IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ARTIFICIAL INTELLIGENCE FOR PHYSICAL-LAYER WIRELESS COMMUNICATIONS

Received March 28, 2019, accepted April 11, 2019, date of publication April 24, 2019, date of current version May 2, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2912420

Enhanced Secure Transmission Against Intelligent Attacks

CHAO LI[1], WEN ZHOU[2], KAI YU[3], LISENG FAN[1], AND JUNJUAN XIA[1]

[1]School of Computer Science, Guangzhou University, Guangzhou 510006, China
[2]College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, China
[3]China Railway Eryuan Engineering Group Co., Ltd., Chengdu 610031, China

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

This work was supported in part by the National Natural Science Foundation of China under Grant 61871139, in part by the Innovation Team Project of Guangdong Province University under Grant 2016KCXTD017, and in part by the Science and Technology Program of Guangzhou under Grant 201807010103.

ABSTRACT In this paper, we proposed an enhanced secure scheme for the wireless communication system threatened by an intelligent attacker, which can work in eavesdropping, jamming, and spoofing modes. The conventional secure scheme is to apply Q-learning-based algorithm to reach a Nash equilibrium (NE) in the framework of a zero-sum game between the transmitter and attacker, which, however, requires the number of antennas at the transmitter to be much larger than that at the attacker. To overcome this limitation, we first consider the scenario where the attacker can flexibly increase the number of antennas in order to increase the attack rate. By adaptively setting the number of antennas at the transmitter and the legitimate receiver equal to that at the attacker, we then apply the beamforming at the transmitter to suppress the eavesdropping and use the filtering at the receiver to prevent the jamming and spoofing. By incorporating the beamforming and filtering, the benefits of the attacker in this game are efficiently restricted. Furthermore, the Q-learning-based power control strategy is used to reach a new NE. The simulation results have been demonstrated to show that the proposed scheme can suppress the intelligent attack efficiently, which outperforms the conventional scheme in the secrecy performance.

INDEX TERMS Reinforcement learning, model-free, beamforming, zero-sum game.

I. INTRODUCTION

In recent years, the demands on the wireless data rate have been explosively increasing [1]–[4], and many wireless techniques have been proposed to meet the requirements [5]–[8]. As a rapidly developing technology, artificial intelligence has been applied in various fields, such as face recognition [9] and observation water levels [10]. The application of artificial intelligence into wireless communication in [11], [12] has gained much attention recently. Many researchers have used the technique of deep learning network proposed in [13], [14] for channel estimation [15], resource allocation [16] and non-orthogonal multiple access [17], [18]. In many cases, an intelligent agent is not just to identify and classify, it still needs to respond to the current state of environment and take appropriate actions adaptively. The reinforcement learning is proposed in [19] to enable the agents to learn a self-adapting strategy. The task of reinforcement learning is often described as a Markov decision process: the agent executes an action at the current state, and meanwhile the environment feeds back a reward to the agent according to the reward function. After continuous trial-error and exploration in the environment, the agent can obtain a learning-based policy which can maximize the long-term reward. However, in practical situations, it is difficult for the agent to detect the state space of the environment and the state transition probability. In order to solve the problem, the authors in [20] have proposed a model-free reinforcement learning algorithm, named Q-learning, which is more suitable for the secure wireless transmission.

Secure wireless transmission is of vital importance for the future mobile communication networks [21], [22]. Intelligent attacker with reinforcement learning ability seriously threatens the security of wireless communication [23], [24]. It is difficult for the transmitter to detect the channel status information (CSI) [25] between the transmitter and receiver, and the transmitter is even unable to predict the action of the intelligent attacker. In the complex radio environment, the transmitter can only adaptively control its transmit power

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

and the number of antennas. Hence it is particularly important for the wireless communication systems to adopt the secure transmission scheme. A power control policy has been proposed in [26] for the wireless communication systems, which works well under a single attack mode. To tackle the challenge of multiple attack modes, the attacker in [27] freely works in eavesdropping, jamming, and spoofing modes. The conventional power control schemes can only work well for the secure communication game with fixed number of antennas, where the number of antennas at the transmitter is much larger than that at the attacker [27]. In the practical communication scenarios, the number of antennas at the transmitter is maybe equal to that at the attacker, in which the conventional power control scheme fails to work. This motivates the research in this paper.

In this paper, we consider a rivalry wireless communication system where an intelligent attacker exists, which reduces the secrecy data rate of the system by flexibly working in eavesdropping, jamming, and spoofing modes. Moreover, the attacker can flexibly increase the number of antennas in order to enhance its attack ability. To deal with the attack, we firstly set the number of antennas at the transmitter and receiver adaptively equal to that at the attacker. We then apply the beamforming at the transmitter to suppress the eavesdropping, and adopt the filtering at the receiver to prevent the jamming and spoofing. We further propose an enhanced secure transmission policy based on the Q-learning, where the transmitter and attacker are considered as two players in a noncooperative zero-sum game framework. In such a game, the attacker chooses to execute one attack mode among eavesdropping, jamming or spoofing, which changes the radio environment of the transmitter from one state to another. Meanwhile, the transmitter computes the secrecy data rate as the feedback reward. By combining the Monte Carlo and dynamic programming methods, the transmitter finally acquires the optimal transmit power to maximize the average secrecy data rate. By incorporating the beamforming and filtering we proposed, the reward of the attacker is efficiently restricted when the attacking modes are executed. Furthermore, we deduce a new Nash equilibrium (NE) of the game. From the simulation results, we find that the secure transmission policy approaches to the NE, and the proposed scheme outperforms the conventional ones significantly.

