IET Communications

Special Section: Cognitive and AI-enabled Wireless and Mobile Communications

Rollout algorithm for light-weight physical-layer authentication in cognitive radio networks

ISSN 1751-8628
Received on 1st December 2019
Revised 5th May 2020
Accepted on 26th May 2020
E-First on 13th July 2020
doi: 10.1049/iet-com.2019.1275
www.ietdl.org

Shengnan Yan[1,2], *Xiaoding Wang*[1,2], *Li Xu*[1,2] ✉

[1]*College of Mathematics and Informatics, Fujian Normal University, Fuzhou, Fujian, 350117, People's Republic of China*

[2]*Key Laboratory of Network Security and Cryptology, Fujian Normal University, Fuzhou, Fujian, 350117, People's Republic of China*

✉ *E-mail: xuli@fjnu.edu.cn*

Abstract: Cognitive radio networks (CRNs) are vulnerable to spoofing attacks due to their wireless and cognitive nature. Since the traditional cryptographic authentication can hardly prevent such attacks in CRNs, the physical-layer authentication has been investigated for recent years. To achieve a light-weight physical-layer authentication, a rollout partially observable Markov decision process-based algorithm, named RoPOMDP, is proposed in this study. In general, RoPOMDP formulates the physical-layer authentication as a zero-sum game, based on which a hypothesis test upon channel vectors is developed. That allows us to design the gains for both spoofers and receivers based on Bayesian risks for the game, in which the spoofing attack probability is predicted by a non-linear function approximation utilising v-support vector regression. Then, a RoPOMDP is employed to estimate the optimal threshold for the test statistic such that spoofing attacks can be detected. The theoretical analysis and simulations indicate that: (i) RoPOMDP improves the spoofing detection accuracy; (ii) as a light-weight algorithm, the complexity of RoPOMDP is lower than contemporary ones.

1 Introduction

Since the introduction of cognitive radios was first proposed by Mitola and Maguire [1] in 1999. Cognitive radio networks (CRNs) have since emerged, where the cognitive users (CUs) can use idle frequency spectrum bands without affecting the primary users, thereby improving the utilisation of spectrum resources and solving the problem of spectrum shortage. However, the wireless and cognitive nature of CRN makes it susceptible to many security threats. For example, some spoofer uses faked media access control (MAC) addresses to send false perception reports [2]. Thus, having an effective authentication mechanism could improve the security of CRNs.

While it remains difficult to fully address spoofing attacks in CRNs, several physical-layer authentication (PHY-authentication) techniques have been proposed to address this challenge. These techniques use properties of channels such as received signal strengths (RSSs) [3–5], channel impulse response [6, 7], and channel state information (CSI) [8–10] to discriminates senders. The accuracy of such physical-layer authentication depends on the authentication threshold. Thus, it is important to choose the right threshold. In [11], Liang *et al*. use a pre-assigned threshold and adaptive threshold to detect spoofing attacks. In [3], the spatial correlation of RSS is used to detect spoofing attacks. The proximity-based authentication in [5] uses the RSS variations in proximity tests at mobile stations. In [8], CSI is used to distinguish radio transmitters with similar signal fingerprints. The time-varying carrier frequency offset between transmit and receive pairs is used in authentication [12]. The channel phase response in a multi-carrier system that can be used for physical layer authentication is introduced in [7]. The principle of indirect reciprocity is applied to solve the wide range of attacks in wireless networks [13]. The frequency hopping strategy in [14] is used by secondary users to update their anti-jamming strategy with incomplete knowledge. In [15], a two-level game model is introduced to study the joint threat of high-level persistent threat attackers and insider. There exist a series of machine learning-based authentication strategies. In [16, 17], the interactions between a receiver and a spoofing attacker are modelled as a zero-sum authentication game. Then a reinforcement-learning is applied to choose the optimal test threshold for improving the accuracy of authentication. However, such a method requires a fixed attack probability. In [18], Bhunia *et al*. propose a defence mechanism of random learning to solve the interference. Ota *et al*. [19] applied reinforcement learning to wireless networks. The event detection algorithm proposed in [20] that speeds up the event detection in wireless sensor networks and actor networks, in which reinforcement learning technology helps each actor move towards the event. In [21], a slope authentication at physical-layer is proposed that overcomes the drawbacks of additional bandwidth requirement and the entire data message corruption in time-division multiplexed tag and authentication with superimposed tag respectively. Xiaoying *et al*. [22] proposed a PHY-layer security authentication scheme that takes advantage of channel randomness to detect spoofing attacks in wireless networks. In [23], a physical layer cheat detector based on Q-learning is proposed, and a physical layer authentication game is constructed. Liang *et al*. [24] proposed a cheat detector based on Dyna-Q to improve the authentication speed. Recently, Liang *et al*. [25] proposed a logistic regression-based authentication to remove the assumption on the known channel model, and thus applicable to more generic wireless networks. In [26], a PHY-layer spoofing detection algorithm for multiple-input multiple-output systems based on Q-learning is proposed, in which the receiver applies the reinforcement learning technique to achieve the optimal test threshold via trials in a dynamic game without knowing the system parameters, such as the channel time variation and spoofing cost. In [27], Pan *et al*. proposed a threshold-free PHY-authentication method based on machine learning, which replaces the traditional threshold-based decision-making with more adaptive classification. Xiaozhen *et al*. [28] proposed a reinforcement learning-based physical authentication scheme to resist rogue edge attackers whose goal is to send spoofing signals to attack vehicle ad hoc networks. While the above-stated techniques are machine learning-based strategies, they are not light-weighted, i.e. these machine learning-based strategies can detect the spoofing attack, but their complexities are too high to apply to users with limited computing resources in CRNs.

