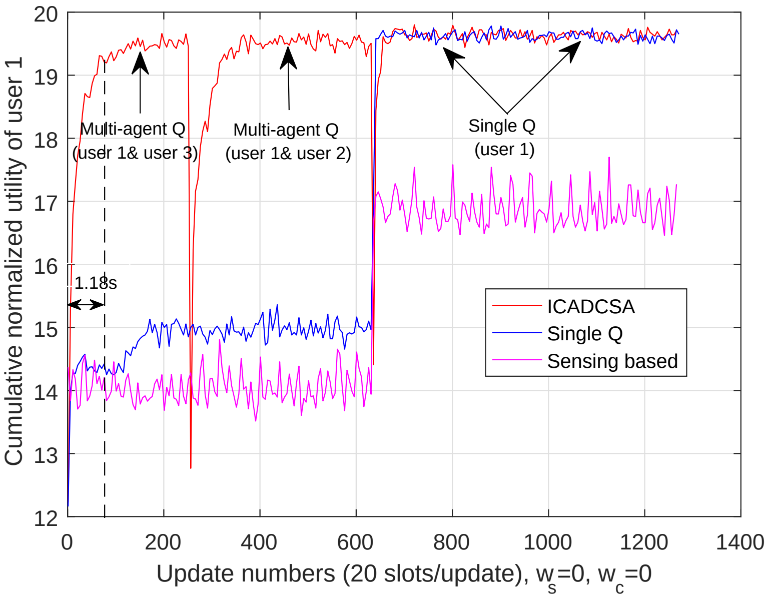


Figure 8. Cumulative normalized utility of User 1. ICADCSA, interference-aware cooperative anti-jamming distributed channel selection algorithm.

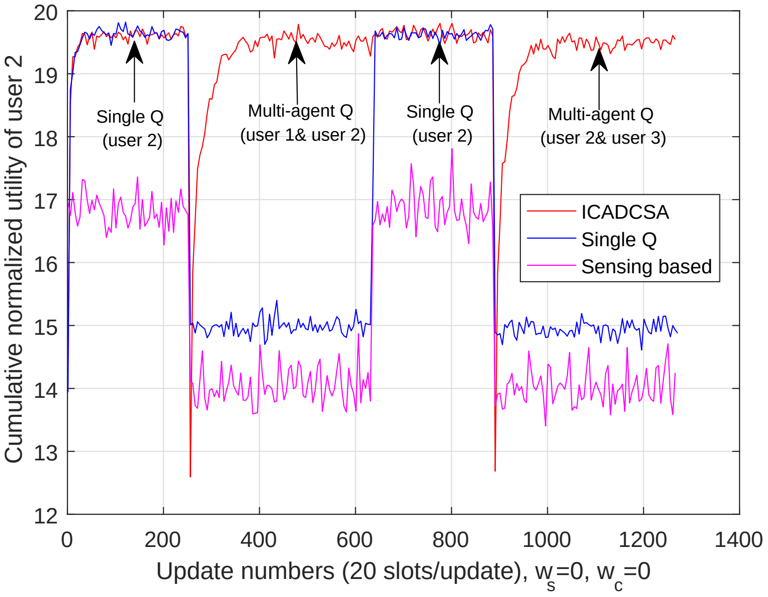


Figure 9. Cumulative normalized utility of User 2.

