Received March 17, 2019, accepted April 3, 2019, date of publication April 11, 2019, date of current version April 24, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2910272

Q-Learning Based Physical-Layer Secure Game Against Multiagent Attacks

YAN XU[1], JUNJUAN XIA[1], HUIJUN WU[2], AND LISENG FAN[1]
[1]School of Computer Science, Guangzhou University, Guangzhou 510006, China
[2]School of Civil Engineering, Guangzhou University, Guangzhou 510006, China

Corresponding author: Liseng Fan (lsfan2019@126.com)

This work was supported by the Yangcheng Scholar Research Project under Grant 1201561560.

ABSTRACT In this paper, we consider a Q-learning-based power allocation strategy for a secure physical-layer system under dynamic radio environments. In such a system, the transmitter sends the information to the receiver threatened by $M(M \geq 2)$ intelligent attackers which have several attack modes and will bring out the severe issue of information security. To safeguard the system security, we formulate the insecure problem as a stochastic game which consists of $M+1$ players: the transmitter which can flexibly choose its transmit power, and M smart attackers that can determine their attack types. Then, the Nash equilibria (NEs) of the physical-layer secure game are derived, and their existence conditions are taken into account. The simulation results show that the proposed power allocation strategy in the stochastic game can efficiently suppress the attack rate of smart attackers even if there exist multiple smart attackers.

INDEX TERMS Q-learning, power allocation, smart attacks, stochastic game.

I. INTRODUCTION

In recent years, there has been a tremendous trend in the development of wireless communication technology [1]–[4], and user's demands for the transmission performance of wireless communication have been continuously increasing [5]–[8]. Hence, it is of vital importance to safeguard the secure transmission from the physical-layer of wireless transmission [9], [10]. Due to the natural openness and broadcast characteristics, wireless communication is vulnerable to be threatened by the attackers in networks [11], [12]. The smart attackers can use smart and programmable radio devices to flexibly select their attack types, such as eavesdropping [13] and jamming [14]–[16]. For example, the attacker can select to overhear the secure message when it is close to the transmitter, while it can select to interfere with the receiver when it is close to the receiver. In this way, the attacker can efficiently attack the legitimate users, which results in the severe issue of information leakage.

In order to safeguard the wireless communication, reinforcement learning has been applied into the wireless networks to improve the secrecy performance [17]–[21]. For example, the authors in [17] studied the reinforcement learning to suppress the smart jamming attack, and the authors in [18] investigated a transmission game against smart attacks such as eavesdropping, jamming and spoofing. The authors in [19] introduced a mobile offloading game, where the security agent could help the mobile user to ensure its offloading rate under the attack from a smart attacker. In addition to the reinforcement learning, deep learning is recently used in the field of wireless communication, which opens a new research direction on the combination of the machine learning and wireless communication [22]–[29]. For example, the authors in [23] investigated a deep learning scheme for super-resolution channel estimation based massive multiple-input multiple-output (MIMO) system, and the authors in [22] and [28] applied deep learning to the non-orthogonal multiple access (NOMA) system.

The works above considered the secure threaten from only one smart attacker. However, in the practical wireless networks, there are maybe multiple attackers, which severely degrade the network secrecy performance. Hence, it is quite of importance to safeguard the network security against multiple smart attackers, which motivates our research. Moreover, we consider some common attack modes and choose to use the two typical attack modes, i.e., eavesdropping and jamming.

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

In this paper, we investigate a Q-learning based physical-layer secure game, in which multiple smart attackers can perform silent, eavesdropping and jamming attacks. The transmit power of the transmitter can be adaptively adjusted to improve the network secrecy performance. To this end, the Q-learning based power allocation strategy is adopted by the transmitter to choose the current transmit power, which is based on the previous actions of all attackers. Especially, the Nash equilibria (NEs) of the physical-layer secure game are analyzed, where the players obtain high utilities, and no one is willing to change the strategy to break the equilibrium. The sequential interaction among the transmitter and attackers is modeled as a Stackelberg game [30], in which the transmitter firstly chooses the transmit power, and then the attacker determines its attack mode based on the transmit power and attack modes of other attackers. In further, we derive the NEs of this secure game and the existing conditions are also given. Simulation results are finally provided to demonstrate that the proposed scheme can efficiently decrease the attack rate of attackers and improve the network secrecy performance.

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

- We formulate a physical-layer secure game against multiagent attacks, and investigate the interaction among the transmitter and multiple smart attackers.
- We derive the NEs of the physical-layer secure game, and provide the existing conditions of the equilibria.
- We apply the Q-learning based power allocation strategy under dynamic radio environments to improve the security and reduce the attackers' attack rate.

We organize the rest of this paper as follows. We present the system model in Section II. In Section III, we formulate the physical-layer secure game, and we investigate a power allocation policy in Section IV. In Section V, we provide the simulation results and give some conclusions in Section VI.

II. SYSTEM MODEL

We consider a secure communication system, shown in Fig. 1, consisting of a transmitter (Alice), a receiver (Bob) and M smart attackers (attacker 1, …, attacker M), in which they are all equipped with a single antenna, due to the limitation in antenna size [31]–[33]. Alice communicates with Bob through the main link, and its transmit power is denoted by $P \in [0, P_{max}]$, where P_{max} is defined as the maximum transmit power. The smart attacker m ($m \in [1, M]$) chooses its attack mode, among the silent, eavesdropping and jamming modes. And we assume that the attackers don't cooperate with each other in the game. Let q_m denote the specific attack mode of the attacker m, where $q_m \in \{0, 1, 2\}$. Specifically, attacker m keeps silent when $q_m = 0$; it overhears the Alice's signals when $q_m = 1$; and it transmits jamming signals to block the transmission of the main link when $q_m = 2$.

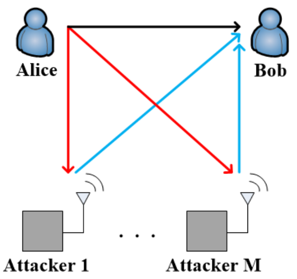


FIGURE 1. A secure communication system under multiple attackers, where the black, red and blue lines represent the main, eavesdropping and jamming links, respectively.

Alice sends a normalized signal x_a to Bob, and then Bob receives the signal y_b,

$$y_b = h_{ab}x_a + n_b, \tag{1}$$

where $h_{ab} \sim \mathcal{CN}(0, \sigma_{ab}^2)$ is the channel parameter of the main link and $n_b \sim \mathcal{CN}(0, \sigma_n^2)$ is the additive white Gaussian noise (AWGN) at Bob. The details about the effect of noise on the transmission performance can be found in the literature, such as works [34]–[37].

Without loss of generality, we consider two smart attackers, Eve1 and Eve2, so that $M = 2$. As q_1 and q_2 vary in {0, 1, 2}, there are totally 9 attack cases, which are detailed as follows:

- When $q_1 = 0$ and $q_2 = 0$ hold, Eve1 and Eve2 choose to keep silent. According to the Shannon theory [38], the secrecy data rate of the main link can be written as

$$R_{0,0} = \log_2(1 + \frac{P|h_{ab}|^2}{\sigma_n^2}). \tag{2}$$

- When $q_1 = 1$ and $q_2 = 0$ hold, Eve1 chooses to overhear the message, while Eve2 selects to keep silent. The received signal at Eve1 is

$$y_{E1} = h_{ae1}x_a + n_{e1}, \tag{3}$$

where $h_{ae1} \sim \mathcal{CN}(0, \sigma_{ae1}^2)$ is the channel parameter from Alice to Eve1 link and $n_{e1} \sim \mathcal{CN}(0, \sigma_n^2)$ is the AWGN at Eve1. Thus, the achievable secrecy data rate can be written as

$$R_{1,0} = \log_2(1 + \frac{P|h_{ab}|^2}{\sigma_n^2}) - \log_2(1 + \frac{P|h_{ae1}|^2}{\sigma_n^2}). \tag{4}$$

- When $q_1 = 2$ and $q_2 = 0$ hold, Eve1 chooses to transmit the jamming signals to block the Alice's signals at Bob, while Eve2 selects to keep silent. The received signal at Bob is

$$y_{J1} = h_{ab}x_a + h_{e1}x_a + n_b, \tag{5}$$

where $h_{e1} \sim \mathcal{CN}(0, \sigma_{e1}^2)$ is the channel parameter from Eve1 to Bob link. Thus, the achievable secrecy data rate can be written as

$$R_{2,0}=\log_2(1+\frac{P|h_{ab}|^2}{\sigma_n^2+P_{J1}|h_{e1}|^2}), \quad (6)$$

where P_{J1} is the jamming power of Eve1.

