This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2652484, IEEE Transactions on Vehicular Technology

Game Theoretic Study on Channel-based Authentication in MIMO Systems

Liang Xiao, *Senior Member, IEEE*, Tianhua Chen, *Student Member, IEEE*, Guoan Han, *Student Member, IEEE*, Weihua Zhuang, *Fellow, IEEE*, and Limin Sun, *Member, IEEE*

***Abstract*—In this paper, we investigate the authentication based on radio channel information in multiple-input multiple-output (MIMO) systems, and formulate the interactions between a receiver with multiple antennas and a spoofing node as a zero-sum physical (PHY)-layer authentication game. In this game, the receiver chooses the test threshold of the hypothesis test to maximize its Bayesian risk based utility in the spoofing detection, while the adversary chooses its attack rate, i.e., how often a spoofing signal is sent. We derive the Nash equilibrium (NE) of the static PHY-layer authentication game and present the condition that the NE exists, showing that both the spoofing detection error rates and the spoofing rate decrease with the number of transmit and receive antennas. We propose a PHY-layer spoofing detection algorithm for MIMO systems based on Q-learning, 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. We also use Dyna architecture and prioritized sweeping (Dyna-PS) to improve the spoofing detection in time-variant radio environments. The proposed authentication algorithms are implemented over universal software radio peripherals and evaluated via experiments in an indoor environment. Experimental results show that the Dyna-PS based spoofing detection algorithm further reduces the spoofing detection error rates and increases the utility of the receiver compared with the Q-learning based algorithm, and both performance improves with more number of transmit or receive antennas.**

***Index Terms*—MIMO, PHY-layer authentication, spoofing detection, game theory, reinforcement learning**

I. Introduction

Multiple-input multiple-output (MIMO) techniques can improve the capacity and reliability of wireless communication systems, and increase secrecy capacities against eavesdropping [1]. However, MIMO transmissions are still vulnerable to spoofing attacks, in which a spoofer claims to be another user by using the latter's identity such as her medium access control (MAC) address. By sending spoofing signals, the adversary can not only obtain illegal advantages, but also launch further attacks such as man-in-the-middle attacks and denial-of-service attacks [2]. Spoofing attacks can be addressed by physical (PHY)-layer authentication techniques that exploit the spatial decorrelation property of PHY-layer features of radio propagations and transmitter devices, such as received signal strengths [2], [3] and channel impulse responses [4] to discriminate radio transmitters.

Game theory is a powerful mathematical tool to analyze the interactions among autonomous players that have the same or different goals [5]. For instance, the interactions between a spoofing node and a receiver that performs the channel-based spoofing detection are formulated in [6] as a PHY-layer authentication game, in which the receiver chooses its test threshold in the authentication to maximize its payoff based on the authentication accuracy, while the spoofer decides how often to send spoofing signals. Game theory also helps develop mechanisms to motivate autonomous individuals to follow the desirable policies and achieve the optimal strategies. For example, the Q-learning based spoofing detection algorithm developed in [6] enables a receiver to derive the optimal test threshold in the dynamic PHY-layer authentication game without being aware of the radio channel model and the spoofing model. However, the PHY-layer authentication game for MIMO systems is more complicated than the game model in [6] and has not been investigated, to the best of our knowledge.

In this work, we extend the PHY-layer authentication game formulated in [6] to a MIMO system, and formulate a channel-based MIMO authentication game. In this game, a spoofing node chooses its attack rate, i.e., how often to send a spoofing signal to maximize its utility based on the Bayesian risk in the spoofing detection. On the other hand, the receiver determines its test threshold in the spoofing detection according to the number of antennas and the size of the frequency samples. The Nash equilibrium (NE) as a solution concept of a two-player non-cooperative game, in which no player can gain by unilaterally changing only his or her own strategy, is derived for the static MIMO authentication game, showing that the test threshold in the detection increases with the number of antennas to avoid rejecting legitimate signals, while the attack rate decreases with it.

We also investigate the dynamic MIMO authentication

This paper was presented in part at IEEE Globecom 2016 [1], and was supported in part by the National Natural Science Foundation of China (61671396), and in part by 863 high technology plan (Grant No. 2015AA01A707).

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).

T. Chen and G. Han 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).

L. Sun is with the Beijing Key Laboratory of IOT Information Security Technology, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China (e-mail: sunlimin@iie.ac.cn).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2652484, IEEE Transactions on Vehicular Technology

game, in which the receiver uses the received signal strength indicators (RSSIs) to discriminate the transmitters in each time slot without knowing the channel time variations and authentication costs. The receiver can employ reinforcement learning algorithms, such as Q-learning as described in [7], to gradually learn the optimal test threshold through the repeat interactions with both the spoofer and the unknown dynamic MIMO environment. The performance of the Q-learning based spoofing detection in time-variant radio environments can be improved by applying the Dyna architecture [8] and prioritized sweeping (Dyna-PS) [9]. More specifically, the Dyna-PS based spoofing detection establishes a learned world model from real experience, and prioritizes the state-action pairs according to their urgency to be explored in the learned world model. We implement both detection schemes in MIMO systems over universal software radio peripherals (USRPs), and perform experiments in an indoor environment to validate their efficacy. Experiment results show that the Dyna-PS based detection exceeds the Q-learning based scheme with a faster learning speed and a higher detection accuracy. Both schemes outperform the benchmark detection with a randomly chosen test threshold.

The contributions of our work can be summarized as follows:

(1) We formulate the interactions between a receiver performing the channel-based spoofing detection and a spoofing node in MIMO systems as a zero-sum PHY-layer authentication game.

(2) We derive the NE of the static PHY-layer authentication game and provide the conditions that the game has no NE.

(2) We investigate the dynamic PHY-layer authentication game and propose a channel-based spoofing detection algorithm based on Q-learning for MIMO systems in a dynamic radio environment. We further improve its performance with Dyna architecture and prioritized sweeping. Experiments over USRPs in an indoor environment are performed to evaluate their detection accuracy and utilities.

The rest of this paper is organized as follows. We review related work in Section II, and present the system model in Section III. We formulate the PHY-layer authentication game in Section IV, and derive the NE of the static game in Section V. We present the Q-learning based spoofing detection algorithm for MIMO systems in Section VI, and develop the Dyna-PS based spoofing detection in Section VII. Experiment results are presented in Section VIII. In Section IX, we conclude this work.

II. Related Work

Channel-based authentication methods exploit the spatial decorrelation property of radio propagation to detect spoofing attacks in wireless systems. For instance, received signal strength measured at the mobile node is compared with the channel record of the claimed transmitter in [3] to detect the spoofing attacks in wireless networks. The spoofing detection algorithm developed in [10] uses the generalized likelihood ratio test to discriminate radio nodes according to their channel frequency responses in MIMO systems. The PHY-layer authentication system proposed in [11] evaluates the estimated channel responses to detect both primary user emulation attacks and Sybil attacks in cognitive radio networks. The spoofing strategies against MIMO systems as evaluated in [12] can be detected by the PHY-layer authentication even in the presence of the optimal spoofing strategy, if the channel estimation is precise. The channel impulse response based PHY-layer authentication designed in [4] introduces a two-dimensional quantization scheme to tolerate the random errors and reduce the spoofing error rates.

Game theory has been used to study wireless security. For instance, a zero-sum jamming game formulated in [13] provides the optimal anti-jamming communication strategy for cognitive radio nodes with perfect channel information. A non-cooperative random access game investigated in [14] addresses jamming attacks in a wireless network with unknown jamming models. The MIMO transmission against a dual-threat attacker that performs both eavesdropping and jamming is formulated as a zero-sum game in [15]. The interaction between a secondary user and a jammer is formulated in [16] as a stochastic game with minimax-Q learning. The transmission over the control channel with reinforcement learning is developed in [17] to achieve the optimal channel allocation in ad hoc networks against jamming.

The interactions between a legitimate receiver and a spoofing node is formulated in [6] as a zero-sum channel-based authentication game. We have extended the work in [6] to formulate a MIMO spoofing detection game in [1], and found that the detection accuracy increases with the number of transmit antennas. Compared with our previous work in [1], we investigate in this paper the NE of the static spoofing detection game in a generic case, and propose a Dyna-PS based detection scheme to improve the detection accuracy of MIMO systems compared with the Q-learning based scheme as proposed in [1] in dynamic radio environments.

