Received February 4, 2020, accepted February 11, 2020, date of publication February 17, 2020, date of current version March 2, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2974233

Intelligent Secure Communication for Cognitive Networks With Multiple Primary Transmit Power

SHIWEI LAI[1], JUNJUAN XIA[1], DAN ZOU[2], AND LISENG FAN[1]

[1]School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou 510006, China
[2]School of Information Engineering, East China Jiaotong University, Nanchang 330013, China

Corresponding authors: Junjuan Xia (xiajunjuan@gzhu.edu.cn) and Liseng Fan (lsfan2019@126.com)

This work was supported in part by the NSFC under Grant 61871139 and Grant 61601275, and in part by the Science and Technology Program of Guangzhou under Grant 201807010103.

ABSTRACT In this paper, we study an intelligent secure communication scheme for cognitive networks with multiple primary transmit power, where a secondary Alice transmits its secrecy data to a secondary Bob threatened by a secondary attacker. The secondary nodes limit their transmit power among multiple levels, in order to maintain the quality of service of the primary networks. The attacker can work in an eavesdropping, spoofing, jamming or silent mode, which can be viewed as the action in the traditional Q-learning algorithm. On the other hand, the system can adaptively choose the transmit power level among multiple ones to suppress the intelligent attacker, which can be viewed as the status of Q-learning algorithm. Accordingly, we firstly formulate this secure communication problem as a static secure communication game with Nash equilibrium (NE) between the main links and attacker, and then employ the Q-learning algorithm to select the transmit power level. Simulation results are finally demonstrated to verify that the intelligent attacker can be effectively suppressed by the proposed studies in this paper.

INDEX TERMS Intelligent secure communication, Q-learning algorithm, Nash equilibrium.

I. INTRODUCTION

In recent years, there have been many progresses in the development of wireless communications [1]–[4], in order to tackle with the increasing challenge of wireless big data [5]–[8] and mobile edge computing [9]–[12]. Among the newly increasing techniques, cognitive technique can be viewed as a novel approach to improve spectrum utilization effectively and also has been recognized as a smart wireless communication technology in the limited of radio spectrum [13]–[17]. When the interference from the secondary users to the primary ones [18] is below a given level or the spectrum is not used, the secondary users are enabled to access the spectrum of the primary users. The channel capacity of the secondary users is limited by the primary users' tolerant interference power in cognitive network [19]–[21]. Thus, it is of vital importance to make sure that the secondary user should make use of the spectrum and reduce interference to the primary user. Most of studies in cognitive network focus on channel identification, detection and management of spectrum and power allocation. In practice, the level of primary transmit power can be single due to the transmission of fixed services. When the transmission services are varying, the primary users should use multiple levels of transmit power, in order to provide better performance [22], [23].

On the other hand, with the rapid development of the wireless networks, wireless networks are closely related to people's privacy communication and so on [24]–[26]. The security of wireless communication network has received more and more attention. Wireless communication security has become an important research topic recently and it mainly focuses on the physical-layer security research of wireless networks. Traditional encryption techniques rely on application-layer operations, which causes much more computational complexity [27]. In contrast, the application of the physical-layer security mechanism can make it more difficult for attackers to decipher the transmitted information. In [28], [29], physical-layer security has been proposed to safeguard data confidentiality in 5G wireless communication networks. Besides the above research, there have been some researches on the newly developed materials, which can be used in wireless networks for both transmission and improving the environments [30]–[33].

The associate editor coordinating the review of this manuscript and approving it for publication was Min Jia.

Most of the research works on the physical-layer security focus on the fixed-mode attacker, which however ignores the fact that the attacker can be change its mode in order to increase the attack rate. In practice, wireless communication networks are more vulnerable to be attacked by intelligent attackers with the rise and development of new intelligent attackers such as unmanned aerial vehicles [34]. Smart attackers can perform many types of attack based on the environment of the wireless network, including eavesdropping, jamming and spoofing [35], [36]. In order to improve the security performance of communication and reduce the security loss caused by failure to detect attacks in time, many researches have focused on the detection of smart attackers and suppressed the attacks. Specifically, in [37], [38], a Q-learning based power allocation algorithm has been applied to strengthen the secrecy capacity under UAV smart attack. The work in [39] has proposed a power control strategy to suppress the intelligent attacks by using some advanced signal processing techniques such as beamforming and filtering.

In this paper, we consider the wireless communication system where there is a secondary user wants to contact with another secondary user under the constraint of the primary user in the cognitive radio network. Meanwhile, a secondary intelligent attacker exists in this network, and it can work in the eavesdropping, jamming and spoofing modes. In order to improve the security performance of the communication system, a static secure communication game with Nash equilibrium (NE) between the main links and attacker is formulated. We further propose a power control strategy based on Q-learning algorithm to select the transmit power level for the secondary user in the range of tolerant interference power of the primary user. The attacker can select its attack mode among eavesdropping, jamming, spoofing or keeping quiet, according to the practical environments and the cost. The transmitter eventually obtains the optimal transmit power to improve the system secrecy capacity by using the Q-learning algorithm. Simulation results validate that the intelligent attacker can be effectively suppressed by the proposed scheme in the cognitive radio network.

The main contributions of this work are summarized as follows:

- A secure transmission of communication problem in cognitive radio networks under a smart attacker is investigated in this paper, and it is formulated as a static secure communication game with NE strategy.
- A Q-learning algorithm is introduced to determine the transmit power of the secondary user which should not be larger than the peak level of the tolerant interference power. The Q-learning algorithm can improve secrecy capacity of the secondary user and suppress smart attacks under the constraint of the primary user.

The outline of this paper is given as follows. Section II describes the system model in the cognitive radio networks. And then in Section III, we study the secure communication game and present the NE of the game. In Section IV, we present the Q-learning algorithm in detail which is used to select the transmit power level in a dynamic game. Section V presents the simulation results followed by the conclusions in Section VI.

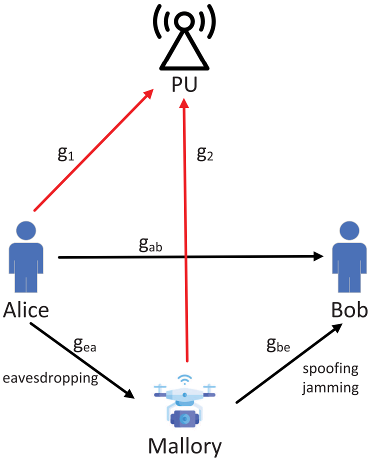


FIGURE 1. Alice communicate with Bob under attack of the smart attacker Mallory in a CR network.

