Learning-Based Defense Against Malicious Unmanned Aerial Vehicles

Minghui Min*†, Liang Xiao*†, Dongjin Xu*, Lianfen Huang*, Mugen Peng‡
*Department of Communication Engineering, Xiamen University, Xiamen, China. Email: lxiao@xmu.edu.cn
†National Mobile Communications Research Laboratory, Southeast University, Nanjing, China.
‡School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China.
Email: pmg@bupt.edu.cn

***Abstract*—Adversary unmanned aerial vehicles (UAVs) seriously threaten public security and user privacy. In this paper, we propose a reinforcement learning (RL) based defense framework to address malicious UAVs close to a target estate such as a company or an institute. This framework uses Q-learning to choose the defense policy such as jamming the global positioning system signals (GPS) and hacking, and laser shooting. According to the defense history and the current security status of the target estate, this scheme can improve the UAV defense performance in the dynamic game without being aware of the UAV attack policy and environment model in the area of interests. Simulation results show that this scheme can reduce the risk rate of the estate and improve the utility compared with the benchmark scheme against malicious UAVs.**

***Index Terms*—Unmanned aerial vehicles, security, reinforcement learning, defense, GPS.**

I. Introduction

Unmanned aerial vehicles (UAVs) facilitate data acquisition at temporal and spatial scales that still remain unachievable for traditional remote sensing platforms. However, they can also threaten the public security and people's privacy. For example, the camera-equipped UAVs sometimes surreptitiously invade user's privacy and intercept data [1]. Besides, the sensitive information carried by UAVs is vulnerable to malicious attackers [2], [3]. In January and May 2015, the Secret Service reported at least two incidents where the UAVs flew into the restricted airspace around the White house. In April 2015, a protester landed a UAV on the roof of the Japanese Prime Minister's office. It carried a container of sand with traces of non-harmful radioactive isotopes [4]. Between January 2013 and August 2015, the report stated, there were 20 suspicious UAV related incidents in or around London alone.

Extensive work has been done to detect and track the malicious UAV, such as radar, acoustic sensing, radio frequency emission sensing and electro-optical sensing [2], [5]. Once a UAV has been identified, the defense systems need a way to destroy it. To minimize the risks to other airspace users and the estate on the ground, several methods have been studied to disable, disrupt or hijack UAVs ranging from physical to digital attacks. The UAV capture and control via Global Positioning System (GPS) signal spoofing is analyzed in [1], [6]. Several methods are studied to take down the malicious UAVs in [5], such as jamming [7], GPS spoofing, hacking, netgun and laser [8]–[10]. However, some drones can apply multiple attack policies and be able to keep flying without the connection to a human on the ground or the GPS navigation signals. Therefore, it is important for the smart malicious UAV defense system to accurately estimate the attack strategy of the malicious UAV (e.g., based on GPS navigation or not) and its fly mode in time.

Instead of being restricted to a specific scheme, we propose a smart malicious UAV defense system that incorporates most existing UAV detetcion/idetification and tracking schemes to improve the defense performance against the malicious UAV. More specifically, in this paper, the repeated interactions between the defense system and the malicious UAV over multiple time slots are formulated as a dynamic malicious UAV defense game, in which the defense system chooses the defense policy according to the state of the malicious UAV and the importance of the estate in the protected area. The malicious UAV defense policy selections in the dynamic game can be approximately formulated as a Markov decision process (MDP) with finite states, in which the defense system observes the state that consists of the previous attack mode of the UAV and the current importance of the target estate.. Therefore, reinforcement learning (RL) techniques can be applied to choose the appropriate defense policy to defend against the malicious UAV in the dynamic game.

As a model-free reinforcement learning technique, Q-learning can derive the optimal strategy in an MDP [11], [12]. By using Q-learning, the malicious UAV system can achieve the optimal defense strategy in the dynamic game without being aware of the attack model of the malicious UAV. The main contributions of this paper are summarized as follows:

(1) We formulate a smart malicious UAV defense system model, in which the defense system applies different intercept methods to prevent the malicious UAV from stealing data.