III. System Model

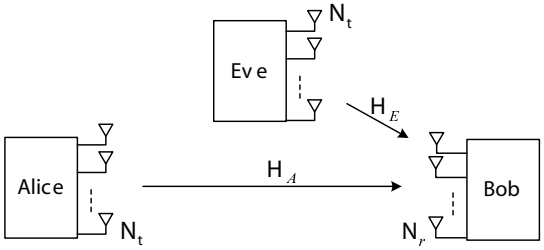


Fig. 1. Illustration of a PHY-layer authentication game consisting of the legal transmitter Alice with N_t antennas, the spoofing node Eve with N_t antennas, and the legitimate receiver Bob with N_r antennas.

As shown in Fig. 1, we consider the spoofing detection of a receiver Bob (B) with N_r receive antennas. Both the legal transmitter Alice (A) and the spoofing node Eve (E) are assumed to use N_t antennas to send signals, each of which contains the MAC address of Alice. Eve takes time and energy to launch spoofing attacks according to a chosen attack rate. Eve will be punished if her spoofing signal is detected by Bob, and will receive illegal payoff otherwise.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2652484, IEEE Transactions on Vehicular Technology

Both Alice and Eve send pilots at M frequencies along with their data symbols in each time slot. For simplicity, we assume that the M pilots are equally separated at the same spectrum with the data symbols at center frequency f_0 and bandwidth W and sent before the data symbols. The channel gain of the signal from the m-th transmit antenna to the n-th receive antenna at frequency i and time slot k is denoted by $h_t^k(m,n,i)$, with $1 \le m \le N_t$, $1 \le n \le N_r$, and $1 \le i \le M$, where the subscript t means the transmitter under test, with $t = A$ if the signal is sent by Alice, and $t = E$ if the transmitter is Eve. For simplicity, we define a $N_t N_r M$-dimensional channel gain vector for the signal under test at time slot k as $\mathbf{H}_t^k = \left[h_t^k\right]$, with $t = A$ or E.

Bob uses the pilots in the channel estimation and obtains the channel estimate denoted by $\tilde{h}_t^k(m,n,i)$ for the channel gain $h_t^k(m,n,i)$ with $t = A$ or E. More specifically, $\tilde{h}_t^k(m,n,i) = h_t^k(m,n,i) + \epsilon^k(m,n,i)$, where $\epsilon^k(m,n,i)$ is the channel estimation error of the receiver and usually modeled as the zero-mean Gaussian distribution. For convenience, we define the $N_t N_r M$-dimensional channel estimate vector $\tilde{\mathbf{H}}_t^k = \left[\tilde{h}_t^k\right]$, and set the channel record of Alice $\hat{\mathbf{H}}_A = \tilde{\mathbf{H}}_A^j$, where j is the time index of the previous signal from Alice that passes the authentication.

Let σ^2 denote the average power gain along the path from Alice to Bob, ρ be the average signal-to-noise ratio (SNR) of the signals received by Bob, α indicate the channel time variation due to environment changes, and β present the power ratio of the Eve's signal to Alice's signal. For ease of reference, the important notations are summarized in TABLE 1.

TABLE I
SUMMARY OF SYMBOLS AND NOTATIONS

α	Channel time variation index
β	Power ratio of Eve's signal to Alice's
ρ	Average SNR of received signals
$N_{t/r}$	Number of transmit/receive antennas
$\mathbf{H}_{A/E}^k$	Channel gain of Alice/Eve at k
$\tilde{\mathbf{H}}_t^k$	Channel vector of the signal under test at k
$\hat{\mathbf{H}}_A$	Channel record of Alice
M	Number of frequency samples
$x \in [0,\infty)$	Test threshold
$y \in [0,1]$	Spoofing rate
$u_{B/E}$	Immediate utility of Bob/Eve
$G_{1/0}$	Gain to accept/reject a signal from Alice/Eve
$C_{1/0}$	Cost to reject/accept a signal from Alice/Eve
C_s	Cost of Eve to send a spoofing signal
$p_{f/m}$	False alarm/miss detection rate
U^k	Expected sum utility of Bob at k
$Q(\mathbf{s},x)$	Q-function at state $\mathbf{s}$ and action x
$V(\mathbf{s})$	Value function at state $\mathbf{s}$
μ	Learning rate
δ	Discount factor
$\Delta(\mathbf{s},x)$	Priority of state-action pair $(\mathbf{s},x)$
Ψ	Priority queue of state-action pairs
Φ	Occurrence count vector
Φ'	Occurrence count vector of the next state
R/R'	Modeled reward funtion/record
Π	State transition probability

IV. PHY-Layer Authentication Game in MIMO Systems

In this section, we apply game theory to investigate the authentication based on the spatial decorrelation of channel frequency responses. At time slot k, either Alice or Eve sends signals to Bob claiming to be Alice. The channel-based spoofing detection establishes a hypothesis test to decide whether or not the signal that Bob receives at time slot k is sent by Alice. The null hypothesis $\mathcal{H}_0$ indicates that the signal is indeed sent by Alice (i.e., the channel gain is $\mathbf{H}_A^k$). In the alternative hypothesis $\mathcal{H}_1$, the claimant node is Eve. Thus the hypothesis test in the spoofing detection is given by

$$\mathcal{H}_0 : \mathbf{H}_t^k = \mathbf{H}_A^k \tag{1}$$

$$\mathcal{H}_1 : \mathbf{H}_t^k \neq \mathbf{H}_A^k. \tag{2}$$

Bob compares the new channel vector $\tilde{\mathbf{H}}_t^k$ with the channel record of Alice $\hat{\mathbf{H}}_A$. If the channel gain $\mathbf{H}_t^k$ is significantly different from the channel record of Alice, Bob chooses the alternative hypothesis and sends a spoofing alarm; otherwise, there is every reason to believe that the signal is sent by Alice.

The test statistic, denoted by L, is chosen as the normalized Euclidean distance between the channel estimate $\tilde{\mathbf{H}}_t^k$ and the channel record of Alice, and is compared with the test threshold x^k at time slot k. As the test statistic L is positive, we have $x^k > 0$. If the test statistic L is less than x^k, Bob accepts the null hypothesis $\mathcal{H}_0$; otherwise, Bob accepts $\mathcal{H}_1$. Thus the PHY-layer authentication is given by

$$L\left(\tilde{\mathbf{H}}_t^k, \hat{\mathbf{H}}_A\right) = \frac{\left\|\tilde{\mathbf{H}}_t^k - \hat{\mathbf{H}}_A\right\|^2}{\left\|\hat{\mathbf{H}}_A\right\|^2} \underset{\mathcal{H}_1}{\overset{\mathcal{H}_0}{\lessgtr}} x^k \tag{3}$$

where $\|\cdot\|$ is the Frobenius norm.

The detection accuracy of the PHY-layer authentication depends on the test threshold x^k. For instance, the successful detection rate decreases with x^k. The false alarm rate of the spoofing detection, denoted by p_f, is defined as the probability that Bob discards Alice's signal by mistake, i.e.,

$$p_f(x) = \Pr\left(\mathcal{H}_1|\mathcal{H}_0\right) = \Pr\left(L\left(\tilde{\mathbf{H}}_A^k, \hat{\mathbf{H}}_A\right) > x\right) \tag{4}$$

where $\Pr(\cdot|\cdot)$ is the conditional probability. The miss detection rate, denoted by p_m, is defined as the probability that a spoofing signal passes the PHY-layer detection and given by

$$p_m(x) = \Pr\left(\mathcal{H}_0|\mathcal{H}_1\right) = \Pr\left(L\left(\tilde{\mathbf{H}}_E^k, \hat{\mathbf{H}}_A\right) \le x\right). \tag{5}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2652484, IEEE Transactions on Vehicular Technology

Once taking the null hypothesis $\mathcal{H}_0$, Bob applies higher-layer authentication methods such as those proposed in [18] and [19], and accepts the signal and updates the channel record with $\hat{\mathbf{H}}_A = \tilde{\mathbf{H}}_t^k$, if the higher-layer authentication also accepts the signal.

