User-Centric View of Unmanned Aerial Vehicle Transmission Against Smart Attacks

Liang Xiao, *Senior Member, IEEE,* Caixia Xie, Minghui Min, *Student Member, IEEE*
and Weihua Zhuang, *Fellow, IEEE*

***Abstract*—Unmanned aerial vehicle (UAV) systems are vulnerable to smart attackers who are selfish and subjective end-users and use smart radio devices to change their attack types and policies based on the ongoing UAV transmission and network states. In this paper, we apply prospect theory to formulate a subjective smart attack game for the UAV transmission, in which a smart attacker Eve makes subjective decisions to choose the attack type such as jamming, spoofing and eavesdropping without knowing the attack detection accuracy of the UAV system and the UAV transmit power on multiple radio channels is chosen to resist smart attacks. Reinforcement learning based UAV power allocation strategies are proposed to achieve the optimal power allocation against smart attacks without knowing the attack model and the channel model in the dynamic game. A deep Q-learning based UAV power allocation strategy combines Q-learning and deep learning to accelerate the learning speed for the case with a large number of channel states and attack modes. Simulation results show that our proposed UAV power allocation strategy can suppress the attack motivation of subjective smart attackers and increase the secrecy capacity and the utility of the UAV system.**

***Index Terms*—Smart attacks, UAV, prospect theory, game theory, reinforcement learning.**

I. Introduction

Unmanned aerial vehicles (UAVs) have proliferation applications such as goods delivery, radio network connectivity enhancement, and the collection of news and surveil events. Due to the broadcast nature of the radio transmission, UAV systems are vulnerable to spoofing, jamming and eavesdropping attacks that are three physical layer attacks in wireless networks. Attackers from the vicinal air or ground aim to obtain illegal network access, threaten user privacy of the UAV systems and launch denial of service (DoS) attacks [1], [2]. In this paper, we focus on the physical-layer security methods to address jamming, eavesdropping and spoofing in UAV systems that are incorporated with the upper-layer security mechanisms, instead of replacing them. For example, the traditional upper-layer encryption methods cannot address all the eavesdropping challenges in UAV systems, due to the dynamic network topology and energy constraint of UAVs, as well as the growing computational capability of eavesdroppers [3]. To this end, the physical-layer encryption protocols are efficient to protect UAV systems from eavesdropping [3], [4].

By applying smart and programmable radio devices such as universal software radio peripherals and wireless open-access research platforms, a smart attacker Eve can observe the ongoing UAV radio transmission status between the UAV Alice and the mobile ground unit (MGU) Bob and then choose the attack type and mode accordingly [5]. For example, a smart attacker Eve sends spoofing signals if she has a similar channel state with Alice [6] and sends jamming signals if she is very close to Bob. Compared with the traditional single-mode passive attackers each performing a single type of attacks, a smart attacker can be more harmful to the UAV transmission by reducing the secrecy capacity of the UAV system and cause more identity based attacks without being detected.

Game theory has been used to study jamming [7] and spoofing [8] in wireless communication networks, assuming that each player is rational and chooses its action to maximize its own expected utility according to the widely used expected utility theory (EUT) [9]. However, an attacker or UAV operator makes a subjective decision, if he or she does not know whether the attacker can be detected successfully and thus the decision sometimes deviates from the EUT results [10]. The Nobel prize-winning prospect theory (PT) applies the probability weighting function and value function to model the subjective decision-making processes such as the probability evaluation distortion and the facts that people tend to be risk averse regarding gains and risk seeking regarding losses. Prospect theory has been used to explain the anti-jamming communication in cognitive radio networks in [11] and the detection of advanced persistent threats launched by subjective attackers in [12].

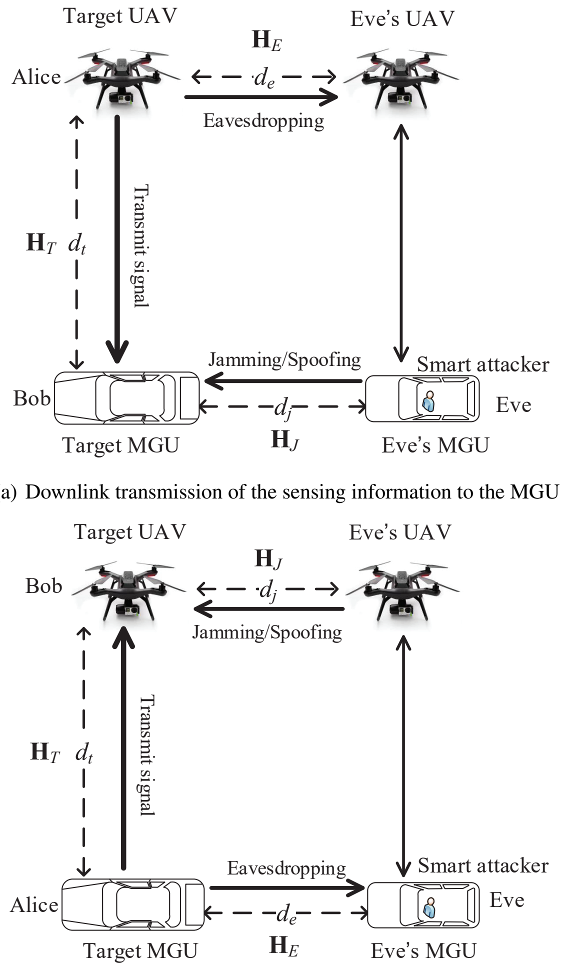
In this paper, we apply prospect theory to investigate smart attacks against the UAV transmission launched by a subjective and selfish attacker Eve. Eve makes subjective decisions regarding the attack policy, as she is not sure whether her attack will be detected by the UAV system. In the PT-based smart attack game, in the first scenario as shown in Fig. 1 (a), the UAV called Alice aims to send messages with sensing information to the MGU Bob. The smart attacker Eve who controls a UAV, a MGU or both, chooses to block Bob from

This paper was presented in part at IEEE ICUWB 2016 [28], and was supported in part by National Natural Science Foundation of China under Grant 61671396 and 91638204, and in part by the open research fund of National Mobile Communications Research Laboratory, Southeast University (No.2018D08).

Liang Xiao is with the Department of Communication Engineering, Xiamen University, Xiamen 361005, China, and also with the National Mobile Communications Research Laboratory, Southeast University, China. E-mail: lxiao@xmu.edu.cn.

Caixia Xie and Minghui Min are with the Department of Communication Engineering, Xiamen University, Xiamen 361005, China.

Weihua Zhuang is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, N2L 3G1. E-mail: wzhuang@uwaterloo.ca.



(a) Downlink transmission of the sensing information to the MGU

(b) Uplink transmission of the control information to the UAV

Fig. 1. The UAV transmission between a target UAV and a MGU against a smart attacker Eve with a compromised UAV and MGU.

receiving Alice's signals, send spoofing signals with Alice's identity or eavesdrop Alice's signals. Alice observes the UAV channel state and the previous transmission performance, and chooses the transmit power accordingly. In the second scenario in Fig. 1 (b), the MGU Alice sends messages with the control information to the UAV Bob against Eve. The Nash equilibria (NEs) of the PT-based smart attack game are derived and the conditions under which the NEs exist are provided to disclose how the utility of the UAV increases with the subjectivity of Eve. Reinforcement learning techniques can derive the optimal UAV power allocation strategy via trial-and-error in the dynamic smart attack game without being aware of the attack model and the UAV network model, as the repeated power allocation can be formulated as a Markov decision process (MDP). A Q-learning based power allocation algorithm is developed based on the quality function or Q-function for each state-action pair of the UAV.

On the other hand, Eve can attack the UAV Alice accordingly once understanding the security mechanism. More specifically, Eve can infer the UAV security mechanism, then deliberately manipulate Alice to use a specific transmission policy, and then attack her accordingly. Therefore, a WoLF-PHC (Win or Learn Faster-Policy Hill Climbing) based power allocation algorithm can be designed to introduce uncertainties in the UAV transmission to fool the smart attackers, to improve the UAV security.

The UAV transmission against smart attackers covers a large number of attack status, radio channel states and feasible transmit power levels, which significantly increases the learning speed required by the Q-leaning or WoLF-PHC based schemes. Therefore, we further propose a deep Q-network (DQN) [13] based power allocation strategy, which introduces deep learning to accelerate the learning speed of the WoLF-PHC based power allocation strategy. More specifically, a combination of Q-learning and deep learning can be used by the UAV to compress the state space of Q-learning to address the high dimension problem in the power allocation against smart attacks. Simulation results show that the DQN-based scheme can increase both the signal-to-interference-plus-noise ratio (SINR) of the UAV signals, the secrecy capacity and the utility of the UAV against smart attacks.

The main contributions of this work can be summarized as follows:

(1) We formulate a PT-based smart attack game for an UAV to provide a user-centric view of smart attacks and provide the NEs of the PT-based game to show the condition that the UAV transmission benefits from the subjectivity of Eve;

(2) We propose a WoLF-PHC based UAV power allocation strategy to address smart attacks in the dynamic game to achieve the optimal power allocation on multiple frequency channels, without the knowledge of smart attack model and UAV channel model. Simulation results show that this scheme can increase the secrecy capacity and the utility of the UAV against subjective smart attackers, as compared with the Q-learning based power allocation strategy;

(3) A DQN-based power allocation strategy is developed to further accelerate the learning speed of the UAV for the case with a large number of frequency channels and transmit power levels.

