PHY-Layer Spoofing Detection With Reinforcement Learning in Wireless Networks

Liang Xiao, *Senior Member, IEEE*, Yan Li, *Student Member, IEEE*, Guoan Han, *Student Member, IEEE*, Guolong Liu, *Student Member, IEEE*, and Weihua Zhuang, *Fellow, IEEE*

***Abstract*—In this paper, we investigate the PHY-layer authentication that exploits radio channel information (such as received signal strength indicators) to detect spoofing attacks in wireless networks. The interactions between a legitimate receiver and spoofers are formulated as a zero-sum authentication game. The receiver chooses the test threshold in the hypothesis test to maximize its utility based on the Bayesian risk in the spoofing detection, whereas the spoofers determine their attack frequencies to minimize the utility of the receiver. The Nash equilibrium of the static authentication game is derived, and its uniqueness is discussed. We also investigate a repeated PHY-layer authentication game for a dynamic radio environment. As it is challenging for the radio nodes to obtain the exact channel parameters in advance, we propose spoofing detection schemes based on Q-learning and Dyna-Q, which achieve the optimal test threshold in the spoofing detection via reinforcement learning. We implement the PHY-layer spoofing detectors over universal software radio peripherals and evaluate their performance via experiments in indoor environments. Both simulation and experimental results have validated the efficiency of the proposed strategies.**

***Index Terms*—Game theory, PHY-layer authentication, reinforcement learning, spoofing detection.**

I. Introduction

WIRELESS networks are vulnerable to spoofing attacks, in which a spoofer claims to be another node by using a faked identity such as the media access control (MAC) address of the latter. Spoofers can obtain illegal advantages and further perform man-in-the-middle attacks and denial-of-service attacks [2]. PHY-layer authentication techniques exploit physical layer properties of wireless communications to detect spoofing attacks. Received signal strengths (RSSs) [2]–[4], channel impulse responses [5], [6], received signal strength indicators (RSSIs), channel state information [7]–[9], and channel frequency responses [10], [11] have been used as the fingerprints of wireless channels to detect spoofing attacks.

As the radio channel responses in wideband wireless communications are difficult to predict and thus to spoof, the channel-based spoofing detector in [10] discriminates transmitters at different locations, in which a hypothesis test compares the channel frequency responses of the messages with the same MAC address. The accuracy of the PHY-layer spoofing detection depends on the test threshold in the hypothesis test performed at the receiver. It is challenging for the receiver to choose a proper test threshold in the spoofing detector without knowing the exact values of the channel parameters in a dynamic radio environment against spoofers that can flexibly choose their attack probabilities to hide and attack effectively.

As both users and attackers have autonomy and flexible control over their transmissions, game theory has shown strengths to improve security in wireless networks, such as jamming attacks [12], [13] and collusion attacks [14]. By applying reinforcement learning techniques, such as Q-learning, a player can achieve the optimal strategies in a dynamic environment without being aware of the system information. For example, the channel selection strategy with Q-learning in [12] can address jamming attacks, and the channel accessing with Q-learning in [15] copes with jammers in a hostile environment. In addition, Dyna-Q increases the learning speed by utilizing the hypothetical experiences from the world model built from real experiences [16].

In this paper, we apply game theory to investigate the PHY-layer authentication in dynamic wireless networks, which compares the channel states of the data packets to detect spoofing attacks. The authentication process is formulated as a zero-sum authentication game consisting of the spoofers and the receiver. The receiver determines the test threshold in the PHY-layer spoofing detection, whereas each spoofer chooses its attack frequency to maximize its utility based on the Bayesian risk [17]. Spoofers cooperatively attack the receiver to avoid collisions. We derive the Nash equilibrium (NE) [18] in the static authentication game based on the channel frequency responses. Both the optimal test threshold in the PHY-layer spoofing detection and the optimal spoofing frequency depend on the relative channel time variation and the ratio of channel gains of the spoofer over the legitimate node. Simulation results show that the PHY-layer spoofing detection at the NE is robust against radio environmental changes.

Manuscript received September 3, 2015; revised November 16, 2015; accepted January 22, 2016. Date of publication February 3, 2016; date of current version December 14, 2016. This work was presented in part in a paper presented at IEEE Globecom 2015. This work was supported by the 863 High Technology Plan under Grant 2015AA01A707. The review of this paper was coordinated by Prof. Y.-B. Lin.

L. Xiao is with the Department Communication Engineering, Xiamen University, Xiamen 361005, China, and also with the Beijing Key Laboratory of IOT Information Security Technology, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China (e-mail: lxiao@xmu.edu.cn).

Y. Li, G. Han, and G. Liu are with the Department Communication Engineering, Xiamen University, Xiamen 361005, China.

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

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2016.2524258

We propose the PHY-layer authentication algorithms by exploiting the channel states of the radio packets, which determine the test threshold based on reinforcement learning, in dynamic wireless networks without knowing complete channel parameters such as the channel time variations. More specifically, the optimal test threshold in the hypothesis test is achieved by trial and error to maximize the long-term utility in the dynamic PHY-layer authentication game. The proposed PHY-layer authentication can be integrated with traditional authentication mechanisms at MAC and network layers. For example, each packet accepted by PHY-layer authentication can be further authenticated at MAC/Internet Protocol levels using the standard schemes based on the specific protocol stack. By combining with higher layer authentication, the proposed PHY-layer authentication scheme can save the time and energy to process most spoofing packets at higher layers. The proposed spoofing detection schemes are implemented over universal software radio peripherals (USRPs), and their performance is verified via field tests in typical indoor environments.

To summarize, the contributions of our work are as follows.

1) We investigate the PHY-layer authentication that exploits the channel states of the radio packets to detect spoofing attacks, as well as formulate the authentication process between the spoofers and the receiver as a static zero-sum game.
2) The NE of the static authentication game based on the channel frequency responses is derived. The condition and the uniqueness of the NE are presented.
3) We propose a spoofing detector that applies Q-learning to achieve the optimal test threshold in the hypothesis test based on the RSSIs of the radio packets in dynamic environments, as well as further improves the detection accuracy and utility of the receiver by learning from the hypothetical experiences via the Dyna architecture. The spoofing detections are implemented over USRPs and verified over experiments in indoor environments.

The rest of this paper is organized as follows. Related work is reviewed in Section II. We describe the system model of the spoofing detection in Section III and formulate the PHY-layer authentication game in Section IV. The NE of the static game is investigated in Section V, and the learning-based spoofing detections are presented in Section VI. Experimental results are presented in Section VII. Finally, Section VIII concludes this paper.

II. Related Work

The spatial correlation of RSS was exploited in [3] to detect spoofing attacks. The proximity-based authentication in [4] uses the RSS variations in proximity tests at mobile stations. The channel impulse responses can be exploited to discriminate transmitters in wireless networks [5]. The channel state information was utilized to distinguish radio transmitters even with similar signal fingerprints in [7]. The time-varying carrier frequency offsets between the transmit–receive pairs were exploited in the authentication [19]. The channel-phase responses in multicarrier systems can be used in the PHY-layer authentication in [20].

The indirect reciprocity principle was applied to address a wide range of attacks in wireless networks [14]. Frequency hopping strategies in [13] were used by secondary users to update their antijamming strategies with imperfect knowledge. The two-layer game model was introduced in [21] to investigate the joint threats from the advanced persistent threat attackers and the insiders. A defense mechanism with stochastic learning was presented in [22] to address jammers. Reinforcement learning has been applied in [23] for the mobility control in wireless networks. The event detection algorithm proposed in [24] accelerates the event detection in wireless sensor and actor networks, in which reinforcement learning techniques help each actor move toward an event.

In [1], we proposed a PHY-layer spoofing detector based on Q-learning and formulated a PHY-layer authentication game. In this paper, we extend the study in [1] by proposing a Dyna-Q-based spoofing detector to improve the authentication speed. In addition, we derive the NE of the static PHY-layer authentication game and evaluate the performance of the PHY-layer spoofing detectors in experiments with USRPs.

III. System Model

We consider a wireless network that consists of a receiver and N transmitters, including the potential spoofers that impersonate another node with a fake MAC address. For simplicity, the MAC address of the ith transmitter is denoted by $\zeta_i \in \mathcal{L}$, where $\mathcal{L}$ is the set of MAC addresses, with $1 \leq i \leq N$. The ith spoofer sends a fake packet in a time slot with a probability denoted by $y_i \in [0, 1]$.