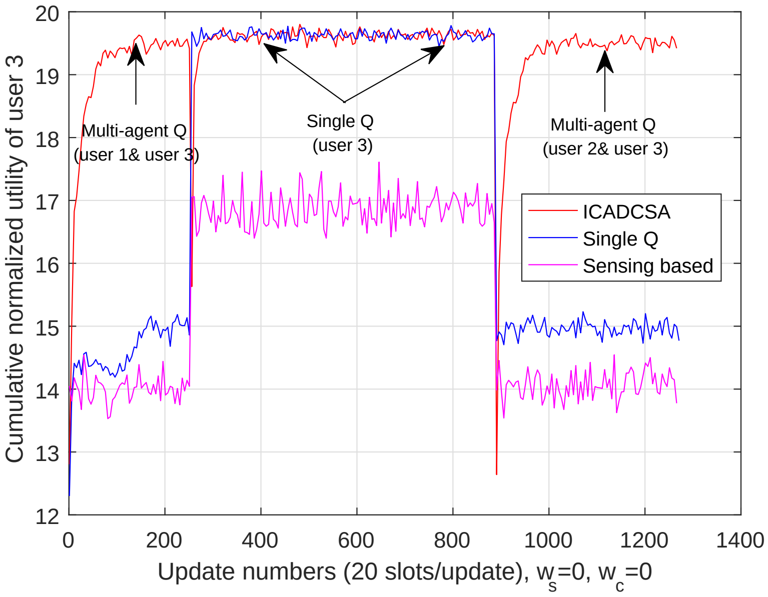


Figure 10. Cumulative normalized utility of User 3.

Furthermore, as can be seen from Figures 8–10, both single-Q learning and multi-agent Q-learning can realize high cumulative utilities within 50 update numbers (about 1.18 s). For the purpose of evaluating the effectiveness of the ICADCSA algorithm, it was compared to the sensing based algorithm and multi-user single Q-learning. In the sensing-based algorithm, users selected channels that were not jammed by the jammer after sensing the current channel states, and in multi-user single Q-learning, each user adopted single Q-learning independently to avoid the jamming channel, while ignoring the existence of mutual interference. Simulation results show that users can achieve higher cumulative normalized utilities U_{cum} using the ICADCSA algorithm when there exists mutual interference between users. The reason is that, in the proposed algorithm, users can learn the actions of jammer and can also adjust their channel selection strategies jointly according to their sensed interference level and shared information. Thus, the users can avoid malicious jamming and co-channel interference simultaneously.

5.3.2. Performance Analysis with Cost

In addition, in Figures 11 and 12, we make comparisons of User 1's cumulative normalized utilities with different channel switching cost and cooperation cost, and in Figure 13, we analyze the relationship between channel switching cost and channel switching number. As is shown in Figure 11, with the increase of channel switching cost, User 1's cumulative normalized utility decreased in the ICADCSA algorithm. Moreover, as is shown in Figure 12, with the increase of cooperation cost, user 1's cumulative normalized utility decreased greatly in the multi-agent Q-learning stages, and the utility kept invariant in the single Q-learning stages. The reason is that in multi-agent Q-learning, users cooperated with each other and shared their joint Q tables, as well as actions, whereas in single Q-learning, users only needed to take actions and update their Q tables independently. Moreover, if the cooperation cost was too high, the cost of cooperation was greater than the negative influence of co-channel interference, which made it unwise for users to cooperate to avoid co-channel interference. In addition, in Figure 13, we make a record of User 1's channel switching number under different channel switching cost units during 0 to 30 s. As an example, during 0 to 6 s, User 1 switched its transmission channel by about 2400 times with no channel switching cost and switched its transmission channel by about 1500 times with $w_s = 0.1$, which depicts that under the influence of channel switching cost, users were willing to decrease the number of channel switching as much as possible. However, if channel switching cost was higher, the channel switching number decreased with a smaller quantity.

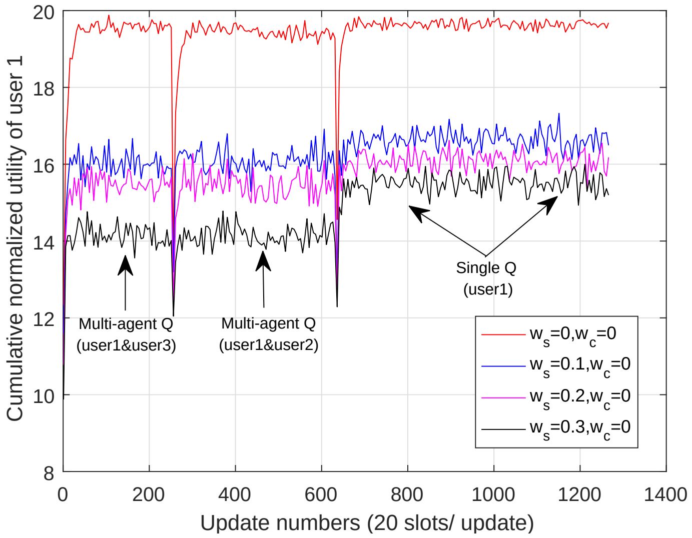


Figure 11. The cumulative normalized utility of User 1 with different channel switching costs.