The remainder of this paper is organized as follows. We review related work in Section II and present the system model in Section III. We formulate the PT-based smart attack game in Section IV, and discuss the NEs of the game in Section V. We propose the reinforcement learning based UAV power allocation strategies in Section VI. Simulation results are discussed in Section VII and conclusions are drawn in Section VIII.

II. Related Work

Smart attacks are investigated with game theory in [5]. An EUT-based mobile offloading game is formulated among a mobile device that chooses the offloading rate, a security agent that selects whether to apply the higher-layer security mechanism, and a smart attacker who can perform either jamming or spoofing. A noncooperative game between a mobile user and a malicious node that can eavesdrop, jam, or both to reduce the capacity of the user is formulated in [14], where a fictitious play-based algorithm is proposed to derive the NE of the mixed-strategy game. The game between a transmitter that chooses to send data or artificial interference, and an adversary that selects to passively eavesdrop or actively jam is investigated in [15]. A stochastic game formulated in [16] provides insights for secret and reliable communication against both jamming and eavesdropping. The interaction between wireless nodes is formulated in [17], where each node is either a selfish user who chooses the transmit power or malicious user who attempts to minimize the throughput of the other node with minimum transmission cost. A two-player general sum Bayesian game as formulated in [18] illustrates how a smart jammer that acts either as a cheater to obtain more network resource or a saboteur to cause serious damage can impact the LTE network.

Prospect theory is applied in [19] to capture the user subjectivity in data pricing and channel allocation in cognitive radio networks. The PT-based jamming game formulated in [11] discloses the impact of subjective views of the jammer and end-user under uncertain channel power gains in cognitive radio networks. A PT-based data pricing model designed in [20] shows that end-user deviation from EUT degrades system throughput performance. The PT-based spectrum investment game in [21] shows that a subjective secondary operator tends to achieve a smaller possible gain with a lower risk, rather than a larger possible gain with a higher risk. A PT-based cloud storage defense game between a defender and an advanced persistent threat attacker is investigated in [12] to disclose the influence of the attacker's subjectivity on the attack and scan intervals. The subjective vendor and attacker game as studied in [22] shows that the subjective decision making processes of the vendor and attacker lead to a longer delivery time in UAV delivery systems.

Multi-agent Q-learning based power allocation for cognitive femtocells as presented in [23] shows that the cooperative learning based scheme outperforms the independent learning with a faster convergence rate and a larger aggregate capacity. The Q-learning based joint resource allocation and power control scheme in [24] reveals the learned strategies with neighbours to accelerate the convergence speed and improve the system capacity in femtocell. The Q-learning based dynamic power management as proposed in [25] manipulates the idle periods of processor cores to achieve a tradeoff between the power consumption and system performance for multicore processors. The neural networks based resource allocation scheme proposed in [26] regulates MAC-layer parameters for the distributed IEEE 802.11 system to improve the throughput. A support vector machine based resource allocation scheme proposed in [27] achieves the quality of service goals and minimizes the incurred cost without in-depth knowledge about the cloud system internals.

The proposed theoretic study on smart attacks in [28] formulated a game between a smart attacker who chooses the attack mode and a mobile user who determines whether to initiate the higher-layer security or only exploit the physical-layer security mechanism. Compared with our previous work in [28], we study the UAV power allocation with multiple channels against a smart attacker Eve who controls both a compromised UAV and a MGU. We investigate how the UAV channel states change the attack mode of Eve, and propose a DQN-based power allocation strategy to improve the secrecy capacity of Alice in the UAV transmission against Eve who makes subjective decisions to choose the attack mode.

III. System Model

A. Network Model

We consider the transmission between a legitimate UAV, Alice, who aims to send or receive surveillance data and transport message to or from the serving mobile ground unit (MGU) Bob located d_t distance away over B radio frequency channels. A smart attacker, Eve, uses a smart and programmable radio device to launch spoofing, eavesdropping or jamming, and chooses the attack pattern such as in terms of the jamming power. Eve cannot simultaneously launch eavesdropping and jamming attacks, because her jamming signals have to be sent at the same frequency with Alice, which prevents her from obtaining Alice's messages.

As shown in Fig. 1, Eve can change her location to attack the transmission between Alice and Bob. For example, Eve can use a compromised UAV that is close to Alice, the target UAV to eavesdrop her downlink transmission of the sensing information to the MGU. In addition, Eve can apply a compromised MGU to jam or spoof the downlink transmission. On the other hand, Eve can launch jamming attacks from the compromised UAV to block the target UAV in the uplink transmission of the control information. We take the downlink UAV transmission of the sensing information from Alice to Bob as an example in this paper, but our work can be extended to the uplink UAV transmission.

Eve can also locate the compromised MGU and UAV to attack more efficiently. For instance, Eve moves the compromised UAV to the neighboring area of Alice to reduce the detection accuracy by Alice, and the compromised MGU to Bob's neighborhood before sending jamming signals. Eve is d_j distance away from Bob and her compromised UAV is d_e away from Alice.

By using the smart radio devices, Eve can estimate the ongoing UAV transmission status and the channel state. Alice decides the transmit power over B radio channels at time k, denoted by $\mathbf{x}^{(k)} = [x_i^{(k)}]_{1\le i\le B} \in \mathbf{X}$ following a total power constraint denoted by P_T, where the power allocation set $\mathbf{X}$ is given by

$$\mathbf{X} = \left\{ [x_i]_{1\le i\le B} \middle| 0 \le x_i \le P_T; \sum_{i=1}^{B} x_i \le P_T \right\}. \tag{1}$$

Eve chooses spoofing, eavesdropping or jamming to prevent reliable and secret communications from Alice to Bob. Eve can also send jamming signals to Bob over B channels with a total power constraint denoted by P_J. The attack mode at time k is denoted by $\mathbf{y}^{(k)} = [y_i^{(k)}]_{1\le i\le B} \in \mathbf{Y}$ with

$$\mathbf{Y} = \left\{ [y_i]_{1\le i\le B} \middle| y_i \le P_J; \sum_{i=1}^{B} y_i \le P_J \right\}. \tag{2}$$

If $\mathbf{y}^{(k)} > 0$, the compromised MGU sends jamming signals with power $y_i^{(k)}$ on channel i; if $\mathbf{y}^{(k)} = -\mathbf{1}$, Eve eavesdrops

TABLE I
LIST OF NOTATIONS

$d_{t/e/j}$	Distance of transmission/wiretap/interference path
ρ	Path loss exponent of radio channel
B	Number of radio channels
$h_{T,i}^{(k)}$	Power gain of transmission channel i between Alice and Bob at time k
$h_{E,i}^{(k)}$	Power gain of at wiretap channel i at time k
$h_{J,i}^{(k)}$	Power gain of interference channel i at time k
$x_i^{(k)}$	Transmit power at channel i at time k
$y_i^{(k)}$	Jamming power at channel i at time k
$P_{T/J}$	Total power constraints of the UAV/attacker
η	Miss detection rate of jamming attack
β	Miss detection rate of spoofing attack
$\alpha_{A/E}$	Objective weight of Alice/Eve
μ	Energy consumption factor
σ	Noise power
C_m	Miss detection cost of spoofing attack
L	Nonzero quantization level of detection accuracy
$U_{A/E}^{EUT/PT}$	EUT/PT-based utility of Alice/Eve

Alice at time k; if $\mathbf{y}^{(k)} = -\mathbf{2}$, Eve spoofs Bob; and if $\mathbf{y}^{(k)} = \mathbf{0}$, Eve does not attack.

B. Channel Model

The channel power gain between Alice and Bob at time k is denoted by $\mathbf{H}_T^{(k)} = [h_{T,i}^{(k)}]_{1\le i\le B}$, where $h_{T,i}^{(k)}$ is the power gain of transmission channel i between Alice and Bob at that time. Similarly, the wiretap channel gain from Alice to the compromised UAV at time k is $\mathbf{H}_E^{(k)} = [h_{E,i}^{(k)}]_{1\le i\le B}$, where $h_{E,i}^{(k)}$ is the power gain of wiretap channel i at time k. The jamming channel gain from Eve to Bob is denoted by $\mathbf{H}_J^{(k)} = [h_{J,i}^{(k)}]_{1\le i\le B}$, where $h_{J,i}^{(k)}$ is the power gain of jamming channel i at time k.

Let d_0 be the reference distance of the UAV transmission and ρ be the path loss exponent of the radio channel model. The path loss of the transmission channel with the transmitter-receiver distance d denoted by PL can be modeled by [29]

$$PL(\mathrm{dB}) = \xi(\mathrm{dB}) + 10\rho\lg(\frac{d}{d_0}), \quad d > d_0 \tag{3}$$

where ξ is a constant due to the transmitter and receiver antenna gains and path loss at d_0, the path loss exponent ρ depends on the radio propagation environment. Both the Alice-Bob link and the Alice-Eve UAV link have $\rho = 2$ due to the approximate free-space propagation, and the Eve-Bob channel has $\rho = 4$ according to a two-ray model. Let $r_i \sim N(0,1)$ follow the normal distribution, and we have $h_{g,i}^{(k)} = r_i^{(k)}/PL_g$, where $g = T, E$ or J. Table I summarizes the notation used in this paper.

