Wang *et al. EURASIP Journal on Wireless Communications and Networking* (2019) 2019:230
https://doi.org/10.1186/s13638-019-1544-8

EURASIP Journal on Wireless Communications and Networking

RESEARCH Open Access

Identification and authentication for wireless transmission security based on RF-DNA fingerprint

Xueli Wang[1*], Yufeng Zhang[1], Hongxin Zhang[2], Xiaofeng Wei[2] and Guangyuan Wang[2]

Abstract

For wireless transmission, radio-frequency device anti-cloning has become a major security issue. Radio-frequency distinct native attribute (RF-DNA) fingerprint is a developing technology to find the difference among RF devices and identify them. Comparing with previous research, (1) this paper proposed that mean (μ) feature should be added into RF-DNA fingerprint. Thus, totally four statistics (mean, standard deviation, skewness, and kurtosis) were calculated on instantaneous amplitude, phase, and frequency generated by Hilbert transform. (2) We first proposed using the logistic regression (LR) and support vector machine (SVM) to recognize such extracted fingerprint at different signal-to-noise ratio (SNR) environment. We compared their performance with traditional multiple discriminant analysis (MDA). (3) In addition, this paper also proposed to extract three sub-features (amplitude, phase, and frequency) separately to recognize extracted fingerprint under MDA. In order to make our results more universal, additive white Gaussian noise was adopted to simulate the real environment. The results show that (1) mean feature conducts an improvement in the classification accuracy, especially in low SNR environment. (2) MDA and SVM could successfully identify these RF devices, and the classification accuracy could reach 94%. Although the classification accuracy of LR is 89.2%, it could get the probability of each class. After adding a different noise, the recognition accuracy is more than 80% when $SNR \geq 5$ dB using MDA or SVM. (3) Frequency feature has more discriminant information. Phase and amplitude play an auxiliary but also pivotal role in classification recognition.

Keywords: RF-DNA, Fingerprint recognition, Logistic regression, Support vector machine, Signal-to-noise ratio, Radio-frequency authentication, Anti-cloning

1 Introduction

In recent years, with the development of mobile communication system equipment and Internet of Things, wireless transmission technology has played more and more important roles in our daily life [1]. Compared with wired network, wireless network is more convenient and concise. Furthermore, in terms of cost, the wireless network greatly eliminates the wiring and decoration costs. However, intrusive attack on electronic devices is growing rapidly. Wireless signals have been often used as a cornerstone of massively malicious attacks, and the broadcast characteristic of the wireless transmission makes the problem worse. It is essential to guarantee the safety of information transmission, urging us to pay more and more attention to security problem and new countermeasures. Physical layer security is the most basic part of wireless transmission security. Many attackers invade security system by copying the device and mocking the signal. For example, the thief gets into the car by imitating the signals like car keys; the intruder enters confidential system by mocking license signal emitted from cloning devices. Countering RF device cloning is an issue that we urgently need to solve. Fortunately, due to slight differences in production, even the "same" devices will have some discrepancy, which is hard for us to observe it directly. But that still gives us opportunity to identify different RF devices and find cloned equipment for malicious attacks. At present, RF-DNA fingerprint technology is a rising technology which is adopted to counter related risk such

*Correspondence: wangxl@bupt.edu.cn
[1] School of Science, Beijing University of Posts and Telecommunications, Xitucheng Road 10, 100876 Beijing, China
Full list of author information is available at the end of the article

as device cloning. Traditional DNA refers to the biological internal attributes and different individuals have different DNA. Similarly, we think each RF device has its own intrinsic physical attributes called RF-DNA fingerprint. In this paper, by calculating the statistics features of many signals emitted from one device, we could get "RF-DNA" of each device. In other words, RF-DNA fingerprint is discriminating features extracted from different RF devices [2], and any two RF devices must have differences. The differences are due to equipment noise and hardware production error [3], reflected in their output signals.

The main mission of RF-DNA fingerprint technology is to distinguish the signal and counter cloning, which could be summarized as identification and authentication. There are some RF devices and an unknown signal. We need to identify which device that the unknown signal came from and this is called identification. As for authentication, an unknown signal claims that it came from one RF device, and we need to find out how credible it is. It is used to prevent the two different devices from using the same RF-DNA.

Similar to biology DNA recognition, RF-DNA fingerprint technology could identify machines, which will have great application in lots of field such as information safety, criminal investigation, and even military command. Once this technology become mature, malicious cloning devices, in all likelihood, will be caught, and our wireless information transmission will become more secure. So far, some related research and progress have been done in this field.

Over past two decades, there are many development opportunities [4] and physical layer challenges [5] in RF-DNA fingerprint issues. At present, it is a mainstream method to classify the amplitude, frequency, and phase features using MDA method [2, 6–9]. Firstly, research [2, 6] enabled both identification and verification device issues and extended the process from the three-class to general N-class problems. By setting a priori distribution of multivariate Gaussian distribution, posteriori probability could be calculated [2, 10] to achieve authentication simulation. Previous studies showed the impact on the number of dimensions [7, 11] and the number of subregion [8] on classification accuracy. The more feature dimensions will get better classification results. Additional, a signal could be divided into midamble region and near-transient region. Choosing near-transient region or inter-manufacturer of mobile communication system could lead to higher classification accuracy [1, 9]. Besides MDA, decision tree algorithm [12] or other classifiers are also good methods to distinguish the signal. Classifier selection is a crucial part of RF-DNA technology [13, 14]. SVM classifier [15] was applied on kernel-independent component nonlinear feature extraction. Research [16, 17] proposed to use probabilistic neural network based on Bayesian classification as classifier. Generalized relevance learning vector quantization-improved [12, 18, 19] is a supervised machine learning algorithm based on MDA, which shows better performance. As for the different feature extraction methods, fast Fourier transform [20, 21], the short-time Fourier transform [22], and discrete Gabor transform [10] could also be used to extract features. Previous research had compared their performance, including on time domain, wavelet domain [23], and spectral domain [24] features. Some novel approaches such as least square estimation [25] and the phase characteristics [26] were proposed to extract transient fingerprint. Also, research [20] separated the features apart and finds more important features. Furthermore, some physiological electrocardiogram signals [15, 17] were classified by emerging artificial neural network [20, 27] directly, especially by recurrent neural network [28] which is worth drawing lessons from.

2 Experimental process

In general, RF-DNA fingerprint technology is divided into the following four steps:

a) Signal collection. Command the RF devices send out a series of unintentional signals and the receivers could collect them. Repeat the above process many times to collect lots of signals. These signals should be considered as security signals and will be used as training set in our classifiers.

b) Feature extraction. Each signal has some own statistical features including on time domain, frequency domain, and some other features. On the one hand, the purpose of extracting features is to reduce dimension. On the other hand, the features might have a more accurate description of this signal. This step is the core step of RF-DNA fingerprint technology. A good feature selection often means good classification accuracy.

c) Set up database. After feature extraction, the feature sequence of each signal is put into the database and should be labeled where it comes from. These features are called RF-DNA fingerprint.

d) Classification. The main mission is to distinguish the label of an unknown signal sample. We could judge by comparing the features of unknown signals with the known samples features in the database.