- When $q_1=0$ and $q_2=1$ hold, Eve1 chooses to keep silent, while Eve2 selects to overhear the message. The received signal at Eve2 is
$$y_{E2}=h_{ae2}x_a+n_{e2}, \quad (7)$$
where $h_{ae2}\sim\mathcal{CN}(0,\sigma_{ae2}^2)$ is the channel parameter from Alice to Eve2 link and $n_{e2}\sim\mathcal{CN}(0,\sigma_n^2)$ is the AWGN at the Eve2. Thus, the achievable secrecy data rate can be written as
$$R_{0,1}=\log_2(1+\frac{P|h_{ab}|^2}{\sigma_n^2})-\log_2(1+\frac{P|h_{ae2}|^2}{\sigma_n^2}). \quad (8)$$
- When $q_1=1$ and $q_2=1$ hold, Eve1 and Eve2 choose to overhear the message. Eve1 receives the signal y_{E1}, and Eve2 receives the signal y_{E2}. Accordingly, the achievable secrecy data rate can be written as
$$R_{1,1}=\log_2(1+\frac{P|h_{ab}|^2}{\sigma_n^2}) - \max[\log_2(1+\frac{P|h_{ae1}|^2}{\sigma_n^2}),\log_2(1+\frac{P|h_{ae2}|^2}{\sigma_n^2})]. \quad (9)$$
- When $q_1=2$ and $q_2=1$ hold, Eve1 chooses to transmit the jamming signals to block the Alice's signals at Bob and block Eve2's eavesdropping signals at the same time. Eve2 selects to overhear the message. The received signal at Eve2 becomes
$$y'_{E2}=h_{ae2}x_a+h_ex_a+n_{e2}, \quad (10)$$
where $h_e\sim\mathcal{CN}(0,\sigma_e^2)$ is the channel parameter from Eve1 to Eve2 link. The achievable secrecy data rate can be written as
$$R_{2,1}=\log_2(1+\frac{P|h_{ab}|^2}{\sigma_n^2+P_{J1}|h_{e1}|^2})-\log_2(1+\frac{P|h_{ae2}|^2}{\sigma_n^2+P_{J1}|h_e|^2}). \quad (11)$$
- When $q_1=0$ and $q_2=2$ hold, Eve1 chooses to keep silent, while Eve2 selects to transmit the jamming signals to block the Alice's signals at Bob. The received signal at Bob becomes
$$y_{J2}=h_{ab}x_a+h_{e2}x_a+n_b, \quad (12)$$
where $h_{e2}\sim\mathcal{CN}(0,\sigma_{e2}^2)$ is the channel parameter from Eve2 to Bob link. Accordingly, the achievable secrecy data rate can be written as
$$R_{0,2}=\log_2(1+\frac{P|h_{ab}|^2}{\sigma_n^2+P_{J2}|h_{e2}|^2}), \quad (13)$$
where P_{J2} is the jamming power of Eve2.
- When $q_1=1$ and $q_2=2$ hold, Eve1 chooses to overhear the message, while Eve2 selects to transmit the jamming signals to block the Alice's signals at Bob and block Eve1's eavesdropping signals at the same time. The received signal at Eve1 becomes
$$y'_{E1}=h_{ae1}x_a+h_ex_a+n_{e1}. \quad (14)$$
The achievable secrecy data rate can be written as
$$R_{1,2}=\log_2(1+\frac{P|h_{ab}|^2}{\sigma_n^2+P_{J2}|h_{e2}|^2}) - \log_2(1+\frac{P|h_{ae1}|^2}{\sigma_n^2+P_{J2}|h_e|^2}). \quad (15)$$
- When $q_1=2$ and $q_2=2$ hold, Eve1 and Eve2 choose to transmit the jamming signals to block the Alice's signals at Bob. The received signal at Bob becomes
$$y_J=h_{ab}x_a+h_{e1}x_a+h_{e2}x_a+n_b. \quad (16)$$
The achievable secrecy data rate can be written as
$$R_{2,2}=\log_2(1+\frac{P|h_{ab}|^2}{\sigma_n^2+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2}). \quad (17)$$

In this paper, we set the noise variance σ_n^2 to unity. And in the subsequent sections, the noise variance σ_n^2 will be replaced by 1 directly.

III. PHYSICAL-LAYER SECURE GAME

In this section, the physical-layer secure game with multiple attackers is studied, and this problem is formulated as a stochastic game. In this game, Alice flexibly selects its transmit power P, Eve1 determines its attack type q_1, and Eve2 determines its attack type q_2.

The cost of Eve1's attack with mode q_1 is given by

$$f(q_1)=\begin{cases}0, & q_1=0,\\ \theta_{e1}, & q_1=1,\\ \theta_{j1}, & q_1=2,\end{cases}$$

where θ_{e1} and θ_{j1} are the cost of Eve1 to perform eavesdropping and jamming, respectively.

Similarly, the cost of Eve2's attack with mode q_2 is written as

$$f(q_2)=\begin{cases}0, & q_2=0,\\ \theta_{e2}, & q_2=1,\\ \theta_{j2}, & q_2=2,\end{cases}$$

where θ_{e2} and θ_{j2} are the cost of Eve2 to perform eavesdropping and jamming, respectively. Let u_a denote the utility of Alice in the static game, given by

$$u_a(P,q_1,q_2)=\ln\,2R_{q_1,q_2}-C_aP, \quad (18)$$

where C_a is the cost of unit transmit power for Alice. For the convenience of calculation, we multiply the secrecy data rate by the coefficient ln 2 in (18).

Let u_{e1} denote the utility of Eve1, which is given by

$$u_{e1}(P, q_1, q_2) = -\ln\ 2R_{q_1,q_2} - f(q_1). \tag{19}$$

Similarly, the utility of Eve2 is given by

$$u_{e2}(P, q_1, q_2) = -\ln\ 2R_{q_1,q_2} - f(q_2). \tag{20}$$

The NE strategy of the stochastic game denoted by (P^*, q_1^*, q_2^*) is given by

$$3u_a(P^*, q_1^*, q_2^*) \geq u_a(P, q_1^*, q_2^*), \quad \forall 0 \leq P \leq P_{max}. \tag{21}$$
$$u_{e1}(P^*, q_1^*, q_2^*) \geq u_{e1}(P^*, q_1, q_2^*), \quad \forall q_1 = 0, 1, 2. \tag{22}$$
$$u_{e2}(P^*, q_1^*, q_2^*) \geq u_{e2}(P^*, q_1^*, q_2), \quad \forall q_2 = 0, 1, 2. \tag{23}$$

From (21)-(23), it is obvious that Alice and all attackers cannot obtain more utility by changing their NE strategies. Thus, no one has the motivation to break the equilibrium of the stochastic game. We can deduce the NE by finding the optimal value, which makes each player achieve the maximum utility under the current environment. An NE $(x^*, 0, 0)$ result is given in the following Lemma 1.

Lemma 1: An NE $(x^*, 0, 0)$ of the physical-layer secure stochastic game is given by

$$\begin{cases} \dfrac{|h_{ab}|^2}{1 + x^*|h_{ab}|^2} = C_a, & \text{(24a)} \\ 0 \leq x^* \leq P_{max}. & \text{(24b)} \end{cases}$$

If

$$\begin{cases} \theta_{E1} \geq \ln(1 + x^*|h_{ae1}|^2), & \text{(25a)} \\ \theta_{J1} \geq \ln(1 + \dfrac{x^* P_{J1}|h_{ab}|^2|h_{e1}|^2}{1 + P_{J1}|h_{e1}|^2 + x^*|h_{ab}|^2}), & \text{(25b)} \\ \theta_{E2} \geq \ln(1 + x^*|h_{ae2}|^2), & \text{(25c)} \\ \theta_{J2} \geq \ln(1 + \dfrac{x^* P_{J2}|h_{ab}|^2|h_{e2}|^2}{1 + P_{J2}|h_{e2}|^2 + x^*|h_{ab}|^2}), & \text{(25d)} \\ \dfrac{|h_{ab}|^2}{1 + P_{max}|h_{ab}|^2} < C_a < |h_{ab}|^2. & \text{(25e)} \end{cases}$$

