

RESEARCH ARTICLE

Specific Emitter Identification Based on Differential Constellation

GUANJIE ZHANG  **AND YANBIN LI**

54th Research Institute of China Electronics Technology Group Corporation (CETC54), Shijiazhuang 050081, China

Corresponding author: Guanjie Zhang (hbu.edu.cn.2005@163.com)


This work was supported in part by the China Postdoctoral Science Foundation under Grant 2021M693002, and in part by the State Key Program of Joint Funds of the National Natural Science Foundation of China under Grant U22B2002.

ABSTRACT The modulator distortion feature of the existing specific emitter identification (SEI) based on constellation contains the inherent carrier frequency offset of the receiver and transmitter, and the influence of the frequency offset on the feature distribution cannot be completely eliminated even by using high-precision frequency synchronization technology. Therefore, this paper presents a novel SEI method based on a differential constellation. On the basis of constructing the modulator distortion signal model, the demodulated signal is differentially processed to form a differential constellation. By comprehensively comparing the difference between the differential demodulation constellation and the ideal constellation, the maximum likelihood method is used to separate the frequency offset in the baseband signal from the modulator distortion feature vector. A new modulator distortion feature representation is designed, which completely eliminates the influence of the carrier frequency offset on the distortion feature distribution of the modulator. Subsequently, a random forest classifier based on a decision tree was constructed to learn the individual differences in the distortion features. Compared with existing identification methods, the fingerprint features extracted by this method are completely independent of the frequency offset of the signal examples, and the influence of the frequency offset on the recognition is eliminated. Our results show that the mean and variance of the feature vector distribution proposed in the method do not change with the frequency offset, the stable and high-precision identification of eight sources can be achieved under different carrier frequency offset conditions, and the accuracy can reach more than 90%.

INDEX TERMS Differential constellation, modulator distortion, radio-frequency fingerprinting (RFF), random forest, specific emitter identification (SEI).

I. INTRODUCTION

Specific emitter identification (SEI) is possible because of unique radio-frequency fingerprinting, which describes the differences between emitters of the same type. Owing to the benign hardware imperfections inherent to the analog components of emitters, the final radio-frequency (RF) signal inevitably parasitizes the unique radio-frequency characteristics of the emitter, which is also called radio-frequency fingerprinting. Radio-frequency fingerprinting is unique, independent of transmission, and difficult to forge; therefore, the identification technique has been widely adopted

The associate editor coordinating the review of this manuscript and approving it for publication was Filbert Juwono .

in military and civilian fields, such as battlefield spectrum management, wireless network security, and the Internet of things (IOT).

SEI is essentially a pattern recognition problem that focuses on fingerprinting extraction. In previous open research, features may be predefined or inferred. Predefined features are related to well-understood signal characteristics known prior to signal recording. The inferred features are extracted from the signals by means of some spectral transformations. According to the different sources of radio-frequency fingerprinting, SEI can be summarized into three categories: information-based identification, transient-based fingerprinting identification, and steady-state fingerprinting identification, as shown in Fig. 1. Information-based

identification belongs to the software-based recognition, which is a method of identification using communication connotation information, e.g. Wright [1], Guo and Chiueh [2]. Transient-based identification and steady-state identification belong to the physical layer waveform identification. Transient-based identification uses the transient waveform difference caused by benign hardware imperfections inherent to the emitters, e.g., Ureten and Serinken [3]. According to the different manifestations of the received signal samples, steady-state identification can be divided into waveform domain recognition, modulation domain recognition, and other domain recognition. Waveform domain techniques use signal samples from the time or frequency domain as the basic blocks of representation, e.g., Dbendorfer et al. [4] and Patel [5]. Modulation domain techniques use IQ samples to represent the most basic signals, e.g., Peng et al. [6], [7]. Other domain techniques use high-order transformation samples as the most basic representation blocks, e.g., Han et al. [8] and Yuan et al. [9].

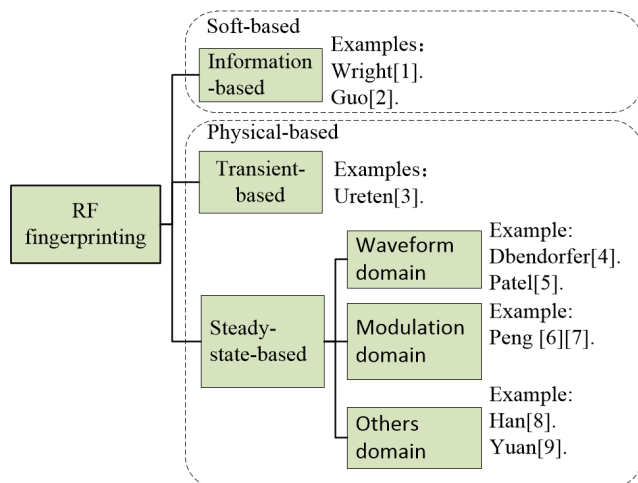


FIGURE 1. RF fingerprinting identification categories diagram.

A. INFORMATION-BASED IDENTIFICATION

In traditional wireless networks, device identification and authentication are primarily realized through mechanisms such as identity authentication. Wright [1], Guo and Chiueh [2], and Hall [10] discussed approaches for detecting the presence of multiple IEEE 802.11 devices using the same MAC address by analyzing frame sequence numbers. Similarly, Franklin et al. proposed a technique to identify devices based on differences in the MAC layer behavior, which depends on the combination of the chipset, firmware, and device driver [11]. Because these approaches are usually configured using bit-level information, they can be circumvented by changing the computer configuration or behavior, thus rendering the wireless network vulnerable to attack.

B. TRANSIENT-BASED IDENTIFICATION

Transient-based identification technology is an important physical layer waveform identification method. In previous

research, different transient-based identification methods have been adopted for different RF signals. Toonstra and Kinsner investigated the transient radiometric signatures of FM signals [12]. In addition, the effectiveness of transient identification methods for Bluetooth, radio frequency identification (RFID), and WSN devices has been demonstrated [13], [14], [15], [16], [17]. Ureten and Serinken constructed a probabilistic neural network to distinguish between eight IEEE 802.11b Wi-Fi cards by using a transient amplitude signal [3]. However, using transients for identification appears to be difficult, as indicated by the imperfect performance of existing schemes even in modest-sized evaluations.

C. STEADY-STATE IDENTIFICATION

Steady-state identification in the waveform, modulation, or other domains also plays an important role in SEI development. Waveform domain techniques use signal samples in the time and frequency domains as the basic blocks of representation, which provides the most flexibility at the cost of complexity. In earlier studies, Remley et al. identified different WLAN transmitters by observing waveform and spectral differences [18], proving the feasibility of waveform domain identification. However, direct recognition based on time-domain communication waveforms is susceptible to modulation symbol interference. Therefore, subsequent studies have demonstrated that the commonly used identification technology for WPAN devices (such as ZigBee and Z-Wave) requires dividing the fixed header portions of the signal (such as the preamble portions) into multiple samples of equal size to calculate the statistical values (such as variance, skewness, kurtosis), instantaneous amplitude, or phase [4], [5], [19], [20], and then using machine-learning classifiers to distinguish devices.

Unlike waveform domain identification, other domain identification techniques use domain-transformed signal samples as basic blocks of representation. Fingerprint features were obtained by exploring higher-order spectral transformation inference. Han et al. focused on a bi-spectrum-based method for the RF fingerprinting identification of communication transmitters [8]. Yuan et al. [9] and Zhang et al. [21] studied the individual identification methods of devices based on the Hilbert-Huang transform. Acosta et al. studied an RFID fingerprinting identification method based on wavelet transform [22]. Satija et al. investigated an emitter identification method based on variational mode decomposition (VMD) and spectrum under Rayleigh fading channels [23].