(2) A Q-learning based malicious UAV defense scheme is developed to achieve the optimal defense strategy, in which the

This work was supported in part by Natural Science Foundation of China under Grant 61671396, in part by the open research fund of National Mobile Communications Research Laboratory, Southeast University (No. 2018D08), and in part by Science and Technology Innovation Project of Foshan City, China (Grant No. 2015IT100095).

defense system chooses the intercept method according to the previous attack model and current importance of the target estate. This scheme enables the defense system to achieve the optimal defense performance without knowing the attack model in a dynamic defense game.

(3) Simulations are performed to show that our proposed reinforcement learning based malicious UAV defense scheme can reduce the risk rate of the estate and improve the utility of the defense system compared with the benchmark scheme.

The rest of this paper is organized as follows: We review the related word in Section II, and present the system model in Section III. A Q-learning based malicious UAV defense scheme is developed in Section IV. Simulation results are provided in Section V and conclusions of this work are drawn in Section VI.

II. Related work

A three-dimensional guidance law for rotary UAV interception proposed in [13] combines proportional navigation based guidance and velocity feedback. The theory and practice of UAV capture and control via GPS signal spoofing are analyzed and demonstrated in [6]. GPS spoofing signal on the autonomous UAV has been verified and assessed through the experimental results in [14], [15], which shows that spoofing signal affects the navigation system of the UAV so that the UAV goes off course or shows an abnormal operation. Several methods that can take down the malicious UAVs are described in [5], such as jamming, GPS spoofing, hacking, netgun, laser and so on.

Reinforcement learning techniques have been used to improve network security. The minimax-Q learning based spectrum allocation as developed in [16] increases the spectrum efficiency in cognitive radio networks [17]. A two-dimensional Q-learning based anti-jamming communication scheme in [18] is proposed to increase the signal-to-interference-plus-noise ratio of secondary users against cooperative jamming in cognitive radio networks.

In this paper, we investigate the malicious UAV defense in a dynamic system. A Q-learning based UAV defense scheme is proposed for the defense system to achieve the optimal defense performance without being aware of the attack model based on the system state that consists of previous attack mode and the current importance of the estate in the protected area. Simulation results show that the proposed scheme can achieve a better performance compared with the benchmark scheme against malicious UAVs.

III. System Model

As shown in Fig. 1, we consider a smart malicious UAV defense system consisting of a defense system and a malicious UAV. The defense system applies multiple intercept methods to prevent the malicious UAV to steal data, such as jamming, GPS spoofing and laser shooting.

To improve the defense performance, we split defense policy into different categories, according to their defense cost and the impacts on the malicious UAV, as shown in table I. Without loss of generality, let Level-i defense policy be stronger to defend against the malicious UAV than Level-j defense policy, with $i > j$. For example, laser shooting can be labelled with a higher level than GPS spoofing, if the former is considered to be more stronger to defend against a given malicious UAV.

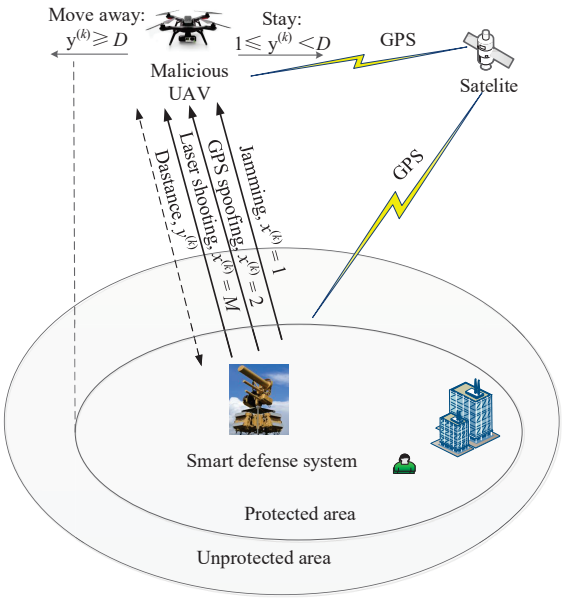


Fig. 1. Smart malicious UAV defense system that uses RL to choose the defense policy $x^{(k)}$, such as jamming, CPS spoofing and laser shooting to defend against the malicious UAV with $y^{(k)}$ meters far away from the protected area of the target estate.

TABLE I
Malicious UAV Defense Policies of the Framework.