Proof: If (25a)-(25d) hold, from (19) and (20), we have

$$\begin{aligned} &u_{e1}(x^*, 0, 0) - u_{e1}(x^*, 1, 0) \\ &\quad = \theta_{E1} - \ln(1 + x^*|h_{ae1}|^2) \geq 0, \\ &u_{e1}(x^*, 0, 0) - u_{e1}(x^*, 2, 0) \\ &\quad = \theta_{J1} - \ln(1 + \frac{x^* P_{J1}|h_{ab}|^2|h_{e1}|^2}{1 + P_{J1}|h_{e1}|^2 + x^*|h_{ab}|^2}) \geq 0, \\ &u_{e2}(x^*, 0, 0) - u_{e2}(x^*, 0, 1) \\ &\quad = \theta_{E2} - \ln(1 + x^*|h_{ae2}|^2) \geq 0, \\ &u_{e2}(x^*, 0, 0) - u_{e2}(x^*, 0, 2) \\ &\quad = \theta_{J2} - \ln(1 + \frac{x^* P_{J2}|h_{ab}|^2|h_{e2}|^2}{1 + P_{J2}|h_{e2}|^2 + x^*|h_{ab}|^2}) \geq 0. \end{aligned}$$

Thus, (22) and (23) hold for $(x^*, 0, 0)$. From (18), we have

$$\frac{\partial u_a(P, 0, 0)}{\partial P} = \frac{|h_{ab}|^2}{1 + P|h_{ab}|^2} - C_a, \tag{26}$$
$$\frac{\partial u_a^2(P, 0, 0)}{\partial P^2} = -\frac{|h_{ab}|^4}{(1 + P|h_{ab}|^2)^2} \leq 0, \tag{27}$$

which indicates that $\partial u_a(P, 0, 0)/\partial P$ monotonically decreases with respect to P. If (25e) holds, from (26), we have

$$\frac{\partial u_a(P, 0, 0)}{\partial P}|_{P=0} = |h_{ab}|^2 - C_a > 0, \tag{28}$$
$$\frac{\partial u_a(P, 0, 0)}{\partial P}|_{P=P_{max}} = \frac{|h_{ab}|^2}{1 + P_{max}|h_{ab}|^2} - C_a < 0. \tag{29}$$

From (27)-(29), it is obvious that there exists a unique solution x^* which satisfies $\partial u_a(P, 0, 0)/\partial P = 0$, $0 \leq x^* \leq P_{max}$. Thus, (24a) is the unique solution of $\partial u_a(P, 0, 0)/\partial P = 0$. From (28)-(29), we can see that $u_a(P, 0, 0)$ increases with P if $P \leq x^*$ while decreases otherwise. Thus, $u_a(P, 0, 0)$ achieves the maximum value at $P = x^*$, and (21) also holds for $(x^*, 0, 0)$. More details about the NE strategy can be found in [39]–[41]. ■

As shown in Lemma 1, if (25a)-(25d) hold (i.e., the costs of eavesdropping and jamming attacks are higher than the transmission costs of Alice), the attack motivation of attackers is suppressed. Otherwise, Alice stops the transmission when (25e) holds. In other words, Alice will stop the transmission under the circumstances that radio channel degradation is serious and the security cannot be guaranteed.

In the following Lemma 2, we give an NE $(P_{max}, 0, 0)$ result. The other NEs results are provided in Appendix.

Lemma 2: The stochastic game has an NE $(P_{max}, 0, 0)$, if

$$\begin{cases} \theta_{E1} \geq \ln(1 + P_{max}|h_{ae1}|^2), & \text{(30a)} \\ \theta_{J1} \geq \ln(1 + \dfrac{P_{max} P_{J1}|h_{ab}|^2|h_{e1}|^2}{1 + P_{J1}|h_{e1}|^2 + P_{max}|h_{ab}|^2}), & \text{(30b)} \\ \theta_{E2} \geq \ln(1 + P_{max}|h_{ae2}|^2), & \text{(30c)} \\ \theta_{J2} \geq \ln(1 + \dfrac{P_{max} P_{J2}|h_{ab}|^2|h_{e2}|^2}{1 + P_{J2}|h_{e2}|^2 + P_{max}|h_{ab}|^2}), & \text{(30d)} \\ \dfrac{|h_{ab}|^2}{1 + P_{max}|h_{ab}|^2} \geq C_a. & \text{(30e)} \end{cases}$$

Proof: Similar to the proof of Lemma 1, if (30a)-(30d) hold, by (19) and (20), we have

$$\begin{aligned} &u_{e1}(P_{max}, 0, 0) - u_{e1}(P_{max}, 1, 0) \\ &\quad = \theta_{E1} - \ln(1 + P_{max}|h_{ae1}|^2) \geq 0, \\ &u_{e1}(P_{max}, 0, 0) - u_{e1}(P_{max}, 2, 0) \\ &\quad = \theta_{J1} - \ln(1 + \frac{P_{max} P_{J1}|h_{ab}|^2|h_{e1}|^2}{1 + P_{J1}|h_{e1}|^2 + P_{max}|h_{ab}|^2}) \geq 0, \\ &u_{e2}(P_{max}, 0, 0) - u_{e2}(P_{max}, 0, 1) \\ &\quad = \theta_{E2} - \ln(1 + P_{max}|h_{ae2}|^2) \geq 0, \\ &u_{e2}(P_{max}, 0, 0) - u_{e2}(P_{max}, 0, 2) \\ &\quad = \theta_{J2} - \ln(1 + \frac{P_{max} P_{J2}|h_{ab}|^2|h_{e2}|^2}{1 + P_{J2}|h_{e2}|^2 + P_{max}|h_{ab}|^2}) \geq 0. \end{aligned}$$

Thus, (22) and (23) hold for $(P_{max}, 0, 0)$. From (27) and (30e), we can easily conclude that $\partial u_a(P, 0, 0)/\partial P$ monotonically decreases with P, and

$$\frac{\partial u_a(P, 0, 0)}{\partial P} \geq \frac{\partial u_a(P, 0, 0)}{\partial P}|_{P=P_{max}} \geq 0, \quad \forall 0 \leq P \leq P_{max}, \tag{31}$$

which indicates that (21) holds for $(P_{max}, 0, 0)$. Hence, $(P_{max}, 0, 0)$ is an NE of the game. ■

As shown in Lemma 2, high attack costs in (30a)-(30d), or low transmission costs in (30e) will make Alice select the maximum transmit power to transmit the signals.

IV. POWER ALLOCATION POLICY UNDER MULTIPLE ATTACKERS

In the dynamic physical-layer secure game, we employ the Q-learning based power allocation strategy for Alice to choose the transmit power flexibly. Moreover, through Q-learning, Eve1 and Eve2 can learn their attack modes to gain a large utility. At the beginning of the game, Alice randomly selects a value for the transmit power as its action, and by observing the Alice's action, Eve1 selects an attack mode as its action. Then, by observing the actions of Alice and Eve1, Eve2 selects an attack mode as its action.

Algorithm 1: Q-Learning Based Power Allocation Algorithm

```
1: Initialize all parameters
2: for each time slot $n$ do
3:     Update the system state $\boldsymbol{s_n} = [q_1^{n-1}, q_2^{n-1}]$
4:     Choose the transmit power $P_n$ using the $\varepsilon$-greedy policy
5:     Observe the attack types of two attackers $\boldsymbol{s_n}$ and the utility of Alice $u_a$
6:     Update the $Q$ function:
       $Q(\boldsymbol{s_n}, P_n) = (1-\alpha)Q(\boldsymbol{s_n}, P_n) + \alpha(u_a(\boldsymbol{s_n}, P_n) + \gamma V(\boldsymbol{s_{n\mathbf{C}1}}))$
7:     Find the optimal value function:
       $V(\boldsymbol{s_n}) = \max_{0 \le P \le P_{max}} Q(\boldsymbol{s_n}, P)$
8: end for
```

As shown in Algorithm 1, the Q-learning based power allocation algorithm for the considered physical-layer secure game is presented. In line 3 of Algorithm 1, the system state at time n is formed by attack modes of all attackers at the previous time, i.e., $\boldsymbol{s_n} = [q_1^{n-1}, q_2^{n-1}]$. Specifically, q_1^n and q_2^n denote the attack modes of the Eve1 and Eve2 at time n, respectively. In line 4, Alice uses the ε-greedy policy to choose the transmit power P_n among $L+1$ levels at time n, namely $P_n \in \{lP_{max}/L\}_{0 \le l \le L}$. Moreover, Alice randomly selects an action with probability ε, while chooses the best action with probability $1-\varepsilon$. In line 6, $Q(\boldsymbol{s_n}, P_n)$ denotes the Q function of Alice, where $\alpha \in [0, 1]$ denotes the learning rate and $\gamma \in [0, 1]$ represents the discount factor. In line 7, $V(\boldsymbol{s_n})$ denotes the optimal value function, which indicates the maximum value of $Q(\boldsymbol{s_n}, P)$.