Once receiving a packet, the receiver estimates the channel states associated with the packet. More specifically, the pilots or the preambles of the packet can be used to estimate the channel response of the corresponding transmitter, which is centered at frequency f_0 with bandwidth W. The receiver samples the channel response at the M tones for each packet in the spectrum $[f_0 - W/2, f_0 + W/2]$. The channel vector (or record) of the kth packet from the ith transmitter is denoted by $\mathbf{r}_i^k = [r_{i,m}^k]_{1 \leq m \leq M}$ (or $\hat{\mathbf{r}}_i^k = [\hat{r}_{i,m}^k]_{1 \leq m \leq M}$), where $r_{i,m}^k$ (or $\hat{r}_{i,m}^k$) is the channel vector (or record) at the mth tone of the kth packet from the ith transmitter.

A hypothesis test is performed to determine whether a packet with the channel vector $\mathbf{r}_i^k$ is indeed sent by the ith transmitter. Let $\phi(\mathbf{r})$ be the MAC address of the node that sends a packet with the channel vector $\mathbf{r}$. The null hypothesis $\mathcal{H}_0$ indicates that the packet is indeed sent by the ith transmitter. The alternative hypothesis $\mathcal{H}_1$ is that the real transmitter of the packet is not i, i.e., the packet is sent by a spoofer in reality. Thus, the spoofing detection is based on the hypothesis test given by

$$\begin{aligned} \mathcal{H}_0 &: \phi\left(\mathbf{r}_i^k\right) = \zeta_i \\ \mathcal{H}_1 &: \phi\left(\mathbf{r}_i^k\right) \neq \zeta_i. \end{aligned} \tag{1}$$

Based on the uniqueness of channel states, the receiver authenticates the kth packet based on the RSSIs. If the channel

vector based on the RSSIs, i.e., $\mathbf{r}_i^k$, and the channel record, i.e., $\hat{\mathbf{r}}_i^k$, are similar, the packet is assumed to be sent from the ith transmitter. Otherwise, the packet is spoofing. The statistic of the hypothesis test in the spoofing detector, which is denoted by $L(\mathbf{r}_i^k, \hat{\mathbf{r}}_i^k)$, is given by

$$L\left(\mathbf{r}_i^k, \hat{\mathbf{r}}_i^k\right) = \frac{\left\|\mathbf{r}_i^k - \hat{\mathbf{r}}_i^k\right\|^2}{\left\|\hat{\mathbf{r}}_i^k\right\|^2} \tag{2}$$

where $\|\cdot\|$ is the Frobenius norm. The test statistic L can be viewed as the normalized Euclidean distance between the channel vector $\mathbf{r}_i^k$ and the channel record $\hat{\mathbf{r}}_i^k$. If the test statistic is less than the test threshold, which is denoted by x, the receiver accepts the null hypothesis $\mathcal{H}_0$. Otherwise, the receiver accepts $\mathcal{H}_1$. Thus, the hypothesis test in the PHY-layer spoofing detection is given by

$$L \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} x. \tag{3}$$

According to (2), we have $L \geq 0$, and thus, $x \geq 0$.

The false alarm rate, which is denoted by P_f, is the probability that a legitimate packet is viewed as a spoofing one, i.e.,

$$P_f = \Pr(\mathcal{H}_1|\mathcal{H}_0) \tag{4}$$

where $\Pr(\cdot|\cdot)$ is the conditional probability. Similarly, the miss detection rate, which is denoted by P_m, is the probability that a spoofing packet passes the detection, which is given by

$$P_m = \Pr(\mathcal{H}_0|\mathcal{H}_1). \tag{5}$$

By (4) and (5), the probability for a receiver to accept a legitimate packet is given by $\Pr(\mathcal{H}_0|\mathcal{H}_0) = 1 - P_f$, and the probability to reject a spoofing packet is $\Pr(\mathcal{H}_1|\mathcal{H}_1) = 1 - P_m$. The detection accuracy of the PHY-layer authentication in (3) depends on the test threshold x: A large test threshold increases the missed detection rate of the spoofing detection, whereas a small x increases the false alarm rate. Therefore, it is critical for the receiver to choose a proper test threshold in the spoofing detection.

The receiver applies the higher layer authentication to process the packets that pass the PHY-layer authentication. A packet is accepted if and only if both the PHY-layer and higher layer authentications accept it. The channel record $\hat{\mathbf{r}}_i^k$ is updated once a packet from transmitter i is accepted by the higher layer authentication, i.e., $\hat{\mathbf{r}}_i^k \leftarrow \mathbf{r}_i^k$. Otherwise, $\hat{\mathbf{r}}_i^k \leftarrow \hat{\mathbf{r}}_i^{k-1}$. For ease of reference, important notations are summarized in Table I.

TABLE I
SUMMARY OF IMPORTANT SYMBOLS

Symbols	Notations
N	Number of spoofers
y_i	Spoofing probability of the i-th spoofer
W	Bandwidth
M	Number of tones
$\mathbf{r}_i^k$	Channel vector of the k-th packet claiming to be sent by the transmitter i
$\hat{\mathbf{r}}_i^k$	Channel record for transmitter i
L	Test statistic
P_f	False alarm rate
P_m	Miss detection rate
x	Test threshold
$G_{1/0}$	Payoff by accepting a legitimate packet (or rejecting a spoofing one)
$C_{1/0}$	Cost if falsely rejecting a legitimate packet (or accepting a spoofing one)
$u_{s/r}(x, \mathbf{y})$	Utility of the spoofer (or receiver)
σ^2	Average power gain from the legitimate transmitter to the receiver
ρ	SINR of the packets sent by the legitimate transmitter
b	Relative change in the channel gain
κ	Ratio of the channel gain of the spoofer to that of the legitimate transmitter
T	Number of packets received in each time slot
U_n	Sum utility of the receiver in n-th time slot
$\mu \in (0, 1]$	Learning rate of Q-learning
$\delta \in (0, 1]$	Discount factor of Q-learning
$\mathbf{s}$	State observed by the receiver
$Q(\mathbf{s}, x)$	Q-function of the receiver choosing threshold x in $\mathbf{s}$
$V(\mathbf{s})$	Best threshold to maximize Q value of the receiver in $\mathbf{s}$
ε	Probability that the receiver chooses the suboptimal actions
Φ	Occurrence count vector of each state-action pair
Φ'	Occurrence count vector of the next state
R	Modeled reward function
R'	Reward record
Π	Probability that the system reaches the next state
J	Additional Q-function updating times

IV. PHY-LAYER SPOOFING DETECTION GAME

The one-slot interaction in the PHY-layer spoofing detection in (3) can be formulated as a zero-sum game consisting of the N spoofers and the receiver. The receiver applies the PHY-layer authentication to detect spoofing attacks, whereas each spoofer sends packets using the MAC address of a legitimate transmitter. The receiver chooses the test threshold $x \in [0, \infty)$ in the hypothesis test to detect the spoofing. The ith spoofer chooses its spoofing frequency, which is denoted by $y_i \in [0, 1]$, $1 \leq i \leq N$. The set of spoofing probabilities of all the N spoofers is denoted by $\mathbf{y} = [y_i]_{1 \leq i \leq N}$. The N cooperative spoofers sent packets without collisions with $\sum_{i=1}^N y_i \leq 1$. As no more than one spoofer attacks in a time slot, the probability for the receiver to obtain a spoofing packet is $\sum_{i=1}^N y_i$.

The utility of the receiver depends on the spoofing detection accuracy. The payoff that a receiver obtains by accepting a legitimate packet (or rejecting a spoofing one) is denoted by G_1 (or G_0). The cost for a receiver to falsely reject a legitimate packet (or accepting a spoofing one) is denoted by C_1 (or C_0). The Bayesian risk [17] of the spoofing detection under a prior

distribution of the spoofing attacks, which is denoted by $\Xi(x, \mathbf{y})$, is given by

$$\Xi(x,\mathbf{y}) = \left(G_1\left(1-P_f(x)\right) - C_1 P_f(x)\right)\left(1-\sum_{i=1}^{N} y_i\right) + \left(G_0\left(1-P_m(x)\right) - C_0 P_m(x)\right)\sum_{i=1}^{N} y_i \quad (6)$$

where the first term corresponds to the gain from a legitimate packet, whereas the second term presents the gain under a spoofing packet.

