Physical Layer Authentication in Wireless Communication Networks: A Survey

Lin Bai, Lina Zhu, Jianwei Liu, Jinho Choi, Wei Zhang

***Abstract*—Physical layer security (PLS) in wireless communication systems has attracted extensive research attentions in recent years. Unlike cryptography-based methods applied in upper-layer in network, PLS methods are applied in physical layers and can provide information-theoretic security by utilizing the randomness of signals and wireless channels. In this survey, we provide a comprehensive review in the domain of physical layer authentication (PLA) in wireless communication systems, including the concepts, several key techniques of typical PLA architectures as well as future challenges and research trends in more sophisticated communication systems. The survey begins with an overview of the background and basic concepts of PLA, such as the general model of wireless security communication system, typical frameworks of key-based/less PLA systems, and the common attack models. We then discuss the major concerns and key techniques that are applied in PLA systems, where three types of authentication schemes are considered, i.e., the authentication based on channel information, radio-frequency and identity watermarks. Basic models and representative research results about key approaches and techniques applied to the authentication systems above are subsequently covered. Finally, the associated challenges and potential research trends of PLA in future communication systems are presented at the end of the survey paper.**

Manuscript received Jul. 09, 2020; revised Aug. 12, 2020; accepted Aug. 13, 2020. This work was supported by National Key Research and Development Program of China (Grant No. 2017YFB0503002) and National Natural Science Foundation of China (Grant No. 61922010). The associate editor coordinating the review of this paper and approving it for publication was W. C. Cheng.

L. Bai, J. W. Liu. School of Cyber Science and Technology, Beihang University, Beijing 100191, China (e-mail: l.bai@buaa.edu.cn; liujianwei@buaa.edu.cn).

L. Zhu. School of Electronic and Information Engineering, Beihang University, Beijing 100191, China (e-mail: zhulina@buaa.edu.cn).

J. Choi. School of Information Technology, Deakin University, Burwood, VIC 3125, Australia (e-mail: jinho.choi@deakin.edu.au).

W. Zhang. School of Electrical Engineering and Telecommunications, The University of New South Wales, Sydney, NSW 2052, Australia (e-mail: w.zhang@unsw.edu.au).

***Keywords*—physical layer authentication (PLA), radio-frequency identification, hypothesis testing, wireless communication**

I. Introduction

A. Physical Layer Security in Wireless Communication Systems

While various of efficient wireless transmission techniques have been proposed to meet the demand of high throughput and reliability in communication, the security has also become a critical issue as people sometimes have to transmit important/private information in wireless networks (such as in unmanned aerial vehicle networks). It is known that the open and broadcasting nature in wireless environments result in the vulnerability in secure communication, so that eavesdropping and impersonation attacks can be carried out more easily in conventional communication networks. Therefore, dedicated methods are required for secure wireless communication.

As shown in Tab. 1, there are two major security concerns in wireless communication networks: confidentiality and authentication. Confidentiality prevents the secret plaintext from being obtained by eavesdroppers, while authentication verifies the identities of users and makes sure that illegitimate users cannot impersonate as legitimate users do. Traditional communication security is usually achieved by the upper-layer security protocol stack using cryptography-based methods, such as symmetric and asymmetrical cryptographic methods, message authentication code (MAC) or digital signature, etc.[8,9] The security of cryptography-based methods relies heavily on the computational complexity and secret keys. However, the development of future Internet of things (IoT) techniques accommodate more limited devices in terms of power and computation resources, which makes it impractical to implement complicated cryptography-based security protocol.

As an alternative security mechanism that compensates the restriction of upper-layer security protocols, physical layer security (PLS) is considered to validate security wireless transmission on the physical layer. By exploring the unique statistical characters of physical channels, PLS-based transmission can support both confidentiality and authentication in the

Table 1 Some of the survey papers that mention about PLA

Works	Brief introduction
[1,2]	Physical-layer identification techniques that are based on unique RFs of different transmitters
[3]	Mainly concentrates on available device features, including feature extracting and classification techniques, and the fingerprinting algorithms in RF fingerprint-based PLA
[4]	Cross-layer PLA design is proposed after briefly reviewing the existing techniques and limitations of current PLA
[5,6], and [7]	Put a major effort on the general and comprehensive review of the confidentiality of PLS, i.e., the secrecy capacity analysis in different transmission scenarios

viewpoint of information-theoretic security. Rather than depending on computational complexity of hard mathematical problems, the reliability of PLS lies on the variation and randomness of wireless channels to limit the information of secret message extracted by attackers. The confidentiality of PLS was firstly considered in 1949 by Shannon, who proposed the first application of information theory to cryptology, which is also known as Shannon's information theoretic secrecy[10]. Then the information theoretic analysis for secure transmissions over insecure channels (wiretap channels) was studied in Ref. [11] by Wyner in 1975, whose approach was further generalized by Csiszar and Korner[12]. One of the most important targets of the physical layer confidentiality studies is to maximize the secret information rate received by the legitimate user in the wiretap channel, which is defined as the secrecy capacity by Wyner[11], with certain constraints on the information attainability to an eavesdropper. Based on the approaches introduced in Refs. [11,13], a number of research works and surveys that aim to implement secure communications in various of systems are presented in the following decades[14,15].

As another key aspect of PLS, physical layer authentication (PLA) plays a nonnegligible role in wireless security communications. In this survey we mainly focus on the concepts, state-of-the-art techniques and future challenges of identity/user authentication in wireless physical layer, which is not only because there are limited systematic works and surveys that concentrate on this field, but also we aim at emphasizing the unique contributions of PLA to the final PLS in the entire wireless communication network.

B. Research Issues and Significance of PLA

Following the spirit of Shannon's work, a series of analysis on information-theoretically secure authentication was studied in Refs. [16-19]. Here, Ref. [16] was considered as the first lower bound results in message authentication, based on which, different information-theoretic lower bounds in authentication theory were studied by Refs. [17,18]. Then in 2000, Maurer proposed a more generalized and simplified cheating probability lower bound for any authentication system by introducing the application of hypothesis testing in authentication[19]. As stated by Maurer, the inherent nature of authentication can be regarded as a binary hypothesis testing problem, i.e., deciding whether a received message is from an authentic transmitter or not. The testing result is obtained by considering the joint probability distribution of the received message and the secret key, which is simply based on the reality that the eavesdropper has no priori knowledge of the secret key. Although the work in Ref. [19] was based on the authentication with secret keys, the idea of hypothesis was generalized, modified and widely applied to various of communication systems afterwards, such as conventional single antenna point-to-point communications[20-24], multiple-input multiple-output systems[25-27], multi-carrier systems[22,23,26-33], 5G systems with handover transmission[34], distributed ad hoc wireless networks[35], etc. The result of hypothesis testing in authentication is usually related to a single value that helps to make a decision by comparing with a preset threshold. How to decide a threshold is also one of the major topics in hypothesis testing based authentication to achieve a good balance between the false alarm and the missing detection rates. The detailed mechanism of hypothesis testing will be introduced in the following sections.

The mostly studied authentication techniques can be classified into two categories: channel-based and radio frequency (RF) fingerprint-based schemes. Channel-based authentication schemes verify the identity of an unknown transmitter by either observing the unique instantaneous/average characteristics of the estimated channel state information (CSI)[28,32,36-38] or comparing the current CSI to the authentic CSI previously reserved during the past transmission[21,27,33,35,39]. The RF fingerprint-based schemes identify a user/device according to unique features of their transmitted waveform, which is also referred to as physical-layer device identification[1]. The works on this field mainly focus on how to extract intended features from the received signals[40] and improve the authentication accuracy by carefully selecting and classifying the features in hand[41]. Besides of the two categories of authentication techniques above, there are also some other authentication schemes proposed, such as the watermark/fingerprint embedding[42-44], multi-attribute multi-observation (MAMO) techniques[4], and so on.

There are already several survey papers that provide a general/partial view on PLA[1-7]. In Refs. [1,2], the authors mainly presented systematic reviews and a summary of physical-layer identification techniques that were based

on unique RFs of different transmitters, while in Ref. [4], a novel MAMO technique with cross-layer PLA design was proposed after briefly reviewing the existing techniques and limitations of current PLA. In Ref. [3], the authors provided a detailed survey of features that can be adopted in wireless device fingerprinting, including the characteristics of available device features, widely used feature extracting and classification techniques, and the fingerprinting algorithms in RF fingerprint-based PLA. Refs. [5-7] put a major effort on the general and comprehensive review of the confidentiality of PLS, i.e., the secrecy capacity analysis in different transmission scenarios. Unfortunately, only limited research results on PLA are summarized and presented in the existing PLS survey papers, which becomes one of the motivations for the appearance of this survey.

C. Criteria of PLA

It is traditionally acknowledged that the goal of authentication is to verify the identity of an entity. For PLA schemes specifically, it is more focused on distinguishing different transmitters. Generally, an effective authentication scheme must meet three properties[45,46]: covertness, robustness, and security. The covertness means that any authentication schemes should not significantly affect the performance of the normal data transmission. It would be better if the PLA application does not occupy too much communication overheads or extra computational resources. Such covertness requests also guarantee that a PLA system framework does no harm to the existing conventional higher-layer cryptographic-based techniques. Robustness requires that the PLA framework is robust enough to mitigate channel fading and noise interference. In other words, the authentication performance should not be severely degraded due to channel dynamics and impairment. Finally, the security is the kernel of PLA systems which represents the ability to prevent the authentication procedure from being interrupted or invaded by eavesdroppers. It is not trivial to achieve all of these three targets when designing a practical authentication scheme, so most of the existing researches on PLA only focus on security. Although there are some efforts on achieving the other two properties[44], still limited works have been done on this topic.

The criteria mentioned above are general but cannot be used as design metrics for PLA systems. Instead, to evaluate the performance (or to be more specific, the accuracy) of a specific authentication framework, detailed metrics are introduced and compared in some PLA frameworks to evaluate the system performance in a more general way. Actually, in early studies about RF fingerprint-based physical-layer device identification[47,48], the authentication procedure is inherently similar to biometric identification systems[49]. Therefore, the metrics established and generally used in classification problems of machine learning can also be used to evaluate PLA systems. According to Ref. [1], the error rates should include the false reject rate (FRR) and the false accept rate (FAR). The former, which is also known as the false alarm rate or type I error, is defined as the probability that the receiver mistakes the legitimate authentication message as non-authentic in the hypothesis testing. The latter is also referred to as the missed detection rate (type II error), which is the probability of successful attack from adversaries (i.e., the chances that a spoofing message coming from the attacker is accepted as authentic). In some works, this metric is replaced by the detection rate or the authentication rate, which is actually the complementary of the missed detection rate. These two types of errors are significantly affected by the design of hypothesis testing schemes, and sometimes conflict with each other. Therefore, a careful authentication design is needed to achieve a good balance between these two metrics subject to performance constraints. However, good authentication and hypothesis testing schemes can still make both false alarm and missed detection errors as lower as possible.

In the security analysis of a certain PLA strategy, it is usually required to make comparison between different authentication systems. One basic and commonly applied idea is to analyze the missed detection and false alarm rates in a receiver operating characteristic (ROC) curve, which shows the FRRs at different FAR levels. Generally, the point where FAR and FRR are equal in ROC is referred to the operating point of ROC, which is named as the equal error rate (EER). The EER is a commonly used metric that can evaluate the priority of a designed classification method or authentication scheme in the range of ROC operating points encompassing the EER. Although the ROC and EER can be generally used as performance metrics of different identification systems, from the perspective of effective authentication schemes designing, minimizing one of the two types of errors (e.g., the missed detection rate) in the constraints of the other metrics (e.g., the false alarm rate) is generally considered.

D. The Paper Structure

The remaining of this paper is organized as follows. Section II presents general architectures of PLA systems, including the security models of key-based/key-less PLA approaches and basic ideas of common attack models. In section III, we investigate some of the major research results on the key techniques in three typical PLA architectures, i.e., channel-based, RF fingerprint-based and watermark embedding-based authentication. Future research issues and challenges are introduced in section IV, and the survey paper is concluded in section V.

The structure and the main contents of this paper are illustrated in Tab. 2.

Table 2 The structure of the survey

Sections	Subsections
Section II: Basic Security Model in PLA	II.A Generalized Wireless Security Communication II.B Key-Based PLA II.C Key-Less PLA II.D Attack Model
Section III: Key Techniques in PLA	III.A Channel-Based PLA III.B RF Fingerprint-Based PLA III.C Watermark Embedding-Based PLA
Section IV: Further Applications and Research Trends of PLA	IV.A Multiuser Communication Networks IV.B The Internet of Things Network

II. Basic Security Model in PLA

Without loss of generality, PLA methods can be roughly classified into two groups: key-based and key-less, according to whether the secret key is used for authentication. We provide basic models and principles of these two groups of PLA methods in this section, based on which some representative research results are briefly introduced.

A. Generalized Wireless Security Communication

For the sake of simplicity, we first introduce a typical wireless communication model that is general and adaptable in most of security communication scenarios.

As illustrated in Fig. 1, there are three different parties: Alice, Bob and Eve in the transmission system, which are borrowed from the conventional terminology of the security community. Alice (A) is a legitimate transmitter and Bob (B) is a receiver, while Eve (E) is an adversary who tries to impersonate the communication between Alice and Bob by pretending that she is the legitimate transmitter. Denote by $\boldsymbol{H}_{\mathrm{AB}}$, $\boldsymbol{H}_{\mathrm{AE}}$, and $\boldsymbol{H}_{\mathrm{BE}}$ the channel coefficients between Alice-Bob, Alice-Eve, and Eve-Bob, respectively. Here, we use the matrix $\boldsymbol{H}$ to represent the equivalent channels for a specific communication scenerio, which can either be an orthogonal frequency division multiplexing (OFDM) system or a MIMO system. Denoting by $\boldsymbol{x}$ the authentication signal vector transmitted from Alice, then the received signal at Bob can be represented by

$$\boldsymbol{y} = \boldsymbol{H}_{\mathrm{AB}}\boldsymbol{x} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{n} = [n_0 \cdots n_{L-1}]^{\mathrm{T}} \sim \mathscr{CN}(\boldsymbol{0}, N_0\boldsymbol{I})$ is the background additive white Gaussian noise (AWGN) vector.

For an OFDM system with multipath environment, if perfect OFDM transmission is assumed, the waveforms between each subcarrier can be viewed as "mutually orthogonal" at the receiver even they undergo multipath fading channels (after insertion and deletion of the CP). So one can describe the OFDM system as a set of L parallel fading channels. Let the lth element of $\boldsymbol{x}$ be the transmitted OFDM symbol by the subcarrier l, and $\boldsymbol{H} = \mathrm{diag}(H_0, \cdots, H_{L-1})$ represents the frequency-domain channel matrix, where H_l is the channel coefficient of subcarrier l, which is given by

$$H_l = \sum_{p=0}^{P-1} h_p \mathrm{e}^{-\frac{\mathrm{j}2\pi lp}{L}}, \quad l = 0, \cdots, L-1. \tag{2}$$

Here, h_p is the channel coefficient of the pth path and P is the length of channel impulse response (CIR).

For a MIMO system, $\boldsymbol{x}$ is the transmit vector, and $\boldsymbol{H}$ represents the channel matrix between the transmitter and the receiver, where the (i, j)th element is the channel coefficient between the transmit antenna j and the receive antenna i. The single-carrier single-input single-output (SISO) system is not mainly considered in this article, as it is just a special case of the above two communication scenarios.

B. Key-Based PLA

1) Authentication by One-Way Transmission: General key-based PLA frameworks are similar to traditional symmetric cryptography or digital signature based systems. The studies on key-based authentication methods appeared in around 1970s[16], where it is aimed to provide proper secret keys by coding schemes to achieve authentication. Then a series of information-theoretic lower bounds in authentication theory were derived in the 1980s by Simmons[50], who provided a more rigorous bound analysis on the probability of successful impersonation over a noiseless channel. It is also confirmed by Simmons that schemes used in authentication theory just aim to achieve uniform spread of the altered messages over the transmitted message set, which can be surely realized by proper channel coding. Therefore, spread spectrum coding[51] and code division multiple access (CDMA) techniques[52] for PLS are exploited decades later. Simmons' bound on impersonation was further modified and enhanced by Johannesson et al. in Ref. [53] who considered the dependence between the message and the encoding rule. The authors in Ref. [19] provided a more generalized key-based authentication scheme by introducing the concept of hypothesis testing. Although the work is based on message authentication, the idea of hypothesis testing-based PLA framework is modified and widely applied to various of communication systems.

A typical key-based authentication framework is illustrated in Fig. 1. The key-based PLA procedure generally includes two phases: one is the identification association phase, when Alice generates keys, assigns identification messages to the authentication tag according to certain generation function and then sends them to Bob. The other one is the identification verification phase, during which the identity information is verified based on the received message and the shared key at Bob. More specifically, the whole authentication mechanism can be summarized as follows.