2.1 Signal collection

There are four RF devices embedded with NRF24LE1 chip shown as Fig. 1 (Fig. 1 contains Fig. 1a and Fig. 1b. Fig. 1b is an enlarged view of Fig. 1a). The only difference among these RF devices is the date of manufacture which are 11th week of 2011, 31st week of 2011, 24th week of 2014, and 48th week of 2015, respectively. The research signals in this paper were collected in May 2018.

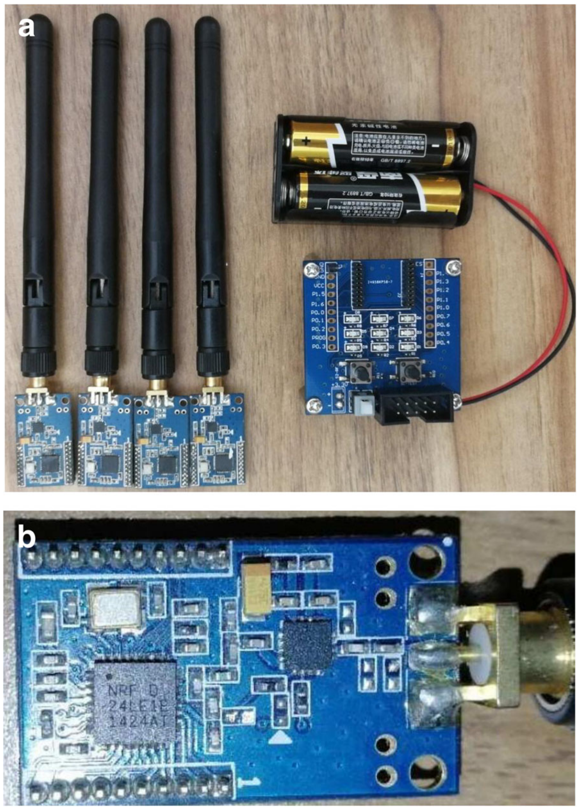

Fig. 1 Four RF device chips and one enlarged view. **a** Our four sampling equipments and their base. We used them to collect original signals. **b** An enlarged view of one RF device produced in 24th week of 2014

The experimental signal acquisition process was shown as Fig. 2. Our personal computer (PC) controlled the RF device to emit unintentional signals, and the detector received the signals displayed in the oscilloscope. The detector recorded the original amplitude signal from these four RF devices once the waveform is stable. Then, the signals were pre-treated by Microsoft Decoder Sample. All signals were obtained under 2.4G bandwidth, and the sampling frequency is $fs = 25.6$MHz. For each RF device, the valid signal lasts about 9 s and has about 230,000,000 sampling points in total.

2.2 Additive white Gaussian noise (AWGN)

The signals were collected in the closed basement, which could be considered as a relatively low noise environment. The distance between RF device and detector is only 5 m. Besides, the outside noise influences were limited as much as possible. However, such experimental scene selection may not have versatility and might not be suitable for practice use. Due to the limited experimental conditions, our experimental scene is unique. The laboratory environment data could not give a convincing result of the performance. Therefore, AWGN was taken in order to evaluate performance under some less ideal conditions and make our results more universal. The SNR was calculated as formula (1).

$$SNR = 10 \times \log_{10}\left(\frac{\text{Signal power}}{\text{Noise power}}\right) \qquad (1)$$

The noise power could be controlled by AWGN while the signal power could be calculated from original amplitude signal. Through analysis and calculation, the SNR of original sampling signal is 30 dB. That means the signal power is 10^3 times than noise power, which could be considered that there is hardly noise in sampling. And after different AWGN, we could get the SNR={0, 1, 3, 5, 7, 10, 15, 20, 25, 30} dB environment, respectively.

2.3 Sample generation

Too short sample leads too poor classification accuracy and too long sample is lack of persuasion, thus, taking $L = 2^{18} = 262,144$ sampling plots as one sample is a plausible choice. Considering that the sampling frequency is $fs = 25.6$ MHz, each sample lasts about 0.01 s, which is in a relatively high precision level.

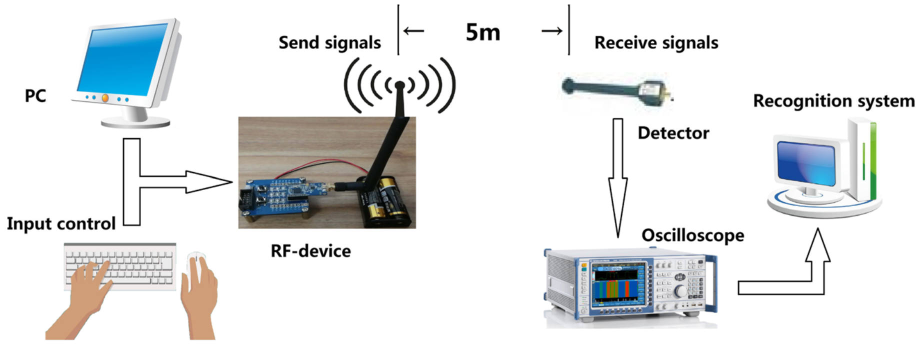

Fig. 2 Experimental signal acquisition process. The PC controlled the RF device to emit unintentional signals and the detector received the signals shown in oscilloscope, and the distance between RF device and detector is 5 meters

For each RF device, we divided the original signal into $N = 2000$ samples. According to the order of the production data, the label catalogs of four RF devices are M1, M2, M3, and M4, respectively. Signal samples from the same device will be marked the same label. In order to ensure the adequacy of training, we randomly take $T = 1600$ samples as training samples from each RF device. The rest 400 samples are set as testing samples to assess performance.

During the operation of signal collection, there are too much bias that we could not fully observed. The system should work directly on the original data with minimal pre-processing. In practice, in order to reduce potential signal collection bias and eliminate the dimension of the data, the original signal sampling sequence $x(n)$ should be normalized. We take linearly normalization method to handle every single original sample before the feature extraction as formula (2).

$$a(\mathrm{n}) = [x(n) - \min(x(n))] / [\max(x(n)) - \min(x(n))] \tag{2}$$

where $x(n)$ is original amplitude sampling signal and $a(n)$ is the normalized signal.

Then, the amplitude of each sample is normalized to the range of [0, 1] as Fig. 3. These four diagrams show the 500 signal points of four machines, respectively. The unstable signal in front was then abandoned. With the naked eye, the signals from the four RF devices are very similar. It is almost impossible to see from the figure that amplitude profiles are visually distinctive. Therefore, we need extract RF-DNA features to identify the differences.