IV. PT-BASED SMART ATTACK GAME

The interaction between Eve and the UAV transmission can be formulated as a PT-based smart attack game, denoted by $\mathcal{G}$, in which Eve chooses her attack type and mode and Alice decides the transmit power over B radio channels. Bob can apply physical-layer security mechanisms to detect spoofing attacks [30] with a miss detection rate denoted by β, and the jamming detection algorithm such as the packet delivery ratio and received signal strength based detection as developed in [31] with a miss detection rate denoted by η. Alice can hardly detect the eavesdropping of Eve.

Alice aims to increase the SINR of the signal received by Bob and thus improve the UAV transmission against the smart attacker Eve. The utility of Alice averaged over the MGU detections is given by

$$u_A^{(k)}(\mathbf{x},\mathbf{y}) = \begin{cases} \sum_{i=1}^{B}\left(h_{T,i}^{(k)} - h_{E,i}^{(k)}\right)x_i \\ \qquad -\mu\sum_{i=1}^{B}x_i, & \mathbf{y} = -\mathbf{1} \quad \text{(4a)} \\ \sum_{i=1}^{B}h_{T,i}^{(k)}x_i - \beta C_m \\ \qquad -\mu\sum_{i=1}^{B}x_i, & \mathbf{y} = -\mathbf{2} \quad \text{(4b)} \\ (1-\eta)\sum_{i=1}^{B}h_{T,i}^{(k)}x_i + \eta\sum_{i=1}^{B}\frac{h_{T,i}^{(k)}x_i}{\sigma + h_{J,i}^{(k)}y_i} \\ \qquad -\mu\sum_{i=1}^{B}x_i, & \mathbf{y} \succeq \mathbf{0} \quad \text{(4c)} \end{cases}$$

where the energy consumption factor μ corresponds to the importance of the transmission energy regarding the security risk, σ is the noise power of Bob, and the UAV cost C_m depends on the miss detection probability of spoofing attacks. The first term in the right-hand-side of Eq. (4a) corresponds to the secrecy capacity of the UAV system.

For simplicity, the detection accuracy of the UAV system is quantized into $L+1$ levels, with the detection accuracy l/L having probability $\beta_l = \Pr(\beta = l/L)$ and $\eta_l = \Pr(\eta = l/L)$. By definition, we have $\beta_l(\eta_l) > 0$ and $\sum_{l=0}^{L}\beta_l(\eta_l) = 1$. The EUT-based utility of Alice averaged over L non-zero detection error rates denoted by U_A^{EUT} is given by

$$U_A^{EUT}(\mathbf{x},\mathbf{y}) = \begin{cases} \sum_{i=1}^{B}\left(h_{T,i}^{(k)}-h_{E,i}^{(k)}\right)x_i & \\ \qquad -\mu\sum_{i=1}^{B}x_i, & \mathbf{y}=-\mathbf{1} \qquad \text{(5a)} \\ \sum_{i=1}^{B} h_{T,i}^{(k)}x_i - \frac{C_m}{L}\sum_{l=0}^{L} l\beta_l & \\ \qquad -\mu\sum_{i=1}^{B}x_i, & \mathbf{y}=-\mathbf{2} \qquad \text{(5b)} \\ \sum_{i=1}^{B} h_{T,i}^{(k)}x_i - \frac{1}{L}\sum_{l=0}^{L} l\eta_l \sum_{i=1}^{B} \frac{h_{T,i}^{(k)}h_{J,i}^{(k)}x_iy_i}{\sigma+h_{J,i}^{(k)}y_i} & \\ \qquad -\mu\sum_{i=1}^{B}x_i, & \mathbf{y}\succeq\mathbf{0}. \qquad \text{(5c)} \end{cases}$$

The expected utility of Eve, denoted by U_E^{EUT}, is the opposite of Alice's utility, i.e., $U_E^{EUT} = -U_A^{EUT}$.

Prospect theory can be used to capture the subjective decision-making processes of Eve and Alice. According to Prelec probability weighting function [32], the subjective probability viewed by Eve (or Alice) denoted by w_E (or w_A) is given by

$$w_r(p) = \exp\left(-(-\ln p)^{\alpha_r} \right) \tag{6}$$

where $r = A, E$, $\alpha_r \in (0, 1]$ is the objective weight of Eve (or Alice), and p is the objective probability. The probability weighting function describes how a subjective player underweighs the high-probability event and overweighs the low probability event. For example, $w_r(p) < p$ if p is close to 1, and $w_r(p) > p$ if p is close to 0.

If Eve and Alice hold subjective views to make decisions under uncertain security performance, their decisions may deviate from the EUT-based results. The PT-based utility of Alice, denoted by U_A^{PT}, is based on the EUT-based utility U_A^{EUT} by replacing the objective probability of the miss detection rate with the subjective probability,

$$U_A^{PT}(\mathbf{x},\mathbf{y}) = \begin{cases} \sum_{i=1}^{B}\left(h_{T,i}^{(k)}-h_{E,i}^{(k)}\right)x_i & \\ \qquad -\mu\sum_{i=1}^{B}x_i, & \mathbf{y}=-\mathbf{1} \qquad \text{(7a)} \\ \sum_{i=1}^{B} h_{T,i}^{(k)}x_i - \frac{C_m}{L}\sum_{l=0}^{L} lw_A(\beta_l) & \\ \qquad -\mu\sum_{i=1}^{B}x_i, & \mathbf{y}=-\mathbf{2} \qquad \text{(7b)} \\ -\frac{1}{L}\sum_{l=0}^{L} lw_A(\eta_l) \sum_{i=1}^{B} \frac{h_{T,i}^{(k)}h_{J,i}^{(k)}x_iy_i}{\sigma+h_{J,i}^{(k)}y_i} & \\ +\sum_{i=1}^{B} h_{T,i}^{(k)}x_i - \mu\sum_{i=1}^{B}x_i, & \mathbf{y}\succeq\mathbf{0}.\text{(7c)} \end{cases}$$

Similarly, the PT-based utility of Eve, denoted by U_E^{PT}, is given by

$$U_E^{PT}(\mathbf{x},\mathbf{y}) = \begin{cases} \sum_{i=1}^{B}\left(h_{E,i}^{(k)}-h_{T,i}^{(k)}\right)x_i & \\ \qquad +\mu\sum_{i=1}^{B}x_i, & \mathbf{y}=-\mathbf{1} \qquad \text{(8a)} \\ -\sum_{i=1}^{B} h_{T,i}^{(k)}x_i + \frac{C_m}{L}\sum_{l=0}^{L} lw_E(\beta_l) & \\ \qquad +\mu\sum_{i=1}^{B}x_i, & \mathbf{y}=-\mathbf{2} \qquad \text{(8b)} \\ \frac{1}{L}\sum_{l=0}^{L} lw_E(\eta_l) \sum_{i=1}^{B} \frac{h_{T,i}^{(k)}h_{J,i}^{(k)}x_iy_i}{\sigma+h_{J,i}^{(k)}y_i} & \\ -\sum_{i=1}^{B} h_{T,i}^{(k)}x_i + \mu\sum_{i=1}^{B}x_i, & \mathbf{y}\succeq\mathbf{0}.\text{(8c)} \end{cases}$$

For simplicity of notation, the time index k is omitted if no confusion results.

V. NE of the PT-Based Smart Game

The subjective smart attacker Eve should make a balance between the risk to be detected and the damage to the UAV system, and she chooses her policy to maximize the PT-based utility instead of the expected utility. The Nash equilibrium of the smart attack game $\mathcal{G}$, denoted by $(\mathbf{x}^*, \mathbf{y}^*)$, provides the best-response of each player if the opponent chooses the NE strategy. By definition, we have

$$U_A^{PT}(\mathbf{x}^*,\mathbf{y}^*) \geq U_A^{PT}(\mathbf{x},\mathbf{y}^*),\ \forall \mathbf{x}\in\mathbf{X} \tag{9}$$

$$U_E^{PT}(\mathbf{x}^*,\mathbf{y}^*) \geq U_E^{PT}(\mathbf{x}^*,\mathbf{y}),\ \forall \mathbf{y}\in\mathbf{Y}. \tag{10}$$

Let $\mathbf{I}_{(m,n)}$ be an m-dimensional all-zero vector except 1 at the n-th element.