- Alice generates and sends an authentication message by using certain association/encrypting function, which is given

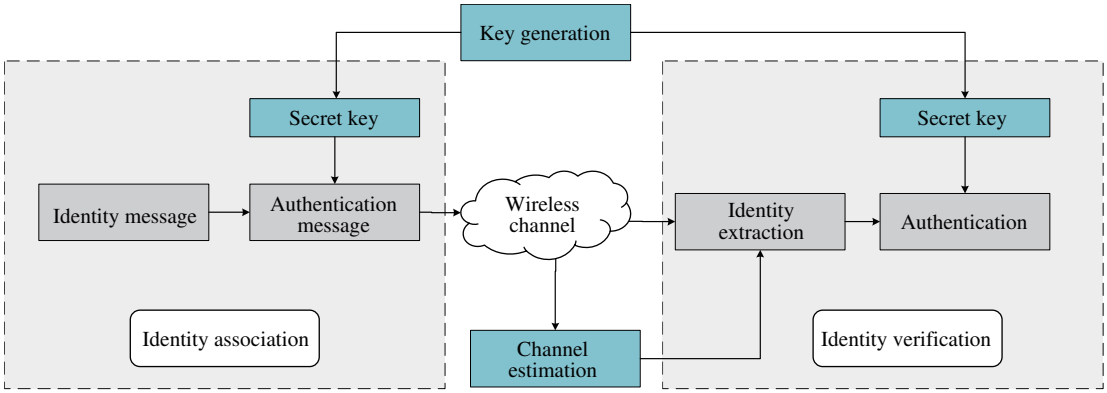


Figure 1 A typical two-phase key-based authentication framework

by $\boldsymbol{x} = \mathscr{F}(\boldsymbol{m}, \boldsymbol{k})$. Here $\boldsymbol{m}$ is the identity information determined by the transmitter and is also known or can be obtained by the receiver (such as the channel state between Alice and Bob or the unique RF information of Alice). $\boldsymbol{k}$ is the shared secret key between Alice and Bob.

- Bob receives the authentication message and extracts the identity information by using pre-designed decrypting function $\mathscr{D}$, i.e., $\boldsymbol{B} = \mathscr{D}(\boldsymbol{k}, \boldsymbol{x})$.
- Bob performs the verification as: if $\boldsymbol{B} = \boldsymbol{m}$, the transmitter is Alice. Otherwise, the authentication fails.

Note that the general authentication method above is based on the assumption that Bob can obtain precise identity information of Alice. However, due to additive noise and imperfect channel estimation, in most cases $\boldsymbol{B} \neq \boldsymbol{m}$. Therefore, binary hypothesis testing is used to decide whether the received message is from Alice or not.

As introduced in Ref. [19], binary hypothesis testing is the task of deciding which of two hypotheses, $\mathscr{H}_0$ or $\mathscr{H}_1$, is true, using test statistics (e.g., the outcome of a measurement). Let $\boldsymbol{T}$ be the identity information with error obtained by the receiver, then the hypothesis testing problem is given by

$$\mathscr{H}_0: \ \boldsymbol{T} \text{ is the estimated version of } \boldsymbol{m}; \tag{3}$$

$$\mathscr{H}_1: \ \boldsymbol{T} \text{ is NOT the estimated version of } \boldsymbol{m}. \tag{4}$$

The null hypothesis, $\mathscr{H}_0$, is accepted when the transmitter is decided to be Alice; otherwise, Bob judges the transmitter is intrusted and simply rejects the communication request. Generally, the hypothesis testing result is dependent on a log-maximum likelihood function

$$\eta = \log \frac{f(\boldsymbol{T}|\boldsymbol{m}, \boldsymbol{k})}{f(\boldsymbol{T}|\boldsymbol{m}_{\mathrm{E}}, \boldsymbol{k}_{\mathrm{E}})}, \tag{5}$$

where $f(\cdot|\cdot)$ is the conditional probability density function (PDF), and $\boldsymbol{m}_{\mathrm{E}}$ and $\boldsymbol{k}_{\mathrm{E}}$ are the spoof message and the key at Eve, respectively. η is also referred to as the test statistic that helps to make authentication decision. It can be observed that larger η indicates a higher probability that the transmitter is authentic. So the authentication system usually chooses hypothesis $\mathscr{H}_0$ if and only if $\eta \geqslant U$, where U is a pre-designed threshold. According to definitions of missed detection and false alarm rate (denoted by α and β, respectively), these two performance metrics can be calculated as

$$\alpha = \mathbb{P}(\eta > U|\mathscr{H}_1) = 1 - F_{\eta_{\mathrm{ev}}}(U), \tag{6}$$

$$\beta = \mathbb{P}(\eta < U|\mathscr{H}_0) = F_{\eta_a}(U), \tag{7}$$

where $F_{\eta_{\mathrm{ev}}}(\cdot)$ is the cumulative probability distribution function (CDF) of (5) while the received message is from Eve, and the CDF, $F_{\eta_a}(\cdot)$, represents the case when Alice is transmitting. Clearly, the setting of U has a strong impact on the probabilities of missed detection and false alarm. A larger U can efficiently decrease the missed detection probability at the cost of higher false alarm rate, and vice versa. (6) is also related to the probability distributions of the identity information $\boldsymbol{m}$ and the secret key $\boldsymbol{k}$, which is to exploit optimal key generation and authentication association schemes. Note that (5) is a theoretical test statistic which may be used in security analysis, but is not practical as it is not realistic for Bob to know the statistic properties of $\boldsymbol{m}_e$ and $\boldsymbol{k}_e$ generated by Eve. Therefore, most works consider simplified test statistics to obtain hypothesis testing results, which will be introduced in detail in the next section.

2) Authentication by Challenge-Response Transmission: The key-based PLA framework illustrated in Fig 1 is performed during single-way transmission. Another kind of widely studied key-based PLA schemes is based on traditional challenge-response architecture (which is referred to as the key-based physical layer challenge-response authentication mechanism (PHY-CRAM))[28,32,37]. The idea of key-based PHY-CRAM is similar to the authentication and key agreement protocol which has been universally used in conventional cryptographic security mechanisms. However, PHY-CRAM can achieve secure transmission by relying on the

randomness of channel characteristics. Besides, the similar mechanism of PHY-CRAM and traditional upper-layer challenge-response protocol makes it easier to consider cross-layer optimization and key submission design between PLA and cryptography-based architectures.

The basic security model of key-based PHY-CRAM is illustrated in Fig. 2 with two legitimate users, Alice and Bob, sharing a secret key. Before security transmission, Alice firstly sends a random message (which is referred to as the challenge signal) to Bob, then Bob returns back a response, which is generally the output of a pre-designed function with the inputs of the received challenge signal and the shared secret key. Once Alice receives the response signal, certain authentication schemes are used to verify the identity of Bob according to the response and the shared secret key. More specifically, the whole authentication mechanism has the following steps.

- Alice generates a random challenge message $\boldsymbol{m}$ from a message codebook $\mathscr{M}$ which can be either a number or a sequence, and sends the message to Bob.
- Bob receives the challenge message and compute the response signal according to the pre-designed function $\mathscr{F}$ and the shared secret key, i.e., $\boldsymbol{R}=\mathscr{F}(\boldsymbol{k},\boldsymbol{m})$. The response signal is sent back to Alice afterwards.
- Alice receives the response signal $\boldsymbol{R}$ and obtain the secret key information by using a decrypting function $\mathscr{D}$ and the local message $\boldsymbol{m}$, i.e., $\boldsymbol{V}=\mathscr{D}(\boldsymbol{m},\boldsymbol{R})$. In some research works, $\mathscr{F}$ is assumed to be a cryptographic hash function, in which case Alice can simply compute the local test response as $\boldsymbol{V}=\mathscr{F}(\boldsymbol{k},\boldsymbol{m})$.
- Alice performs the verification as follows. If $\boldsymbol{V}=\boldsymbol{k}$ (or $\boldsymbol{V}=\boldsymbol{R}$), the transmitter is Bob. Otherwise, the authentication fails.

According to the authentication procedure above, it can be observed that the security of key-based PHY-CRAM is ensured by the randomness of the secret key and the channel characteristics which can help to hide the challenge message and the response signal. Just as the case of single-transmission authentication scheme, the PHY-CRAM introduced above is only based on ideal transmission scenarios with perfect channel estimation and decoding. Therefore, the hypothesis testing is also used in key-based PHY-CRAM frameworks, which can be shown as

$$\mathscr{H}_0:\ \boldsymbol{V}\ \text{is the estimated version of}\ \boldsymbol{k}(\boldsymbol{R}), \tag{8}$$

$$\mathscr{H}_1:\ \boldsymbol{V}\ \text{is NOT the estimated version of}\ \boldsymbol{k}(\boldsymbol{R}). \tag{9}$$

For the case that $\boldsymbol{V}$ is the estimated version of $\boldsymbol{k}$, the test statistic can be chosen as

$$\eta=\log\frac{f(\boldsymbol{V}|\boldsymbol{m},\boldsymbol{k})}{f(\boldsymbol{V}|\boldsymbol{m}_{\mathrm{E}},\boldsymbol{k}_{\mathrm{E}})}. \tag{10}$$

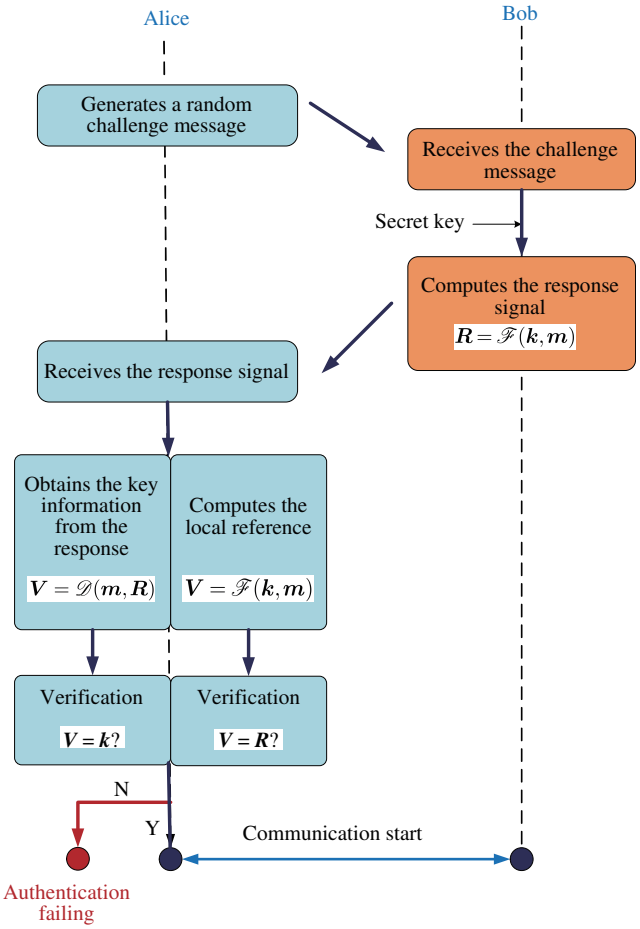


Figure 2 A basic key-based PHY-CRAM system

On the basis of the fundamental PHY-CRAM, there have already been some studies with modifying and further enhancing[28,32,36,37]. The works in Refs. [28,37] are all based on CSI and short-term channel reciprocity. More specifically, Bob should firstly estimate the CSI from the received challenge message $\boldsymbol{m}$, and then generates the response signal according to the estimated CSI and the shared key. The authors of Ref. [37] extended the work in Ref. [28] to the worst case of static channels by introducing artificial noise to conventional PHY-CRAM proposed in Ref. [28].

We have already emphasized the nonnegligible role of hypothesis testing in key-based PLA systems. Unlike the basic frameworks above, authentication can be performed without hypothesis testing. For example, the authors of Ref. [32] proposed a PHY-CRAM framework simply by utilizing channel reciprocity and verifying the identity by checking the Euclidean distance between the response and the locally generated contrast signal. In Ref. [36], it is considered channel coding based authentication schemes, where the shared key and CSI between two legitimate nodes are combined to make an adversary's attack ineffective. Detailed description of such authentication schemes is provided in the next section.

C. Key-Less PLA

A significant part of studies on channel-based authentication and almost all kinds of RF fingerprint-based authentication schemes can be classified as key-less PLA[21,27,33,35,39,54,55]. One of the first works considering the key-less channel-based authentication scheme is the architecture established by Xiao et al. in Ref. [54], where an approach based on the channel frequency response and hypothesis testing is presented to determine whether the current and prior communication attempts are made by the same user. Similar ideas are proposed in Refs. [21,26,27,33,35,39,55-58] in different wireless transmission systems. For example, in Refs. [55,57], the authors extended the approach in Ref. [54] to time-variant wireless channels, and in Refs. [35,56,58], more complicated communication scenarios such as the cellular Internet of things (CIoT) network, the ad hoc networks, and relay networks are considered to achieve channel-based authentication. The works in Refs. [21,27,33,39] utilize typical machine learning techniques to properly track and emulate the channels similarity between adjacent transmission time slots, where the authors of Refs. [21] and [27] focused on exploiting optimal testing threshold of the channel similarities of legitimate users and the adversary, and in Refs. [33,39], different channel characteristics are observed and classified into different clusters to track the channel dynamics rather than rely only on the last trusted channel. The detailed descriptions are provided in the next section.

By employing the typical Alice-Bob-Eve security model, we can introduce a basic key-less channel-based authentication framework as follows, which generally consists of three stages.

- Initialization stage. Alice sends the perfectly authenticated Alice-Bob channel information, denoted by $\boldsymbol{H}_{\mathrm{AB}}$, to Bob at the beginning of communication, i.e., $\boldsymbol{H}(0) = \boldsymbol{H}_{\mathrm{AB}}$.
- At transmission time slot t, Bob receives the data packet with the training sequence from Alice, and estimates the channel information $\hat{\boldsymbol{H}}(t)$ based on the training data block.
- Bob performs the verification as follows. If $\hat{\boldsymbol{H}}(t) \simeq \boldsymbol{H}(t-1)$, the received data come from the same source (i.e., from Alice) as those in the previous time slot; otherwise, the authentication fails. Once the authentication succeeds, $\hat{\boldsymbol{H}}$ is stored as the reference channel estimation. Considering the varing environment and the channel estimation error introduced by the noise, $\hat{\boldsymbol{H}}(t)$ is not equal to $\boldsymbol{H}(t-1)$. Therefore, Bob should use hypothesis testing to differentiate the channels of legitimate transmitter and adversary.

In comparison with the channel-based authentication, the works on RF fingerprint-based authentication (or device identification) schemes appear much earlier. There have already been some works focusing on identifying different devices according to the unique characteristics or imperfections of their analog (radio) circuitry[47,59-62]. Generally, the fingerprinting features will be extracted from the received signal at the receiver side by observing the radio communication, and then be compared with the local reference fingerprints associated to the device under identification. Therefore, reliable fingerprint database and feature selection strategies play a key role in the authentication[63,64]. Based on these early studies, more sophisticated RF fingerprint-based authentication schemes were proposed in Refs. [40,41,46,65-67]. A basic key-less RF fingerprint-based authentication framework is illutrated in Fig. 3, which includes the following steps.

- Training stage. Features of different legitimate devices are collected and selected by using a certain selection method at the receiver side. The goal of this stage is to establish reliable fingerprints database that maps the features of received signals to the identity of an authenticated user.
- The receiver receives an authentication signal from an unknown transmitter and then extracts useful features from the received signal.
- The receiver maps the selected radiometric features to the identity fingerprints by using a pre-determined mapping function, and then retrieves the reference fingerprints from the database and compares them against the obtained fingerprints to verify the identity (or the class) of the transmitter.

In comparison with the channel reciprocity requirement in key-less channel-based authentication schemes, RF fingerprint-based authentication is more stable with time as it is practically impossible to arbitrarily change hardware-level RF features in a short time. However, the RF features between different devices are usually very slight, which forces the receiver to execute complicated feature selection and analyzing algorithms to extract the subtle differences in the signals. Besides, some researchers argue that RF fingerprint-based authentication is vulnerable to impersonation attack[68].

D. Attack Model

Clear definitions and extensive studies on the attack models of potential adversaries are essential in security analysis of a designed PLA scheme. As clever malicious users always exhaust all power and computational resources in hand to achieve their goals, it is necessary to consider all possible attacks the system might face. Depending on the number of adversaries, there are two big classifications of attack models: the single-attacker model and the multiple-attacker model. Most of the previous studies on PLA mainly focus on the single-attacker model, i.e., only one active eavesdropper is considered during the security communication.

In most of the security analysis of communication systems, Eve, the adversary, is an aware active user who has full (read and write) access to the communication channel and also knows the authentication scheme that Alice and Bob use. From the perspective of ensuring higher security, the hypo-

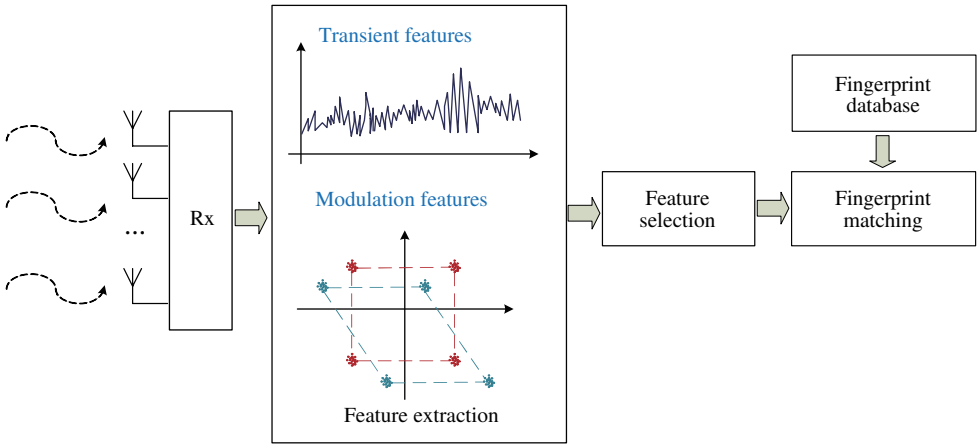


Figure 3 A basic key-less RF fingerprint-based PLA framework

thetical attacker is usually assumed to be as powerful as possible. There are usually three basic assumptions about the hypothetical attacker.

- The adversary has the full knowledge of CSI of the entire communication network, and is also equipped with very powerful hardware equipments with unlimited transmission power and computational ability.
- The adversary has partial knowledge of CSI of the entire communication network (probably the statical characteristics of channels), but is equipped with very powerful hardware equipments with unlimited transmission power and computational ability.
- The adversary has partial knowledge of CSI of the entire communication network (probably the statical characteristics of channels), and its transmission power and computational ability is also restricted due to hardware limitation.