Modulation-domain identification techniques use I/Q signal examples and demodulation information to realize radiometric identification. This technique does not require signal samples to have a fixed header, and is not affected by random symbol modulation. Brik et al. proposed a passive radiometric device identification system (PARADIS) that differentiated 138 wireless devices with an accuracy of > 99% [24]. The characteristics are imperfections in the modulation domain,

such as frequency offset, I/Q origin offset, error vector magnitude, and phase errors.

Huang and Zheng investigated and analyzed RF fingerprinting based on constellation errors, constructed modulator distortion models, and accurately identified seven TDMA satellite terminals [25]. Pan investigated and analyzed RF fingerprinting based on signal trajectory images and identified seven model terminals [26]. Peng et al. proposed a ZigBee device identification algorithm based on a differential constellation trace figure (DCTF) to achieve 93.8% accuracy on 54 devices under the condition of 15 dB SNR [6], [7]. Liu and Doherty studied the recognition problem based on nonlinear power amplifier fingerprinting and constructed a nonlinear power amplifier model using Taylor series [27]. Polak and Goeckel focused on the recognition problem based on oscillator phase noise [28]. Zhang and Li proposed a novel SEI method based on feature diagram superposition, which can better reflect the fingerprint features of different emitters [29].

In recent years, deep learning has achieved a series of breakthroughs in machine vision and speech recognition, which has motivated scholars to use deep neural networks for modulation recognition [30], [31] and radar waveform recognition [32], [33]. As the differences between individual transmitters are subtle, SEI using deep learning is still at a nascent stage. Merchant et al. focused on a framework for training convolutional neural networks using time-domain complex baseband error signals and identified seven ZigBee devices with a recognition accuracy of 92.29% [34]. Restuccia et al. proposed a DeepRadioID identification system based on deep learning, which improves the accuracy of identification by dynamically optimizing the wireless channel [35]. Sankhe et al. [36], Riyaz et al. [37] proposed an ORACLE recognition system that directly trains a convolutional neural network using raw IQ signals and achieved 99% recognition accuracy for 16-bit similar devices under laboratory conditions. Pan et al. investigated a technique using Hilbert-Huang transformed examples to train a residual neural network to achieve high identification [38]. Liu et al. proposed a method that treats the signals at different times as signals of separate domains, and resolves the influence of time on the SEI by eliminating the factors of different domains [39]. It should be noted that the architecture of a CNN is more suitable than waveforms for extracting features from images. This implies that the aforementioned methods cannot maximize the powerful self-learning capabilities of CNN.

In existing steady-state recognition methods, radiometric identification technology based on the waveform domain usually uses only the header part of the RF signal to avoid interference from the modulation symbol. Radiometric identification technology based on other domains not only relies on the existing tools of signal processing but also faces the problem of excessive computation of feature extraction. Radiometric identification technology based on deep learning also faces the interference of random symbol modulation and

cannot physically interpret the recognition results; therefore, it cannot maximize identification performance. Unlike the aforementioned methods, modulation domain identification technology adopts the method of demodulation before extraction, converts the feature extraction object into constellation, and then compares the differences between demodulation constellation and ideal constellation to achieve the extraction of subtle features. The existing methods usually directly extract the modulator distortion feature after compensating for the carrier frequency deviation, but none of the carrier frequency offset compensation methods can completely eliminate the frequency offset.

To overcome the influence of carrier frequency offset on identification performance, this paper proposes a novel identification algorithm based on modulator distortion characteristics, which belongs to modulation domain identification technology. By comprehensively comparing the difference between the differential demodulation constellation and the ideal constellation, a new representation vector of the distortion characteristics of the modulator is designed based on the received signal preprocessing, which ensures that the distortion feature vector is independent of the carrier frequency offset. Then, a random forest classifier based on a decision tree was constructed to learn the individual differences in the distortion fingerprinting vector of the modulator, and individual recognition based on the IQ distortion feature was realized.

The main contributions are summarized as follows:

1) The differential constellation is used as the basic representation block of feature extraction, which lays the foundation for RF fingerprinting extraction of the modulator independent of the carrier frequency offset. To the best of our knowledge, this is the first attempt to use a differential constellation to extract the distortion fingerprints of the modulator.

2) By comprehensively comparing the difference between the difference decomposition constellation and the ideal constellation, the carrier frequency offset in the baseband signal is separated from the distortion characteristics of the modulator, a new modulator distortion vector is designed, and its calculation process is derived in detail for the first time.

3) We further investigated the performance of our algorithm in the presence of different carrier frequency offsets. The simulation results show that the proposed approach has a stable accuracy under different carrier frequency offsets and strong practicability and robustness.

Table 1 summarizes the most relevant related works, including data on the scale of evaluation.

The remainder of this paper is organized as follows. Section II briefly describes the proposed signal model. Section III presents an identification method based on the distortion characteristics of the modulator. Section IV presents the experimental results. Finally, we conclude the paper in Section V.

TABLE 1. Comparison of this paper with related ideas.

Technique	Feature Type	Classifier type	Identity Model	Evaluation scale
Franklin et. al.[11]	Soft-based meas.	N/A	Compliance with 802.11 standard	17 802.11 NICs
Toonstra et al.[12]	RF fingerprinting	Neural Networks	Transient properties	5 radio transmitters
V. Brik et. al.[24]	RF fingerprinting	Machine learning	Modulation accuracy; Constellation	138 802.11 NICs
Y Huang et. al.[25]	RF fingerprinting	Machine learning	Modulation accuracy; Constellation	7 TDMA satellite terminals
L. Peng et. al.[6]	RF fingerprinting	Deep learning	Modulation accuracy; Differential Constellation	54 ZigBee devices
X. Zhang et. al.[29]	RF fingerprinting	Deep learning	Modulation accuracy; Constellation	7 model terminals
Y. Pan [26]	RF fingerprinting	Machine learning	Modulation accuracy; Constellation	7 model terminals
This paper	RF fingerprinting	Machine learning	Modulation accuracy; Differential Constellation	8 model terminals

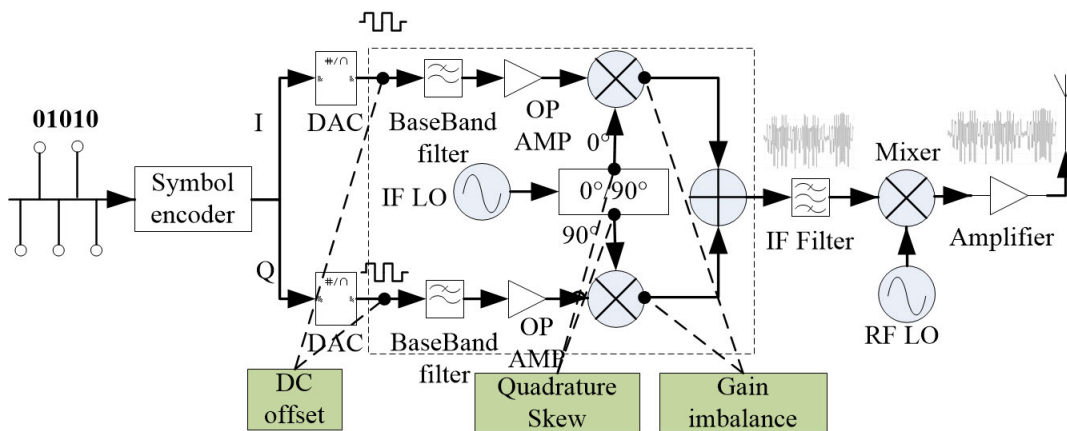


FIGURE 2. Common transmitter impairments and their sources.