In the zero-sum game, the utility of the spoofer (or receiver), which is denoted by $u_s(x, \mathbf{y})$ (or $u_r(x, \mathbf{y})$), follows $u_s(x, \mathbf{y}) = -u_r(x, \mathbf{y})$. Based on the Bayesian risk in (6), the utilities of the players are given by

$$\begin{aligned} u_r(x,\mathbf{y}) =& \; -u_s(x,\mathbf{y}) = \Xi(x,\mathbf{y}) \\ =& \; (G_0 - G_1)\sum_{i=1}^{N} y_i - (G_0 + C_0)P_m(x)\sum_{i=1}^{N} y_i \\ & - (G_1 + C_1)P_f(x)\left(1-\sum_{i=1}^{N} y_i\right) + G_1. \end{aligned} \quad (7)$$

A. Authentication Based on Channel Frequency Responses

As a concrete example, we consider the spoofing detector in [10] based on the channel frequency responses associated with the packets instead of their RSSIs. In this case, the receiver obtains a channel vector for the kth packet from transmitter i, which is denoted by H_i^k, and stores the channel records for transmitter i, which are denoted by $\hat{H}_i^k$. Assume small channel time variations, small estimation errors, zero phase drift between the channel measurements, and the frequency-selective Rayleigh channel models. The generalized likelihood ratio test, which is denoted by L', is chosen in [10] by

$$L' = \left\| H_i^k - \hat{H}_i^k \right\|^2. \quad (8)$$

The false alarm rate and the missed detection rate of the hypotheses test as in the spoofing detection in [10] are given by

$$P_f(x) = 1 - F_{\chi^2_{2M}}\left(\frac{2x\rho}{2\sigma^2 + b\rho\sigma^2}\right) \quad (9)$$

$$P_m(x) = F_{\chi^2_{2M}}\left(\frac{2x\rho}{2\sigma^2 + (1+\kappa)\rho\sigma^2}\right) \quad (10)$$

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

The PHY-layer spoofing detection game based on the channel frequency responses with a single spoofer, which is denoted by $\mathbf{G} = \langle \{r, s\}, \{x, y\}, \{u_r, u_s\} \rangle$, consists of a receiver r and a spoofer s. The receiver chooses its test threshold $x \in [0, \infty)$, whereas the spoofer determines its attack frequency $y \in [0, 1]$. The utilities of the players in the zero-sum game are given by (7).

V. Nash Equilibrium in the Authentication Game

The NE of a game consists of the best response strategies, i.e., no player can increase its utility by unilaterally choosing a different strategy [18]. The NE of the static PHY-authentication game $\mathbf{G}$ is denoted by (x^*, y^*). The receiver chooses the test threshold x^* to maximize its utility $u_r(x, y^*)$ in the spoofing detection, whereas the spoofer aims to maximize its utility $u_s(x^*, y)$. Therefore, we have

$$x^* = \arg\max_{x \geq 0} u_r(x, y^*) \quad (11)$$

$$y^* = \arg\max_{0 \leq y \leq 1} u_s(x^*, y). \quad (12)$$

Lemma 1: If $b \leq \kappa + 1$, the unique NE of the static PHY-authentication game $\mathbf{G}$ is given by

$$x^* = \mathrm{sol}_x\left(G_1 - G_0 - P_f(x)(G_1 + C_1) + P_m(x)(G_0 + C_0) = 0\right) \quad (13)$$

where $\mathrm{sol}(\cdot)$ is the solution of the equation, and

$$y^* = \frac{1}{1 + \frac{G_0 + C_0}{G_1 + C_1}\left(\frac{2+b\rho}{2+(1+\kappa)\rho}\right)^M e^{\frac{(\kappa+1-b)\rho^2 x^*}{\sigma^2(2+b\rho)(2+(1+\kappa)\rho)}}}. \quad (14)$$

Otherwise, if $b > \kappa + 1$, no NE exists in the game $\mathbf{G}$.

Proof: See the Appendix. □

According to Lemma 1, the NE of the game depends on the relative channel time variation b and the ratio of the channel gains of the spoofer to the legitimate node, i.e., κ. Under small relative channel time variations (i.e., $b \leq \kappa + 1$), both the receiver and the spoofer choose their strategies according to the payoffs of the receiver in the spoofing detection. Otherwise, under large channel time variations (i.e., $b > \kappa + 1$), the channel responses cannot be used as the fingerprint of radio channels to detect spoofing, and thus, the receiver does not have an optimal test threshold.

Corollary 1: If $b = \kappa + 1$, the spoofing detection performance of the static authentication game $\mathbf{G}$ at the NE is given by

$$P_f = 1 - P_m = \frac{G_1 + C_0}{G_1 + C_1 + G_0 + C_0}. \quad (15)$$

Proof: If $b = \kappa + 1$, by (9) and (10), we have $P_f(x^*) = 1 - P_m(x^*)$. By (13), we have (15). □

As indicated in (15), the false alarm rate increases with both G_1 and C_0, whereas the missed detection rate decreases with them. Under the high cost to accept a spoofing packet and the high payoff to accept a legitimate packet, the spoofer attacks more frequently, and thus, the receiver chooses a lower test

threshold to reject the spoofing packets, resulting in a high false alarm rate. Similarly, if both the cost to reject a legitimate packet and the payoff to reject a spoofing packet are high, the test threshold increases, yielding a small spoofing probability. Consequently, the false alarm rate decreases with G_0 and C_1, whereas the missed detection rate increases with them.

VI. Spoofing Detection With Reinforcement Learning

Reinforcement learning techniques such as Q-learning and Dyna-Q can be used to find the optimal strategy in a dynamic environment with incomplete information [16]. In a dynamic radio environment with receivers unaware of the channel model and spoofing frequency, the optimal test threshold in the spoofing detection can be achieved by the receiver via trial and error. In general, the optimal authentication threshold x^* decreases with the attack frequency.

A. Authentication With Q-Learning

As a simple reinforcement learning technique, Q-learning enables each agent to learn its optimal strategy in dynamic environments. In the spoofing detection with Q-learning, the receiver builds the hypothesis test in (3) to determine the sender for each of the T packets received in the time slot. The test threshold x is chosen from $K+1$ levels, i.e., $x \in \{l/K\}_{0\le l\le K}$. The state observed by the receiver at time n, which is denoted by $\mathbf{s}_n$, consists of the false alarm rate and the missed detection rate of the spoofing detection at time $n-1$, i.e., $\mathbf{s}_n = [P_f^{n-1}, P_m^{n-1}] \in \mathbf{S}$, where $\mathbf{S}$ is the set of the states observed by the receiver. For simplicity, both error rates are quantized into $X+1$ levels, i.e., $P_f, P_m \in \{l/X\}_{0\le l\le X}$. The receiver chooses its action x_n based on the state $\mathbf{s}_n$ to maximize its expected sum utility, denoted by U_n, which is given by

$$U_n = \sum_{k=(n-1)T+1}^{nT} u_r^k(x, \mathbf{y}) \tag{16}$$

where u_r^k is the immediate utility given by (7).

The spoofing detection with Q-learning depends on the learning rate, which is denoted by $\mu \in (0,1]$, which indicates the weight of the current Q-function, which is denoted by $Q(\mathbf{s}_n, x_n)$. The discount factor, which is denoted by $\delta \in (0,1]$, represents the uncertainty on the rewards in the future. The value of the state $\mathbf{s}$, which is denoted by $V(\mathbf{s})$, is the maximum value of the Q-function. The receiver updates its Q-function as follows:

$$Q(\mathbf{s}_n, x_n) \leftarrow (1-\mu)Q(\mathbf{s}_n, x_n) + \mu\left(U_n + \delta V(\mathbf{s}_{n+1})\right) \tag{17}$$

$$V(\mathbf{s}_n) \leftarrow \max_{x \in \left\{\frac{l}{K}\right\}_{0\le l\le K}} Q(\mathbf{s}_n, x). \tag{18}$$

The optimal test threshold x^* is chosen by

$$x^* = \arg \max_{x \in \left\{\frac{l}{K}\right\}_{0\le l\le K}} Q(\mathbf{s}_n, x). \tag{19}$$

Based on the ε-greedy policy, the receiver chooses the suboptimal actions with a small probability ε, whereas the optimal action that maximizes the utility is chosen with probability $1-\varepsilon$. Thus

$$\Pr(x = \dot{x}) = \begin{cases} 1-\varepsilon, & \dot{x} = x^* \\ \frac{\varepsilon}{K}, & \dot{x} \in \left\{\frac{l}{K}\right\}_{0\le l\le K}, \dot{x} \ne x^* \end{cases}. \tag{20}$$

The spoofing detection with Q-learning is summarized in Algorithm 1.

