Anti-jamming Power Control Game in Unmanned Aerial Vehicle Networks

Shichao Lv*† Liang Xiao‡ Qing Hu*† Xiaoshan Wang*† Changzhen Hu§ Limin Sun*†
*School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China.
†Beijing Key Laboratory of IOT Information Security Technology, IIE CAS, Beijing 100093, China.
‡Dept. of Communication Engineering, Xiamen University, Xiamen 361005, China
§School of Software, Beijing Institute of Technology, Beijing 100081, China
Emails: {lvshichao,huqing,wangxiaoshan,sunlimin}@iie.ac.cn, lxiao@xmu.edu.cn, czhoo@bit.edu.cn

***Abstract*—In this paper, the anti-jamming issue in unmanned aerial vehicle (UAV) networks is analyzed in a static game and a dynamic game. We investigate the effect of wireless channel fading characteristics from a UAV to a ground station and flying cost on the performance of a closed-form Nash equilibrium (NE) in the static game. Besides, in a Stackelberg dynamic game, wherein the system model is hard to determine, we propose a Q-learning based anti-jamming scheme and evaluate its performance via exhaustive simulations, which can achieve relatively higher average utility and Signal to Interference plus Noise Ratio (SINR) than a benchmark method.**

***Index Terms*—unmanned aerial vehicle, anti-jamming, game**

I. Introduction

Due to the low cost and high convenience, unmanned aerial vehicles (UAVs) has gained rapid promotion and development in the civil and commercial applications in recent years [1]. The emerging applications can be substantially divided into four classes: disaster response, cargo delivery, monition area or wireless network coverage and project construction. All of these applications are based on wireless communication technologies, such as navigation (e.g., Global Positioning System, GPS), short distance unlicensed wireless communications (e.g., WiFi or ZigBee), long distance licensed wireless communications (e.g., LTE) [2]. Once those wireless links are jammed, the communication between the UAV and the ground station or the remote control terminal will be interrupted or unavailable. The consequence of jamming in UAV networks is serious since the UAV may be compromised by an adversary. For example, a portable shoulder-mounted rifle called Drone Defender can thoroughly invalidate the GPS or WiFi signals of a UAV, solely employing directional electromagnetic wave. Thereby, the UAV will have to hover in place or directly land to the specified ground under the control of an attacker [3].

Besides, compared to other wireless applications, high mobility and speed are the most distinctive features of the UAV equipment. It is mainly because that the movement trajectories of them are in a three-dimensional space in the air rather than in a two-dimensional one confined to a highway (e.g., vehicles in a vehicular ad hoc network) [2]. With respect to these two features, they will directly affect the fading characteristics of Air-Ground channel (e.g., resulting much more path loss) and the network quality (e.g., the dynamic change of network topology will lead to an intermittent connection) [4]. In addition, if there is a jamming attack, the reliability of wireless connection will be much more intractable. Whereas, thanks to the high velocity and flexible maneuverability, defending against or evading a wireless jamming attack may become easier. More specifically, facing a jamming attacker, the UAV can quickly increase its transmitting power or fly toward to the legitimate receiver, observably improving the received signal quality. Otherwise, the UAV can promptly evade the heavy jamming area via lots of aircraft motions (such as level flight, banking turn, climb or descent) to avoid wasting too much unnecessary energy. Hence, as for how to resist jamming attacks, there should be a complicated trade-off.

Game theory has been demonstrated to be an extremely effective method to study the jamming and anti-jamming issue in wireless networks [5]–[8]. In this paper, we carry out a theoretic study on anti-jamming power control game in UAV networks. The interactions between a jammer and a UAV are formulated as a jamming game. Under a transmitting power limitation, a UAV sends its message to a legitimate receiver. First, a static one-shot jamming interaction scenario is studied and a closed-form Nash equilibrium (NE) is derived. Second, in a dynamic Stackelberg jamming scenario, an optimal strategy of anti-jamming power control for the UAV is derived via reinforcement learning algorithm [9]. Exhaustive simulations are utilized to evaluate the performance of anti-jamming power control.

The key contributions of this paper include:

1) Extending jamming games to UAV networks, a closed-form NE expression is obtained for a static jamming game. Considering the factors of flying cost and wireless channel fading characteristics from a UAV to a ground station, we discuss its performance of the derived NE.
2) Confronting the high mobility, dynamic link distance and uncertain wireless channel power gains of the UAV network, a multiple time slot learning-based optimal

This work was supported in part by the National Key Research and Development Program of China (2016YFB0800202), Key Research Program of Chinese MIIT (JCKY2016602B001), National Defense Science and Technology Innovation Foundation of the Chinese Academy Sciences (CXJJ-16Z234), National Natural Science Foundation of China (61671396), CCF-Venustech Hongyan Research Initiative (2016-010). *(Corresponding author: Limin Sun).*

strategy is proposed and evaluated.