The first two assumptions are not practical, of course, however it can be seen that such assumptions simplify general theoretical analysis about the performance of the proposed authentication schemes in the following sections. The last assumption is more practical, which is usually used as the typical attacker model to evaluate the security performance of a particular authentication scheme. Although the adversary may change attacking strategies due to different communication environment and authentication schemes, the two most commonly used attack modes are impersonation and substitution attacks. Here we mainly focus on these two attack modes in the following discussion.

1) Impersonation Attack: Through impersonation attack, the goal of Eve is to create and send a fraudulent message which is hoped to be accepted by the receiver, i.e., she attempts to imitate Alice. These attackers may forge massive fake identities, or embezzle other legitimate nodes' identities, such as Sybil attacks[69]. There are two ways to perform such attacks: one is an active attack, where the adversary will randomly send the forge message to Bob with the knowledge of authentication scheme and maybe partial CSI of the communication network. Usually the probability that Eve's message is successfully authenticated depends on the security of authentication test that the receiver uses. The other form is a passive attack, where the adversary monitors all kinds of message streams pass through the network during authentication, and tries to learn the knowledge of any useful authentication information (i.e., the shared secret key, the channel information between Alice and Bob, or the identity features or fingerprints) from whatever it gets. Impersonation attack is the most commonly concerned attack mode during the design of authentication framework, either in key-based[32,36,46], or key-less schemes[26,39]. For example, in Refs. [36] and [46], passive attack and key equivocation were considered, based on which the authors of Ref. [46] provided the theoretical lower-bound on the impersonation attack success probability by using the similar approach in Ref. [19]. Besides, in Ref. [36] the authors also considered the case of active attack and derive a closed-form expression for the probability of successful attack. The authors of Ref. [32] emulated the security of PHY-CRAM in relay systems in the scenarios of passive and active attack. In Ref. [39] the eavesdropper tried to impersonate the communication by emulating a large number of different channel responses to improve the passibility of successful attack, and in Ref. [26] besides of single attack, the authors further analyzed multiple attack strategies where Eve can transmitted a sequence of messages in order to break the authentication system. In some works such as Ref. [32], the authors also briefly analyzed the cases of replay attack, including signal replay attack and feature replay attack[60-62]. Generally, relay communication networks are more vulnerable in replay attack due to its inherent characteristics coming from entrusted relay nodes, as it is easier for malicious nodes to store or collect radio features of the waveforms passing by. However, there is little work that focuses on such an attack

mode.

2) Substitution Attack: In substitution attack, Eve will intercept the message coming from Alice and replace it with an originally generated message which she hopes to be accepted and also correctly decoded by the receiver. Therefore, an adversary can be considered successful only when the forge message is successfully decoded by the receiver. Substitution attack is mainly considered in the message authentication systems, though, authentication systems such as that based on tag/fingerprint embedding[46], and relay systems[70] also pay some efforts on security analysis in the scenario of substitution attack. In Ref. [46], the authors also derived the theoretical lower-bound on substitution attack success probability just as that of impersonation attack. A more detailed analysis about substitution attack in relay systems was provided in Ref. [70] where the channel conditions is obtained under which substitution attacks performed by relay nodes can be detected.

III. Key Techniques in PLA

On the basis of the basic security models introduced in the previous section, we will describe the latest and the most frequently studied techniques that may be implemented in PLA frameworks in this section. The concepts, challenges and the recent works on the following two basic authentication categories, namely channel-based authentication and RF-based authentication, are introduced in detail. Besides, other authentication techniques including watermark/fingerprint embedding schemes will also be discussed in this section.

A. Channel-Based PLA

Channel-based authentication schemes are universally studied in PLA which intuitively take advantage of the special nature of wireless environments. Most channel-based PLA frameworks are established on the channel assumption of the well-known Jakes uniform scattering model[71], i.e., the typical frequency-selective wireless scenarios with rich multipath environment. It has been proved that the transceivers in such environment is location-specific, which means the following.

- The channel can be specified by a number of complex samples either in the frequency domain (a set of complex gains at a set of frequencies) or the time domain (a set of impulse response samples at different time delays).
- Such sets of numbers vary from one transmit-receive path to another if the paths are separated by the order of an RF wavelength or more.

Based on the assumptions above, CSI can be utilized as a special kind of fingerprint that represents and discriminates different user identities in the total network. Note that in

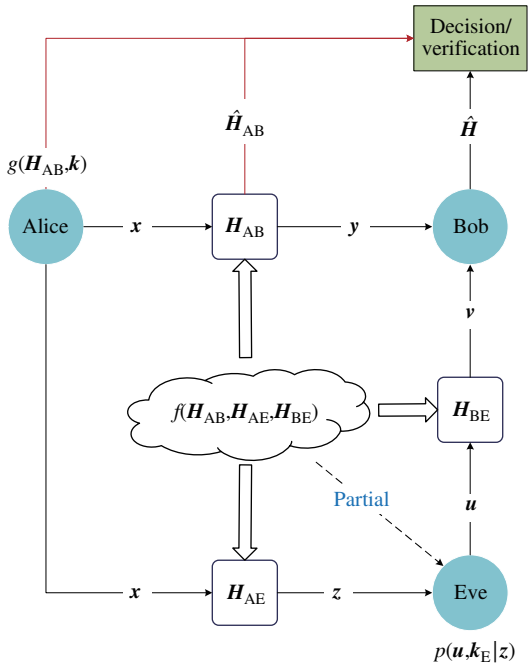


Figure 4 A general abstract information flow in channel-based authentication schemes

OFDM systems the correlations between adjacent subcarriers should also be considered and controlled, as the spectrum decorrelation improves the robustness and the security of PLA systems[36]. Based on the typical Alice-Bob-Eve security model introduced in the previous section, Fig. 4 illustrates a general abstract information flow in channel-based authentication schemes.

As is shown in Fig. 4, $f(\boldsymbol{H}_{\mathrm{AB}},\boldsymbol{H}_{\mathrm{AE}},\boldsymbol{H}_{\mathrm{BE}})$ is the joint PDF of the communication channels, and $g(\boldsymbol{H}_{\mathrm{AB}},\boldsymbol{k}_A)$ represents the PDF of $\boldsymbol{H}_{\mathrm{AB}}$ and the shared secret key $\boldsymbol{k}_A$. Note that for key-less PLA systems, $g(\boldsymbol{H}_{\mathrm{AB}},\boldsymbol{k})$ is reduced to $g(\boldsymbol{H}_{\mathrm{AB}})$. Denote by $\boldsymbol{x}$, $\boldsymbol{u}$ the authentication signal and forge message sent by Alice and Eve, respectively. $\boldsymbol{y}/\boldsymbol{v}$ and $\boldsymbol{z}$ are the received signals in Bob and Eve. Here, $\boldsymbol{x}$ sent by Alice is generated with the knowledge of $g(\boldsymbol{H}_{\mathrm{AB}},\boldsymbol{k})$ and the secret key (if possible), which will also be observed as $\boldsymbol{z}$ by Eve after transmitting through the channel $\boldsymbol{H}_{\mathrm{AE}}$. Assume that Eve is able to acquire partial knowledge of the network channel information, she should learn and grab useful information from $\boldsymbol{z}$ and the limited CSI, after which the forge message $\boldsymbol{u}$ is created by $p(\boldsymbol{u},\boldsymbol{k}_{\mathrm{E}}|\boldsymbol{z})$ and transmitted to Bob. Therefore, Bob may receive either the authentication message $\boldsymbol{y}$, or the forge message $\boldsymbol{v}$ after transmitting through $\boldsymbol{H}_{\mathrm{AB}}$ or $\boldsymbol{H}_{\mathrm{EB}}$, respectively. To complete authentication, Bob firstly performs channel estimation based on the received signal and obtains $\hat{\boldsymbol{H}}$. Then certain authentication scheme $\mathscr{D}$ is used by referring to a local authenticated channel sample $\hat{\boldsymbol{H}}_{\mathrm{AB}}$, which is previously obtained during the initial transmission. For a general channel-based PLA framework, the following hypothesis test-

ing method can be used in the authentication phase

$$\mathscr{H}_0: \quad \hat{\boldsymbol{H}} \simeq \hat{\boldsymbol{H}}_{\mathrm{AB}}, \tag{11}$$

$$\mathscr{H}_1: \quad \hat{\boldsymbol{H}} \simeq \hat{\boldsymbol{H}}_{\mathrm{EB}}. \tag{12}$$

There are three major assumptions in the channel-based authentication scheme introduced above, which are listed as follows.

• The channel coefficient between Alice and Bob should vary slowly or is highly correlated in time so that the estimation results can be almost the same between adjacent time slots. Therefore, the basic method should be modified and enhanced in fast fading environment or time-variant channels[55,57,72].

• The initialization stage must be included to verify the identity of the first transmission source. Generally, the channel information during the first transmission time slot is assumed to be perfectly authenticated by using other authentication schemes (e.g., a cryptographic-based authentication at the application layer). Note that there are limited works that consider the authentication problem during the first transmission.

• The reciprocity assumption is satisfied for PHY-CRAM, i.e., as during the coherence time, the observed channel impulse responses at two geographically separated communicating terminals are the same.

There are two ways of utilizing channel information for identification: one is observing channel impause/frequecy response correlations between adjacent time, the other way is completing authentication directly based on the current channel impause/frequecy response. For the former, we call it as CSI difference-based PLA, while the latter is named as CSI-based PLA for convenience. The representative research works based on these two kinds of methods are summarized in Tab. 3, which will be introduced in detail in the following subsections.

1) CSI Difference-Based PLA: The basic idea of the scheme given by (11) is only valid in time-invariant channel environment. As has already been mentioned in the previous section, in some works like Refs. [21,26,27,33,39,57], the hypothesis testing given by (11) is modified by comparing the channel responses between two adjacent transmission time slots. A threshold δ_h is decided for the following hypothesis testing

$$\mathscr{H}_0: \quad |\hat{\boldsymbol{H}}(t) - \hat{\boldsymbol{H}}(t-1)|^2 < \delta_h, \tag{13}$$

$$\mathscr{H}_1: \quad |\hat{\boldsymbol{H}}(t) - \hat{\boldsymbol{H}}(t-1)|^2 \geqslant \delta_h. \tag{14}$$

The total authentication procedure is similar to that introduced in subsection II.C.

The concept of CSI difference-based PLA is straightforward, though, it has additional challenges in practical use. Specifically, the robustness and security of such authentication schemes are severely affected by the channel inherent characteristics. As CSI difference-based PLA utilizes the differences between a measured (test) channel response and a prior channel response to discriminate between transmitters at different locations, in high dynamic communication networks such as mobile communication and ad hoc networks, the channel response from the same transmitter can fluctuate violently due to the rapid movement of devices. Furthermore, in urban area with intensive buildings, due to the rapid spatial decorrelation properties of the wireless multipath channel, even a minor movement of a mobile can lead to a quite different channel response. It can be seen that the performance of the aforementioned CSI difference-based PLA is greatly degraded due to the mismatch between adjacent channel response, especially the false alarm rate. Therefore, enhanced schemes are required to overcome the performance degradation due to channel dynamics. Therefore, the major challenge or the most important technique of CSI difference-based PLA is the channel-tracking methods in high dynamic communication networks.

A. Channel tracking methods based on generalized time-varying multipath channel models

There have been a number of works to handle the PLA challenge in time-varying multipath communication scenarios, some of which were studied by Xiao's group[29-31,55], and the works in Refs. [57] and [22].

In Ref. [55], a generalized time-variant channel frequency response sample was built in the terms of three parts: the fixed average channel response over time and contains the spatial variability information, the variable part with zero mean, and the receiver noise. The characteristics of these three parameters directly reflect the channel correlation and variance in spatial and time domains, therefore affect the authentication performance in time-varying multipath communications. The hypothesis testing method proposed in Ref. [55] is enhanced in Ref. [30] with a more generalized channel model. The approaches in Ref. [57] use the similar channel model, which introduces a robust channel-based PLA by exploiting the noise-mitigated CIR difference and tracking the significant channel taps.

Above all, for a multipath OFDM system defined in the previous section, due to the correlation of adjacent CIRs on the same path, an autoregressive (AR) model of order 1 (AR-1)[30] is utilized to describe the temporal process of $h_p(k)$ at time k:

$$h_p(k) = \lambda h_p(k-1) + \sqrt{(1-\lambda^2)\sigma_l^2}\varepsilon_p(k-1), \tag{15}$$

where the AR coefficient λ is the correlation of two successive CIRs, ε_p is a zero-mean complex Gaussian random variable with variance 1, and σ_p^2 is the variance of h_p. Therefore, the

Table 3 Representative works on channel-based PLA techniques

Works	Channel model	Authentication mode	Major contributions
[20,21]	Single-carrier, time-invariant and frequency-selective	One-way, CSI difference-based with hypothesis testing	1. Enhanced PLA via residual testing or time-domain CSI comparison[20] 2. Machine learning schemes such as SVM and linear Fisher discriminant analysis (LFDA) are used to exploit the channel features and enhance the PLA performance[21]
[28,37]	Multi-carrier, time-invariant/variant and frequency-selective	Challenge-response, CSI-based with hypothesis testing	1. PLA by exploiting both the reciprocity and randomness of the phase responses over multi-carrier channels[28] 2. A modified artificial-noise aided PLA is proposed to alleviate the discontinuity of phase responses at far separated time slots[37]
[22,29-31,55]	Multi-carrier, time-variant and frequency-selective	One-way, CSI difference-based with hypothesis testing	1. PLA based on a generalized channel frequency response with both spatial and temporal variability and correlations[55] 2. An enhanced PLA scheme which is more robust against terminal mobility is proposed based on inter-burst authentication and intra-burst authentication[29] 3. A PLA scheme with a practical generalized likelihood ratio test (GLRT), that requires no a priori knowledge of channel parameters, is proposed and analyzed in frequency-selective Rayleigh channels[30] 4. A hybrid authentication protocol is proposed to integrate the PLA algorithm into any existing higher-layer security mechanism without assuming a reliable reference channel estimation[31] 5. A novel channel-based PLA is proposed by exploiting and mathematically modeling both of time-varying channel amplitude and propagation delay[22]
[23]	General channel model (not specific)	One-way, CSI-based with hypothesis testing	An outer bound on the error probability region in terms of the attacker strategy is derived
[32]	Multi-carrier, time-invariant and frequency-selective	Challenge-response, CSI-based without hypothesis testing	A novel PHY-CRAM that doesn't require any channel estimation or training

test statistic under the hypothesis testing (13) can be given by

$$T_l(k) = H_l(k) - H_{AB,l}(k-1). \tag{16}$$

According to the distributions of both $H_l(k)$ and the noise $n_p(k)$, the PDF of the test statistic T_l under the two hypothesis $\mathscr{H}_0$ and $\mathscr{H}_1$ can be represented, based on which the threshold δ_h is determined adaptively subject to certain constraints of false alarm or missed detection rate.

Based on the channel tracking model introduced above, enhanced CSI difference-based PLA schemes are studied. The work in Ref. [29] focuses on mobile communication scenarios, and an enhanced scheme is proposed to overcome the problem of rapid spatial decorrelation properties of wireless multipath channels in mobile communication. More specifically, two parts of authentication, inter-burst authentication and intra-burst authentication are considered to relax the limit on user displacement between two bursts. The inter-burst authentication is carried out using the first frame of each data burst to determine whether the current transmitter is still Alice or not, where similar time-varying channel tracking model as that described by Ref. [55] is used with the aid of Neyman-Pearson test. During the inter-burst process both Alice and Bob save at least one channel response from the last burst as the key, which is sent for re-verification in the first frame of each burst. In this way, the authentication with decorrelation channels between data bursts is realized. A cross-layer authentication framework was proposed in Ref. [31]. Besides of the dynamic environment, the effect of unreliable reference channel estimation is also considered to analyze the CSI difference-based PLA scheme proposed in the previous works.

However, the performance of this channel tracking method is strongly related to the AR coefficient. In other words, in the extremely high dynamic communication networks where the AR coefficient tends to zero, the aforementioned channel tracking method may have difficulty in differentiate channel responses of legitimate user and the adversary.

B. Channel tracking methods based on clustering and machine learning

In Ref. [39], the authors have proposed an authentication technique based on Gaussian processes (GPs) to track the channel dynamics by clustering the observations into trajectories. More specifically, Bob measures and stores the N most recent frequency responses of the channel between (a presumed) Alice and himself at time n, which is denoted by $\hat{\boldsymbol{H}}_{AB} = [\hat{\boldsymbol{h}}_{\mathrm{AB},n-N+1}, \hat{\boldsymbol{h}}_{\mathrm{AB},n-N+2}, \cdots, \hat{\boldsymbol{h}}_{\mathrm{AB},n}]$. With the newly observed channel response $\boldsymbol{h}_{n+1}$, the hypothesis testing is performed based on the following test statistic

$$\gamma = \|\boldsymbol{h}_{n+1} - \hat{\boldsymbol{h}}_{\mathrm{AB},n+1}\|, \tag{17}$$

where $\hat{\boldsymbol{h}}_{\mathrm{AB},n+1}$ represents the prediction of Alice-Bob channel according to the predesigned channel tracking algorithm. Here, the authors use the overlapping mixture of Gaussian processes (OMGP) multi-target tracking algorithm from Ref. [73] to obtain $\hat{\boldsymbol{h}}_{\mathrm{AB},n+1}$ with the input observed sample $\hat{\boldsymbol{H}}_{\mathrm{AB}}$. The work in Ref. [33] followed a similar idea.