Algorithm 1 Spoofing detection with Q-learning

Initialize: ε, μ, δ, $Q(\mathbf{s}, x) = \mathbf{0}$, $V(\mathbf{s}) = \mathbf{0}$,$\forall x \in \{l/K\}_{0\le l\le K}$
for $n = 1, 2, 3, \ldots$ **do**
  Observe the current state $\mathbf{s}_n$
  Select a threshold x_n via (20)
  for $k = 1$ **to** T **do**
    Read the MAC address ζ_i of the kth packet
    Extract $\mathbf{r}_i$ and $\hat{\mathbf{r}}_i$
    Calculate L via (2)
    if ($L \le x_n$ **and** the kth packet passes the higher layer authentication) **then**
      $\hat{\mathbf{r}}_i \leftarrow \mathbf{r}_i$
      Accept the kth packet
    else
      Send spoofing alarm
    end if
  end for
  Observe $\mathbf{s}_{n+1}$ and U_n
  Update $Q(\mathbf{s}_n, x_n)$ via (17)
  Update $V(\mathbf{s}_n)$ via (18)
end for

B. Authentication With Dyna-Q

As an extension of Q-learning, Dyna-Q applies the Dyna architecture in [16] to formulate a learned world model that consists of the major functions of the online planning receiver. By obtaining the hypothetical experiences from the world model, Dyna-Q increases the learning speed in a dynamic environment with unknown parameters. The spoofing detection with Dyna-Q can improves the performance over Q-learning. As shown in Algorithm 2, the receiver first applies Q-learning to obtain real experiences in the spoofing detection.

At time n, the receiver observes the false alarm rate and the missed detection rate of the spoofing detection at the last time slot, i.e., $\mathbf{s}_n = [P_f^{n-1}, P_m^{n-1}] \in \mathbf{S}$. Based on state $\mathbf{s}_n$, the receiver applies the test threshold x_n chosen based on ε-greedy policy in (20) in the spoofing detection for the T packets received in a time slot and then observes the resulting security performance to estimate its immediate utility and update its sum utility U_n, as in (16). The real experiences in the nth time slot are given by $(\mathbf{s}_n, x_n, \mathbf{s}_{n+1}, U_n)$. Similar to Q-learning, the receiver updates the Q-function via (17) and (18).

After obtaining real experiences, the receiver updates the experience records for each state–action pair, which consist of the occurrence count vector Φ, the occurrence count vector of the next state Φ', the modeled reward function R, the reward record R', and the state transition probability $\mathbf{\Pi}$. The count vector Φ' for the current experience increases by 1, i.e.,

$$\Phi'(\mathbf{s}_n, x_n, \mathbf{s}_{n+1}) \leftarrow \Phi'(\mathbf{s}_n, x_n, \mathbf{s}_{n+1}) + 1. \tag{21}$$

The occurrence count vector Φ consists of all the possible realizations of $\mathbf{s}_{n+1}$, i.e.,

$$\Phi(\mathbf{s}_n, x_n) \leftarrow \sum_{\mathbf{s}' \in \mathbf{S}} \Phi'(\mathbf{s}_n, x_n, \mathbf{s}'). \tag{22}$$

The reward record R' is based on the sum utility in (16), which is given by

$$R'\left(\mathbf{s}_n, x_n, \Phi(\mathbf{s}_n, x_n)\right) \leftarrow U_n. \tag{23}$$

Thus, the modeled reward function is the average over all the occurrence realizations

$$R(\mathbf{s}_n, x_n) \leftarrow \frac{1}{\Phi(\mathbf{s}_n, x_n)} \sum_{\psi=1}^{\Phi(\mathbf{s}_n, x_n)} R'(\mathbf{s}_n, x_n, \psi). \tag{24}$$

The state transition probability from the current state $\mathbf{s}_n$, with action x_n, to the next state $\mathbf{s}_{n+1}$, maps the current state–action pair $(\mathbf{s}_n, x_n)$ to the distribution of the next state $\mathbf{s}_{n+1}$. More specifically, the probability for the system to reach the state $\mathbf{s}_{n+1}$ from the state–action pair $(\mathbf{s}_n, x_n)$, which is denoted by $\Pi(\mathbf{s}_n, x_n, \mathbf{s}_{n+1}) \in \mathbf{\Pi}$, is updated by

$$\Pi(\mathbf{s}_n, x_n, \mathbf{s}_{n+1}) \leftarrow \frac{\Phi'(\mathbf{s}_n, x_n, \mathbf{s}_{n+1})}{\Phi(\mathbf{s}_n, x_n)}. \tag{25}$$

Algorithm 2 Spoofing detection with Dyna-Q

1: **Initialize**: ε, μ, δ, J, $Q(\mathbf{s}, x) = \mathbf{0}$, $V(\mathbf{s}) = \mathbf{0}$, $\Phi(\mathbf{s}, x) = \mathbf{0}$, $\Phi'(\mathbf{s}, x, \mathbf{s}') = \mathbf{0}$, $R'(\mathbf{s}, x, \Phi(\mathbf{s}, x, \mathbf{s}')) = \mathbf{0}$, $R(\mathbf{s}, x) = \mathbf{0}$, $\Pi(\mathbf{s}, x_j, \mathbf{s}') = \mathbf{0}$, $\forall x \in \{l/K\}_{0 \le l \le K}, \mathbf{s}, \mathbf{s}' \in \mathbf{S}$
2: **for** $n = 1, 2, 3, \ldots$ **do**
3: Authenticate the received T packets as steps 3–18 in Algorithm 1
4: Update $\Phi'(\mathbf{s}_n, x_n, \mathbf{s}_{n+1})$ via (21)
5: Update $\Phi(\mathbf{s}_n, x_n,)$ via (22)
6: Update the state transition probability function $\Pi(\mathbf{s}_n, x_n, \mathbf{s}_{n+1})$ via (25)
7: Update the reward record $R'(\mathbf{s}_n, x_n, \Phi(\mathbf{s}_n, x_n))$ via (23)
8: Update the reward function $R(\mathbf{s}_n, x_n)$ via (24)
9: **for** $j = 1$ **to** J **do**
10: Randomly select a state–action pair $(\mathbf{s}_j, x_j)$
11: Select the next state $\mathbf{s}_{j+1}$ via $\Pi(\mathbf{s}_j, x_j, \mathbf{s})$
12: Obtain the reward $U_j = R(\mathbf{s}_j, x_j)$
13: Update $Q(\mathbf{s}_j, x_j)$ via (26)
14: Update $V(\mathbf{s}_j)$ via (27)
15: **end for**
16: **end for**

Next, the receiver uses these real experience records (Φ, Φ', R, R', $\mathbf{\Pi}$) to build the Dyna architecture to obtain the hypothetical experience. In the Dyna architecture, the receiver performs J additional Q-function updating processes in the modeled environment. In the jth updating process, the receiver first randomly and repeatedly chooses a state–action pair $(\mathbf{s}_j, x_j)$ to try all the actions at all the states. The modeled system updates its state based on the state transition probability $\mathbf{\Pi}$ in (25). Based on the reward function R in (24), the receiver updates its modeled reward U_j in the state–action pair $(\mathbf{s}_j, x_j)$. According to the next state and the modeled reward, the receiver updates its Q-function by

$$Q(\mathbf{s}_j, x_j) \leftarrow (1-\mu) Q(\mathbf{s}_j, x_j) + \mu \left(U_j + \delta V(\mathbf{s}_{j+1})\right) \tag{26}$$

$$V(\mathbf{s}_j) \leftarrow \max_{x \in \left\{\frac{l}{K}\right\}_{0 \le l \le K}} Q(\mathbf{s}_j, x). \tag{27}$$

By repeating the Q-function updating processes J times, $Q(\mathbf{s}, x)$ converges faster than that with Q-learning. The parameter J weighs the real experiences. If J is so large that the real experiences are negligible, the modeled Dyna architecture may deviate from the real wireless system. On the other hand, the convergence speed of the detection is slow under a small J.