The remainder of this paper is organized as follows. Section II reviews the related work. Section III presents our system model. We formulate a static and a dynamic game in Section IV and V. Next, we provide numerical and simulation results in Section VI. Finally, Section VII concludes the whole paper.

II. Related Work

In [5], the static and dynamic jamming games between a jammer and underwater sensors were investigated, where the closed-form NE and reinforcement learning based optional anti-jamming strategy are present. In [6], a Bayesian Stackelberg game was formulated for anti-jamming transmission with incomplete information, wherein the Stackelberg equilibrium is obtained via a backward induction analysis. In [7], regarding AWGN and frequency selective fading channels, the two-party zero-sum game was analyzed, focusing on the achieved channel capacity of the authorized user when there is a hostile jammer. [8] present a jointly optimizing constrained zero-sum Markov game of transmission rate adaptation and frequency hopping for the wireless system in the presence of a jammer.

In the UAV-assisted drone delivery system for online purchases transportation, a network interdiction game [10] was investigated, in which the vendor player chooses the optimal flying path of a UAV to securely and timely deliver goods whilst the cyber-physical attacker aims to interdict the potential path and thwart this delivery. The NE of it is derived by solving the standard linear programming problems and the subjective behavior of players is analyzed via prospective theory. In a scenario, wherein two UAVs encounter a jamming UAV, as described in [11], based on the differential game theory, the jamming problem was described as a zero-sum pursuit-evasion game. Using Isaacs approach, the necessary conditions for the saddle point strategy are obtained. Besides, when there is one UAV, aiming to communicate with some relay nodes under the jamming attack coming from another malicious UAV, [12] also translated the game between the legitimate UAV and jamming UAV into a zero-sum pursuit-evasion game via a set of Isaacs equations. However, none of these works take the characteristics of the UAV-to-Ground (also named Air-ground) radio propagation or flying cost into account in a static game. Besides, none of the above works consider the impact of variable link distance and channel fading characteristics as well as their consequence in dynamic anti-jamming games. Those are the focus of this paper.

III. System Model and Background

A. System Model

In this work, we consider a UAV wireless communication network, consisting of a legitimate UAV U, a jamming attacker J and a ground station GS. GS and J are located on the ground, whereas U is flying in the air according to its flying track as depicted by the blue dotted line in Fig. 1. The relative horizontal link distance between J and GS is fixed and denoted by D_J, and the maximal communication range

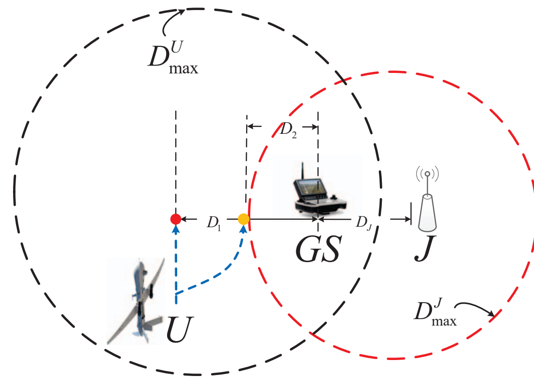


Fig. 1. System model.

TABLE I
Summary of Symbols and Notation

U	UAV
J	Jammer
GS	Ground staion
H	Channel power gain
H_L	Large scale channel power gain
H_S	Small scale channel power gain
$PL(d)$	Path loss of link distance d
$K(d)$	Ricean K-factor of link distance d
$a(b)$	Transmitting power of U (J)
$\mathbf{p}(\mathbf{q})$	Mixed strategy of U (J)
C_U	Unit cost of transmitting power of U
C_J	Unit cost of transmitting power of J
C_M	Loss of the task if U does not transmit any signal
C_F	Cost of flying power

of J is D^J_{max}. At time slot k, U can transmit its signal either at location L_1 (or L_2), which are denoted by yellow (or red) dot. Thus, the relative horizontal link distance between U and GS is D_U, wherein D_U equals either D_1 (or D_2) as shown in Fig. 1. $D_2 = D^J_{max} - D_J$ is the closest distance U can achieve. When $D_U = D_2$, U is exactly out of the jamming range and can not be shot by J via directional electromagnetic wave tool such as Drone Defender. The shorter the link distance is, the smaller the path loss is. Without loss of generality, it is assumed that the path loss of location L_2 is significantly less than the location L_1. Hence, in the downlink from U to GS, to improve GS's received signal strength, U can either increase its transmitting power at location L_1 or fly toward to GS and transmit its signal at location L_2.

