Physical-Layer Authentication Based on Extreme Learning Machine

Ning Wang, Ting Jiang, Shichao Lv and Liang Xiao, *Senior Member, IEEE*

***Abstract*—Most physical-layer authentication techniques use hypothesis tests to compare the radio channel information with the channel record of Alice to detect spoofer Eve in wireless networks. However, the test threshold in the hypothesis test is not always available, especially in dynamic networks. In this paper, we propose a physical-layer authentication scheme based on extreme learning machine that exploit multi-dimensional characters of radio channels and use the training data generated from the spoofing model to improve the spoofing detection accuracy. Simulation results show that our proposed technique can significantly improve the authentication accuracy compared with the state-of-the-art method.**

***Index Terms*—Physical layer authentication, Wireless security, Extreme learning machine.**

I. Introduction

PHYSICAL layer authentication (PLA) exploits the physical-layer properties of wireless channels, such as received signal strength (RSS) and channel impulse response (CIR) to detect spoofing attacks in wireless communication [1]–[3].

Consider the Alice-Bob-Eve model (i.e., a legitimate receiver Bob, legitimate transmitter Alice, and illegitimate potential transmitter Eve that impersonates another node with a fake MAC address), the problem of PLA can be formulated as binary hypothesis testing [4], [5]: Suppose Bob receives two messages at times indexed by k and $k+1$. The first message, at time k, has been confirmed that it was indeed sent by Alice, and Bob seeks to determine whether the second message, at time $k+1$, still belongs to Alice, i.e.,

1) $H_0 : \mathbf{H}(k+1) = \mathbf{H}_E(k+1)$,
2) $H_1 : \mathbf{H}(k+1) = \mathbf{H}_A(k+1)$,

where $\mathbf{H}_A$ and $\mathbf{H}_E$ denote Alice's and Eve's channel information received by Bob, respectively. The null hypothesis H_0 represents that the real transmitter, at times $k+1$, is Eve, while the alternative hypothesis H_1 indicates that the message, at times $k+1$, is indeed sent by Alice.

Past approaches have explicitly utilized a threshold segmentation method to distinguish whether the channel characteristics of current transmissions is the same as that of the previous one [4]–[7]. In these techniques, based on the Neyman-Pearson theory, a suitable threshold is selected to achieve the detection in the hypothesis test in which the performance of the detection depends on this threshold. Particularly, [4] presented a dynamic threshold selection scheme that uses reinforcement learning techniques and games theory to obtain the optimal test threshold for dynamic wireless networks. However, even when the most optimal threshold is selected, the detection performance of these thresholding methods is still limited, especially when the correlation of legitimate channels is low, where the measurement performance of H_1 is easily intertwined with that of H_0. This makes them unsuitable for some specific occasions, such as high-speed mobile, where the channel correlation coefficient of two successive legitimate channels may be small.

Without requiring a threshold, a classification algorithm based on machine learning can consider multiple features to achieve a remarkable detection performance. Particularly, extreme machine learning (ELM), a feedforward neural network for classification or regression, can offer significant advantages such as fast learning speed, ease of implementation and minimal human intervention [8]. Because the hidden nodes in ELM can be randomly generated and weight of the output solved by a least square solution, a unique optimal solution can be obtained with minimal human intervention. In contrast to multiple-layer neural networks or deep networks, the calculation in ELM does not need iteration. Hence, the training and calculation time is greatly reduced. Therefore, due to the simple structure and fast performance, ELM has strong potential as a viable detection technique in a low-complexity scenario.

In this paper, we apply ELM to investigate the PLA in which a two-dimensional measure space is considered and pseudo adversary model is used to generate the training data in terms of a legitimate user message. Furthermore, the performance of our proposed scheme is verified via simulations. The contributions of this work lie in two aspects: (1) Utilizing a machine-learning classification algorithm to achieve PLA is investigated. The advantages of multidimensional measure are used to improve the detection performance; (2) A pseudo adversary model is proposed to obtain the training data for the ELM-based authentication, which can solve the problem of the application of the supervised learning algorithm in practice.

The rest of this paper is organized as follows: Section II details the proposed authentication scheme based on ELM,

N Wang and T Jiang are with the Key Laboratory of Universal Wireless Communication Ministry of Education, Beijing University of Posts and Telecommunications. (Emails: wangning8566@bupt.edu.cn, Tjiang@bupt.edu.cn).