3 Statistic fingerprint generation

3.1 Divide sample into sub-regions

Figure 4 elaborates the whole fingerprint generation process. The first two boxes have been introduced in Section 2. That is , there were $k = 4$ RF devices, and we collected $N = 2000$ signal samples for each RF device. However, a relatively ideal condition to extract features is on a steady signal. Hence, we decided to divide one signal sample into N_R equal length sub-regions and thus each region could be considered more stable in comparison. Additionally, the benefit of doing this is that you can increase the dimension of fingerprint features. Bihl et al. [7] showed that the increase of feature dimension may increase the accuracy of classification. Figure 5 demonstrates the sub-region allocation process. Then, we got N_R sub-regions and one complete sample region, totally $(N_R + 1)$ regions. We extracted the features separately in these $(N_R + 1)$ regions. Cobb et al. [8] analyzed the performance of parameters N_R value, and we take $N_R = 16$ which is a reasonable trade-off.

3.2 Feature extraction using Hilbert transform

The most straightforward method to extract features is using the original amplitude signal as our features. However, our results show that such classification accuracy is less than 40% using only amplitude feature directly without any transforms. Hence, we need to find more feature dimension information. Using Hilbert transform, instantaneous amplitude (IA) noted by $a(n)$, instantaneous phase (IP) noted by $\varphi(n)$, and instantaneous frequency (IF) noted by $f(n)$, totally $N_F = 3$ features could be extracted from the given real-valued time domain signal.

Firstly, the IA signal was converted into I-Q sample $S_C(n) = H(a(n)) = s_I(n) + s_Q(n)$. Next, the IP $\varphi(n)$ and the IF $f(n)$ were calculated as formula (3).

$$\varphi(n) = \tan^{-1}\left[\frac{s_Q(n)}{s_I(n)}\right], f(n) = \frac{1}{2\pi}\left[\frac{d\phi(n)}{dt}\right] \tag{3}$$

3.3 Calculate statistical fingerprint

Compared with the previous method, we propose mean (μ) feature could be added to our statistical RF-DNA fingerprint. Taking IA feature $a(n)$ as an example, mean, standard deviation (σ), skewness(γ), and kurtosis(κ) were calculated as formula (4)–(7). That is, N_S=4 statistical fingerprint were calculated in each (N_R+1) regions and each N_F feature sequence. The hot picture of normalized statistical fingerprint was shown as Fig. 6, which calculated average from 2000 signals for each RF device. It can be intuitively seen from the diagram that M4 has more difference from other three RF devices. That will cause M4 to be more easily identified which is consistent with to our results.

$$\text{Mean} : \mu = \frac{1}{L}\sum_{\mathrm{n}=1}^{L} a(n) \tag{4}$$

$$\text{Variance} : \sigma^2 = \frac{1}{L}\sum_{\mathrm{n}=1}^{L} (a(n) - \mu)^2 \tag{5}$$

$$\text{Skewness} : \gamma = \frac{1}{L\sigma^3}\sum_{\mathrm{n}=1}^{L} (a(n) - \mu)^3 \tag{6}$$

$$\text{Kurtosis} : \kappa = \frac{1}{L\sigma^4}\sum_{\mathrm{n}=1}^{L} (a(n) - \mu)^4 \tag{7}$$

where $a(n)$ denotes the normalized sample signal sequence and L denotes the number of sampling points and standard deviation σ is $\sqrt{\sigma^2}$.

Overall, for one sample i ($i = 1, 2, \ldots, N; N = 2000$), the way we generate fingerprint can be summarized as the following three steps. (a) Divide the original signal sample into N_R=16 equal length sub-region and one total region. Then, we got the vector as formula (8). (b) Calculate $N_F = 3$ features signal within $(N_R + 1) = 17$ regions as formula