The downlink jamming interactions are assumed in a slotted-time process, including two players in the air: U and J. The transmit power of U (or J) is equably quantized into N_U (or N_J) levels with a unit transmit power ΔP. Then the action of U (or J) denoted by a (or b), which is normalized transmit power with a power constraint $0 \leq a \leq N_U$ (or $0 \leq b \leq N_J$). Hence, the actual transmitting power of U (or J) equals $a\Delta P$ (or $b\Delta P$). In addition, let $\tilde{H}_U$ (or $\tilde{H}_J$) be the normalized channel power gain between U (or J) and GS. $H_U = \tilde{H}_U \Delta P$ (or $H_J = \tilde{H}_J \Delta P$) denotes the normalized channel power gain of unit transmit power level for U (or J) within the dwell time of the current time slot. H_J can be described via a classical two-ray model channel whilst H_U is depicted by a Air-Ground fading channel given in the following subsection. And at location L_1 and L_2, H_U can be written by H_{U_1} and H_{U_2}, respectively. For the sake of reference, the commonly used notation is summarized in Table I.

B. Air-Ground Large and Small Scale Fading Characteristics

According to [13], the large scale fading character of path loss for Air-Ground wireless channel between U and GS, complies with a modified log-distance model as

$$PL(d)(dB) = A_0 + 10nlog_{10}(d/d_{min}) + \varsigma F + X \quad (1)$$

where A_0 is measured in dB, indicating the value of path loss at the link range of d_{min}; n is a dimensionless path loss exponent; X is a Gaussian random variable with zero-mean and standard deviation σ_X; $\varsigma \in [-1, 1]$ denotes the flight direction, when the UAV flies toward GS, $\varsigma = -1$, otherwise $\varsigma = 1$; F is a adjustable magnitude in dB regarding the flight direction.

Since the path loss $PL(d)$ is defined as the difference between the transmitting signal power P_t and the received signal power P_r in logarithmic units, i.e., $P_r = P_t - PL(d)$. For the sake of brevity, the relationship between the large scale channel power gain H_L and the path loss can be given by,

$$H_L(dB) \approx \frac{P_t - PL(d)}{P_t}. \quad (2)$$

Then the linear form of H_L can be given by $H_L = 10^{\frac{H_L(dB)}{10}}$. From (1) and (2), we can see that H_L can be roughly estimated by the transmitting power P_t and the link range d.

Besides, in [13], the Air-Ground small scale fading character is well modeled via the Ricean fading, in which a linear model of Ricean K-factor is given in the following,

$$K(d) = K_{d_{min}} + n_k(d - d_{min}) + Y \quad (3)$$

where $K_{d_{min}}$ is the value of K at link distance d_{min}, n_k denotes a slope while Y is a Gaussian random variable with zero mean and standard deviation σ_Y.

If the flying velocity is assumed to be known, then we can compute the maximal doppler frequency shift. Along with the above Ricean K-factor, the small scale fading channel power gain H_S can also be determined. After getting H_L and H_S, the Air-Ground channel power gain H can be simultaneously described via large and small scale fading characteristics, e.g., $H = H_L H_S$.

IV. Jamming Games in UAV Wireless Networks

The contest between U and J is described as a jamming game, in which both U and J take actions to choose their transmit power. In this section, we analyze a one-shot interaction static jamming game $\mathbb{G}$, focusing on the uncertainty of opponent's behavior when the large and small scale fading characteristics are assumed to be precisely estimated.

A. Game Model

In the static game $\mathbb{G}$, the strategy space for U (or J) can be defined as

$$\begin{cases} \mathcal{A} = \{(a) | 0 \le a \le N_U\} & (4a) \\ \mathcal{B} = \{(b) | 0 \le b \le N_J\} & (4b) \end{cases}$$

where the number of elements of $\mathcal{A}$ and $\mathcal{B}$ are $(N_U + 1)$ and $(N_J + 1)$, respectively.

If $a = 0$, U does not transmit any signal to GS, then U will lose C_M. If $a > 0$, U can transmit at location L_1 or L_2 according to the following rules. U pre-defines a SINR threshold η and a flying power cost threshold δ. If the SINR at location L_1 is greater than or equal to η (i.e., $\frac{aH_{U_1}}{\sigma + bH_J} \ge \eta$), then U will transmit at location L_1 with power $a = \frac{\eta(\sigma + N_J H_J)}{H_{U_1}}$. However, if the SINR at location L_1 is less than η (i.e., $\frac{aH_{U_1}}{\sigma + bH_J} < \eta$), then U will have two choices, flying toward to GS and transmitting at location L_2 with power $a = \frac{\eta(\sigma + N_J H_J)}{H_{U_2}}$ when $0 < C_F \le \delta$, otherwise, U transmits at location L_1 with a larger power $a = \frac{H_{U_2}}{H_{U_1}} \cdot \frac{\eta(\sigma + N_J H_J)}{H_{U_1}}$ when $C_F > \delta$.