As a general convention, we use the following notations throughout the paper: the real part of the complex number, $\Re\{\cdot\}$, conjugate, $\{\cdot\}^*$, Transpose, $\{\cdot\}^T$, and sum, $\text{sum}(\cdot)$.

II. SIGNAL MODEL

Owing to the benign hardware imperfections inherent to the analog components of the transmitter, the signal contains hardware fingerprinting of the transmitter. Fig. 2 presents a typical transmitter design and illustrates the likely causes of common impairments. The portions before the digital-to-analog transition of the transmitter are digital components and their imperfections do not exist. The portions after the digital-to-analog converter are analog components, which have benign hardware imperfections inherent to normal variations in the physical properties of such components. This section focuses on the distortion of the IQ modulator and describes the sources of the distortion of the modulator by constructing a signal-generation model of the IQ modulator with impairments.

The output signal of an ideal modulator has characteristics of equal amplitude, complete quadrature, and zero mean. However, in the actual transmitter path, distortion of the IQ modulator causes IQ imbalance and DC, which is unavoidable. Therefore, the distortion of the modulator can characterize the differences in the emitters. Impairments from the modulator include (1) I/Q gain mismatch, where the amplitudes of the I/Q signals are different; (2) quadrature errors, where the phase difference between the IQ signals

is not equal to 90°; and (3) DC offset, generated by carrier leakage from the mixer.

The output signal of the I/Q modulator with impairments is expressed as:

$$Z(t) = s_I(t) \cos(2\pi f_c t + \varsigma/2) - s_Q(t) \sin(2\pi f_c t - \varsigma/2) \tag{1}$$

where

$$s_I(t) = G_{I/Q} \sum_{k=-\infty}^{\infty} I_k h(t - kT) + O_I(t),$$

$$s_Q(t) = \sum_{k=-\infty}^{\infty} Q_k h(t - kT) + O_Q(t),$$

$s_I(t)$ and $s_Q(t)$ denote the baseband signals. f_c is the carrier frequency of the transmitter. ς denotes quadrature error. The $G_{I/Q} = G_I/G_Q$ is an I/Q gain imbalance. T is the symbol period. $h(t)$ represents the shaping pulse. I_k and Q_k are encoded symbols on the I and Q paths.

$O_I(t)$ and $O_Q(t)$ are the DC offsets of paths I and Q respectively. Normally, the DC offsets $O_I(t)$ and $O_Q(t)$ are constants, that is, $O_I(t) \equiv O_I, O_Q(t) \equiv O_Q$.

The output complex signal representation respectively is

$$Z(t) = \text{Re} \left\{ (\mu_1 \rho(t) + \mu_2 \rho^*(t) + \xi) e^{j(2\pi f_c t + \phi)} \right\} \tag{2}$$

where,

$$\mu_1 = 0.5 (G_{I/Q} + 1) \cos(\varsigma/2) + 0.5j (G_{I/Q} - 1) \sin(\varsigma/2),$$

$$\mu_2 = 0.5 (G_{I/Q} - 1) \cos(\varsigma/2) + 0.5j (G_{I/Q} + 1) \sin(\varsigma/2),$$

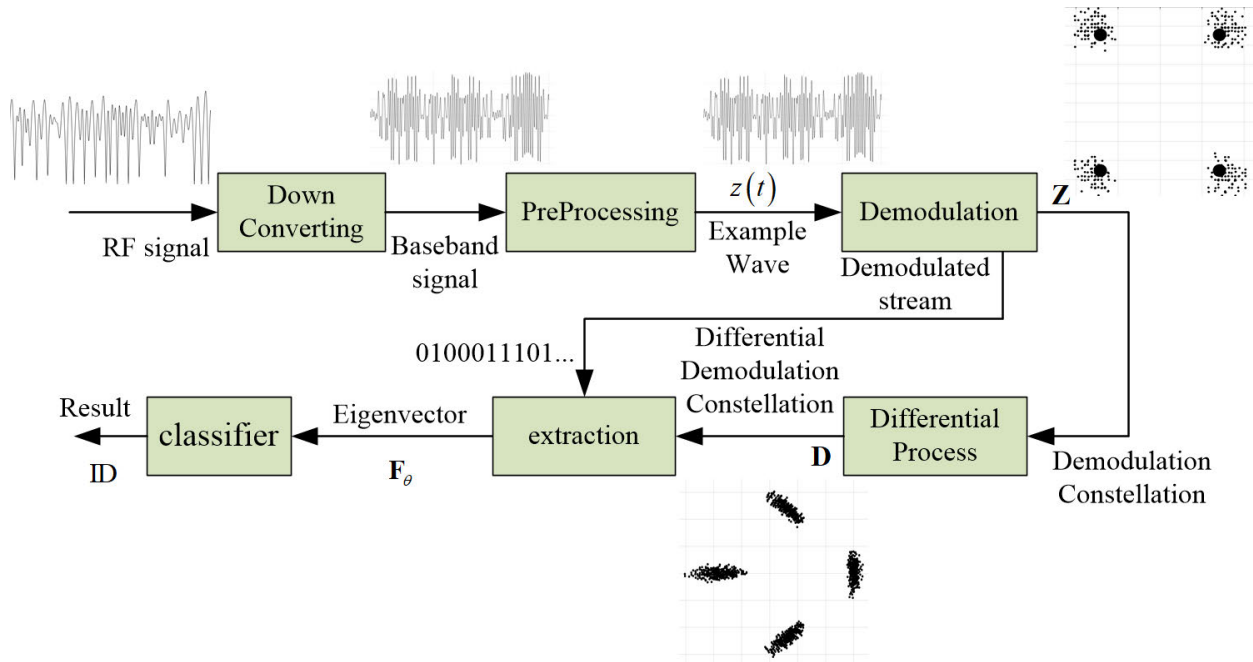


FIGURE 3. Block diagram of classifier based on RF fingerprinting.

$\xi = G_{I/Q} \cdot O_I + jO_Q$ is a complex direct current(DC) offset. ϕ is the initial phase. $\rho(t)$ denotes the complex baseband signal, for MPSK signals, it is expressed as:

$$\rho(t) = \sum_{k=-\infty}^{\infty} c_k h(t - kT - \tau),$$

$c_k = I_k + j \cdot Q_k$ is a complex modulation symbol.

Equation (2) represents a complex signal-generation model for a modulator with impairments. We observed from (2) that the output signal is closely related to the transmission frequency f_c , complex baseband signal $\rho(t)$, time t , and modulator impairment characteristics denoted by μ_1 , μ_2 and ξ .

III. FEATURES EXTRACTION AND IDENTIFICATION

RF signals are affected by many factors during the propagation process, resulting in differences in the amplitude, frequency, and phase of the signal. In this study, timing, synchronization, and difference processing are carried out in the process of feature extraction, and we design a new feature vector of the modulator, whose distribution is free from the influence of the carrier frequency offset, random symbol modulation, and other factors.