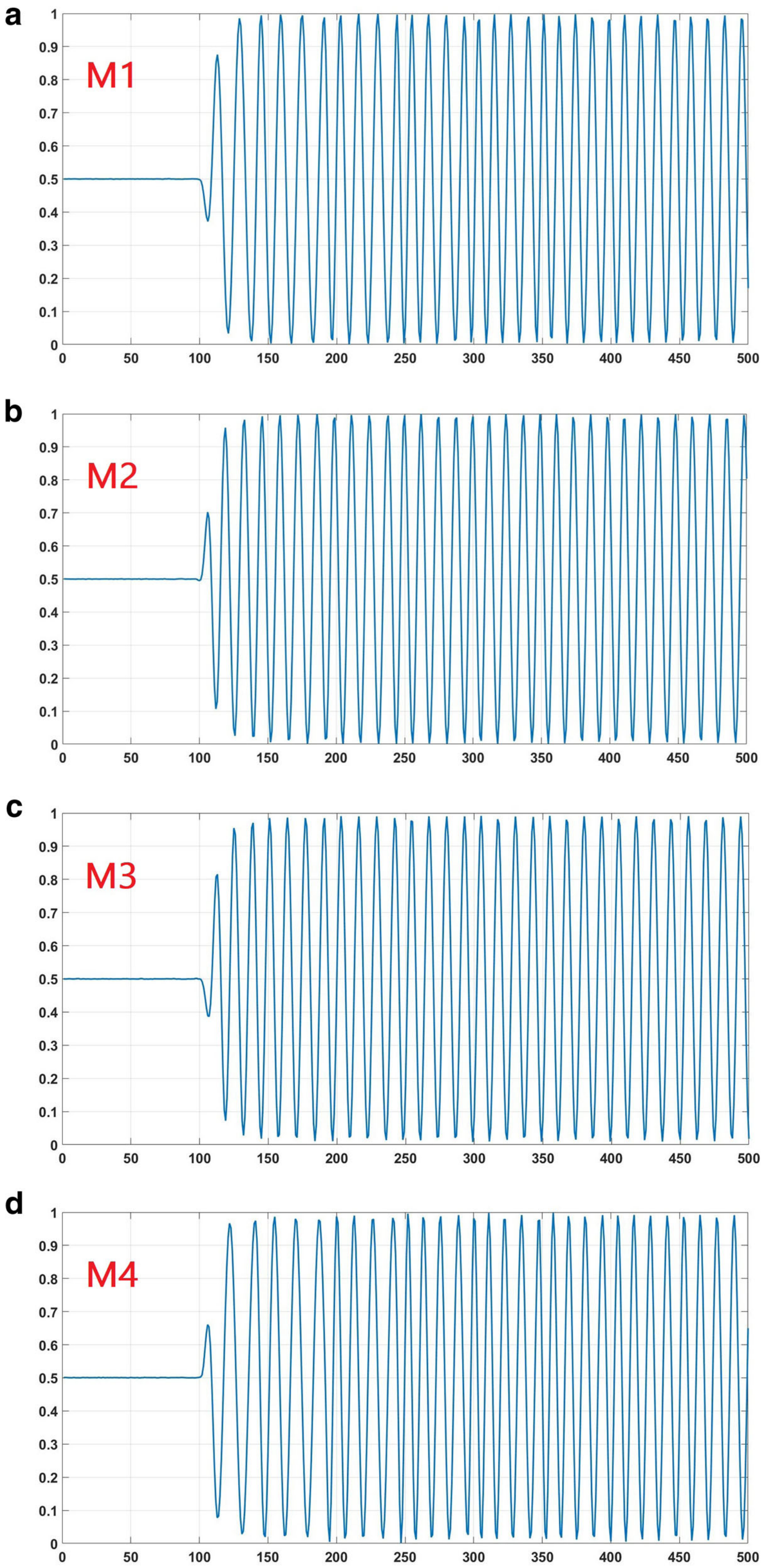

Fig. 3 Five hundred normalized signal points of four RF devices. We show 500 original samples of four RF devices. Those signals were normalized to the range [0, 1]. The unstable signal in front was then abandoned. It is almost impossible to see from the figure that amplitude profiles are visually distinctive. Therefore, we need extract RF-DNA features to identify the differences

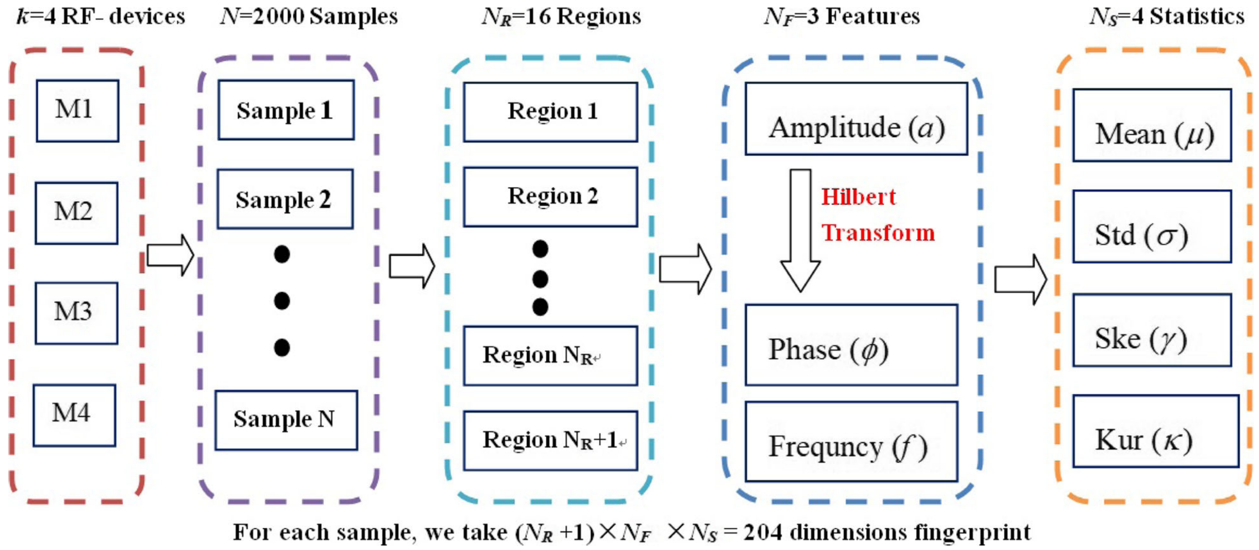

Fig. 4 The whole process of fingerprint generation. The whole RF-DNA fingerprint generation process including five steps: (1) four RF devices, (2) collect $N = 2000$ samples, (3) divide one sample into N_R=16 equal length sub-region, (4) calculate $N_F = 3$ features, and (5) extract $N_S = 4$ statistics features

(9). (c) Extract $N_S = 4$ statistics and generate 1×204 dimension single sample fingerprint as formula (10).

$$F_{R^i} = \left[F_{R_1^i} \vdots F_{R_2^i} \vdots F_{R_3^i} \vdots \ldots \vdots F_{R^i_{(N_R+1)}} \right]_{1\times(N_R+1)} \tag{8}$$

$$F_i^x = \left[(F_{R^i})^{a} \vdots (F_{R^i})^{\varphi} \vdots (F_{R^i})^{f} \right]_{1\times(N_R+1)\cdot N_F} \tag{9}$$

$$F_i = \left[\mu(F_i^x) \vdots \sigma(F_i^x) \vdots \gamma(F_i^x) \vdots \kappa(F_i^x) \right]_{1\times(N_R+1)\cdot N_F\cdot N_S=1\times 204} \tag{10}$$

Finally, for each RF device, the training matrix composed of $T = 1600$ separately training fingerprint sets is

$$Tr = [F_1, F_2, ..., F_T,]'_{T\times 204} \tag{11}$$

4 Classification methods

Previous research [1, 2, 6–9] mostly used MDA as classifier for fingerprint recognition. MDA is an extension to Fisher's linear discriminant in multivariate statistical analysis when there are more than two RF devices needed to be classified. It effectively reduces the input data dimensionality by projecting it into a lower-dimensional space. We

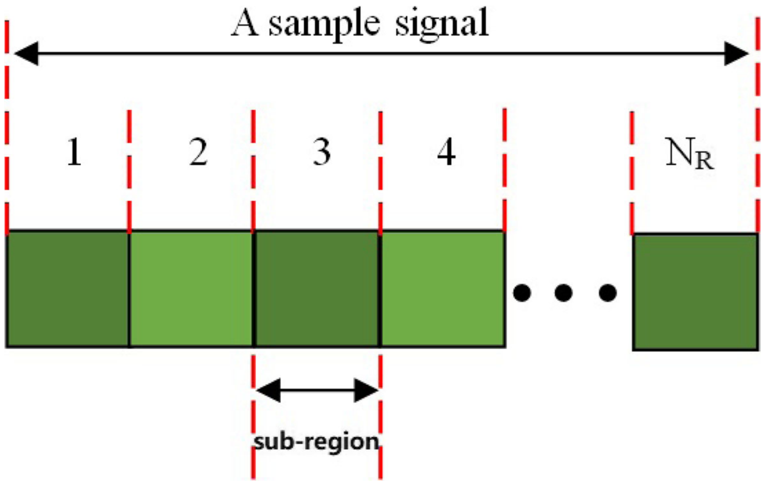

Fig. 5 One sample signal was divided into N_R equal length sub-regions. A schematic that we divided the signal equally. We divided one signal sample into N_R equal length sub-regions and thus each region could be considered more stable in comparison. Additional, the benefit of doing this is that you can increase the dimension of fingerprint features