Theorem 1. *The PT-based smart attack game $\mathcal{G}$ has an NE $(P_T\mathbf{I}_{(B,j^*)}, -\mathbf{1})$ if*

$$\begin{cases} \max_{1\leq i\leq B}\{h_{T,i}-h_{E,i}\} > \mu & \text{(11)} \\ Lh_{E,j^*}P_T > \max\Bigg\{ C_m\sum_{l=0}^{L} lw_E(\beta_l), & \\ \qquad \frac{h_{T,j^*}h_{J,j^*}P_TP_J}{\sigma+h_{J,j^*}P_J}\sum_{l=0}^{L} lw_E(\eta_l)\Bigg\} & \text{(12)} \end{cases}$$

where

$$j^* = \arg\max_{1\leq v\leq B}(h_{T,v}-h_{E,v}). \tag{13}$$

Proof: By (7a), if (11) holds, we have

$$U_A^{PT}(P_T\mathbf{I}_{(B,j^*)}, -\mathbf{1}) = P_T \max_{1\leq i\leq B}\{h_{T,i}-h_{E,i}-\mu\}$$
$$\geq \sum_{i=1}^{B}(h_{T,i}-h_{E,i}-\mu)x_i = U_A^{PT}(\mathbf{x},-\mathbf{1}). \tag{14}$$

If (12) holds, we have

$$\begin{aligned}
&U_E^{PT}(P_T\mathbf{I}_{(B,j^*)},-\mathbf{1}) = -P_T\left(h_{T,j^*}-h_{E,j^*}-\mu\right)\\
&> \max\Bigg\{(\mu-h_{T,j^*})P_T+\frac{C_m}{L}\sum_{l=0}^{L} lw_E(\beta_l),\\
&\quad(\mu-h_{T,j^*})P_T+\frac{1}{L}\frac{h_{T,j^*}h_{J,j^*}P_TP_J}{\sigma+h_{J,j^*}P_J}\sum_{l=0}^{L} lw_E(\eta_l)\Bigg\}\\
&= \max\left\{U_E^{PT}(P_T\mathbf{I}_{(B,j^*)},-\mathbf{2}),U_E^{PT}(P_T\mathbf{I}_{(B,j^*)},\mathbf{y}\succeq\mathbf{0}\right\}.
\end{aligned}\tag{15}$$

Thus (9) and (10) hold, indicating that $(P_T\mathbf{I}_{(B,j^*)},-\mathbf{1})$ is an NE of $\mathcal{G}$. ■

Corollary 1. *If (11) and (12) hold, the utility of Alice in the PT-based smart attack game $\mathcal{G}$ is*

$$U_A^{EUT} = P_T\max_{1\le i\le B}\{h_{T,i}-h_{E,i}\}-\mu P_T. \tag{16}$$

Remark: If the wiretap channel is in a good condition, the smart attacker Eve eavesdrops the communications between Alice and Bob, which can avoid being detected by the detection system. If there exists a transmission channel better than the wiretap channel, i.e., $\max_i\{h_{T,i}-h_{E,i}\}>\mu$, Alice allocates all her power to the channel with maximum channel gain gap between the transmission channel and the wiretap channel to maximize the secrecy capacity.

Theorem 2. *The PT-based smart attack game $\mathcal{G}$ has an NE $(P_T\mathbf{I}_{(B,q^*)},-\mathbf{2})$ if*

$$\begin{cases}
\max\limits_{1\le i\le B}\{h_{T,i}\}>\mu & (17)\\
C_m\sum\limits_{l=0}^{L} lw_E(\beta_l)>\max\Bigg\{Lh_{E,q^*}P_T, & \\
\qquad\dfrac{h_{T,q^*}h_{J,q^*}P_TP_J}{\sigma+h_{J,q^*}P_J}\sum\limits_{l=0}^{L} lw_E(\eta_l)\Bigg\} & (18)
\end{cases}$$

where

$$q^* = \arg\max_{1\le v\le B} h_{T,v}. \tag{19}$$

Proof: By (7b), if (17) holds, we have

$$\begin{aligned}
&U_A^{PT}(P_T\mathbf{I}_{(B,q^*)},-\mathbf{2}) = P_T(h_{T,q^*}-\mu)-\frac{C_m}{L}\sum_{l=0}^{L} lw_A(\beta_l)\\
&\ge\sum_{i=1}^{B}(h_{T,i}-\mu)x_i-\frac{C_m}{L}\sum_{l=0}^{L} lw_A(\beta_l)=U_A^{PT}(\mathbf{x},-\mathbf{2}).
\end{aligned}\tag{20}$$

If (18) holds, we have

$$\begin{aligned}
&U_E^{PT}(P_T\mathbf{I}_{(B,q^*)},-\mathbf{2}) = (\mu-h_{T,q^*})P_T+\frac{C_m}{L}\sum_{l=0}^{L} lw_E(\beta_l)\\
&> \max\Bigg\{(\mu-h_{T,q^*})P_T+\frac{1}{L}\frac{h_{T,q^*}h_{J,q^*}P_TP_J}{\sigma+h_{J,q^*}P_J}\sum_{l=0}^{L} lw_E(\eta_l),\\
&\quad(\mu+h_{E,q^*}-h_{T,q^*})P_T\Bigg\}\\
&= \max\left\{U_E^{PT}(P_T\mathbf{I}_{(B,q^*)},\mathbf{y}\succeq\mathbf{0},U_E^{PT}(P_T\mathbf{I}_{(B,q^*)},-\mathbf{1})\right\}.
\end{aligned}\tag{21}$$

Thus (9) and (10) hold, indicating that $(P_T\mathbf{I}_{(B,q^*)},-\mathbf{2})$ is an NE of $\mathcal{G}$. ■

Corollary 2. *If (17) and (18) hold, the utility of Alice in the PT-based smart attack game is*

$$U_A^{EUT} = P_T\max_{1\le i\le B}\{h_{T,i}\}-\mu P_T-\frac{C_m}{L}\sum_{l=0}^{L} l\beta_l. \tag{22}$$

Remark: Eve spoofs Bob if she overviews the spoofing cost of the UAV system due to miss detection is large. In this case, Alice at the NE should allocate all her transmit power to the best frequency channel.

Theorem 3. *The PT-based smart attack game $\mathcal{G}$ with $B=1$ has an NE (P_T,P_J), if*

$$\begin{cases}
\dfrac{h_Th_JP_J}{\sigma+h_JP_J}\sum\limits_{l=0}^{L} lw_E(\eta_l)>\max\Bigg\{Lh_E, & \\
\qquad\qquad\dfrac{C_m}{P_T}\sum\limits_{l=0}^{L} lw_E(\beta_l)\Bigg\} & (23)\\
\dfrac{h_Th_JP_J}{\sigma+h_JP_J}\sum\limits_{l=0}^{L} lw_A(\eta_l)<L(h_T-\mu). & (24)
\end{cases}$$

Proof: By (8a) $\sim$ (8c) and (23), as $B=1$, we have

$$\begin{aligned}
&U_E^{PT}(P_T,P_J) = (\mu-h_T)P_T+\frac{1}{L}\frac{h_Th_JP_TP_J}{\sigma+h_JP_J}\sum_{l=0}^{L} lw_E(\eta_l)\\
&> \max\Bigg\{\left(\mu+h_E-h_T\right)P_T,(\mu-h_T)P_T+\frac{C_m}{L}\sum_{l=0}^{L} lw_E(\beta_l)\Bigg\}\\
&= \max\left\{U_E^{PT}(P_T,-1),U_E^{PT}(P_T,-2)\right\}.
\end{aligned}\tag{25}$$

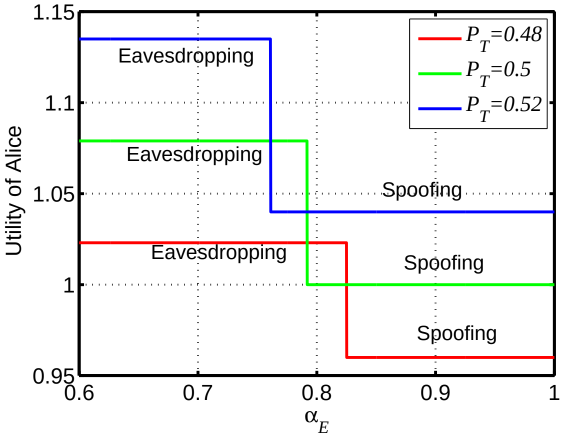
If (24) holds, we have

$$\frac{\partial U_A^{PT}(x,y)}{\partial x} = h_T-\mu-\frac{1}{L}\frac{h_Th_Jy}{\sigma+h_Jy}\sum_{l=0}^{L} lw_A(\eta_l)>0. \tag{26}$$

Thus (9) and (10) hold, indicating that (P_T,P_J) is an NE of the game $\mathcal{G}$. ■

Corollary 3. *If $B=1$, and (23) and (24) hold, the utility of Alice in the PT-based smart attack game is*

$$U_A^{EUT} = \left(h_T-\mu\right)P_T-\frac{1}{L}\frac{h_Th_JP_TP_J}{\sigma+h_JP_J}\sum_{l=0}^{L} l\eta_l. \tag{27}$$



Utility of Alice

Fig. 2. Performance of the subjective smart attack game under uncertain miss detection rate, with $B = 3$, $C_m = 1.5$, $\mu = 0.2$, $\mathbf{H}_T = [3\ 2\ 1]$, $\mathbf{H}_E = [0.8\ 0.5\ 0.2]$, $\mathbf{H}_J = [1\ 2\ 3]$, $\sigma = 1$, $[\beta_l]_{0\leq l\leq L} = [0.1\ 0.8\ 0.05\ 0.03\ 0.02\ 0]$, $[\eta_l]_{0\leq l\leq L} = [0.1\ 0.6\ 0.1\ 0.05\ 0.05\ 0.1]$, and $\alpha_A = 1$.

Remark: If Eve overweighs the miss detection rate, she is more likely to attack the signal transmissions. On the other hand, if Alice is confident regarding the detection accuracy, she transmits signals with full power.