The interaction between Bob and Eve at time slot k can be formulated as a static PHY-layer spoofing detection game, in which Eve chooses how often to send spoofing signals in the time slot, denoted by $y^k \in [0,1]$, while Bob determines his test threshold in the detection $x^k \in [0,\infty)$. We consider a zero-sum MIMO authentication game, and omit the superscript k if no confusion occurs. The utilities of Bob and Eve in the time slot are denoted by u_B and u_E, respectively, and satisfy $u_E = -u_B$. The payoff for Bob to accept a legitimate signal (or reject a spoofing one) is denoted by G_1 (or G_0). On the other hand, the cost for Bob to falsely reject a legitimate signal (or accept a spoofing one) is denoted by C_0 (or C_1). The cost for Eve to send a spoofing signal is denoted by C_s. Based on the detection accuracy and security cost, the utility of each player in the static game can be defined according to the Bayesian risk of the detection as

$$u_B(x,y) = -u_E(x,y) = \Big(G_1\big(1-p_f\left(x\right)\big) - C_1 p_f(x)\Big)\left(1-y\right) + \Big(G_0\big(1-p_m\left(x\right)\big) - C_0 p_m(x) + C_s\Big)y. \tag{6}$$

V. NE of the PHY-Layer Authentication Game

We consider the Nash equilibrium of the static PHY-layer authentication game, denoted by (x^*, y^*), in which neither Bob nor Eve can increase his or her utility by unilaterally choosing another strategy, i.e.,

$$x^* = \arg\max_{x\geq 0}\ u_B\left(x, y^*\right) \tag{7}$$

$$y^* = \arg\min_{0\leq y\leq 1} u_B\left(x^*, y\right). \tag{8}$$

As a concrete example, we assume zero phase shift between the channel measurements at neighboring time slots and frequency-selective Rayleigh channel models. For convenience, the test statistic of the generalized likelihood ratio test in (3), is replaced by $L' = \left\|\tilde{\mathbf{H}}_t^k - \hat{\mathbf{H}}_A\right\|^2$. In this case, if the channel responses at M frequencies from each of the $N_t \times N_r$ antenna pairs are independent and identically distributed, we have

$$L'\left(\tilde{\mathbf{H}}_t^k, \hat{\mathbf{H}}_A\right) = \left\|\tilde{\mathbf{H}}_t^k - \hat{\mathbf{H}}_A\right\|^2 \sim \chi^2\left(2N_tN_rM\right) \tag{9}$$

where $\chi^2(m)$ is a Chi-square distribution with m degrees of freedom.

The detection accuracy of the spoofing detection depends on the test threshold x, the average power gain along the path σ^2, the average SNR of the signal received by Bob ρ, the channel time variation α, and the average ratio of the SNR between Eve's and Alice's signals β. By [10], we have

$$p_f(x) = 1 - F_{\chi^2}\left(\frac{2x\rho}{\sigma^2\left(2+\alpha\rho\right)}, 2N_tN_rM\right) \tag{10}$$

$$p_m(x) = F_{\chi^2}\left(\frac{2x\rho}{\sigma^2(2+\rho+\beta\rho)}, 2N_tN_rM\right) \tag{11}$$

where $F_{\chi^2}(\cdot, m)$ is the cumulative distribution function of a Chi-square distribution with m degrees of freedom. In the following, we present the NE of the authentication game in this simplified scenario.

Theorem 1. *If the channel gains over M frequencies are independent and identically distributed, and*

$$\begin{cases} C_s < G_1 + C_0 & \text{(12a)} \\ \alpha \leq \beta + 1 & \text{(12b)} \end{cases}$$

the static PHY-layer authentication game for the $N_t \times N_r$ MIMO systems has a unique NE given by

$$\left(G_1+C_1\right)F_{\chi^2}\left(\frac{2\rho x^*}{\sigma^2\left(2+\alpha\rho\right)}, 2N_tN_rM\right) = G_0 + C_s + C_1 - \left(G_0+C_0\right)F_{\chi^2}\left(\frac{2\rho x^*}{\sigma^2(2+\rho+\beta\rho)}, 2N_tN_rM\right) \tag{13}$$

$$y^* = \left(1 + \frac{G_0+C_0}{G_1+C_1}\left(\frac{2\rho^{-1}+\alpha}{2\rho^{-1}+1+\beta}\right)^{N_tN_rM} \exp\left(\frac{\left(\beta+1-\alpha\right)\rho^2 x^*}{\sigma^2\left(2+\alpha\rho\right)\left(2+\rho+\beta\rho\right)}\right)\right)^{-1}. \tag{14}$$

Proof: By (6) (10) and (11), if $G_1 + C_0 > C_s$, we have

$$\frac{\partial u_E(0,y)}{\partial y} = -G_0 - C_1 - C_s < 0 \tag{15}$$

$$\lim_{x\to\infty}\frac{\partial u_E(x,y)}{\partial y} = G_1 + C_0 - C_s > 0. \tag{16}$$

Let $\hat{x}$ be the solution of $\partial u_E(x,y)/\partial y = 0$. As u_E is a linear function of y, we have $x^* = \hat{x}$, with $\hat{x}$ given by (13) after simplification. By (10) and (11), we have

$$\frac{\partial p_f(x)}{\partial x} = -\frac{x^{N_tN_rM-1}\exp\left(-\frac{x\rho}{\sigma^2(2+\alpha\rho)}\right)}{\left(\sigma^2(2\rho^{-1}+\alpha)\right)^{N_tN_rM}\Gamma\left(N_tN_rM\right)} \tag{17}$$

$$\frac{\partial p_m(x)}{\partial x} = \frac{x^{N_tN_rM-1}\exp\left(-\frac{x\rho}{\sigma^2(2+\rho+\beta\rho)}\right)}{\left(\sigma^2(2\rho^{-1}+1+\beta)\right)^{N_tN_rM}\Gamma\left(N_tN_rM\right)} \tag{18}$$

where $\Gamma(\cdot)$ is the Gamma function [20]. Based on (6), (17) and (18), we have

$$\frac{\partial^2 u_E}{\partial y \partial x} = \frac{x^{N_tN_rM-1}\rho^{N_tN_rM}}{\sigma^{2N_tN_rM}\Gamma\left(N_tN_rM\right)}\left(\frac{\left(G_1+C_1\right)\exp\left(-\frac{x\rho}{\sigma^2(2+\alpha\rho)}\right)}{(2+\alpha\rho)^{N_tN_rM}} + \frac{\left(G_0+C_0\right)\exp\left(-\frac{x\rho}{\sigma^2(2+\rho+\beta\rho)}\right)}{(2+\rho+\beta\rho)^{N_tN_rM}}\right) \geq 0 \tag{19}$$

showing that $\hat{x}$ is unique and positive. If $x > \hat{x}$, we have $\partial u_E(x,y)/\partial y > 0$; otherwise, if $0 \leq x < \hat{x}$, we have $\partial u_E(x,y)/\partial y < 0$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2652484, IEEE Transactions on Vehicular Technology

By (17), (18) and (6), we have

$$\frac{\partial u_B}{\partial x} = \frac{x^{N_t N_r M-1} \exp\left(-\frac{x\rho}{\sigma^2(2+\alpha\rho)}\right)}{\sigma^{2N_t N_r M}\Gamma\left(N_t N_r M\right)} \left(\frac{(G_1+C_1)(1-y)}{(2\rho^{-1}+\alpha)^{N_t N_r M}} - \frac{(G_0+C_0)y \exp\left(\frac{(\beta+1-\alpha)x\rho^2}{(2+\alpha\rho)(2+\rho+\beta\rho)\sigma^2}\right)}{(2\rho^{-1}+1+\beta)^{N_t N_r M}} \right). \quad (20)$$

As $\partial u_E(\hat{x},y)/\partial y = 0$, $u_E(\hat{x},y)$ is constant for any $y \in [0,1]$. Let $\hat{y}$ be the solution of $\partial u_B(\hat{x},y)/\partial x = 0$, which can be simplified by (20) into (14). By (20), if $\beta+1 \geq \alpha$, we have $\partial u_B(x,\hat{y})/\partial x \geq 0$ for $0 < x \leq \hat{x}$ and $\partial u_B(x,\hat{y})/\partial x \leq 0$ for $x > \hat{x}$. Thus, (7) and (8) hold for $(x^*,y^*) = (\hat{x},\hat{y})$.