II. SYSTEM MODEL

Fig. 1 shows the system model of the cognitive radio network, where there is one secondary Alice wants to send information to the secondary Bob under the constraint of the primary user. The intelligent attacker Mallory exists in the network, and it can work in the **eavesdropping**, **jamming**, **spoofing**, or **silent** mode, depending on the instantaneous channel state and system settings. Specifically, if the channel between the Alice and Mallory is in good condition, the Mallory may tend to eavesdrop the confidential signal from the Alice. On the other hand, if the channel between the Bob and Mallory is in good condition, the Mallory may tend to perform spoofing or jamming. When all the channels associated with the Mallory are in poor condition, the intelligent attacker may select to keep silent, as it cannot achieve good result in performing eavesdropping, spoofing or jamming. In this work, we use $q \in \{0, 1, 2, \ldots, K\}$ to denote the Mallory action mold, and K is the total number of attack modes. In particular, $K = 3$ corresponds to that the action mode of Mallory consists of keeping silent, eavesdropping, jamming and spoofing, and $q = 0, 1, 2$ or 3 represents the silent, eavesdropping, jamming and spoofing modes, respectively.

To maintain the quality of service for the primary network, the transmit power of the secondary nodes should be limited. In this work, we consider a practical cognitive communication scenario where there exist multiple level of interference transmit power. In particular, we use $I_P \in [0, I_{P,max}]$ to denote the tolerant interference from the primary user, which $I_{P,max}$ is the tolerant peak interference. Moreover, suppose that the primary user has $(L+1)$ levels of tolerant interference power, and the primary user can use the l-th level of the tolerant interference power, denoted by $I_{P,l}$, where $l \in [0, L]$. Note that $I_{P,l}$ is equal to $\frac{lI_{P,max}}{L}$. When the l-th level of the tolerant interference power is used, the transmit power at the Alice is given by

$$P_{Alice} = \frac{I_{P,l}}{|g_1|^2}, \tag{1}$$

where $g_1 \sim \mathcal{CN}(0, \sigma_{i1}^2)$ denote the channel parameters of the link from the Alice to the primary user. Here we use $P_{q,Mallory}$ to denote the transmit power at the Mallory. From P_{Alice} and $P_{q,Mallory}$, we will discuss the secure data transmission process, as follows.

A. WHEN MALLORY KEEPS SILENT

When $q = 0$ holds where the Mallory keeps silent, Alice communicates with Bob by sending signal X_a and then Bob receives a signal Y_B, given by

$$Y_B = g_{ab}\sqrt{P_{Alice}}X_a + n_b, \tag{2}$$

where the channel parameter of the Alice-Bob link is denoted by $g_{ab} \sim \mathcal{CN}(0, \sigma_{ab}^2)$, and $n_b \sim \mathcal{CN}(0, \sigma_n^2)$ is the additive white Gaussian noise at the Bob [40]–[42], where the noise effect on the communication systems can be found in the works [43]–[46]. Note that the Alice can utilize the spectrum resources of the primary networks, as long as its interference is tolerated, which can help improve the system spectrum efficiency significantly. Based on the Shannon theory, we can write the capacity of the Alice-Bob link named by R, given by [47]–[49]

$$R = \log_2\left(1 + \frac{I_{P,l}|g_{ab}|^2}{\sigma_n^2|g_1|^2}\right). \tag{3}$$

B. WHEN MALLORY PERFORMS EAVESDROPPING

When $q = 1$ holds where Mallory performs eavesdropping on the signal X_a of Alice, it obtains a signal Y_E, given by

$$Y_E = g_{ea}\sqrt{P_{Alice}}X_a + n_e, \tag{4}$$

where the channel parameter of the Mallory-Alice link is denoted by $g_{ea} \sim \mathcal{CN}(0, \sigma_{ea}^2)$, and $n_e \sim \mathcal{CN}(0, \sigma_n^2)$ is the additive white Gaussian noise at the Mallory. In this case, the secrecy capacity under eavesdropping can be written as

$$R_E = \log_2\left(1 + \frac{I_{P,l}|g_{ab}|^2}{\sigma_n^2|g_1|^2}\right) - \log_2\left(1 + \frac{I_{P,l}|g_{ea}|^2}{\sigma_n^2|g_1|^2}\right). \tag{5}$$

C. WHEN MALLORY PERFORMS JAMMING

When $q = 2$ holds where Bob is disturbed by the Mallory's jamming signal denoted by Z_J, Bob receives a signal Y_J which consists of both the desired signal and the jamming signal, given by

$$Y_J = g_{ab}\sqrt{P_{Alice}}X_a + g_{be}\sqrt{P_{q,Mallory}}Z_J + n_b, \tag{6}$$

where the channel parameter of the Mallory-Bob link is denoted by $g_{be} \sim \mathcal{CN}(0, \sigma_{be}^2)$.To limit the interference on the primary user, $P_{q,Mallory}$ is given by

$$P_{q,Mallory} = \frac{P_J}{|g_2|^2}, \tag{7}$$

where P_J is the peak interference when the Mallory performs jamming. And $g_2 \sim \mathcal{CN}(0, \sigma_{i2}^2)$ denote the channel parameters of the link from Mallory to the primary user.

In this case, the transmission capacity is given by,

$$\begin{aligned} R_J &= \log_2\left((1 + \frac{I_{P,l}|g_{ab}|^2}{\sigma_n^2|g_1|^2})(1 + \frac{P_J|g_{be}^2|}{\sigma_n^2|g_2|^2})^{-1}\right) && (8)\\ &= \log_2\left(1 + \frac{I_{P,l}|g_{ab}|^2|g_2|^2}{|g_1|^2(\sigma_n^2|g_2|^2 + P_J|g_{be}^2|)}\right) && (9)\end{aligned}$$

D. WHEN MALLORY PERFORMS SPOOFING

When Mallory selects to be a spoofer, it transmits a spoofing signal Z_S with the power P_S to lie to Bob. Then Bob gets a signal Y_S, given by

$$Y_S = g_{be}\sqrt{P_{q,Mallory}}Z_S + n_b \tag{10}$$

where $P_{q,\text{Mallory}}$ is limited by

$$P_{q,Mallory} = \frac{P_S}{|g_2|^2}, \tag{11}$$

in which P_S is the peak interference when the Mallory performs spoofing.

The capacity under spoofing is denoted by R_S. The more spoofing messages Bob receives, the greater it loses. Hence, the secrecy data rate is modeled as a liner function. Note that the intention of the spoofer is to send a spoofing message to Bob, instead of preventing Alice's transmission. Therefore, if Mallory chooses to perform as a spoofing attack, it only sends a signal when the Alice is silent. The secrecy data rate of Alice which is attacked can be formulated as

$$R_S = \log_2(1 + \frac{I_{P,l}|g_{ab}|^2}{\sigma_n^2|g_1|^2}) - \gamma\log_2(1 + \frac{P_S|g_{be}|^2}{\sigma_n^2|g_2|^2}) \tag{12}$$

where γ is the impact factor of each spoofing signal, and γ is in the range of [0.1] .

III. SECURE TRANSMISSION GAME

In this work, we study the condition of secure communication which is under the environment of CR network. We model this problem as a non-cooperative static security game. When the secondary user, Alice and Mallory find the spectrum hole, they can use the spectrum resources of primary users, without affecting the communication of primary networks. Therefore, Alice can select to send signals with transmit power in the range of $[0, \frac{I_{P,max}}{|g_1|^2}]$. The intelligent attacker Mallory chooses its attack mode according to the actual situation, that is, $q = 0, 1, 2$ or 3, which corresponds to keeping quiet, eavesdropping, jamming and spoofing, respectively. These attack modes will destroy the Alice's secure communication rate while ensuring that they are not discovered. Alice, instead, needs to maximize its communication security performance, i.e., R_E, R_J and R_S.