Our contribution. In this study, we propose a light-weight physical layer authentication algorithm [rollout partially observable Markov decision process-based algorithm (RoPOMDP)]. In general, RoPOMDP consists of a v-support vector regression (v-

SVR)-based spoofing packet prediction and a rollout-based hypothesis test for physical layer authentication. Thus, the detail of our contribution is listed as follows:

(i) To detect spoofing packets in CRNs, we first formulate the PHY-authentication as a zero-sum game, based on which a hypothesis test upon channel vectors is developed. Then, we obtain the gains of both spoofers and receivers based on Bayesian risks for the PHY-authentication game, in which the spoofing attack probability is predicted by a non-linear function approximation utilising v–SVR. Owing to the reason that both channel model and spoofing model are unknown by the receiver, a RoPOMDP is employed to estimate the optimal threshold for the test statistic in the PHY-authentication game based on observation of the radio environment.
(ii) The theoretical analysis and simulations indicate that (i) RoPOMDP improves the spoofing detection accuracy and (ii) the complexity of RoPOMDP is lower than contemporary algorithms, i.e. the Q-learning-based PHY-authentication algorithm.

The rest of the paper is organised as follows. The system model is introduced in Section 2. The strategies are elaborated in Section 3. Section 4 gives the theoretical analysis of the complexity of RoPOMDP, with the results of validation experiments presented. Section 5 gives the concluding remarks of this paper.

2 System model

In this study, a CRN that consists of a receiver and N transmitters is considered. The MAC address of the ith transmitter (node) is denoted by MAC$_i$. Accordingly, the node of MAC$_i$ is denoted by n_i, while the spoofing one that claims to have MAC$_i$ is denoted by n_i'. Each node is assumed to have the information of the centre frequency f_0 and bandwidth W. The channel response is sampled at M different tones in frequency $f \in [f_0 - W/2, f_0 + W/2]$. The RSS indicator (RSSI) vector of the kth packet claimed to have the MAC address i is denoted by $r_i^k = [r_{i,m}^k]$, where $r_{i,m}^k$ is the RSSI of the mth tone for the ith transmitter. Owing to the fact that potential spoofers can impersonate another node with a fake MAC address, the receiver requires a PHY-authentication technique, which usually adopts the hypothesis test on channel response to detect spoofing attacks, i.e. an elaborately designed hypothesis test can determine whether a packet with a channel vector $\boldsymbol{r}_i^k$ is sent by the node with MAC address i. Based on RSSI, a hypothesis test can be constructed as follows.

Let null hypothesis H_0 indicate that the node with MAC address i sends the kth packet, while the alternative hypothesis H_1 represents that the spoofer sends the kth packet with MAC address i. Thus, we have

$$H_0 \colon g(\boldsymbol{r}_i^k) = i \tag{1}$$

$$H_1 \colon g(\boldsymbol{r}_i^k) \neq i, \tag{2}$$

where $g(\boldsymbol{r}_i^k)$ denotes the MAC address of the node that sends the kth packet with channel vector $\boldsymbol{r}_i^k$. The probability of a legitimate packet being considered as a spoofing one, which is the false alarm rate α, is given by

$$\alpha = \Pr(H_1 | H_0) \tag{3}$$

Similarly, the miss detection rate β is the probability that a spoofing packet is classified as a legitimate one, which is given by

$$\beta = \Pr(H_0 | H_1) \tag{4}$$

Then, the probability of accepting a legitimate packet is $\Pr(H_0 | H_0) = 1 - \alpha$, while rejecting a spoofing packet is denoted by $\Pr(H_1 | H_1) = 1 - \beta$. In this study, we consider the spoofing detector w.r.t channel frequency responses instead of RSSIs. In this scenario, the receiver obtains a channel vector denoted by $\boldsymbol{H}_i^k$ from transmitter i for the kth packet and then stores the channel records denoted by $\hat{H}_i^k$ for transmitter i. Note that interferences from other sources are considered as noises. Thus, $\boldsymbol{H}_i^k$ and $\hat{H}_i^k$ are noisy versions of the true channel response $\boldsymbol{H}_i^k$ and the measured one $\hat{\boldsymbol{H}}_i^k$, respectively, i.e.

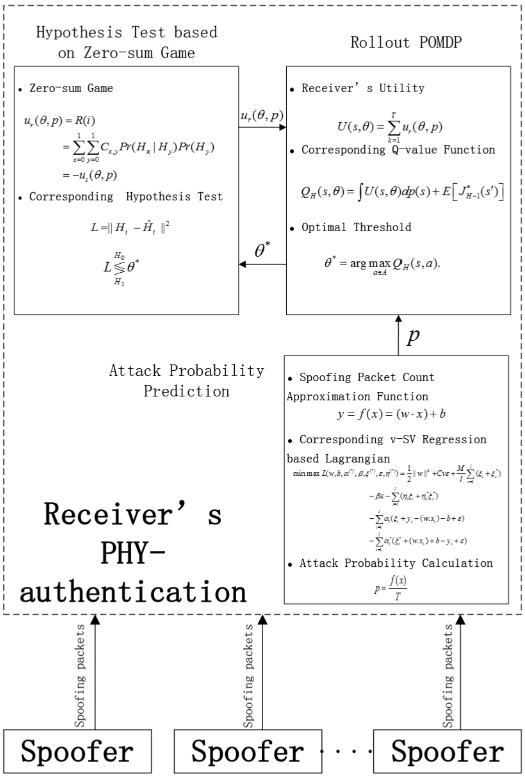
Hypothesis Test based on Zero-sum Game
Rollout POMDP
• Zero-sum Game
$u_r(\theta,p) = R(i) = \sum_{x=0}^{1}\sum_{y=0}^{1} C_{x,y} Pr(H_x | H_y) Pr(H_y) = -u_s(\theta,p)$
• Corresponding Hypothesis Test
$L = \| H_i - \hat{H}_i \|^2$
$L \underset{H_1}{\overset{H_0}{\lessgtr}} \theta^*$
$u_r(\theta,p)$
θ^*
• Receiver' s Utility
$U(s,\theta) = \sum_{k=1}^{T} u_r(\theta,p)$
• Corresponding Q-value Function
$Q_H(s,\theta) = \int U(s,\theta) dp(s) + E\left[J_{H-1}^*(s')\right]$
• Optimal Threshold
$\theta^* = \arg\max_{a \in A} Q_H(s,a).$
p
Attack Probability Prediction
• Spoofing Packet Count Approximation Function
$y = f(x) = (w \cdot x) + b$
• Corresponding v-SV Regression based Lagrangian
$\min\max L(w,b,\alpha^{(*)},\beta,\xi^{(*)},\varepsilon,\eta^{(*)}) = \frac{1}{2}\|w\|^2 + Cv\varepsilon + \frac{M}{l}\sum_{i=1}^{l}(\xi_i + \xi_i^*) - \beta\varepsilon - \sum_{i=1}^{l}(\eta_i\xi_i + \eta_i^*\xi_i^*) - \sum_{i=1}^{l}\alpha_i(\xi_i + y_i - (w.x_i) - b + \varepsilon) - \sum_{i=1}^{l}\alpha_i^*(\xi_i^* + (w.x_i) + b - y_i + \varepsilon)$
• Attack Probability Calculation
$p = \frac{f(x)}{T}$
Receiver' s PHY-authentication
Spoofing packets
Spoofing packets
Spoofing packets
Spoofer
Spoofer
· · ·
Spoofer