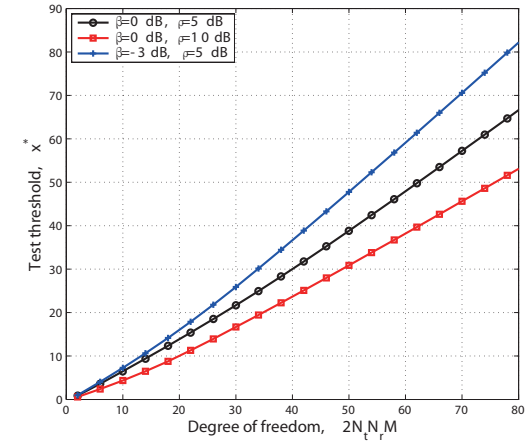
Next, we prove the uniqueness of the NE by assuming that there exists another NE $(x',y') \neq (x^*,y^*)$. If $0 \leq x' < x^*$, we have $\partial u_E(x',y)/\partial y < 0$ and thus $y' = 0$. By (20), we have $\partial u_B(x,y')/\partial x \geq 0$, indicating that $u_B(x,y')$ increases with x. Thus $u_B(x',y') < u_B(x^*,y')$, contradicting with the assumption that (x',y') is an NE. If $x' > x^*$, we have $\partial u_E(x',y)/\partial y > 0$, yielding $y' = 1$. By (20), we have $\partial u_B(x,y')/\partial x \leq 0$, and thus $u_B(x',y') < u_B(x^*,y')$, contradicting with the assumption. Thus, $(x^*,y^*) = (\hat{x},\hat{y})$ is a unique NE in the game. □

The NE of the PHY-layer authentication game given by (13) and (14) depends on the spoofing cost and the relative channel time variation. If both the spoofing cost and channel time variation are small (i.e., (12a) and (12b)), Bob chooses his test threshold and Eve decides her spoofing rate based on the gains of correct detections, the costs of detection errors, and the channel parameters given by (13) and (14).

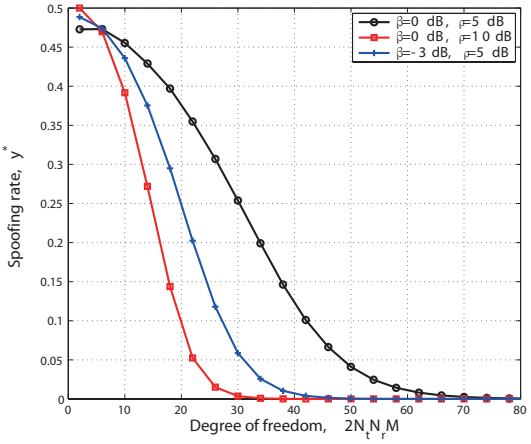
As a concrete example, we evaluate the NE of the game in Fig. 2 with $G_1 = 9$, $G_0 = 10$, $C_1 = 1$, $C_0 = 5$, $C_s = 0.1$, $\sigma = 1$ and $\alpha = 0.3$, showing that the optimal test threshold increases with the degree of freedom (i.e., $2N_tN_rM$), and the spoofing rate decreases with it, e.g., the spoofing rate is negligible if $N_tN_rM = 30$. As shown in Fig. 3, the detection accuracy increases with the degree of freedom. For instance, the false alarm rate of the PHY-layer authentication is below 4×10^{-4} and the miss detection rate is less than 0.07, if $2N_tN_rM \geq 40$, $\beta = -3$dB and $\rho = 5$dB. Consequently, the utility of Bob increases with the degree of freedom before converging to 9. The spoofing rate and the error rates of the proposed detection algorithms decrease with the average SNR of the signal. For example, the spoofing rate decreases from 0.35 to 0.05, and the miss detection rate decreases from 0.012 to 0.075, as ρ increases from 5 dB to 10 dB, at $2N_tN_rM = 20$, $\alpha = 0.3$ and $\beta = 0$ dB. The spoofing rate decreases with the power ratio of the Eve's signal to Alice's signal β, and the spoofing detection accuracy increases with it. For example, the spoofing rate decreases from 0.4 to 0.3, and the false alarm rate decreases from 0.75 to 0.001, as β changes from 0 dB to -3 dB.

Theorem 2. *The static MIMO PHY-layer authentication game does not have an NE, if*

$$C_s \geq G_1 + C_0 \quad (21)$$



(a) Test threshold in spoofing detection



(b) Spoofing rate

Fig. 2. NE of the static PHY-layer spoofing detection game with $G_1 = 9$, $G_0 = 10$, $C_1 = 1$, $C_0 = 5$, $C_s = 0.1$, $\sigma = 1$ and $\alpha = 0.3$.

or

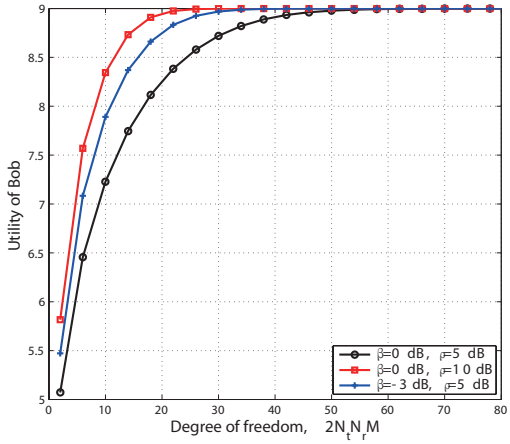
$$\begin{cases} C_s < G_1 + C_0 & (22a) \\ \alpha > \beta + 1. & (22b) \end{cases}$$

Proof: If $C_s \geq G_1 + C_0$, we have $\partial u_E(x,y)/\partial y < 0$, thus $y^* = 0$. By (20), we have $\partial u_B(x,0)/\partial x \geq 0$ and $x^* \to \infty$, thus no NE exists.

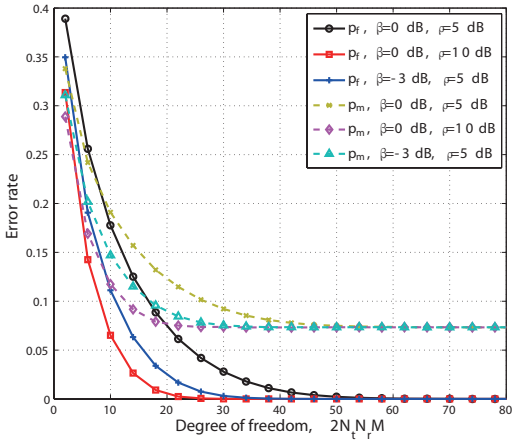
If $C_s < G_1 + C_0$ and $\alpha > \beta + 1$, the NE is given by (13) and (14) as shown in Theorem 1. By (20) and (14), we have $\partial u_B(x,y^*)/\partial x < 0$, $\forall 0 < x < x^*$, and $\partial u_B(x,y^*)/\partial x > 0$, $\forall x > x^*$, contradicting with (7) and (8). Thus, no NE exists in this case. □

If the channel time variation is large (i.e., (22b)), the receiver can not use the channel responses to build a stable radio fingerprint. On the other hand, if the spoofing cost is large (i.e., (21)), the attack motivation is suppressed and thus Bob chooses a high test threshold to avoid false alarm in the spoofing detection.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2652484, IEEE Transactions on Vehicular Technology



(a) Utility of Bob



(b) Spoofing detection error rate

Fig. 3. Performance of the PHY-layer spoofing detection game at the NE with $G_1 = 9$, $G_0 = 10$, $C_1 = 1$, $C_0 = 5$, $C_s = 0.1$, $\sigma = 1$ and $\alpha = 0.3$.

VI. PHY-Layer Spoofing Detection with Q-Learning

We formulate the repeated interactions between Bob and Eve in a dynamic radio environment as a dynamic PHY-layer authentication game, in which Bob builds the hypothesis test as shown in (3) to detect spoofing attacks in each time slot. For simplicity, we assume that either Alice or Eve sends T signals in a time slot. The expected sum utility of Bob is denoted by U^k and defined as

$$U^k = \sum_{n=(k-1)T+1}^{kT} u_B^n(x^k, y^k) \tag{23}$$

where u_B^n is the immediate utility from the n-th signal as defined in (6). In the dynamic game, Bob estimates the false alarm rate and miss detection rate of the T signals in a time slot and for simplicity quantizes them into X non-zero levels, i.e., p_f^k and $p_m^k \in \{l/X\}_{0 \le l \le X}$. The state observed by Bob at time slot k, denoted by $\mathbf{s}^k$, consists of both the false alarm rate and the miss detection rate at the last time slot, i.e., $\mathbf{s}^k = \left[p_f^{k-1}, p_m^{k-1}\right] \in \{l/X, m/X\}_{0 \le l,m \le X}$. The feasible test threshold in the spoofing detection is quantized into K non-zero levels, i.e., $x \in \{l/K\}_{0 \le l \le K}$.