Shichao Lv is with the Beijing Key Laboratory of IOT Information Security Technology, IIE CAS. (Email: lvshichao@iie.ac.cn).

L. Xiao is with Dept. Communication Engineering, Xiamen University, 361005 China. (E-mail: lxiao@xmu.edu.cn).

This work was supported in part by National Natural Science Foundation of China under Grant 61671396 and CCF-Venustech Hongyan Research Initiative (2016-010).

Manuscript received April 19, 2005; revised August 26, 2015.

and Section III shows the simulation results, followed by the conclusion in Section IV.

II. THE PROPOSED AUTHENTICATION SCHEME

A. Measure space

To enhance the distinguishable of the physical layer properties of wireless channels, a two-dimensional measure space is established, which contains two different metrics. Suppose $\mathbf{S}_A(k)$ and $\mathbf{S}_\Delta(k+1)$ are sampling from $\mathbf{H}_A(k)$ and $\mathbf{H}_\Delta(k+1)$, respectively, where $\Delta \in \{A, E\}$. The Euclid distance is employed as the first metric:

$$D(\mathbf{S}_A(k), \mathbf{S}_\Delta(k+1)) = \|\mathbf{S}_A(k) - \mathbf{S}_\Delta(k+1)\|^2, \quad (1)$$

where $\|\cdot\|$ is the Frobenius norm, $\mathbf{S}_A(k)$ denotes Alice's signal at time k and $\mathbf{S}_\Delta(k+1)$ is the signal sent by Alice or Eve at time $k+1$. Next, the Pearson correlation coefficient is considered as the second metric, given by

$$R(\mathbf{S}_A(k), \mathbf{S}_\Delta(k+1)) = \frac{\sum\limits_{i=1}^{n} \left(S_{Ai}(k) - \overline{\mathbf{S}_A(k)}\right)\left(S_{\Delta i}(k+1) - \overline{\mathbf{S}_\Delta(k+1)}\right)}{\sqrt{\sum\limits_{i=1}^{n} \left(S_{Ai}(k) - \overline{\mathbf{S}_A(k)}\right)^2}\sqrt{\sum\limits_{i=1}^{n} \left(S_{\Delta i}(k+1) - \overline{\mathbf{S}_\Delta(k+1)}\right)^2}}. \quad (2)$$

where $S_{Ai}(k)$ and $S_{\Delta i}(k)$ are the values of each part, and $\overline{\mathbf{S}_A(k)}$ and $\overline{\mathbf{S}_\Delta(k+1)}$ are the mean values respectively. Thus, based on (1) and (2), the feature measure space can be constructed as follows:

Definition 1: The feature measure space is defined as a two-dimensional characteristic space that contains Euclid distance and Pearson correlation coefficient, given by

$$F(\mathbf{S}_A(k), \mathbf{S}_\Delta(k+1)) = [\mathbf{D}, \mathbf{R}]. \quad (3)$$

B. Pseudo Adversary Model

In practical terms, the acquisition of training data is the foundation problem for a machine learning algorithm. For the legitimate users, the training data of a legitimate communication is relatively easy to obtain, such as in [4], the training data is acquired at the previous period. However, because the attackers are unknown and unpredictable, it is difficult for a legitimate user to accurately obtain the attack training data. To solve this problem, we present a pseudo adversary model to help legitimate users acquire positive and negative training data only based on its own communication data. The definition of the pseudo adversary model is presented by Definition 2, followed by Theorem 1 and Lemma 1. In addition, the pseudo adversary model is shown in Fig.1.

Definition 2 (Pseudo adversary model): We define a random sequence $\mathbf{S}_E^{(P)}(k+1)$ as the pseudo adversary model, which has the same distribution, however, is uncorrelated with $\mathbf{S}_A(k)$.

Theorem 1: Let a denote the channel correlation coefficient of two successive legitimate channels. When $0 < a < 1$, the feature of $F(\mathbf{S}_A(k), \mathbf{S}_E^{(P)}(k+1))$ can be determined by $\mathbf{S}_A(k)$ and distinguished from $F(\mathbf{S}_A(k), \mathbf{S}_A(k+1))$. Only if $a = 0$, we have $F(\mathbf{S}_A(k), \mathbf{S}_E^{(P)}(k+1)) = F(\mathbf{S}_A(k), \mathbf{S}_A(k+1))$.