V. SIMULATION RESULTS

In this section, we provide some simulation results to show the impact of multiple smart attackers on the system secrecy performance. As a commonly-used setting, the average channel gain of the main link, σ_{ab}^2, is set to 1.2, the average channel gains of the eavesdropping link, σ_{ae1}^2 and σ_{ae2}^2, are set to 0.5 and 1, the average channel gains of the jamming link, σ_{e1}^2 and σ_{e2}^2, are set to 2 and 1.5, and the average channel gain from the Eve1 to Eve2 link, σ_e^2, is set to 1.6. Moreover, the jamming power at the Eve1, P_{J1}, is set to 3.2, and the jamming power at the Eve2, P_{J2}, is set to 2.9. In further, the costs of Eve1 and Eve2 to perform the eavesdropping mode, θ_{e1} and θ_{e2}, are set to 2.5 and 2.8, respectively. Similarly, the costs of Eve1 and Eve2 to perform the jamming mode, θ_{j1} and θ_{j2}, are set to 3.2 and 2.9, respectively. Furthermore, the cost of unit transmit power for Alice, C_a, is set to 0.1.

Fig. 2 shows the average utility of Alice and the two attackers with Q-learning, where the time slot ranges from 0 to 8000. We perform 10^4 times of the experiments, and use the average value to show in Fig. 2. As observed from Fig. 2, we can find that all the utility curves increase sharply from 0 to 3000 time slots. After 3000 time slots, the lines trend to be steady, which means that all players have obtained high utilities, and the secure game has reached a steady state. Moreover, the average utility of Eve1 is larger than that of Eve2. This is because that Eve1 is closer to Alice, and it is easier to overhear the Alice's signals.

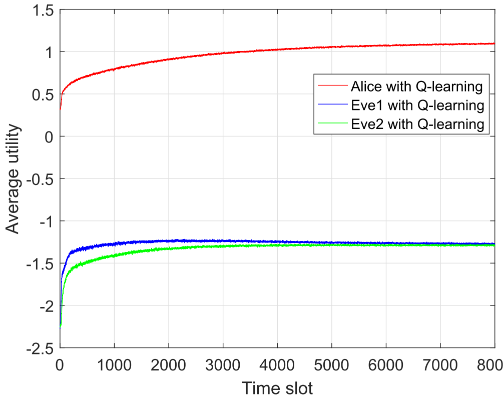


FIGURE 2. Average utility of Alice and the two attackers with Q-learning.

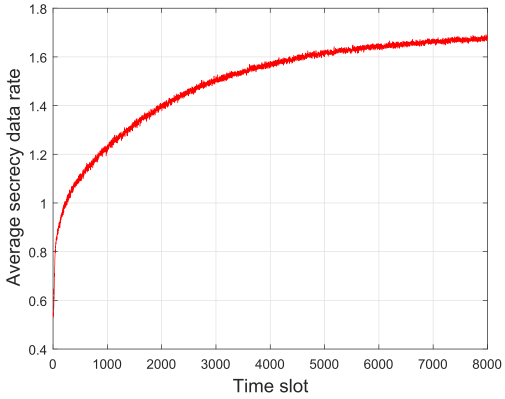


FIGURE 3. Average secrecy data rate of the considered system.

Fig. 3 illustrates the average secrecy data rate of the considered system over 8000 time slots, where we perform 10^4 times

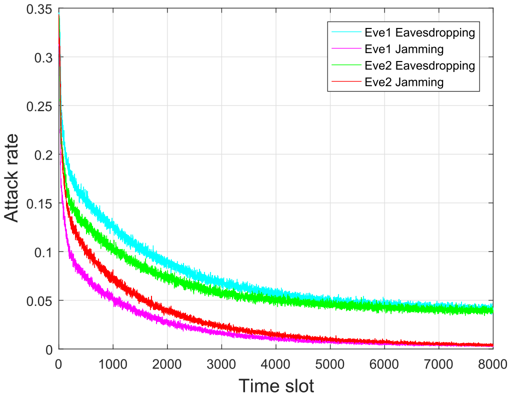


FIGURE 4. Attack rates of Eve1 and Eve2 over time slot.

of simulation to obtain the average value. From Fig. 3, we can find that the system secrecy data rate becomes larger when the time slot increases. Moreover, the secrecy data rate can reach to about 1.7 bps/Hz, when the time slot is large sufficiently. This indicates that the Q-learning based power allocation strategy can guarantee the security of the considered system, even if there exist multiple smart attackers.

Fig. 4 demonstrates the average attack rates of Eve1 and Eve2 over 8000 time slots, where we perform 10^4 times of simulation to obtain the average rate. As shown in Fig. 4, we can find that the probability that Eve1 chooses to perform eavesdropping or jamming becomes smaller with a larger time slot. In particular, the probability becomes convergent to a steady value when the time slot is larger than 5000. Moreover, all the attack rates of Eve1 and Eve2 at the time slot of 8000 are quite low, indicating that the proposed policy works well. In further, the eavesdropping rate of Eve1 is higher than that of Eve2, since Eve1 is closer to Alice compared with Eve2. Furthermore, the jamming rate of Eve1 is lower than that of Eve2, as Eve2 is closer to Bob compared with Eve1.

VI. CONCLUSIONS

In this paper, we investigated the stochastic physical-layer secure game with multiple smart attackers, and the NEs of the game and their existence conditions were derived. Moreover, we employed the Q-learning based power allocation algorithm to help Alice choose its transmit power flexibly, and the attackers could learn how to choose their attack modes better through Q-learning. Simulation results showed that the proposed scheme can efficiently improve the secrecy data rate and suppress the smart attackers' attack rate, even if there exist multiple smart attackers. The advantage of this algorithm is that we can find the optimal value through Q-table. Correspondingly, the disadvantages are that the learning time is long and the convergence speed is slow. Especially, when there are many state-action pairs in the game, the Q-table will become large, and it is complex to implement this algorithm. In the future works, we will improve the algorithm by reducing the implementation complexity, and consider the location change of attackers [42]–[44]. Moreover, other intelligent algorithms [45]–[47] will be considered for further enhancing the transmission security for this system.

APPENDIX

Lemma 3: An NE $(x^*, 1, 0)$ of the stochastic game is given by

$$\begin{cases} \dfrac{|h_{ab}|^2}{1+x^*|h_{ab}|^2} - \dfrac{|h_{ae1}|^2}{1+x^*|h_{ae1}|^2} = C_a, & \text{(32a)} \\ 0 \le x^* \le P_{max}. & \text{(32b)} \end{cases}$$

If

$$\begin{cases} \theta_{E1} \le \ln(1+x^*|h_{ae1}|^2), & \text{(33a)} \\ \theta_{E1} - \theta_{J1} \le \ln(\dfrac{1+x^*|h_{ae1}|^2}{1+x^*|h_{ab}|^2}) + \ln(1+\dfrac{x^*|h_{ab}|^2)}{1+P_{J1}|h_{e1}|^2}), & \text{(33b)} \\ \theta_{E2} \ge \max[1, \ln(1+\dfrac{x^*|h_{ae2}|^2 - x^*|h_{ae1}|^2}{1+x^*|h_{ae1}|^2})], & \text{(33c)} \\ \theta_{J2} \ge \ln(1+\dfrac{x^*|h_{ae1}|^2}{1+P_{J2}|h_e|^2}) - \ln(1+\dfrac{x^*|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2}) - \ln(\dfrac{1+x^*|h_{ae1}|^2}{1+x^*|h_{ab}|^2}), & \text{(33d)} \\ |h_{ae1}|^2 \le |h_{ab}|^2, \\ \dfrac{|h_{ab}|^2}{1+P_{max}|h_{ab}|^2} - \dfrac{|h_{ae1}|^2}{1+P_{max}|h_{ae1}|^2} < C_a < |h_{ab}|^2 - |h_{ae1}|^2. & \text{(33e)} \end{cases}$$