Basically, we can summarize the basic architecture of such kinds of authentication schemes as "Channel Prediction based PLA", as the kernel of tracking techniques is very similar to channel prediction. The methods introduced above are typical solutions to overcome the channel dynamics. However, note that the properties of multipath and channel dynamics don't always do harm to the system security performance, sometimes the time variations even improve the authentication. Works by Xiao's group all aim to provide advanced PLA schemes with hypothesis testing and sophistic performance analysis based on more generalized channel model with dynamic environment variance, multipath delay, correlation in time and frequency domain, and practical technique limitation such as channel estimation error. For example, Ref. [55] examined the ability of channel-based PLA to authenticate transmitters in a more practical time-variant environment based on a generalized channel response with both spatial and temporal variability, and considers correlations among the time, frequency and spatial domains. The theoretical and numerical results confirm that the minimum average miss rate is a tradeoff between the positive impact of the time variation and its negative impact resulting from the rise of the test threshold.

2) CSI-Based PLA: There are two ways of utilizing CSI directly for authentication, which are listed as follows.

• CSI white-list authentication. It is a natural idea to complete authentication by building the user profile using CSI for a specific user identity before authentication, just as that introduced in Ref. [74]. The mean amplitudes of CSI measurements are analyzed and partitioned into two groups by using K-means algorithm during the data pre-processing stage, then the framework deposits the pre-processed legitimate CSI samples as the user profile for future matching. Obviously, CSI white-list authentication is vulnerable to channel dynamics, which makes the method less practical.

• CSI-key joint authentication. Another way of utilizing the characteristics of channel states is by generating authentication information according to both of the channel and the shared secret key. In Ref. [24], the authors introduced two key-based authentication frameworks: asymmetric channel-based cryptographic authentication (A-CBCA) and symmetric channel-based cryptographic authentication (S-CBCA). Both of these authentication schemes generate the secret keys (private/public key couple for A-CBCA and the shared key for S-CBCA) based on CSI estimations at Alice by using certain mapping strategies such as the SKA protocol[15], which is used to generate the authentication signature and encrypts the message package in both of the authentication frameworks.

CSI-key joint authentication can also be applied to PHY-CRAM systems. Considering the PHY-CRAM based generalized system model given in the previous section, if Bob is able to obtain a good estimation of the Alive-Bob channel $\boldsymbol{H}_{AB}$, then he can compute the response signal $\boldsymbol{R}$ as

$$R_l = K_l \mathrm{e}^{\mathrm{j}\theta_l}, \tag{18}$$

where θ_l is the channel phase response of the subcarrier l. Therefore, the received signal at Alice in the lth subcarrier is given by

$$V_l = R_l \| H_{AB,l} \| \mathrm{e}^{-\mathrm{j}\theta_l} + n_l = K_l \| H_{AB,l} \| + n_l. \tag{19}$$

Hypothesis testing similar to that given by (8) with the test statistic (10) can be considered to determine whether or not $\boldsymbol{V}$ is transmitted by Bob. However, a more realistic way is to use the following simplified method

$$\mathscr{H}_0: \quad |\boldsymbol{K}_A^{\dagger} \boldsymbol{V}| \geqslant \delta_h, \tag{20}$$

$$\mathscr{H}_1: \quad |\boldsymbol{K}_A^{\dagger} \boldsymbol{V}| < \delta_h. \tag{21}$$

Just as CSI difference-based PLA, CSI-based PLA also suffers from various factors including fading and multipath environment, channel estimation error, correlation between subcarriers in OFDM system, etc.

• Channel estimation error. Channel estimation is the first and one of the most important proceeding block during the authentication phase at the receiver. As CSI-based PLA frameworks usually rely on the authentication information extracted from channel responses, channel estimation error can directly affect the successful authentication probability. The problem aggravates for PHY-CRAM based systems, as the receivers of challenge and response signals may have different channel estimations. The authors of Ref. [28] have provided the conditional PDF of the channel phase estimation error.

• Channel correlations of different carriers. For the authentication in OFDM systems, the correlations of different sub-channel responses tends to cause an increase of successful passive attack probability. It has been further proved in Ref. [36] that the phases of channel coefficients of OFDM subcarriers are highly correlated when the number of multipath is extremely small. Therefore, the eavesdropper is able to estimate the secret key through passive attack by using the observed information between adjacent subcarriers, and successfully impersonate the communication between Alice and Bob. To overcome this problem, the subcarriers in a single transmission should be with sufficient spacing, so that the channel coefficients are independent with each other. However, this approach may not be reasonable if the channel does not have a sufficient number of multipath[36].

Therefore, robust authentication that can overcome the impact of fading and multipath environment as well as the channel estimation error in multi-carrier systems is the key research topic in CSI-based PLA design.

a. Robust CSI-insensitive relaxed authentication

One natural idea to overcome the challenge is to design and apply efficient channel estimation methods before authentication procedure. However, sophisticated estimation methods reduce the channel estimation error at the cost of higher computational complexity or longer pilot, which brings extra hardware burden on small devices in future IoT networks.

A promising idea is to consider relaxed authentication schemes, i.e., try to reduce the sensitivity of the authentication message to the channel estimation error. However, any sorts of relaxed authentication schemes design should result in the security performance degradation, because it inherently increases the risk of suffering successful attack. The author of Ref. [36] proposed the idea of generating the challenge signal by combining the information of the shared key and CSI between two legitimate nodes, which is named as key-channel based randomization method. Based on the OFDM system discussed in the previous section, the authentication consists of two steps as follows.

- In the first step, Alice sends a challenge signal to Bob by using a random message, $\boldsymbol{m}$, which is shared by both sides of Alice and Bob (and also in Eve). At Bob, channel estimation ($\hat{\boldsymbol{H}}_b$) is carried out before recovering the message m.

As for Bob, he can determine the corresponding codeword as $\boldsymbol{c} = \phi(\boldsymbol{m})$, where $\phi(\cdot)$ is a channel encoding function. Using the shared secret key, $\boldsymbol{k}$, and the estimated CSI $\hat{\boldsymbol{H}}_b$, Bob can determine the set of selected subcarrier indices ($\mathscr{I}_b$) and obtain a randomized key $\boldsymbol{e}$. Basically, $\boldsymbol{e}$ is a binary sequence whose positions of '1' are related to the subcarriers with relatively lower channel gain. The response signal at Bob is represented by $\boldsymbol{b} = \boldsymbol{c} \oplus \boldsymbol{e}$.

- In the second stage, Alice receives the response signal and obtain estimated CSI $\hat{\boldsymbol{H}}_a$. By using similar key-channel based randomization method, she generates her own key $\bar{\boldsymbol{e}}$ and performs soft-decision decoding with log-likelihood ratio (LLR) to recover $\bar{\boldsymbol{c}}$. If $\bar{\boldsymbol{c}} = \boldsymbol{c}$, Alice can easily verify that the signal is transmitted by Bob for authentication. according to the proposed key-channel based randomization.

By observing the authentication method above, it can be seen that the relationship between authentication message $\bar{\boldsymbol{c}}$ and the channel states is relaxed with the help of $\boldsymbol{e}$ and soft-decision decoding. The author clearly addresses that because $\boldsymbol{e}$ is only related to the subcarriers' indices of the relatively small channel gains. Even though $\boldsymbol{H}_b$ and $\boldsymbol{H}_a$ are different due to the estimation error, their impact on Alice's decoding performance based on LLR in may not be significant due to their low channel gains.

b. Robust authentication without channel estimation

Another solution to the problem of channel estimation error is more straightforward, which is to derive authentication methods without channel estimation. The authors in Ref. [32] introduced a general PHY-CRAM based framework which can achieve secure authentication by exchanging unencrypted shared secrets, such as a random number, and a secret key among participants. The verifier at the receiver is able to verify the secrets without knowing the CSI. More specifically, considering a simple Alice-Bob-Eve security model, suppose that the channel between Alice and Bob is denoted by $\boldsymbol{H}_{\mathrm{AB}}$, which is block fading and remains the same during one challenge-response time duration (the channel reciprocity assumption). The total authentication procedure is similar to the general PHY-CRAM framework given in the previous section.

- Just as mentioned before, Alice first generates a random challenge message $\boldsymbol{m}$ from a message codebook $\mathscr{M}$ which can be either a number or a sequence, and sends the message to Bob. The received message can be represented by $\boldsymbol{y} = \boldsymbol{H}_{\mathrm{AB}}\boldsymbol{m} + \boldsymbol{n}_1$, where $\boldsymbol{n}_1$ is the AWGN.
- Bob receives the challenge message and tries to compute the response signal. Instead of extracting the secret message by channel estimation, Bob obtains the response signal $\boldsymbol{R}$ by simply reversing the noisy challenge signal and multiplexing it with the secret key $\boldsymbol{K}$, i.e.,

$$\boldsymbol{R} = \frac{\boldsymbol{K}}{\boldsymbol{H}_{\mathrm{AB}}\boldsymbol{m} + \boldsymbol{n}_1}. \tag{22}$$

- The response signal received by Alice can be represented by

$$\boldsymbol{V} = \boldsymbol{H}_{\mathrm{AB}}\boldsymbol{R} + \boldsymbol{n}_2 = \frac{\boldsymbol{H}_{\mathrm{AB}}\boldsymbol{K}}{\boldsymbol{H}_{\mathrm{AB}}\boldsymbol{m} + \boldsymbol{n}_1} + \boldsymbol{n}_2. \tag{23}$$

Alice verifies the secret key information by computing the following test statistic

$$\eta = \|\boldsymbol{V} \odot \boldsymbol{m} - \boldsymbol{K}\|. \tag{24}$$

A hypothesis testing method is used by defining the threshold δ

$$\mathscr{H}_0: \quad \eta < \delta, \tag{25}$$

$$\mathscr{H}_1: \quad \eta \geqslant \delta, \tag{26}$$

where Alice performs the verification as: if $\mathscr{H}_0$ is accepted, the transmitter is Bob. Otherwise, the authentication fails.

It can be concluded that since the method above does not need to estimate CSI, the training and synchronization sequences in the header are eliminated, which prevents the attackers from probing the channel and increases security strength. Besides, in comparison with other traditional challenge-response authentication, channel coding and frequency offset compensation are not used in the scheme. Due to its simplicity and good concealment of CSI, the "noisy

channel inversion" based PLA-CRAM can be extended to relay systems, just as discussed in Ref. [38].

c. Robust authentication for OFDM systems

Authentication design of multi-carrier systems, such as OFDM systems, is essential for the practical utilization of PHA in communication systems under multipath environment.

There are two promising tracks to handle the problem: one is by adding artificial interferences to the secret information at each subcarrier; the other is considering time diversity, i.e., by allocating more carriers at sufficiently-separated time-epoches larger than the coherence time. Both of these two methods are considered in Ref. [37].

For an OFDM PHY-CRAM system with Alice-Eve channel matrix $\boldsymbol{H}_{\mathrm{AE}}$, denote by $\boldsymbol{m} = [a_0 \mathrm{e}^{\mathrm{j}\theta_0}, \cdots, a_{L-1}\mathrm{e}^{\mathrm{j}\theta_{L-1}}]$ the authentication message. The received signal at Eve in the lth subcarrier after adding artificial phase noise is given by

$$x_l = H_{\mathrm{AE},l} a_l \mathrm{e}^{\mathrm{j}\theta_l} \mathrm{e}^{\mathrm{j}v_l} + n_l = H_{\mathrm{AE},l} a_l \mathrm{e}^{\mathrm{j}(\theta_l + v_l)} + n_l. \tag{27}$$

Although $\mathrm{e}^{\mathrm{j}\theta_l}$ may be strongly correlated between OFDM subchannels, it has been proved in Ref. [36] that due to the artificial interference $\mathrm{e}^{\mathrm{j}v_l}$, Eve may not be able to recover the authentication message $\boldsymbol{m}$ by observing channel phase response in adjacent subcarriers. To remove the effect of the artificial interference, Bob should design different test statistic for efficient hypothesis testing. The authors in Ref. [37] simply used the same test statistic as that defined in the previous work[28], and claim that a strictly-positive key equivocation can be ensured even for the worst case scenario. For the case of active attack, the effect of the extra phase noise $\mathrm{e}^{\mathrm{j}v_l}$ on the performance of hypothesis testing should be carefully considered, which is however not clearly stated in Ref. [37].

B. RF Fingerprint-Based PLA

Due to certain imperfections inherent in the hardware components caused by various manufacturing and environmental factors, the device-dependent bias to the nominal hardware specification can be used to create a waveform signature (fingerprint or transceiverprint) that can uniquely determine the transmitter identification. Such imperfections are generally regarded as RF features.

A typical RF fingerprinting-based system consists of preprocessing, detection, feature extraction, and classification stages[75]. The general system model has been given in Fig. 3 except the preprocessing block. The goal of the preprocessing stage is to generate complex-valued analytic functions from real-valued data, which has been extensively studied. Therefore, we mainly focus on the research results of detection, feature extraction, and classification techniques.

There are two types of widely studied RF features: one is based on the signal transients[60,75-80] where the transient behaviour of RF signal with respect to instantaneous frequency and amplitude is utilized as device identity. The other one mainly considers modulation domain techniques that represent signals at the most basic level in terms of I/Q samples, whose interpretation depends on the underlying modulation scheme[40,41,61,65,66,81]. Most of works on transients-based authentication schemes focus on the power, timing, amplitude, phase or frequency of RF waveforms, while the modulation-based schemes are based on the carrier center frequency (frequency offset), the error in the symbol clock of the transmitted signal, the I/Q imperfections, etc. An extensive study about RF fingerprints of wireless devices was provided in Ref. [3], including the taxonomy of wireless features, a review on fingerprint algorithms, and some open research problems in the field.

1) Transient Detection: The objective of the transient detection stage is to determine the exact time instant at which the transmitter is turned on[75], i.e., the turn-on transients. In the following analysis, we also refer to transient detection as "transient extraction", which represents the similar meaning in some works[82]. Both amplitude-based and phase-based techniques have been previously investigated for detecting signal transients[75,83,84]. Here are the brief introduction on typical transient detection techniques.

• Bayesian detection/Amplitude detection. Bayesian detection is a classical turn-on transients detection strategy widely used in the early studies for RF fingerprinting-based PLA[85,86]. For amplitude-based schemes, the technique is based on a posteriori probability of a simple step change point detector which determines the instant at which the received power level exhibits a sudden increase[85]. Here, the Bayesian approach relies on Bayes' theorem for describing the learning process, by which prior information is updated. However, the received power level increases gradually due to the practical Wi-Fi standard. If the change detector lags behind the actual starting point, characteristics important for classification may be lost. The authors of Ref. [75] proposed a Bayesian ramp change detector[86], which is able to estimate the time instant at which the signal power starts its gradual increase.

• Phase-based detection. The idea of phase-based detection is based on the fact that the slope of the phase associated with the start of transient is linear. Therefore, a phase-based detection technique is carried out by computing the corresponding difference in phase variance, which was firstly given in Ref. [83]. The method is proved to be more susceptible to noise and interference. The works in Ref. [87] followed the phase-based detection approach proposed in Ref. [83].

• Preamble detection. As noted in Ref. [88], the power amplifier's control mode of ramp-on may lead to unstable aligned envelope profiles. Therefore, an enhanced detection scheme utilising a fixed preamble and its periodicity was proposed by the authors of Ref. [88], which is verified with a direct se-

quence spread spectrum-preamble.

• Hybrid detection. A natural idea to enhance detection accuracy can be considered with reliable judging means that provide a-priori whether amplitude-based or phase-based features will provide best performance. The authors in Ref. [84] proposed a variance trajectory-based generalized detection approach which permits evaluation of transient detection performance using both amplitude and phase features.

Besides, there are some other works searching for efficient detection methods based on other features, such as Ref. [82] where the envelope rather than the amplitude features of the transient signal is utilized to achieve lower sampling rate, which is more suitable for more resource constrained transceivers.

2) Feature Extraction/Selection: The specific signal characteristics used to differentiate various transients or transmitters are called transient features. The works of Ellis and Serinken[89] analyzed both amplitude and phase information that can be used as RF fingerprints. For the amplitude information, the authors of Ref. [75] observed several most commonly used features. The phase information is directly obtained after detecting the complex signal[83], such as the standard deviation of phase/normalized phase, standard deviation of normalized in-phase data, etc. The authors of Ref. [87] considered both of the amplitude and phase features.

The basic structures of modulation-based authentication systems are similar to those of transient-based ones, while the major difference falls in the features extraction.

a. Fixed features without selection

A good signal feature has a low intra radio variability (from sample to sample in the same radio) but a high inter radio variability (between different radios)[87,90].

There have already been plenty of works studying the authentication based on fixed transient-based or modulation-based RF features. For transient-based schemes, the generally used features include received signal strength information[77,78], the signal amplitude or phase angle[75,83,87,90], power spectral density (PSD) features[84], etc. Besides, the authors of Refs. [91] and [92] developed approaches for device identification based on the imperfections of RF oscillator of a transmitter, i.e., frequency offset and phase noise. In Ref. [91], the RF features were extracted from variations in the components that comprise the phased-locked-loop circuit of RF oscillators, while in Ref. [92], the feature are corresponded to the variability in control voltage.

For modulation-based schemes, the most commonly used modulation feature is carrier frequency offset (CFO), which is a salient metric with great flexibility for device identification. As has been demonstrated in Ref. [65], CFO can further enhance the authentication accuracy in comparison with other radiometric signatures. Besides, CFO estimation and compensation are commonly embedded functions for signal recovery in wireless systems, therefore no extra computational resource is required. In Ref. [65], the authors concentrated on the static wireless environment with constant CFO, which is extracted and estimated by using training sequences. However, due to the mobility-induced Doppler frequency shift in more realistic mobile communications, time-varying CFO should also be considered. The authors of Ref. [66] further extended the CFO-based authentication scheme in Ref. [65] to the time-varying scenario and propose a continuous PLA scheme for mobile communications by exploiting the characteristics of varying CFO. More specifically, the combined CFO model ia modeled as an AR random process. Then Kalman filtering is used to track the variation pattern in the predicted sequential CFO estimates, which is compared with the current actual estimated CFO in the hypothesis testing.