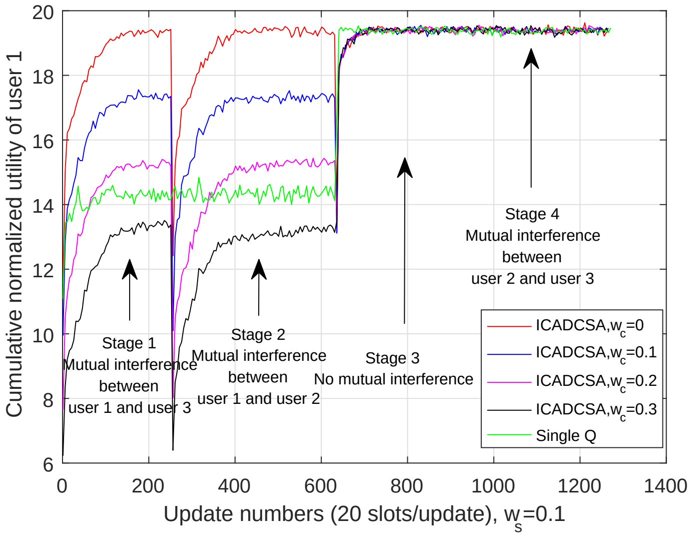


Figure 12. The cumulative normalized utility of User 1 with different cooperation costs.

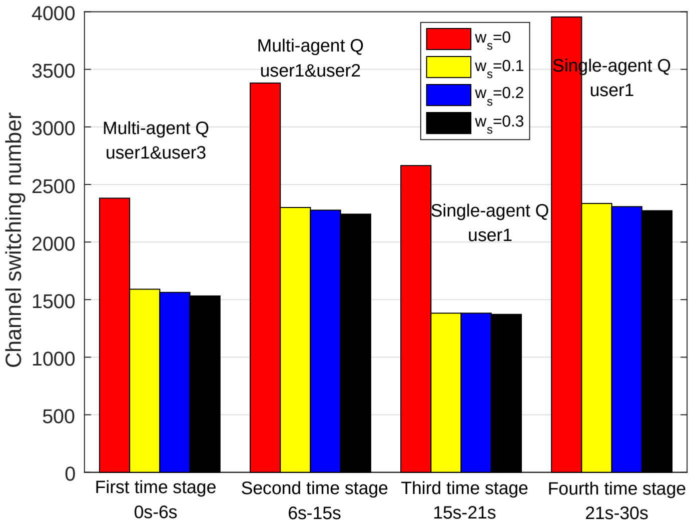


Figure 13. The channel switching number of User 1 with different channel switching costs.

5.3.3. Quick Decision under the Dynamic Environment

Furthermore, we notice that in Figure 9, User 2 adopted single Q-learning in the first time stage and adopted multi-agent Q-learning to cooperate with User 1 in the second time stage; while in the third stage, User 2 employed single Q-learning method again, and it could achieve high cumulative normalized utility via using the training experience of the first stage. The proposed ICADCSA algorithm only needed to converge once in one certain case, as each user maintained different Q-table for different cases. When users encountered the same case that had been trained before, they could take advantage of the previous experience and make quick decisions, which means the proposed algorithm worked well in UAV communication networks with high mobility.

6. Conclusions

This paper investigated the anti-jamming channel selection problem in UAV communication networks. Via constructing a cooperative anti-jamming mechanism, users can cooperate with each other 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' utilities, 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 the 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 ICADCSA algorithm can avoid malicious jamming and co-channel interference effectively.

In the future, we will try to investigate the cooperative anti-jamming distributed channel selection with more jamming patterns. Moreover, as the strategy space of Q-learning is limited, we will investigate the cooperative anti-jamming channel selection problem via multi-agent deep Q-learning to extend the strategy space.

Author Contributions: Y.X., G.R. and J.C. conceived of and designed the model. Y.X., L.J. and L.K. performed the theoretical analysis and simulation. X.Z. analyzed the simulation result. Y.X. wrote the paper. G.R., J.C., X.Z., L.J. and L.K. 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.