The main contributions of this paper are summarized as follows:

- We propose an enhanced secure transmission system against the intelligent attacker, which can work in eavesdropping, jamming, and spoofing modes. Moreover, we consider the attacker can flexibly increase the number of antennas in order to increase the attack probability.
- We apply the beamforming at the transmitter to suppress the eavesdropping, and use the filtering at the receiver to prevent the jamming and spoofing.
- We adopt a Q-learning based algorithm for the secure transmission to optimize the power control scheme, and we derive a new NE solution, which is proven to be optimal for the secure transmission game.

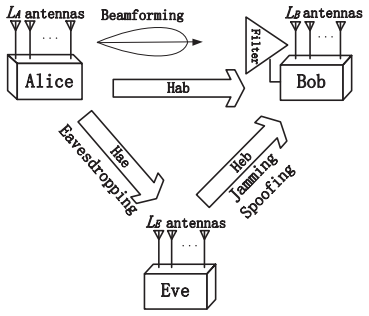

FIGURE 1. System model of a rivalry communication system with an intelligent attacker.

II. BEAMFORMING AND FILTERING PROCESS

To discuss the beamforming and filtering process in the rivalry game, we need to first introduce the system model, which is depicted in Fig. 1. This system consists of an intelligent transmitter Alice and a smart attacker Eve. The Eve is equipped with multiple antennas, and it can flexibly select some of them to attack the secure communication. In addition, Eve can select to execute several modes of attack, and we use m to denote the attack type. If the beamforming is not used, the Eve has better opportunity to overhear the secure message from the Alice. If the filtering is not used, the Eve has better opportunity to perform jamming or spoofing. Specifically, $m = 0, 1, 2$ and 3 represent the silent, eavesdropping, jamming and spoofing modes at the Eve, respectively. To prevent the smart attack, the intelligent transmitter Alice executes in a rivalry way, by adaptively adjusting its number of antennas to that used at the Eve, and meanwhile consciously changing its transmit power p_t. The value of p_t varies in the range of $[0, P_{\max}]$, where $P_{\max}$ is the maximum transmit power at the Alice.

As both the Alice and Eve are intelligent which have the reinforcement learning ability, the rivalry of them in the process of transmission can be formulated as a secure game. In such a game, the Eve can flexibly increase the number of used antennas to help strengthen the attack. To tackle this problem, the beamforming at the Alice and the filtering at the Bob are adopted in this work to prevent the attack. Specifically, let L_A, L_B and L_E denote the number of used antennas at the Alice, Bob and Eve, respectively, where L_E can be flexibly increased to help strengthen the attack. We use $\mathbf{H}_{AB} \sim \mathcal{CN}(0, \alpha\mathbf{I})$, $\mathbf{H}_{AE} \sim \mathcal{CN}(0, \beta\mathbf{I})$ and $\mathbf{H}_{EB} \sim \mathcal{CN}(0, \varepsilon\mathbf{I})$ to represent the channel parameters of the Alice-Bob, Alice-Eve and Eve-Bob links, respectively.

Moreover, we use $\mathbf{w}_A$ and $\mathbf{w}_B$ to represent the beamforming and filtering vectors, respectively. Note that these two vectors are maybe not optimal. Then, the Alice chooses a value for the transmit power p_t to send the beamformed signal $\mathbf{w}_A s_A$, where s_A is normalized to unity. After that, the received signal at the Bob is input to the filter, and the resultant signal is

$$y_{m,B} = \mathbf{w}_B^H \mathbf{H}_{AB} \mathbf{w}_A \sqrt{p_t} s_A + \mathbf{w}_B^H \mathbf{n}_B, \tag{1}$$

where $m = 0$ and 1 indicate the silent and eavesdropping modes, respectively. The term $\mathbf{n}_B \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$ is the additive white Gaussian noise (AWGN) at the Bob. The details about the noise effect on the wireless communication systems can be found in [28]–[30]. When $m = 1$ holds, the eavesdropping signal at the Eve is

$$y_E = \mathbf{H}_{AE}\mathbf{w}_A\sqrt{p_t}s_A + \mathbf{n}_E, \tag{2}$$

where $\mathbf{n}_E \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$ is the AWGN at the Eve. When the Eve selects to send a jamming signal s_J with $m = 2$, the output signal of the filter at the Bob becomes

$$y_{m,B} = \mathbf{w}_B^H\mathbf{H}_{AB}\mathbf{w}_A\sqrt{p_t}s_A + \mathbf{w}_B^H\mathbf{H}_{EB}\sqrt{p_J}s_J + \mathbf{w}_B^H\mathbf{n}_B, \tag{3}$$

where p_J is the jamming power at the Eve. When $m = 3$ holds, the Alice does not transmit while the Eve transmits the spoofing signal s_S. The output signal of the filter at the Bob becomes

$$y_{m,B} = \mathbf{w}_B^H\mathbf{H}_{EB}\sqrt{p_S}s_S + \mathbf{w}_B^H\mathbf{n}_B, \tag{4}$$

where p_S is the spoofing power at the Eve.