After receiving the RF signal, the receiver performs pre-processing such as down-conversion, normalization, filtering, and time synchronization to convert the RF signal into a baseband waveform signal. Different baseband signals inevitably carry different carrier frequency offsets. The baseband signal is sent to the forward demodulation module to obtain the demodulation code stream and the corresponding demodulation constellation carrying carrier frequency offsets, which are processed differently between symbols to obtain the

differential constellation such that the feature extraction object is converted from the baseband signal waveform into a differential constellation. Then, the demodulation code stream and differential demodulation constellation are used to calculate the maximum likelihood estimation, and the distortion characteristic vector of the IQ modulator is obtained independently of the carrier frequency offset. Finally, the distortion features were sent to the trained random forest classifier for identification. The recognition process is illustrated in Fig. 3.

A. PREPROCESSING

In the process of signal propagation and reception, unstable factors such as channel changes, carrier frequency changes, and differences in the acquisition time. These unstable factors affect the amplitude, delay, frequency, and phase of a signal at different times. These influences may cause instability in fingerprinting generated by the same emitter. Preprocessing can eliminate the influence of these unstable factors and improve the stability and resolution of the fingerprinting.

Different feature extraction methods require different pre-processing steps. The preprocessing in this study included down-conversion, signal detection normalization, time synchronization, and filtering to form a time-aligned baseband signal example. A preprocessing block diagram for this study is shown in Fig. 4.

Because the various frequency impairments between the receiver and transmitter are not zero, the baseband signal after preprocessing inevitably has a carrier frequency offset. In the white Gaussian channel, the preprocessed complex baseband

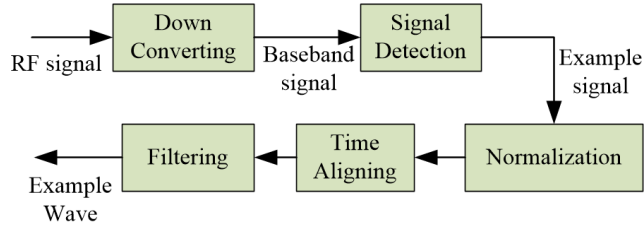


FIGURE 4. Preprocessing block diagram.

signal for a burst signal can be expressed as:

$$z(t) = (\mu_1 \rho(t) + \mu_2 \rho^*(t) + \xi) e^{j(2\pi f_\Delta t + \varphi)} + \eta(t) \quad (3)$$

where $f_\Delta = f_{cR} - f_{cT}$ is the deviation between the receiver frequency f_{cR} and transmitter frequency f_{cT} . Owing to manufacturing imperfections, f_{cR} and f_{cT} deviate with a slight frequency offset. φ is the phase deviation. $\eta(t)$ is zero-mean Gaussian white noise.

From (3), we observed that the preprocessed baseband signal is a non-zero frequency signal with a carrier frequency of f_Δ .

B. FEATURE EXTRACTION

The purpose of the feature extraction is to obtain a map that maps the signal waveform into a vector that can characterize different transmitters. In this section, the preprocessed signal is demodulated, differentiated, and extracted, and then the maximum likelihood estimation algorithm is used to map the sample signal to a modulator distortion vector independent of the carrier frequency offset.

The first is demodulation processing, in which the demodulation constellation can be expressed as:

$$z_n = (\mu_1 c_n + \mu_2 c_n^* + \xi) e^{j(2\pi f'_\Delta n + \phi)} + \eta_n \quad (4)$$

where z_n is the coordinate of the demodulation constellation point, $(\mu_1 c_n + \mu_2 c_n^* + \xi)$ denotes the amplitude of the demodulated constellation point, and c_n can be obtained by demodulation and assuming that the demodulation is error-free. η_n is zero-mean Gaussian white noise. f'_Δ is the frequency offset after carrier synchronization. Where n denotes the sampling time. If a single signal contains a total of N modulation symbols, then the constellation \mathbf{Z} after the demodulation of a single signal example has a total of N sampling points, that is, $\mathbf{Z} = \{z_n\}, n = 0, 1, 2, \dots, N - 1$.

From (4), we observe that the signal example after demodulation contains a phase rotation factor, $e^{j(2\pi f'_\Delta n + \phi)}$. Because the frequency offset f'_Δ is non-zero, the example signal after demodulation will rotate and accumulate with the increase in sampling time n , which leads to the correlation of the characteristic distribution of the modulator distortion fingerprinting extracted by the existing method with the carrier frequency offset. Therefore, the influence of carrier frequency offset should be considered when extracting the distortion features of the modulator.

The demodulated constellation \mathbf{Z} is treated with equal-interval differential processing. Because the Gaussian

white noise is much smaller than the amplitude of the demodulated signal, ignoring the cross term between the signal and noise, the differential demodulation constellation of the example can be expressed as:

$$D_n = z_{n+\Delta n} \cdot z_n^* = \begin{pmatrix} \mu_1 \mu_1^* c_{n+\Delta n} c_n^* + \mu_1 \mu_2^* c_{n+\Delta n} c_n \\ + \mu_1 \xi^* c_{n+\Delta n} + \mu_2 \mu_1^* c_{n+\Delta n} c_n^* \\ + \mu_2 \mu_2^* c_{n+\Delta n} c_n + \mu_2 \xi^* c_{n+\Delta n} \\ + \xi \mu_1^* c_n^* + \xi \mu_2^* c_n + \xi \xi^* \end{pmatrix} e^{j(2\pi f'_\Delta \Delta n)} + v_n \quad (5)$$

Here, $v_n = \eta_n \cdot \eta_n^*$. Because the preprocessed complex baseband signal $z(t)$ obeys a Gaussian distribution, the demodulated signal z_n obeys a Gaussian distribution, and the differential demodulation signal D_n obeys a Gaussian distribution. Assuming that the differential interval is $\Delta n = 1$, there are $N - 1$ samples of the differential demodulation signal of a single example, that is, $\mathbf{D} = \{D_n\}, n = 0, 1, 2, \dots, N - 2$.

To extract the modulator distortion eigenvector, the differential demodulation signal \mathbf{D} is represented as a matrix as follows:

$$\mathbf{D} = \mathbf{G}\boldsymbol{\theta} + \mathbf{v} \quad (6)$$

where,

$\mathbf{D} = [D_0 D_1 D_2 \dots D_{N-2}]^T$ is a $(N - 1) \times 1$ vector, representing a differential demodulation constellation.

$\mathbf{G} = [\mathbf{c}_1 \cdot \mathbf{c}_0^* \mathbf{c}_1 \cdot \mathbf{c}_0 \mathbf{c}_1 \mathbf{c}_1^* \cdot \mathbf{c}_0^* \mathbf{c}_1^* \cdot \mathbf{c}_0 \mathbf{c}_1^* \mathbf{c}_0^* \mathbf{c}_0 \mathbf{1}]$ is a $(N - 1) \times 9$ matrix composed of modulation symbols.

$\mathbf{c}_0 = [c_0 c_1 c_2 \dots c_{N-2}]^T$ is an $(N - 1) \times 1$ vector, and $\mathbf{c}_1 = [c_1 c_2 c_3 \dots c_{N-1}]^T$ is an $(N - 1) \times 1$ vector. $\mathbf{1}$ is a $(N - 1) \times 1$ vector.

$\boldsymbol{\theta} = e^{j(2\pi f'_\Delta)} [\mu_1 \mu_1^* \mu_1 \mu_2^* \mu_1 \xi^* \mu_2 \mu_1^* \mu_2 \mu_2^* \mu_2 \xi^* \xi \mu_1^* \xi \mu_2^* \xi \xi^*]$ is a 9×1 vector.