Proof: The proof can be referred to that of Lemma 1. ■

Lemma 4: The stochastic game has an NE $(P_{max}, 1, 0)$, if

$$\begin{cases} \theta_{E1} \le \ln(1+P_{max}|h_{ae1}|^2), & \text{(34a)} \\ \theta_{E1} - \theta_{J1} \le \ln(\dfrac{1+P_{max}|h_{ae1}|^2}{1+P_{max}|h_{ab}|^2}) + \ln(1+\dfrac{P_{max}|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2}), & \text{(34b)} \\ \theta_{E2} \ge \max[1, \ln(\dfrac{1+P_{max}|h_{ae2}|^2}{1+P_{max}|h_{ae1}|^2})], & \text{(34c)} \\ \theta_{J2} \ge \ln(1+\dfrac{P_{max}|h_{ae1}|^2}{1+P_{J2}|h_e|^2}) - \ln(1+\dfrac{P_{max}|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2}) - \ln(\dfrac{1+P_{max}|h_{ae1}|^2}{1+P_{max}|h_{ab}|^2}), & \text{(34d)} \\ |h_{ae1}|^2 \le |h_{ab}|^2, \\ \dfrac{|h_{ab}|^2}{1+P_{max}|h_{ab}|^2} - \dfrac{|h_{ae1}|^2}{1+P_{max}|h_{ae1}|^2} \ge C_a. & \text{(34e)} \end{cases}$$

Proof: Similar to the proof of Lemma 2. ■

Lemma 5: An NE $(x^*, 2, 0)$ of the stochastic game is given by

$$\begin{cases} \dfrac{|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2 x^*|h_{ab}|^2} = C_a, & (35a) \\ 0 \le x^* \le P_{max}. & (35b) \end{cases}$$

If

$$\begin{cases} \theta_{J1} \le \ln(1+\dfrac{x^* P_{J1}|h_{ab}|^2|h_{e1}|^2}{1+P_{J1}|h_{e1}|^2+x^*|h_{ab}|^2}), & (36a) \\ \theta_{E1}-\theta_{J1} \ge \ln(\dfrac{1+x^*|h_{ae1}|^2}{1+x^*|h_{ab}|^2}) + \ln(1+\dfrac{x^*|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2}), & (36b) \\ \theta_{E2} \ge \ln(1+\dfrac{x^*|h_{ae2}|^2}{1+P_{J1}|h_{e1}|^2}), & (36c) \\ \theta_{J2} \ge \ln(1+\dfrac{x^*|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2}) - \ln(1+\dfrac{x^*|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2}), & (36d) \\ \dfrac{|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{max}|h_{ab}|^2} < C_a < \dfrac{|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2}. & (36e) \end{cases}$$

Proof: The proof can be referred to that of Lemma 1. ■

Lemma 6: The stochastic game has an NE $(P_{max}, 2, 0)$, if

$$\begin{cases} \theta_{J1} \le \ln(1+\dfrac{P_{max} P_{J1}|h_{ab}|^2|h_{e1}|^2}{1+P_{J1}|h_{e1}|^2+P_{max}|h_{ab}|^2}), & (37a) \\ \theta_{E1}-\theta_{J1} \ge \ln(\dfrac{1+P_{max}|h_{ae1}|^2}{1+P_{max}|h_{ab}|^2}) + \ln(1+\dfrac{P_{max}|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2}), & (37b) \\ \theta_{E2} \ge \ln(1+\dfrac{P_{max}|h_{ae2}|^2}{1+P_{J1}|h_{e1}|^2}), & (37c) \\ \theta_{J2} \ge \ln(1+\dfrac{P_{max}|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2}) - \ln(1+\dfrac{P_{max}|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2}), & (37d) \\ \dfrac{|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{max}|h_{ab}|^2} \ge C_a. & (37e) \end{cases}$$

Proof: Similar to the proof of Lemma 2. ■

Lemma 7: An NE $(x^*, 0, 1)$ of the stochastic game is given by

$$\begin{cases} \dfrac{|h_{ab}|^2}{1+x^*|h_{ab}|^2} - \dfrac{|h_{ae2}|^2}{1+x^*|h_{ae2}|^2} = C_a, & (38a) \\ 0 \le x^* \le P_{max}. & (38b) \end{cases}$$

If

$$\begin{cases} \theta_{E1} \ge \max[\ln(\dfrac{1+x^*|h_{ae1}|^2}{1+x^*|h_{ae2}|^2}), 1], & (39a) \\ \theta_{J1} \ge \ln(1+\dfrac{x^*|h_{ae2}|^2}{P_{J1}|h_e|^2}) - \ln(1+\dfrac{x^*|h_{ab}|^2}{P_{J1}|h_{e1}|^2}) \\ \qquad - \ln(\dfrac{1+x^*|h_{ae2}|^2)}{1+x^*|h_{ab}|^2}), & (39b) \\ \theta_{E2} \le \ln(1+x^*|h_{ae2}|^2), & (39c) \\ \theta_{E2}-\theta_{J2} \le \ln(\dfrac{1+x^*|h_{ae2}|^2}{1+x^*|h_{ab}|^2}) + \ln(1+\dfrac{x^*|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2}), & (39d) \\ |h_{ae2}|^2 \le |h_{ab}|^2, \\ \dfrac{|h_{ab}|^2}{1+P_{max}|h_{ab}|^2} - \dfrac{|h_{ae2}|^2}{1+P_{max}|h_{ae2}|^2} < C_a < |h_{ab}|^2 - |h_{ae2}|^2. & (39e) \end{cases}$$

Proof: The proof can be referred to that of Lemma 1. ■

Lemma 8: The stochastic game has an NE $(P_{max}, 0, 1)$, if

$$\begin{cases} \theta_{E1} \ge \max[\ln(\dfrac{1+P_{max}|h_{ae1}|^2}{1+P_{max}|h_{ae2}|^2}), 1], & (40a) \\ \theta_{J1} \ge \ln(1+\dfrac{P_{max}|h_{ae2}|^2}{P_{J1}|h_e|^2}) - \ln(1+\dfrac{P_{max}|h_{ab}|^2}{P_{J1}|h_{e1}|^2}) \\ - \ln(\dfrac{1+P_{max}|h_{ae2}|^2}{1+P_{max}|h_{ab}|^2}), & (40b) \\ \theta_{E2} \le \ln(1+P_{max}|h_{ae2}|^2), & (40c) \\ \theta_{E2}-\theta_{J2} \le \ln(\dfrac{1+P_{max}|h_{ae2}|^2}{1+P_{max}|h_{ab}|^2}) + \ln(1+\dfrac{P_{max}|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2}), & (40d) \\ |h_{ae2}|^2 \le |h_{ab}|^2, \\ \dfrac{|h_{ab}|^2}{1+P_{max}|h_{ab}|^2} - \dfrac{|h_{ae2}|^2}{1+P_{max}|h_{ae2}|^2} \ge C_a. & (40e) \end{cases}$$

Proof: Similar to the proof of Lemma 2. ■

Lemma 9: An NE $(x^*, 1, 1)$ of the stochastic game is given by

$$\begin{cases} \dfrac{|h_{ab}|^2}{1+x^*|h_{ab}|^2} - \max[\dfrac{|h_{ae1}|^2}{1+x^*|h_{ae1}|^2}, \dfrac{|h_{ae2}|^2}{1+x^*|h_{ae2}|^2}] = C_a, & (41a) \\ 0 \le x^* \le P_{max}. & (41b) \end{cases}$$

If

$$\begin{cases} \theta_{E1} \leq \max[\ln(\frac{1+x^*|h_{ae1}|^2}{1+x^*|h_{ae2}|^2}), 0], & (42a)\\ \theta_{E1}-\theta_{J1} \leq \max[\ln(\frac{1+x^*|h_{ae1}|^2}{1+x^*|h_{ab}|^2}), \ln(\frac{1+x^*|h_{ae2}|^2}{1+x^*|h_{ab}|^2}) \\ \quad -\ln(1+\frac{x^*|h_{ae2}|^2}{1+P_{J1}|h_e|^2})+\ln(1+\frac{x^*|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2})], & (42b)\\ \theta_{E2} \leq \max[0, \ln(\frac{1+x^*|h_{ae2}|^2}{1+x^*|h_{ae1}|^2})], & (42c)\\ \theta_{E2}-\theta_{J2} \leq \max[\ln(\frac{1+x^*|h_{ae1}|^2}{1+x^*|h_{ab}|^2}), \ln(\frac{1+x^*|h_{ae2}|^2}{1+x^*|h_{ab}|^2}) \\ \quad -\ln(1+\frac{x^*|h_{ae1}|^2}{1+P_{J2}|h_e|^2})+\ln(1+\frac{x^*|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2})], & (42d)\\ |h_{ae1}|^2 \leq |h_{ab}|^2, \quad |h_{ae2}|^2 \leq |h_{ab}|^2, \\ \frac{|h_{ab}|^2}{1+P_{max}|h_{ab}|^2}-\max[\frac{|h_{ae1}|^2}{1+P_{max}|h_{ae1}|^2}, \frac{|h_{ae2}|^2}{1+P_{max}|h_{ae2}|^2}] \\ < C_a < |h_{ab}|^2-\max[|h_{ae1}|^2, |h_{ae2}|^2]. & (42e) \end{cases}$$