VII. Performance Evaluation

Simulations were performed to evaluate the NE strategy of the static spoofing detection game under various scenarios. Experiments over USRPs also have been performed to evaluate the spoofing detections in dynamic spoofing detection games.

A. Numerical Results

The NE of the static spoofing detection game $\mathbf{G}$ was evaluated via simulations with $G_1 = G_0 = 6$, $C_1 = 2$, $C_0 = 4$, $\sigma = 1$, and $M = 10$. As shown in Fig. 1(a), the optimal test threshold increases with the channel time variation b to avoid rejecting legitimate packets, as the test statistic in (2) increases with it. As the test statistic in the presence of spoofer increases with the channel time variation index b, a spoofer is more likely to fail under large channel variations. Thus, the optimal attack probability y^* decreases with b, as shown in Fig. 1(b). In Fig. 1(c), both the false alarm rate and the missed detection rate of the spoofing detection at the NE increase with the channel variation index b, because it is challenging to discriminate transmitters according to their channel states under significant radio environmental changes. If the ratio of the channel gains of the spoofer to the legitimate node is high, the spoofer and the legitimate transmitter have quite different power levels, which improves the detection accuracy. As the channel estimation error at the receiver decreases with SINR, both the false alarm rate and the missed detection rate of the PHY-layer spoofing detection decrease with SINR. Even if $\kappa = -3$ dB, $b = 0.2$, and $\rho = 10$ dB, the spoofing detection still achieves good performance with $P_f = 1.5\%$ and $P_m = 1.19\%$. Fig. 1(d) shows that the utility of the receiver increases with κ and ρ, as the detection accuracy increases correspondingly. For example, if $b = 0.2$ and $\rho = 10$ dB, the utility of the receiver for $\kappa = 0$ dB increases to 5.96 from 5.88 for $\kappa = -3$ dB.

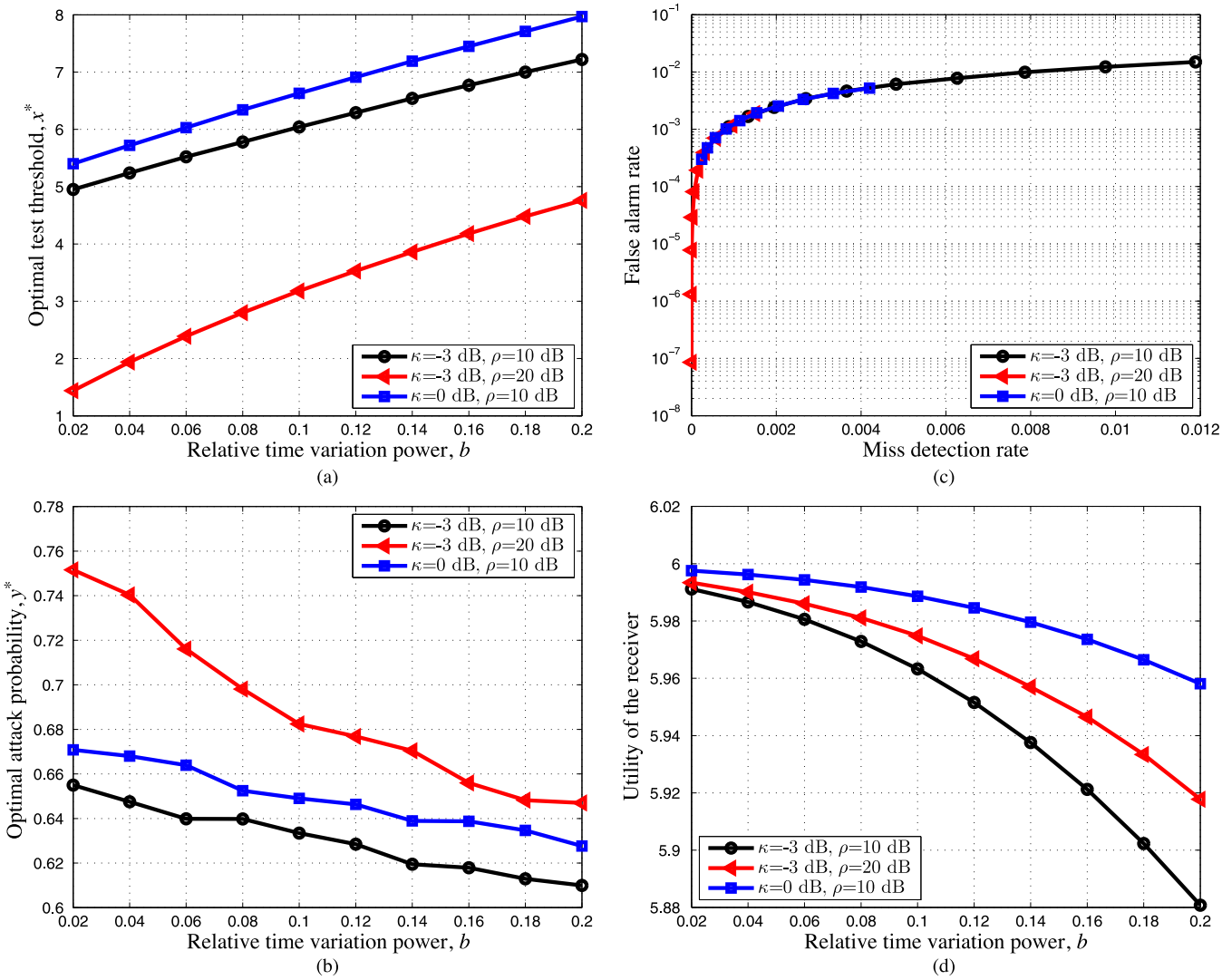


Fig. 1. Performance of the spoofing detection game at the NE. (a) Optimal test threshold x^*. (b) Optimal attack probability y^*. Performance of the spoofing detection game at the NE. (c) Error rates in the detection. (d) Utility of the receiver.

B. Experimental Results

The proposed spoofing detection schemes were implemented on USRPs, and experiments were performed in an indoor environment. As benchmark, we considered the spoofing detection with a fixed threshold. In the experiments, we considered 12 transmitters in a $12 \times 9.5 \times 3$ m^3 office room, as shown in Fig. 2, with $G_1 = C_1 = 6$, $G_0 = 9$, $C_0 = 4$, $f_0 = 2.4$ GHz, $M = 5$, $y = 0.25$, $\mu = 0.8$, $\delta = 0.7$, $\varepsilon = 0.1$, $X = 29$, and $J = 10$. As shown in Fig. 3, the maximum test statistic of the legitimate packets in an experiment with four transmitters is 0.1325, whereas the minimum test statistic of spoofing packets is 0.0185. Thus, the action set of the receiver in the experiment was set between 0.01 and 0.14.

As shown in Fig. 4(b), the spoofing detection with Q-learning reaches a stable utility that is higher than the fixed threshold soon after the start of the game. For instance, the utility of the receiver with Q-learning is 14.7% higher than that with the fixed test threshold after 1500 time slots with $W = 200$ MHz and $M = 5$. Fig. 4(a) shows that the optimal threshold achieved by the spoofing detection with Q-learning is about 0.023.

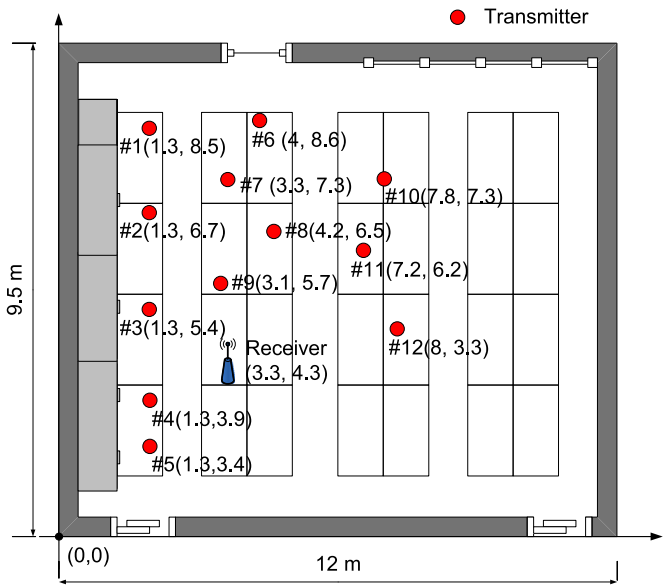


Fig. 2. Network topology of the experiments in a $12 \times 9.5 \times 3$ m^3 office room, consisting of 12 transmitters and a receiver.