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

References

1. Gupta, L.; Jain, R.; Vaszkun, G. Survey of important issues in UAV communication networks. *IEEE Commun. Surv. Tutor.* **2016**, *18*, 1123–1152. [CrossRef]
2. Liu, D.; Xu, Y.; Wang, J.; Xu, Y.; Anpalagan, A.; Wu, Q.; Wang, H.; Shen, L. Self-organizing relay selection in UAV communication networks: A matching game perspective. *arXiv* **2018**, arXiv:1805.09257.
3. Zou, Y.; Zhu, J.; Wang, X.; Hanzo, L. A survey on wireless security: Technical challenges, recent advances, and future trends. *Proc. IEEE* **2016**, *104*, 1727–1765. [CrossRef]
4. Chen, C.; Song, M.; Xin, C.; Backens, J. A game-theoretical anti-jamming scheme for cognitive radio networks. *IEEE Netw.* **2013**, *27*, 22–27. [CrossRef]
5. Zhang, L.; Guan, Z.; Melodia, T. United against the enemy: Anti-jamming based on cross-layer cooperation in wireless networks. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 5733–5747. [CrossRef]
6. Zhu, H.; Fang, C.; Liu, Y.; Chen, C.; Li, M.; Shen, X.S. You can jam but you cannot hide: Defending against jamming attacks for geo-location database driven spectrum sharing. *IEEE J. Sel. Areas Commun.* **2016**, *34*, 2723–2737. [CrossRef]