Let $f(q)$ be the cost of attack mode q caused by Mallory in this paper. When $q = 1$ holds which represents eavesdropping, the attack cost $f(q)$ is equal to θ_E. Similarly, when $q = 2$ and 3 holds which represents jamming and spoofing, the corresponding attack costs $f(q)$ are equal to θ_J and θ_S, respectively.

In this non-cooperative static secure game, the utility of Alice is related to the confidential capacity and transmit power, and it can be formulated as

$$U_a(P_{Alice}, q) = R_q \ln 2 - C_a P_{Alice}, \tag{13}$$

where C_a represents the Alice's cost by unit transmit power. The q-th element of the secrecy capability vector $[R, R_E, R_J, R_S]$ is denoted by R_q. Take this data rate and multiply by the coefficient $ln 2$, for simplicity. As same as above, the utility of the Mallory is related to the confidential capacity and transmit power, and it can be formulated as

$$U_e(P_{Alice}, q) = -R_q \ln 2 - f(q). \tag{14}$$

In general, the NE strategy of this game is expressed as (P^*_{Alice}, q^*). In order to maximize the Alice's own utility U_a, it needs to choose the transmit power P_{Alice} appropriately. Meanwhile, Mallory needs to select its attack mode to maximize its own utility U_e combined with the actual transmit power of Alice. Neither Alice nor Mallory will benefit from changing the strategy alone. Therefore, in order to maximize the own interest, neither party is willing to change its strategy. From this, we can get the following inequality.

$$U_a(P^*_{Alice}, q^*) \geq U_a(P_{Alice}, q^*), \forall 0 \leq P_{Alice} \leq P^{max}_{Alice} \tag{15}$$
$$U_e(P^*_{Alice}, q^*) \geq U_e(P^*_{Alice}, q), \forall q = 0, 1, 2, 3 \tag{16}$$

Lemma 1: The static secure game has an NE$(x^*_{Alice}, 0)$ *given by*

$$\frac{1}{\sigma_n^2/|g_{ab}|^2 + x^*_{Alice}} = C_a, \tag{17a}$$
$$0 \leq x^*_{Alice} \leq P^{max}_{Alice}, \tag{17b}$$

If $\theta_E \in \mathcal{Y}_1$, $\theta_J \in \mathcal{Y}_2$, $\theta_S \in \mathcal{Y}_3$, *and* $C_a \in \mathcal{Y}_4$ *hold, where*

$$\mathcal{Y}_1 = \left[\ln\left(1 + \frac{x^*_{Alice}|g_{ea}|^2}{\sigma_n^2}\right), \infty\right) \tag{18a}$$

$$\mathcal{Y}_2 = \left[\ln\left(\frac{\sigma^2|g_2|^2 + P_J|g_{be}|^2 + x^*_{Alice}\frac{|g_{ab}|^2}{\sigma^2}(|g_2|^2 + P_J|g_{be}|^2)}{\sigma^2|g_2|^2 + P_J|g_{be}|^2 + x^*_{Alice}|g_{ab}|^2|g_2|^2}\right), \infty\right) \tag{18b}$$

$$\mathcal{Y}_3 = \left[\gamma \ln\left(1 + \frac{P_S|g_{be}|^2}{\sigma_n^2|g_2|^2}\right), \infty\right) \tag{18c}$$

$$\mathcal{Y}_4 = \left[\frac{1}{\sigma_n^2/|g_{ab}|^2 + P^{max}_{Alice}}, \leq \frac{1}{\sigma_n^2/|g_{ab}|^2}\right] \tag{18d}$$

Proof: See Appendix A. □

It can be seen from Lemma 1 that when the cost of attack is much higher than the transmission lost, the incentive to attack disappears (i.e., eqs. (18a)-(18c)). Moreover, in the case of poor channel communication environments and serious information leakage (i.e., eqs. (18d)), Alice will stop transmission.

Lemma 2: The static secure game has an NE $(P^{max}_{Alice}, 0)$, *if* $\theta_E \in \mathcal{Y}'_1$, $\theta_J \in \mathcal{Y}'_2$, $\theta_S \in \mathcal{Y}'_3$, *and* $C_a \in \mathcal{Y}'_4$ *hold, where*

$$\mathcal{Y}'_1 = \left[\ln\left(1 + \frac{P^{max}_{Alice}|g_{ea}|^2}{\sigma_n^2}\right), \infty\right) \tag{19a}$$

$$\mathcal{Y}'_2 = \left[\ln\left(\frac{\sigma^2|g_2|^2 + P_J|g_{be}|^2 + P^{max}_{Alice}\frac{|g_{ab}|^2}{\sigma^2}[|g_2|^2 + P_J|g_{be}|^2]}{\sigma^2|g_2|^2 + P_J|g_{be}|^2 + P^{max}_{Alice}|g_{ab}|^2|g_2|^2}\right), \infty\right) \tag{19b}$$

$$\mathcal{Y}'_3 = \left[\geq \gamma \ln\left(1 + \frac{P_S|g_{be}|^2}{\sigma_n^2|g_2|^2}\right), \infty\right) \tag{19c}$$

$$\mathcal{Y}'_3 = \left[0, \frac{1}{\sigma_n^2/|g_{ab}|^2 + P^{max}_{Alice}}\right] \tag{19d}$$

Proof: See Appendix B. □

Lemma 2 illustrates that Alice prefers to transmit with the maximum power in the case of low transmission cost or high attack cost.

IV. POWER ALLOCATION STRATEGY IN DYNAMIC GAME

In practical communication environments, it is difficult for Alice to predict the attack mode and channel information of Mallory in a certain period of time under the constraint of primary user. Q-learning is a classic and widely used algorithm, which can derive the solution of the non-convex problem. In this work, Alice can learn how to select the optimal transmit power by the Q-learning algorithm when it communicates with Bob in the range of tolerant interference power of primary user to suppress the attack from Mallory effectively. Meanwhile, Mallory chooses the corresponding attack mode according to the cost and the choice of Alice.

As shown in Algorithm 1, Q-learning is a value-based and off-policy algorithm. Let $Q(s, P_{Alice})$ denote the Q-function of Alice, in which s is the system state and the action P_{Alice} is the transmit power of Alice. P_{Alice} is limited by the primary user which should be not larger than the peak level of the tolerant interference power. The Q-function $Q(s, P_{Alice})$ is the expected discounted long-term reward of Alice. The value function $V(s)$ is the maximum of $Q(s, P_{Alice})$.