Proof: The proof can be referred to that of Lemma 1. ■

Lemma 10: The stochastic game has an NE $(P_{max}, 1, 1)$, if

$$\begin{cases} \theta_{E1} \leq \max[\ln(\frac{1+P_{max}|h_{ae1}|^2}{1+P_{max}|h_{ae2}|^2}), 0], & (43a)\\ \theta_{E1}-\theta_{J1} \leq \max[\ln(\frac{1+P_{max}|h_{ae1}|^2}{1+P_{max}|h_{ab}|^2}), \ln(\frac{1+P_{max}|h_{ae2}|^2}{1+P_{max}|h_{ab}|^2}) \\ \quad -\ln(1+\frac{P_{max}|h_{ae2}|^2}{1+P_{J1}|h_e|^2})+\ln(1+\frac{P_{max}|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2})], & (43b)\\ \theta_{E2} \leq \max[0, \ln(\frac{1+P_{max}|h_{ae2}|^2}{1+P_{max}|h_{ae1}|^2})], & (43c)\\ \theta_{E2}-\theta_{J2} \leq \max[\ln(\frac{1+P_{max}|h_{ae1}|^2}{1+P_{max}|h_{ab}|^2}), \ln(\frac{1+P_{max}|h_{ae2}|^2}{1+P_{max}|h_{ab}|^2}) \\ \quad -\ln(1+\frac{P_{max}|h_{ae1}|^2}{1+P_{J2}|h_e|^2})+\ln(1+\frac{P_{max}|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2})], & (43d)\\ |h_{ae1}|^2 \leq |h_{ab}|^2, \\ |h_{ae2}|^2 \leq |h_{ab}|^2, \\ \frac{|h_{ab}|^2}{1+P_{max}|h_{ab}|^2}-\max[\frac{|h_{ae1}|^2}{1+P_{max}|h_{ae1}|^2}, \\ \frac{|h_{ae2}|^2}{1+P_{max}|h_{ae2}|^2}] \geq C_a. & (43e) \end{cases}$$

Proof: Similar to the proof of Lemma 2. ■

Lemma 11: An NE $(x^*, 2, 1)$ of the stochastic game is given by

$$\begin{cases} \frac{|h_{ab}|^2}{1+P_{J1}x^*|h_{e1}|^2+x^*|h_{ab}|^2}- \\ \frac{|h_{ae2}|^2}{1+P_{J1}x^*|h_e|^2+x^*|h_{ae2}|^2}=C_a, & (44a)\\ 0 \leq x^* \leq P_{max}. & (44b) \end{cases}$$

If

$$\begin{cases} \theta_{J1} \leq \ln(\frac{1+x^*|h_{ab}|^2}{1+x^*|h_{ae2}|^2})+\ln(1+\frac{x^*|h_{ae2}|^2}{1+P_{J1}|h_e|^2}) \\ \quad -\ln(1+\frac{x^*|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2}), & (45a)\\ \theta_{J1}-\theta_{E1} \leq \ln(1+\frac{x^*|h_{ae2}|^2}{1+P_{J1}|h_e|^2})-\ln(1+\frac{x^*|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2}) \\ \quad -\max[\ln(\frac{1+x^*|h_{ae1}|^2}{1+x^*|h_{ab}|^2}), \ln(\frac{1+x^*|h_{ae2}|^2}{1+x^*|h_{ab}|^2})], & (45b)\\ \theta_{E2} \leq \ln(1+\frac{x^*|h_{ae2}|^2}{1+P_{J1}|h_e|^2}), & (45c)\\ \theta_{E2}-\theta_{J2} \leq \ln(1+\frac{x^*|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2}) \\ \quad +\ln(1+\frac{x^*|h_{ae2}|^2}{1+P_{J1}|h_e|^2})-\ln(1+\frac{x^*|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2}), & (45d)\\ \frac{|h_{ae2}|^2}{1+P_{J1}|h_e|^2} \leq \frac{|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2}, \\ \frac{|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{max}|h_{ab}|^2}-\frac{|h_{ae2}|^2}{1+P_{J1}|h_e|^2+P_{max}|h_{ae2}|^2} \\ < C_a < \frac{|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2}-\frac{|h_{ae2}|^2}{1+P_{J1}|h_e|^2}. & (45e) \end{cases}$$

Proof: The proof can be referred to that of Lemma 1. ■

Lemma 12: The stochastic game has an NE $(P_{max}, 2, 1)$, if

$$\begin{cases} \theta_{J1} \leq \ln(\frac{1+P_{max}|h_{ab}|^2}{1+P_{max}|h_{ae2}|^2})+\ln(1+\frac{P_{max}|h_{ae2}|^2}{1+P_{J1}|h_e|^2}) \\ \quad -\ln(1+\frac{P_{max}|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2}), & (46a)\\ \theta_{J1}-\theta_{E1} \leq \ln(1+\frac{P_{max}|h_{ae2}|^2}{1+P_{J1}|h_e|^2})-\ln(1+\frac{P_{max}|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2}) \\ \quad -\max[\ln(\frac{1+P_{max}|h_{ae1}|^2}{1+P_{max}|h_{ab}|^2}), \ln(\frac{1+P_{max}|h_{ae2}|^2}{1+P_{max}|h_{ab}|^2})], & (46b)\\ \theta_{E2} \leq \ln(1+\frac{P_{max}|h_{ae2}|^2}{1+P_{J1}|h_e|^2}), & (46c)\\ \theta_{E2}-\theta_{J2} \leq \ln(1+\frac{P_{max}|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2}) \\ \quad +\ln(1+\frac{P_{max}|h_{ae2}|^2}{1+P_{J1}|h_e|^2})-\ln(1+\frac{P_{max}|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2}), & (46d)\\ \frac{|h_{ae2}|^2}{1+P_{J1}|h_e|^2} \leq \frac{|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2}, \\ \frac{|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{max}|h_{ab}|^2} \\ \quad -\frac{|h_{ae2}|^2}{1+P_{J1}|h_e|^2+P_{max}|h_{ae2}|^2} \geq C_a. & (46e) \end{cases}$$

Proof: Similar to the proof of Lemma 2. ■

Lemma 13: An NE (x^*, 0, 2) of the stochastic game is given by

$$\left\{\begin{aligned} &\frac{|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2x^*|h_{ab}|^2}=C_a, && (47a)\\ &0\leq x^*\leq P_{max}. && (47b)\end{aligned}\right.$$

If

$$\left\{\begin{aligned} &\theta_{E1}\geq \ln(1+\frac{x^*|h_{ae1}|^2}{1+P_{J2}|h_e|^2}), && (48a)\\ &\theta_{J1}\geq \ln(1+\frac{x^*|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2})-\ln(1+\frac{x^*|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2}), && (48b)\\ &\theta_{J2}\leq \ln(1+\frac{x^*P_{J2}|h_{ab}|^2|h_{e2}|^2}{1+P_{J2}|h_{e2}|^2+x^*|h_{ab}|^2}), && (48c)\\ &\theta_{J2}-\theta_{E2}\leq \ln(\frac{1+x^*|h_{ab}|^2}{1+x^*|h_{ae2}|^2})-\ln(1+\frac{x^*|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2}), && (48d)\\ &\frac{|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2+P_{max}|h_{ab}|^2}<C_a<\frac{|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2}. && (48e)\end{aligned}\right.$$