7. Akyildiz, I.F.; Lee, W.Y.; Vuran, M.C.; Mohanty, S. Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey. *Comput. Netw.* **2016**, *50*, 2127–2159. [CrossRef]
8. Xu, Y.; Wang, J.; Wu, Q.; Anpalagan, A.; Yao, Y.D. Opportunistic spectrum access in unknown dynamic environment: A game-theoretic stochastic learning solution. *IEEE Trans. Wirel. Commun.* **2012**, *11*, 1380–1391. [CrossRef]
9. Wu, Q.; Xu, Y.; Wang, J.; Shen, L.; Zheng, J.; Anpalagan, A. Distributed channel selection in time-varying radio environment: Interference mitigation game with uncoupled stochastic learning. *IEEE Trans. Veh. Technol.* **2013**, *62*, 4524–4538.
10. Niyato, D.; Saad, W. *Game Theory in Wireless and Communication Networks*; Cambridge University Press: Cambridge, UK, 2012.
11. Xu, Y.; Wang, J.; Wu, Q.; Du, Z.; Shen, L.; Anpalagan, A. A game-theoretic perspective on self-organizing optimization for cognitive small cells. *IEEE Commun. Mag.* **2015**, *53*, 100–108. [CrossRef]
12. Sun, Y.; Wang, J.; Sun, F.; Zhang, J. Energy-aware joint user scheduling and power control for two-tier femtocell networks: A hierarchical game approach. *IEEE Syst. J.* **2017**, *12*, 2533–2544. [CrossRef]
13. Jia, L.; Xu, Y.; Sun, Y.; Feng, S.; Anpalagan, A. Stackelberg game approaches for anti-jamming defence in wireless networks. *arXiv* **2018**, arXiv:1805.12308.
14. Yang, D.; Xue, G.; Zhang, J.; Richa, A.; Fang, X. Coping with a smart jammer in wireless networks: A stackelberg game approach. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 4038–4047. [CrossRef]
15. Li, Y.; Xiao, L.; Liu, J.; Tang, Y. Power control stackelberg game in cooperative anti-jamming communications. In Proceedings of the 2014 5th International Conference on Game Theory for Networks, Beijing, China, 25–27 November 2014; pp. 1–6.
16. Xiao, L.; Chen, T.; Liu, J.; Dai, H. Anti-jamming transmission stackelberg game with observation errors. *IEEE Commun. Lett.* **2015**, *19*, 949–952. [CrossRef]
17. Yao, F.; Jia, L.; Sun, Y.; Xu, Y.; Feng, S.; Zhu, Y. A hierarchical learning approach to anti-jamming channel selection strategies. *Wirel. Netw.* **2017**, 1–13. [CrossRef]
18. Xu, Y.; Wang, J.; Wu, Q.; Zheng, J.; Shen, L.; Anpalagan, A. Dynamic spectrum access in time-varying environment: Distributed learning beyond expectation optimization. *IEEE Trans. Commun.* **2017**, *65*, 5305–5318. [CrossRef]
19. Xiao, L.; Lu, X.; Xu, D.; Tang, Y.; Wang, L.; Zhuang, W. UAV relay in VANETs against smart jamming with reinforcement learning. *IEEE Trans. Veh. Technol.* **2018**, *67*, 4087–4097. [CrossRef]
20. Chen, J.; Xu, Y.; Wu, Q. Distributed channel selection for multicluster FANET based on real-time trajectory: A Potential game approach. *IEEE Trans. Veh. Technol.* **2018**, submitted.
21. Hu, Q.; Yue, W. *Markov Decision Processes with Their Applications*; Springer: New York, NY, USA, 2007.
22. Puterman, M.L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*; John Wiley & Sons: Hoboken, NJ, USA, 2009.
23. Watkins, C.J.C.H.; Dayan, P. Q-learning. *Mach. Learn.* **1992**, *8*, 279–292. [CrossRef]
24. Slimeni, F.; Chtourou, Z.; Scheers, B.; Le Nir, V.; Attia, R. Cooperative Q-learning based channel selection for cognitive radio networks. *Wirel. Netw.* **2018**, 1–11. [CrossRef]
25. Kong, L.; Xu, Y.; Zhang, Y. A reinforcement learning approach for dynamic spectrum anti-jamming in fading environment. In Proceedings of the 2018 18th IEEE International Conference on Communication Technology (ICCT 2018), Chongqing, China, 8–11 October 2018; pp. 1–7.
26. Liu, X.; Xu, Y.; Jia, L.; Wu, Q.; Anpalagan, A. Anti-jamming communications using spectrum waterfall: A deep reinforcement learning approach. *IEEE Commun. Lett.* **2018**, *22*, 998–1001. [CrossRef]
27. Busoniu, L.; Babuska, R.; De Schutter, B. A comprehensive survey of multi-agent reinforcement learning. *IEEE Trans. Syst. Man Cybern.* **2008**, *38*, 156–172. [CrossRef]
28. Aref, M.A.; Jayaweera, S.K.; Machuzak, S. Multi-agent reinforcement learning based cognitive anti-jamming. In Proceedings of the 2017 IEEE Wireless Communications and Networking Conference (WCNC), San Francisco, CA, USA, 19–22 March 2017; pp. 1–6.
29. Aref, M.A.; Jayaweera, S.K. A novel cognitive anti-jamming stochastic game. In Proceedings of the Cognitive Communications for Aerospace Applications Workshop 2017, Cleveland, OH, USA, 27–28 June 2017; pp. 1–4.
30. Aref, M.A.; Jayaweera, S.K. A cognitive anti-jamming and interference-avoidance stochastic game. In Proceedings of the 2017 IEEE 16th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC), Oxford, UK, 26–28 July 2017; pp. 520–527.

31. Yao, F.; Jia, L. A collaborative multi-agent reinforcement learning anti-jamming algorithm in wireless networks. *arXiv* **2018**, arXiv:1809.04374.
32. Xu, Y.; Ren, G.; Chen, J.; Luo, Y.; Jia, L.; Liu, X.; Yang, Y.; Xu, Y. A one-leader multi-follower Bayesian-Stackelberg game for anti-jamming transmission in UAV communication networks. *IEEE Access* **2018**, *6*, 21697–21709. [CrossRef]
33. Cao, L.; Zheng, H. Distributed rule-regulated spectrum sharing. *IEEE J. Sel. Areas Commun.* **2008**, *26*, 130–145. [CrossRef]
34. Vlassis, N. A concise introduction to multi-agent systems and distributed artificial intelligence. *Synth. Lect. Artif. Intell. Mach. Learn.* **2007**, *1*, 1–71. [CrossRef]
35. Xu, Y.; Wu, Q.; Shen, L.; Wang, J.; Anpalagan, A. Opportunistic spectrum access with spatial reuse: Graphical game and uncoupled learning solutions. *IEEE Trans. Wirel. Commun.* **2013**, *12*, 4814–4826. [CrossRef]

© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).