Action Level	Defense policy
1	Send jamming signals
2	GPS spoofing
$M-1$	Send hacking siganls
M(strongest)	Laser shooting

Once a malicious UAV has been identified, the smart malicious UAV defense system chooses the defense policy at time k, denoted by $x^{(k)} \in X = \{0, 1, 2, ..., M\}$, where X is the action set of the defense system. The smart malicious UAV defense system keeps no action if $x = 0$, and defends against the UAV with Level-x defense policy if $x > 0$.

The state of the malicious UAV at time k is denoted by $y^{(k)} \in Y = \{0, 1, 2, ..., D\}$, in which the malicious UAV is crashed if $y^{(k)} = 0$, and remains $y^{(k)}$ distance away from the defense system if $1 \leq y^{(k)} \leq D$. The malicious UAV choose to move away (i.e., $y^{(k)} \geq D$), or stay in the interest area (i.e., $1 \leq y^{(k)} < D$). The state of the malicious UAV at next time is updated according to the current defense policy and the current state of the malicious UAV (i.e., the distance between the malicious UAV and the protected environment). The state

TABLE II
SUMMARY OF SYMBOLS AND NOTATIONS

$x^{(k)} \in X$	Defense policy at time k
$y^{(k)} \in Y$	UAV attack strategy
$C^{(k)}$	Defense cost
$G^{(k)}$	Defense gain
$U^{(k)}$	Utility of the defense system

transfer probability is denoted by $P_{x,y,y'} = \Pr(y'|x,y)$, where y' is the next state of the malicious UAV if the defense system applies Level-x defense policy against the UAV at state y.

Intuitively, the defense system obtains more security gains if a closer UAV is crashed that is more likely to steal sensitive information. For simplicity, the gain of the malicious UAV defense system at time k denoted by $G^{(k)}$ is modeled as a linear function of the distance between the malicious UAV and the protected area at last time slot, i.e., $G^{(k)} = A - By^{(k-1)}$, where A and B are constant. The cost of the malicious UAV defense system at time k denoted by $C^{(k)}$ is a function of the defense policy x and the distance between the malicious UAV and the protected area y. We model the cost of defense system as $C^{(k)} = \eta_1 x + \eta_2 y$, where η_1 and η_2 are constant.

Let $\mathrm{I}(\sigma)$ be the indicator function, which equals 1 if σ is true and 0 otherwise. The utility of the smart malicious UAV defense system at time k based on the security gain and the defense cost is denoted by $U^{(k)}$ and is given by

$$\begin{aligned} U^{(k)}(x,y) &= \mathrm{I}(y=0)\, G^{(k)}\phi^{(k)} - C^{(k)} \\ &= \mathrm{I}(y=0)\left(A - By^{(k-1)}\right)\phi^{(k)} - \eta_1 x - \eta_2 y, \end{aligned} \quad (1)$$

where $\phi^{(k)} \in (0,1]$ indicates the importance of the estate in the protected area at time k.

For ease of reference, the commonly used notations are summarized in Table II. The time index k in the superscript is omitted if no confusion occurs.

IV. Q-learning Based Malicious UAV Defense Scheme

In the dynamic malicious UAV defense system, the malicious UAV system chooses the defense policy based on the system state, which consists of the previous attack strategy and current importance of the target estate. The next system state observed by the defense system is independent of the previous states and actions, for a given system state and UAV defense strategy in the current time slot. Therefore, the malicious UAV defense process can be viewed as an MDP, in which the Q-learning technique, can derive the optimal policy without being aware of the attack model. We propose a Q-learning based defense policy selection scheme to defend against the malicious UAV in the dynamic game, as illustrated in Fig. 2. We initialize the distance between the malicious UAV and the protected area, y, and the learning parameters, α and γ are set to achieve a good defense performance. At time k, the smart malicious UAV defense system observes the state of the defense system, $s^{(k)}$, which consists of the previous state of the malicious UAV and the current importance of the target estate, i.e., $s^{(k)} = \left[y^{(k-1)}, \phi^{(k)}\right] \in \Lambda$, where Λ is the state set of the defense system. Based on the system state, the malicious UAV defense system chooses the defense policy $x^{(k)} \in X$.