Fig. 1 *Flowchart of proposed strategy*

$$\boldsymbol{H}_i^k = \boldsymbol{H}_i^k + N_1 \tag{5}$$

$$\hat{H}_i^k = \hat{\boldsymbol{H}}_i^k + N_2, \tag{6}$$

where N_1 and N_2 are i.i.d complex Gaussian noise samples $C(0, \sigma^2)$. In addition, we also assume small channel time variations, small estimation errors, zero phase drift between channel measurements and frequency-selective Rayleigh channel models. Thus, the generalised likelihood ratio test L chosen is given by

$$L = \| \boldsymbol{H}_i^k - \hat{H}_i^k \|^2 . \tag{7}$$

Thus, the hypothesis test performed by the receiver is given by

$$L \underset{H_1}{\overset{H_0}{\lessgtr}} \theta . \tag{8}$$

In this study, we focus on finding the optimal threshold θ^* for (8).

3 Proposed PHY-authentication strategy

The proposed strategy, RoPOMDP, consists of (i) an attack probability prediction and (ii) a hypothesis test based on the zero-sum game, both of which collaborate to detect spoofing packets. The flowchart of RoPOMDP is given in Fig. 1.

3.1 Attack probability prediction

It is essential to predict the number of spoofing packets by a specific timeslot so as to calculate the attack probability p [23], in which the v-SVR is employed. In general, the number of spoofing packets y_i at the timeslot t_i among T packets is given in a linear function as

$$y_i = f(x_i) = (w \cdot x_i) + b, \tag{9}$$

where $w = \{w_1, w_2, \dots, w_d\}$ and $x_i \in R^d$ represents the state at the timeslot t_i that consists of historical observations. The v-SVR is trained based on samples $\{x_i, y_i\}$ by minimising the regularised risk function

$$\frac{1}{2} \parallel w \parallel^2 + \mathscr{C} \cdot \frac{1}{l} \sum_{i=1}^{l} \left| y_i - f(x_i) \right|_{\varepsilon} \tag{10}$$

to obtain the optimal w, where $\mathscr{C}$ is a constant determining the trade-off between minimising training errors and minimising the model complexity term $\parallel w \parallel^2$. To estimate function (9) from training samples, we proceed as follows [29]. At each point x_i, we allow an error of ε. Everything above ε is captured in slack variables $\xi_i^{(*)}$ ($(*)$ implies both the variables with and without asterisks)), which are penalised in the objective function via a regularisation constant M, to be chosen a priori. The size of ε is traded off against model complexity and slack variables via a constant v as

$$\min.\, \tau(w, \xi^{(*)}, \varepsilon) = \frac{1}{2} \parallel w \parallel^2 + M[v\varepsilon + \frac{1}{l} \sum_{i=1}^{l} (\xi_i + \xi_i^*)] \tag{11}$$

$$\text{s.t. } (wx_i + b) - y_i \le \varepsilon + \xi_i, \tag{12}$$

$$y_i - (wx_i + b) \le \varepsilon + \xi_i^*, \tag{13}$$

$$\xi_i^* \ge 0, \varepsilon \ge 0. \tag{14}$$

For the constraints, we introduce multipliers $\alpha_i, \alpha_i^{(*)}, \eta_i^{(*)}, \beta \ge 0$, and obtain the Lagrangian

$$\begin{aligned} L(w, b, \alpha^{(*)}, \beta, \xi^{(*)}, \varepsilon, \eta^{(*)}) = {} & \frac{1}{2} \parallel w \parallel^2 + Mv\varepsilon + \frac{M}{l} \sum_{i=1}^{l} (\xi_i + \xi_i^*) \\ & - \beta\varepsilon - \sum_{i=1}^{l} (\eta_i \xi_i + \eta_i^* \xi_i^*) \\ & - \sum_{i=1}^{l} \alpha_i (\xi_i + y_i - (w \cdot x_i) - b + \varepsilon) \\ & - \sum_{i=1}^{l} \alpha_i^* (\xi_i^* + (w \cdot x_i) + b - y_i + \varepsilon) \end{aligned} \tag{15}$$

In fact, minimising (11) equals to

$$\min_{w, \varepsilon, b, \xi_i^{(*)}} \max_{\alpha_i^{(*)}, \beta, \eta_i^{(*)}} L(w, b, \alpha^{(*)}, \beta, \xi^{(*)}, \varepsilon, \eta^{(*)}), \tag{16}$$

the dual of which is given by

$$\max_{\alpha_i^{(*)}, \beta, \eta_i^{(*)}} \min_{w, \varepsilon, b, \xi_i^{(*)}} L(w, b, \alpha^{(*)}, \beta, \xi^{(*)}, \varepsilon, \eta^{(*)}), \tag{17}$$