From (6), it can be seen that $\boldsymbol{\theta}$ can be calculated from the differential demodulation constellation \mathbf{D} , carrier frequency offset f'_Δ , and demodulation symbol \mathbf{c}_0 and \mathbf{c}_1 . Because the differential demodulation signal D_n obeys a Gaussian distribution, the maximum likelihood estimation method was used to estimate the parameter $\boldsymbol{\theta}$ as follow:

$$\hat{\boldsymbol{\theta}} = (\mathbf{G}^H \mathbf{G})^{-1} \mathbf{G}^H (\mathbf{D}) \quad (7)$$

The estimated parameter $\hat{\boldsymbol{\theta}}$ contains the phase rotation factor $e^{j(2\pi f'_\Delta)}$. To avoid the influence of the distortion eigenvector by the phase rotation factor $e^{j(2\pi f'_\Delta)}$, $\hat{\boldsymbol{\theta}}$ was normalized in this study. The eigenvector of the modulator distortion is

$$\mathbf{F}_\theta = \hat{\boldsymbol{\theta}} / \text{sum}(\hat{\boldsymbol{\theta}}) = \frac{[\mu_1 \mu_1^* \mu_1 \mu_2^* \mu_1 \xi^* \mu_2 \mu_1^* \mu_2 \mu_2^* \mu_2 \xi^* \xi \mu_1^* \xi \mu_2^* \xi \xi^*]^T}{\text{sum} \begin{pmatrix} \mu_1 \mu_1^* + \mu_1 \mu_2^* + \mu_1 \xi^* + \mu_2 \mu_1^* + \mu_2 \mu_2^* + \mu_2 \xi^* + \xi \mu_1^* + \xi \mu_2^* + \xi \xi^* \end{pmatrix}} \quad (8)$$

The eigenvector \mathbf{F}_θ is a 9-dimensional column vector, and is only related to the distortion of the modulator; therefore,

this study uses \mathbf{F}_θ as the eigenvector of modulator distortion to identify.

C. CLASSIFIER MODEL

In recent years, the study of the random forest classification method has become a research hotspot in the field of machine learning and has been widely used in many fields, such as finance, ecology, and network security. The random forest classifier is an ensemble-learning method based on a decision tree. Some scholars have optimized and improved the standard random forest classification algorithm to improve the classification. Paul et al. proposed an improved random forest classifier for classification with minimal trees to reduce classification errors [40]. Liu et al. proposed a random forest algorithm integrating decision trees and optimal trees, which introduced the ID3 algorithm to improve the classification accuracy [41]. Yuan et al. proposed an overlap-imbalance-sensitive random forest (OIS-RF) method that optimizes random forest performance [42]. The research focus of this study was not on the classifier model; therefore, this study adopted the standard random forest classification method for recognition. To achieve a better recognition performance, this study built a complete random forest classifier based on 10 decision trees. The input eigenvector of the classifier is a complex vector \mathbf{F}_θ of the size 9×1 :

IV. PERFORMANCE ANALYSIS

To make the experiments easy to understand and repeatable, we conducted experiments using simulated RF signals of multiple emitters and used the recognition method in this study to achieve accurate identification. In addition, the superiority of the proposed method is verified by comparing it with the recognition performance of the modulator distortion feature vector mentioned in [26].

The experimental conditions were set as follows: according to the distortion signal model in Section II, eight emitter signals were generated, each emitter signal was added to the carrier frequency offset in the range of -6 Hz to 6 Hz, and the frequency offset interval was 2 Hz, resulting in a total of seven frequency offsets. Each emitter generated 200 examples under a single carrier frequency offset for a total of 11200 examples. QPSK modulation, number of symbols per example $L=1000$, symbol rate of 100k Baud, carrier frequency of 350kHz, sampling rate of 10 MHz, raised cosine shaping filter with a roll-off factor of 0.35 was used, and the signal-to-noise ratio(SNR) was 20 dB. Thus each example contains 100000 sampling points. The parameters of the distortion model are listed in Table 2.

In this study, the frequency offset refers to the base-band signal frequency after frequency compensation. The frequency offset after carrier synchronization is typically on the order of Hz. According to the signal in this section, in a single example, a frequency offset of 6 Hz can cause a 21.6° phase rotation. It is sufficient to prove the influence of frequency offset on recognition performance, therefore, the carrier frequency offset is set from -6 Hz to 6 Hz.

A. FEATURES DISTRIBUTION WITH DIFFERENT FREQUENCY OFFSETS

To analyze the influence of different frequency offsets on the distortion vector \mathbf{F}_θ of the modulator in this study, we analyzed the vector distribution and second-order statistics of the eigenvector to compare the fingerprint distribution of the emitters under different carrier frequency offset conditions.

First, the distribution of the eigenvectors at different frequency offsets was observed. Because all components of vector are Gaussian distribution.

Consider the second component of \mathbf{F}_θ , denoted as $F_{\theta 2}$, as an example. This section compares the distribution of $F_{\theta 2}$ when the frequency offset is -6 Hz and 0 Hz, as shown in Fig. 5. Each point in Fig. 5 represents a feature extracted from an example, where points of the same shape represent the same emitter. Therefore, each figure in Fig. 5(a) and Fig. 5(b) contains eigenvalues for a total of 1600 examples from eight emitters, each with 200 eigenpoints.

Comparing Fig. 5(a) and Fig. 5(b), it can be seen that, taking the 2nd emitter as an example, regardless of whether the carrier frequency offset is -6 Hz or 0 Hz, the real part of the eigenvalues is distributed between 0.25-0.35, and the imaginary part is distributed between -0.06-0.08. In other words, when the carrier frequency offset changes, the distribution of the IQ distortion features in this study does not change. In addition, it can be seen from Fig. 5 that regardless of whether the frequency offset is -6 Hz or 0 Hz, the IQ distortion of different emitters is distributed in different regions; that is, different emitters can be distinguished by the characteristic vector of each emitter.

Next, the change in the mean and variance of eigenvectors $F_{\theta 2}$ with carrier frequency offset was analyzed. Fig. 6(a) shows the mean of the eigenvector $F_{\theta 2}$ for each emitter at different frequency offsets. Each point in the figure represents the mean of all 200 signal examples for an emitter at a fixed frequency offset, and the points with the same shape represent the mean of the fingerprint characteristics of the same emitter, that is, Fig. 6(a) contains a total of 56 points. Fig. 6(b) shows the variance of each emitter characteristic vs. the carrier frequency offset, where each point represents the variance of all the features of an emitter at different frequency offsets, and the points of the same shape represent the variance of the same emitter.

As shown in Fig. 6(a), the mean values of the different frequency offsets of the same emitter are clustered together. and do not change with the carrier frequency offset. As shown in Fig. 6(b), the variance of the features of the same emitter does not change with the frequency offset; that is, fingerprinting of the same emitter under different carrier frequency offset conditions does not diverge with the frequency offset. In summary, the feature distribution did not shift or spread with a change in the carrier frequency offset. This is because the feature vector proposed in this paper theoretically ensures that it is independent of the carrier frequency offset, which solves the problem that the frequency offset encountered by the traditional method cannot be completely eliminated and

TABLE 2. Parameters of different transmitter.