We now turn to solve the beamforming and filtering vectors $\mathbf{w}_A$ and $\mathbf{w}_B$. To this end, we use the singular value decomposition (SVD) to decompose $\mathbf{H}_{AB} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^H$, where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_{L_B}]$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_{L_A}]$ are two unitary matrices, and $\mathbf{\Lambda} = \text{diag}[\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots \boldsymbol{\lambda}_{\min(L_A,L_B)}]$ is the singular matrix in which the singular values are ordered in a descender order. Similarly, H_{EB} is decomposed as $\mathbf{H}_{BE} = \mathbf{U}_E\mathbf{\Lambda}_E\mathbf{V}_E^H$, where $\mathbf{U}_E = [\mathbf{u}_{E,1}, \mathbf{u}_{E,2}, \ldots \mathbf{u}_{E,L_B}]$ and $\mathbf{V}_E = [\mathbf{v}_{E,1}, \mathbf{v}_{E,2}, \ldots \mathbf{v}_{E,L_E}]$ are two unitary matrices, and $\mathbf{\Lambda}_E = \text{diag}[\boldsymbol{\lambda}_{E,1}, \boldsymbol{\lambda}_{E,2}, \ldots \boldsymbol{\lambda}_{E,\min(L_E,L_B)}]$ is the singular matrix in which the singular values are ordered in a descender order. From these two decompositions, we can set $\mathbf{w}_A$ and $\mathbf{w}_B$ as

$$\mathbf{w}_A = \mathbf{v}_1, \tag{5}$$

$$\mathbf{w}_B = \mathbf{u}_{E,L_B}, \tag{6}$$

in which (5) can maximize the equivalent channel gain of the main link, while (6) can minimize the equivalent channel gain of the jamming and spoofing links. There are some more efficient beamforming and filtering schemes in the literature [31]–[33], which will be studied in our future works.

The secrecy data rate of the system is denoted by C_0, C_1, C_2 and C_3, for $m = 0$, 1, 2 and 3, respectively. From (1)-(6), the secrecy data rate is given by

$$C_0 = \log_2(1 + \widetilde{p}_t|\mathbf{w}_B^H\mathbf{H}_{AB}\mathbf{w}_A|^2), \tag{7a}$$

$$C_1 = C_0 - \log_2(1 + \widetilde{p}_t(\mathbf{H}_{AE}\mathbf{w}_A)^H(\mathbf{H}_{AE}\mathbf{w}_A)), \tag{7b}$$

$$C_2 = \log_2(1 + \frac{\widetilde{p}_t|\mathbf{w}_B^H\mathbf{H}_{AB}\mathbf{w}_A|^2}{1 + \widetilde{p}_J(\mathbf{w}_B^H\mathbf{H}_{EB})(\mathbf{w}_B^H\mathbf{H}_{EB})^H}), \tag{7c}$$

$$C_3 = C_0 - \zeta\log_2(1 + \widetilde{p}_S(\mathbf{w}_B^H\mathbf{H}_{EB})(\mathbf{w}_B^H\mathbf{H}_{EB})^H), \tag{7d}$$

where $\widetilde{p}_t = p_t/\sigma^2$, $\widetilde{p}_J = p_J/\sigma^2$ and $\widetilde{p}_S = p_S/\sigma^2$ are the transmit, jamming and spoofing power normalized by the average noise power. In addition, $\zeta \in (0, 1)$ in (7d) is the probability of the influence of spoofing message. Note that in eqs. (7c) and (7d), we only list the secrecy data rate expression of all possible attacker modes. In practice, the system need not know the transmission power and modes of the Eve. The proposed Q-learning based algorithm can protect the secure communication from the smart attacker which can arbitrarily change its mode and transmit power. Moreover, when m = 3 holds, the Alice does not transmit any signal, and then the attacker selects to perform spoofing. The secure data rate in eq. (7d) is used to characterize how the spoofing signal looks like to actual useful signal from the Alice. In other words, eq. (7d) expresses the similarity between the spoofing signal and the actual useful signal from the Alice. The details about eq. (7d) can be found in the literature, such as Ref. [27].

III. POLICY OF SECURE COMMUNICATION GAME

In this work, both the Alice and Eve have the reinforcement learning ability, and hence the rivalry of them in the transmission process is formulated as a secure communication game. In such a game, one agent's environment is the result of the other's action, i.e., the Alice and Eve determine their actions by observing the rival's behaviors. The process of the game is performed in a time sequence, and for each time slot, the two agents choose an action according to the state of the current moment and obtain a reward gain. Specifically, the Alice chooses a transmit power p_t and adaptivity adjusts the number of transmit antennas L_A, while the Eve selects to perform the action mode $m \in \{0, 1, 2, 3\}$ with variable L_E. The target of the Alice and Eve is to select some optimal action in order to maximize their holistic reward after many time tests. After some interval, the Eve checks the probability of keeping silent. If the silent probability is higher than a given threshold, then the Eve will select to increase the number of antennas by one, in order to break the equilibrium and increase the attack rate. In accordance with L_E, the Alice adaptively adjusts its antenna number L_A equivalent to L_E. Note that the system can estimate the channel parameters of the Eve when it is active in the network. By analyzing the dimension of the channel matrix, the system can know the antenna numbers of the Eve.