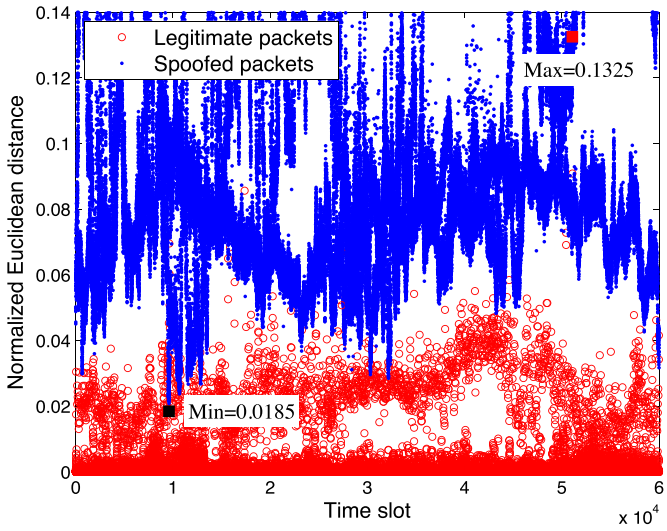
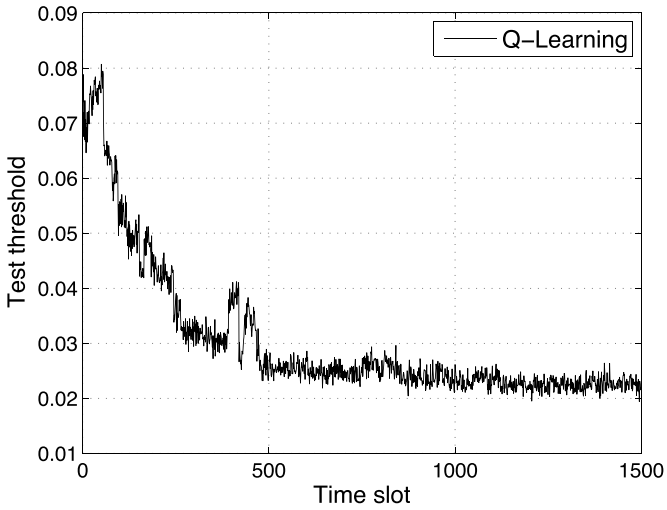
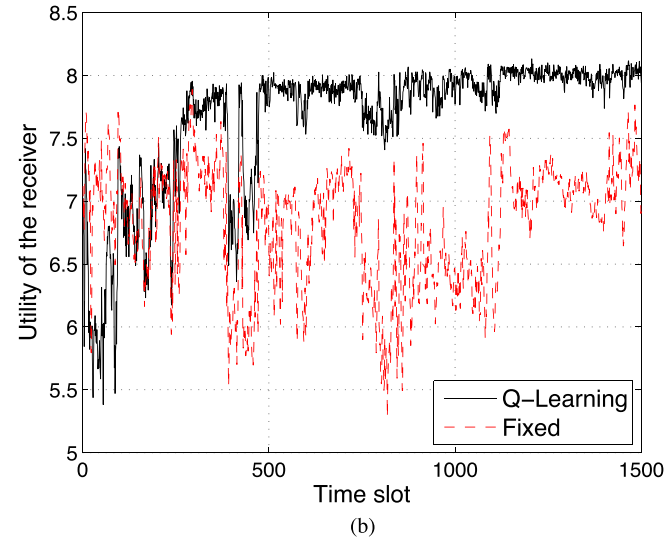


Fig. 3. Test statistic of the PHY-layer spoofing detection with $W = 200$ MHz and $M = 5$, in which the nodes located at #2, #7, #10, and #12 in the topology shown in Fig. 2 send spoofing packets.

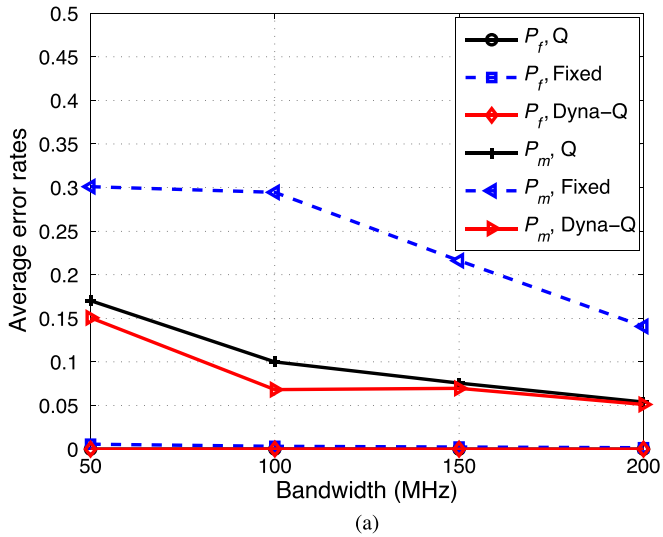


(a)

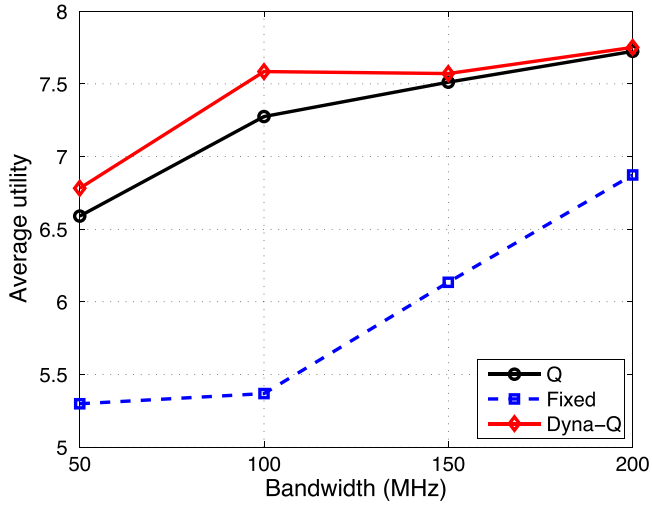


(b)

Fig. 4. Performance of the spoofing detection with $W = 200$ MHz, $M = 5$, and the legitimate node located at #10 and spoofers located at #2, #7, and #12 in the topology in Fig. 2. (a) Threshold in the spoofing detection. (b) Utility of the receiver.



(a)



(b)

Fig. 5. Performance of the PHY-layer spoofing detector in the indoor environment with $M = 5$, the legitimate node located at #10, and spoofers located at #2, #7, and #12 in the topology shown in Fig. 2. (a) Average error rates. (b) Average utility of the receiver.

As shown in Fig. 5(a), the spoofing detection with Dyna-Q has lower error rates than that with Q-learning, and both have better detection accuracy than that with a fixed threshold. For example, the missed detection rate and the false alarm rate of the spoofing detection with Q-learning are reduced by 61.72% and 93.33% compared with those of the fixed threshold strategy with $W = 200$ MHz and $M = 5$. As shown in Fig. 5(b), the detection with Dyna-Q leads to the highest utility of the receiver and that with Q-learning also improves the utility compared with the benchmark detector. For instance, the average utility of the receiver with Q-learning increases by 22.44% from that of the fixed threshold, if $W = 150$ MHz and $M = 5$, which is further improved by 0.79% by Dyna-Q, as Dyna-Q increases the learning speed of the detector.