Sources ID	1	2	3	4	5	6	7	8
ζ	0.0209	-0.0122	-0.0087	-0.0035	0.0006	0.005	0.0139	0.0174
G_I	1.565	1.68	1.64	1.64	1.5	1.475	1.45	1.41
G_Q	1.65	1.565	1.565	1.6	1.49	1.5	1.5	1.5
O_I	-0.02	-0.0128	-0.0083	-0.0038	0.0007	0.0052	0.0097	0.0142
O_Q	-0.01	-0.0123	-0.0078	-0.0033	0.0012	0.0057	0.0102	0.0147

makes the vector better reflect the distortion characteristics of the emitters.

B. IDENTIFICATION PERFORMANCE WITH THE SAME CARRIER FREQUENCY OFFSET

To investigate the identification performance when the carrier frequency offset is constant, this experiment takes the dataset with a frequency offset of -6 Hz as the benchmark dataset and uses the random forest classifier for identification. There were 1600 examples of eight emitters in the benchmark dataset, 70% of which were randomly selected as the training set, and the remaining examples were used as the test set. Because the samples in the test set were randomly selected, the total number of samples for each emitter in the test set was inconsistent, and the numbers of examples in the test set for the 1st to 8th emitters were 58, 56, 69, 62, 56, 65, 53, and 61, respectively. Characterization of emitter identification performance based on the precision and overall accuracy of a single emitter.

The overall accuracy of the classification recognition is the ratio of the number of correctly classified examples to the total number of examples in the dataset.

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad (9)$$

where TP represents true positives, TN represents true negatives, FP represents false positives, and FN represents false negatives. where $TP+TN$ is the number of correctly classified examples. $TP + TN + FP + FN$ represents the total number of examples in a dataset.

Precision is represented as:

$$Precision = TP / (TP + FP) \quad (10)$$

For the identification performance of emitters with a frequency offset of -6Hz, the overall accuracy of the eight emitters was 94.1%, that is, the identification method proposed in this study can accurately identify the emitter under the condition that the carrier frequency offset is unchanged. As shown in Fig. 7, the correct identification rate of 1st emitter was 93%, the probability of error identification as 8th emitter was 3%, and the probability of error identification as 7th emitter was 4%, which corresponds to the feature distribution in Fig. 5(b). In Fig. 5(b), the feature distributions of 1st emitter and 8th emitter partially overlap, which easily causes the classifier to identify errors. Comparing Fig. 7 and Fig. 5(b),

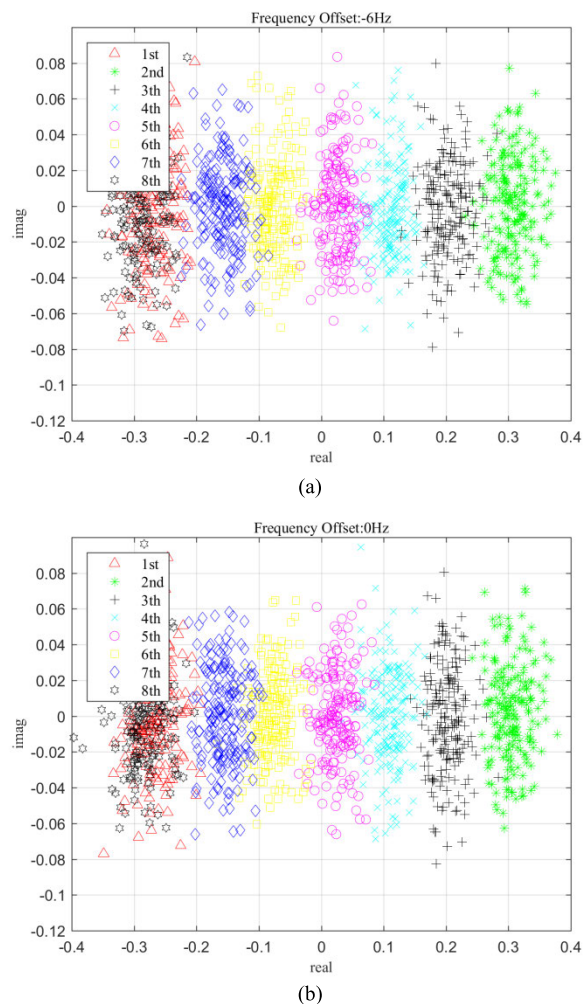


FIGURE 5. Distribution of F_{θ_2} with different frequency offsets: (a) -6Hz frequency offset and (b) 0Hz frequency offset.

it can be seen that when the frequency offset is constant and the characteristic distribution of the emitters overlaps, the performance of the classifier decreases.

C. INFLUENCE OF FREQUENCY OFFSET ON IDENTIFICATION PERFORMANCE

To investigate the recognition performance of the identification method proposed in this paper when the carrier frequency

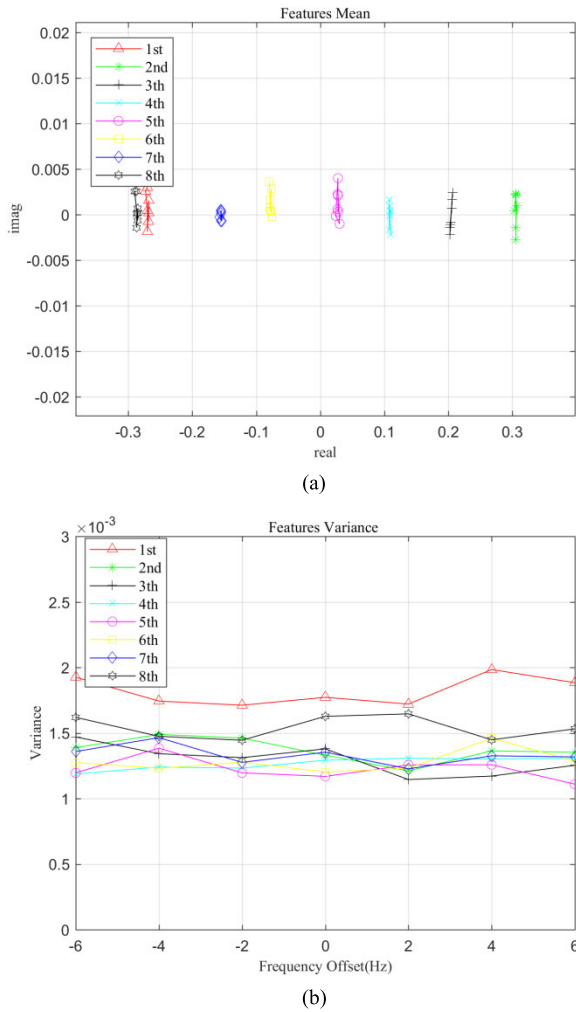


FIGURE 6. Statistics for different frequency offsets: (a) mean, and (b) variance.

offset changes. In this experiment, Pan [26] and the direct classification of raw data were compared with the method used in this study. In [26], the traditional frequency offset compensation method was used to extract the vector of modulator distortion for classification. The direct classification method for raw data involves directly identifying the demodulated constellation without feature extraction. The method proposed in this study extracts a modulator distortion vector that is independent of the carrier frequency offset for classification. The classifiers of the three recognition methods are random forest classifiers, and all classifiers use 70% of all examples with a carrier frequency offset of -6 Hz as the training set to complete the training. The remaining 30% of examples with a carrier frequency offset of -6 Hz and all examples with a carrier frequency offset of -4 Hz, -2 Hz, 0 Hz, 2 Hz, 4 Hz, and 6 Hz were used as the test set to complete the recognition accuracy test. In this experiment, the overall accuracy was used to characterize the identification performance of all emitters under frequency-offset