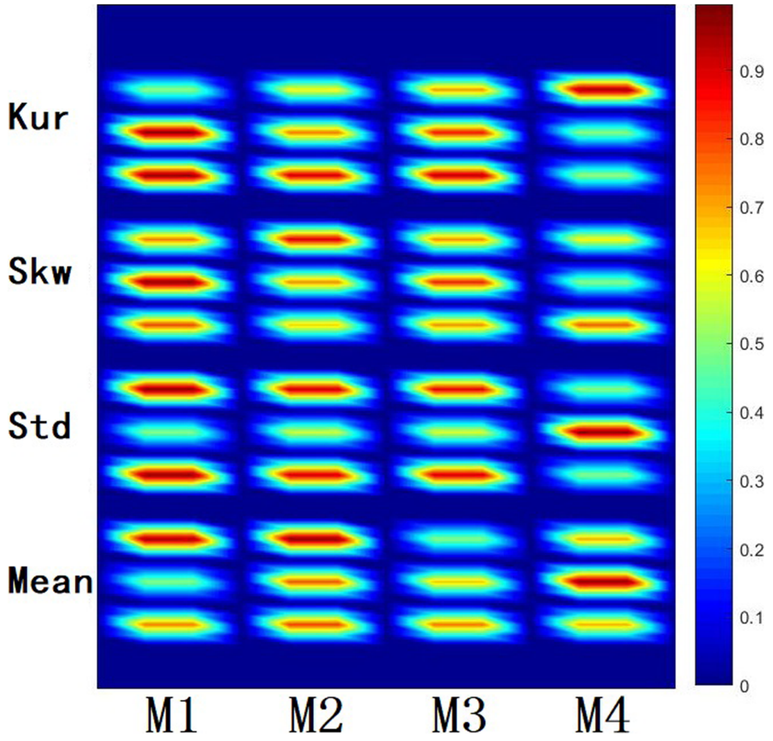

Fig. 6 Average 2000 fingerprint for four RF devices at SNR = 30 dB. The heat map of normalized statistical fingerprint for four RF devices at SNR = 30 dB. That is our RF-DNA fingerprint

need to find projection vector $b = (b_1, b_2, ..., b_{204})'$. After projecting, the aim is to maximize λ which is the ratio of between-group to within-group sum of squares defined as formula (12).

$$\lambda = \frac{b' S_b b}{b' S_w b} \tag{12}$$

where S_b is the between-group scatter matrix and S_w is the within-group scatter matrix. We could calculate that the projection vector b is eigenvector of $S_w^{-1} S_b$, and λ is the associated eigenvalue reflecting group separation.

Based on the extracted 204-dimension fingerprint, we apply two statistic methods which are SVM and LR as classifiers to deal with Hilbert transform features for the first time. The limitation of previous research is that the classifiers can only identify the unknown sample belongs to which device. LR can give the probability of belonging to each class, which could be used to achieve RF authentication mission.

4.1 Support vector machine

Traditional SVM can only solve the two classifications problem. The training fingerprint samples have been extracted as formula (11). The training samples and their labels set is $S = \left\{ (F_1, y_1), (F_2, y_2), ..., (F_T, y_T) \right\}, F_\mathrm{i} \in R^{204}, y_i \in \{+1, -1\}$, where $T = 1600$ is the training samples size for each RF device and y_i is class category. Through maximizing the interval or the equivalent method as formula (13), we can find separating hyperplane $\omega' F + b = 0$.

$$\begin{aligned} &\min_{w,b,\xi} \frac{1}{2} (\omega)'(\omega) + C \sum_{i=1}^{T} \xi_i \\ s.t. \quad &y_i(\omega' \cdot F_i + b) \geq 1 - \xi_i \\ &\xi_i \geq 0 \qquad i = 1, 2, \ldots, T \end{aligned} \tag{13}$$

where C is penalty coefficient, and we set $C = 100$ using tenfold cross-validation. ω and b are parameters of separating hyperplane. ξ is the distance between fingerprint sample F_i and the separating hyperplane. Finally, we would take the unknown sample fingerprint into this separating hyperplane. Through positive or negative of the obtained value, we could classify this unknown sample.

However, SVM is designed to deal with binary classification problems. In this experiment, we used one-against-one method which could extend SVM to k classes. Design sub-classifiers between any two classes, and thus, we could get $k(k-1)/2 = 6$ sub-classifiers ($k = 4$). For example, the SVM sub-classifier of class c_α and class c_β is established. If the unknown sample is classified into class c_α, then class c_α scores one point; otherwise, class c_β scores one point. After six times classification, the unknown signal sample finally belongs to the class which gets the highest score.

4.2 Logistic regression

The traditional logistic regression is also used to solve the problem of two classifications. Similarly, we extend logistic regression to k classes. Since there are four RF devices, we assume that $P(y = c_\alpha|F)$ $(\alpha = 0, 1, 2, 3)$ represent the probability of belonging to class c_α. We set $y = 0$ as the reference group and covariant variable is F $=[F^{(1)}, F^{(2)}, \ldots, F^{(204)}]$. Set up disordered logistic regression models.

$$\begin{aligned} g_\alpha(F) &= \ln\left[\frac{P(y = c_\alpha|F)}{P(y = c_0|F)}\right] \\ &= w_{\alpha,0} + w_{\alpha,1}F^{(1)} + \ldots + w_{\alpha,204}F^{(204)} \end{aligned} \tag{14}$$

where $\alpha = 0, 1, 2, 3$ and obviously $g_0(F) = 0$. Equally, the conditional probability of label y is:

$$P(y = c_\alpha|F) = \frac{e^{g_\alpha(F)}}{1 + \sum_{j=1}^{3} e^{g_j(F)}} \tag{15}$$

In identification mission, we could infer that the unknown fingerprint sample belongs to the largest probability class c_α, that is

$$P(y = c_\alpha|F) > P(y = c_\beta|F) \qquad \forall\beta \neq \alpha; \quad \alpha, \beta = 0, 1, 2, 3 \tag{16}$$

In authentication mission, a signal will claim that it is emitted from a security RF device. We could authenticate this signal by setting probability verification threshold P_0.

$$P(y = c_\alpha|F) \geq P_0 \qquad \alpha = 0, 1, 2, 3 \tag{17}$$

where c_α is the RF device class which the unknown sample claims to belong. The decision for this authentication mission is a binary result. If the probability $P(y = c_\alpha|F)$ meets the threshold P_0 as formula (17), we will accept this fingerprint and deem it as security signal. Otherwise, we will refuse and take it as a security signal and deem it as an imposter. For example, if the probability of an unknown sample belonging to a security device is the largest, but the probability is less than the threshold, then we still do not regard it as a security signal.

There are two values to measure the selection of the threshold which are true positive (TP) and true negative (TN). TP denotes the probability that a security signal comes and you accept it. TN denotes the probability that an imposter signal comes and you refuse it. The larger the two values, the better the authentication effect.