Thus, (15) should be minimised over the primal variables $w, \varepsilon, b, \xi_i^{(*)}$. Let the derivatives with respect to the primal variables equal to zero yields:

$$w = \sum_i (\alpha_i^* - \alpha_i) x_i \tag{18}$$

$$M \cdot v - \sum_i (\alpha_i + \alpha_i^*) - \beta = 0 \tag{19}$$

$$\sum_{i=1}^{l} (\alpha_i^* - \alpha_i) = 0 \tag{20}$$

$$\frac{M}{l} - \alpha_i^{(*)} - \eta_i^{(*)} = 0 \tag{21}$$

Substituting (18)–(21) into (15), we then rewrite the constraints to describe the *v-SVR* optimisation problem

$$\begin{aligned} \max.\; W(\alpha^{(*)}) = {} & \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) y_i \\ & - \frac{1}{2} \sum_{i,j=1}^{l} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) k(x_i, x_j), \end{aligned} \tag{22}$$

$$\text{s.t. } \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) = 0, \tag{23}$$

$$\alpha_i^{(*)} \in [0, \frac{M}{l}], \tag{24}$$

$$\sum_{i=1}^{l} (\alpha_i^* + \alpha_i) \le M \cdot v, \tag{25}$$

where kernel $k(x, y) = \exp(- \parallel x - y \parallel^2 / 2\sigma^2)$ is substituted for the dot product in some feature space-related to input space via a non-linear map Φ as

$$k(x, y) = (\Phi(x) \cdot \Phi(y)). \tag{26}$$

Eventually, the *v-SVR* can be deduced as

$$y_i = \sum_{i=1}^{l} (\alpha_i^* - \alpha_i) k(x_i, x) + b. \tag{27}$$

In this study, we assume the receiver receives T packets in each timeslot. Thus, the probability that a received packet is a spoofing one at timeslot t_i is estimated as

$$p_i = \frac{f(x_i)}{T}. \tag{28}$$

3.2 Hypothesis test based on zero-sum game

We formulate the PHY-authentication as a zero-sum game to discover the optimal threshold θ^* for the test statistic L given in (7). Similar to the literature [23], let $C_{x,y}$ denote the payoff for the receiver choosing hypothesis H_x in the case of hypothesis H_y, $x, y \in \{0, 1\}$. We denote the gain of accepting a packet sent by the node with MAC address i as g_i^l, while the gain of rejecting a spoofing packet claimed to have MAC address i is denoted by g_i^s. Let γ_i denote the cost of rejecting a legitimate packet sent by the node with MAC address i. Thus, we have

$$C_{0,0} = g_i^l - G - C \tag{29}$$

$$C_{0,1} = -G - C \tag{30}$$

$$C_{1,0} = -\gamma_i - C \tag{31}$$

$$C_{1,1} = g_i^s - C. \tag{32}$$

where C and G represent costs of physical-layer and high-layer authentications, respectively. Accordingly, the Bayesian risk $R(i)$ of the spoofing detection for a packet from transmitter i is then given by

$$
\begin{aligned}
R(i) &= \sum_{x=0}^{1}\sum_{y=0}^{1} C_{x,y}\Pr(H_x \mid H_y)Pr(H_y) \\
&= (g_i^l - G - C)(1-\alpha)(1-p) - (G+C)\beta p \\
&\quad -(\gamma_i + C)\alpha(1-p) + (g_i^s - C)(1-\beta)p,
\end{aligned} \tag{33}
$$

where p is estimated by v-SVR as we described in the previous section. Since the PHY-authentication is formulated as a zero-sum game, we have $u_r(\theta, p) = R(i) = -u_s(\theta, p)$, where $u_r(\theta, p)$ and $u_s(\theta, p)$ denote the utility of the receiver and spoofer, respectively. Note that the Nash equilibrium (NE) of a game consists of the best response strategies such that no player can increase its utility by unilaterally choosing a different strategy [23]. Let the NE of the PHY-authentication game is denoted by (θ^*, y^*). This indicates the receiver chooses the test threshold θ^* to maximise his/her utility $u_r(\theta^*, p^*)$ in the spoofing detection, while the spoofer aims to maximise his/her utility $u_s(\theta^*, p^*)$. Thus, the optimal threshold θ^* satisfies

$$
\theta^* = \arg\max_{\theta \geq 0} u_r(\theta, p). \tag{34}
$$

However, θ^* depends on the channel model and spoofing model, which are not always known by the receiver. To solve this problem, we apply a partially observable Markov decision process (POMDP) to calculate θ^* based on observations of the radio environment. First, we denote the state observed by the receiver in the nth time slot as $s_n = [\alpha_{n-1}, \beta_{n-1}] \in S$, which consists of the false alarm rate and miss detection rate of authentication in the previous time slot, while S denotes the state set. Accordingly, the action set denoted by $A = [\theta_l]_{1 \leq l \leq K}$ consists of K level thresholds, in which each action $a_n \in A$ is chosen based on state s_n. In addition, each receiver is assumed to obtain T packets in each time slot. Thus, the utility after receiving T packets at state s_n with action θ_n, denoted by $U(s_n, \theta_n)$, is then given by

$$
U(s_n, \theta_n) = \sum_{k=(n-1)T+1}^{nT} u_r^k(\theta_n, p_n). \tag{35}
$$

Note that several Q-value approximation methods [30, 31] have been proposed for large state-space Markov decision processes. In this study, we consider the policy rollout [30]. In general, the policy-rollout method estimates the Q-value for each belief state and each action by averaging the evaluated accumulated costs from several Monte–Carlo simulation runs using a given base policy rather than calculating the expectation of Q-value over the entire state space. To apply the rollout framework to POMDP, we first give the expected utility over a horizon of H steps as

$$
J_H(s_0) = E\left(\sum_{k=0}^{H-1} \int U(s_k, \theta)\mathrm{d}p(s_k)\right), \tag{36}
$$

while the optimal value of which is denoted by $J_H^*(s_0)$. Accordingly, the Q-value function of state s is then given by

$$
Q_H(s, \theta) = \int U(s, \theta)\mathrm{d}p(s) + E[J_{H-1}^*(s')], \tag{37}
$$