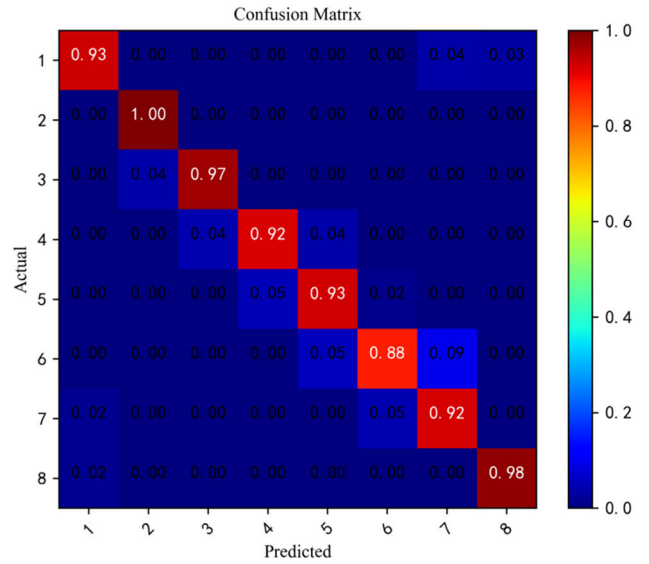


FIGURE 7. Confusion matrix for identification of 8 emitters without frequency offset.

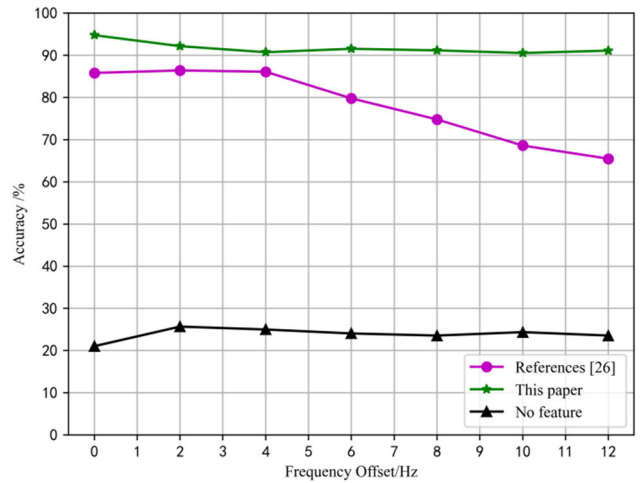


FIGURE 8. Overall accuracy with different carrier frequency offsets.

conditions. Through the recognition accuracy under different frequency offset conditions, the emitter recognition performance of different carriers was verified.

Fig. 8 shows the influence of different carrier frequency offsets on the identification performance. The horizontal axis represents the frequency offset between the example in the test set and that in the training set. The vertical axis represents the overall identification accuracies of the eight emitters.

It can be seen from Fig. 8 that the recognition accuracy of the direct classification method of raw data is stable between 20%-30%, although the raw data contain all the fingerprint information, and its recognition accuracy does not change with the change in frequency offset. However, the overall recognition accuracy of the method is poor, mainly because the original data after demodulation is directly used without feature extraction for identification, and the original data

cannot explicitly present all fingerprint information; therefore, the recognition accuracy rate is low overall. The method in [26] adopts the modulator distortion vector after frequency compensation, and although frequency compensation is carried out in the feature extraction process, the influence of frequency offset cannot be completely eliminated. Therefore, when the frequency offset between the recognition and training examples exists, the recognition accuracy rate decreases, and the larger the offset, the lower the recognition accuracy rate, which affects the actual performance of the recognition model. For the method proposed in this paper, when the difference between the carrier frequency of the example to be identified and the carrier frequency of the training set is within 12 Hz, the recognition accuracy is stable at approximately 90%, and does not change with the increase in the frequency offset; therefore, the recognition method in this study does not need to consider the actual change in the carrier frequency offset.

V. CONCLUSION

This paper proposes a specific emitter identification method based on differential constellation and applies differential processing technology to separate and extract modulator distortion features. We design a new modulator distortion fingerprint vector that is independent of the carrier frequency offset. In contrast to the existing modulator distortion fingerprint extraction method which requires high-precision carrier synchronization, this technique does not need to accurately compensate for the frequency offset of the constellation. As long as the demodulation symbol and demodulation signal can be obtained, the modulator distortion feature can be extracted, and the training of the random forest classifier can be completed to achieve accurate identification of multiple emitters. Experimental results show that the proposed technique can significantly improve the SEI performance compared to existing algorithms.