5 Results and discussion

Table 1 shows classification confusion matrix in three different classifiers. At SNR=30 dB environment, the test accuracy of MDA and SVM algorithm is beyond 94% on average. Therefore, we could believe that the method of feature extraction in the time domain is effective. When SNR=0 dB, the noise power is equal to the signal power, which could be considered in a very high noise environment. In such simulated environment, the classification accuracy will be significantly reduced and any two of the four RF devices may be confused. The environment noise does have a great influence on discrimination.

Table 1 Classification accuracy in different classifiers at $SNR = 30, 0$ dB

Actual class	M1 (%)	M2 (%)	M3 (%)	M4 (%)	Accuracy %)
	Confusion matrix (SNR=30)				
Classifer: MDA					
M1	90.75	0.00	9.25	0.00	
M2	0.00	99.50	0.50	0.00	94.50
M3	12.25	0.00	87.75	0.00	
M4	0.00	0.00	0.00	100.00	
Classifer: SVM					
M1	89.50	0.00	10.50	0.00	
M2	0.00	99.50	0.50	0.00	94.44
M3	10.75	0.25	89.00	0.00	
M4	0.25	0.00	0.00	99.75	
Classifer: LR					
M1	82.25	0.00	17.00	0.75	
M2	0.00	99.25	0.75	0.00	89.19
M3	23.50	0.75	75.75	0.00	
M4	0.50	0.00	0.00	99.50	
	Confusion matrix (SNR=0)				
Classifer: MDA					
M1	42.00	11.75	31.50	14.75	
M2	11.00	70.75	16.50	1.75	54.63
M3	31.25	20.5	37.50	10.75	
M4	18.50	0.75	12.50	68.25	
Classifer: SVM					
M1	33.25	12.00	35.00	19.75	
M2	10.25	67.25	19.00	3.50	48.75
M3	26.75	19.75	37.75	15.75	
M4	20.00	2.75	20.50	56.75	
Classifer: LR					
M1	19.00	19.25	23.50	38.25	
M2	11.75	60.00	17.25	11.00	40.25
M3	20.50	25.50	24.00	30.00	
M4	12.00	11.75	18.25	58.00	

Figure 7 created by Matlab R2016a shows the tend of classification accuracy at different SNR. Obviously, as the

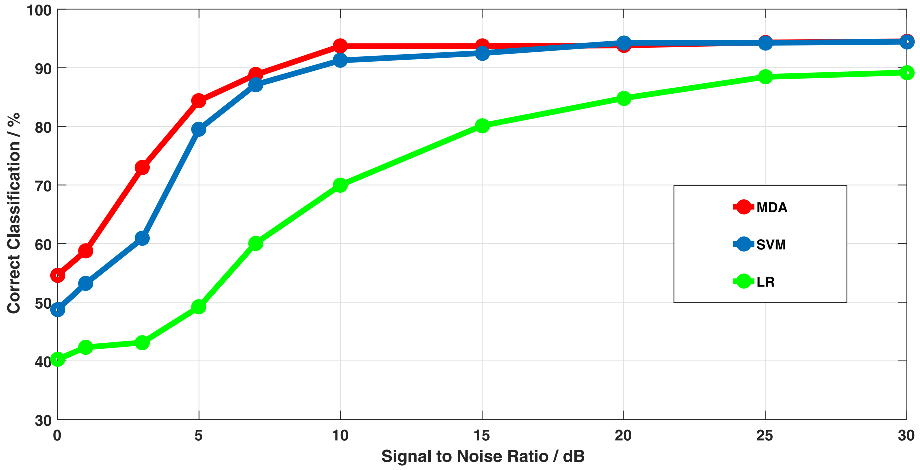

Fig. 7 Classification accuracy under different classifiers. With the change of SNR, the tend of classification accuracy under MDA, SVM, and LR

SNR increases, the classification accuracy is increasing. When SNR is below 5 dB, the classification accuracy is less than 80% and begins to decline significantly. Both MDA and SVM show a better classification performance than LR.

Then, we listed the classification accuracy of four RF devices separately under MDA classifier shown as Fig. 8. We can find that M2 and M4 maintain a relatively high accuracy. That is because the differences between RF devices are uncertain and we cannot observe it directly. We could only observe the difference indirectly that M2 and M4 have a more significant fingerprint features; hence, they could be easier classified. Similarly, the difference between M1 and M3 are small; thus, they could be easier confused and have relatively lower classification accuracy.

Due to the characteristics of LR, we achieve authentication simulation shown as Table 2. In our experiment, three RF devices were designed as cloning devices to send malicious attack signal, and there is one security RF device. We set different threshold P_0 from 0.2 to 0.8. Then, TP and TN were calculated in different SNR environment. Take 13.0 and 99.9 in the upper left corner as an example. When we set $P_0 = 0.8$ and SNR=30 dB, due to the higher threshold, only 13.0% security signal could be accepted, but 99.9% imposter signal will be refused. As the decrease of threshold, more security signals are accepted and less impostor signals are refused. Besides, as the increase of SNR, both TP and TN are increasing. The external noise showed a great effect on RF authentication. For the three possible cloning RF devices and one security RF device in our experiment, the best probability threshold could be

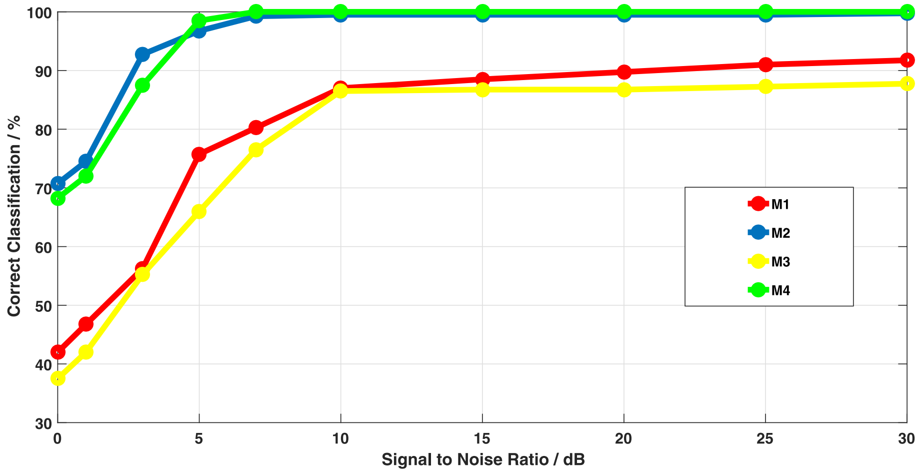

Fig. 8 Classification accuracy of each RF device individually under MDA. We observe the classification accuracy of each machine and find out the similarity between four RF devices