At time n, the attack mode of Mallory is denoted by q^n. Alice uses the Mallory's attack mode q^{n-1} in the last slot as the system state at time n, which is given by $s^n = q^{n-1}$. In Algorithm 1, we select Alice's action by using the ε-greedy algorithm in a time slot. We randomly explore an action with probability ε and exploit the best action in highest reward Q with probability $1 - \varepsilon$. The learning rate is denoted by β, which determines how much the error is learned in this time slot. And β is a number less than 1. Meanwhile, the decay value of the future rewards is denoted by the discount factor δ, which is the range of [0,1]. In this trial-and-error process, Alice selects its transmit power to maximize its long-term reward, and can adaptively suppress the Mallory's smart attack.

Note that in the secure game involving two users, the action of one user can be regarded as the state of the other user.

Algorithm 1:Power Control With Q-learning

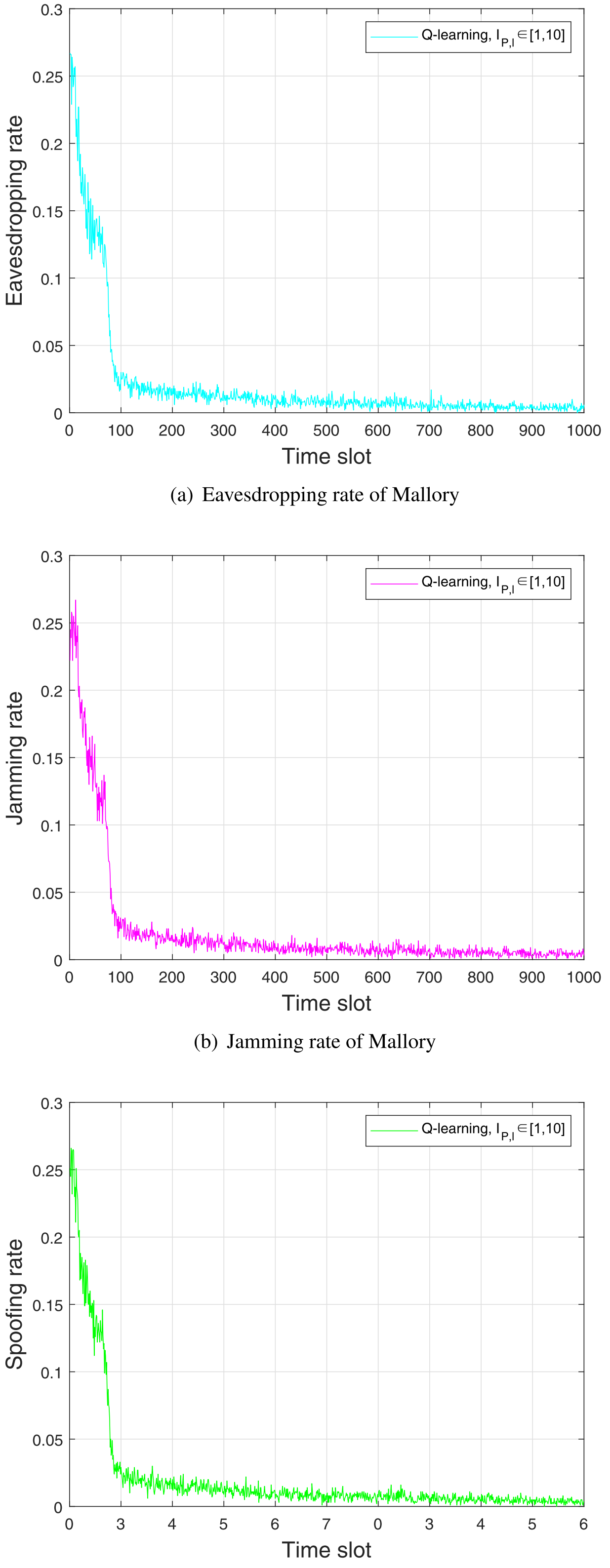
```
Initialize $Q(s, P_{Alice}){=}0$, $V(s){=}0$, $q^0{=}0$, $\forall s, P_{Alice}$;
for $n = 1, 2, 3, \ldots$ do
    Update the state $s^n = q^{n-1}$;
    Choose transmit power $P^n_{Alice}$ by using the $\varepsilon$-greedy algorithm
    Transmit with power $P^n_{Alice}$
    Observe the attack mold $q^n$ and the utility of Alice $U_a$
    Update the value function and Q function:
        $Q(s^n, P^n_{Alice}) = (1-\beta)Q(s^n, P^n_{Alice}) +$
        $\beta(U_a(s^n, P^n_{Alice}) + \delta V(s^{n+1}))$
    $V(s^n){=}\max_{0 \le P \le P^{max}_{Alice}} Q(s^n, P_{Alice})$
end for
```

Accordingly, we regard attacking mode of Mallory, i.e., the action of Mallory, as the state space of Alice, which is denoted as $q = 0, 1, 2$ and 3. Moreover, we discretize the maximum transmit power P_{max} into $L+1$ levels, and define the transmit power level $P_{Alice} \in \{lP_{max}/L\}_{0 \le l \le L}$ as the action of Alice, which is also regarded as the state of Mallory. In further, the state transition probability of the Markov states is not known prior, and hence we use the Q-learning algorithm to solve the secure game, which does not need the state transition probability.

In order to execute the Q-learning algorithm, the system needs to observe the attacking mode of Mallory and the utility of Alice. Although such information is maybe difficult to obtain in practice, it is meaningful to study with known information of Mallory's mode and Alice's utility, in the following three folds. Firstly, such information can be obtained through some signal processing methods, such as using some pilot signals in the system to estimate the required channel parameters and the Mallory's mode. Secondly, if we cannot obtain the accurate data of the required information, we can try to obtain some statistical value, through some estimation methods, such as estimating the location of the Mallory. Thirdly, even if we cannot obtain any information of the Mallory's mode and Alice's utility, the study in our work can still serve as a useful benchmark, and help obtain some insights on the secure system.

V. SIMULATION RESULTS

The performance of the proposed Q-learning algorithm was evaluated in this section. In order to implement the algorithm and simulate the practical communication environments, we set $\sigma^2_{ab} = 1.2$, $\sigma^2_{ea} = 0.1$, $\sigma^2_{be} = 3$, $\sigma^2_{g1} = 6$, and $\sigma^2_{g2} = 4.2$ as the average channel gains [1] [50]. We denote the peak interference power when the Mallory performs jamming and spoofing by $P_J = 7.4$ and $P_S = 7.2$, respectively. The cost of transmit power for Alice C_a is set to 0.1, and the impact factor of each spoofing signal γ is set to 0.5.