Obviously, the basic idea of the aforementioned authentication is very similar to that of "Channel Prediction based PLA" in the previous subsection, which is actually known as "CFO Predection based PLA".

b. Multiple features with selection

Besides of fixed features based PLA, there are quite a few works on PLA schemes achieved by observing multiple features[40,41,81]. As suggested by Ref. [89], the feature set which will yield the best fingerprint is highly dependent on the type of radio being fingerprinted. The identification accuracy is intuitively maximized when using all available features for model training and application, though, this leads to increased computational complexity in the training phase. An efficient feature extraction algorithm should minimize the length of the feature vector without losing the necessary components for classification[75]. Therefore, it is important to select the set of radiometric features that can effectively discriminate between different devices before exploring efficient classification algorithms.

Early works on transient feature selection rely strongly on experiments, i.e., the performance of different feature combinations are tested using all kinds of experimental setup. In Ref. [90], the authors created 6 data sets for each sensor node in a wireless sensor network, each of which consists of 100 samples. The works in Ref. [79] suggested that the transients can be reduced to approximately 50 to 200 feature values from 2 000 to 3 000 complex samples, where the exact sizes of the window and feature vector are experimentally determined by trial and error to give the best classification performance.

The extensive study on machine learning in the past decades shows that the efficiency of feature selection can be enhanced by applying different machine learning techniques. A natural idea applied in Ref. [87] was to determine the two classes of features that have low intra-transceiver variability and high inter-transceiver variability, respectively, by using Euclidian distance and clustering techniques of multivariate

analysis. Further, in Ref. [75] a multivariate procedure, principal component analysis, was used to reduce the dimensionality of the feature space by rotating the data such that maximum variability is projected onto the axes. The authors of Ref. [75] claimed that there is no definite pattern from phase profiles that will help to discriminate among the transmitters, as the dynamic range of phase variations is very small. Therefore, the phase profiles are not included as part of the feature vectors.

As mentioned before, the principle of feature selection for transient-based RF fingerprinting is to select the features set that varies slightly for the same device while has higher variability between different radios. To realize the same target in modulation-based authentication systems, designing specific metrics that quantify the relevance between the features and identities of devices and the redundancy of different features is considered. The works in Refs. [41] and [63] used the mutual information of the feature g_i and the device identifier ID, which can be a unique integer number known previously, to measure the relevance between features and users identities

$$I(g_i,\mathrm{ID}) = \int \sum_{\mathrm{ID}\in\mathscr{I}} p(g_i,\mathrm{ID}) \log \frac{p(g_i,\mathrm{ID})}{p(g_i)p(\mathrm{ID})} \mathrm{d}g_i, \tag{28}$$

where $p(g_i,\mathrm{ID})$, $p(g_i)$, and $p(\mathrm{ID})$ are joint PDF and partial PDFs of g_i and ID, respectively. Similarly, redundancy can also be represented by the mutual information of two different features, which is defined as the average of the mutual information of each pair of features

$$W_S = \frac{1}{N} \sum_{g_i,g_j\in\mathscr{G}} I(g_i,g_j), \tag{29}$$

where $N = |\mathscr{G}|$. Obviously, a good features set $\mathscr{G}$ should maximizes the relevance while minimizing the redundancy. In Refs. [41], [63], and [64], the authors developed and implemented an enhanced feature selection algorithm based on minimizing redundancy-maximal-relevance to order the observed features by their effectiveness in device discrimination. Besides of the algorithm mentioned above, there are other researches that develop the feather selection framework directly based on the statistic properties of candidate features. In Ref. [81], the authors developed a signature extraction (learning phase) framework to compute each single-characteristic classifier by using data-driven density formation and maximum likelihood (ML) classification.

Above all, Tab. 4 is a nonexhaustive conclusion of relative works.

3) Classification: The target of classification in authentication is to verify the identity of a user by analyzing and matching the transient features to the local legitimate RF fingerprint data base. Depending on whether training phases are involved, the classification methods that have been studied for RF fingerprinting-based PLA can be further classified into two categories: supervised and unsupervised classification.

a. Supervised classification

Most of the classification schemes used in RF identification are based on supervised learning. For supervised classification, collection of training data sets is needed and has a great impact on the performance of the classifiers. Here we summarize some typical supervised classification schemes that are widely applied in device identification.

• Feature distance. It is a natural idea to perform a "one-to-many" comparison between current received RF fingerprint and the stored reference models based on the predesigned feature distance function. The authors of Ref. [62] propose a transient-based identification method by using Mahalanobis matching, where Mahalanobis distance is used to see the similarity between the reference and test features. It is shown that the method achieves an equal error rate as low as 0.0024 (0.24%). The authors of Ref. [94] used the multitude of sensors on a modern smartphone to generate a robust device fingerprint in mobile communication. Two classification schemes, Euclidean distance-based and the maximum-likelihood estimation methods, are considered and compared in experiment setups using frequency response graph. It was concluded in Ref. [94] that the identification precision of simple Euclidean distance-based classification approach is degraded due to the variation at some frequencies, which causes penalty for distances calculation.

• Fisher-based MDA-ML classification process. Multiple discriminant analysis with maximum likelihood (MDA-ML) is an extension of Fisher's linear discriminant process, where higher dimensional data is projected onto a 2-dimensional "Fisher plane" that maximizes inter-class distances while minimizing intra-class distances. Therefore, classification is performed using unknown data and the trained 2-dimensional decision boundaries calculated from ML distributions. In Ref. [95], fisher-based MDA-ML is used for device classification.

• Support vector machines algorithm (SVM). SVM is a widely studied supervised algorithm which learns to classify the data sets from a set of training examples (e.g., the reference fingerprints) and makes it a non-probabilistic binary linear classifier. SVM achieves a good separation on a multi-dimensional surface by using a separating function, therefore, distance between the nearest training-data point of any class (so-called functional margin) becomes the largest. The advantage of the SVM for fingerprint classification is that it is well known for its high level of accuracy and robustness against outliers. In Ref. [61], the authors proposed an identification technique referred to as the passive radiometric device identification system (PARADIS) that when applied to a large set of network interface cards (NICs) achieves ex-

Table 4 Representative works on different feature extraction/selection techniques

Feature selection	Brief introduction	Works
Experimenting	Evaluating the performance of different feature combinations	[79,90]
Redundancy-maximal-relevance minimization algorithm	By calculating the mutual information between features and devices ID, finding the feature set that maximizes the relevance while minimizing the redundancy of two different features	[41], [63,64]
Principal component analysis	Finding a compact feature subspace that contains most of the total covariance in the data	[93]
Euclidian distance and clustering techniques of multivariate analysis	Determining the intra-transceiver and inter-transceiver variability	[87]
Data-driven density formation and maximum likelihood classification	Combining the results of the single-characteristic classifiers based on weighted voting	[81]
Probabilistic neural network	Following an approach developed in statistics called Bayesian classification	[75,83]

cellent identification accuracy. During the training stage of the PARADIS-SVM classifier, a number of matrices are constructed which return a measure of similarity between a given signature and the known ones. The classifier finally returns the best-matching identity along with a measure of similarity in identification.

• K-nearest-neighbor (KNN). KNN is a typical nonparametric method used for classification and regression, where the input k closest training examples in the feature space are assigned to the majority category label among its k nearest neighbors. The KNN classifier is commonly based on the Euclidean distance between a test sample and the specified training samples. There have been extensive works about KNN including new rejection approaches, refinements with respect to Bayes error rate, distance weighted approaches, soft computing methods, and fuzzy methods. In Ref. [61], KNN was also considered a classification scheme for comparison during classifier training and identification stages, where the radiometric signatures were obtained from frames as training data sets.

• Probabilistic neural network. In Refs. [75] and [83], probabilistic neural network provided a general solution to classification problems by following Bayesian classification, which is an approach developed in statistics that takes into account the relative likelihood of events and uses a priori information to improve prediction. The authors of Ref. [75] provided the performance analysis of the proposed identification schemes through different classification tests, including the benchmark test, transient duration test, and the dimension reduction test by using probabilistic neural network. It is shown that the classification error is around 2%∼4% for standard Wi-Fi radios[83]. Provided the performance test results based on the infrastructure which captures signals from Bluetooth wireless PC cards, Bluetooth test radios and 802.11 wireless LAN adapters. The results showed that the detection algorithm proposed in the paper achieves an overall success rate of about 89.5%.

• Bootstrap aggregating. Bootstrap aggregating (or bagging) is an enhanced machine learning scheme whose basic idea is to increase classification accuracy by generating and aggregating multiple classifiers[96]. The major priority of the bootstrap aggregating algorithm is that it can achieve virtually identical accuracy with reduced computational cost in comparison with traditional machine learning schemes. In Ref. [41], the authors proposed an identification algorithm based on bootstrap aggregating method, where the training data is divided into subsets and then each subset is used to train a nonlinear kernel predictor. The kernel used in Ref. [96] was based on the C4.5 decision tree algorithm.

b. Unsupervised classification

Compared with supervised methods, unsupervised methods have the advantage of differentiating devices without the knowledge of device fingerprints ahead of time. In Ref. [77] the received signal strength (RSS) vector, which is related to the environment and the locations of devices, was measured by surrounding access points, and in Ref. [78] the RSS readings were modeled as a Gaussian mixture model. There are limited works on the applications of unsupervised classification methods in RF fingerprinting-based PLA, and here is a brief introduction about some well-known techniques.

• K-means clustering. As a typical unsupervised learning technique, K-means clustering is one of the most widely used clustering algorithm in classification problems. The aim of the K-means algorithm is to partition a number of N-dimension points into several clusters. In Ref. [97], the authors proposed a K-means cluster based approach to analyze RSS and detect spoofing attacks, which is general for almost all RSS-based localization algorithms.

• Unsupervised Bayesian learning. Unsupervised Bayesian learning is the application of Bayesian networks to unsupervised learning. The technique is based on the principle that the feature space can be modeled as a nonparametric Bayesian model when the number of classes is undetermined. For example, in Ref. [98], a nonparametric

Bayesian model was employed based on reliable RF features to enhance the classification performance. More specifically, the goal of the identification is to determine the parameters of the distributions and the indicators to classify the features based on an infinite Gaussian mixture model. Finally, a modified collapsed Gibbs sampling method is proposed to sample from the distribution and find the class label with the maximum a posteriori probability.

• Principal component analysis. Principal component analysis is a widely used unsupervised dimension reduction technique that aims to extract the main information from an original data set, usually by finding the best low rank approximation of the data via the singular value decomposition. Principal component analysis is generally used as a dimension reduction technique along with other unsupervised learning schemes such as K-means clustering. In Refs. [75] and [67], principal component analysis was used to reduce the dimensions of feature space for fingerprinting definition.

C. Watermark Embedding-Based PLA

The idea of watermark embedding-based PLA is inspired by conventional cryptography-based authentication methods. For conventional authentication methods, a tag (which can be referred to as an MAC or digital signature) is generated and transmitted along with the data message by the transmitter. Generally, the tag is a function of the data (for the MAC) or the unique identity (for the digital signature), and a secret key that is only shared between the legitimate transmitter and receiver, which is a separate authentication message appended to the information message. Clearly, as the tag is available to an adversary, the security of such kinds of methods are guaranteed by the secret key and a complicated tag generating function, which ensures that an adversary cannot easily recover the key given the message and the associated tag. However, appending the separate tag consumes additional bandwidth.

In the physical layer, the only difference is that the tag, which conveys unique watermark fingerprint, is superimposed on the transmitted information stream. The resulting methods are referred to as "authentication with superimposed tag (Auth-SUP)", which are the most widely studied schemes in the early PLA works.

A typical framework of Auth-SUP is illustrated in Fig. 5. Given transmitted message $\boldsymbol{m}$ and secret key $\boldsymbol{t}$ at Alice, she generates an authentication tag by using the message and a previously shared secret key. i.e., $\boldsymbol{t} = \mathscr{H}(\boldsymbol{m}, \boldsymbol{k}_A)$, where $\mathscr{H}$ is a cryptographic hash function. Then, the transmit signal is created by superimposing the tag to the original information message. To verify the identity of Alice, Bob should first extracts the authentication tag from the received signal and then compares it with a locally generated one. Clearly, the kernels of Auth-SUP are tag superimposing and testing methods. There are two superimposing ways: add the authentication tag to the data frame[42,43,45,46] or to the training pilot frame[44]. For the sake of simplicity, the authentication methods based on the formal superimposing way is called SUP-based PLA, while that based the latter way is referred to as blind SUP (BSUP)-based PLA[44]. These two categories of watermark embedding-based PLA will be introduced in detail in the following subsections.

1) SUP-Based PLA: The basic procedure of SUP-based PLA is illustrated in Fig. 5. Here we discuss more detailed transceiver methods.

a. Tag superimposing

Alice usually generates the transmitted signal using the information message $\boldsymbol{m}$ and the authentication tag $\boldsymbol{t}$ as

$$\boldsymbol{s} = \sqrt{p_s}\boldsymbol{m} + \sqrt{p_t}\boldsymbol{t}, \tag{30}$$

where p_t and p_t are the powers allocated to the information message and authentication tag, respectively. Typically, $p_t \ll p_s$ so that the adversary, Eve, can only obtain an extremely noisy version of tag, which is challenging for her to extract correct information about the secret key. Let $\boldsymbol{p}$ be the pilot sequence used for channel estimation, the final transmitted data frame has the following information

$$\boldsymbol{x} = \{\boldsymbol{p}, \boldsymbol{s}\}. \tag{31}$$

b. Tag testing

The received signal at Bob after channel transmission is given by

$$\boldsymbol{y} = \boldsymbol{H}_{\mathrm{AB}}\boldsymbol{x} + \boldsymbol{n}. \tag{32}$$

To perform authentication, Bob should first decode $\hat{\boldsymbol{s}}$ after channel estimation to obtain information message $\hat{\boldsymbol{m}}$. Then a residual signal $\boldsymbol{r}$ is extracted from the received signal $\boldsymbol{y}$ by $\boldsymbol{r} = Ext(\boldsymbol{y}, \hat{\boldsymbol{m}})$, based on which a hypothesis testing is considered:

$$\mathscr{H}_0: \quad \text{Tag } \boldsymbol{t} \text{ is contained in } \boldsymbol{r}, \tag{33}$$

$$\mathscr{H}_1: \quad \text{Tag } \boldsymbol{t} \text{ is NOT contained in } \boldsymbol{r}. \tag{34}$$

Based on different security system structures, various of efficient residual signal extracting and hypothesis testing schemes are studied. One of the most generally used schemes is correlation analysis[43]. Assume that Bob can always recover the information message perfectly (i.e., $\hat{\boldsymbol{m}} = \boldsymbol{m}$), he will compute the expected tag and the residual signal by $\boldsymbol{t} = \mathscr{H}(\boldsymbol{m}, \boldsymbol{k})$ and $\boldsymbol{r} = \frac{1}{p_t}(\boldsymbol{y} - p_s\boldsymbol{m}, \boldsymbol{k})$. The presence of the expected tag $\boldsymbol{t}$ is tested by designing the test statistic

$$\eta = \mathscr{R}(\langle \boldsymbol{r}, \boldsymbol{t} \rangle), \tag{35}$$

which is the cross-correlation between the expected tag and the residual signal. The hypothesis testing is performed by comparing (35) with a predetermined threshold. According

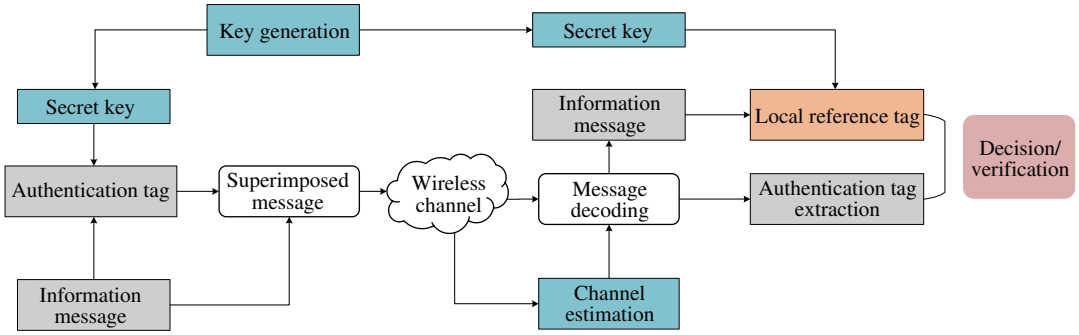


Figure 5 A typical framework of Auth-SUP

to[43], the threshold is determined by minimizing missing detection rate subject to certain false alarm rate (the approach is similar to that applied in channel-based PLA).

c. Performance analysis

Clearly, the authentication performance is strongly affected by channel noise and message recovering error, as it is trivial that a slight change in the message will result in the nonsuccess of tag recovery at the receiver, and hence authentication will fail with a high probability. The impact of imperfect message recovery has been analyzed in Ref. [43], where the probability of successful authentication is given by

$$P_{\text{auth}} = \rho P(C) + \beta(1 - P(C)). \tag{36}$$

Here, ρ is the conditional probability of authentication given correct message and $P(C)$ denotes the probability of correct message recovery. β is the false alarm rate as noted in the previous sections. To simplify the analysis, it is generally assumed that $\beta \to 0$, so that the receiver will always reject imperfectly recovered message frames.