In each time slot, the test threshold x^k is chosen based on state $\mathbf{s}^k$, and the ϵ-greedy policy with $0 < \epsilon \le 1$ is applied to avoid being trapped at the beginning of the game. More specifically, the test threshold that achieves the highest expected long-term utility or Q-function denoted by $Q()$ for the current state is chosen with a high probability $1-\epsilon$, while each of the other values is selected with a low probability ϵ/K to encourage explorations. The test threshold is chosen as

$$\Pr\left(x^k = \dot{x}\right) = \begin{cases} 1-\epsilon, & \dot{x} = \arg \max\limits_{x \in \{l/K\}_{0 \le l \le K}} Q\left(\mathbf{s}^k, x\right) \\ \frac{\epsilon}{K}, & o.w. \end{cases} \tag{24}$$

Upon choosing the test threshold x^k, Bob calculates the test statistic according to (9) and applies the hypothesis test in (3) to determine whether the transmitter that sent the T signals at time slot k is indeed Alice. If the alternative hypothesis is chosen, Bob sends a spoofing alarm, and calculates miss detection rate of the spoofing detection at time slot k. Otherwise, Bob updates the channel record of Alice and calculates the false alarm rate accordingly.

According to this experience, denoted by $\left(\mathbf{s}^k, x^k, U^k, \mathbf{s}^{k+1}\right)$, Bob updates the Q-function of the state-action pair $\left(\mathbf{s}^k, x^k\right)$ as follows:

$$Q\left(\mathbf{s}^k, x^k\right) \leftarrow (1-\mu)\, Q\left(\mathbf{s}^k, x^k\right) + \mu\left(U^k + \delta V\left(\mathbf{s}^{k+1}\right)\right) \tag{25}$$

$$V\left(\mathbf{s}^k\right) \leftarrow \max_{x \in \{\frac{l}{K}\}_{0 \le l \le K}} Q\left(\mathbf{s}^k, x\right) \tag{26}$$

where $\mu \in (0, 1]$ is the learning rate that represents the weight of the current Q-function in the learning process, and $\delta \in (0, 1]$ is the discount factor that indicates the uncertainty of the rewards in the future game. The PHY-layer spoofing detection with Q-learning is summarized in Algorithm 1.

VII. PHY-Layer Spoofing Detection with Dyna-PS

To improve the performance of the Q-learning based spoofing detection in time-variant radio environments, we apply the Dyna architecture as described in [8] that formulates a learned world model from real experience, and use prioritized sweeping [9] that prioritizes the backup state-action pairs according to their urgencies. More specifically, the Dyna-PS based spoofing detection prioritizes the state-action pairs, and remembers the predecessors of each state, i.e., the states that have a non-zero transition probability to a given state. The spoofing detection maintains a queue for each state-action pair, and prioritizes the state-action pairs according to the change of their values. The scheme first applies Q-learning to obtain real experiences, and then establishes the Dyna architecture for the hypothetical experience corresponding to the change of the Q-function of the state-action pairs in the state updates.

More specifically, at time slot k, Bob evaluates the error rates of the spoofing detection at the last time slot to form the state $\mathbf{s}^k = \left[p_f^{k-1}, p_m^{k-1}\right] \in \mathbf{S}$, where $\mathbf{S}$ is the feasible

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2652484, IEEE Transactions on Vehicular Technology

Algorithm 1 MIMO spoofing detection with Q-learning.

Initialize ε, μ, δ, $Q(\mathbf{s}, x) = \mathbf{0}$, and $V(\mathbf{s}) = \mathbf{0}$, $\forall x \in \{l/K\}_{0 \le l \le K}$
for $k = 1, 2, 3, ...$ **do**
  Choose x^k via (24)
  for $n = 1$ **to** T **do**
    Receive signal n at time slot k
    Channel estimation to obtain $\tilde{\mathbf{H}}_t^n$
    Calculate L via (9)
    if $L \le x^k$ **then**
      Perform the higher-layer authentication
      if signal n is accepted **then**
        $\acute{\mathbf{H}}_A \leftarrow \tilde{\mathbf{H}}_t^n$
      else
        Send spoofing alarm for signal n
      end if
    end if
  end for
  Observe p_f^k and p_m^k to form $\mathbf{s}^{k+1} = [p_f^k, p_m^k]$
  Update $Q\left(\mathbf{s}^k, x^k\right)$ via (25)
  Update $V\left(\mathbf{s}^k\right)$ via (26)
end for

state set, and chooses the test threshold x^k according to the ε-greedy policy as shown in (24). Then, the hypothesis test is performed according to (3) to determine whether or not the transmitter that sent the T signals at time slot k is Alice. The security performance is then evaluated to obtain both the immediate utility and the sum utility U^k in Eq. (23). Bob then updates his Q-function and value function in this real experience, respectively, via (25) and (26).

The real experience at time slot k consists of four elements, $(\mathbf{s}^k, x^k, \mathbf{s}^{k+1}, U^k)$. The receiver establishes a queue denoted by Ψ, which consists of the priority of each state-action pair, denoted by Δ. The priority is defined as the change of the Q-function of the state-action pair, set as zero at the beginning, and then updated by

$$\Delta\left(\mathbf{s}^k, x^k\right) \leftarrow \max\{\left|U^k + \delta V\left(\mathbf{s}^{k+1}\right) - Q\left(\mathbf{s}^k, x^k\right)\right|, \Delta\left(\mathbf{s}^k, x^k\right)\}. \tag{27}$$

The receiver only deals with the state-action pair whose priority is higher than a given threshold θ in the learning process.

Based on the above real experience, the spoofing detection updates its experience record for each state-action pair, which consists of the occurrence count vector denoted by Φ, the occurrence count vector of the next state denoted by Φ', the modeled reward function denoted by R, the reward record denoted by R', and the state transition probability denoted by Π. More specifically, the count vector Φ' for the current experience increases by 1 at each time slot, i.e.,

$$\Phi'\left(\mathbf{s}^k, x^k, \mathbf{s}^{k+1}\right) \leftarrow \Phi'\left(\mathbf{s}^k, x^k, \mathbf{s}^{k+1}\right) + 1. \tag{28}$$

The occurrence count vector Φ that consists of all the possible realizations of $\mathbf{s}^{k+1}$ is given by

$$\Phi\left(\mathbf{s}^k, x^k\right) = \sum_{\mathbf{s}' \in \mathbf{S}} \Phi'\left(\mathbf{s}^k, x^k, \mathbf{s}'\right). \tag{29}$$

The reward record R' is the sum utility give by

$$R'\left(\mathbf{s}^k, x^k, \Phi\left(\mathbf{s}^k, x^k\right)\right) = U^k. \tag{30}$$

The modeled reward function R is defined as the average of R' over all the occurrence realizations, i.e.,

$$R\left(\mathbf{s}^k, x^k\right) = \frac{1}{\Phi\left(\mathbf{s}^k, x^k\right)} \sum_{\psi=1}^{\Phi\left(\mathbf{s}^k, x^k\right)} R'\left(\mathbf{s}^k, x^k, \psi\right). \tag{31}$$

The state transition probability from the current state-action pair $\left(\mathbf{s}^k, x^k\right)$ to the next state $\mathbf{s}^{k+1}$, denoted by Π, is defined as

$$\Pi\left(\mathbf{s}^k, x^k, \mathbf{s}^{k+1}\right) = \frac{\Phi'\left(\mathbf{s}^k, x^k, \mathbf{s}^{k+1}\right)}{\Phi\left(\mathbf{s}^k, x^k\right)}. \tag{32}$$

Algorithm 2 MIMO spoofing detection with Dyna-PS.