We repeat the above game for many times, and the two agents will learn the corresponding optimal policy (p_t^*, m^*). Nash equilibrium of this game is a set that satisfies the two agents' policy, which is defined as

$$R_A(p_t^*, m^*) \geq R_A(p_t, m^*) \quad \forall 0 \leq p_t \leq P_{\max}, \tag{8}$$

$$R_E(p_t^*, m^*) \geq R_E(p_t^*, m) \quad \forall m = 0, 1, 2, 3, \tag{9}$$

where R_A and R_E denote the reward functions of the Alice and Eve, respectively. We can compute R_A based on the secrecy date rate and the transmission cost. Note that the increase in the number of antennas will cause more cost, and we set the total transmission cost of Alice to $p_t L_A \mu$, where μ is the cost of unit power. Accordingly, we can compute the reward function of Alice, $R_A(p_t, m)$, as

$$R_A(p_t, m) = \ln 2C_m - p_t L_A \mu, \tag{10}$$

where C_m is the secrecy data rate when the Eve selects to perform the m-th action mode. The logarithmic basis in eq. (10) is the natural basis, instead of the basis of 2. This can facilitate the partial operation in the subsequent derivation process. In the practical scenarios, we expect that only the Alice can win in this confrontation between the Alice and Bob. Therefore, we consider the secure communication game as a zero-sum game, where the reward value function of Eve $R_E(p_t, m)$ can be given by

$$R_E(p_t, m) = -\ln 2C_m - L_E \nu_m, \tag{11}$$

where ν_m is the cost of a single antenna at the Eve to execute the action mode m. Similarly, the logarithmic basis in eq. (10) is the natural basis, instead of the basis of 2. This can facilitate the partial operation in the subsequent derivation process.

We next derive the Nash equilibrium (NE) solution of the secure communication game based on $R_A(p_t, m)$ and $R_E(p_t, m)$. According to the definition of NE, if one player between the Alice and Eve keeps adopting the NE policy, the other one cannot obtain more reward gain by changing its policy. We find that it is most beneficial for the Eve to keep silent when the Alice transmits a signal with power x. Therefore, we show the NE solution $(x, 0)$ in the following Lemma 1.

Lemma 1: The secure communication game has one NE policy $= (x, 0)$, i.e., Eve chooses the no-attack policy when Alice uses the optimal transmit power x, which is computed by

$$x = \frac{|\mathbf{w}_B^H \mathbf{H}_{AB} \mathbf{w}_A|^2 - L_A \mu}{L_A \mu |\mathbf{w}_B^H \mathbf{H}_{AB} \mathbf{w}_A|^2} \quad \exists 0 \le x \le P_{\max}, \tag{12}$$

if the following conditions are satisfied:

$$\frac{1}{L_A/|\mathbf{w}_B^H \mathbf{H}_{AB} \mathbf{w}_A|^2 + P_{\max} L_A} < \mu < \frac{\mathbf{w}_B^H \mathbf{H}_{AB} \mathbf{w}_A}{L_A} \tag{13a}$$

$$\nu_1 \ge \frac{\ln(1 + x(\mathbf{H}_{AE} \mathbf{w}_A)^H (\mathbf{H}_{AE} \mathbf{w}_A))}{L_E} \tag{13b}$$

$$\nu_2 \ge \frac{\ln(1 + \frac{\widetilde{p}_J x |\mathbf{w}_B^H \mathbf{H}_{AB} \mathbf{w}_A|^2 (\mathbf{w}_B^H \mathbf{H}_{EB})(\mathbf{w}_B^H \mathbf{H}_{EB})^H}{1 + x|\mathbf{w}_B^H \mathbf{H}_{AB} \mathbf{w}_A|^2 + \widetilde{p}_j (\mathbf{w}_B^H \mathbf{H}_{EB})(\mathbf{w}_B^H \mathbf{H}_{EB})^H})}{L_E} \tag{13c}$$

$$\nu_3 \ge \zeta \ln(1 + \widetilde{p}_s (\mathbf{w}_B^H \mathbf{H}_{EB})(\mathbf{w}_B^H \mathbf{H}_{EB})^H)/L_E \tag{13d}$$

Proof: According to (10), when $m = 0$ holds, we take the partial derivative of $R_A(p_t, m)$ with respect to p_t, and have

$$\frac{\partial R_A(p_t, 0)}{\partial p_t} = \frac{1}{p_t + 1/|\mathbf{w}_B^H \mathbf{H}_{AB} \mathbf{w}_A|^2} - L_A \mu, \tag{14}$$

from which we easily find

$$\frac{\partial R_A^2(p_t, 0)}{\partial p_t^2} = -\left(\frac{1}{p_t + 1/|\mathbf{w}_B^H \mathbf{H}_{AB} \mathbf{w}_A|^2}\right)^2 < 0. \tag{15}$$