REFERENCES

- [1] J. Wright, "Detecting wireless LAN MAC address spoofing," *Comput. Sci.*, White Paper, Jan. 2003.
- [2] F. Guo and T. Chiueh, "Sequence number-based MAC address spoof detection," in *Proc. RAID*, Seattle, WA, USA, 2006, pp. 309–329.
- [3] O. Ureten and N. Serinken, "Wireless security through RF fingerprinting," *Can. J. Electr. Comput. Eng.*, vol. 32, no. 1, pp. 27–33, 2007.
- [4] C. K. Dbendorfer, B. W. Ramsey, and M. A. Temple, "An RF-DNA verification process for ZigBee networks," in *Proc. MILCOM*, Orlando, FL, USA, 2012, pp. 1–6.
- [5] H. Patel, "Non-parametric feature generation for RF-fingerprinting on ZigBee devices," in *Proc. IEEE CISDA*, Verona, NY, USA, May 2015, pp. 1–5.
- [6] L. Peng, J. Zhang, M. Liu, and A. Hu, "Deep learning based RF fingerprint identification using differential constellation trace figure," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 1091–1095, Jan. 2020.
- [7] L. Peng, A. Hu, J. Zhang, Y. Jiang, J. Yu, and Y. Yan, "Design of a hybrid RF fingerprint extraction and device classification scheme," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 349–360, Feb. 2019.
- [8] J. Han, T. Zhang, D. Ren, and X. Zheng, "Communication emitter identification based on distribution of bispectrum amplitude and phase," *IET Sci., Meas. Technol.*, vol. 11, no. 8, pp. 1104–1112, Nov. 2017.
- [9] Y. Yuan, Z. Huang, H. Wu, and X. Wang, "Specific emitter identification based on Hilbert–Huang transform-based time-frequency-energy distribution features," *IET Commun.*, vol. 8, no. 13, pp. 2404–2412, Sep. 2014.
- [10] J. Hall, "Detection of rogue devices in wireless networks," Ph.D. dissertation, School Comput. Sci., Dept. Ottawa-Carleton Inst. Comput. Sci., Carleton Univ., Ottawa, ON, Canada, Tech. Rep., 2006.
- [11] J. Franklin, D. McCoy, P. Tabriz, V. Neagoe, J. V. Randwyk, and D. Sicker, "Passive data link layer 802.11 wireless device driver fingerprinting," in *Proc. USENIX Secur. Symp.*, Vancouver, BC, Canada, 2006, pp. 167–178.
- [12] J. Toonstra and W. Kinsner, "Transient analysis and genetic algorithms for classification," in *Proc. Commun., Power, Comput., Conf.*, 1995, pp. 432–437.
- [13] B. Danev and S. Capkun, "Transient-based identification of wireless sensor nodes," in *Proc. Int. Conf. Inf. Process. Sensor Netw.*, San Francisco, CA, USA, 2009, pp. 25–36.
- [14] H. P. Romero, K. A. Remley, D. F. Williams, and C.-M. Wang, "Electromagnetic measurements for counterfeit detection of radio frequency identification cards," *IEEE Trans. Microw. Theory Techn.*, vol. 57, no. 5, pp. 1383–1387, May 2009.
- [15] D. A. Knox and T. Kunz, "Practical RF fingerprints for wireless sensor network authentication," in *Proc. 8th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Aug. 2012, pp. 531–536.
- [16] M. Barbeau, J. Hall, and E. Kranakis, "Detecting impersonation attacks in future wireless and mobile networks," in *Proc. Int. Workshop Secure Mobile Ad-Hoc Netw. Sensors*, Singapore, 2005, pp. 80–95.
- [17] J. Hall, M. Barbeau, and E. Kranakis, "Enhancing intrusion detection in wireless networks using radio frequency fingerprinting," in *Proc. CIIT*, 2004, pp. 201–206.
- [18] K. A. Remley, C. A. Grosvenor, R. T. Johnk, D. R. Novotny, P. D. Hale, M. D. McKinley, A. Karygiannis, and E. Antonakakis, "Electromagnetic signatures of WLAN cards and network security," in *Proc. 5th IEEE Int. Symp. Signal Process. Inf. Technol.*, Dec. 2005, pp. 484–488.
- [19] T. J. Bihl, K. W. Bauer, and M. A. Temple, "Feature selection for RF fingerprinting with multiple discriminant analysis and using ZigBee device emissions," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 8, pp. 1862–1874, Aug. 2016.
- [20] C. Dubendorfer, B. Ramsey, and M. Temple, "ZigBee device verification for securing industrial control and building automation systems," in *Proc. Int. Conf. Critical Infrastructure*, Washington, DC, USA, 2013, pp. 47–61.
- [21] J. Zhang, F. Wang, O. A. Dobre, and Z. Zhong, "Specific emitter identification via Hilbert–Huang transform in single-hop and relaying scenarios," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 6, pp. 1192–1205, Jun. 2016.
- [22] C. Bertoincini, K. Rudd, B. Nousain, and M. Hinders, "Wavelet fingerprinting of radio-frequency identification (RFID) tags," *IEEE Trans. Ind. Electron.*, vol. 59, no. 12, pp. 4843–4850, Dec. 2012.
- [23] U. Satija, N. Trivedi, G. Biswal, and B. Ramkumar, "Specific emitter identification based on variational mode decomposition and spectral features in single hop and relaying scenarios," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 3, pp. 581–591, Mar. 2019.
- [24] V. Brik, S. Banerjee, M. Gruteser, and S. Oh, "Wireless device identification with radiometric signatures," in *Proc. 14th ACM Int. Conf. Mobile Comput. Netw.*, San Francisco, CA, USA, Sep. 2008, pp. 116–127.
- [25] Y. Huang and H. Zheng, "Radio frequency fingerprinting based on the constellation errors," in *Proc. 18th Asia-Pacific Conf. Commun. (APCC)*, Oct. 2012, pp. 900–905.
- [26] Y. Pan, "Research on key technologies of communication emitter identification," Ph.D. dissertation, Dept. Inf. Syst. Eng., Inf. Eng. Univ., Zhengzhou, Henan, China, 2019.
- [27] M.-W. Liu and J. F. Doherty, "Specific emitter identification using nonlinear device estimation," in *Proc. IEEE Sarnoff Symp.*, Princeton, NJ, USA, Apr. 2008, pp. 1–5.
- [28] A. C. Polak and D. L. Goeckel, "Wireless device identification based on RF oscillator imperfections," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 12, pp. 2492–2501, Dec. 2015.
- [29] X. Zhang and T. Li, "Specific emitter identification based on feature diagram superposition," in *Proc. 7th Int. Conf. Integr. Circuits Microsystems (ICICM)*, Xi'an, China, Oct. 2022, pp. 703–707.
- [30] T. J. O'Shea, T. Roy, and T. C. Clancy, "Over-the-air deep learning based radio signal classification," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 168–179, Feb. 2018.

- [31] M. Kulin, T. Kazaz, I. Moerman, and E. De Poorter, "End-to-end learning from spectrum data: A deep learning approach for wireless signal identification in spectrum monitoring applications," *IEEE Access*, vol. 6, pp. 18484–18501, 2018.
- [32] C. Wang, J. Wang, and X. Zhang, "Automatic radar waveform recognition based on time-frequency analysis and convolutional neural network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, New Orleans, LA, USA, Mar. 2017, pp. 2437–2441.
- [33] Z. Zhou, G. Huang, H. Chen, and J. Gao, "Automatic radar waveform recognition based on deep convolutional denoising auto-encoders," *Circuits, Syst., Signal Process.*, vol. 37, no. 9, pp. 4034–4048, Jan. 2018.
- [34] K. Merchant, S. Revay, G. Stantchev, and B. Noursain, "Deep learning for RF device fingerprinting in cognitive communication networks," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 1, pp. 160–167, Feb. 2018.
- [35] F. Restuccia, S. D'Oro, A. Al-Shawabka, M. Belgiovine, L. Angioloni, S. Ioannidis, K. Chowdhury, and T. Melodia, "DeepRadioID: Real-time channel-resilient optimization of deep learning-based radio fingerprinting algorithms," in *Proc. 20th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, Jul. 2019, pp. 51–60.
- [36] K. Sankhe, M. Belgiovine, F. Zhou, S. Riyaz, S. Ioannidis, and K. Chowdhury, "ORACLE: Optimized radio classification through convolutional neural networks," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, Paris, France, Apr. 2019, pp. 370–378.
- [37] S. Riyaz, K. Sankhe, S. Ioannidis, and K. Chowdhury, "Deep learning convolutional neural networks for radio identification," *IEEE Commun. Mag.*, vol. 56, no. 9, pp. 146–152, Sep. 2018.
- [38] Y. Pan, S. Yang, H. Peng, T. Li, and W. Wang, "Specific emitter identification based on deep residual networks," *IEEE Access*, vol. 7, pp. 54425–54434, 2019.
- [39] J. Liu, J. Li, J. Wang, and H. Huang, "Specific emitter identification at different time based on multi-domain migration," in *Proc. Asia-Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Nov. 2022, pp. 917–922.
- [40] A. Paul, D. P. Mukherjee, P. Das, A. Gangopadhyay, A. R. Chintla, and S. Kundu, "Improved random forest for classification," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4012–4024, Aug. 2018.
- [41] Y. Liu, L. Liu, Y. Gao, and L. Yang, "An improved random forest algorithm based on attribute compatibility," in *Proc. IEEE 3rd Inf. Technol., Netw., Electron. Autom. Control Conf. (ITNEC)*, Chengdu, China, Mar. 2019, pp. 2558–2561.
- [42] B.-W. Yuan, Z.-L. Zhang, X.-G. Luo, Y. Yu, X.-H. Zou, and X.-D. Zou, "OIS-RF: A novel overlap and imbalance sensitive random forest," *Eng. Appl. Artif. Intell.*, vol. 104, Sep. 2021, Art. no. 104355.



research interests include signal processing, artificial intelligence, and specific emitter identification.



YANBIN LI was born in Shijiazhuang, Hebei, China, in 1966. He received the B.S. degree from Tianjin University, Tianjin, China, in 1985, the M.S. degree from the 54th Research Institute of China Electronic Technology Group Corporation (CETC54), Shijiazhuang, in 1988, and the Ph.D. degree from Shanghai Jiaotong University, Shanghai, China, in 1995. He is currently a Chief Expert with China Electronic Technology Group Corporation and a Chief Scientist of CETC54. His current research interests include electronic countermeasures (ECM) and cognitive electronic warfare.

• • •