Power allocation strategy is another kernel impact factor to the authentication performance of Auth-SUP. As noted above, to maintain the privacy of the secret key, the tag has relatively much lower power than the data. However, the probability of successful tag detection generally decreases as p_t decreases. It is observed that the tag power should be properly chosen so that it can be distinguished from the noise during tag testing, while it does not cause severe performance degradation to message detection at the same time. Therefore, a tradeoff between the performance of tag testing (also known as the system robustness) and message recovery (which reflects the system covertness) was considered in Refs. [43,45,46,99]. In Ref. [43,45,46], the system robustness, covertness, and security are analyzed separately. The authors of Ref. [45] mainly focused on PLA with single-antenna fingerprinting. In Ref. [46], the framework was extended to MIMO system, where multi-antenna precoding and channel mode power allocation applied to both the data and the fingerprint are also studied. It was concluded in both Refs. [45] and [46] that when the tag is superimposed at relatively very low power (small than 1 percent of the total power), the impact on the bit error rate of data is negligible. In this case, an improvement of authentication performance can be achieved by increasing the tag size, e.g., increasing the frame length, which helps to achieve a trade-off between the desired authentication performance and security. In Ref. [99], the authors propose a new systematic metric that combines the three properties together to fairly compare the performance of different authentication schemes.

2) BSUP-Based PLA: The basic procedure of SUP-based BPLA is illustrated in Fig. 5. Here we discuss more detailed transceiver methods.

a. Tag superimposing

Unlike the SUP-based PLA scheme described above, Alice generates the authentication tag by using pilot $\boldsymbol{p}$ and previously shared secret key $\boldsymbol{k}$. i.e., $\boldsymbol{t} = \mathscr{H}(\boldsymbol{p}, \boldsymbol{k})$. The tag is then superimposed on the pilot as

$$\boldsymbol{s} = \sqrt{p_s}\boldsymbol{p} + \sqrt{p_t}\boldsymbol{t}, \tag{37}$$

and the final data frame is transmitted with the following information.

$$\boldsymbol{x} = \{\boldsymbol{s}, \boldsymbol{m}\}. \tag{38}$$

b. Tag testing

The received signal at Bob during the training period is given by

$$\boldsymbol{y}_p = \boldsymbol{H}_{\text{AB}}\boldsymbol{s} + \boldsymbol{n} = \boldsymbol{H}_{\text{AB}}(\sqrt{p_s}\boldsymbol{p} + \sqrt{p_t}\boldsymbol{t}) + \boldsymbol{n}. \tag{39}$$

As Bob has perfect knowledge of pilot $\boldsymbol{p}$ and key $\boldsymbol{k}$, he can easily recover the authentication tag $\boldsymbol{t}$. To detect whether the tag is included in the received signal, a hypothesis testing is considered which is similar to (33). However, due to the known interference of $\boldsymbol{p}$, it is not straightforward to remove the effect of $\boldsymbol{p}$ without a good estimation of CSI. Actually, the process of cancelling the interference part in (39) can be regarded as a problem of blind known-interference cancellation (BKIC). The principle of the BKIC is introduced in Ref. [100], which is based on the assumption that the wireless channel generally remains similar between adjacent symbols.

Therefore the interference in $\boldsymbol{y}_p$ can be cancelled by using the pilot in its adjacent symbol. The output of BKIC can be summarized as

$$\boldsymbol{r}_0|\mathscr{H}_0 = \boldsymbol{H}\sqrt{p_t}\boldsymbol{t} + \boldsymbol{n} + \boldsymbol{\epsilon}, \tag{40}$$

$$\boldsymbol{r}_1|\mathscr{H}_1 = \boldsymbol{n} + \boldsymbol{\epsilon}, \tag{41}$$

where $\boldsymbol{\varepsilon}$ is the residual noise caused by the pilot cancellation processing of the BKIC module. Observing the similarity between adjacent channels, the effects of channel coefficients can be removed[44]. In Ref. [44], the method above is enhanced so that it can be extended to multi-path channels, as the knowledge of the maximum delay of all paths is assumed to be known at the transceiver.

Unlike the conventional SUP-based method, BSUP-based PLA requires neither channel estimation nor message recovery, which greatly saves computational resources. As the tag is superimposed on the pilot, the receiving SNR of the message also increases. The whole authentication framework is more reasonable and streamlined than conventional ones. However, as the BSUP scheme corrupts the pilot structure, the accuracy of channel estimation is deteriorated in message recovery for an unaware receiver.

3) Security Analysis: For watermark embedding-based PLA systems, in most research works only substitution or impersonate attack is considered, i.e., the target of the adversary, Eve, is to impersonate the communication between Alice and Bob by generating legitimate authentication tag. Therefore, a successful attack performed by Eve is dependent on whether she can recover correct authentication tag hidden in the transmission signal. To achieve this, she should have the ability of extracting key information by observing the information package passing by. Key equivocation[101] quantifies the uncertainty about the secret key with Eve's observation of the transmitted signal $\boldsymbol{y}_E$, which is given by the binary entropy $\mathbb{H}(\boldsymbol{k}|\boldsymbol{y})$. Assuming that Eve has perfect knowledge of pilot/data information and the hash function $\mathscr{H}(\cdot)$, generally, the key equivocation can be focused as

$$\mathbb{H}(\boldsymbol{k}|\boldsymbol{y}_E) \cong \mathbb{H}(\boldsymbol{k}|\boldsymbol{I},\boldsymbol{t}) = -\sum_{\boldsymbol{k}_i \in \mathscr{K}} f(\boldsymbol{k}_i|\hat{\boldsymbol{t}}) \log f(\boldsymbol{k}_i|\hat{\boldsymbol{t}}), \tag{42}$$

where $\hat{\boldsymbol{t}}$ is the estimate of tag at Eve and $f(\cdot|\cdot)$ is the conditional PDF ($\boldsymbol{I} = \boldsymbol{m}$ for SUP and $\boldsymbol{I} = \boldsymbol{p}$ for BSUP). As indicated by Ref. [101], it is proven in Refs. [45] and [44] that if the tag is observed without error, a positive key equivocation is achieved if and only if multiple keys $\boldsymbol{k}_i$ map the information data $\boldsymbol{m}$ to the same tag $\boldsymbol{t}$ for SUP-based PLA (for BSUP scheme, the information data is replaced by pilot $\boldsymbol{p}$). Otherwise, each tag corresponds to a unique key and $\mathbb{H}(\boldsymbol{k}|\boldsymbol{y}_E) = 0$, Eve has full knowledge of the secret key. However, as the tag is always observed with noise and interference, there is a chance of $\hat{\boldsymbol{t}} \neq \boldsymbol{t}$, where $f(\boldsymbol{k}_i|\hat{\boldsymbol{t}})$ is always nonzero and positive key equivocation is achieved. Therefore, as noted in Refs. [44], [45], and [99], the security of Auth-SUP based PLA mostly depends on how noisy the observed tag is at Eve. Especially, the key equivocation approaches its upper bound $\mathbb{H}(\boldsymbol{k})$ as the noise becomes increasingly powerful. The work in Ref. [45] provided a basic security analysis for SUP-based PLA systems, where the simulation about equivocation of the binary tag with varying tag-to-noise ratio (TNR) is performed. The results indicate that the noise significantly expands the search space for the secret key at Eve, which deceases the probability of successful key recovery by using the brute force method. The authors of Ref. [44] extended the analyzing results to the BSUP-based PLA framework. It is concluded in Ref. [44] that both of the power allocation ratio and TNR directly affect the equivocation. The simulation results indicate that although SUP has higher value than equivocation, the gap is narrow and both schemes have similar security levels.

D. A Brief Comparison Between the Mentioned PLA Techniques

According to the discussion above, it can be conclude that the authentication performances of different PLA techniques are strongly affected by the channel conditions. We have summarized the performance and applications of the three PLA techniques mentioned above in Tab. 5.

IV. Further Applications and Research Trends of PLA

A. Multiuser Communication Networks

As aforementioned, works on PLA are mostly based on the traditional Alice-Bob-Eve model, i.e., the communication scenario with only a pair of legitimate transceiver and a single adversary. A more general communication system can includes multiple base stations (BSs), access points, devices and terminals with complicated topology. Besides, there exist multiple attackers that aim to disturb the system by stealthy wireless impersonation attacks. Therefore, secure authentication between multiple transmitters and users should be taken into account. There are some existing works that focus on PLA in mission critical machine-type communication (MTC) and cognitive radio networks[33,58,102-105].

To understand the advantages of PLA for a more generalized multiuser communication network, several problems and techniques should be clearly stated, some of which are as follows.

- The applications of distributed multiple antennas in PLA
- Delay control in mission critical MTC
- Reliable authentication in cognitive radio networks
- The applications of deep learning in multiuser PLA

Table 5 The performance comparison of three PLA techniques

Name	Strengths	Weaknesses	Applications
Channel-based PLA	The idea is straightforward and the security is well guaranteed by taking advantage of the special nature of wireless environments	1. The CSI information should be renewed each time when channel changes, and its performance is degraded in high dynamic environments 2. The security cannot be guaranteed in LOS channel systems	Urban communication, indoor communication networks, etc.
RF fingerprint-based PLA	The authentication is completed by exploiting the unique variations in the RF chain of radios, which is usually stable with time. So RF fingerprint-based PLA can be applied in systems with high dynamic	The RF difference between different devices are usually slight, so a large number of features may need to be selected and classified to uniquely identify transmitters, which causes high computation complexity	High dynamic communication systems
Watermark embedding-based PLA	The idea is similar to conventional cryptography-based authentication methods, which makes it easier for cross-layer authentication design optimization. The authentication method is also adapted to time-varying environment as the tag can be generated without CSI	1. The authentication performance of SUP-based PLA is constrained by message recover error 2. For the BSUP-based PLA, the accuracy of channel estimation is deteriorated in message recovery for an unaware receiver as it scheme corrupts the pilot structure	Small-scale networks with relatively good channel conditions

1) The Applications of Distributed Multiple Antennas in PLA: MIMO techniques have been widely used in the current wireless communication systems. Compared with single-antenna PLA structure, the authentication performance can be further enhanced in multiuser communication networks by collecting and jointly estimating physical-layer information received by multiple distributed users/landmarks at different locations. Especially for channel-based PLA schemes, the spatial resolution of the radio channel increases. In Ref. [106], an RSS-based authentication system was proposed to detect multiple spoofing attacks by using spatial correlation information of multiple landmarks that improve the spoofing detection accuracy. SVM method is also considered to improve the accuracy of determining the number of attackers. The authors of Ref. [34] provided a message authentication scheme for cellular Internet of things (CIoT) networks based on the multiple channel characteristics between the source nodes and the anchor nodes, which aims to decide whether or not the current channel responses correspond to a certain legitimate source. Note that the systems mentioned above can be regarded as typical distributed multi-antenna authentication systems, where multiple landmarks are equipped with single antenna. In Ref. [104], the authors established a more practical authentication framework with unknown channel model by evaluating channel information collected by multi-landmarks. Based on the general system model, logistic regression with the Frank-Wolfe (FW) algorithm is applied to exploit the received signal strength indicators collected at multiple antennas by multiple landmarks to discriminate transmitters even though the channel distribution is not aware of, where the FW algorithm aims to estimate the coefficient of the logistic regression model. Furthermore, to reduce the overall communication overhead between the landmarks and the security agent, distributed FW algorithm is applied instead of the FW algorithm. Based on the works in Ref. [104], another efficient authentication scheme was proposed in Ref. [105] by using an online method called incremental aggregated gradient to solve the logistic regression problem for the same authentication system with multi-landmarks. The authors confirm that the IAG-based PLA further reduces the computation overhead while achieving a higher detection accuracy in comparison with the distributed FW-based scheme.

The current works on the PLA in distributed multi-antenna systems are mostly based on the framework with single authentication entity, which is supported by multiple distributed assisted nodes to collect physical-layer information. However, for multiuser networks, collaborative communication can be applied between distributed devices. Therefore, each node in the network can be either the authentication entity or the assisted node, and proper nodes selection and scheduling methods are required to achieve higher network authentication accuracy with limited energy consuming. In other words, more researches on the application of distributed multi-antenna schemes in PLA are needed in future study.

2) Delay Control in Mission Critical MTC: The applications of mission mission critical MTC to future wireless systems is becoming an increasingly attractive research topic in the area of wireless communication. Compared to common communication applications such as IEEE 802.11 based wireless systems or cellular 4G LTE networks, mission critical MTC has much higher requirements regarding reliability, availability and especially latency[107]. A. Weinand et al. firstly studied possible PLA schemes for mission critical MTC in Refs. [33] and [107], which are robust to active attacks such as spoofing or replying attack. The authentication approach is

basically channel-based, where the channel estimations at the receiver are clustered according to a Gaussian mixture model. However, the works in Refs. [33] and [107] are based on simple Alice-Bob-Eve security model and traditional machine-learning schemes. The requirements of mission critical MTC, such as the network delay, are not analyzed.

The authors of Ref. [102] concentrated on the delay performance of PLA for an uplink centralized MTC network with multiple wireless devices and an access point. The AP is assumed to be equipped with multiple antennas, which is responsible for processing feature-based authentication that compares the CSI associated with each transmission to a pre-stored feature database. The authors introduce a infinite-buffer queueing model to represent the data flow from each device to the AP, based on which the upper-bounds on delay violation probability performance is developed as a function of FAR and MDR by using the stochastic network calculus framework. It is claimed in Ref. [102] that despite some cost of increase in delay violation probability, the proposed PLA can effectively confront disassociation and Sybil attacks and ensure authentication security in mission-critical MTC systems.

3) Reliable Authentication in Cognitive Radio Networks: As a software defined radio, cognitive radio was proposed to utilize the vacant spectrum space amongst the crowded channel efficiently by spectrum sensing. Security issues on cognitive radio have also attracted extensive attention from the researchers in this field since decades ago. For example, Burbank[108] and Clancy et al.[109] provided early surveys in 2008 that introduced key security concerns in cognitive radio networks, including the new threats and challenges, and related evolution to mitigation attacks.

Among all of the attack modes, primary user emulation attack (PUEA) is one of the major threats to the spectrum sensing, where the malicious user mimics the primary user's signal characteristics to make secondary users (SUs) erroneously identify the attacker as the legitimate primary user. PUEA can introduce extra interference in spectrum sensing and occupy the channel resources of legitimate SUs. Therefore, reliable PLA-based methods are required to overcome PUEA and improve the performance of cognitive radio networks.

The methods proposed in Refs. [110,111] are all based on the wireless channel characteristics. More specifically, in Ref. [110], the authors introduced an advanced PUEA scheme with estimation and learning method, based on which a defense solution is proposed by estimating the invariant of a communication channel (i.e., the variance of the received signal power) to distinguish the identities of different transmitters. The authors assumed a channel model with the path loss and the log-normal shadowing, thus only energy sensing is considered in Ref. [110]. The authors of Ref. [111] proposed a channel-based detection method for OFDM cognitive radio networks in multipath Rayleigh fading channels to distinguish the primary user emulate attacker from the primary user (PU) by using Neyman-Pearson test.

In comparison with channel-based methods, the applications of RF-based PLA techniques in cognitive radio networks are more widely studied. An RF tag-based authentication method was proposed in Ref. [112] where the one-way hash chain is used to generate the tag which is then embedded in quadrature phase shift keying (QPSK) modulation scheme and quadrature amplitude modulation. The authors of Ref. [113] utilized the transmitter location fingerprints extracted from estimation of the PSD.

There are also few works considering the scenario with multiple SUs[114,115]. Here, the information received by different SUs can be jointly used to enhance the detection performance, which is similar to the idea of distributed multi-antenna techniques. For example, in Ref. [115], local information about the PU is analyzed by the SUs, and then they are exchanged among the SUs. A cooperative spectrum sensing system is established to combat PUEA based on the energy characteristics of a weighted sum of the signals received from cooperative spectrum sensing system SUs in fusion center[114]. Rather than detecting attack users, the scheme aims to detect the presence of legitimate PUs with existing PUEA signals.

Above all, despite plenty of research works focusing on the issues of PLA in cognitive radio networks, there are still several open problems that need to be investigated. On one hand, the types of attack modes considered in most studies are simple. Except for the PUEA, limited studies are carried out about other attacks as in Ref. [110]. On the other hand, the structures of cognitive radio networks are assumed to be homoplastic. The studies about more dynamic and complex network structures are required as future research works.

4) The Applications of Deep Learning in Multiuser PLA: Deep learning (DL), as one of the hottest machine learning research topics, has been expansively studied and applied in artificial intelligence fields like computer vision, image classification, and multiuser communications[116,117]. There are already several research works that concentrate on the applications of DL algorithms to improve the wireless network security. For example, in Ref. [118] the authors presented a novel DL-based indoor fingerprinting based method to realize indoor localization by using CSI. DL can also be used for channel prediction, such as the Rayleigh fading channel prediction approach with deep neural networks proposed in the work [119]. In Ref. [120], the authors developed a DL-based PLA framework to enhance the security of industrial wireless sensor networks (IWSNs). The authentication system includes multiple sensor nodes in the different locations of the industrial scene, which has been identified by the upper layer authentication to

facilitate labeling the corresponding CSI before communication. Based on the system model, the authors proposed a rapid deep neural network (DNN) based PHY-layer authentication algorithm to meet the low latency requirements of industrial wireless sensor networks. It is clarified and demonstrated by the authors the proposed methods can enhance the security of the industrial wireless network without sacrificing communication resources.

DL-based authentication methods are also used to enhance security in mobile edge computing (MEC). For example, the authors of Ref. [121] proposed a DL-based PLA scheme which exploits CSI to enhance the security of MEC system via detecting spoofing attacks in wireless networks. After the communication node sends the identity information to the MEC server, the MEC server should require for the upper layer protocol authentication from the authentication center. Then the uplink PHY-layer authentication is performed on the MEC server after the protocol authentication is confirmed. The DL approach is applied to complete channel authentication and spoofer detection in an MEC system, where the MEC server estimates the CSIs of each edge nodes by pilots, and processes the CSIs to the input samples of DNN. The authors demonstrated that the proposed method can achieve efficient multi-user authentication with smaller computation overheads and lower energy consumptions compared with the conditional hypothesis test approach.

In a word, DL-based PLA is becoming a promising authentication approach that can be applied in multiuser communication networks. However, the works focusing on DL-based PLA are still limited, which remains for further study.