where $J_{H-1}^*(s')$ is the optimal value over $H-1$ steps starting at the next state s'. Thus, the upper bound on the $Q_H(s, \theta)$ can be obtained by maximising

$$
\hat{Q}_H(s, \theta) = \frac{1}{N}\sum_{i=1}^{N}\left\{g(s^{(i)}, \theta) + \hat{J}_{H-1}(s^{'(i)})\right\}, \tag{38}
$$

where $s^{(i)}$ and $s^{'(i)}$ represent the sample for state s and s' during the Monte–Carlo simulation. Accordingly, the optimal action, which is the optimal threshold θ^*, at timeslot k is chosen as

$$
\theta_k^* = \arg\max_{\theta \in A} Q_H(s_k, \theta). \tag{39}
$$

It is worth mentioning that $p(s_k)$ can be obtained by the particle filter [32] and the parameter initiation of which can be obtained from real experiences. For example, the receiver updates the experience records for each state–action pair, which consists of the occurrence count vector $\boldsymbol{\tau}$, the occurrence count vector of the next state $\boldsymbol{\tau}'$, and the state transition probability Ψ. The count vector $\boldsymbol{\tau}'$ for the current experience increases by 1, i.e.

$$
\boldsymbol{\tau}'(s_n, a_n, s_{n+1}) \leftarrow \boldsymbol{\tau}'(s_n, a_n, s_{n+1}) + 1. \tag{40}
$$

The occurrence count vector $\boldsymbol{\tau}$ consists of all the possible realisations of s_{n+1} as

$$
\boldsymbol{\tau}(s_n, a_n) \leftarrow \sum_{s' \in S} \boldsymbol{\tau}'(s_n, a_n, s'). \tag{41}
$$

The state transition probability from the current state s_n to the next state s_{n+1} by action a_n maps state action pair (s_n, a_n) to the distribution of state s_{n+1}. Accordingly, the transition probability, which is denoted by $\Psi(s_n, a_n, s_{n+1})$, is given as

$$
\Psi(s_n, x_n, s_{n+1}) \leftarrow \frac{\boldsymbol{\tau}'(s_n, x_n, s_{n+1})}{\boldsymbol{\tau}(s_n, x_n)}. \tag{42}
$$

Once the optimal threshold θ^* is obtained, one could employ (8) to determine whether the packet is a spoofing one or not in the physical layer. If both the PHY-layer and higher-layer authentications accept, then the packet is accepted. Also, the reference channel vector $\hat{\boldsymbol{H}}_i$ is updated as $\hat{\boldsymbol{H}}_i^k = \boldsymbol{H}_i^k$ once packets of node i is accepted; otherwise, $\hat{\boldsymbol{H}}_i^k = \hat{\boldsymbol{H}}_i^{k-1}$. Then, we summarise the RoPOMDP in Algorithm 1 (see Fig. 2).

4 Performance analysis

4.1 Theoretical analysis

In this section, we are going to prove the advantage of RoPOMDP over Q-learning-based strategies [14, 18–20, 23, 24] denoted by Q_A in terms of complexity.

Theorem 1: The complexity of RoPOMDP is lower than Q_A.

Proof: RoPOMDP consists of three efficient algorithms which are the spoofing packet prediction, the hypothesis test construction, and the PHY-authentication game. Since the spoofing packets differ by time, the non-linear function approximation utilising v-SVR is employed to accurately estimate the attack probability p at a

```
Initialize: $Q(s, \theta) = 0, \forall\theta \in A$.
  for n=1,2,3,..., $|S|$ do
    for i=1,2,3,..., H do (for each time slot)
      Select a threshold $\theta_n$ via (39).
      for k=1,2,...,T do (for each packet)
        Authenticate the $k$-th packet:
          Extract $H_i$ and $\hat{H}_i$ from the $k$-th packet with MAC
  address $i$.
          Calculate test statistic $L$ via (7).
        if $L \leq \theta_n$ and the higher-layer authentication is passed
  then
          set $\hat{H}_i^k = H_i^k$ and accept this packet
        else set $\hat{H}_i^k = H_i^{k-1}$ and send a spoofing alarm.
        end if
      end for
      Observe next state $s_{n+1}$ and Utility $U(s_n, \theta_n)$.
      Update $Q(s_n, \theta_n)$ via (38).
    end for
  end for
```

Fig. 2 *Algorithm 1: RoPOMDP-based spoofing detection algorithm*

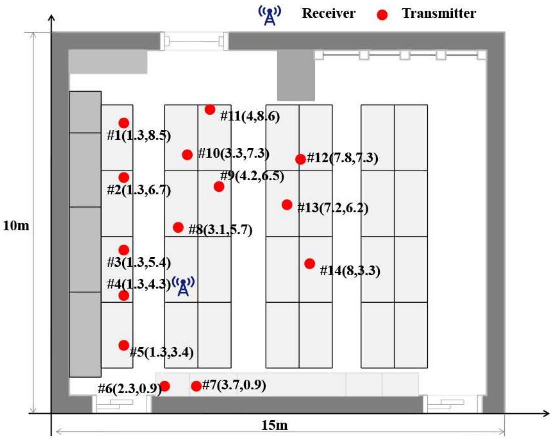

Fig. 3 *Network topology of the experiments in a $10 \times 15 \times 3m^3$ office room, consisting of 14 transmitters and a receiver*