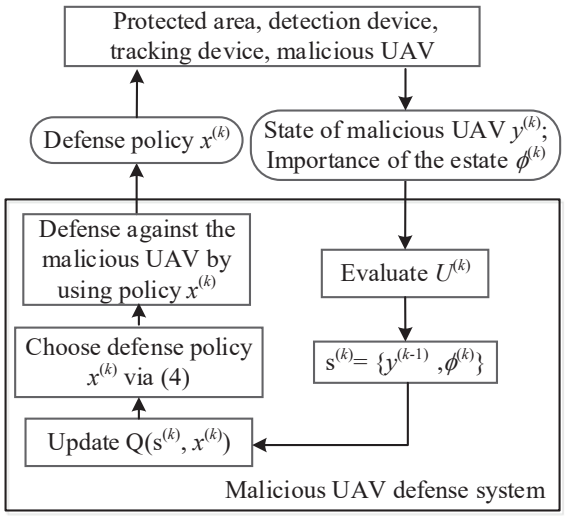


Fig. 2. Q-learning based malicious UAV defense scheme.

Algorithm 1 Q-learning based Malicious UAV Defense Scheme.

1: Initialize $y^{(0)}$, $\phi^{(1)}$, α, and γ
2: $\mathbf{Q} = \mathbf{0}$
3: Randomly choose $\mathbf{s}^{(0)} \in \Lambda$
4: **for** $k = 1, 2, 3, \ldots$ **do**
5: $s^{(k)} = \left[y^{(k-1)}, \phi^{(k)}\right]$
6: Choose $x^{(k)}$ via (4)
7: Apply the defense policy $x^{(k)}$
8: Measure the UAV distance $y^{(k)}$
9: Estimate the importance of the data in the target estate $\phi^{(k+1)}$
10: Evaluate $U^{(k)}$ via (1)
11: Update $Q(s^{(k)}, x^{(k)})$ via (2)
12: Update $V(s^{(k)})$ via (3)
13: **end for**

The malicious UAV system evaluates its reward or utility $U^{(k)}$ based on the state of malicious UAV, the importance of the target estate in the protected area, and the selected defense policy. The Q-learning based malicious UAV defense system maintains a Q-function for each action-state pair, denoted by $Q(s,x)$, which is the expected discounted long-term reward observed by the malicious UAV defense system. The Q-function is updated at time k according to the iterative Bellman

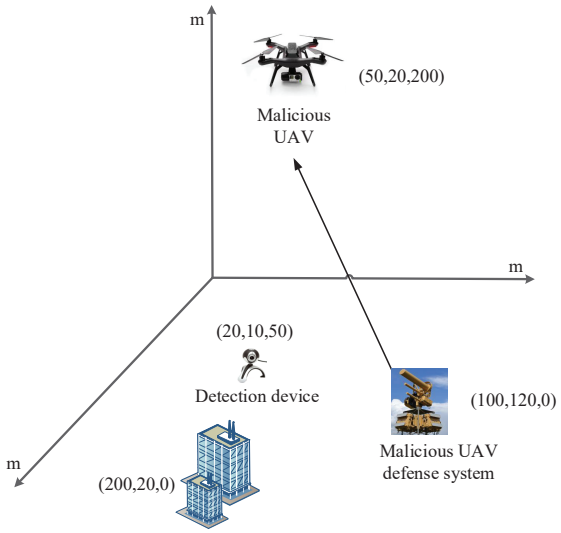


Fig. 3. Initial topology of the defense system in the simulation.

equation as follows:

$$Q\left(s^{(k)},x^{(k)}\right)\leftarrow(1-\alpha)\,Q\left(s^{(k)},x^{(k)}\right)+\alpha\left(U^{(k)}+\gamma V\left(s'\right)\right), \tag{2}$$

where s' is the next state if the defense system applies the Level-x defense policy against the UAV at state $s^{(k)}$, the learning rate $\alpha\in(0,1]$ is the weight of the current experience, the discount factor $\gamma\in[0,1]$ indicates the uncertainty of the defense system on the future reward, and the value function $V(s)$ maximizes $Q(s,x)$ over the action set given by