As shown in Fig. 2, the expected utility of the UAV system increases with P_T, and has a sharp decrease at $\alpha_E = 0.792$ if $P_T = 0.5$, because Eve tends to eavesdrop Alice rather than spoofing Bob. Eve tends to spoof Bob with higher transmit power.

VI. PT-Based Dynamic Smart Attack Game

In the dynamic PT-based smart attack game, a smart attacker (Eve) and a UAV (Alice) repeat their interactions without being aware of the environment model, such as the current channel conditions. To derive the optimal power allocation strategy of the UAV, reinforcement learning techniques can be applied to defend against smart attackers. Here, we propose Q-learning, WoLF-PHC and DQN-based power allocation strategies for the UAV.

A. Q-learning Based Power Allocation

In the Q-learning based power allocation, Alice sends a signal to Bob against Eve in each time slot with the chosen transmit power based on the transmission history and the network state. The Q-learning algorithm, as a model-free reinforcement learning algorithm, depends on the quality function or Q-function denoted by $Q\big(\mathbf{s}^{(k)},\mathbf{x}^{(k)}\big)$. The function is the expected discount long-term utility if transmitting with power $\mathbf{x}$ in state $\mathbf{s}$ at time slot k, $\mathbf{s}$ is the system state consisting of the attack mode at time slot $k-1$. The transmit power of the UAV is quantified into $10P_T + 1$ levels, i.e., $x_i \in \{0, 0.1, 0.2, ..., P_T\}$. The value function denoted by $V(\mathbf{s})$ represents the maximum value of the Q-function in state $\mathbf{s}$. According to the iterative Bellman equation, Alice updates her Q-function at time slot k as follows:

$$Q\big(\mathbf{s}^{(k)},\mathbf{x}^{(k)}\big) \leftarrow (1-\tau)Q\big(\mathbf{s}^{(k)},\mathbf{x}^{(k)}\big) + \tau\Big(u\big(\mathbf{s}^{(k)},\mathbf{x}^{(k)}\big) + \gamma V\big(\mathbf{s}^{(k+1)}\big)\Big) \quad (28)$$

$$V\big(\mathbf{s}^{(k)}\big) = \max_{\mathbf{x}^{(k)}\in\mathbf{X}} Q\big(\mathbf{s}^{(k)},\mathbf{x}^{(k)}\big) \quad (29)$$

where learning factor $\tau \in (0, 1]$ represents the learning rate of Alice, and discount factor $\gamma \in [0, 1]$ represents how Alice views the importance of future rewards.

According to the ϵ-greedy strategy, Alice chooses the transmit power that maximizes its Q-function with a high probability, $1-\epsilon$, and chooses each of the other power with a small probability, where $\epsilon \in (0, 1)$ is a small positive value. The Q-learning based power allocation strategy against smart attacks is summarized in Algorithm 1.

Algorithm 1 Q-learning based power allocation.

```
Initialize $\tau$, $\gamma$, $\epsilon$, $\mathbf{s}^{(1)}$.
$Q(\mathbf{s},\mathbf{x}) = \mathbf{0}$, $V(\mathbf{s}) = \mathbf{0}$, $\forall \mathbf{s},\mathbf{x}$.
For $k = 1, 2, 3, ...$
    Choose $\mathbf{x}^{(k)}$ with $\epsilon$-greedy strategy;
    Transmit signals with power $\mathbf{x}^{(k)}$;
    Observe $\mathbf{y}^{(k)}$, SINR$^{(k)}$ and $u^{(k)}$;
    Update $Q(\mathbf{s}^{(k)},\mathbf{x}^{(k)})$ and $V(\mathbf{s}^{(k)})$ via (28) and (29);
    $\mathbf{s}^{(k+1)} = \mathbf{y}^{(k)}$.
End for
```

B. WoLF-PHC Based Power Allocation

Alice can use a mixed strategy to determine the transmit power allocation with randomness to confuse Eve. The WoLF-PHC algorithm extends the Q-learning algorithm to the mixed-strategy game, and uses the win or learn fast principle and a learning parameter for better convergence [33]. More specifically, the transmit strategy is chosen according to a mixed-strategy table denoted by $\pi(\mathbf{s}^{(k)},\mathbf{x})$, with $\sum_{\mathbf{x}\in\mathbf{X}} \pi(\mathbf{s}^{(k)},\mathbf{x}) = 1$. Two learning parameters δ_w and $\delta_l \in (0, 1]$ are chosen to update $\pi(\mathbf{s}^{(k)},\mathbf{x})$ for different cases. Alice has to learn faster if she is losing and to learn faster about Eve's policy, i.e., the losing learning speed δ_l has to be larger than the winning speed δ_w.

The average mixed-strategy table denoted by $\bar{\pi}$ is updated by

$$\bar{\pi}\big(\mathbf{s}^{(k)},\mathbf{x}\big) \leftarrow \bar{\pi}\big(\mathbf{s}^{(k)},\mathbf{x}\big) + \frac{1}{C(\mathbf{s}^{(k)})}\Big(\pi\big(\mathbf{s}^{(k)},\mathbf{x}\big) - \bar{\pi}\big(s^{(k)},\mathbf{x}\big)\Big), \quad \forall \mathbf{x} \in \mathbf{X} \quad (30)$$

where $C(\mathbf{s}^{(k)})$ is the number of state $\mathbf{s}^{(k)}$. If the current policy has a higher expected value than the average mixed-strategy, Alice wins and updates the mixed-strategy table with the learning parameter δ_w; Otherwise, the UAV loses and updates the mixed-strategy table with a higher rate δ_l. That is

$$\delta^{(k)} = \begin{cases} \delta_w, & \sum_{\mathbf{x}\in\mathbf{X}} \pi\big(\mathbf{s}^{(k)},\mathbf{x}\big)Q\big(\mathbf{s}^{(k)},\mathbf{x}\big) \\ & \quad > \sum_{\mathbf{x}\in\mathbf{X}} \bar{\pi}\big(\mathbf{s}^{(k)},\mathbf{x}\big)Q\big(\mathbf{s}^{(k)},\mathbf{x}\big) \\ \delta_l, & \text{o.w.} \end{cases} \quad (31)$$

Algorithm 2 WoLF-PHC based power allocation.

```
Initialize $\tau$, $\gamma$, $\delta_l$, $\delta_w$, $\epsilon$, $\mathbf{s}^{(1)}$.
$Q(\mathbf{s},\mathbf{x}) = \mathbf{0}$, $V(\mathbf{s}) = \mathbf{0}$, $\pi(\mathbf{s},\mathbf{x}) = \frac{\mathbf{1}}{|\mathbf{X}|}$, $C(\mathbf{s}) = \mathbf{0}$.
For $k = 1, 2, 3, ...$
    Choose $\mathbf{x}^{(k)}$ via (33);
    Transmit signals with power $\mathbf{x}^{(k)}$;
    Observe $\mathbf{y}^{(k)}$, SINR$^{(k)}$ and $u^{(k)}$;
    Update $Q(\mathbf{s}^{(k)}, \mathbf{x}^{(k)})$ and $V(\mathbf{s}^{(k)})$ via (28) and (29);
    Update $\bar{\pi}(\mathbf{s}^{(k)})$ via (30);
    Update $\pi(\mathbf{s}^{(k)})$ via (32);
    $\mathbf{s}^{(k+1)} = \mathbf{y}^{(k)}$.
End for
```

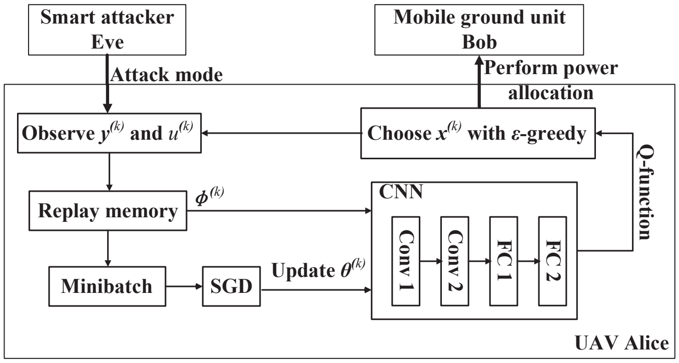


Fig. 3. DQN-based UAV power allocation scheme.

The mixed-strategy table π is updated with the learning parameter $\delta^{(k)}$ as follows,

$$\pi\big(\mathbf{s}^{(k)}, \mathbf{x}\big) \leftarrow \pi\big(\mathbf{s}^{(k)}, \mathbf{x}\big) + \begin{cases} -\min\left(\pi\big(\mathbf{s}^{(k)}, \mathbf{x}\big), \frac{\delta^{(k)}}{|\mathbf{X}|-1}\right), & \text{if } \mathbf{x} \neq \arg\max_{\hat{\mathbf{x}} \in \mathbf{X}} Q\big(\mathbf{s}^{(k)}, \hat{\mathbf{x}}\big) \\ \sum_{\mathbf{x} \neq \hat{\mathbf{x}}} \min\left(\pi\big(\mathbf{s}^{(k)}, \mathbf{x}\big), \frac{\delta^{(k)}}{|\mathbf{X}|-1}\right), & \text{o.w.} \end{cases} \tag{32}$$

The UAV transmit power $\mathbf{x}^{(k)}$ is chosen according to the mixed-strategy table, i.e.,

$$\Pr\left(\mathbf{x}^{(k)} = \hat{\mathbf{x}}\right) = \pi\big(\mathbf{s}^{(k)}, \hat{\mathbf{x}}\big), \quad \forall \hat{\mathbf{x}} \in \mathbf{X}. \tag{33}$$

Alice observes Eve's response, the receiver Bob measures the bit error rate or packet loss rate of Alice's signals and then infers the SINR and the attack policy at that time, which can be sent to Alice via the feedback channel. Then Alice uses the SINR of the message and the secrecy capacity to evaluate her utility and then updates the Q-function via (28) and (29) as shown in Algorithm 2.