Proof: First, we consider that $F(\mathbf{S}_A(k), \mathbf{S}_E^{(P)}(k+1))$ can be determined by $\mathbf{S}_A(k)$. Here, $\mathbf{S}_A(k)$ is used to derive the statistical features of $F(\mathbf{S}(k), \mathbf{S}_E^{(P)}(k+1))$. According to Definition 2, $\mathbf{S}_A(k) = S_{A1}, ..., S_{At}$ and $\mathbf{S}_E^{(P)}(k+1) = S_{E1}^{(P)}(k+1), ..., S_{Et}^{(P)}(k+1)$ are with the same distribution function, i.e., $\mathbf{S}_A(k)$ and $\mathbf{S}_E^{(P)}(k+1) \sim N_t(\mu, \sigma^2)$ [5]. From Definition 1, we know that $F(\mathbf{S}(k), \mathbf{S}(k+1))$ includes two metrics: Euclid distance and Pearson correlation coefficient, herein, each will be discussed, respectively. For the first metric, because $\mathbf{S}_A(k)$ and $\mathbf{S}_E^{(P)}(k+1)$ are two independent normal distributions, $Z_i = S_{Ai}(k) - S_{Ei}^{(P)}(k+1)$ follows a bivariate random normal distribution, i.e., $\mathbf{Z} \sim N_t(0, 2\sigma^2)$. Thus, the statistic, (i.e., Euclidean distance)

$$Q_E^{(P)} = \sqrt{\sum_{i=1}^{t} \left(\frac{Z_i}{2\sigma_i}\right)^2}, \quad (4)$$

is distributed according to the chi distribution with t degrees of freedom, and the probability density function is

$$f(Z_E^{(P)}; t) = \frac{2^{1-\frac{t}{2}} Z^{t-1} e^{-\frac{Z^2}{2}}}{\Gamma(\frac{t}{2})}, \quad (5)$$

where $\Gamma(*)$ is the Gamma function. Then, the expectation and variance value of $Q_E^{(P)}$ can be calculated, i.e.,

$$\begin{aligned} E(Q_E^{(P)}) &= \sqrt{2}\tfrac{\Gamma((t+1)/2)}{\Gamma(t/2)}, \\ \mathrm{var}(Q_E^{(P)}) &= t - E(Q_E^{(P)})^2. \end{aligned} \quad (6)$$

For the second metric, because $\mathbf{S}_A(k)$ and $\mathbf{S}_E^{(P)}(k+1)$ are uncorrelated, we can consider that the true Pearson correlation coefficient $\rho_E^{(P)}$ is equal to 0, and the sample Pearson correlation coefficient is $R_{E1}^{(P)}, ..., R_{En}^{(P)}$, where n is the sample number. Then, the distribution of $R_E^{(P)}$ can be written as [9]

$$f(R_E^{(P)}) = \frac{\left(1 - \left(R_E^{(P)}\right)^2\right)^{\frac{t-4}{2}}}{B(\frac{1}{2}, \frac{t-2}{2})}, \quad (7)$$

where $B(*)$ is the beta function. In practice, the expectation and variance value of $R_E^{(P)}$ can be described by

$$\begin{aligned} E(R_E^{(P)}) &= arctanh(\rho_E^{(P)}) \to 0, \\ \mathrm{var}(R_E^{(P)}) &= \tfrac{1}{n-3}, \end{aligned} \quad (8)$$

where $arctanh(*)$ is the Fisher transformation.

As a result, according to (4)-(8), the statistical characteristic of $F(\mathbf{S}_A(k), \mathbf{S}_E^{(P)}(k+1))$ can be calculated based on $\mathbf{S}_A(k)$.