$$V\left(s^{(k)}\right)=\max_{x'\in X}Q\left(s^{(k)},x'\right). \tag{3}$$

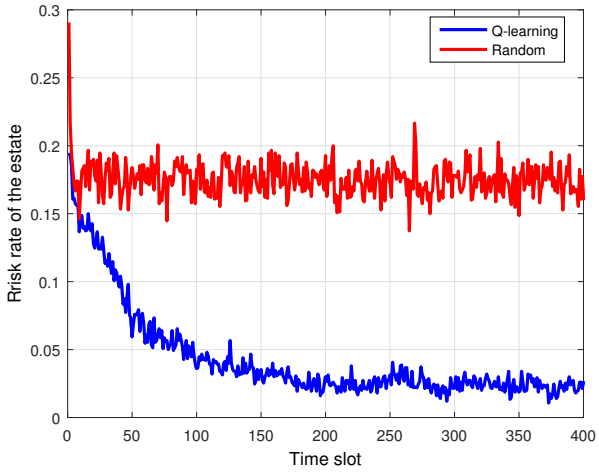
To make a tradeoff between exploitation and exploration, the malicious UAV defense policy is chosen according to the ε-greedy policy. More specifically, the malicious UAV defense policy $x^{(k)}$ that maximizes the Q-function is chosen with a high probability $1-\varepsilon$, and other actions are selected with a low probability to avoid staying in the local maximum, i.e.,

$$\Pr\left(x^{(k)}=\hat{x}\right)=\begin{cases}1-\varepsilon, & \hat{x}=\arg\max_{x'}Q\left(s^{(k)},x'\right)\\ \frac{\varepsilon}{|X|-1}, & \text{o.w.}\end{cases} \tag{4}$$

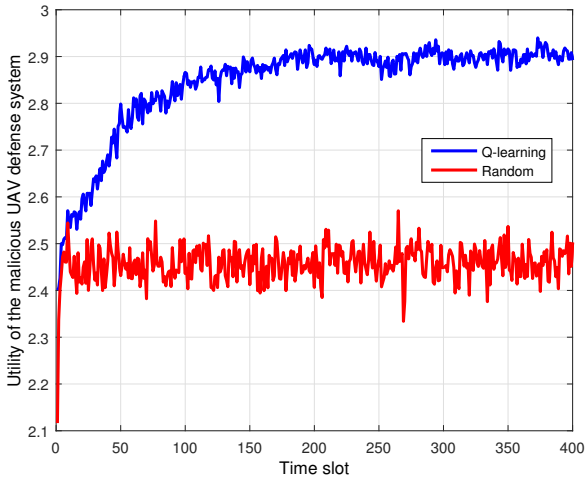
The learning process is shown as in Algorithm 1.

V. Simulation Results

Simulations have been performed to evaluate the proposed RL-based malicious UAV defense scheme for the defense system with topologies as shown in Fig. 3. In the simulations, we set $\alpha=0.9$, $\gamma=0.5$, $A=5$, $B=4$, $\eta_1=2$ and $\eta_2=3$, if not specified otherwise. The malicious UAV is more likely to be intercepted with a stronger defense policy and a shorter distance between the malicious UAV and the protected area.



(a) Risk rate of the estate



(b) Utility of the defense system

Fig. 4. Malicious UAV defense performance in the simulations against a malicious UAV with settings as shown in Fig. 3, with $\alpha=0.9$, $\gamma=0.5$, $A=5$, $B=4$, $\eta_1=2$ and $\eta_2=3$.

As a special case, we set $P_{x,y,y'}$ as follow,

$$P_{x,y,y'}=\Pr(y'|x,y)=\begin{cases}\frac{e^{\frac{mx}{ny}}}{1+e^{\frac{mx}{ny}}}, & y'=0\\ \frac{1}{(|Y|-1)\left(1+e^{\frac{mx}{ny}}\right)}, & \text{o.w.}\end{cases} \tag{5}$$

where m and n represent the impact weights of the defense policy x and the distance y on the defense results, respectively. We evaluate the Q-learning based malicious UAV defense performance. As shown in Fig. 4 (a), the risk rate decreases over time with our proposed Q-learning based malicious UAV defense scheme, and converges to 0.02% after about 200 time slots, which is about 88.6% lower than that of the random strategy. Consequently, as shown in Fig. 4 (b), the utility of the defense system increases quickly after the start of the learning process, and converges to a certain value that is much higher than the benchmark strategy. For example, the utility of the

defense system with our proposed scheme exceeds the random strategy by 18.4% at time 200, because the defense system adjusts the defense policy via trials-and-errors.