(a) Eavesdropping rate of Mallory

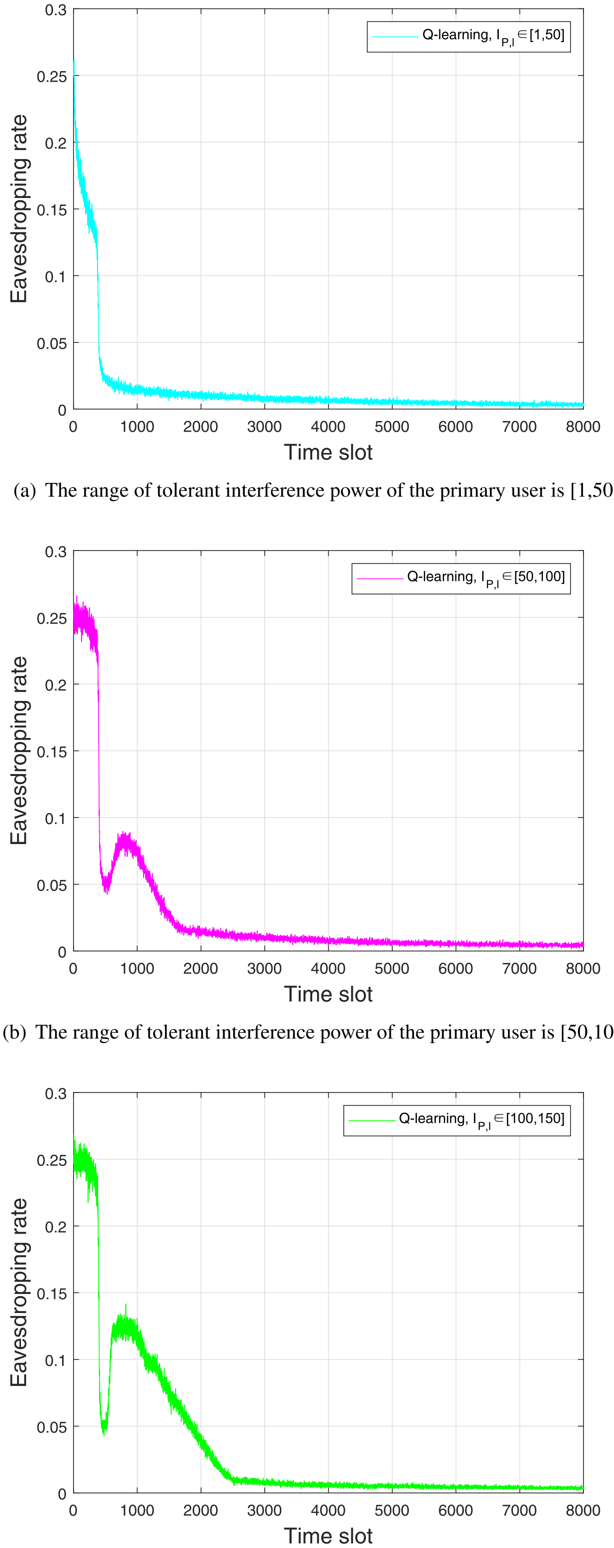
(b) Jamming rate of Mallory

(c) Spoofing rate of Mallory

FIGURE 2. Anti-attack performance when the tolerant interference power of PU is in the range of [1,10].

Fig. 2 shows the anti-attack performance when the tolerant interference power of the primary user ranges from 1 to

[1] Note that the node location is actually used in the secure game, since the statistical channel gains are related to the distance between the nodes.

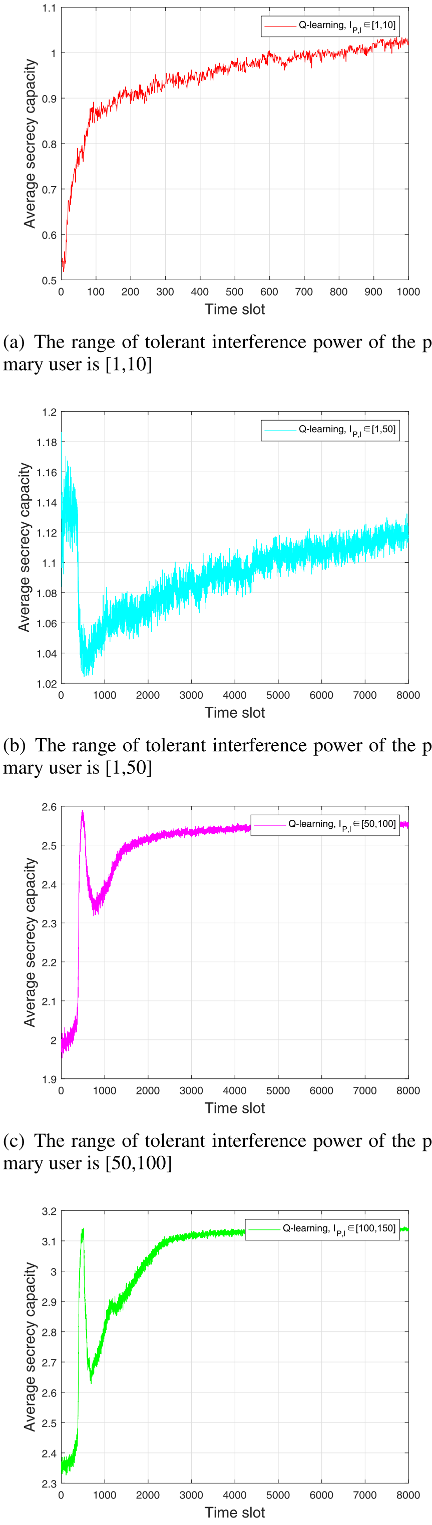


(a) The range of tolerant interference power of the primary user is [1,50]

(b) The range of tolerant interference power of the primary user is [50,100]

(c) The range of tolerant interference power of the primary user is [100,150]

FIGURE 3. The eavesdropping rate of Mallory with respect to the tolerant interference power of the primary user.



(a) The range of tolerant interference power of the primary user is [1,10]

(b) The range of tolerant interference power of the primary user is [1,50]

(c) The range of tolerant interference power of the primary user is [50,100]

(d) The range of tolerant interference power of the primary user is [100,150]

FIGURE 4. The average secrecy capacity of Alice with respect to the tolerant interference power of the primary user.

10. Specifically, Fig. 2 (a), (b) and (c) are associated with the eavesdropping rate, jamming rate and spoofing rate of Mallory, respectively. As can be seen from Fig. 2, there is a decreasing trend in the attack rate of Mallory after many times of training and learning, and it tends to zero gradually. For example, as can be seen from Fig. 2 (a), there is an evident

decline in eavesdropping rate from 25% at the beginning to almost zero after 1000 time slots. Similarly, both the jamming rate and the spoofing rate of Mallory decrease significantly and finally tend to zero. Fig. 2 indicates that Alice can learn how to select the transmit power when it contacts with Bob by the training of Q-learning algorithm in the specific range. In this situation, Mallory tends not to attack because the cost of the attack is too high. In other words, Alice can suppress the attack behavior of Mallory when it communicates with Bob, which further proves that the proposed Q-learning algorithm can achieve the purpose of secure communication.

Fig. 3 shows the eavesdropping rate of Mallory with respect to the range of tolerant interference power of the primary user. In particular, Fig. 3 (a), (b) and (c) correspond to the tolerant interference power of primary user in the range of [1,50], [50,100] and [100,150], respectively. We can see from Fig. 3 that the proposed Q-learning algorithm can reduce the eavesdropping rate effectively. For instance, as shown in Fig. 3 (a), the eavesdropping rate falls below 0.025 after 500 time slots and it tends to zero as the time slot increases when the tolerant interference power of PU is in the range of [1,50]. Similarly, the eavesdropping rate begins to show a downward trend after 1500 time slots when the tolerant interference power of PU is in the range of [50,100] and [100,150], respectively. Finally, they all converge to zero when the time slots are larger than 3000. Simulation result in Fig.3 validates that the proposed Q-learning algorithm can make Alice select the optimal transmit power so that it can suppress the attack rate of the attacker in any range of the tolerant interference power of the primary user. In further, the jamming rate and the spoofing rate also decrease significantly and converge to zero in the same range of tolerant interference power.