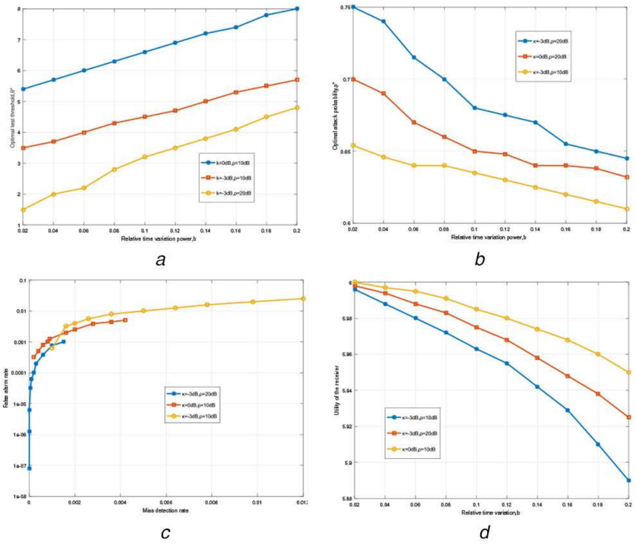

Fig. 4 *Performance of the spoofing detection game at the NE in*
(a) Optimal test threshold θ^*, ***(b)*** Optimal attack probability p^*, ***(c)*** False alarm rate in the detection, ***(d)*** Utility of the receiver

Table 1 Experiment setup

Par.	Des.	Val.
C	physical-layer authentication cost	1
G	high-layer authentication cost	3
g_l^i	gain of accepting a legitimate packet	10
r^i	cost of rejecting a legitimate packet	20
g_i^s	gain of rejecting a spoofing packet	10
f_0	information of center frequency	2.4 GHz
W	channel bandwidth	[50, 200] MHz
M	number of different tones	5
θ	threshold	[0.01, 0.14]
b	relative time variation power	[0.02, 0.2]
κ	channel gain ratio	$(0, -3)$dB
ρ	SINR of packets from n_i	(10, 20) dB

given time. The PHY-authentication game is developed utilising a policy RoPOMDP to calculate the optimal threshold for the hypothesis test to detect spoofing packets. The complexity of the RoPOMDP depends on the number of state $|S|$, the step number N' and Monte–Carlo simulation runs N, which is $O(|S|)$. Compared with Q_A, the complexity of which is at least $O(|A \parallel S|)$, RoPOMDP is more light-weighted.□

4.2 Validation experiment

4.2.1 Parameter setup: Experiments were implemented on universal software radio peripherals (USRPs), each of which equipped with a single antenna is used to operate using the IEEE 802.11a/g standard working at 2.4 GHz with a bandwidth within $[50, 200]$ MHz and performed in an indoor environment (see Fig. 3) to analyse the performance of the proposed spoofing detection scheme RoPOMDP. Owing to the generalised likelihood ratio test (7), the corresponding false alarm rate and the missed detection rate of the hypotheses test as in the spoofing detection [33] are given by

$$\alpha(\theta) = 1 - F_{\chi^2_{2M}}\left(\frac{2\theta\rho}{2\sigma^2 + b\rho\sigma^2}\right) \tag{43}$$

$$\beta(\theta) = 1 - F_{\chi^2_{2M}}\left(\frac{2\theta\rho}{2\sigma^2 + (1+\kappa)\rho\sigma^2}\right) \tag{44}$$

where σ^2 is the average power gain from the legitimate transmitter at the receiver, κ represents the ratio of the channel gain of the spoofer to that of the legitimate transmitter, $F_{\chi^2_{2M}}(.)$ denotes the cumulative distribution function of the chi-square distribution with $2M$ degrees of freedom, ρ is the signal-to-interference-plus-noise ratio (SINR) of the packets sent by the legitimate transmitter, and b is the relative change in the channel gain due to environmental changes.

First, we give the performance evaluation of the spoofing detection game at NE (see Fig. 4) of the proposed strategy in optimal test threshold θ^*, optimal attack probability p^*, false alarm rate in the detection and utility of the receiver, respectively. Then, performance comparisons are implemented between the proposed rollout-based strategy, the fixed threshold based PHY-authentication strategy (i.e. making decisions based on (8) w.r.t a fixed threshold), and the Q-learning based PHY-authentication strategy [24] in terms of average error rates and utility first. The experiment parameters are listed in Table 1. Note that the action set is chosen based on the assumption that the attack probability is chosen as $p = 0.75$.

4.2.2 Experiment results: In Fig. 4*a*, it is clear that the optimal test threshold increases with the channel time variation b to avoid rejecting. As shown in Fig. 4*b*, a spoofer tends to fail with large channel variations, as the test statistic increases with b in the presence of spoofer. Both false alarm rate and missed detection rate at the NE increase with b as shown in Fig. 4*c*. This is because it is a challenge to distinguish transmitters according to their channel states under significant radio environmental changes. In addition, both false alarm rate and missed detection rate decrease with SINR, when the channel estimation error at the receiver decreases with SINR. It is clear that the spoofing detection still achieves good performance with $\alpha = 0.02$ and $\beta = 0.01$ as $\kappa = 3\,\text{dB}$, $b = 0.2$, and $\rho = 10\,\text{dB}$. Fig. 4*d* shows that the detection accuracy increases as the utility of the receiver increases with κ and ρ. For example, if $b = 0.2$ and $\rho = 10\,\text{dB}$, the utility of the receiver for $\kappa = 0\,\text{dB}$ increases to 5.95 from 5.89 for $\kappa = -3\,\text{dB}$.

Figs. 5 and 6 show the performance comparison between our proposed strategy and baseline approaches with nodes 2, 12, and 14 pretending to be node 10 in a topology as shown in Fig. 3. In Fig. 5, both false alarm rate and miss detection rate for all schemes decrease with the bandwidth. It is clear that if the bandwidth is 200 MHz, the false alarm rate and miss detection rate of the proposed strategy approach 0.01. The false alarm rate and miss detection rate of the fixed threshold strategy are up to 0.041 and 0.0296, respectively, as the bandwidth is 150 MHz compared with that of the proposed strategy of only 0.0005 and 0.0007. It is clear that the proposed rollout-based strategy performs better than others. As shown in Fig. 6, the average utility of the receiver increases at least by 0.0323 if the bandwidth is 150 MHz. Obviously, the proposed rollout-based strategy outperforms baseline approaches. Note that a higher utility suggests a better threshold referring to (39). That