Subsequently, the distinguishability of $F(\mathbf{S}_A(k), \mathbf{S}_E^{(P)}(k+1))$ from $F(\mathbf{S}_A(k), \mathbf{S}_A(k+1))$ is considered. We note that for

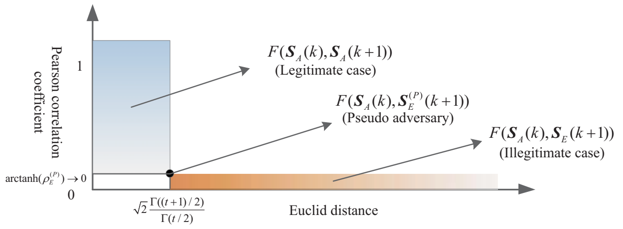

Pearson correlation coefficient
1
0
arctanh($\rho_E^{(P)}$) → 0
$\sqrt{2}\frac{\Gamma((t+1)/2)}{\Gamma(t/2)}$
Euclid distance
$F(\boldsymbol{S}_A(k), \boldsymbol{S}_A(k+1))$
(Legitimate case)
$F(\boldsymbol{S}_A(k), \boldsymbol{S}_E^{(P)}(k+1))$
(Pseudo adversary)
$F(\boldsymbol{S}_A(k), \boldsymbol{S}_E(k+1))$
(Illegitimate case)

Fig. 1 Illustration of the performance of the pseudo adversary model, legitimate case and illegitimate case in the measure space.

the condition of identical distribution, the feature of the case of two uncorrelated sequences is stable, which can be determined according to (6) and (8). However, for the condition of two related sequences, the statistical characteristic of the measure space depends on the parameter a. Let Q_A and ρ_A denote the Euclid distance and Pearson correlation coefficient between $\mathbf{S}_A(k)$ and $\mathbf{S}_A(k+1)$, respectively. Obviously, if $0 < a < 1$, we have $E(\rho_A) > E(\rho_E^{(P)})$ $(E(\rho_A) \to a$ while $E(\rho_E^{(P)}) \to 0)$ and $E(Q_A) < E(Q_E^{(P)})$ $(E(Q_A) < \sqrt{2}\frac{\Gamma((t+1)/2)}{\Gamma(t/2)}$ while $E(Q_E^{(P)}) = \sqrt{2}\frac{\Gamma((t+1)/2)}{\Gamma(t/2)})$. Only if $a = 0$, $\mathbf{S}_A(k)$ and $\mathbf{S}_A(k+1)$ are independent of each other, then $F(\mathbf{S}_A(k), \mathbf{S}_E^{(P)}(k+1)) = F(\mathbf{S}_A(k), \mathbf{S}_A(k+1))$, otherwise, $F(\mathbf{S}_A(k), \mathbf{S}_E^{(P)}(k+1))$ is distinct from $F(\mathbf{S}_A(k), \mathbf{S}_A(k+1))$. Thus, the proof is completed.

Lemma 1: Let us consider the illustration in Fig. 1. Under the real attacks, the upper bound of $F(\mathbf{S}_A(k), \mathbf{S}_E(k+1))$ is $F(\mathbf{S}_A(k), \mathbf{S}_E^{(P)}(k+1))$, in contrast, $F(\mathbf{S}_A(k), \mathbf{S}_E^{(P)}(k+1))$ is the lower bound of $F(\mathbf{S}_A(k), \mathbf{S}_A(k+1))$.

Proof: Similar to Theorem 1, let $\mathbf{S}_E(k+1) = S_{E1}(k+1), ..., S_{Et}(k+1)$ and $\mathbf{S}_A(k) = S_{A1}, ..., S_{At}$ are independently with different distribution, i.e., $\mathbf{S}_A(k) \sim N_t(\mu_1, \sigma_1^2)$ and $\mathbf{S}_E(k+1) \sim N_t(\mu_2, \sigma_2^2)$. Thus, $Z_i = S_{Ai}(k) - S_{Ei}(k+1)$ follows a bivariate random normal distribution, i.e., $\mathbf{Z} \sim N_t(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$. Hence, the expectation value of the Euclid distance between $\mathbf{S}_E(k+1)$ and $\mathbf{S}_A(k)$ can be written as

$$E(Q_E) = \sqrt{\frac{\pi}{2}} L_{1/2}^{(1/t-1)}\left(\frac{-\lambda}{2}\right), \tag{9}$$

where $\lambda = \sum\limits_{i=1}^{t} (\frac{\mu_{1i}-\mu_{2i}}{\sigma_{1i}+\sigma_{2i}})^2$. Only if $\mu_1 - \mu_2 = 0$ and $\sigma_1 = \sigma_2$, i.e., when a non-central chi distribution becomes a central chi distribution, we have $E(Q_E) = E(Q_E^{(P)})$. Otherwise, the expectation value of the non-central chi distribution is greater than that of central chi distribution, i.e., $E(Q_E) > E(Q_E^{(P)})$.