B. The Internet of Things Network

As the next wave of technological evolution, the IoT has increasingly attracting attention due to its ability of pervasively supporting wireless devices for diverse end-to-end (E2E) communication services via the Internet. Due to the hardware restraints of most IoT devices, such as limited battery lifetime and computational capability in wireless sensor networks, standardized transmission and conventional authentication framework seem to be too expensive and redundant. Therefore, PLA is a promising substitution to conventional authentication scheme to guarantee secure communication. Current studies on PLA are mostly based on the traditional Alice-Bob-Eve model. However, several extra problems should be considered before commercial applications of PLA in practical IoT network, some of which are listed as follows.

- Extending the device-to-device PLA to E2E PLA.
- PLA handover in dynamic environment, such as ad hoc or 5G heterogeneous networks.
- The integration of PLA and cryptography-based authentication structure.

1) Cross-Layer Authentication for E2E Networks: In Ref. [4] the authors introduce a general architecture for cross-layer authentication in E2E communication in future complex heterogeneous networks. A cross-layer authentication framework is further studied in detail in Ref. [40] by the same research group, where an enhanced E2E authentication framework is proposed to realize seamless integration of PLA into traditional asymmetric cryptography-based authentication schemes. More specifically, the authors consider a typical IoT network consisting of a source node, some collaborative nodes, and a destination node at the other communication end. The secure authentication consists of two phases. During the first phase (which is referred to as the registration phase), multiple RF features (CFO and IQI) of the source node are received and estimated by the selected collaborative nodes. The kernel of the registration phase is to generate a unique physical-ID by using the information of quantified CFO/IQI along with the upper-layer public key, which also provides additional physical entropy to protect keys. As the physical features are observed by different collaborative nodes, the technique MAMO is also used. The second phase is authentication, where both of physical and cryptography-based authentication are proceeded in the selected collaborative nodes and destination node. The main process of physical-layer authentication is extracting and verifying the quantified CFO/IQI of the transmitted node to exclude the easily detectable illegitimate source nodes. Then cryptography-based authentication is performed based on the typical one-way hash digital signature scheme. The authors of Ref. [40] proposed a physical-layer-enhanced identity-based cryptography (IBC) system with the PHY-IBC-based key protection for E2E communication in IoT networks.

2) Reliable Authentication Handover in Heterogeneous Networks: Besides of the challenges above, with the rapid growth of the wireless infrastructure scale in future complex heterogeneous networks, frequent and seamless authentication handover is also required when mobile users switch between different BSs or access points. The work in Ref. [40] mainly focused on the PLA of a single fixed IoT communication link. In Ref. [4], the authors only provided a general thought of authentication handover for 5G heterogeneous networks, where prediction and sharing of multiple physical attributes are used as security context to monitor and track the real-time moving of the user. Therefore, the authentication server can generate physical-key information and send it to the serving AP of the next cell in advance. The authors of Ref. [34] explored an efficient PLA scheme for software-defined-radio (SDN) enabled heterogeneous 5G network handover test with the application of the nonparametric Kolmogorov-Smirnov test. The authentication management module is installed in the SDN controller. Similarly, the key

technique of authentication is monitoring and predicting the positions of mobile users for seamless handover authentication between relevant access points and BSs. Such authentication handover method in Ref. [34] is actually an extension of channel-based PLA, but relies on the characteristics of more channel-based or RF-based features.

The works above are some of the attempts on PLA handover. However, more detailed techniques, such as user location predicting algorithm and PHA information sharing agreement, are still limited. The applications of seamless authentication handover still remains as a challenging research topic for the scholars in this field.

3) Adaptive Authentication for Dynamic Environment: For time-varying wireless communication environment such as ad hoc networks, the dynamic conditions of the network and the mobility of devices usually result in non-symmetrical observations at the transmitter and receiver, and therefore give rise to challenges on PLA. As discussed in section III.A, some researchers have already proposed potential solutions to the problem, mostly by using channel-tracking[33,35,39,55,57,73]. For example, the authors of Ref. [35] proposed a simple PLA method for ad hoc wireless sensor networks with the aid of symmetric cryptography. The PLA is based on the channel information, where the channel response in the current time is compared with a series of channel information received and stored previously. To adapt to the channel variation, the time interval of two consecutive authentication procedures is compared with channel's coherence time. If the former one is larger, the first frame of the message should be re-authenticated via the conventional cryptography-based digital signature scheme. The method above is straightforward, but the authentication performance is poor and inefficient in more complex communication environment. Basically, the variations of physical layer attributes are unknown for the designers and become harder to predict and track in extremely dynamic environment. To further enhance the authentication performance, multiple physical layer attributes may be taken into account, which also aggravate the challenge above.

In Ref. [122], the authors proposed an adaptive PLA framework for dynamic time-varying environment by tracking multiple physical layer attributes based on the kernel-based machine learning technique. Different from the channel-tracking methods in section III.A, here the variations of physical attributes and the time-varying environment is learnt by the kernel machine learning, which is an intelligent process to achieve reliable authentication through discovering the complex dynamic environment encountered. More specifically, let $\boldsymbol{H}(t) = [H_1(t), H_2(t), \cdots, H_N(t)]$ be the N imperfect observations of physical attributes at the receiver during the tth transmit duration, then a hypothesis testing problem is formulated by comparing $\mathscr{F}(\boldsymbol{H}(t) - \boldsymbol{H}(t-1))$ with a predetermined threshold δ, where $\mathscr{F}(\cdot)$ is the intelligent adaptive function promptly updated according to the variations of the environment. Clearly, the proposed adaptive authentication scheme enhances the authentication performance in time-varying environment by tracking and studying the differences of multiple physical attributes between adjacent authentication time rather than the physical attributes themselves. Compared with traditional multiple RF-based schemes and channel-tracking methods based on simple time-varying channel models, the method proposed in Ref. [122] extends the thoughts of channel-based and multiple RF-based PLA to a general MAMO authentication framework considering more complex communication environment with multiple time-varying physical attributes. Meanwhile, the machine learning method is more complicated as an optimization problem should be solved during each training period. The proposed authentication method in Ref. [122] only consider three physical attributes (i.e., the carrier CFO, CIR, and RSS indicator) using Gaussian kernel function. Therefore, feature selections with more physical attributes and the application of more advanced and efficient feature-tracking schemes can be studied to develop enhanced adaptive authentication methods for dynamic environment.

V. Conclusion

This paper has provided a comprehensive survey of the field of PLA in wireless communication networks. We firstly investigated the background, fundamentals and attack models of key-based and key-less PLA, based on which representative research results and key approaches of three typical PLA architectures: the authentication based on channel information, RF features of devices, and identity watermarks are introduced in the following. It is observed that channel-based PLA is more sensitive to the variance of communication environment and device mobility, while RF fingerprint-based schemes facing the challenges of reliable feature selection and classification problems in complex networks. Watermark embedding-based PLA has the similar framework as traditional cryptography-based methods such as digital signature, but the security is guaranteed according to key equivocation. Finally, we discussed potential research trends of PLA in future multiuser communication networks such as cognitive radio and IoT, and demonstrated that PLA is a promising solution to cope with security challenges in future device communications and enhance the authentication security in conjunction with traditional upper-layer authentication frameworks.

References

[1] DANEV B, ZANETTI D, CAPKUN S. On physical-layer identification of wireless devices[J]. ACM Comput. Surv, 2012, 45(1).

[2] BALDINI G, STERI G. A survey of techniques for the identification of mobile phones using the physical fingerprints of the built-in components[J]. IEEE Commun. Surveys Tuts, 2017, 19(3): 1761-1789.
[3] XU Q, ZHENG R, SAAD W, et al. Device fingerprinting in wireless networks: Challenges and opportunities[J]. IEEE Commun. Surveys Tuts, 2015, 18(1): 94-104.
[4] WANG X, HAO P, HANZO L. Physical-layer authentication for wireless security enhancement: Current challenges and future developments[J]. IEEE Commun. Mag, 2016, 54(6): 152-158.
[5] MUKHERJEE A, FAKOORIAN S A, HUANG J, et al. Principles of physical layer security in multiuser wireless networks: A survey[J]. IEEE Commun. Surveys Tuts, 2014, 16(3): 1550-1573.
[6] LIU Y, CHEN H H, WANG L. Physical layer security for next generation wireless networks: Theories, technologies, and challenges[J]. IEEE Commun. Surveys Tut, 2017, 19(1): 347-376.
[7] JORSWIECK E, TOMASIN S, SEZGIN A. Broadcasting into the uncertainty: Authentication and confidentiality by physical-layer processing[J]. Proc. IEEE, 2015, 103(10): 1702-1724.
[8] MASSEY J L. An introduction to contemporary cryptology[J]. Proc. IEEE, 1988, 76(5): 533-549.
[9] BARENGHI A, BREVEGLIERI L, KOREN I, et al. Fault injection attacks on cryptographic devices: Theory, practice, and countermeasures[J]. Proc. IEEE, 2012, 100(11): 3056-3076.
[10] SHANNON C E. Communication theory of secrecy systems[J]. Bell Syst. Tech. J., 1949, 28(4): 656-715.
[11] WYNER A D. The wire-tap channel[J]. Bell Syst. Tech. J., 1975, 54(8): 1355-1367.
[12] CSISZAR I, KORNER J. Broadcast channels with confidential messages[J]. IEEE Trans. Inf. Theory, 1979, 24(3): 339-348.
[13] MAURER U M. Secret key agreement by public discussion from common information[J]. IEEE Trans. Inf. Theory, 1993, 39(3): 733-742.
[14] LIANG Y, POOR H V, SHAMAI S. Information theoretic security[J]. Foundations and Trends in Communications and Information Theory, 2009, 5(4-5): 355-580.
[15] BLOCH M, BARROS J. Physical-layer security: From information theory to security engineering[M]. Cambridge: Cambridge University Press, 2011.
[16] GILBERT E N, MACWILLIAMS F J, SLOANE N J A. Codes which detect deception[J]. Bell Syst. Tech. J., 1974, 53(3): 405-424.
[17] BRICKELL E F. A few results in message authentication[J]. Congr. Numerantium, 1984, 43: 141-154.
[18] SGARRO A. Blind coding: Authentication frauds from the point of view of rate-distortion theory[J]. Journal of Discrete Mathematical Sciences and Cryptography, 2001, 4(2-3).
[19] MAURER U. Authentication theory and hypothesis testing[J]. IEEE Trans. Inf. Theory, 2000, 46(4): 1350-1356.
[20] TUGNAIT J K, KIM H. A channel-based hypothesis testing approach to enhance user authentication in wireless networks[C]//Proceedings of IEEE International Conference on Communication systems and networks (COMSNETS). Piscataway: IEEE Press, 2010: 1-9.
[21] PEI C, ZHANG N, SHEN X S, et al. Channel-based physical layer authentication[C]//Proceedings of IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2014: 4114-4119.
[22] LIU J, WANG X. Physical layer authentication enhancement using two-dimensional channel quantization[J]. IEEE Trans. Wireless Commun, 2016, 15(6): 4171-4182.
[23] FERRANTE A, LAURENTI N, MASIERO C, et al. On the error region for channel estimation-based physical layer authentication over Rayleigh fading[J]. IEEE Trans. Inf. Forensics Security, 2015, 10(5): 941-952.
[24] TOMASIN S. Analysis of channel-based user authentication by keyless and key-based approaches[J]. IEEE Trans. Wireless Commun, 2018, 17(9): 5700-5712.
[25] XIAO L, GREENSTEIN L, MANDAYAM N, et al. MIMO-assisted channel-based authentication in wireless networks[C]//Proceedings of 42nd Annual Conference on Information Sciences and Systems (CISS). Piscataway: IEEE Press, 2008: 642-646.
[26] BARACCA P, LAURENTI N, TOMASIN S. Physical layer authentication over MIMO fading wiretap channels[J]. IEEE Trans. Wireless Commun, 2012, 11(7): 2564-2573.
[27] PAN F, WEN H, LIAO R, et al. Physical layer authentication based on channel information and machine learning[C]//Proceedings of IEEE Conference on Communications and Network Security (CNS). Piscataway: IEEE Press, 2017: 364-365.
[28] WU X, YANG Z. Physical-layer authentication for multi-carrier transmission[J]. IEEE Commun. Lett, 2015, 19(1): 74-77.
[29] XIAO L, GREENSTEIN L, MANDAYAM N, et al. A physical-layer technique to enhance authentication for mobile terminals[C]//Proceedings of IEEE International Conference on Communications. Piscataway: IEEE Press, 2008: 1520-1524.
[30] XIAO L, GREENSTEIN L J, MANDAYAM N B, et al. Channel-based spoofing detection in frequency-selective Rayleigh channels[J]. IEEE Trans. Wireless Commun, 2009, 8(12): 5948-5956.
[31] XIAO L, REZNIK A, TRAPPE W, et al. PHY-authentication protocol for spoofing detection in wireless networks[C]//Proceedings of IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2010: 1-6.
[32] SHAN D, ZENG K, XIANG W, et al. PHY-CRAM: Physical layer challenge-response authentication mechanism for wireless networks[J]. IEEE J. Sel. Areas Commun, 2013, 31(9): 1817-1827.
[33] WEINAND A, KARRENBAUER M, LIANGHAI J, et al. Physical layer authentication for mission critical machine type communication using Gaussian mixture model based clustering[C]//Proceedings of IEEE Vehicular Technology Conference (IEEE VTC Spring). Piscataway: IEEE Press, 2017: 1-5.
[34] MA T, HU F, MA M. Fast and efficient physical layer authentication for 5G HetNet handover[C]//Proceedings of International Telecommunication Networks and Applications Conference (ITNAC). Piscataway: IEEE Press, 2017.
[35] WEN H, HO P H, QI C, et al. Physical layer assisted authentication for distributed ad hoc wireless sensor networks[J]. IET Inf. Secur, 2010, 4(4): 390-396.
[36] CHOI J. A coding approach with key-channel randomization for physicallayer authentication[J]. IEEE Trans. Inf. Forensics Security, 2019, 14(1): 175-185.
[37] WU X, YANG Z, LING C, et al. Artificial-noise-aided physical layer phase challenge-response authentication for practical OFDM transmission[J]. IEEE Trans. Wireless Commun, 2016, 15(10): 6611-6625.
[38] DU X, SHAN D, ZENG K, et al. Physical layer challenge-response authentication in wireless networks with relay[C]//Proceedings of IEEE INFOCOM. IEEE: Press, 2014: 1276-1284.
[39] VAERENBERGH S V, GONZÁLEZ Ó, VÍA J, et al. Physical layer authentication based on channel response tracking using Gaussian processes[C]//Proceedings of IEEE Acoust. Speech Signal Process. (ICASSP). Piscataway: IEEE Press, 2014: 2410-2414.
[40] HAO P, WANG X, SHEN W. A collaborative PHY-aided technique for end-to-end IoT device authentication[J]. IEEE Access, 2018, 6: 42279-42293.
[41] SHI Y, JENSEN M A. Improved radiometric identification of wireless devices using MIMO transmission[J]. IEEE Trans. Inf. Forensics Security, 2011, 6(4): 1346-1354.

[42] YU P L, VERMA G, SADLER B M. Wireless physical layer authentication via fingerprint embedding[J]. IEEE Commun. Mag, 2015, 53(6): 48-53.

[43] VERMA G, YU P, SADLER B. Physical layer authentication via fingerprint embedding using software-defined radios[J]. IEEE Access, 2015, 3: 81-88.

[44] XIE N, ZHANG S. Blind authentication at the physical layer under time-varying fading channels[J]. IEEE J. Sel. Areas Commun, 2018, 36(7): 1465-1479.

[45] YU P L, BARAS J S, SADLER B M. Physical-layer authentication[J]. IEEE Trans. Inf. Forensics Security, 2008, 3(1): 38-51.

[46] YU P L. SADLER B M. MIMO authentication via deliberate fingerprinting at the physical layer[J]. IEEE Trans. Inf. Forensics Security, 2011, 6(3): 606-615.

[47] DANEV B, HEYDT-BENJAMIN T, ČAPKUN S. Physical-layer identification of RFID devices[C]//Proceedings of 18th Conf. USENIX Security Symp. Piscataway: IEEE Press, 2009: 199-214.

[48] MAIO D, MALTONI D, CAPPELLI R, et al. FVC2000: Fingerprint verification competition[J]. IEEE Trans. Pattern Anal. Machine Intell, 2002, 24: 402-412.

[49] BOLLE R M, CONNELL J H, PANKANTI S, et al. Guide to biometrics[M]. Berlin: Springer-Verlag, 2003.

[50] SIMMONS G J. Authentication theory/coding theory[C]//Proceedings of Advances in Cryptology (CRYPTO'84). 1984: 411-431.

[51] SIMON M K, OMURA J K, SCHOLTZ R A, et al. Spread spectrum communications[M]. 3 vols. Rockville: Computer Science Press, 1985.

[52] VITERBI A J. CDMA: Principles of spread spectrum communication, Addison-Wesley wireless communications series[M]. MA: Addison-Wesley Publishing Company, 1995.

[53] JOHANNESSON R, SGARRO A. Strengthening Simmons' bound on impersonation[J]. IEEE Trans. Inform. Theory, 1991, 37: 1182-1185.

[54] XIAO L, GREENSTEIN L, MANDAYAM N, et al. Fingerprints in the ether: Using the physical layer for wireless authentication[C]//Proceedings of IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2007: 4646-4651.

[55] XIAO L, GREENSTEIN L, MANDAYAM N, et al. Using the physical layer for wireless authentication in time-variant channels[J]. IEEE Trans. Wireless Commun, 2008, 7(7): 2571-2579.

[56] LIU J, WANG X, TANG H. Physical layer authentication enhancement using maximum SNR ratio based cooperative AF relaying[J]. Wireless Communications and Mobile Computing, 2017, 2017: 1-16.