Fig. 4 shows the average secrecy capacity of Alice with respect to the tolerant interference power of the primary user. Fig. 4 (a), (b), (c) and (d) are associated with the tolerant interference power of primary user in the range of [1,10], [1,50], [50,100] and [100,150], respectively. As observed from Fig. 4, we can find that the Alice's average secrecy capacity increases on the whole as the number of training increases. For example, as shown in Fig. 4 (a), the average secrecy capacity of Alice based on the Q-learning algorithm increases dramatically with a rise of around 50%. We can observe from Fig. 4 (b) that there is an obvious increase after 1000 time slots in the secrecy capacity. Similarly, the average secrecy capacity of Alice continues to rise after a short period of decline both in Fig. 4 (c) and (d). In further, the secrecy capacity in Fig. 4 (c) and (d) is much more stable than that in Fig. 4 (b) after 3000 time slots. Simulation result in Fig. 4 demonstrates that the average secrecy capacity of Alice can be improved after learning and can move towards maximization.

VI. CONCLUSION

In this work, we have investigated the secure transmission problem under the smart attack in cognitive networks. The secondary users, Alice and Mallory, were allowed to utilize spectrum resources which were also used by primary user. The attacker, Mallory, had three attack modes including eavesdropping, jamming, and spoofing. We formulated an NE strategy game to maximize the utility of the transmitter and meanwhile minimized its damage from the attacker. The Q-learning algorithm was utilized to control the transmit power of the transmitter and determine the attack mode of the smart attacker. The employed Q-learning algorithm enabled the transmitter to obtain the optimal transmit power during the learning stage in the range of tolerant interference power of the primary user which hence suppressed the attacker eventually. Simulation results were provided to show that the algorithm could effectively and clearly achieve the expected target, which suppressed the attack behavior of the attacker. In future works, we will consider some learning based algorithms [51], [52], especially the deep learning based algorithms [53]–[55], to the considered system, in order to enhance the system performance.

APPENDIXES

APPENDIX A

PROOF OF LEMMA 1

If eqs. (18a)-(18c) hold, from (14), we have Thus, (16) holds for $(x^*_{Alice}, 0)$. From (13), we have (A.1a), (A.1b), and (A.1c) as shown at the top of the next page,

$$\frac{\partial U_a(P_{Alice}, 0)}{\partial P_{Alice}} = \frac{1}{\sigma_n^2/|g_{ab}|^2 + P_{Alice}} - C_a, \tag{A.2}$$

$$\frac{\partial^2 U_a(P_{Alice}, 0)}{\partial P^2_{Alice}} = -\Big(\frac{1}{\sigma_n^2/|g_{ab}|^2 + P_{Alice}}\Big)^2 \leq 0. \tag{A.3}$$

The above formulas show that $\partial U_a(P_{Alice}, 0)/\partial P_{Alice}$ decreases monotonically with respect to P_{Alice}. Thus, if (18d) holds, from (A.2) we have

$$\frac{\partial U_a(P_{Alice},0)}{\partial P_{Alice}}\bigg|_{P_{Alice}=0} = \frac{1}{\sigma_n^2/|g_{ab}|^2} - C_a > 0, \tag{A.4}$$

$$\frac{\partial U_a(P_{Alice},0)}{\partial P_{Alice}}\bigg|_{P_{Alice}=P^{max}_{Alice}} = \frac{1}{\sigma_n^2/|g_{ab}|^2 + P^{max}_{Alice}} - C_a < 0, \tag{A.5}$$

indicating that $\partial U_a(P_{Alice}, 0)/\partial P_{Alice} = 0$ has only a sole solution because of the formula (17a). From (A.2)-(A.4) we can know that when P_{Alice} is smaller than x^*_{Alice}, $U_a(P_{Alice}, 0)$ monotonically increases. On the contrary, when P_{Alice} is larger than x^*_{Alice}, $U_a(P_{Alice}, 0)$ monotonically decreases, which means that $U_a(P_{Alice}, 0)$ has a maximum value. Thus, (15) holds and $(x^*_{Alice}, 0)$ is an NE of this game. In this way, we have completed the proof of Lemma 1.

APPENDIX B

PROOF OF LEMMA 2

Similar to the proof in Lemma 1, if eqs. (19a)-(19c) hold, from (14), we have

$$U_e(P^{max}_{Alice}, 0) - U_e(P^{max}_{Alice}, 1) = \theta_E - \ln\left(1 + \frac{P^{max}_{Alice}|g_{ea}|^2}{\sigma_n^2|g_1|^2}\right) \geq 0. \tag{B.1}$$

In the same way, $U_e(P^{max}_{Alice}, 0) \geq U_e(P^{max}_{Alice}, 2)$ and $U_e(P^{max}_{Alice}, 0) \geq U_e(P^{max}_{Alice}, 3)$ imply that (16) holds. The

$$U_e(x^*,0)-U_e(x^*,1)=\theta_E-\ln\left(1+\frac{x^*_{Alice}|g_{ea}|^2}{\sigma_n^2}\right)\geq 0, \tag{A.1a}$$

$$U_e(x^*,0)-U_e(x^*,2)=\theta_J-\ln\left(\frac{\sigma^2|g_2|^2+P_J|g_{be}|^2+x^*_{Alice}\frac{|g_{ab}|^2}{\sigma^2}(|g_2|^2+P_J|g_{be}|^2)}{\sigma^2|g_2|^2+P_J|g_{be}|^2+x^*_{Alice}|g_{ab}|^2|g_2|^2}\right), \tag{A.1b}$$

$$U_e(x^*,0)-U_e(x^*,3)=\theta_S-\gamma\ln\left(1+\frac{P_S|g_{be}|^2}{\sigma_n^2|g_2|^2}\right)\geq 0. \tag{A.1c}$$

above formulas, show that $\partial U_a(P_{Alice},0)/\partial P_{Alice}$ decreases monotonically with respect to P_{Alice}, and we have

$$\frac{\partial U_a(P_{Alice},0)}{\partial P_{Alice}}\geq\frac{\partial U_a(P_{Alice},0)}{\partial P_{Alice}}\bigg|_{P_{Alice}=P^{max}_{Alice}}\geq 0,$$
$$\forall 0\leq P_{Alice}\leq P^{max}_{Alice}. \tag{B.2}$$

This implies that (15) holds and $(P^{max}_{Alice},0)$ is an NE of this game. In this way, we have completed the proof of Lemma 2.

REFERENCES

[1] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, "Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3393–3406, Jul. 2018.