Table 2 Authentication mission threshold decision at different SNR

P_0	0.8		0.7		0.6		0.5		0.4		0.3		0.2	
SNR	TP	TN	TP	TN	TP	TN	TP	TN	TP	TN	TP	TN	TP	TN
30	13.0	99.9	34.8	99.9	61.5	98.9	78.5	96.1	92.5	88.8	96.8	80.6	99.3	72.7
20	1.75	99.8	14.8	99.4	32.3	98.8	55.5	95.8	77.8	87.8	91.3	79.5	97.8	72.1
15	0.3	99.8	4.5	99.2	14.5	98.8	34.5	95.4	61.8	87.2	84.8	78.8	96.5	70.2
10	0.0	99.7	2.5	98.9	10.0	98.7	29.5	94.2	47.3	87.0	72.3	77.3	92.0	68.3
5	0.0	99.5	1.8	98.9	7.0	97.9	15.5	93.3	36.3	86.9	56.0	76.8	79.3	60.5
0	0.0	99.5	0.0	98.8	0.0	97.9	1.0	93.0	10.3	86.0	25.3	69.2	62.8	45.6

set from 0.4 to 0.5, where the sum of TP and TN is relative high.

In addition, we extra extracted mean features as RF-DNA fingerprint. The performance was shown as Fig. 9 that +Mean refers to the fingerprint with mean feature and −Mean refers to the fingerprint without mean feature. In the high SNR environment, due to the precision of the classification is already high, mean feature can only play a small role. But in the low SNR environment, mean feature conducts a significant improvement in the classification accuracy. Therefore, it is meaningful for RF-DNA to extract the mean feature.

Furthermore, previous studies only focused on the integrity of features including frequency, phase, and amplitude. They did not study which features had more discriminant information. We propose to only extract the frequency or phase or amplitude sub-feature alone. The number of feature dimension changed from 204 to 204/3=68. Figure 10 shows the classification accuracy to single features under MDA classifier. The red line "All" means that all three features are adopted which are 204 dimensions.

We found that the frequency sub-feature has the highest classification accuracy. In other words, the frequency feature information has the largest effect on classification recognition in our experiment. Amplitude and phase feature information play an auxiliary role. However, the classification accuracy between red line "All" and blue line "Frequency" is still very different. When SNR$\leq$ 5 dB, frequency feature also loses its identification ability. That is, amplitude and phase features information indeed have a great contribution on classification. They are crucial to enhance classification accuracy. In summary, the success classification of RF-DNA fingerprint was due to the joint action of all three features.

6 Conclusions

Recently, using cloning equipment to obtain illegal access authentication seriously affects the security of information transmission. RF-DNA fingerprint is a rising concept to mark every RF device, thus could be used to identify malicious attack cloning RF devices. In our experiment, 2.4 G bandwidth signal from four RF devices were collected. Results show that the optimal classification

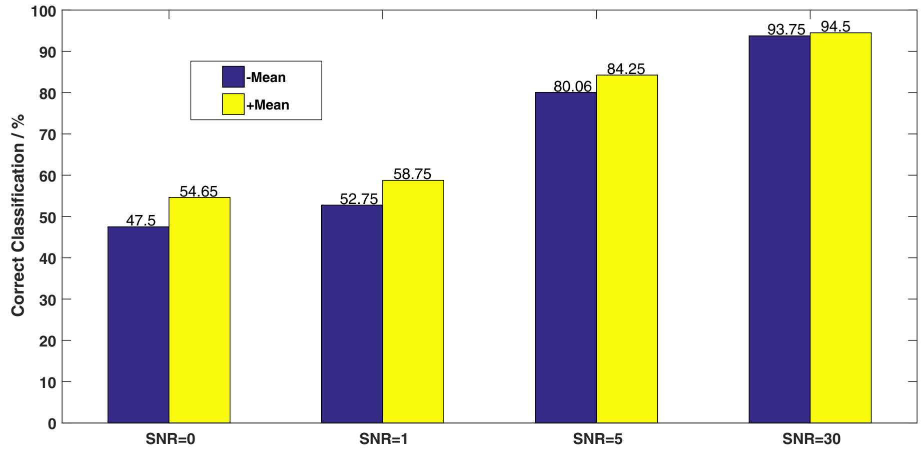

Fig. 9 Effect of mean feature under MDA. We extra extracted mean features as RF-DNA fingerprint. The recognition accuracy was improved. +Mean, with mean feature; −Mean, without mean feature

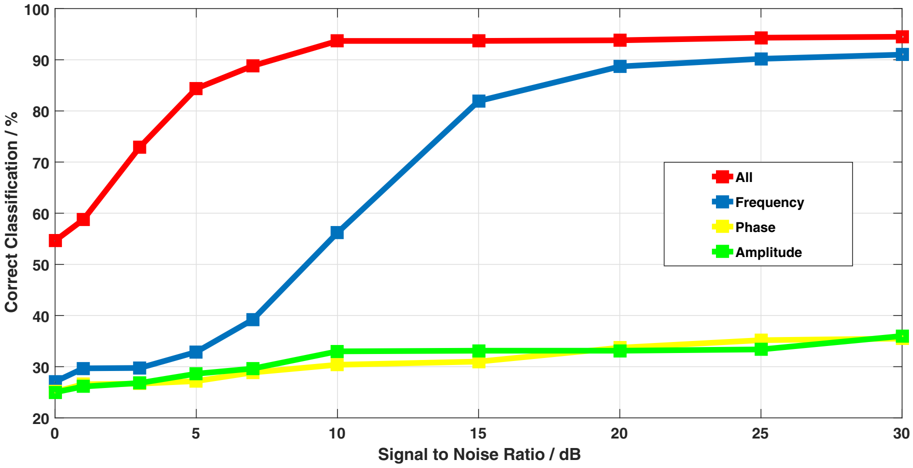

Fig. 10 Classification accuracy for single sub-features under MDA. We propose to only extract the frequency or phase or amplitude sub-feature alone to find which one has the highest classification accuracy. The red line "All" means that all three features are adopted which are 204 dimensions

accuracy could reach 94%. The reason why we achieved satisfactory results is RF-DNA fingerprint of each RF device is unique, just like DNA in living beings, and the difference among similar RF devices could be discovered. This paper demonstrates that using our extracted fingerprint to distinguish RF devices is successful. Meanwhile, we analyzed the performance under some unsatisfactory conditions. With the decrease of SNR, the classification accuracy is also decreasing. That makes our experimental results more universal and persuasive in real application. Although the accuracy of LR is not as good as SVM or MDA, it could achieve authentication mission and find reasonable threshold setting. Besides, adding mean and separating the sub-features are also innovations of this paper, which will be some special applications in practice.