The above equation shows that $R_A(p_t, 0)$ is a convex function. We solve the following equation when $(p_t, 0) = (x, 0)$ holds

$$\frac{\partial R_A(x, 0)}{\partial x} = \frac{1}{x + 1/|\mathbf{w}_B^H \mathbf{H}_{AB} \mathbf{w}_A|^2} - L_A \mu = 0, \tag{16}$$

from which we have

$$x = \frac{|\mathbf{w}_B^H \mathbf{H}_{AB} \mathbf{w}_A|^2 - L_A \mu}{L_A \mu |\mathbf{w}_B^H \mathbf{H}_{AB} \mathbf{w}_A|^2} \quad 0 \le p_t \le P_{\max}. \tag{17}$$

Hence, $R_A(p_t, 0)$ achieves the local maximum value when $(p_t, 0) = (x, 0)$. By letting the following two equations hold,

$$\frac{\partial R_A(p_t, 0)}{\partial p_t}|_{p_t = 0} = \frac{1}{p_t + 1/|\mathbf{w}_B^H \mathbf{H}_{AB} \mathbf{w}_A|^2} - L_A \mu > 0, \tag{18}$$

$$\frac{\partial R_A(p_t, 0)}{\partial p_t}|_{p_t = P_{\max}} = \frac{1}{p_t + 1/|\mathbf{w}_B^H \mathbf{H}_{AB} \mathbf{w}_A|^2} - L_A \mu < 0, \tag{19}$$

we find that the inequality in (13a) is a tenable condition where $R_A(p_t, 0)$ achieves the global maximum value. Hence, it has been proven that $(x, 0)$ satisfies (8).

Then, we can write the following inequalities from (11) as

$$\begin{aligned}
&R_E(x, 0) - R_E(x, 1) \\
&= L_E \nu_1 - \ln(1 + x(\mathbf{H}_{AE} \mathbf{w}_A)^H (\mathbf{H}_{AE} \mathbf{w}_A)) \ge 0, \\
&R_E(x, 0) - R_E(x, 2) \\
&= L_E \nu_2 - \ln(1 + \frac{\widetilde{p}_J x |\mathbf{w}_B^H \mathbf{H}_{AB} \mathbf{w}_A|^2 (\mathbf{w}_B^H \mathbf{H}_{EB})(\mathbf{w}_B^H \mathbf{H}_{EB})^H}{1 + x|\mathbf{w}_B^H \mathbf{H}_{AB} \mathbf{w}_A|^2 + \widetilde{p}_j (\mathbf{w}_B^H \mathbf{H}_{EB})(\mathbf{w}_B^H \mathbf{H}_{EB})^H}) \\
&\ge 0, \\
&R_E(x, 0) - R_E(x, 3) \\
&= L_E \nu_3 - \zeta \ln(1 + \widetilde{p}_s (\mathbf{w}_B^H \mathbf{H}_{EB})(\mathbf{w}_B^H \mathbf{H}_{EB})^H) \ge 0.
\end{aligned}$$

Thus, the inequalities in (13b)-(13d) are the tenable conditions where $(x, 0)$ satisfies (9).

In summary, $(x, 0)$ is the strict NE policy of the secure communication game. The details about the NE policy can be found in the literature, such as the works in [34]–[36]. In this way, the proof of Lemma 1 is completed. ■

IV. POWER CONTROL ALGORITHM

We employ a power control algorithm for the Alice based on Q-learning, which is widely used in artificial intelligent filed as a typical and powerful model-free reinforcement learning method. The main motivation of adopting the Q-learning is that it is difficult for the Alice to detect the channel state information and the state transition probability. To solve this problem, Q-learning is used for the Alice to control the transmit power, in order to achieve the optimal policy and action.

The power control algorithm is based on the temporal-difference in essence, and it is a combination of the Monte Carlo and dynamic programming. Overall, the feature of the temporal-difference algorithm is that we use many times of test to perform the attack and protection. By using the temporal-difference process, the algorithm can find the best awards for both sides. During the process, the test results of previous time slots are saved in a table, based on which the current state is updated. Firstly, we randomly initialize the Q-table $Q(s, a)$, which consists of the state-action pairs (s, a). At each episode of training, Alice explores the environment

from an initial state s to the terminal state. At time t, the action mode of Eve is m, which can be regarded as the state of Alice, denoted by $s_t = m$. The Alice then chooses an action a_t[1] according to the state s_t, and meanwhile acquires the state of the next time s_{t+1} and the reward gain R_A. We jointly use the reward value R_A and action-value function of the next state $Q(s_{t+1}, a)$ to update the Q-table, and we can write the process as