For the second metric, the Pearson correlation coefficient, because $\mathbf{S}_E(k+1)$ and $\mathbf{S}_A(k)$ are uncorrelated, the performance of the Pearson correlation coefficient between $\mathbf{S}_E(k+1)$ and $\mathbf{S}_A(k)$ is the same as that between $\mathbf{S}_E^{(P)}(k+1)$ and $\mathbf{S}_A(k)$. As a result, we have $E(Q_E) \geq E(Q_E^{(P)}) \geq E(Q_A)$ and $E(\rho_A) \geq E(\rho_E^{(P)}) = E(\rho_E) \to 0$, and the equality is sustained, if and only if, $a = 0$, $\mu_1 = \mu_2$, and $\sigma_1 = \sigma_2$. Thus, Lemma 1 can be confirmed.

C. ELM-based PLA authentication

In this subsection, ELM-based PLA is presented, and the algorithm process is given in Algorithm 1.

For an ELM, the input data can be mapped to an L-dimensional ELM random feature space, and the network output is

$$f_L(\mathbf{x}) = \sum_{i=1}^{L} \beta_i \theta_i(\mathbf{x}) = \theta(\mathbf{x})\beta, \tag{10}$$

where $\beta = [\beta_1, ..., \beta_L]^T$ is the output weight matrix between the hidden nodes and output ones. $\theta(\mathbf{x}) = [g_1(\mathbf{x}), ..., g_L(\mathbf{x})]$ are the hidden node outputs (i.e., random hidden features) for input $\mathbf{x}$, and $g_i(\mathbf{x})$ is the output of ith hidden node. Given N training samples $\{(x_i, t_i)\}_{i=1}^{N}$, the ELM can resolve the following learning problem:

$$\mathbf{\Theta}\beta = \mathbf{T}, \tag{11}$$

where $\mathbf{T} = [t_1, ..., t_N]^T$ are target labels, and $\Theta = [\theta^T(x_1), ..., \theta^T(x_N)]^T$. The output weights β can calculated from

$$\beta = \mathbf{\Theta}^{\dagger}\mathbf{T}, \tag{12}$$

where $\Theta^{\dagger}$ is the Moore-Penrose generalized inverse of the matrix Θ.

To be specific, in our authentication scheme, with $F(\mathbf{S}_A(k), \mathbf{S}_{\Delta}(k+1))$ as the detection object, the detection contains two patterns: legitimate case $F(\mathbf{S}_A(k), \mathbf{S}_A(k+1))$ and illegitimate case $F(\mathbf{S}_A(k), \mathbf{S}_E(k+1))$, which are corresponding to binary hypothesis testing H_1 and H_0 respectively. Thus, based on training data, Bob can establish the design matrix $\mathbf{X}_{TR} \leftarrow [\mathbf{S}_A(k), \mathbf{S}_{\Delta}(k+1)]$ and response variables $\mathbf{Y}_{TR} \leftarrow F(\mathbf{S}_A(k), \mathbf{S}_{\Delta}(k+1))$. Then, the ELM training is

$$\beta^{(ELM)} = g(\mathbf{W}\mathbf{X}_{TR})^{+}\mathbf{Y}_{TR}, \tag{13}$$

where g is the activation function, $\mathbf{W}$ is the matrix of input-to-hidden-layer weights filled by Gaussian random noise, $(*)^{+}$ means the operation of pseudoinverse. Based on (13), the predicted value $\mathbf{Y}_{PR} \in \{Alice, Eve\}$ can be calculated according to

$$\mathbf{Y}_{PR} = \beta^{(ELM)} g(\mathbf{W}[\mathbf{S}_A(k), \mathbf{S}_{\Lambda}(k+1)]), \tag{14}$$

where $\Lambda \in \{A, E\}$ is the true identity of the current message.