[2] J. Xia, L. Fan, W. Xu, X. Lei, X. Chen, G. K. Karagiannidis, and A. Nallanathan, "Secure cache-aided multi-relay networks in the presence of multiple eavesdroppers," *IEEE Trans. Commun.*, vol. 67, no. 11, pp. 7672–7685, Nov. 2019.

[3] X. Hu, C. Zhong, X. Chen, W. Xu, and Z. Zhang, "Cluster grouping and power control for angle-domain MmWave MIMO NOMA systems," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 5, pp. 1167–1180, Sep. 2019.

[4] W. Xu, J. Liu, S. Jin, and X. Dong, "Spectral and energy efficiency of multi-pair massive MIMO relay network with hybrid processing," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3794–3809, Sep. 2017.

[5] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.

[6] C. Lu, W. Xu, S. Jin, and K. Wang, "Bit-level optimized neural network for multi-antenna channel quantization," *IEEE Wireless Commun. Lett.*, vol. 9, no. 1, pp. 87–90, Jan. 2020.

[7] J. Xia, D. Deng, Y. Rao, D. Li, F. Zhu, and L. Fan, "When distributed switch-and-stay combining meets buffer in IoT relaying networks," *Phys. Commun.*, vol. 38, Feb. 2020, Art. no. 100920.

[8] Z. Zhao, R. Zhao, J. Xia, X. Lei, D. Li, C. Yuen, and L. Fan, "A novel framework of three-hierarchical offloading optimization for MEC in industrial IoT networks," *IEEE Trans Ind. Informat.*, to be published.

[9] L. Fan, N. Zhao, X. Lei, Q. Chen, N. Yang, and G. K. Karagiannidis, "Outage probability and optimal cache placement for multiple Amplify-and-Forward relay networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 12373–12378, Dec. 2018.

[10] X. Lin, Y. Tang, X. Lei, J. Xia, Q. Zhou, H. Wu, and L. Fan, "MARL-based distributed cache placement for wireless networks," *IEEE Access*, vol. 7, pp. 62606–62615, 2019.

[11] J. Xia, C. Li, X. Lai, S. Lai, F. Zhu, D. Deng, and L. Fan, "Cache-aided mobile edge computing for B5G wireless communication networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2020, no. 1, pp. 1–5, Jan. 2020.

[12] Y. Guo, Z. Zhao, R. Zhao, S. Lai, Z. Dan, J. Xia, and L. Fan, "Intelligent offloading strategy design for relaying mobile edge computing networks," *IEEE Access*, to be published.

[13] J. Zhao, X. Guan, and X. P. Li, "Power allocation based on genetic simulated annealing algorithm in cognitive radio networks," *Chin. J. Electron.*, vol. 22, no. 1, pp. 177–180, Jan. 2013.

[14] J. Zhao, S. Ni, L. Yang, Z. Zhang, Y. Gong, and X. You, "Multiband cooperation for 5G HetNets: A promising network paradigm," *IEEE Veh. Technol. Mag.*, vol. 14, no. 4, pp. 85–93, Dec. 2019.

[15] Z. Junhui, Y. Tao, G. Yi, W. Jiao, and F. Lei, "Power control algorithm of cognitive radio based on non-cooperative game theory," *China Commun.*, vol. 10, no. 11, pp. 143–154, Nov. 2013.

[16] J. Zhao, Q. Li, Y. Gong, and K. Zhang, "Computation offloading and resource allocation for cloud assisted mobile edge computing in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7944–7956, Aug. 2019.

[17] S. Ni, J. Zhao, H. H. Yang, and Y. Gong, "Enhancing downlink transmission in MIMO HetNet with wireless backhaul," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6817–6832, Jul. 2019.

[18] X. Lai, W. Zou, D. Xie, X. Li, and L. Fan, "DF relaying networks with randomly distributed interferers," *IEEE Access*, vol. 5, pp. 18909–18917, 2017.

[19] M. Jia, X. Wang, Q. Guo, I. W.-H. Ho, X. Gu, and F. C.-M. Lau, "Performance analysis of cooperative non-orthogonal multiple access based on spectrum sensing," *IEEE Trans. Veh. Technol.*, vol. 68, no. 7, pp. 6855–6866, Jul. 2019.

[20] Z. Na, Y. Wang, X. Li, J. Xia, X. Liu, M. Xiong, and W. Lu, "Subcarrier allocation based simultaneous wireless information and power transfer algorithm in 5G cooperative OFDM communication systems," *Phys. Commun.*, vol. 29, pp. 164–170, Aug. 2018.

[21] Z. Na, J. Lv, M. Zhang, B. Peng, M. Xiong, and M. Guan, "GFDM based wireless powered communication for cooperative relay system," *IEEE Access*, vol. 7, pp. 50971–50979, 2019.

[22] M. Jia, X. Zhang, X. Gu, Q. Guo, Y. Li, and P. Lin, "Interbeam interference constrained resource allocation for shared spectrum multibeam satellite communication systems," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6052–6059, Aug. 2019.

[23] Z. Chen, F. Gao, X. Zhang, J. C. F. Li, and M. Lei, "Sensing and power allocation for cognitive radio with multiple primary transmit powers," *IEEE Wireless Commun. Lett.*, vol. 2, no. 3, pp. 319–322, Jun. 2013.

[24] X. Lai, L. Fan, X. Lei, J. Li, N. Yang, and G. K. Karagiannidis, "Distributed secure Switch-and-Stay combining over correlated fading channels," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 8, pp. 2088–2101, Aug. 2019.

[25] J. Yang, D. Ruan, J. Huang, X. Kang, and Y.-Q. Shi, "An embedding cost learning framework using GAN," *IEEE Trans. Inf. Forensics Security*, vol. 15, no. 1, pp. 839–851, 2020.

[26] F. Shi, J. Xia, Z. Na, X. Liu, Y. Ding, and Z. Wang, "Secure probabilistic caching in random multi-user multi-UAV relay networks," *Phys. Commun.*, vol. 32, pp. 31–40, Feb. 2019.

[27] M. Jia, Z. Yin, D. Li, Q. Guo, and X. Gu, "Toward improved offloading efficiency of data transmission in the IoT-cloud by leveraging secure truncating OFDM," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4252–4261, Jun. 2019.

[28] N. Yang, L. Wang, G. Geraci, M. Elkashlan, J. Yuan, and M. D. Renzo, "Safeguarding 5G wireless communication networks using physical layer security," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 20–27, Apr. 2015.

[29] M. Jia, D. Li, Z. Yin, Q. Guo, and X. Gu, "High spectral efficiency secure communications with nonorthogonal physical and multiple access layers," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 5954–5961, Aug. 2019.

[30] S. Pan, Z. Liu, and W. Lu, "Synthesis of naked plasmonic/magnetic Au/Fe3O4 nanostructures by plasmon-driven anti-replacement reaction," *Nanotechnology*, vol. 30, no. 6, Dec. 2018, Art. no. 065605.

[31] S. Pan, X. Zhang, W. Lu, and S. F. Yu, "Plasmon-engineered anti-replacement synthesis of naked cu nanoclusters with ultrahigh electrocatalytic activity," *J. Mater. Chem. A*, vol. 6, no. 38, pp. 18687–18693, 2018.