Proof: The proof can be referred to that of Lemma 1. ■

Lemma 14: The stochastic game has an NE (P_{max}, 0, 2), if

$$\left\{\begin{aligned} &\theta_{E1}\geq \ln(1+\frac{P_{max}|h_{ae1}|^2}{1+P_{J2}|h_e|^2}), && (49a)\\ &\theta_{J1}\geq \ln(1+\frac{P_{max}|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2})-\ln(1+\frac{P_{max}|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2}), && (49b)\\ &\theta_{J2}\leq \ln(1+\frac{P_{max}P_{J2}|h_{ab}|^2|h_{e2}|^2}{1+P_{J2}|h_{e2}|^2+P_{max}|h_{ab}|^2}), && (49c)\\ &\theta_{J2}-\theta_{E2}\leq \ln(\frac{1+P_{max}|h_{ab}|^2}{1+P_{max}|h_{ae2}|^2})-\ln(1+\frac{P_{max}|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2}), && (49d)\\ &\frac{|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2+P_{max}|h_{ab}|^2}\geq C_a. && (49e)\end{aligned}\right.$$

Proof: Similar to the proof of Lemma 2. ■

Lemma 15: An NE (x^*, 1, 2) of the stochastic game is given by

$$\left\{\begin{aligned} &\frac{|h_{ab}|^2}{1+P_{J2}x^*|h_{e2}|^2+x^*|h_{ab}|^2}-\frac{|h_{ae1}|^2}{1+P_{J2}x^*|h_e|^2+x^*|h_{ae1}|^2}=C_a, && (50a)\\ &0\leq x^*\leq P_{max}. && (50b)\end{aligned}\right.$$

If

$$\left\{\begin{aligned} &\theta_{E1}\leq \ln(1+\frac{x^*|h_{ae1}|^2}{1+P_{J2}|h_e|^2}), && (51a)\\ &\theta_{E1}-\theta_{J1}\leq \ln(1+\frac{x^*|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2})\\ &\quad+\ln(1+\frac{x^*|h_{ae1}|^2}{1+P_{J2}|h_e|^2})-\ln(1+\frac{x^*|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2}), && (51b)\\ &\theta_{J2}\leq \ln(\frac{1+x^*|h_{ab}|^2}{1+x^*|h_{ae1}|^2})+\ln(1+\frac{x^*|h_{ae1}|^2}{1+P_{J2}|h_e|^2})\\ &\quad-\ln(1+\frac{x^*|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2}), && (51c)\\ &\theta_{J2}-\theta_{E2}\leq \ln(1+\frac{x^*|h_{ae1}|^2}{1+P_{J2}|h_e|^2})-\ln(1+\frac{x^*|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2})\\ &\quad-\max[\ln(\frac{1+x^*|h_{ae1}|^2}{1+x^*|h_{ab}|^2}),\ln(\frac{1+x^*|h_{ae2}|^2}{1+x^*|h_{ab}|^2})], && (51d)\\ &\frac{|h_{ae1}|^2}{1+P_{J2}|h_e|^2}\leq\frac{|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2},\\ &\frac{|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2+P_{max}|h_{ab}|^2}-\frac{|h_{ae1}|^2}{1+P_{J2}|h_e|^2+P_{max}|h_{ae1}|^2}\\ &\quad<C_a<\frac{|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2}-\frac{|h_{ae1}|^2}{1+P_{J2}|h_e|^2}. && (51e)\end{aligned}\right.$$

Proof: The proof can be referred to that of Lemma 1. ■

Lemma 16: The stochastic game has an NE (P_{max}, 1, 2), if

$$\left\{\begin{aligned} &\theta_{E1}\leq \ln(1+\frac{P_{max}|h_{ae1}|^2}{1+P_{J2}|h_e|^2}), && (52a)\\ &\theta_{E1}-\theta_{J1}\leq \ln(1+\frac{P_{max}|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2})\\ &\quad+\ln(1+\frac{P_{max}|h_{ae1}|^2}{1+P_{J2}|h_e|^2})-\ln(1+\frac{P_{max}|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2}), && (52b)\\ &\theta_{J2}\leq \ln(\frac{1+P_{max}|h_{ab}|^2}{1+P_{max}|h_{ae1}|^2})+\ln(1+\frac{P_{max}|h_{ae1}|^2}{1+P_{J2}|h_e|^2})\\ &\quad-\ln(1+\frac{P_{max}|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2}), && (52c)\\ &\theta_{J2}-\theta_{E2}\leq \ln(1+\frac{P_{max}|h_{ae1}|^2}{1+P_{J2}|h_e|^2})-\ln(1+\frac{P_{max}|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2})\\ &\quad-\max[\ln(\frac{1+P_{max}|h_{ae1}|^2}{1+P_{max}|h_{ab}|^2}),\ln(\frac{1+P_{max}|h_{ae2}|^2}{1+P_{max}|h_{ab}|^2})], && (52d)\\ &\frac{|h_{ae1}|^2}{1+P_{J2}|h_e|^2}\leq\frac{|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2},\\ &\frac{|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2+P_{max}|h_{ab}|^2}\\ &\quad-\frac{|h_{ae1}|^2}{1+P_{J2}|h_e|^2+P_{max}|h_{ae1}|^2}\geq C_a. && (52e)\end{aligned}\right.$$

Proof: Similar to the proof of Lemma 2. ■

Lemma 17: An NE (x^*, 2, 2) of the stochastic game is given by

$$\begin{cases} \dfrac{|h_{ab}|^2}{1+P_{J1}x^*|h_{e1}|^2+P_{J2}x^*|h_{e2}|^2+x^*|h_{ab}|^2}=C_a, & \text{(53a)}\\ 0\leq x^*\leq P_{max}. & \text{(53b)}\end{cases}$$

If

$$\begin{cases} \theta_{J1}\leq \ln(1+\dfrac{x^*|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2})-\ln(1+\dfrac{x^*|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2}), & \text{(54a)}\\ \theta_{J1}-\theta_{E1}\leq \ln(1+\dfrac{x^*|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2})-\ln(1+\dfrac{x^*|h_{ae1}|^2}{1+P_{J2}|h_{e}|^2}) \\ \quad -\ln(1+\dfrac{x^*|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2}), & \text{(54b)}\\ \theta_{J2}\leq \ln(1+\dfrac{x^*|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2}) \\ \quad -\ln(1+\dfrac{x^*|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2}), & \text{(54c)}\\ \theta_{J2}-\theta_{E2}\leq \ln(1+\dfrac{x^*|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2})-\ln(1+\dfrac{x^*|h_{ae2}|^2}{1+P_{J1}|h_{e}|^2}) \\ \quad -\ln(1+\dfrac{x^*|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2}), & \text{(54d)}\\ \dfrac{|h_{ae1}|^2}{1+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2+P_{max}|h_{ab}|^2} \\ \quad < C_a < \dfrac{|h_{ae1}|^2}{1+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2}. & \text{(54e)}\end{cases}$$

Proof: The proof can be referred to that of Lemma 1. ■

Lemma 18: The stochastic game has an NE (P_{max}, 2, 2), if

$$\begin{cases} \theta_{J1}\leq \ln(1+\dfrac{P_{max}|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2})-\ln(1+\dfrac{P_{max}|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2}), & \text{(55a)}\\ \theta_{J1}-\theta_{E1}\leq \ln(1+\dfrac{P_{max}|h_{ab}|^2}{1+P_{J2}|h_{e2}|^2})-\ln(1+\dfrac{P_{max}|h_{ae1}|^2}{1+P_{J2}|h_{e}|^2}) \\ \quad -\ln(1+\dfrac{P_{max}|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2}), & \text{(55b)}\\ \theta_{J2}\leq \ln(1+\dfrac{P_{max}|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2}) \\ \quad -\ln(1+\dfrac{P_{max}|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2}), & \text{(55c)}\\ \theta_{J2}-\theta_{E2}\leq \ln(1+\dfrac{P_{max}|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2})-\ln(1+\dfrac{P_{max}|h_{ae2}|^2}{1+P_{J1}|h_{e}|^2}) \\ \quad -\ln(1+\dfrac{P_{max}|h_{ab}|^2}{1+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2}), & \text{(55d)}\\ \dfrac{|h_{ae1}|^2}{1+P_{J1}|h_{e1}|^2+P_{J2}|h_{e2}|^2+P_{max}|h_{ab}|^2}\geq C_a. & \text{(55e)}\end{cases}$$

Proof: Similar to the proof of Lemma 2. ■

REFERENCES

[1] L. Fan *et al.*, "Secure cache-aided multi-relay networks in the presence of multiple eavesdroppers," *IEEE Trans. Commun.*, to be published.

[2] C. Li *et al.*, "Physical-layer secure game against smart attacks in NOMA networks," *Phys. Commun.*, to be published.