At time slot k, let the normalized channel power gains of U and J are H_U and H_J, wherein H_U can be either H_{U_1} or H_{U_2}. Besides, the Additive White Gaussian Noise (AWGN) power is σ. When both players take strategy $a \in \mathcal{A}$ or $b \in \mathcal{B}$, the utility of U and J, which are based on the Signal to Interference plus Noise Ratio (SINR) of GS's received signal, can be severally given by

$$\begin{aligned} & u_U(a, b) \\ & = \frac{aH_U}{\sigma + bH_J} \\ & - [I(\frac{aH_{U_1}}{\sigma + bH_J} \ge \eta) \frac{\eta(\sigma + N_J H_J)}{H_{U_1}} C_U \\ & + I(\frac{aH_{U_1}}{\sigma + bH_J} < \eta) I(0 < C_F < \delta)(\frac{\eta(\sigma + N_J H_J)}{H_{U_2}} C_U + C_F) \\ & + I(\frac{aH_{U_1}}{\sigma + bH_J} < \eta) I(C_F > \delta) \frac{H_{U_2}}{H_{U_1}} \frac{\eta(\sigma + N_J H_J)}{H_{U_1}} C_U \\ & + I(a == 0) C_M] \\ & + bC_J \end{aligned} \quad (5)$$

$$u_J(a, b) = -u_U(a, b) \quad (6)$$

where C_U (or C_J) is the cost of unit power of U (or J). C_F is the flying cost of U while C_M is the loss of the task if U does not transmit any signal. η is a threshold of SINR used to assess the wireless channel quality. δ is a threshold of flying cost used to decide whether flying toward to GS or not. $I(\cdot)$ is an indication function, if the event is true, then $I(\cdot) = 1$ and $I(\cdot) = 0$ otherwise. H_U (or H_J) is the wireless channel power gain of U (or J), considering both the large and small scale channel fading characteristics. Note that in the first item of (5), the value of H_U depends on the transmitting location of U. If U transmits at location L_1 (or L_2), then $H_U = H_{U_1}$ (or $H_U = H_{U_2}$).

Mixed strategy is used in $\mathbb{G}$, and the probability of choosing its strategy x (or y) is defined as $\mathbf{p} = [p_a]_{0\le a\le N_U}$ (or $\mathbf{q} = [q_b]_{0\le b\le N_J}$). Hence, the expected utility of U (or J) is denoted as

$$u_U^{EUT}(\mathbf{p},\mathbf{q}) = \sum_{0\le a\le N_U}\sum_{0\le b\le N_J} p_a q_b u_U(a,b). \tag{7}$$

$$u_J^{EUT}(\mathbf{p},\mathbf{q}) = \sum_{0\le a\le N_U}\sum_{0\le b\le N_J} p_a q_b u_J(a,b). \tag{8}$$

B. Nash equilibrium

The Nash equilibrium (NE) of $\mathbb{G}$ can be given by $(\mathbf{p}^*,\mathbf{q}^*)$, indicating the best response strategy of each player as long as the opponent follows this NE. Hence, the NE of $\mathbb{G}$ can be given by

$$\begin{cases} \mathbf{p}^* = \operatorname*{argmax}_{\mathbf{p}} u_U(\mathbf{p},\mathbf{q}^*) & \text{(9a)}\\ \mathbf{q}^* = \operatorname*{argmax}_{\mathbf{q}} u_J(\mathbf{p}^*,\mathbf{q}) & \text{(9b)}\\ s.t. \sum_{0\le a\le N_U} p_a = 1, \sum_{0\le b\le N_J} q_b = 1 & \text{(9c)}\\ \mathbf{p}\succeq 0, \mathbf{q}\succeq 0. & \text{(9d)} \end{cases}$$

Theorem 1: An NE of the static jamming game $\mathbb{G}$ is given by

$$\begin{cases} \sum_{0\le b\le N_J} u_U(a,b)\mathbf{q}_b^* = \lambda_U[1,...,1]^T, 0\le a\le N_U & \text{(10a)}\\ \sum_{0\le a\le N_U} u_J(a,b)\mathbf{p}_a^* = \lambda_J[1,...,1]^T, 0\le b\le N_J & \text{(10b)}\\ s.t. \sum_{0\le a\le N_U} p_a = 1, \sum_{0\le b\le N_J} q_b = 1 & \text{(10c)}\\ \mathbf{p}\succeq 0, \mathbf{q}\succeq 0. & \text{(10d)} \end{cases}$$

where λ_U and λ_J are Lagrange multipliers as long as there is a solution. While $[1,...,1]^T$ is a column vector with $((N_U+1)*(N_J+1))$-row.