C. DQN-Based Power Allocation

The Q-function has to be evaluated in the PHC-based UAV power allocation and the learning time required increases rapidly with the dimension of the state-action space, which in turn depends on the number of the radio channels B, the transmit power constraint P_T and the jamming power constraint P_J. Therefore, we present a DQN-based power allocation that uses a neural network to accelerate the learning process.

A neural network called DQN with weights $\theta^{(k)}$ is used to estimate the Q value for each power allocation strategy, i.e., $Q\big(\mathbf{s}^{(k)}, \mathbf{x}^{(k)}; \theta^{(k)}\big) \approx Q^*\big(\mathbf{s}^{(k)}, \mathbf{x}^{(k)}\big)$. As shown in Fig. 3, the neural network consists of two convolutional layers with rectified linear units, followed by two fully connected layers. The first convolutional layer convolves 20 filters of 3×3 with stride 1, and the second one convolves 40 filters of 2×2 with stride 1. The two fully connected layers consists of 180 and $|\mathbf{X}|$ rectifier units respectively.

Alice's experience denoted by $e^{(k)} = \big(\varphi^{(k)}, \mathbf{x}^{(k)}, u^{(k)}, \varphi^{(k+1)}\big)$ is stored in a data set $\mathcal{D} = \big\{e^{(j)}\big\}_{1 \le j \le k}$, where $\varphi^{(k)}$ denotes the state sequence at time k, which consists of the current system state and the previous W state-action pairs, i.e., $\varphi^{(k)} = \big(\mathbf{s}^{(k-W)}, \mathbf{x}^{(k-W)}, ..., \mathbf{x}^{(k-1)}, \mathbf{s}^{(k)}\big)$. In the first $W-1$ time slots, Alice chooses the power allocation strategy randomly. After that, Alice selects a power allocation strategy with maximum Q-value according to ϵ-greedy policy. Then, the attack mode of Eve is observed, and the utility of Alice is obtained.

When training the DQN, random minibatches from the data set are used instead of the most recent transition. More specifically, Alice chooses an experience, $e^{(d)}$, from the data set $\mathcal{D}$ randomly for T times to calculate the Q-function. The CNN weight, $\theta^{(k)}$, is obtained by minimizing the loss function denoted by $L(\theta)$ following the stochastic gradient descent (SGD) as summarized in Algorithm 3, i.e.,

$$\begin{aligned} \theta^{(k)} &= \arg\min_{\theta} L(\theta) \\ &= \arg\min_{\theta} \mathbb{E}_{\varphi, \mathbf{x}, u, \varphi'} \Bigg[\Big(u^{(k)} + \gamma \max_{\mathbf{x}'} Q\big(\varphi', \mathbf{x}'; \theta^{(k-1)}\big) \\ &\qquad - Q\big(\varphi, \mathbf{x}; \theta\big)\Big)^2 \Bigg] \end{aligned} \tag{34}$$

where φ' is the next state sequence.

As clarified in Section VI of the revision, the DQN-based power allocation scheme does not always outperform the other two methods. More specifically, the DQN based scheme that can increase the secrecy capacity and the utility of the UAV system has the highest computational complexity and takes much longer time to make a decision in each time compared with the WoLF-PHC based scheme. Therefore, the DQN based power allocation is applicable to the UAVs with sufficient computation resources. On the other hand, UAVs with restricted computational resource cannot afford the complicated DQN algorithm, and have to resort to the WoLF-PHC based scheme with less complexity and computation costs to choose the transmission strategy in time. For example, the WoLF-PHC algorithm takes 4% less time on average to choose the transmission strategy in a time slot compared with the DQN algorithm in the experiment.

VII. Simulation Results

Simulations are carried out to evaluate the performance of the proposed power allocation strategies against a smart

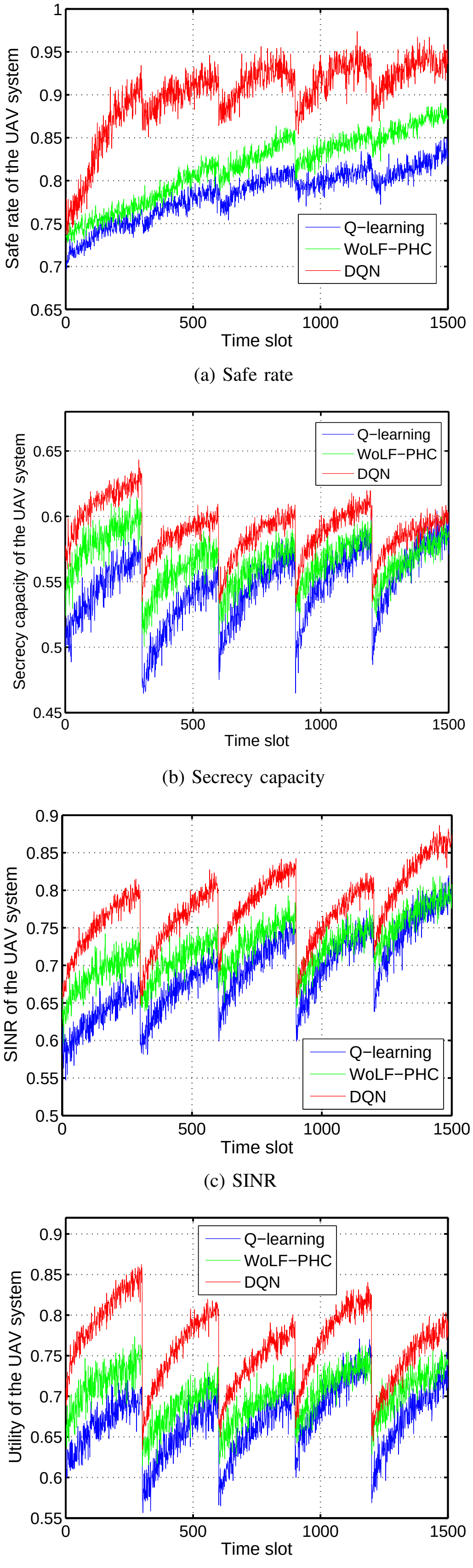
Algorithm 3 DQN-based power allocation.

```
Initialize $\theta$, $\gamma$, $\epsilon$, $\mathcal{D}$, $W$, $T$.
For $k = 1, 2, ...$
    If $k < W$
        Choose $\mathbf{x}^{(k)}$ randomly;
    Else
        Obtain the output $Q(\varphi^{(k)}, \mathbf{x}; \theta^{(k)})$;
        Choose $\mathbf{x}^{(k)}$ with $\epsilon$-greedy policy;
    End if
    Transmit signals with power $\mathbf{x}^{(k)}$;
    Observe $\mathbf{y}^{(k)}$ and $u^{(k)}$;
    $\mathbf{s}^{(k+1)} = \mathbf{y}^{(k)}$;
    $\varphi^{(k+1)} = (\mathbf{s}^{(k-W+1)}, \mathbf{x}^{(k-W+1)}, ..., \mathbf{x}^{(k)}, \mathbf{s}^{(k+1)})$;
    $e^{(k)} = (\varphi^{(k)}, \mathbf{x}^{(k)}, u^{(k)}, \varphi^{(k+1)})$;
    $\mathcal{D} \leftarrow e^{(k)} \cup \mathcal{D}$;
    For $d = 1, 2, ..., T$;
        Select $e^{(d)} \in \mathcal{D}$ randomly;
    End for
    Update $\theta^{(k)}$ via (34);
End for
```

attacker Eve with Q-learning based attack strategy, in which Eve chooses to launch jamming, spoofing or eavesdropping attacks. The system parameters are chosen for some typical scenarios similar to [34] with $P_T = P_J = 0.4$, $\sigma = 1$, $C_m = -0.5$, $[\beta_l]_{0\leq l\leq 5} = [0.1\ 0.8\ 0.05\ 0.03\ 0.02\ 0]$, $[\eta_l]_{0\leq l\leq 5} = [0.1\ 0.6\ 0.1\ 0.05\ 0.05\ 0.1]$, $\alpha_E = 0.8$, $\alpha_A = 1$, $\tau = 0.95$, $\gamma = 0.7$, $d_0 = 10$ m and $\epsilon = 0.9$, if not specified otherwise. In the downlink UAV transmission, $d_e = 30$ m, $d_t = d_j = 50$ m, $\xi = 0.02$, 0.075 and 0.0032 for the Alice-Bob, Eve-Bob and Alice-Eve's UAV links, respectively.