Without loss of generality, we assume that the receiver obtains M packets in each time slot. The spoofing detection process is shown in detail in Algorithm 1.

Algorithm 1 ELM-based PHY-layer authentication

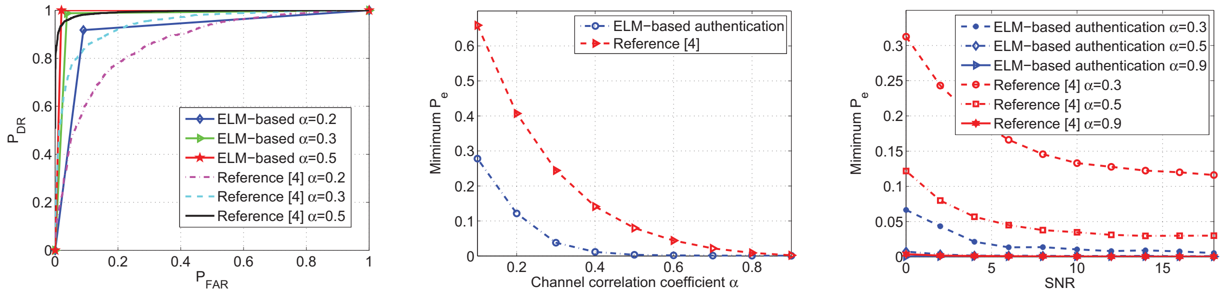
```
Initialization: Activate function $g$ and the hidden nodes number $L$.
Repeat (for each episode)
Obtain the training data based on Theory 1 and Lemma 1.
Calculate the ELM network model $\beta^{(ELM)}$ via (13).
    For $t = 1, 2, ...M$(for each packet)
      Extract $\mathbf{S}_A(t)$ and $\mathbf{S}_{\Lambda}(t+1)$.
      Calculate measure space $\mathbf{F}(\mathbf{S}_A, \mathbf{S}_{\Lambda}(t+1))$ via (3).
      Calculate the predictive value $\mathbf{Y}_{PR}$ via (14).
        If $\mathbf{Y}_{PR} = Alice$
          Update $\mathbf{S}_A(t) = \mathbf{S}_A(t+1)$, and accept this message.
        Else
          Keep $\mathbf{S}_A(t) = \mathbf{S}_A(t-1)$, and send an alarm.
        End If
    End For
End Repeat
```

III. Simulation Results

Simulations are performed to evaluate our proposed scheme relative to the thresholding method, where the reference method is presented in [4], and in which standard Euclidean distance is used.

Following [5], we can simplify two successive channel state information as

$$\mathbf{H}_A(k+1) = a\mathbf{H}_A(k) + \omega(k), \tag{15}$$



(a) P_{DR} VS. P_{FAR} at SNR=3dB (b) Minimum P_e VS. a at SNR=3dB (c) Minimum P_e VS. SNR

Fig. 2 Detection performance of PLA based on ELM with 2000 training data in the OFDM system with 256 sub-carrier under Rayleigh channel.

where a is the correlation parameter between channel gain samples, and $\omega(k)$ is an i.i.d. zero-mean complex Gaussian process, which is independent of $H_A(k)$. The variance of $\omega(k)$ is defined as

$$\sigma_\omega^2 = (1 - a^2)\sigma_A^2. \tag{16}$$

In practical terms, the channel correlation coefficient a can be identified as the term $J_0(2\pi v T/\lambda)$, where λ is the RF wavelength, ν is the moving speed of Alice and J_0 represents the Bessel function of the first kind and zero-th order [10].

In the simulations, a can be given by the channel model based on (15) and (16). Sigmoid function is considered as the activate function, and the node number of the hidden layer is 200. Moreover, the different test threshold will lead to different detection performance in PLA (such as [6] [7]). For consistency, we consider the minimum Bayes risk, i.e., $P_e^{(min)}$, as a performance indicator, given by

$$P_e^{(min)} = \min\{P(H_0\,|H_1)P(H_1) + P(H_1\,|H_0)P(H_0)\}, \tag{17}$$

where all of possible threshold traversals are considered.