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \theta[R_A + \gamma max_a Q(s_{t+1}, a) - Q(s_t, a_t)], \quad (20)$$

where $\theta \in (0, 1]$ is the learning rate, and it determines the speed of policy update. The larger of θ, the greater the weight of the retaining current experience. The discount factor $\gamma \in [0, 1]$ represents the probability that Alice attaches importance to the reward in memory. We do not know the state transition probability of Alice, so we need to take many episodes to get the expected action-value function, just like the idea of Monte Carlo. After enough experiments, the learning based Q-table approaches to the optimal Q^*-table. However, if the Alice exploits the current optimal action in the Q-table every time, it will probably approach to a local optimal policy. In order to obtain the global optimal policy, we employ the ϵ-greedy policy where the Alice selects an action in this algorithm to reach a tradeoff between the exploration and exploitation, i.e., the Alice exploits the current optimal action with probability ϵ, or randomly selects an action with probability $(1 - \epsilon)$. After some interval, the Eve checks the probability of keeping silent. If the silent probability is higher than a given threshold, then the Eve will increase the number of antenna to strengthen the attack ability. To deal with this problem, the Alice adaptively adjusts its antennas number L_A equal to L_E.

As summarized above, Alice can learn the optimal policy and action according to the power control algorithm, which is given in the following algorithm,

Algorithm 1 Power Control Algorithm

1: Input parameters: θ, γ, ϵ
2: Initialize $Q(s, a)$, for all $s \in 0, 1, 2, 3$, $a \in [0, P_{\max}]$ at random
3: Loop for each episode:
4: Initialize s
5: loop for each time slot of episode:
6: Choose a_t from s_t using policy derived from $Q(\epsilon - greedy)$
7: Take action a, observe R_A, s_{t+1}
8: $Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \theta[R_A + \gamma max_a Q(s_{t+1}, a) - Q(s_t, a_t)]$
9: $s_t \leftarrow s_{t+1}$
10: Observe Eve's antennas L_E, and set $L_A = L_E$
11: until s is terminal

V. SIMULATION RESULTS

In this section, the power control algorithm we proposed for the secure communication game was evaluated via simulation. According to the tenable conditions of NE in $(13a) - (13d)$, we set the system parameters as: $\{\alpha, \beta, \varepsilon\} = \{1.2, 0.5, 2\}$, $\mu = 0.1$, $\nu_{m=\{0,1,2,3\}} = \{0, 2.5, 3.2, 3\}$, $\zeta = 0.5$, $p_j = 3.2$ and $p_s = 3$. To make the results more clear, we simply assume that the Eve observes the silent probability every 10000 time slots throughout the whole process, and increases one antenna at every turn if the silent probability is higher than 90%. Therefore, the whole process which consists of 40000 time slots is divided into four phases.

Fig. 2 illustrates the mode probabilities of the Eve versus the time slot varying from 0 to 40000, where the number of

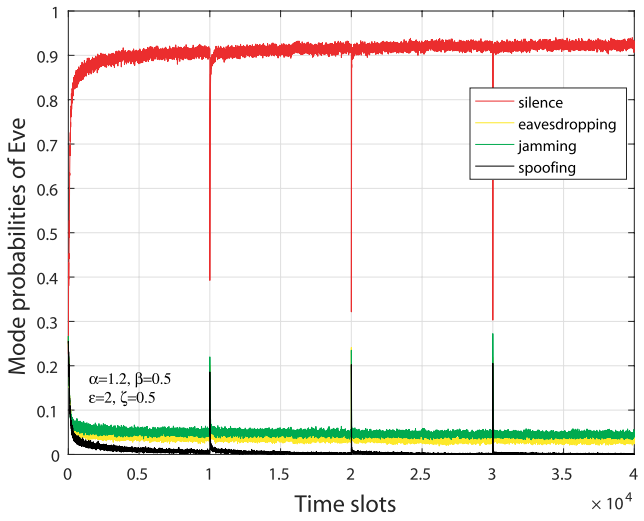

FIGURE 2. The mode probabilities of the Eve with variable number of antennas.

the used antennas at the Eve adaptively increases from 1 to 4. In the first phase, the average silent probability sharply rises up to 90% from 0 to about 3000 time slots. From 3000 to 6000 time slots, the silent probability grows up very slowly, and it approaches to a steady level of 91% after 6000 time slots. This indicates that a balance is almost achieved between the attack and protection. On the contrary, the eavesdropping, jamming and spoofing probabilities sharply fall below 5%. Then, at the point of 10000 time slot, the Eve checks the silent probability. If the probability is higher than 90%, the Eve increases the number of antennas by one. The Alice and Bob adaptively set the number of antennas equal to that at the Eve. For the second phase with the time slot in the range of [10000, 20000], the mode probability becomes convergent much more quickly. This also indicates that a balance is almost achieved between the attack and protection. Furthermore, the silent probability is higher than that in the first phase. These observations also hold for the third and fourth phases.

Fig. 3 shows the secrecy data rate of the secure transmission system versus the time slot varying from 0 to 40000,

[1] In our work, the Alice chooses an action a among all possible values of transmit power. To simplify the selection, we the transmit power into limited levels. Then, the Alice can choose an action a among the limited levels of transmit power.