Initialize: ε, μ, δ, θ, J, $Q = \mathbf{0}$, $V = \mathbf{0}$, $\Delta = \mathbf{0}$, $\Phi = \mathbf{0}$, $\Phi' = \mathbf{0}$, $\Psi = \mathbf{0}$, $R' = \mathbf{0}$, $R = \mathbf{0}$, and $\Pi = \mathbf{0}$
for $k = 1, 2, 3, ...$ **do**
  Authenticate the received T signals and update $Q\left(\mathbf{s}^k, x^k\right)$ and $V\left(\mathbf{s}^k\right)$ by Step 3-19 in Algorithm 1
  Obtain the real experience $\left(\mathbf{s}^k, x^k, \mathbf{s}^{k+1}, U^k\right)$
  Calculate $\Delta\left(\mathbf{s}^k, x^k\right)$ via (27)
  Insert $\left(\mathbf{s}^k, x^k\right)$ into Ψ with priority $\Delta\left(\mathbf{s}^k, x^k\right)$
  Update $\Phi'\left(\mathbf{s}^k, x^k, \mathbf{s}^{k+1}\right)$ via (28)
  Update $\Phi\left(\mathbf{s}^k, x^k, \right)$ via (29)
  Update $R'\left(\mathbf{s}^k, x^k, \Phi\left(\mathbf{s}^k, x^k\right)\right)$ via (30)
  Update $R\left(\mathbf{s}^k, x^k\right)$ via (31)
  Update $\Pi\left(\mathbf{s}^k, x^k, \mathbf{s}^{k+1}\right)$ via (32)
  Set $j = 0$
  while $j < J$ **do**
    Select the state-action pair $(\mathbf{s}^j, x^j)$ with the highest priority in Ψ
    if $\Delta(\mathbf{s}^j, x^j) > \theta$ **then**
      Set $\Delta(\mathbf{s}^j, x^j) = 0$
      Obtain the reward $U^j = R\left(\mathbf{s}^j, x^j\right)$
      Update $Q\left(\mathbf{s}^j, x^j\right)$ via (34)
      Update $V\left(\mathbf{s}^j\right)$ via (35)
      for all the predecessors $(\bar{\mathbf{s}}, \bar{x})$ of state $\mathbf{s}^j$ **do**
        Calculate $\Delta(\bar{\mathbf{s}}, \bar{x})$ via (36)
        Insert $(\bar{\mathbf{s}}, \bar{x})$ into Ψ with priority $\Delta(\bar{\mathbf{s}}, \bar{x})$
      end for
      j++
    else
      break
    end if
  end while
end for

According to the above experience record $(\Phi, \Phi', R, R', \Pi)$, the spoofing detection builds the Dyna architecture and obtains the hypothetical experience. Based on the Dyna architecture, this scheme performs additional Q-function updating processes

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2652484, IEEE Transactions on Vehicular Technology

if the priority queue Ψ is not empty. More specifically, in the j-th additional updating process, the detection first chooses a state-action pair $(\mathbf{s}^j,\ x^j)$ with the highest priority in Ψ, and then sets its priority $\Delta(\mathbf{s}^j, x^j) = 0$. The modeled reward U^j is set as follows

$$U^j(\mathbf{s}^j, x^j) \leftarrow R\left(\mathbf{s}^j, x^j\right). \tag{33}$$

According to the state transition probability Π and the modeled reward U^j, the spoofing detection then updates the Q-function and value function, respectively, by

$$Q(\mathbf{s}^j, x^j) \leftarrow (1-\mu)Q(\mathbf{s}^j, x^j) + \mu\left(U^j + \delta \sum_{\mathbf{s}'\in\mathbf{S}} \Pi\left(\mathbf{s}^j, x^j, \mathbf{s}'\right) V(\mathbf{s}')\right) \tag{34}$$

$$V\left(\mathbf{s}^j\right) \leftarrow \max_{x\in\{\frac{l}{K}\}_{0\le l\le K}} Q\left(\mathbf{s}^j, x\right). \tag{35}$$

The change of the Q-function for each predecessor of state $\mathbf{s}^j$, denoted by $\Delta(\bar{\mathbf{s}}, \bar{x})$ is then updated as

$$\Delta(\bar{\mathbf{s}}, \bar{x}) \leftarrow \max\left\{\left|U^j + \delta V\left(\mathbf{s}^j\right) - Q(\bar{\mathbf{s}}, \bar{x})\right|, \Delta(\bar{\mathbf{s}}, \bar{x})\right\}. \tag{36}$$

The state-action pair $(\bar{\mathbf{s}}, \bar{x})$ is inserted into the priority queue Ψ with priority given by $\Delta(\bar{\mathbf{s}}, \bar{x})$. By updating the Q-function in the learned world model, the spoofing detection as summarized in Algorithm 2 has a faster convergence speed than the Q-learning based detection.

In summary, the Dyna-PS based spoofing detection requires more memory to store the experience record than the Q-learning based algorithm, and performs J more authentication processes than the latter in each time slot. On the other hand, the Dyna-PS based detection has a faster convergence rate by using the hypothetical experiences generated by the Dyna architecture in addition to the real experiences.

VIII. Experimental Results

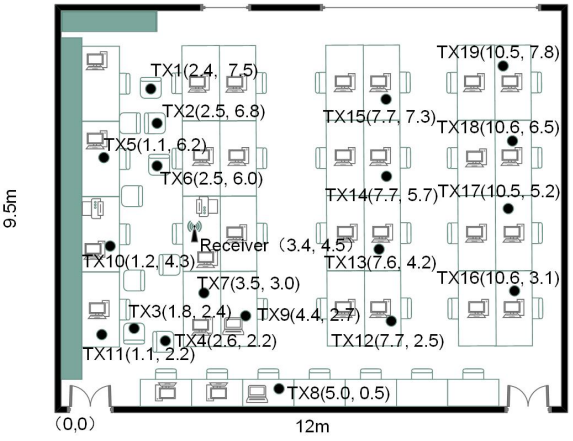
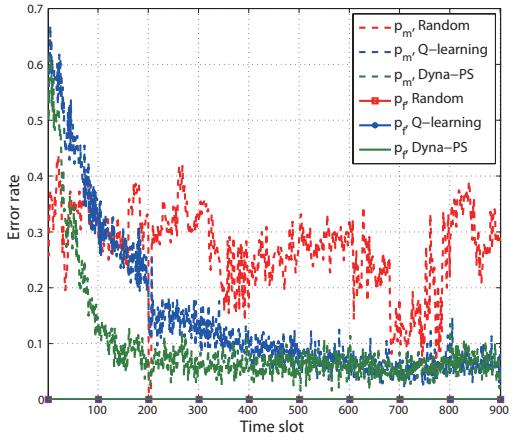
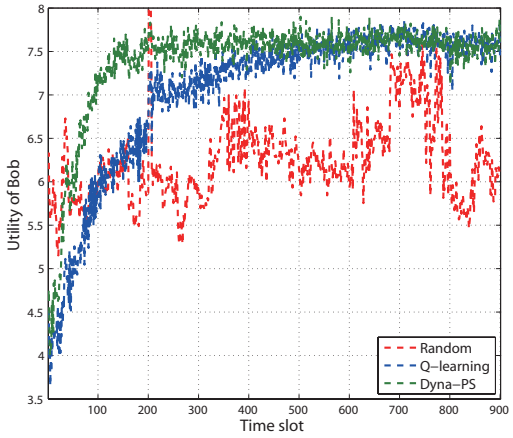


Fig. 4. Network topology of the experiments in a $12 \times 9.5 \times 3$ m^3 office room, consisting of 19 transmitters each with up to 5 antennas to act as Alice or Eve, and the receiver with up to 5 antennas.

The proposed reinforcement learning based PHY-layer spoofing detection algorithms were implemented over USRPs.



(a) Spoofing detection error rate



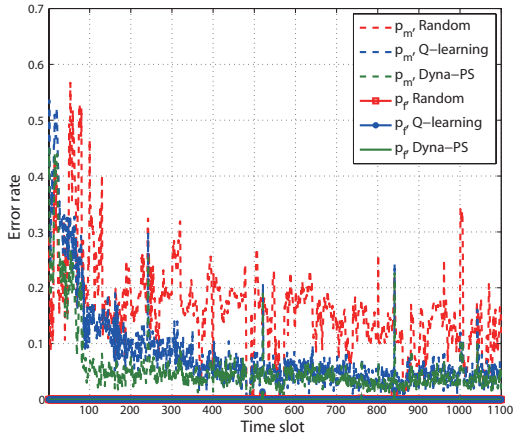
(b) Utility of Bob

Fig. 5. Performance of the PHY-layer authentication in the 5×3 MIMO system in the experiment, in which Alice and Eve were located at TX 3 and TX 7 respectively in the office room with topology shown in Fig. 4.