[32] Y. K. Wang, Z. M. Xie, M. M. Wang, H. W. Deng, J. F. Yang, Y. Jiang, T. Zhang, X. P. Wang, Q. F. Fang, and C. S. Liu, "The superior thermal stability and tensile properties of hot rolled W-HfC alloys," *Int. J. Refractory Met. Hard Mater.*, vol. 81, pp. 42–48, Jun. 2019.

[33] M. M. Wang, Z. M. Xie, H. W. Deng, J. F. Yang, Y. K. Wang, T. Zhang, Y. Xiong, X. P. Wang, Q. F. Fang, and C. S. Liu, "Grain size effects of tungsten powder on the micro-structure and mechanical properties of tungsten-based alloys," *Mater. Sci. Eng. A*, vol. 754, pp. 216–223, Apr. 2019.

[34] X. Lin, J. Xia, and Z. Wang, "Probabilistic caching placement in UAV-assisted heterogeneous wireless networks," *Phys. Commun.*, vol. 33, pp. 54–61, Apr. 2019.

[35] J. Xia, Y. Xu, D. Deng, Q. Zhou, and L. Fan, "Intelligent secure communication for Internet of Things with statistical channel state information of attacker," *IEEE Access*, vol. 7, pp. 144481–144488, 2019.

[36] Y. Xu, J. Xia, H. Wu, and L. Fan, "Q-learning based physical-layer secure game against multiagent attacks," *IEEE Access*, vol. 7, pp. 49212–49222, 2019.

[37] C. Li, Y. Xu, J. Xia, and J. Zhao, "Protecting secure communication under UAV smart attack with imperfect channel estimation," *IEEE Access*, vol. 6, pp. 76395–76401, 2018.

[38] L. Xiao, H. Zhang, Y. Xiao, X. Wan, S. Liu, L.-C. Wang, and H. V. Poor, "Reinforcement learning-based downlink interference control for ultra-dense small cells," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 423–434, Jan. 2020.

[39] C. Li, W. Zhou, K. Yu, L. Fan, and J. Xia, "Enhanced secure transmission against intelligent attacks," *IEEE Access*, vol. 7, pp. 53596–53602, 2019.

[40] C. Li, Z. Gao, J. Xia, D. Deng, and L. Fan, "Cache-enabled physical-layer secure game against smart UAV-assisted attacks in b5G NOMA networks," *EURASIP J. Wireless Commun. Netw.*, vol. 2020, no. 1, pp. 1–5, Jan. 2020.

[41] B. Lu, "Interference suppression by exploiting wireless cache in relaying networks for b5g communications," *Phys. Commun.*, to be published.

[42] W. Huang, "Multi-antenna processing based cache-aided relaying networks for b5g communications," *Phys. Commun.*, to be published.

[43] R. Zhao, "Deep reinforcement learning based mobile edge computing for intelligent Internet of Things," *IEEE Access*, to be published.

[44] Z. Zhao, W. Zhou, D. Dan, J. Xia, and L. Fan, "Intelligent mobile edge computing with pricing in Internet of Things," *IEEE Access*, to be published.

[45] D. Deng, J. Xia, L. Fan, and X. Li, "Link selection in buffer-aided cooperative networks for green IoT," *IEEE Access*, vol. 8, pp. 30763–30771, 2020.

[46] X. Wang, "Joint resource allocation for cognitive OFDM-NOMA systems with energy harvesting in green IoT," *IEEE Access*, to be published.

[47] J. Li, Y. Peng, Y. Yan, X.-Q. Jiang, H. Hai, and M. Zukerman, "Cognitive radio network assisted by OFDM with index modulation," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 1106–1110, Jan. 2020.

[48] J. Li, S. Dang, M. Wen, X.-Q. Jiang, Y. Peng, and H. Hai, "Layered orthogonal frequency division multiplexing with index modulation," *IEEE Syst. J.*, vol. 13, no. 4, pp. 3793–3802, Dec. 2019.

[49] J. Li, M. Wen, X. Cheng, Y. Yan, S. Song, and M. H. Lee, "Generalized precoding-aided quadrature spatial modulation," *IEEE Trans. Veh. Technol.*, vol. 66, no. 2, pp. 1881–1886, Feb. 2017.

[50] M. Jia, L. Wang, Q. Guo, X. Gu, and W. Xiang, "A low complexity detection algorithm for fixed up-link SCMA system in mission critical scenario," *IEEE Internet Things J.*, vol. 5, no. 5, pp. 3289–3297, Oct. 2018.

[51] J. Xia, "A MIMO detector with deep learning in the presence of correlated interference," *IEEE Trans. Veh. Technol.*, to be published.

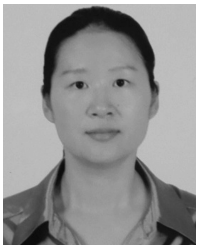
[52] G. Liu, Y. Xu, Z. He, Y. Rao, J. Xia, and L. Fan, "Deep learning-based channel prediction for edge computing networks toward intelligent connected vehicles," *IEEE Access*, vol. 7, pp. 114487–114495, 2019.

[53] M.-S. Baek, S. Kwak, J.-Y. Jung, H. M. Kim, and D.-J. Choi, "Implementation methodologies of deep learning-based signal detection for conventional MIMO transmitters," *IEEE Trans. Broadcast.*, vol. 65, no. 3, pp. 636–642, Sep. 2019.

[54] K. He, "Ultra-reliable MU-MIMO detector based on deep learning for 5G/B5G-enabled IoT," *EURASIP J. Wireless Commun. Netw.*, to be published.

[55] K. He, Z. Wang, W. Huang, D. Deng, J. Xia, and L. Fan, "Generic deep learning-based linear detectors for MIMO systems over correlated noise environments," *IEEE Access*, vol. 8, pp. 29922–29929, 2020.

SHIWEI LAI received the bachelor's degree in computer science and technology from the Guangdong University of Education, in June 2019. She is currently a Graduate Student with the School of Computer Science and Cyber Engineering, Guangzhou University.



JUNJUAN XIA received the bachelor's degree from the Department of Computer Science, Tianjin University, in 2003, and the master's degree from the Department of Electronic Engineering, Shantou University, in 2015. She is currently a Laboratory Assistant with the School of Computer Science and Cyber Engineering, Guangzhou University. Her current research interests include wireless caching, physical-layer security, cooperative relaying, and interference modeling.

DAN ZOU received the M.S. degree from East China Jiaotong University, in 2008, where she is currently pursuing the Ph.D. degree. Her research interests include vehicular networks, mobile edge computing, and resource allocation.

LISENG FAN received the Ph.D. degree from the Tokyo Institute of Technology, Tokyo, in 2008. He is currently a Professor with the School of Computer Science and Cyber Engineering, Guangzhou University. He has published more than 40 articles on the IEEE Journal and IEEE conferences. His main research interests include the information security, wireless networks, and the artificial intelligence. His recent research interest is the application of artificial intelligence into the wireless networks.