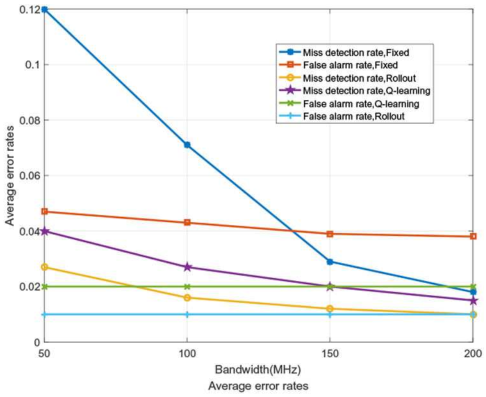

Fig. 5 *Performance comparison in average error rates while varying bandwidth*

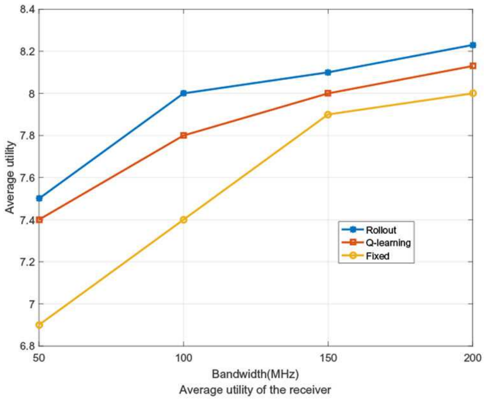

Fig. 6 *Performance comparison in average utility while varying bandwidth*

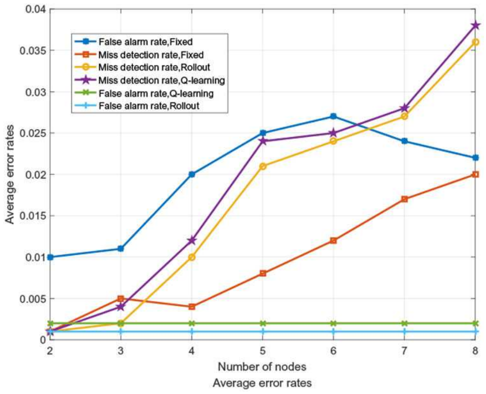

Fig. 7 *Performance comparison in average error rates while varying number of nodes*

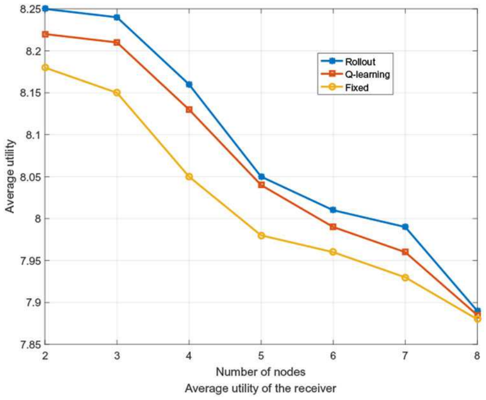

Fig. 8 *Performance comparison in average utility while varying number of nodes*

explains why the proposed strategy has a lower miss detection rate and a lower false alarm rate as well.

It is obvious that the miss detection rate increases with the number of nodes (see Fig. 7) except the rollout-based strategy. The reason behind that is the optimal threshold θ^* could not match the optimal one for a specific spoofer. If the cost of accepting a spoofing packet is smaller than that to reject a legitimate packet, the receiver tends to restrict the false alarm in the learning process. Although the Q-learning-based strategy performs better than the fixed threshold one, it is still no match for the proposed strategy. As shown in Fig. 8, the average utility of the receiver for all strategies decreases with the number of nodes as the detection error rate increases. The reason behind that is there could be more spoofers involved as the number of nodes increases. Although the proposed strategy outperforms baseline approaches, the optimal threshold is hard to obtain such that the utility drops.

5 Conclusions

Owing to the wireless and cognitive nature of CRNs, spoofers intend to send false perception reports with fake MAC addresses for impersonating other honest CUs. To solve this problem, a lightweight physical-layer authentication algorithm RoPOMDP is developed based on channel frequency responses. Be specific, RoPOMDP first formulates the PHY-authentication as a zero-sum game, based on which a hypothesis test upon channel vectors is developed. Accordingly, the gains of both spoofers and receivers of the game are obtained w.r.t Bayesian risks, where the spoofing attack probability is predicted by applying a *v*-*SVR*-based approximation function. Then, a RoPOMDP is designed to estimate the optimal threshold for the test statistic in the PHY-authentication game. The theoretical analysis and simulations show that RoPOMDP outperforms the *Q*-learning-based authentication algorithms in both spoofing detection accuracy and complexity.

6 Acknowledgments

The authors would like to thank the National Science Foundation of China (nos. U1905211,61771140, 61702103, and 61702100) and the Fujian Province University Industry Cooperation of Major Science and Technology Project (no. 2017H6005).

7 References

[1] Mitola, J.I., Maguire, G.Q.: 'Cognitive radio: making software radios more personal', *IEEE Pers. Commun.*, 1999, **6**, (4), pp. 13–18

[2] Nasnin, F., Islam, M.N., Chakrabarty, A.: 'Security analysis on cognitive radio network', 2019

[3] Zeng, K., Govindan, K., Mohapatra, P.: 'Non-cryptographic authentication and identification in wireless networks [security and privacy in emerging wireless networks]', *IEEE Wirel. Commun.*, 2010, **17**, (5), pp. 56–62

[4] Jie, Y., Yingying, C., Trappe, W., *et al.*: 'Detection and localization of multiple spoofing attackers in wireless networks', *IEEE Trans. Parallel Distrib. Syst.*, 2013, **24**, (1), pp. 44–58

[5] Kalamandeen, A., Scannell, A., Lara, E., *et al.*: 'Ensemble: cooperative proximity-based authentication'. Proc. Int. Conf. on Mobile Systems, Applications and Services, San Francisco, CA, USA, 2010, pp. 331–344

[6] Jinliang, L., Liang, X., Guolong, L.: 'Active authentication with reinforcement learning based on ambient radio signals', *Multimedia Tools Appl.*, 2014, **76**, (3), pp. 3979–3998