VI. Conclusion

In this paper, we have proposed a smart defense system to defend malicious UAVs that uses Q-learning to improve the defense performance without being aware of the attack policy in the dynamic game. Simulation results show that the proposed RL based malicious UAV defense scheme can reduce the risk rate of the target estate of the protected area, and increase the utility of the system compared with the benchmark scheme with random policy selection. For instance, the risk rate of the estate is 88.6% lower and the utility of the defense system is 18.4% higher, compared with the benchmark scheme after 200 time slots.

References

[1] D. He, S. Chan, and M. Guizani, "Communication security of unmanned aerial vehicles," *IEEE Wireless Commun.*, vol. 24, no. 4, pp. 134–139, Dec. 2016.

[2] D. Sathyamoorthy, "A review of security threats of unmanned aerial vehicles and mitigation steps," *J. Defence and Security*, vol. 6, no. 1, pp. 81–97, 2015.

[3] Y. Xiao, V. K. Rayi, B. Sun, X. Du, F. Hu, and M. Galloway, "A survey of key management schemes in wireless sensor networks," *Computer communications*, vol. 30, no. 11-12, pp. 2314–2341, Sep. 2007.

[4] W. Ripley, "Drone with radioactive material found on japanese prime minister's roof." Available online at: http://edition.cnn.com/2015/04/22/asia/japan-prime-minister-rooftop-drone (Last access date: 14 July 2015).

[5] T. Humphreys, "Statement on the security threat posed by unmanned aerial systems and possible countermeasures," *Technical Report, Oversight and Management Efficiency Subcommittee, Homeland Security Committee*, Apr. 2015.

[6] A. J. Kerns, D. P. Shepard, J. A. Bhatti, and T. E. Humphreys, "Unmanned aircraft capture and control via GPS spoofing," *J. Field Robotics*, vol. 31, no. 4, pp. 617–636, Apr. 2014.

[7] B. Han, J. Li, J. Su, M. Guo, and B. Zhao, "Secrecy capacity optimization via cooperative relaying and jamming for wanets," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 4, pp. 1117–1128, Apr. 2015.

[8] X. Lu, D. Xu, L. Xiao, L. Wang, and W. Zhuang, "Anti-jamming communication game for UAV-aided VANETs," in *Proc. IEEE Global Commun. Conf. (GLOBECOM), Singapore, Dec. 2017*.

[9] S. Lv, L. Xiao, Q. Hu, X. Wang, C. Hu, and L. Sun, "Anti-jamming power control game in unmanned aerial vehicle networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM), Singapore, Dec. 2017*.

[10] L. Xiao, Y. Li, G. Han, G. Liu, and W. Zhuang, "PHY-layer spoofing detection with reinforcement learning in wireless networks," *IEEE Trans. Vehicular Technology*, vol. 65, no. 12, pp. 10037–10047, Dec. 2016.

[11] C. J. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3-4, pp. 279–292, May 1992.

[12] L. Xiao, Y. Li, G. Han, H. Dai, and H. V. Poor, "A secure mobile crowdsensing game with deep reinforcement learning," *IEEE Trans. Information Forensics & Security*, Jul. 2017.

[13] B. Zhu, A. H. B. Zaini, and L. Xie, "Distributed guidance for interception by using multiple rotary-wing unmanned aerial vehicles," *IEEE Trans. Industrial Electronics*, vol. 64, no. 7, pp. 5648–5656, Jul. 2017.

[14] S. Seo, B. Lee, S. Im, and G. Jee, "Effect of spoofing on unmanned aerial vehicle using counterfeited GPS signal," *J. Positioning, Navigation, and Timing*, vol. 4, no. 2, pp. 57–65, May 2015.

[15] X. Du, Y. Xiao, M. Guizani, and H.-H. Chen, "An effective key management scheme for heterogeneous sensor networks," *Ad Hoc Networks*, vol. 5, no. 1, pp. 24–34, Jan. 2007.

[16] B. Wang, Y. Wu, K. R. Liu, and T. C. Clancy, "An anti-jamming stochastic game for cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 4, pp. 877–889, Apr. 2011.

[17] C. Zhang and W. Zhang, "Spectrum sharing for drone networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 1, pp. 136–144, Jan. 2017.

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