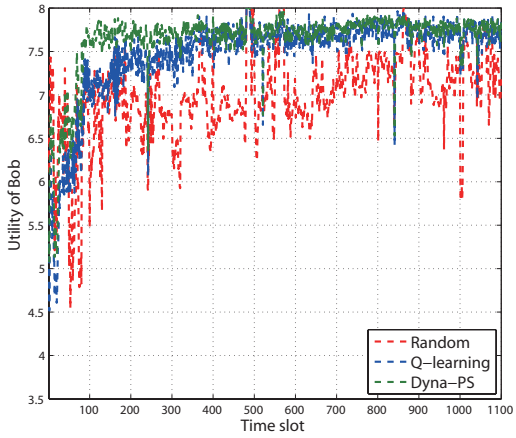
Experiments were performed in a $12 \times 9.5 \times 3$ m^3 office room to evaluate their performance in dynamic games, in which each transmitter was equipped with up to 5 antennas, and Bob had 3 receive antennas. As a benchmark, we considered the PHY-layer spoofing detection with a randomly selected test threshold. If not specified otherwise, we set in the experiments $G_1 = C_1 = 6$, $G_0 = 9$, $C_0 = 4$, $C_s = 1$, $y = 0.5$, $\mu = 0.8$, $\delta = 0.7$, $\epsilon = 0.1$, $\theta = 0.2$, and $M = 3$ frequencies at 2.41 GHz, 2.42 GHz and 2.43 GHz.

As shown in Fig. 5 (a), the miss detection rate of the Q-learning based spoofing detection in the 5×3 MIMO system decreases over time from 0.54 at the beginning to 0.07 after 600 time slots, and the convergent error rate is 83.3% lower than the benchmark scheme. The spoofing detection with Dyna-PS further reduces the miss detection rate, which only takes 200 time slots to converge to 0.07, which is 1/3 of the time required by the Q-learning based scheme. The false alarm rates of both schemes are mostly below 3×10^{-4}. In addition, as shown in Fig. 5 (b), the utility of Bob with the Q-learning based detection increases over time to reach 7.5 after 600 time

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2652484, IEEE Transactions on Vehicular Technology

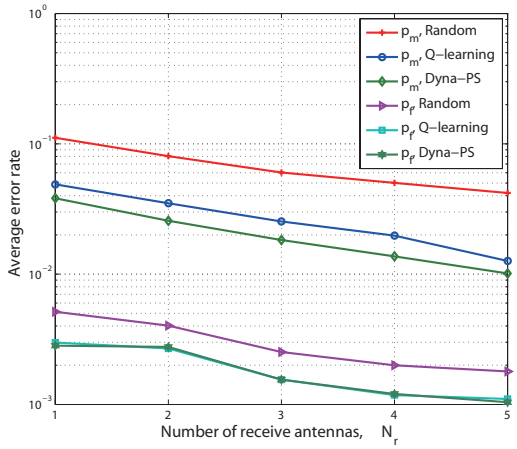


(a) Spoofing detection error rate

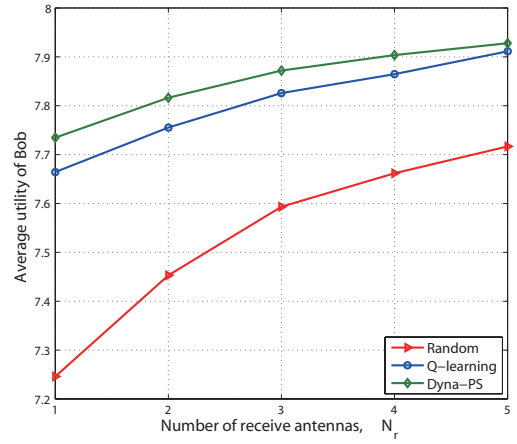


(b) Utility of Bob

Fig. 6. Performance of the PHY-layer authentication in the 5×3 MIMO system in the experiment, in which Alice and Eve were located at TX3 and TX16 respectively in the office room with topology shown in Fig. 4.



(a) Average spoofing detection error rate



(b) Average utility of Bob

Fig. 7. Average spoofing detection performance of the $5 \times N_r$ MIMO system in the experiments with topology shown in Fig. 4.

slots, which is 25% higher than the benchmark scheme. The detection with Dyna-PS has an even faster learning speed and takes 1/3 of time to reach a utility as high as 7.5 compared with the detection with Q-learning.

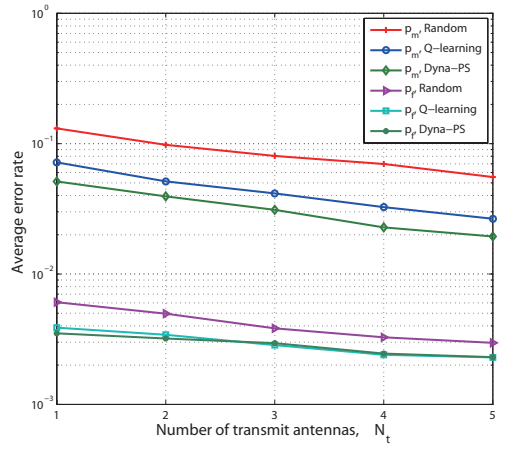
If the channel time variance is large, Fig. 6 illustrates that the miss detection rate of the Q-learning based detection converges fast even after a large channel variation. For example, at time slot 240 (the channel gains change significantly), the miss detection rate quickly decreases to 4%. The Dyna-PS based scheme takes 1/4 of time to reach the same accuracy, and obtains a utility 2.6% higher than the Q-learning based scheme.

Fig. 7 illustrates the impacts of the number of receive antennas on the spoofing detection of the MIMO system with 5 transmit antennas. For example, the miss detection rate of the Q-learning based detection decreases by 42.9% to 1.9%, and the false alarm rate decreases by 55.6% to 0.1%, if the number of receive antennas changes from 2 to 4. The spoofing detection with the Dyna-PS based detection has an even lower error rate, e.g., the miss detection rate of the 5×4 MIMO system is 1.3%, which is 31.5% lower than the Q-learning based scheme. Consequently, as shown in Fig. 7 (b), the average utility of Bob with the Q-learning based detection increases by 2.6% compared with the benchmark, and is further improved by 3.1% with Dyna-PS in the 5×4 MIMO system.

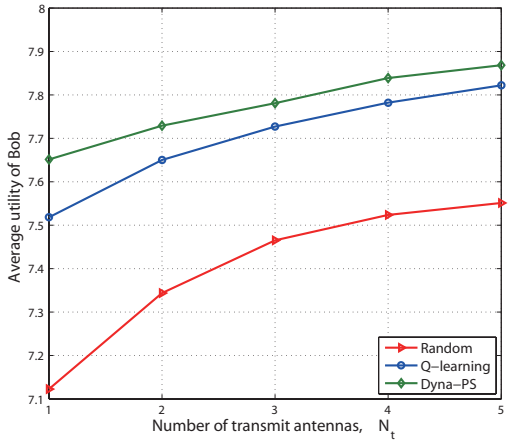
Fig. 8 shows that the miss detection rate with $N_r = 3$ decreases from 5.1% to 3.2%, and the false alarm rate decreases from 0.3% to 0.2%, if there are 4 transmit antennas instead of 2. The detection with Dyna-PS is more accurate, e.g., the miss detection rate is 2.3%, which is 28.1% lower than the Q-learning based scheme in the 4×3 MIMO system. Consequently, as shown in Fig. 8 (b), the average utility of Bob increases with the number of transmit antennas. For example, the average utility of Bob with the Q-learning based detection increases by 4%, if the number of transmit antennas increases from 1 to 5.