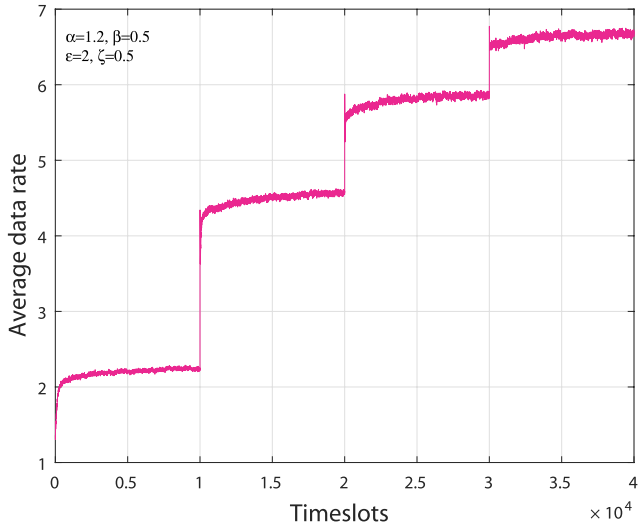

FIGURE 3. The secrecy data rate of the proposed secure communication policy with variable number of antennas.

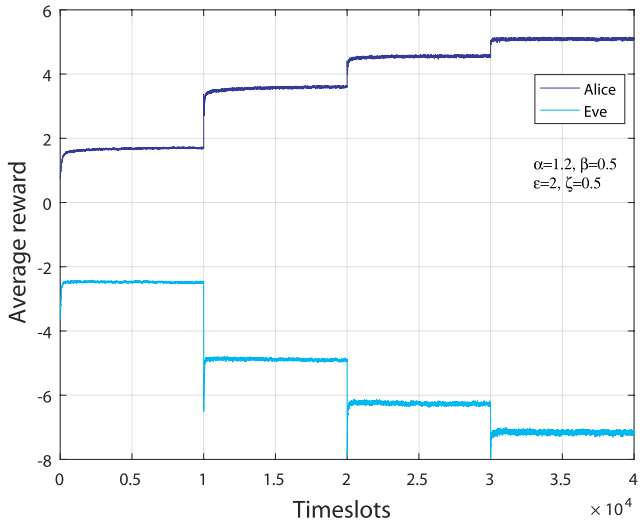

FIGURE 4. The average reward of the proposed secure communication policy with variable number of antennas.

where the number of the used antennas at the Eve adaptively increases from 1 to 4. In the first phase, we find that the average secrecy data rate sharply increases from 0 to 3000 time slots. From 3000 to 6000 time slots, the secrecy data rate rises up very slowly, and it approaches to a steady level after 6000 time slot. This indicates that a balance is almost achieved between the attack and protection. At the point of 10000 time slot, both the Eve and Alice increase the number of antennas by one. For the second phase with the time slot in the range of [10000, 20000], the secrecy data rate dramatically rises up, and it becomes convergent more quickly compared with the first phase. These observations also hold for the third and fourth phases.

In Fig. 4, the two curves show the average reward values of the Alice and Eve versus the time slot varying from 0 to 40000, where the number of the used antennas at the Eve adaptively increases from 1 to 4. In the first phase,

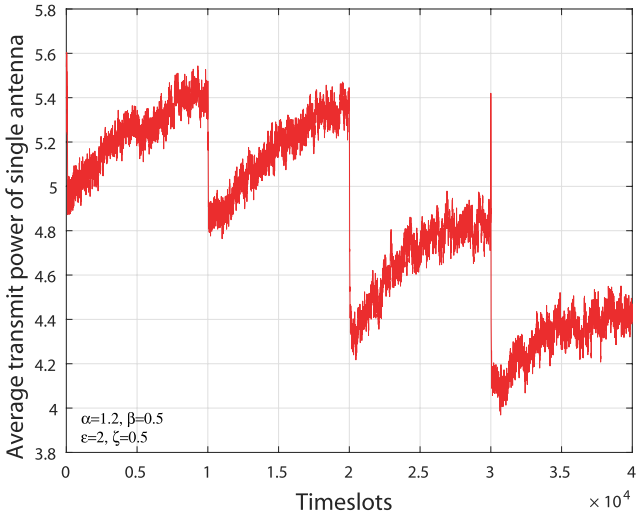

FIGURE 5. The average transmit power of the proposed secure communication policy with variable number of antennas.

the reward values of the Alice and Eve grow up quickly, and they both approach to a steady level, which indicates that a balance is almost achieved between the attack and protection. In the second phase, for the reason of the increasing number of antennas by one, the reward of the Alice rapidly increases by 100%, and meanwhile the reward of the Eve decreases by 100%. For the subsequent phases, the convergent reward value of the Alice becomes much higher when the number of the antennas increases.

Fig. 5 illustrates the transmit power of the Alice versus the time slot varying from 0 to 40000, where the number of the used antennas at the Eve adaptively increases from 1 to 4. It is apparent that the transmit power of the Alice gradually grows up when the time slot increases from 0 to 10000, and it approaches to the peak value of 5.5, where a balance is almost achieved between the attack and protection. For the second phase with the time slot in the range of [10000, 20000], the transmit power of Alice falls to a temporary value of 4.8 at the point of 10000 time slot because of the increasing number of the antennas, and then it gradually rises to 5.4. These similar observations also hold for the third and fourth phases.