[3] Z. Na *et al.*, " Turbo receiver channel estimation for GFDM-based cognitive radio networks," *IEEE Access*, vol. 6, pp. 9926–9935, 2018.

[4] Y. Liu, K.-Y. Lam, S. Han, and Q. Chen, "Mobile data gathering and energy harvesting in rechargeable wireless sensor networks," *Inf. Sci.*, no. 482, pp. 189–209, May 2019.

[5] 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.

[6] 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.

[7] Z. Na *et al.*, "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.

[8] Y. Liu, Q. Chen, and X. Tang, "Adaptive buffer-aided wireless powered relay communication with energy storage," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 2, pp. 432–445, Jun. 2018.

[9] 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*, to be published.

[10] C. Li and W. Zhou, "Enhanced secure transmission against intelligent attacks," *IEEE Access*, to be published.

[11] D. Kapetanovic, G. Zheng, and F. Rusek, "Physical layer security for massive MIMO: An overview on passive eavesdropping and active attacks," *IEEE Commun. Mag.*, vol. 53, no. 6, pp. 21–27, Jun. 2015.

[12] 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.

[13] W.-J. Lei and S.-F. Lan, "An anti-eavesdropping secure transmission scheme using artificial noise with spatial modulation," *J. Univ. Electron. Sci. Technol. China*, vol. 47, no. 1, pp. 13–18, 2018.

[14] 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. Inf. Forensics Security*, vol. 10, no. 12, pp. 2578–2590, Dec. 2015.

[15] 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.

[16] W. Xu, W. Trappe, Y. Zhang, and T. Wood, "The feasibility of launching and detecting jamming attacks in wireless networks," in *Proc. 6th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2005, pp. 46–57.

[17] L. Xiao, X. Lu, D. Xu, Y. Tang, L. Wang, and W. Zhuang, "UAV relay in VANETs against smart jamming with reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4087–4097, May 2018.

[18] Y. Li, L. Xiao, H. Dai, and H. V. Poor, "Game theoretic study of protecting MIMO transmissions against smart attacks," in *Proc. IEEE Int. Conf. Commun.*, May 2017, pp. 1–6.

[19] 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, 2016.

[20] 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.

[21] X. Lin and J. Xia, "MARL-based distributed cache placement for wireless networks," *IEEE Access*, to be published.

[22] G. Gui, H. Huang, Y. Song, and H. Sari, "Deep learning for an effective nonorthogonal multiple access scheme," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8440–8450, Sep. 2018.

[23] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and doa estimation based massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018.

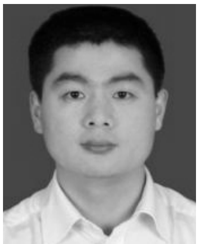
[24] Y. Li, X. Cheng, and G. Gui, "Co-robust-ADMM-net: Joint ADMM framework and DNN for robust sparse composite regularization," *IEEE Access*, vol. 6, pp. 47943–47952, 2018.

[25] Y. Li *et al.*, "Musai-$L_{1/2}$: MUltiple sub-wavelet-dictionaries-based adaptively-weighted iterative half thresholding algorithm for compressive imaging," *IEEE Access*, vol. 6, pp. 16795–16805, 2018.

[26] T. Zhou, S. Yang, L. Wang, J. Yao, and G. Gu, "Improved cross-label suppression dictionary learning for face recognition," *IEEE Access*, vol. 6, pp. 48716–48725, 2018.
[27] J. Pan, Y. Yin, J. Xiong, W. Luo, G. Gui, and H. Sari, "Deep learning-based unmanned surveillance systems for observing water levels," *IEEE Access*, vol. 6, pp. 73561–73571, 2018.
[28] M. Liu, T. Song, and G. Gui, "Deep cognitive perspective: Resource allocation for NOMA based heterogeneous IoT with imperfect SIC," *IEEE Internet Things J.*, to be published.
[29] M. Liu, T. Song, J. Hu, J. Yang, and G. Gui, "Deep learning-inspired message passing algorithm for efficient resource allocation in cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 641–653, Jan. 2019.
[30] J. Zhang and Q. Zhang, "Stackelberg game for utility-based cooperative cognitiveradio networks," in *Proc. 10th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2009, pp. 23–32.
[31] X. Liu, Y. Wang, S. Liu, and J. Meng, "Spectrum resource optimization for NOMA-based cognitive radio in 5g communications," *IEEE Access*, vol. 6, pp. 24904–24911, 2018.
[32] Z. Na, J. Lv, F. Jiang, M. Xiong, and N. Zhao, "Joint subcarrier and subsymbol allocation based simultaneous wireless information and power transfer for multiuser GFDM in IoT," *IEEE Internet Things J.*, to be published.
[33] X. Lan, Q. Chen, X. Tang, and L. Cai, "Achievable rate region of the buffer-aided two-way energy harvesting relay network," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11127–11142, Nov. 2018.
[34] Z. Na, J. Lv, M. Zhang, and M. Xiong, "GFDM based wiereless powered communications for cooperative relay system," *IEEE Access*, to be published.
[35] H. Wang, M. Cheng, Q. Chen, X. Tang, and Q. Huang, "Enhanced adaptive network coded cooperation for wireless networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 11988–12002, Dec. 2018.
[36] Y. Liu, Q. Chen, X. Tang, and L. X. Cai, "On the buffer energy aware adaptive relaying in multiple relay network," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 6248–6263, Sep. 2017.
[37] J. Xia, "When distributed switch-and-stay combining meets buffer in IoT relaying networks," *Phys. Commun.*, to be published.
[38] C. Shannon and W. Weaver, *A Mathematical Theory of Communication*. USA: The Bell System Technical Journal, 1949.
[39] J. Yang *et al.*, "Numerical and experimental study on the thermal performance of aerogel insulating panels for building energy efficiency," *Renew. Energy*, no. 138, pp. 445–457, Aug. 2019.
[40] P. Li, H. Wu, Y. Liu, J. Yang, Z. Fang, and B. Lin, "Preparation and optimization of ultra-light and thermal insulative aerogel foam concrete," *Construct. Building Mater.*, vol. 205, pp. 529–542, Apr. 2019.
[41] Y. Lv *et al.*, "Quantitative research on the influence of particle size and filling thickness on aerogel glazing performance," *Energy Buildings*, vol. 174, pp. 190–198, Sep. 2018.
[42] X. Liu, F. Li, and Z. Na, "Optimal resource allocation in simultaneous cooperative spectrum sensing and energy harvesting for multichannel cognitive radio," *IEEE Access*, vol. 5, pp. 3801–3812, 2017.
[43] X. Liu, M. Jia, Z. Na, W. Lu, and F. Li, "Multi-modal cooperative spectrum sensing based on dempster-shafer fusion in 5G-based cognitive radio," *IEEE Access*, vol. 6, pp. 199–208, 2017.
[44] X. Liu *et al.*, "A multichannel cognitive radio system design and its performance optimization," *IEEE Access*, vol. 6, pp. 12327–12335, 2018.
[45] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive MIMO for hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3027–3032, Mar. 2019. doi: 10.1109/TVT.2019.2893928.
[46] 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.
[47] N. Kato *et al.*, "The deep learning vision for heterogeneous network traffic control: Proposal, challenges, and future perspective," *IEEE Wireless Commun.*, vol. 24, no. 3, pp. 146–153, Jun. 2017.

YAN XU received the bachelor's degree in electronic information science and technology from Hubei Normal University, in 2017. She is currently pursuing the degree with the School of Computer and Education Software, Guangzhou University. Her research interests include wireless cooperative communications, physical-layer security, and multiagent machine learning algorithm.

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 Educational Software, Guangzhou University. Her current research interests include wireless caching, physical-layer security, cooperative relaying, and interference modeling.



HUIJUN WU received the B.E. degree from Tsinghua University, China, in 2000, and the Ph.D. degree from the South China University of Technology, in 2005. He is the Head of the Department of Building Services Engineering, Guangzhou University, China, where he is currently a Professor of building services engineering. His research interests include heat transfer engineering, thermal insulation material, HVAC, data mining, and building energy efficiency.

LISENG FAN received the Ph.D. degree from the Tokyo Institute of Technology, Tokyo, in 2008. He is currently a Teacher with the School of Computer Science, Guangzhou University. His main research interests include the information security, wireless networks, and the artificial intelligence. His current research interest includes the application of artificial intelligence into the wireless networks.