[57] LIU F J, WANG X, TANG H. Robust physical layer authentication using inherent properties of channel impulse response[C]//Proceedings of Military Communications Conference (MILCOM). Piscataway: IEEE Press, 2011: 538-542.

[58] CAPARRA G, CENTENARO M, LAURENTI N, et al. Energy-based anchor node selection for iot physical layer authentication[C]//Proceedings of IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2016: 1-6.

[59] SHAW D, KINSNER W. Multifractal modeling of radio transmitter transients for classification[C]//Proceedings of IEEE WESCANEX 97 Communications, Power and Computing. Piscataway: IEEE Press, 1997: 306-312.

[60] KNOX D A, KUNZ T. Secure authentication in wireless sensor networks using RF fingerprints[C]//Proceedings of IEEE/IFIP International Conference on Embedded and Ubiquitous Computing. Piscataway: IEEE Press, 2008, 1: 230-237.

[61] BRIK V, BANERJEE S, GRUTESER M, et al. Wireless device identification with radiometric signatures[C]//Proceedings of ACM MobiCom. New York: ACM, 2008.

[62] DANEV B, CAPKUN S. Transient-based identification of wireless sensor nodes[C]//Proceedings of International Conference on Information Processing in Sensor Networks. Piscataway: IEEE Press, 2009: 25-36.

[63] PENG H, LONG F, DING C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy[J]. IEEE Trans. Pattern Anal. Mach. Intell, 2005, 27(8): 1226-1238.

[64] DING C, PENG H. Minimum redundancy feature selection from microarray gene expression data[C]//Proceedings of IEEE Bioinformatics Conf. Piscataway: IEEE Press, 2003: 523-528.

[65] HOU W, WANG X, CHOUINARD J Y. Physical layer authentication in OFDM systems based on hypothesis testing of CFO estimates[C]//Proceedings of IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2012: 3559-3563.

[66] HOU W, WANG X, CHOUINARD J, et al. Physical layer authentication for mobile systems with time-varying carrier frequency offsets[J]. IEEE Trans. Commun, 2014, 62(5): 1658-1667.

[67] REISING D R, TEMPLE M A, JACKSON J A. Authorized and rogue device discrimination using dimensionally reduced RF-DNA fingerprints[J]. IEEE Trans. Inf. Forensics Security, 2015, 10(6): 1180-1192.

[68] DANEV B, LUECKEN H, CAPKUN S, et al. Attacks on physical-layer identification[C]//Proceedings of the 3rd ACM Conference on Wireless Network Security (WISEC). New York: ACM, 2010: 89-98.

[69] DOUCEUR J R. The Sybil attack[C]//Proceedings of International Workshop on Peer-to-Peer Systems (IPTPS). Cambridge: Springer, 2002: 251-260.

[70] CAO R, GRAVES E, WONG T F, et al. Detecting substitution attacks against non-colluding relays[C]//Proceedings of IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2013: 1856-1861.

[71] CLARKE R. A statistical theory of mobile-radio reception[J]. Bell System Technical Journal, 1968, 47(6): 957-1000.

[72] LIU J, REFAEY A, WANG X, et al. Reliability enhancement for CIR-based physical layer authentication[J]. Security Commun. Netw, 2014.

[73] LÁZARO-GREDILLA M, VAN VAERENBERGH S, LAWRENCE N D. Overlapping mixtures of Gaussian processes for the data association problem[J]. Pattern Recognition, 2012, 45(4): 1386-1395.

[74] LIU H, WANG Y, LIU J, et al. Practical user authentication leveraging channel state information[C]//Proceedings of ACM Symp. Inf. Comput. Commun. Security. New York: ACM, 2014: 389-400.

[75] URETEN O, SERINKEN N. Wireless security through RF fingerprinting[J]. Canadian Journal of Electrical and Computer Engineering, 2007, 32(1): 27-33.

[76] BARBEAU M, HALL J, KRANAKIS E. Detecting impersonation attacks in future wireless and mobile networks[J]. Lecture Notes in Computer Science, Springer-Verlag, 2006, 4074: 80-95.

[77] FARIA D B, CHERITON D R. Detecting identity-based attacks in wireless networks using signalprints[C]//Proceedings of the ACM Workshop on Wireless Security (WiSe). New York: ACM, 2006: 43-52.

[78] SHENG Y, TAN K, CHEN G, et al. Detecting 802.11 MAC-layer spoofing using received signal strength[C]//Proceedings of IEEE INFOCOM. Piscataway: IEEE Press, 2008: 1768-1776.

[79] TEKBAS ʻO H, SERINKEN N, ʻURETEN O. An experimental performance evaluation of a novel transmitter identification system under varying environmental conditions[J]. Can. J. Elect. Comput. Eng, 2004, 29(3): 203-209.

[80] DANEV B, CAPKUN S, MASTI R J, et al. Towards practical identi-

fication of HF RFID devices[C]//Proceedings of ACM TISSEC. New York: ACM, 2012.

[81] CANDORE A, KOCABAS O, KOUSHANFAR F. Robust stable radiometric fingerprinting for wireless devices[C]//Proceedings of IEEE Int. Workshop Hardware-Oriented Security Trust. Piscataway: IEEE Press, 2009: 43-49.

[82] REHMAN S UR, SOWERBY K, COGHILL C. RF fingerprint extraction from the energy envelope of an instantaneous transient signal[C]//Proceedings of Austral. Commun. Theory Workshop (AusCTW), 2012: 90-95.

[83] HALL J, BARBEAU M, KRANAKIS E. Detection of transient in radio frequency fingerprinting using phase characteristics of signals[C]//Proceedings of the 3rd IASTED International Conference on Wireless and Optical Communications (WOC). Piscataway: IEEE Press, 2003: 13-18.

[84] SUSKI W, TEMPLE M, MENDENHALL M, et al. Using spectral fingerprints to improve wireless network security[C]//Proceedings of IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2008: 1-5.

[85] URETEN O, SERINKEN N. Bayesian detection of radio transmitter turn-on transients[C]//Proceedings of NSIP99. Piscataway: IEEE Press, 1999: 830-834.

[86] URETEN O, SERINKEN N. Bayesian detection of WiFi transmitter RF fingerprints[J]. IET Electronics Letters, 2005, 41(6): 373-374.

[87] HALL J, BARBEAU M, KRANAKIS E. Enhancing intrusion detection in wireless networks using radio frequency fingerprinting (extended abstract)[C]//Proceedings of 3rd IASTED International Conference on Communications, Internet and Information Technology (CIIT). Piscataway: IEEE Press, 2004.

[88] YUAN H L, HU A Q. Preamble-based detection of Wi-Fi transmitter RF fingerprints[J]. IET Electron. Lett, 2005, 46(16): 1165-1167.

[89] ELLIS K J, SERINKEN N. Characteristics of radio transmitter fingerprints[J]. Radio Sci, 2001, 36(4): 585-597.

[90] RASMUSSEN K B, CAPKUN S. Implications of radio fingerprinting on the security of sensor networks[C]//Proceedings of Third International Conference on Security and Privacy in Communications Networks and the Workshops (SecureComm). Piscataway: IEEE Press, 2007: 331-340.

[91] POLAK A C, GOECKEL D L. Wireless device identification based on RF oscillator imperfections[C]//Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2014: 2679-2683.

[92] AZARMEHR M, MEHTA A, RASHIDZADEH R. Wireless device identification using oscillator control voltage as RF fingerprint[C]//Proceedings of IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE). Piscataway: IEEE Press, 2017: 1-4.

[93] GUYON I, ELISSEEFF A. An introduction to variable and feature selection[J]. J. Mach. Learn. Res, 2003, 3: 1157-1182.

[94] BOJINOV H, MICHALEVSKY Y, NAKIBLY G, et al. Mobile device identification via sensor fingerprinting[J]. arXiv:1408.1416, 2014.

[95] KLEIN R W, TEMPLE M A, MENDENHALL M J, et al. Sensitivity analysis of burst detection and RF fingerprinting classification performance[C]//Proceedings of IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2009: 1-5.

[96] QUINLAN J R. Bagging, boosting, and C4.5[C]//Proceedings of 13th Nat. Conf. Artifitial Intelligence. California: AAAI Press, 1996, 1: 725-730.

[97] CHEN Y, TRAPPE W, MARTIN R P. Detecting and localizing wireless spoofing attacks[C]//Proceedings of 2007 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks. Piscataway: IEEE Press, 2007.

[98] NGUYEN N T, et al. Device fingerprinting to enhance wireless security using nonparametric Bayesian method[C]//Proceedings of IEEE INFOCOM. Piscataway: IEEE Press, 2011: 1404-12.

[99] XIE N, CHEN C, ZHONG M. Security model of authentication at the physical layer and performance analysis over fading channels[J]. IEEE Trans. Dependable Secure Comput, 2018: 1-1.

[100] ZHANG S, LIEW S C, WANG H. Blind known interference cancellation[J]. IEEE J. Sel. Areas Commun, 2013, 31(8): 1572-1582.

[101] COVER T, THOMAS J. Elements of Information Theory[M]. Hoboken: Wiley-Interscience, 1991.

[102] FORSSELL H, THOBABEN R, AL-ZUBAIDY H, et al. Physical layer authentication in mission-critical MTC networks: A security and delay performance analysis[J]. IEEE J. Sel. Areas Commun, 2019, 37(4): 795-808.

[103] ANDREWS S, GERDES R M, LI M. Towards physical layer identification of cognitive radio devices[C]//Proceedings of IEEE Conference on Communications and Network Security (CNS). Piscataway: IEEE Press, 2017: 1-9.

[104] WAN X, XIAO L, LI Q, et al. PHY-layer authentication with multiple landmarks with reduced communication overhead[C]//Proceedings of IEEE International Conference on Communications (ICC). Piscataway: IEEE Press, 2017: 1-6.

[105] XIAO L, WAN X, HAN Z. PHY-layer authentication with multiple landmarks with reduced overhead[J]. IEEE Trans. Wireless Commun, 2018, 17(3): 1676-1687.

[106] YANG J, CHEN Y, TRAPPE W, et al. Detection and localization of multiple spoofing attackers in wireless networks[J]. IEEE Trans. Parallel Distrib. Syst, 2013, 24(1): 44-58.

[107] WEINAND A, KARRENBAUER M, SATTIRAJU R, et al. Application of machine learning for channel based message authentication in mission critical machine type communication[C]//Proceedings of 23th European Wireless Conference. Dresden: VDE, 2017: 1-5.

[108] BURBANK J L. Security in cognitive radio networks: the required evolution in approaches to wireless network security[C]//Proceedings of International Conference on Cognitive Radio Oriented Wireless Networks and Communications. Piscataway: IEEE Press, 2008: 1-7.

[109] CLANCY T C, GOERGEN N. Security in cognitive radio networks: Threats and mitigation[C]//Proceedings of International Conference on Cognitive Radio Oriented Wireless Networks and Communications. Piscataway: IEEE Press, 2008: 1-8.

[110] CHEN Z, COOKLEV T, CHEN C, et al. Modeling primary user emulation attacks and defenses in cognitive radio networks[C]//Proceedings of IEEE 28th International Performance Computing and Communications Conference. Piscataway: IEEE Press, 2009: 208-215.

[111] CHIN W L, TSENG C L, TSAI C S, et al. Channel-based detection of primary user emulation attacks in cognitive radios[C]//Proceedings of IEEE Vehicular Technology Conference (VTC Spring). Piscataway: IEEE Press, 2012: 1-5.

[112] BORLE K M, CHEN B, DU W. A physical layer authentication scheme for countering primary user emulation attack[C]//Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2013: 2935-2939.

[113] ZHAO C, XIE L, JIANG X, et al. A PHY-layer authentication approach for transmitter identification in cognitive radio networks[C]//Proceedings of International Conference on Communications and Mobile Computing (CMC). Piscataway: IEEE Press, 20120, 2: 154-158.

[114] YANG J, CHEN Y, SHI W, et al. Cooperative spectrum sensing against attacks in cognitive radio networks[C]//Proceedings of IEEE

International Conference on Information and Automation (ICIA). Piscataway: IEEE Press, 2014: 1-6.

[115] YUAN Z, NIYATO D, LI H, et al. Defense against primary user emulation attacks using belief propagation of location information in cognitive radio networks[C]//Proceedings of IEEE Communications and Networking Conference. Piscataway: IEEE Press, 2011: 599-604.

[116] ZHANG M, WEN C, JIN S, et al. A model-driven deep learning network for quantized GFDM receiver[J]. Journal of Communications and Information Networks, 2019, 4(3): 53-59.

[117] ZHAO Y, LIU S, XUE F, et al. DeepCount: Crowd counting with Wi-Fi using deep learning[J]. Journal of Communications and Information Networks, 2019, 4(3): 38-52.

[118] WANG X, GAO L, MAO S, et al. CSI-based fingerprinting for indoor localization: A deep learning approach[J]. IEEE Trans. Veh. Technol, 2017, 66(1): 763-776.

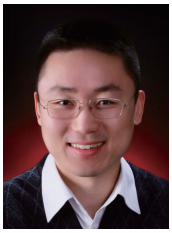
[119] LIAO R F, WEN H, WU J, et al. The Rayleigh fading channel prediction via deep learning[J]. Wireless Commun. Mobile Comput, 2018, 2018.

[120] LIAO R F, WEN H, WU J, et al. Deep-learning-based physical layer authentication for industrial wireless sensor networks[J]. Sensors, 2019, 19(11).

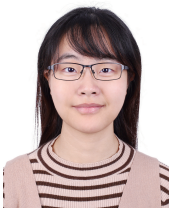
[121] LIAO R F, WEN H, WU J, et al. Security enhancement for mobile edge computing through physical layer authentication[J]. IEEE Access, 2019, 7:116390-116401.

[122] FANG H, WANG X, HANZO L. Learning-aided physical layer authentication as an intelligent process[J]. IEEE Trans. Commun, 2018, 67(3): 2260-2273.

About the Authors

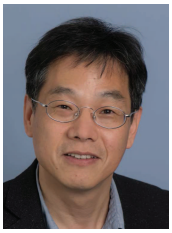


Lin Bai (M'13-SM'17) received his B.Sc. degree in electronic and information engineering from Huazhong University of Science and Technology, Wuhan, China, in 2004, his M.Sc. degree (Hons.) in communication systems from University of Wales, Swansea, U.K., in 2007, and his Ph.D. degree in advanced telecommunications from the School of Engineering, Swansea University, U.K., in 2010. Since 2011, he has been with Beihang University (Beijing University of Aeronautics and Astronautics, BUAA), Beijing, China, where he is currently a Professor at the School of Cyber Science and Technology. His research interests include multiple-input multiple-output (MIMO), Internet-of-things, and unmanned aerial vehicle communications. He authored two books published by Springer in 2012 and 2014. He was invited to serve as a symposium Co-Chair of IEEE GLOBECOM 2019 and a Tutorial Co-Chair of IEEE/CIC ICCC 2019. He has served as a Lead Guest Editor for IEEE Wireless Communications and a Guest Editor for IEEE Internet of Things Journal. Currently, he is on the Editorial Board of several journals, including IEEE Transactions on Signal Processing and IEEE Transactions on Wireless Communications, and serves as the Managing Editor of Journal of Communications and Information Networks. Prof. Bai is a Distinguished Lecturer of the IEEE Vehicular Technology Society and a Senior Member of the IEEE.

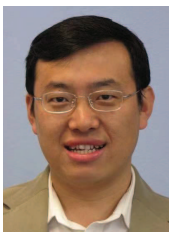


Lina Zhu (S'17) [corresponding author] received her B.Sc. degree in electronics and information engineering from Beihang University, Beijing, China, in 2016, and she is currently pursuing her Ph.D. degree in traffic science and communication engineering with the School of Electronic and Information Engineering, Beihang University, Beijing. Her current research interests include areas of wireless communications, including satellite communications and non-orthogonal-multiple-access (NOMA) systems.

Jianwei Liu received his B.Sc. and M.Sc. degrees in electronic and information from Shandong University, Shandong, China in 1985 and 1988. He received his Ph.D. degree in communication and electronic system from Xidian University, Shaanxi, China in 1998. Currently, he is a Professor with the School of Cyber Science and Technology, Beihang University, Beijing, China. His research interests include wireless communication network, cryptography, and network security.



Jinho Choi (SM'02) was born in Seoul, Republic of Korea. He received his B.E. (magna cum laude) degree in electronics engineering in 1989 from Sogang University, Seoul, and M.S.E. and Ph.D. degrees in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST) in 1991 and 1994, respectively. He is with the School of Information Technology, Burwood, Deakin University, Australia, as a Professor. Prior to joining Deakin University in 2018, he was with Swansea University, the United Kingdom, as a Professor/Chair in Wireless, and Gwangju Institute of Science and Technology (GIST), Korea, as a Professor. His research interests include the IoT, wireless communications, and statistical signal processing. He authored two books published by Cambridge University Press in 2006 and 2010. Prof. Choi received the 1999 Best Paper Award for Signal Processing from EURASIP, 2009 Best Paper Award for Signal Processing from EURASIP, 2009 Best Paper Award from WPMC (Conference), and is a Senior Member of IEEE. Currently, he is an Editor of IEEE Transactions on Communications and IEEE Wireless Communications Letters, and a Division Editor of Journal of Communications and Networks (JCN). He also served as an Associate Editor or Editor of other journals including IEEE Communications Letters, JCN, IEEE Transactions on Vehicular Technology, and ETRI journal.



Wei Zhang (F'15) is a Professor at University of New South Wales, Sydney, Australia. His current research interests include UAV communications, mmWave communications, space information networks, and massive MIMO. He is Editor-in-Chief of Journal of Communications and Information Networks. He also serves as Chair for IEEE Wireless Communications Technical Committee and Vice Director of IEEE Communications Society Asia Pacific Board. He is a Member of Board of Governors of IEEE Communications Society.