In summary, the experiment is performed in the indoor environment in Figs. 5 and 6 characterized by radio channel variations due to the movement of people in the office room. The proposed spoofing detection schemes are shown

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2652484, IEEE Transactions on Vehicular Technology



(a) Average spoofing detection error rate



(b) Average utility of Bob

Fig. 8. Average spoofing detection performance of the $N_t \times 3$ MIMO system in the experiments with topology shown in Fig. 4.

to improve the detection accuracy and thus the utility of the receiver. Similar performance can be observed in the other cases, not shown here, which demonstrates the robustness of the schemes.

IX. Conclusions

In this work, we have investigated the PHY-layer authentication in MIMO systems and presented the NEs of the static authentication game, showing that the receiver chooses its test threshold and adversary decides its attack probability based on the SNR, attack costs, channel conditions, and channel gain time variation. We have also proposed the Q-learning and Dyna-PS based spoofing detections, in which the test threshold in the spoofing detection is chosen via the reinforcement learning techniques. The USRP-based experiments have been performed, and the results demonstrate that the proposed authentication methods can efficiently improve the authentication performance even in time-variant radio environments. For example, in the MIMO system with 5 transmit antennas and 20 MHz bandwidth, the miss detection rate of the Q-learning based detection decreases by 42.9% to 1.9%, and the false alarm rate decreases by 55.6% to 0.1%, if the number of receive antennas changes from 2 to 4. The Dyna-PS based spoofing detection further reduces the miss detection rate to 1.3% in the 5×4 MIMO system. The utility of the receiver with the Q-learning based authentication is 2.6% higher than the benchmark scheme, which is further improved by 3.1% with Dyna-PS.

References

[1] L. Xiao, T. Chen, G. Han, Y. Li, W. Zhuang, and L. Sun, "Channel-based authentication game in MIMO systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Washington, DC, Dec. 2016.

[2] K. Zeng, K. Govindan, and P. Mohapatra, "Non-cryptographic authentication and identification in wireless networks," *IEEE Trans. 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 and Distributed Systems*, vol. 24, no. 1, pp. 44–58, Jan. 2013.

[4] F. J. Liu, X. Wang, and S. L. Primak, "A two dimensional quantization algorithm for CIR-based physical layer authentication," in *Proc. IEEE Int'l Conf. Commun. (ICC)*, pp. 4724–4728, Budapest, Jun. 2013.

[5] J. Chen, Q. Yu, P. Cheng, Y. Sun, Y. Fan, and X. Shen, "Game theoretical approach for channel allocation in wireless sensor and actuator networks," *IEEE Trans. Automatic Control*, vol. 56, no. 10, pp. 2332–2344, Aug. 2011.

[6] L. Xiao, Y. Li, G. Han, G. Liu, and W. Zhuang, "PHY-layer spoofing detection with reinforcement learning in wireless networks," *IEEE Trans. Vehicular Technology*, vol. 65, no. 12, pp. 10037–10047, Feb. 2016.

[7] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3, pp. 279–292, May 1992.

[8] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, Sep. 1998.

[9] A. W. Moore and C. G. Atkeson, "Prioritized sweeping: Reinforcement learning with less data and less time," *Machine Learning*, vol. 13, no. 1, pp. 103–130, Oct. 1993.

[10] L. Xiao, L. Greenstein, N. Mandayam, and W. Trappe, "MIMO-assisted channel-based authentication in wireless networks," in *Proc. IEEE Information Sciences and Systems (CISS)*, pp. 642–646, Princeton, Mar. 2008.

[11] H. Wen, S. Li, X. Zhu, and L. Zhou, "A framework of the PHY-layer approach to defense against security threats in cognitive radio networks," *IEEE Network*, vol. 27, no. 3, pp. 34–39, May 2013.

[12] P. Baracca, N. Laurenti, and S. Tomasin, "Physical layer authentication over MIMO fading wiretap channels," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2564–2573, Jul. 2012.

[13] C. Chen, M. Song, C. S. Xin, and J. Backens, "A game-theoretical anti-jamming scheme for cognitive radio networks," *IEEE Netw.*, vol. 27, no. 3, pp. 22–27, May 2013.

[14] Y. E. Sagduyu and A. Ephremides, "A game-theoretic analysis of denial of service attacks in wireless random access," *Wireless Netw.*, vol. 15, no. 5, pp. 651–666, Apr. 2009.

[15] A. Mukherjee and A. L. Swindlehurst, "Optimal strategies for countering dual-threat jamming/eavesdropping-capable adversaries in MIMO channels," in *Proc. IEEE Military Commun. Conf. (MILCOM)*, pp. 1695–1700, San Jose, Oct. 2010.

[16] B. Wang, Y. Wu, K. J. R. Liu, and T. C. Clancy, "An anti-jamming stochastic game for cognitive radio networks," *IEEE J. Sel. Areas in Commun.*, vol. 29, no. 4, pp. 877–889, Apr. 2011.

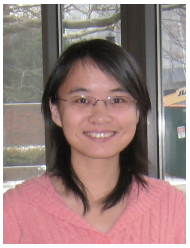
[17] B. F. Lo and I. F. Akyildiz, "Multiagent jamming-resilient control channel game for cognitive radio ad hoc networks," in *Proc. IEE Int'l Conf. Commun. (ICC)*, pp. 1821–1826, Ottawa, Jun. 2012.

[18] R. Lu, X. Lin, H. Zhu, X. Liang, and X. Shen, "BECAN: A bandwidth-efficient cooperative authentication scheme for filtering injected false data in wireless sensor networks," *IEEE Trans. Parallel and Distributed Systems*, vol. 23, no. 1, pp. 32–43, Jan. 2012.

[19] D. He, C. Chen, S. Chan, and J. Bu, "Secure and efficient handover authentication based on bilinear pairing functions," *IEEE Trans. Wireless Communications*, vol. 11, no. 1, pp. 48–53, Jan. 2012.

[20] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Dover, 1964.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2017.2652484, IEEE Transactions on Vehicular Technology



Liang Xiao (M'09, SM'13) received the B.S. degree in communication engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China; the M.S. degree in electrical engineering from Tsinghua University, Beijing, China; and the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 2000, 2003, and 2009, respectively. She is currently a Professor with the Department of Communication Engineering, Xiamen University, Fujian, China. Her current research interests include smart grids, network security, and wireless communications.

Tianhua Chen (S'16) received the B.S. degree in communication engineering from Xiamen University, Xiamen, China, in 2014, where she is currently pursuing the M.S. degree with the Department of Communication Engineering. Her research interests include network security and wireless communications.

Guoan Han (S'16) received the B.S. degree in communication engineering from Southwest Jiaotong University, Chengdu, China, in 2015, where he is currently pursuing the M.S. degree with the Department of Communication Engineering. His research interests include network security and wireless communications.

Weihua Zhuang (M'93, SM'01, F'08) has been with the Department of Electrical and Computer Engineering, University of Waterloo, Canada, since 1993, where she is a Professor and a Tier I Canada Research Chair in Wireless Communication Networks. Her current research focuses on resource allocation and QoS provisioning in wireless networks, and on smart grid. She is a co-recipient of several best paper awards from IEEE conferences. Dr. Zhuang was the Editor-in-Chief of IEEE Transactions on Vehicular Technology (2007-2013), and the Technical Program Chair/Co-Chair of the IEEE VTC Fall 2017/2016. She is a Fellow of the IEEE, a Fellow of the Canadian Academy of Engineering, a Fellow of the Engineering Institute of Canada, and an elected member in the Board of Governors and VP Publications of the IEEE Vehicular Technology Society.

Limin Sun received his B.S., M.S., and D.Sc. degree in College of Computer, National University of Defense Technology in China in 1988, 1995, and 1998, respectively. He is currently a Professor in Institute of Information Engineering at Chinese Academy of Sciences, chair of Beijing Key Lab of Internet of Things Security, deputy chair of China Computer Federation Technical Committee on Internet of Things. He is also a member of IEEE and a senior member of China Computer Federation (CCF). His research interests include Internet of Things and its security, and intelligent transportation systems. He led dozens of important projects, including National 973 projects, key projects of Natural Science Foundation of Chinese, the National Major Projects and the Pilot projects of Chinese Academy of Sciences. He published 5 academic books and 150 papers on journals such as IEEE TPDS, TOSN, TMC, and on international conferences such as SENSYS, MOBISYS, MOBICOM, ICDCS, INFOCOM. He was granted more than 50 patents and software copyrights.