[7] Xiaofu, W., Zhen, Y.: 'Physical-layer authentication for multi-carrier transmission', *IEEE Commun. Lett.*, 2015, **19**, (1), pp. 74–77

[8] Liu, H., Wang, Y., Liu, J., *et al.*: 'Practical user authentication leveraging channel state information(CSI)'. Proc. ACM Symp. on Information, Computer and Communication Security, Kyoto, Japan, 2014, pp. 389–400

[9] Tugnait, J.: 'Wireless user authentication via comparison of power spectral densities', *IEEE J. Selected Areas Commun.*, 2013, **31**, (9), pp. 1791–1802

[10] Jiang, Z., Zhao, J., Li, X., *et al.*: 'Rejecting the attack: source authentication for WIFI management frames using CSI information'. Proc. IEEE Int. Conf. on Computer Communications (INFOCOM), Turin, Italy, 2013, pp. 2544–2552

[11] Liang, X., Reznik, A., Trappe, W., *et al.*: 'PHY-authentication protocol for spoofing detection in wireless networks'. Global Telecommunications Conf., Miami, FL, USA, 2010

[12] Hou, W., Wang, X., Chouinard, J.Y., *et al.*: 'Physical layer authentication for mobile systems with time-varying carrier frequency offsets', *IEEE Trans. Commun.*, 2014, **62**, (5), pp. 1658–1667

[13] Xiao, L., Chen, Y., Lin, W.S., *et al.*: 'Indirect reciprocity security game for large-scale wireless networks', *IEEE Trans. Inf. Forensics Sec.*, 2012, **7**, (4), pp. 1368–1380

[14] Wu, Y., Wang, B., Liu, K, *et al.*: 'Anti-jamming games in multi-channel cognitive radio networks', *IEEE J. Sel. Areas Commun.*, 2012, **30**, (1), pp. 4–15

[15] Hu, P., Li, H., Fu, H., *et al.*: 'Dynamic defense strategy against advanced persistent threat with insiders'. IEEE INFOCOM 2015–IEEE Conf. on Computer Communications, Hong Kong, China, 2015

[16] He, F., Li, X., Xianbin, W.: 'Coordinated multiple relays based physical layer security improvement: a single leader-multiple followers Stackelberg game scheme', *IEEE Trans. Inf. Forensics Sec.*, 2018, **13**, (1), pp. 197–209

[17] He, F., Xianbin, W., Lajos, H.: 'Learning-aided physical layer authentication as an intelligent process', *IEEE Trans. Commun.*, 2019, **67**, (3), pp. 2260–2273

[18] Bhunia, S., Sengupta, S., Vázquez-Abad, F.: 'CR-Honeynet: a learning and decoy based sustenance mechanism against jamming attack in CRN'. Military Communications Conf., Baltimore, MD, USA, 2014

[19] Ota, K., Dong, M., Cheng, Z., *et al.*: 'ORACLE: mobility control in wireless sensor and actor networks', *Comput. Commun.*, 2012, **35**, (9), pp. 1029–1037

[20] Dong, M., Ota, K., Li, H., *et al.*: 'RENDEZVOUS: towards fast event detecting in wireless sensor and actor networks', *Computing*, 2014, **96**, (10), pp. 995–1010

[21] Ning, X., Changsheng, C.: 'Slope authentication at the physical layer', *IEEE Trans. Inf. Forensics Sec.*, 2018, **13**, (6), pp. 1579–1594

[22] Xiaoying, Q., Ting, J., Sheng, W., *et al.*: 'Physical layer authentication enhancement using a Gaussian mixture model', *IEEE Access*, 2018, **6**, pp. 53583–53592

[23] Liang, X., Yan, L., Guolong, L., *et al.*: 'Spoofing detection with reinforcement learning in wireless networks'. GLOBECOM 2015–2015 IEEE Global Communications Conf., San Diego, CA, USA, 2015

[24] Liang, X., Yan, L., Guoan, H., *et al.*: 'PHY-layer spoofing detection with reinforcement learning in wireless networks', *IEEE Trans. Veh. Technol.*, 2016, **65**, (12), pp. 10037–10047

[25] Liang, X., Xiaoyue, W., Zhu, H.: 'PHY-layer authentication with multiple landmarks with reduced overhead', *IEEE Trans. Wirel. Commun.*, 2017, **17**, (3), pp. 1676–1687

[26] Liang, X., Tianhua, C., Guoan, H., *et al.*: 'Game theoretic study on channel-based authentication in MIMO systems', *IEEE Trans. Veh. Technol.*, 2017, **66**, (8), pp. 7474–7484

[27] Pan, F., Zhibo, P., Hong, W., *et al.*: 'Threshold-free physical layer authentication based on machine learning for industrial wireless CPS', *IEEE Trans. Ind. Inf.*, 2019, **15**, (12), pp. 6481–6491

[28] Xiaozhen, L., Liang, X., Tangwei, X., *et al.*: 'Reinforcement learning based PHY authentication for VANETs', *IEEE Trans. Veh. Technol.*, 2020, **69**, (3), pp. 3068–3079

[29] Schölkopf, B., Bartlett, P., Smola, A., *et al.*: 'Support vector regression with automatic accuracy control'. Int. Conf. on Artificial Neural Networks, London, 1998, pp. 111–116

[30] Bertsekas, D.P., Castanon, D.A.: 'Rollout algorithms for stochastic scheduling problems', *J. Heuristics*, 1999, **5**, (1), pp. 89–108

[31] Wu, G., Chong, E.K.P., Givan, R.: 'Burst-level congestion control using hindsight optimization', *IEEE Trans. Autom. Control*, 2002, **47**, (6), pp. 979–991

[32] Doucet, A., de Freitas, N., Gordon, G.: '*Sequential Monte Carlo methods in practice*' (Springer-Verlag, New York, 2001)

[33] Liang, X., Greenstein, L., Mandayam, N., *et al.*: 'Fingerprints in the ether: using the physical layer for wireless authentication'. Proc. IEEE ICC, Beijing, China, 2007, pp. 4646–4651