Proof: Following (7) and (9a), the Lagrangian of L_U is given by

$$\begin{aligned} L_U(\mathbf{p},\lambda_U,\mu) = u_U^{EUT}(\mathbf{p},\mathbf{q}^*) - \lambda_U(\sum_{0\le a\le N_U} p_a - 1)\\ + \sum_{0\le a\le N_U}\mu_a p_a. \end{aligned} \tag{11}$$

where the Lagrange multipliers λ_U and μ, satisfying Karush-Kuhn-Tucker (KTT) conditions, are defined by

$$\begin{cases} \dfrac{\partial L_U(\mathbf{p},\lambda_U,\mu)}{\partial p_a} = 0\\ p_a\ge 0, \mu_a\ge 0, \mu_a p_a = 0, \quad 0\le a\le N_U\\ \sum_{0\le a\le N_U} p_a - 1 = 0. \end{cases} \tag{12}$$

Based on (7) and (11), applying the above complementary slackness KKT condition (12), we can get

$$\begin{cases} \sum_{0\le b\le N_J} u_U(a,b)\mathbf{q}_b^* - \lambda_U = 0, \quad 0\le a\le N_U\\ \sum_{0\le a\le N_U} p_a = 1\\ \lambda_U \ge 0 \end{cases} \tag{13}$$

Thus, $\mathbf{q}_b^*$ is obtained. Similarly, $\mathbf{p}_a^*$ can be derived as (10b). ■

Corollary 1: If $N_U = N_J = 1$, a unique NE of the static game $\mathbb{G}$, which is given by

$$\begin{cases} q_0^* = p_0^* = \dfrac{u_U(1,1)-u_U(0,1)}{u_U(1,1)-u_U(0,1)+u_U(0,0)-u_U(1,0)} & \text{(14a)}\\ q_1^* = p_1^* = \dfrac{u_U(0,0)-u_U(1,0)}{u_U(1,1)-u_U(0,1)+u_U(0,0)-u_U(1,0)} & \text{(14b)} \end{cases}$$

with conditions I_1 : $u_U(0,0)u_U(1,1) \ne u_U(1,0)u_U(0,1)$, $u_J(0,0)u_J(1,1) \ne u_J(1,0)u_J(0,1)$ and I_2 : $\frac{u_U(1,1)-u_U(0,1)}{u_U(0,0)-u_U(1,0)} \ge 0$ or $\frac{u_U(0,0)-u_U(1,0)}{u_U(1,1)-u_U(0,1)} \ge 0$.

Proof: If $N_U = N_J = 1$, by (13) we can get

$$\begin{bmatrix} u_U(0,0) & u_U(0,1)\\ u_U(1,0) & u_U(1,1)\end{bmatrix}\begin{bmatrix} q_0^*\\ q_1^*\end{bmatrix} = \begin{bmatrix}\lambda_U\\ \lambda_U\end{bmatrix} \tag{15}$$

Since $q_1^* = 1 - q_0^*$, we obtain

$$\frac{q_0^*}{1-q_0^*} = \frac{u_U(1,1)-u_U(0,1)}{u_U(0,0)-u_U(1,0)}. \tag{16}$$

Thus, q_0^* is derived as (14a), then (14b) can also be computed via $q_1^* = 1 - q_0^*$. Similarly, p_0^* and p_1^* are obtained as (14a) and (14b). In a single time slot, once the channel power gain H_{U_1}, H_{U_2} and H_J are fixed, then there are five cases, i.e.,

- $0 < \eta \le \frac{H_{U_1}}{\sigma+H_J} < \frac{H_{U_1}}{\sigma}$
- $0 < \frac{H_{U_1}}{\sigma+H_J} < \eta \le \frac{H_{U_1}}{\sigma}$, $C_F > \delta$
- $0 < \frac{H_{U_1}}{\sigma+H_J} < \eta \le \frac{H_{U_1}}{\sigma}$, $0 < C_F < \delta$
- $0 < \frac{H_{U_1}}{\sigma+H_J} < \frac{H_{U_1}}{\sigma} < \eta$, $C_F > \delta$
- $0 < \frac{H_{U_1}}{\sigma+H_J} < \frac{H_{U_1}}{\sigma} < \eta$, $0 < C_F < \delta$

In each case, the values of q_0^* (or p_0^*) and q_1^* (or p_1^*) are various due to different transmitting location choices of U.

Next, we prove the uniqueness of q_1^* and q_0^*. If there is another different solution $(q_1')^*$ and $(q_0')^*$, we will get

$$\begin{aligned} u_U(0,0)(q_0^* - (q_0')^*) + u_U(0,1)(q_1^* - (q_1')^*) = 0\\ u_U(1,0)(q_0^* - (q_0')^*) + u_U(1,1)(q_1^* - (q_1')^*) = 0. \end{aligned} \tag{17}$$

Since $q_0^* \ne (q_0')^*$ and $q_1^* \ne (q_1')^*$, then we can get $u_U(0,0) = u_U(0,1) = u_U(1,0) = u_U(1,1) = 0$. Furthermore, $u_U(0,0)u_U(1,1) = u_U(1,0)u_U(0,1) = 0$, which is contradictory with condition I_1. Hence, (14a) and (14b) are unique. Similarly, the uniqueness of p_0^* and p_1^* can be proved. ■

V. Learning-Based Anti-jamming Power Control

In this section, the static jamming game is extended to a Stackelberg dynamic game $\mathbb{G}'$, including two players: a leader (i.e., a foresighted UAV U) and a follower (i.e., a myopic jammer J). Taking the variation of link distance and consequent fading channels into account, pure strategies of U and J are the power selections at different time slots. We propose a practical Q-learning based method for obtaining the optimal power control strategy in $\mathbb{G}'$.