By summarizing the above analysis, we can conclude that the secure transmission policy we proposed can enable the Alice to approach to the learning based optimal policy, and it can efficiently enhance the secrecy data rate and meanwhile reduce the attack probabilities, irrespective of the number of the used antennas at the Eve.

VI. CONCLUSIONS

In this paper, we presented the enhanced secure transmission against the attacker which flexibly increased the number of antennas in order to help strength the attack. In the transmission process, the number of antennas at the transmitter and legitimate receiver was adaptively set equal to that at the attacker. Then, the beamforming at the transmitter was employed to suppress the eavesdropping, and the filtering at

the receiver was used to prevent the jamming and spoofing. In this way, the benefits of the attacker were efficiently restricted, and a new NE was reached by the Q-learning based power control strategy. By simulation, we confirmed that the performance of the proposed scheme incorporating the beamforming and filtering could outperform the conventional schemes. In the future works, we will incorporate some other wireless transmission techniques such as resource allocation [37]–[40] and IOT techniques [41]–[44] to further enhance the security of the considered system.

REFERENCES

[1] 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.
[2] Z. Na *et al.*, " Turbo receiver channel estimation for GFDM-based cognitive radio networks," *IEEE Access*, vol. 6, pp. 9926–9935, 2018.
[3] 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.
[4] X. Liu, X. Zhang, M. Jia, L. Fan, W. Lu, and X. Zhai, "5G-based green broadband communication system design with simultaneous wireless information and power transfer," *Phys. Commun.*, vol. 28, pp. 130–137, Jun. 2018.
[5] 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.
[6] X. Liu *et al.*, "A multichannel cognitive radio system design and its performance optimization," *IEEE Access*, vol. 6, pp. 12327–12335, 2018.
[7] 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.
[8] 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, 2013.
[9] T. Zhou, S. Yang, L. Wang, J. Yao, and G. Gui, "Improved cross-label suppression dictionary learning for face recognition," *IEEE Access*, vol. 6, pp. 48716–48725, 2018.
[10] J. Pan *et al.*, "Deep learning-based unmanned surveillance systems for observing water levels," *IEEE Access*, vol. 6, pp. 73561–73571, 2018.
[11] 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.
[12] 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.
[13] 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.
[14] Y. Li *et al.*, "MUSAI-ℓ_1: MUltiple sub-wavelet-dictionaries-based adaptively-weighted iterative half thresholding algorithm for compressive imaging," *IEEE Access*, vol. 6, pp. 16795–16805, 2018.
[15] 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.
[16] M. Liu, J. Yang, T. Song, J. Hu, 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.
[17] G. Guan, 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.
[18] 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.
[19] R. S. Sutton and A. G. Barto, "Reinforcement learning," *A Bradford Book*, vol. 15, no. 7, pp. 665–685, 1998.
[20] 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.
[21] 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.
[22] 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.
[23] 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.
[24] Y. Xu and J. Xia, "Q-learning based physical-layer secure game against multi-agent attacks," *IEEE Access*, vol. 7, pp. 49212–49222, 2019.
[25] X. Lin and J. Xia, "MARL-based distributed cache placement for wireless networks," *IEEE Access*, to be published.
[26] M. Liu and L. Yuan, "Power allocation for secure SWIPT systems with wireless-powered cooperative jamming," *IEEE Commun. Lett.*, vol. 21, no. 6, pp. 1353–1356, Jun. 2017.
[27] 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, Nov. 2018.
[28] C. Li, "Physical-layer secure game against smart attacks in NOMA networks," *Phys. Commun.*, to be published.
[29] 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.
[30] 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.
[31] J. Xia, "When distributed switch-and-stay combining meets buffer in IOT relaying networks," *Phys. Commun.*, to be published.
[32] 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.
[33] 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.
[34] 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.
[35] 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. 1, no. 205, pp. 529–542, 2019.
[36] Y. Lv *et al.*, "Quantitative research on the influence of particle size and filling thickness on aerogel glazing performance," *Energy Buildings*, vol. 174, no. 1, pp. 190–198, 2018.
[37] Y. Liu, K.-Y. Lam, S. Han, and Q. Chen, "Mobile data gathering and energy harvesting in rechargeable wireless sensor networks," *Inf. Sci.*, vol. 482, pp. 189–209, May 2019.
[38] L. Fan, X. Lei, P. Fan, and R. Q. Hu, "Outage probability analysis and power allocation for two-way relay networks with user selection and outdated channel state information," *IEEE Commun. Lett.*, vol. 16, no. 5, pp. 638–641, May 2012.
[39] 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, 2018.
[40] 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.
[41] 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.
[42] L. Fan *et al.*, "Secure cache-aided multi-relay networks in the presence of multiple eavesdroppers," *IEEE Trans. Commun.*, to be published.
[43] 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.
[44] Z. Na, J. Lv, M. Zhang, and M. Xiong, "GFDM based wiereless powered communication for cooperative relay system," *IEEE Access*, to be published.

Authors' photographs and biographies not available at the time of publication.