In the first experiment, the channel gains change randomly every 300 time slots. As shown in Fig. 4, the safe rate of the UAV system that is the probability of not being attacked increases with time. For instance, the safe rate increases by 25% after 1500 time slots in the dynamic game. The DQN-based strategy has the highest safe rate, and followed by the WoLF-PHC and Q-learning based strategy. The safe rate of the DQN-based strategy is 93%, which is 7% and 11% higher than that of the WoLF-PHC and Q-learning based scheme, respectively. The DQN algorithm has a higher secrecy capacity than the WoLF-PHC and Q-learning algorithms. Besides, the DQN-based strategy has a higher SINR than the WoLF-PHC and Q-learning based strategies. The utility of Alice first decreases if the channel gains change, and then rapidly learns thereafter. For instance, the DQN-based transmission increases the utility of Alice by 0.84, which is 13% higher than WoLF-PHC, and 22% higher than Q-learning. As shown in Fig. 5, the DQN-based strategy reduces the influence of spoofing attacks compared with Q-learning and WoLF-PHC.

The subjectivity view of Eve improves the UAV transmission as shown in Fig. 6. For instance, the average safe rate decreases by 16.3% if the objectivity of Eve α_E changes from 0.6 to 1 with the Q-learning based strategy. The DQN-based strategy exceeds the WoLF-PHC and Q-learning strategies with a higher safe rate. For instance, the average safe rate



(a) Safe rate

(b) Secrecy capacity

(c) SINR

(d) Utility

Fig. 4. Performance of the learning based power allocation strategy against smart attacks with $P_T = P_J = 0.4$, $L = 5$, $\sigma = 1$, $C_m = -0.5$, $\alpha_E = 0.8$ and $\alpha_A = 1$.

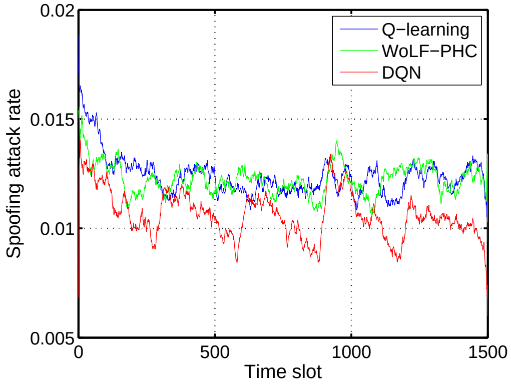
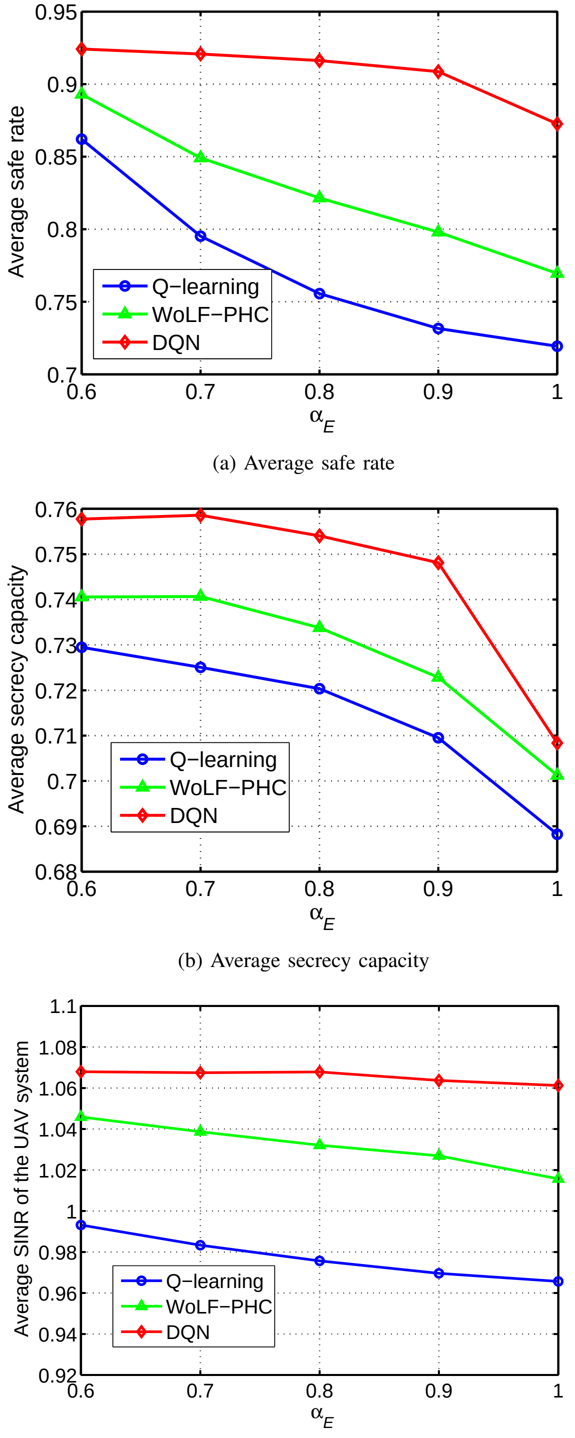


Fig. 5. The influence of the spoofing attack of the UAV system with $P_T = P_J = 0.4$, $L = 5$, $\sigma = 1$, $C_m = -0.5$, $\alpha_E = 0.8$ and $\alpha_A = 1$.



(a) Average safe rate

(b) Average secrecy capacity

(c) Average SINR of the UAV system

Fig. 6. Average performance of the learning based power allocation strategy against smart attacks with $P_T = P_J = 0.4$, $L = 5$, $\sigma = 1$, $C_m = 1.5$ and $\alpha_A = 1$.

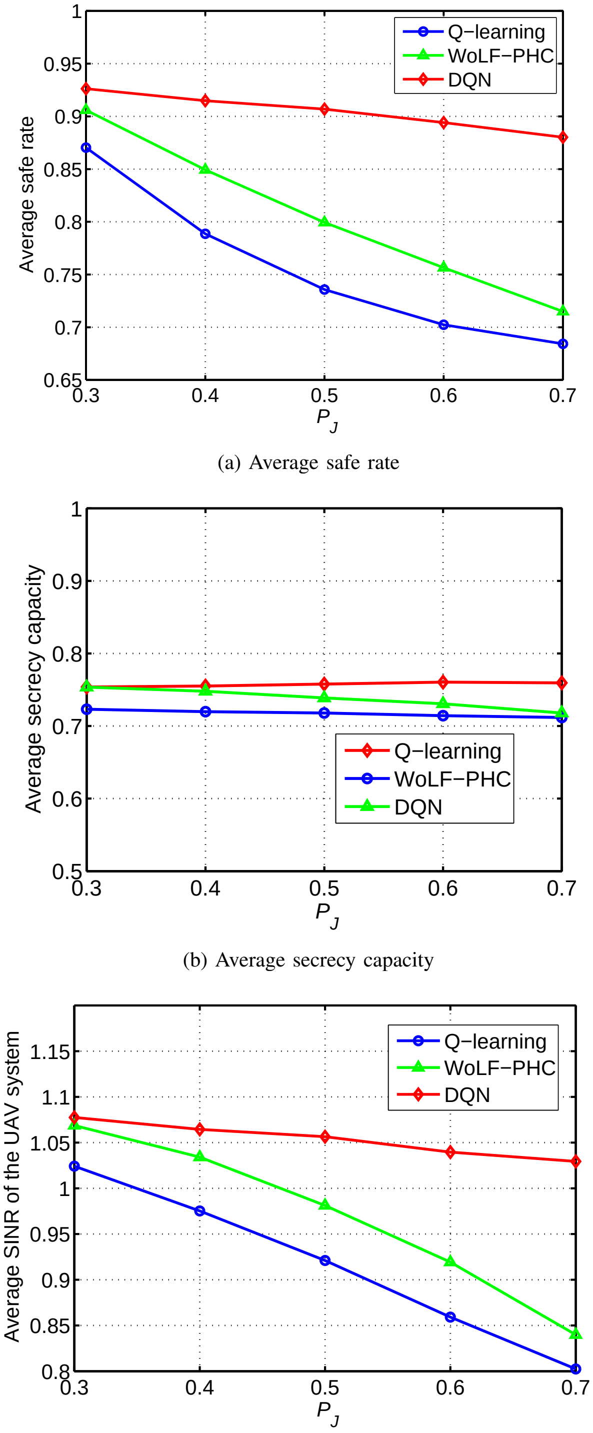
with the DQN is 11% higher than that of the WoLF-PHC and 18% higher than that of Q-learning algorithm if $\alpha_E = 0.9$. The average secrecy capacity decreases with α_E. For example, the average secrecy capacity decrease by 5.3% when α_E changes from 0.9 to 1 by DQN. The average SINR decreases with α_E. For instance, if $\alpha_E = 0.9$, the DQN can increase the SINR by 10.3% compared with the Q-learning strategy.

We illustrate the influence of the total power constraint of the smart attacker Eve with $P_T = 0.4$ in Fig. 7. The average safe rate decreases with the total power of the attacker. For instance, the average safe rate decreases 20.5% by Q-learning if the total power of the attacker P_J changes from 0.3 to 0.7, because the attacker has more power to prevent the reliable and secret communications between Alice and Bob. The DQN-based strategy has the highest average SINR, e.g., the average SINR of the DQN based strategy is 22.3% and 29.3% higher than WoLF-PHC and Q-learning, respectively, if $P_J = 0.7$.