At a single time slot k, as a leader in this game, the foresighted U first chooses its action, focusing on both instantaneous utility $u_U^{(k)}$ as given in (18) and its future long-term reward. Then after observing U's action $a^{(k)}$, as a follower, the myopic jammer J chooses its action $b^{(k)}$, just aiming to maximize its instantaneous utility at time slot k as defined in (19).

The action of U is selected based on the system state at time slot k denoted by $s^{(k)} = [b^{(k-1)}, H_U^{(k-1)}, H_J^{(k-1)}]$, indicating the J's power choice of $b^{(k-1)}$, and the wireless channels power gains $H_U^{(k-1)} = (H_U)_L^{(k-1)}(H_U)_S^{(k-1)}$ as well as $H_J{}^{(k-1)} = (H_J)_L^{(k-1)}(H_J)_S^{(k-1)}$. The value of H_U can be either H_{U_1} or H_{U_2}. At different time slots, D_U changes significantly, then the corresponding H_U and received signal power of GS will vary greatly, making them difficult to be accurately determined and the optimal power strategy intractability to be obtained. Therefore, it is challenging to derive a closed-form expression for the optimal transmitting power strategy.

$$\begin{aligned}
&u_U^{(k)}(a^{(k)}, b^{(k)}) \\
&= \frac{a^{(k)} H_U^{(k)}}{\sigma + b^{(k)} H_J^{(k)}} \\
&- [I(\frac{a^{(k)} H_{U_1}^{(k)}}{\sigma + b^{(k)} H_J^{(k)}} \geq \eta) u_1 \\
&+ I(\frac{a^{(k)} H_{U_1}^{(k)}}{\sigma + b^{(k)} H_J^{(k)}} < \eta) I(0 < C_F < \delta) u_2 \\
&+ I(\frac{a^{(k)} H_{U_1}^{(k)}}{\sigma + b^{(k)} H_J^{(k)}} < \eta) I(C_F > \delta) u_3 \\
&+ I(a == 0) C_M] \\
&+ b^{(k)} C_J
\end{aligned} \tag{18}$$

$$u_J^{(k)}(a^{(k)}, b^{(k)}) = -u_U^{(k)}(a^{(k)}, b^{(k)}). \tag{19}$$

where $u_1 = \frac{\eta(\sigma + N_J H_J^{(k)})}{H_{U_1}^{(k)}} C_U$, $u_2 = \frac{\eta(\sigma + N_J H_J^{(k)})}{H_{U_2}^{(k)}} C_U + C_F$, $u_3 = \frac{H_{U_2}^{(k)}}{H_{U_1}^{(k)}} (\frac{\eta(\sigma + N_J H_J^{(k)})}{H_{U_1}^{(k)}} C_U)$.

Q-learning is a model-free value iteration algorithm, which is useful when a system model is hard to construct, hence, we will employ it to find out the optimal strategy for maximizing U's long-term utility. We employ a Q-learning based algorithm with a learning rate β, which is described in Algorithm 1. At time slot k, $Q(s^{(k)}, a^{(k)})$ denotes the Q-function when the system state is $s^{(k)} = [b^{(k-1)}, H_U^{(k-1)}, H_J^{(k-1)}]$ and the action of U is $a^{(k)}$. Besides, the highest value of the system state is denoted as $V(s^{(k)})$. U updates its Q-function as follows,

$$\begin{aligned}
Q(\mathbf{s}^{(k)}, a^{(k)}) \leftarrow (1-\beta) Q(\mathbf{s}^{(k)}, a^{(k)}) + \\
\beta[u_U^{(k)} + \gamma V(\mathbf{s}^{(k+1)})].
\end{aligned} \tag{20}$$

$$V(\mathbf{s}^{(k)}) = \max_{a \in \mathcal{A}} Q(\mathbf{s}^{(k)}, a). \tag{21}$$

With respect to the choice of transmission power of U at time slot k, for obtaining the locally optional power control strategy, it is assumed that it follows a ϵ-greedy policy, whose probability distribution is as the following,

$$Pr(a^{(k)} = a^*) = \begin{cases} 1-\epsilon, & a^* = \underset{a \in \mathcal{A}}{\operatorname{argmax}} Q(\mathbf{s}^{(k)}, a) \\ \frac{\epsilon}{N_U}, & o.w. \end{cases} \tag{22}$$