In Fig. 2(a), the detection performance of authentication is represented by receiver operating characteristic curve (ROC), where P_{DR} denotes the detection rate and P_{FAR} is the false alarm rate. We can see that under different channel correlation coefficients, ELM-based PLA is closer to the top left corner, compared with the reference method. For example, $a = 0.2$, the point closest to the top-left corner of the ELM-based curve, is in the vicinity of (0.1,0.9), while that is about (0.2,0.8) under the reference method. This implies our proposed strategy can reach a low identification error.

Fig. 2(b) shows the performance of minimum Bayes risk with a larger scale of channel correlation coefficient a. It is more obvious that ELM-based PLA can dramatically reduce the minimum Bayes risk. For example, $a = 0.1$, the $P_e^{(min)}$ is less than 0.3 in the ELM-based authentication, while it is higher than 0.6 under the reference method. The minimum Bayes risk was reduced by roughly 50%. Furthermore, from Fig. 3(b), we can see that the performance of ELM-based authentication is acceptable ($P_e^{(min)} < 0.15$) when $1 \geq a \geq 0.2$.

Fig. 2(c) illustrates the fluctuation of detection performance under difference signal-noise ratios (SNRs). Under a higher channel correlation coefficient, e.g., $a = 0.9$, both curves under the ELM-based authentication and reference method become flat and the $P_e^{(min)}$ exceeds 0. However, when the channel correlation is low, e.g., $a = 0.3$, the detection performance will be influenced by noise when the SNR is less than 10dB. Nevertheless, $P_e^{(min)}$ is less than 0.1 in the ELM-based authentication, even if SNR = 0. While under the same conditions, $P_e^{(min)}$ will be higher than 0.3, based on [4].

These show that our ELM-based PLA can present a better detection performance compared with [4] in terms of the minimum Bayes risk, and the coverage area of the available performance can reach $\{a\,|0.2 \leq a \leq 1\}$ at least.

IV. Conclusion

In this paper, we have proposed an ELM-based PLA with a two-dimensional measure space. The feasibility of utilizing a classification approach to detect spoofing attacks was investigated, and a pseudo adversary model was established to acquire the training data. Our proposed scheme can significantly reduce the minimum Bayes risk compared with the previous methods.

References

[1] X. Wang, P. Hao, and L. Hanzo, “Physical-layer authentication for wireless security enhancement: current challenges and future developments,” *IEEE Communications Magazine*, pp. 1–7, 2016.

[2] L. Xiao, T. Chen, G. Han, W. Zhuang, and L. Sun, “Game theoretic study on channel-based authentication in mimo systems,” *IEEE Transactions on Vehicular Technology*, 2017.

[3] X. Wan, L. Xiao, Q. Li, and Z. Han, *Phy-layer Authentication with Multiple Landmarks with Reduced Communication Overhead*. in Proc. IEEE International Conference on Communications (ICC), Paris, May, 2017.

[4] L. Xiao, Y. Li, G. Han, G. Liu, and W. Zhuang, “Phy-layer spoofing detection with reinforcement learning in wireless networks,” *IEEE Trans. Veh. Technol*, vol. 65, no. 12, pp. 10 037–10 047, 2016.

[5] L. Xiao, L. J. Greenstein, N. B. Mandayam, and W. Trappe, “Channel-based spoofing detection in frequency-selective rayleigh channels,” *IEEE Trans. Wireless Commun*, vol. 8, no. 12, pp. 5948–5956, 2009.

[6] J. Yang, Y. Chen, W. Trappe, and J. Cheng, “Detection and localization of multiple spoofing attackers in wireless networks,” *IEEE Transactions Parallel and Distributed Syst*, vol. 24, no. 1, pp. 44–58, 2013.

[7] J. K. Tugnait, “Wireless user authentication via comparison of power spectral densities,” *IEEE J. Sel. Areas in Commun*, vol. 31, no. 9, pp. 1791–1802, 2013.

[8] G. Huang, S. Song, J. N. Gupta, and C. Wu, “Semi-supervised and unsupervised extreme learning machines,” *IEEE transactions on cybernetics*, vol. 44, no. 12, pp. 2405–2417, 2014.

[9] F. Kenney and E. Keeping, “Mathematics of statistics-part two,” pp. 209–221, 1951.

[10] W. C. Jakes and D. C. Cox, *Microwave mobile communications*. Wiley-IEEE Press, 1994.