VIII. Conclusion

In this paper, we have investigated the PT-based smart attack game between a subjective smart attacker and a UAV who allocates power on multiple channels. The NEs of the static game have been derived to show the impact of the subjectivity of the attacker. We have proposed the DQN-based power allocation strategy for a UAV to address smart attacks without the knowledge of the attack model and the attack detection accuracy of the communication system. Simulation results demonstrate that the proposed strategy accelerates the learning rate and the secrecy capacity of the UAV system. For example, the proposed scheme increases the secrecy capacity of the UAV system by 16% and the utility by 22%, as compared with the Q-learning based scheme, if $P_T = P_J = 0.4$, $\alpha_E = 0.8$ and $\alpha_A = 1$.

This work assumes the simplified system model as shown in Fig. 1. If several UAVs cooperate to address smart attacks, the network model becomes more complicated and the UAV systems can apply multi-agent reinforcement learning techniques to improve the security performance. The multi-agent reinforcement learning based security technique requires more computational overhead than our proposed scheme, because



(a) Average safe rate

(b) Average secrecy capacity

(c) Average SINR of UAV system

Fig. 7. Average performance of the learning based power allocation strategy against smart attacks with $P_T = 0.4$, $L = 5$, $\sigma = 1$, $C_m = -0.5$, $\alpha_E = 0.8$ and $\alpha_A = 1$.

each cooperative UAV can observe more state information and the resulting state space is much larger. We will investigate this technique in our future work.

References

[1] C. Xie and L. Xiao, "User-centric view of smart attacks in wireless networks," in *Proc. IEEE Int'l Conf. Ubiquitous Wireless Broadband (ICUWB)*, pp. 1–6, Nanjing, China, Oct. 2016.

[2] A. Y. Javaid, W. Sun, V. K. Devabhaktuni, and M. Alam, "Cyber security threat analysis and modeling of an unmanned aerial vehicle system," in *IEEE Conf. Technologies for Homeland Security (HST)*, pp. 585–590, Waltham, MA, Nov. 2012.

[3] J. Xu, L. Duan, and R. Zhang, "Proactive eavesdropping via jamming for rate maximization over Rayleigh fading channels," *IEEE Wireless Commun. Lett.*, vol. 5, no. 1, pp. 80–83, Feb. 2016.

[4] M. H. Yılmaz and H. Arslan, "A survey: Spoofing attacks in physical layer security," in *IEEE Local Computer Networks Conf. Workshops (LCN Workshops)*, pp. 812–817, Clearwater Beach, FL, Oct. 2015.

[5] Y.-S. Shiu, S. Y. Chang, H.-C. Wu, S. C.-H. Huang, and H.-H. Chen, "Physical layer security in wireless networks: A tutorial," *IEEE wireless Commun.*, vol. 18, no. 2, pp. 66–74, Apr. 2011.

[6] L. Xiao, C. Xie, T. Chen, H. Dai, and H. V. Poor, "A mobile offloading game against smart attacks," *IEEE Access*, vol. 4, pp. 2281–2291, May 2016.

[7] L. Xiao, L. J. Greenstein, N. B. Mandayam, and W. Trappe, "Channel-based spoofing detection in frequency-selective Rayleigh channels," *IEEE Trans. on Wireless Commun.*, vol. 8, no. 12, pp. 5948–5956, Dec. 2009.

[8] M. G. Oskoui, P. Khorramshahi, and J. A. Salehi, "Using game theory to battle jammer in control channels of cognitive radio ad hoc networks," in *IEEE Int'l Conf. Commun. (ICC)*, pp. 1–5, Kuala Lumpur, Malaysia, May 2016.

[9] R. W. Thomas, B. J. Borghetti, R. S. Komali, and P. Mahonen, "Understanding conditions that lead to emulation attacks in dynamic spectrum access," *IEEE Commun. Magazine*, vol. 49, no. 3, pp. 32–37, Mar. 2011.

[10] A. Fielder, E. Panaousis, P. Malacaria, C. Hankin, and F. Smeraldi, "Game theory meets information security management," in *IFIP Int'l Info. Security Conf.*, pp. 15–29, Marrakech, Morocco, Jun. 2014.

[11] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econometrica*, vol. 47, no. 2, pp. 263–291, Mar. 1979.

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

[13] L. Xiao, D. Xu, C. Xie, N. B. Mandayam, and H. V. Poor, "Cloud storage defense against advanced persistent threats: A prospect theoretic study," *IEEE Journal on Selected Areas in Commun.*, vol. 35, no. 3, pp. 534–544, Mar. 2017.

[14] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.

[15] Q. Zhu, W. Saad, Z. Han, H. V. Poor, and T. Basar, "Eavesdropping and jamming in next-generation wireless networks: A game-theoretic approach," in *IEEE Military Commun. Conf. (MILCOM)*, pp. 119–124, Baltimore, MD, Nov. 2011.

[16] A. Mukherjee and A. L. Swindlehurst, "Jamming games in the MIMO wiretap channel with an active eavesdropper," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 82–91, Jan. 2013.

[17] A. Garnaev, M. Baykal-Gursoy, and H. V. Poor, "A game theoretic analysis of secret and reliable communication with active and passive adversarial modes," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 2155–2163, Mar. 2016.

[18] Y. E. Sagduyu, R. Berry, and A. Ephremides, "MAC games for distributed wireless network security with incomplete information of selfish and malicious user types," in *Proc. IEEE Int'l Conf. Game Theory for Networks*, pp. 130–139, Istanbul, Turkey, May 2009.

[19] F. M. Aziz, J. S. Shamma, and G. L. Stber, "Jammer-type estimation in LTE with a smart jammer repeated game," *IEEE Trans. on Vehicular Technology*, vol. 66, no. 8, pp. 7422–7431, Aug. 2017.

[20] Y. Yang, L. T. Park, N. B. Mandayam, I. Seskar, A. L. Glass, and N. Sinha, "Prospect pricing in cognitive radio networks," *IEEE Trans. Cognitive Commun. and Networking*, vol. 1, no. 1, pp. 56–70, Mar. 2015.

[21] T. Li and N. B. Mandayam, “When users interfere with protocols: Prospect theory in wireless networks using random access and data pricing as an example,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 1888–1907, Apr. 2014.

[22] J. Yu, M. H. Cheung, and J. Huang, “Spectrum investment with uncertainty based on prospect theory,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, pp. 1620–1625, Sydney, NSW, Australia, Jun. 2014.

[23] A. Sanjab, W. Saad, and T. Basar, “Prospect theory for enhanced cyber-physical security of drone delivery systems: A network interdiction game,” pp. 1–6, Paris, France, May 2017.

[24] H. Saad, A. Mohamed, and T. ElBatt, “Cooperative Q-learning techniques for distributed online power allocation in femtocell networks,” *Wireless Commun. and Mobile Computing*, vol. 15, no. 15, pp. 1929–1944, Oct. 2015.

[25] A. Shahid, S. Aslam, H. S. Kim, and K.-G. Lee, “A docitive Q-learning approach towards joint resource allocation and power control in self-organised femtocell networks,” *Trans. Emerging Telecommunications Technologies*, vol. 26, no. 2, pp. 216–230, Feb. 2015.

[26] R. Ye and Q. Xu, “Learning-based power management for multicore processors via idle period manipulation,” *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 33, no. 7, pp. 1043–1055, Jul. 2014.

[27] C. Wang and W.-H. Kuo, “A utility-based resource allocation scheme for IEEE 802.11 WLANs via a machine-learning approach,” *Wireless Networks*, vol. 20, no. 7, pp. 1743–1758, Oct. 2014.

[28] P. Lama and X. Zhou, “Aroma: Automated resource allocation and configuration of mapreduce environment in the cloud,” in *Proc. ACM Int’l Conf. Autonomic Computing*, pp. 63–72, San Jose, CA, Sept. 2012.

[29] T. S. Rappaport *et al.*, *Wireless communications: Principles and practice*, vol. 2. Prentice Hall PTR, NJ, 1996.

[30] S. Mathur, A. Reznik, C. Ye, R. Mukherjee, A. Rahman, *et al.*, “Exploiting the physical layer for enhanced security,” *IEEE Wireless Commun.*, vol. 17, no. 5, pp. 63–70, Oct. 2010.

[31] A. Marttinen, A. M. Wyglinski, and R. Jantti, “Statistics-based jamming detection algorithm for jamming attacks against tactical MANETs,” in *IEEE Military Commun. Conf.*, pp. 501–506, Baltimore, MD, Oct. 2014.

[32] D. Prelec, “The probability weighting function,” *Econometrica*, vol. 66, no. 3, pp. 497–527, May 1998.

[33] M. Bowling and M. Veloso, “Multiagent learning using a variable learning rate,” *Artificial Intelligence*, vol. 136, no. 2, pp. 215–250, Apr. 2002.

[34] A. Garnaev and W. Trappe, “The eavesdropping and jamming dilemma in multi-channel communications,” in *Proc. IEEE Int’l Conf. Commun. (ICC)*, pp. 2160–2164, Budapest, Jun. 2013.