Algorithm 1 Q-learning based power control scheme

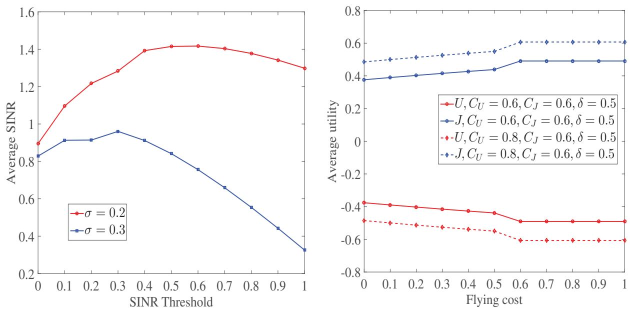
1: Initialize $\beta = 0.8$, $\gamma = 0.75$, $\epsilon = 0.1$, $b^0 = 0$, $(H_U)^0 = 0$, $(H_J)^0 = 0$, $Q(s^0, a^0) = 0$.
2: **for** k=1,2,... **do**
3: Update the system state $s^{(k)} = [b^{(k-1)}, H_U^{(k-1)}, H_J^{(k-1)}]$;
4: Choose $a^{(k)}$ according to ϵ-greedy;
5: Observe $b^{(k)}$, $H_U^{(k-1)}$, $H_J^{(k-1)}$ and obtain instantaneous utility $u_U^{(k)}(a^{(k)}, b^{(k)})$;
6: Update $Q(\mathbf{s}^{(k)}, a^{(k)})$ via (20);
7: Update $V(\mathbf{s}^{(k)})$ via (21);
8: **end for**

VI. Numerical Simulation Results

Numerical simulations are utilized to evaluate the performance of derived NE in the static game $\mathbb{G}$ and the optimal anti-jamming strategy in the dynamic game $\mathbb{G}'$. First, paying attention to the influence of chosen SINR threshold η and flying cost C_F on the obtained NE, the performance of NE is examined under different cases. Then, we also evaluate the proposed Q-learning based anti-jamming strategy, compared with a ϵ-greedy benchmark strategy.

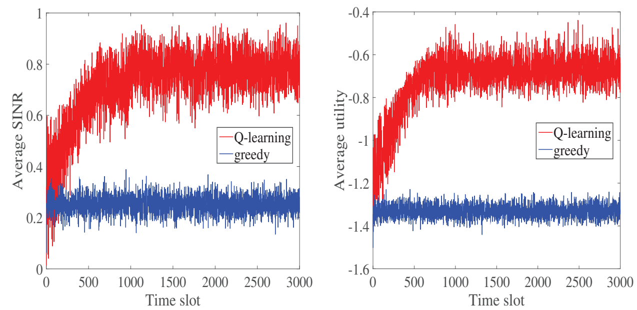
In the first numerical simulation, we set the channel parameters according to the C-band (5060MHz) Mountainous, Telluride scenario in [13], $a = b = 10$ (dBW), $D_1 = 16$, $D_2 = 6$ (km), $A_0 = 119.7$, $n = 1.7$, $\sigma_X = 2.8$, $d_{min} = 3.4$ (km), $K_{d_{min}} = 29.9$, $n_k = -0.02$, $\sigma_Y = 2.2$, the flying velocity of U is $v_U = 80$ (m/s), then the obtained Air-Ground channel power gain H_{U_1} and H_{U_2} are normalized into the range of $[0, 1]$. Besides, the transmitting cost of U and J are set as two cases: 1) $C_U = C_J = 0.6$, 2) $C_U = 0.8, C_J = 0.6$.

As shown in Fig. 2(a), the average SINR of U is examined according to the chosen SINR threshold η. With the increase



(a) SINR of U. (b) Utility of U.

Fig. 2. Performance of NE in the static jamming game $\mathbb{G}$.



(a) SINR of U. (b) Utility of U.

Fig. 3. Performance of the Stackelberg dynamic game $\mathbb{G}'$, with $\beta = 0.8$, $\gamma = 0.75$, $\epsilon = 0.1$, $v_U = 80$, $C_U = 0.8, C_J = 0.6$, $C_F = 0.6$, $\eta = 0.6$, $\delta = 0.5$.

of η, the average SINR first increases and then decreases. It indicates that the value of pre-defined SINR threshold is critical for the average SINR of U. Besides, an intermediate value of η is better for U to get a larger SINR. In addition, we investigate the performance of derived NE, focusing on the relationship between the utility of U (or J) and the flying cost C_F. As shown in Fig. 2(b), in a single time slot, when C_U and C_J are given, the utility of U decreases as C_F grows, meanwhile the utility of J gets improved. This phenomenon indicates that if the location L_2 is too far away from the location L_1, then the flying cost is too high. Thus, transmitting the signal at L_1 may be wise for U to obtain a larger utility.