As shown in Fig. 6, the average error rates and utility of the spoofing detectors decrease graciously with the number of nodes. For example, if $W = 200$ MHz, $M = 5$, and $N = 8$, the missed detection rate of the spoofing detector with Dyna-Q is 6.9%, whereas that with Q-learning is 7.9%, with the false

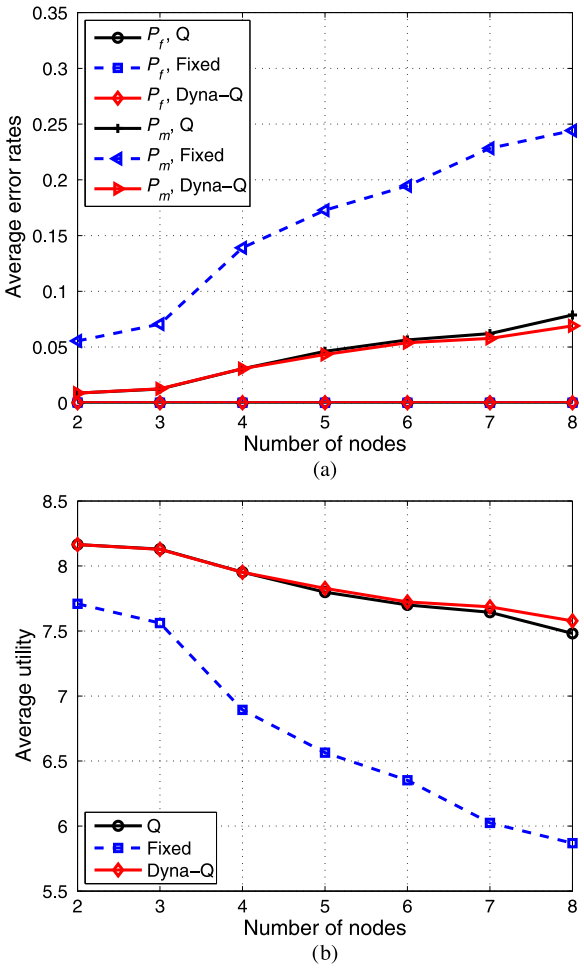


Fig. 6. Performance of the PHY-layer spoofing detector in the indoor environment with $W = 200$ MHz, $M = 5$, and the legitimate node located at #10 and spoofers located at #1–12 in the topology shown in Fig. 2. (a) Average error rates. (b) Average utility of the receiver.

alarm rates being less than 5%. The utility of the receiver with Dyna-Q is 1.28% higher than that with Q-learning in this case.

VIII. Conclusion

In this paper, we have formulated the interactions between a receiver and spoofers as a zero-sum spoofing detection game. We have derived the NE in the static spoofing detection game and discussed the uniqueness of the NE. The PHY-authentication method based on Q-learning and Dyna-Q was proposed for a dynamic radio environment. Simulation results show that the spoofing detection is robust against environmental changes. Experimental results over USRPs in an indoor environment demonstrate that the proposed spoofing detection scheme with Q-learning can efficiently improve the authentication performance. For example, the average error rate of the proposed scheme is less than 5%, whereas that with a fixed threshold is more than 14%, when the bandwidth is 200 MHz and the number of nodes is 4. As the spoofing detection scheme with Dyna-Q increases the learning speed over with Q-learning, the authentication performance is efficiently improved. For example, when the bandwidth is 100 MHz and the number of nodes is 4, the average utility of the receiver with the threshold strategy based on Q-learning is improved by 4.3% by the Dyna-Q-based detector.

Appendix
Proof of Lemma 1

According to (9) and (10), we have $P_f(0) = 1$, $P_m(0) = 0$, $\lim_{x\to\infty} P_f(x) = 0$, $\lim_{x\to\infty} P_m(x) = 1$, and

$$\frac{dP_f(x)}{dx} = -\frac{x^{M-1}e^{-\frac{x}{\sigma^2(2/\rho+b)}}}{(\sigma^2(2/\rho+b))^M\,\Gamma(M)} \tag{28}$$

$$\frac{dP_m(x)}{dx} = \frac{x^{M-1}e^{-\frac{x}{\sigma^2(2/\rho+1+\kappa)}}}{(\sigma^2(2/\rho+1+\kappa))^M\,\Gamma(M)} \tag{29}$$

where $\Gamma(\tau) = \int_0^\infty \varphi^{\tau-1}e^{-\varphi}d\varphi$.

By (7), we have

$$\frac{\partial u_s(x,y)}{\partial y} = G_1 - G_0 - P_f(x)(G_1 + C_1) + P_m(x)(G_0 + C_0) \tag{30}$$

indicating that $u_s(x,y)$ is a linear function of y. By (30), we have $\partial u_s(0,y)/\partial y = -G_0 - C_1 < 0$ and $\lim_{x\to\infty} \partial u_s(x,y)/\partial y = G_1 + C_0 > 0$.

By (28)–(30), we have

$$\frac{\partial^2 u_s(x,y)}{\partial y \partial x} = \frac{x^{M-1}e^{\frac{-x}{\sigma^2\left(\frac{2}{\rho}+b\right)}}}{\sigma^{2M}\Gamma(M)} \times \left(\frac{G_1 + C_1}{\left(\frac{2}{\rho}+b\right)^M} + \frac{(G_0 + C_0)e^{\frac{(\kappa+1-b)x/\sigma^2}{\left(\frac{2}{\rho}+b\right)\left(\frac{2}{\rho}+1+\kappa\right)}}}{\left(\frac{2}{\rho}+1+\kappa\right)^M}\right) \geq 0 \tag{31}$$

indicating that $\partial u_s(x,y)/\partial y$ increases with x. As $\partial u_s(0,y)/\partial y < 0$ and $\lim_{x\to\infty} \partial u_s(x,y)/\partial y > 0$, the solution of $\partial u_s(x,y)/\partial y = 0$, which is denoted by $\hat{x}$, is unique and positive. If $0 \leq x < \hat{x}$, we have $\partial u_s(x,y)/\partial y < 0$. Otherwise, if $x > \hat{x}$, we have $\partial u_s(x,y)/\partial y > 0$.

Next, by (7), (28), and (29), we have

$$\frac{\partial u_r(x,y)}{\partial x} = \frac{x^{M-1}e^{\frac{-x/\sigma^2}{\frac{2}{\rho}+b}}}{\sigma^{2M}\Gamma(M)}\left(\xi_1 - \xi_2 e^{\frac{(\kappa+1-b)x/\sigma^2}{\left(\frac{2}{\rho}+b\right)\left(\frac{2}{\rho}+1+\kappa\right)}}\right) \tag{32}$$

where $\xi_1 = (G_1 + C_1)(1-y)/(2/\rho+b)^M$, and $\xi_2 = (G_0 + C_0)y/(2/\rho+1+\kappa)^M$.

1) If $b < \kappa + 1$: As $\partial u_s(\hat{x},y)/\partial y = 0$, $u_s(\hat{x},y)$ is constant for any $y \in [0,1]$, i.e., y^* can be any value in [0,1]. Let $\hat{y}$ be the solution of $\partial u_r(\hat{x},y)/\partial x = 0$, which is given by (14) after simplification. If $y = \hat{y}$, we have $\xi_1 > \xi_2$. By (32), we have $\partial u_r(x,\hat{y})/\partial x > 0$ for $0 < x < \hat{x}$ and $\partial u_r(x,\hat{y})/\partial x < 0$ for $x > \hat{x}$, indicating that $x^* = \hat{x}$, if $y^* = \hat{y}$. Thus, we have $(x^*, y^*) = (\hat{x}, \hat{y})$.

2) If $b = \kappa + 1$: As $\partial u_s(\hat{x}, y)/\partial y = 0$, $u_s(\hat{x}, y)$ is constant with $y \in [0, 1]$. It is clear by (14) that $\partial u_r(x, \hat{y})/\partial x = 0$, indicating that $u_r(x, \hat{y})$ is constant with $x \geq 0$. Thus, we have $(x^*, y^*) = (\hat{x}, \hat{y})$.

   The uniqueness of the NE for $b \leq \kappa + 1$ is proved by contradictions. Assume that there exists another NE $(x_1, y_1) \neq (x^*, y^*)$. If $x_1 < x^*$, we have $\partial u_s(x_1, y)/\partial y < 0$, and thus, $y_1 = 0$. By (32), we have $\partial u_r(x, 0)/\partial x \geq 0$, i.e., $u_r(x, y_1)$ increases with x. Thus, $u_r(x_1, y_1) < u_r(x^*, y_1)$, contradicting the assumption that (x_1, y_1) is NE. If $x_1 > x^*$, we have $\partial u_s(x_1, y)/\partial y > 0$, yielding $y_1 = 1$. By (32), we have $\partial u_r(x, y_1)/\partial x \leq 0$, i.e., $u_r(x, y_1)$ decreases with x. Thus, $u_r(x_1, y_1) < u_r(x^*, y_1)$, contradicting to the assumption. Thus, (x^*, y^*) is the unique NE in this game.

3) If $b > \kappa + 1$: Similar to Case 2, if $x^* \neq \hat{x}$, no NE exists. Otherwise, if $x^* = \hat{x}$, we have $\partial u_r(x^*, y^*)/\partial x = 0$. In addition, $\partial u_r(x, y^*)/\partial x > 0 \quad \forall 0 < x < \hat{x}$, and $\partial u_r(x, y^*)/\partial x < 0 \ \forall x > \hat{x}$. However, by (14), we have $\partial u_r(x^*, y^*)/\partial x = 0$. By (32), we have $\partial u_r(x, y^*)/\partial x < 0$ for $0 < x < \hat{x}$ and $\partial u_r(x, y^*)/\partial x > 0$ for $x > \hat{x}$, indicating that $x^* \neq \hat{x}$. Thus, no NE exists in this case.

In summary, we have Lemma 1.

REFERENCES

[1] L. Xiao, Y. Li, G. Liu, Q. Li, and W. Zhuang, "Spoofing detection with reinforcement learning in wireless networks," in *Proc. IEEE GLOBECOM*, San Diego, CA, USA, 2015, pp. 1–6.

[2] K. Zeng, K. Govindan, and P. Mohapatra, "Non-cryptographic authentication and identification in wireless networks," *IEEE Wireless Commun.*, vol. 17, no. 5, pp. 56–62, Oct. 2010.

[3] J. Yang, Y. Chen, W. Trappe, and J. Cheng, "Detection and localization of multiple spoofing attackers in wireless networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 1, pp. 44–58, Jan. 2013.

[4] A. Kalamandeen, A. Scannell, E. Lara, A. Sheth, and A. LaMarca, "Ensemble: Cooperative proximity-based authentication," in *Proc. Int. Conf. Mobile Syst., Appl. Services*, 2010, pp. 331–344.

[5] F. Liu, X. Wang, and H. Tang, "Robust physical layer authentication using inherent properties of channel impulse response," in *Proc. IEEE MILCOM*, 2011, pp. 538–542.

[6] F. Liu, X. Wang, and S. Primak, "A two dimensional quantization algorithm for CIR-based physical layer authentication," in *Proc. IEEE ICC*, 2013, pp. 4724–4728.

[7] H. Liu, Y. Wang, J. Liu, J. Yang, and Y. Chen, "Practical user authentication leveraging channel state information (CSI)," in *Proc. ACM Symp. Inf., Comput. Commun. Security*, 2014, pp. 389–400.

[8] J. Tugnait, "Wireless user authentication via comparison of power spectral densities," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1791–1802, Sep. 2013.

[9] Z. Jiang, J. Zhao, X. Li, J. Han, and W. Xi, "Rejecting the attack: Source authentication for WiFi management frames using CSI information," in *Proc. IEEE INFOCOM*, 2013, pp. 2544–2552.

[10] L. Xiao, L. Greenstein, N. Mandayam, and W. Trappe, "Fingerprints in the ether: Using the physical layer for wireless authentication," in *Proc. IEEE ICC*, 2007, pp. 4646–4651.

[11] L. Xiao *et al.*, "PHY-authentication protocol for spoofing detection in wireless networks," in *Proc. IEEE GLOBECOM*, 2010, pp. 1–6.

[12] W. Conley and A. Miller, "Cognitive jamming game for dynamically countering ad hoc cognitive radio networks," in *Proc. IEEE MILCOM*, 2013, pp. 1176–1182.

[13] Y. Wu, B. Wang, K. Liu, and T. Clancy, "Anti-jamming games in multi-channel cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 1, pp. 4–15, Jan. 2012.

[14] L. Xiao, Y. Chen, W. Lin, and K. Liu, "Indirect reciprocity security game for large-scale wireless networks," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 4, pp. 1368–1380, Aug. 2012.

[15] Y. Gwon, S. Dastangoo, C. Fossa, and H. Kung, "Competing mobile network game: Embracing anti-jamming and jamming strategies with reinforcement learning," in *Proc. IEEE Conf. Commun. Netw. Security*, 2013, pp. 28–36.

[16] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 1998.

[17] A. Sage and J. Melsa, *Estimation Theory With Applications to Communications and Control*. New York, NY, USA: McCraw-Hill, 1971.

[18] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA, USA: MIT Press, 1991.

[19] W. Hou, X. Wang, J. Chouinard, and A. Refaey, "Physical layer authentication for mobile systems with time-varying carrier frequency offsets," *IEEE Trans. Commun.*, vol. 62, no. 5, pp. 1658–1667, May 2014.

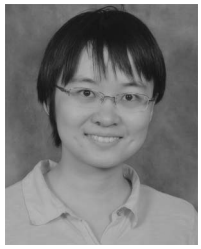
[20] X. Wu and Z. Yang, "Physical-layer authentication for multi-carrier transmission," *IEEE Commun. Lett.*, vol. 19, no. 1, pp. 74–77, Jan. 2015.

[21] P. Hu, H. Li, H. Fu, D. Cansever, and P. Mohapatra, "Dynamic defense strategy against advanced persistent threat with insiders," in *Proc. IEEE INFOCOM*, 2015, pp. 747–755.

[22] S. Bhunia, S. Sengupta, and F. Vazquez-Abad, "CR-honeynet: A learning and decoy based sustenance mechanism against jamming attack in CRN," in *Proc. IEEE MILCOM*, 2014, pp. 1173–1180.

[23] K. Ota *et al.*, "ORACLE: Mobility control in wireless sensor and actor networks," *Comput. Commun.*, vol. 35, no. 9, pp. 1029–1037, 2012.

[24] M. Dong *et al.*, "RENDEZVOUS: Towards fast event detecting in wireless sensor and actor networks," *Computing*, vol. 96, no. 10, pp. 995–1010, Oct. 2014.



Liang Xiao (M'09–SM'13) received the B.S. degree in communication engineering from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2000; the M.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2003; and the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 2009.

She is currently a Professor with the Department of Communication Engineering, Xiamen University, Xiamen, China. She is also currently with the Beijing Key Laboratory of IOT Information Security Technology, Institute of Information Engineering, Chinese Academy of Sciences, Beijing. Her current research interests include network security and wireless communications.

Yan Li (S'15) received the B.S. degree in electronic and information engineering from Heilongjiang Bayi Agricultural University, Daqing, China, in 2013. She is currently working toward the M.S. degree in the Department of Communication Engineering, Xiamen University, Xiamen, China.

Her research interests include network security and wireless communications.

Guoan Han (S'15) received the B.S. degree in communication engineering from Southwest Jiaotong University, Chengdu, China, in 2015. He is currently working toward the M.S. degree with the Department of Communication Engineering, Xiamen University, Xiamen, China.

His research interests include network security and wireless communications.

Guolong Liu (S'15) received the B.S. degree in communication engineering from Hunan Institute of Science and Technology, Yueyang, China, in 2013. He is currently working toward the M.S. degree with the Department of Communication Engineering, Xiamen University, Xiamen, China.

His research interests include network security and wireless communications.

Weihua Zhuang (M'93–SM'01–F'08) received the B.Sc. and M.Sc. degrees from Dalian Maritime University, Dalian, China, and the Ph.D. degree from the University of New Brunswick, Fredericton, NB, Canada, all in electrical engineering.

Since 1993, she has been with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Canada, where she is currently a Professor and a Tier I Canada Research Chair in Wireless Communication Networks. Her current research focuses on resource allocation and quality-of-service provisioning in wireless networks and on smart grid.

Dr. Zhuang is a Fellow of the Canadian Academy of Engineering and the Engineering Institute of Canada and an Elected Member in the Board of Governors and VP Publications of the IEEE Vehicular Technology Society. She is a Technical Program Cochair for the IEEE VTC Fall 2016. She was the Editor-in-Chief of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY during 2007–2013. She has been a corecipient of several best paper awards from IEEE conferences.