In the second simulation, to get a proper performance, we set the simulation parameters as the following, $\beta = 0.8$, $\gamma = 0.75$, $\epsilon = 0.1$, $v_U = 80$, $C_U = 0.8, C_J = 0.6$, $C_F = 0.6$, $\eta = 0.6$, $\delta = 0.5$ and the normalized values of $a, b, H_{U_1}, H_{U_2}, H_J \in [0, 1]$. A ϵ-greedy strategy is used as a benchmark. In the ϵ-greedy benchmark algorithm, based on the jamming history of J, the anti-jamming power choice of U is to maximize its instantaneous utility other than the long-term ones. Meanwhile, all the jamming power choice of J are chosen for maximize J's immediate utility regarding the last time slot system state.

As shown in Fig. 3(a), in the Q-learning based method, the average SINR of U converges to 0.76 after 1000 time slots. Whereas in the ϵ-greedy method, it is about 0.25. In Fig. 3(b), we can get that the average utility of two methods are -0.68 and -1.37, respectively. Hence, the performance of our proposed Q-learning based anti-jamming scheme is relatively better than the ϵ-greedy benchmark method.

VII. Conclusion

In this work, we have proposed a static jamming game and a Stackelberg dynamic game in UAV systems. In the static game, the characteristics of the Air-Ground radio propagation, flying cost and SINR threshold were specially considered. A closed-form NE was derived and its uniqueness was also proved. Static NE simulation results revealed that some intermediate values of SINR threshold are better for the legitimate UAV. Meanwhile, it is unwise to resort to improve the downlink wireless channel via solely flying toward to the ground station when the flying cost is relatively high. In the second game, we employed a Q-learning algorithm to obtain an optimal strategy in a situation that the system parameters are hard to determine. The performance of Q-learning based power control scheme was slightly improved, comparing with a classical ϵ-greedy benchmark method.

References

[1] US Department of Transportation, “Unmanned aircraft system (UAS) service demand 2015 - 2035: literature review & projections of future usage,” *Tech. rep., v.1.0, DOT-VNTSC-DoD-13-01*, 2014.

[2] S. Hayat, E. Yanmaz, and R. Muzaffar, “Survey on unmanned aerial vehicle networks for civil applications: a communications viewpoint,” *IEEE Communications Surveys and Tutorials*, vol. 18, no. 4, pp. 2624–2661, 2016.

[3] Battelle, “Battelle DroneDefender. https://www.battelle.org/government-offerings/national-security/aerospace-systems/counter-UAS-technologies/dronedefender,” 2016.

[4] Y. Zeng, R. Zhang, and T. J. Lim, “Wireless communications with unmanned aerial vehicles: opportunities and challenges,” *Communications Magazine*, vol. 54, no. 5, pp. 36–42, 2016.

[5] L. Xiao, J. Liu, Q. Li, N. B. Mandayam, and H. V. Poor, “User-centric view of jamming games in cognitive radio networks,” *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 12, pp. 2578–2590, 2015.

[6] L. Jia, F. Yao, Y. Sun, Y. Niu, and Y. Zhu, “Bayesian stackelberg game for antijamming transmission with incomplete information,” *IEEE Communications Letters*, vol. 20, no. 10, pp. 1991–1994, 2016.

[7] T. Song, W. E. Stark, T. Li, and J. K. Tugnait, “Optimal multiband transmission under hostile jamming,” *IEEE Transactions on Communications*, vol. 64, no. 9, pp. 4013–4027, 2016.

[8] M. K. Hanawal, M. J. Abdel-Rahman, and M. Krunz, “Joint adaptation of frequency hopping and transmission rate for anti-jamming wireless systems,” *IEEE Transactions on Mobile Computing*, vol. 15, no. 9, pp. 2247–2259, 2016.

[9] C. J. Watkins and P. Dayan, “Q-learning,” *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[10] A. Sanjab, W. Saad, and T. Başar, “Prospect theory for enhanced cyber-physical security of drone delivery systems: A network interdiction game,” *arXiv preprint arXiv:1702.04240*, 2017.

[11] S. Bhattacharya and T. Başar, “Game-theoretic analysis of an aerial jamming attack on a uav communication network,” in *Proc. of IEEE American Control Conference*, 2010.

[12] J. Parras, J. del Val, S. Zazo, J. Zazo, and S. V. Macua, “A new approach for solving anti-jamming games in stochastic scenarios as pursuit-evasion games,” in *Proc. of IEEE Statistical Signal Processing Workshop*, 2016.

[13] R. Sun and D. Matolak, “Air-Ground channel characterization for unmanned aircraft systems-Part II: hilly & mountainous settings,” *Transactions on Vehicular Technology*, vol. 66, no. 3, pp. 1913–1925, 2017.