In this paper, we used NRF24LE1 chip as RF devices in closed basement scenario. Other scenarios could be implemented through different simulators. And our experimental results can be extended to many scenarios, such mobile phone signal [9], remote sensing signal, and military radar signal. Even some human signals such as electrocardiogram, electroencephalogram, or electromyogram could use RF-DNA fingerprint technology to identify human health.

The limitation of this paper is that the extracted features are relatively less, only 204 dimensions. Increasing feature extraction may improve identification accuracy significantly. Additionally, only four RF devices are classified in our experiment. We should find some advanced methods if we need to deal with a large number of cloning RF devices. And the future work could also focus on the method of extracting fingerprint and the classifier chosen. For example, recent researches took short-time Fourier transforms [22] and discrete Gabor transform [10] to generate RF-DNA fingerprint. Besides, some neural network model [17, 20] could also be used as classifiers. Combining appropriate statistical algorithms, finding meaningful RF-DNA fingerprint features can improve the recognition accuracy prominently. Furthermore, our research extract only one kind of fingerprint, and the combination of multiple fingerprint could be a rising area of future research.

Abbreviations

AWGN: Additive white Gaussian noise; IA: Instantaneous amplitude; IP: Instantaneous phase; IF: Instantaneous frequency; LR: Logistic regression; MDA: Multiple discriminant analysis; N_S: Number of statistics; N_F: Number of features; N_R: Number of regions; PC: Personal computer; RF-DNA: Radio-frequency distinct native attribute; SNR: Signal-to-noise ratio; SVM: Support vector machine; TN: True negative; TP: True positive

Authors' contributions

XW conceived the idea. XW and YZ designed the algorithm experiments. XW, YZ, HZ ,and XW performed the model and experiments. XW, YZ, HZ, and GW developed the post-processing treatments of the experimental data. XW, YZ, and HZ carried out the numerical calculations and figures. XW and YZ wrote the paper. All authors contributed to scientific discussion and critical revision of the article. All authors read and approved the final manuscript.

Funding

This research is supported by National Natural Science Foundation of China (No.11471053), Laboratory of Universal Wireless Communications (KFKT-2015103), and Beijing Natural Science Foundation (No.3182028).

Availability of data and materials

We declared that materials described in the manuscript, including all relevant raw data, will be freely available to any scientist wishing to use them for non-commercial purposes, without breaching participant confidentiality. The datasets used or analyzed during the current study are available from the corresponding author on reasonable request. The original sampling data could be downloaded through the following hyperlink: https://pan.baidu.com/s/1Bl7GYQEB1R_azrT0WyReWg

Competing interests

The authors declare that they have no competing interests.

Author details

[1]School of Science, Beijing University of Posts and Telecommunications, Xitucheng Road 10, 100876 Beijing, China. [2]School of Electronic Engineering,

Beijing University of Posts and Telecommunications, Xitucheng Road 10, 100876 Beijing, China.

Received: 6 September 2018 Accepted: 27 August 2019
Published online: 18 September 2019

References

1. D. R. Reising, M. A. Temple, M. J. Mendenhall, in *Wireless Communications and NETWORKING Conference*. Improving intra-cellular security using air monitoring with RF fingerprints (IEEE, 2010), pp. 1–6. https://doi.org/10.1109/wcnc.2010.5506229
2. W. E. Cobb, E. D. Laspe, R. O. Baldwin, M. A. Temple, C. K. Yong, Intrinsic physical-layer authentication of integrated circuits. IEEE Trans. Inf. Forensics Secur. **7**(1), 14–24 (2012)
3. J. Hall, M. Barbeau, E. Kranakis, in *IASTED International Multi-Conference on Wireless and Optical Communications*. Detection of transient in radio frequency fingerprinting using signal phase, (2003)
4. Q. Xu, R. Zheng, W. Saad, Z. Han, Device fingerprinting in wireless networks: challenges and opportunities. IEEE Commun. Surv. Tutor. **18**(1), 94–104 (2016)
5. S. U. Rehman, K. W. Sowerby, S. Alam, I. Ardekani, in *Communications and Network Security*. Radio frequency fingerprinting and its challenges (IEEE, 2014), pp. 496–497. https://doi.org/10.1109/cns.2014.6997522
6. C. K. Dubendorfer, B. W. Ramsey, M. A. Temple, in *Military Communications Conference, 2012 - Milcom*. An RF-DNA verification process for ZigBee networks, (2013), pp. 1–6
7. T. J. Bihl, K. W. Bauer, M. A. Temple, Feature selection for RF fingerprinting with multiple discriminant analysis and using ZigBee device emissions. IEEE Trans. Inf. Forensic Secur. **11**(8), 1862–1874 (2017)
8. W. E. Cobb, E. W. Garcia, M. A. Temple, R. O. Baldwin, in *Military Communications Conference, 2010 - Milcom*. Physical layer identification of embedded devices using RF-DNA fingerprinting (IEEE, 2010), pp. 2168–2173. https://doi.org/10.1109/milcom.2010.5680487
9. M. D. Williams, M. A. Temple, D. R. Reising, in *Global Telecommunications Conference*. Augmenting bit-level network security using physical layer RF-DNA fingerprinting (IEEE, 2010), pp. 1–6. https://doi.org/10.1109/glocom.2010.5683789
10. D. R. Reising, M. A. Temple, in *IEEE International Conference on Communications*. Wimax mobile subscriber verification using Gabor-based RF-DNA fingerprints (IEEE, 2012), pp. 1005–1010. https://doi.org/10.1109/icc.2012.6364039
11. D. R. Reising, M. A. Temple, J. A. Jackson, Authorized and rogue device discrimination using dimensionally reduced RF-DNA fingerprints. IEEE Trans. Inf. Forensic Secur. **10**(6), 1180–1192 (2015)
12. H. J. Patel, M. A. Temple, R. O. Baldwin, Improving ZigBee device network authentication using ensemble decision tree classifiers with radio frequency distinct native attribute fingerprinting. IEEE Trans. Reliab. **64**(1), 221–233 (2015)
13. S. Manel, J. M. Dias, S. J. Ormerod, Comparing discriminant analysis, neural networks and logistic regression for predicting species distributions: a case study with a himalayan river bird. Ecol. Model. **120**(2), 337–347 (1999)
14. P. K. Harmer, D. R. Reising, M. A. Temple, in *IEEE International Conference on Communications*. Classifier selection for physical layer security augmentation in cognitive radio networks (IEEE, 2013), pp. 2846–2851. https://doi.org/10.1109/icc.2013.6654972
15. H. Li, H. Liang, L. Cao, L. Cao, X. Feng, C. Tang, E. Li, Novel ecg signal classification based on kica nonlinear feature extraction. Circ. Syst. Signal Process. **35**(4), 1187–1197 (2016)
16. O. Ureten, N. Serinken, Wireless security through rf fingerprinting. Can. J. Electr. Comput. Eng. **32**(1), 27–33 (2007)
17. B. Abedi, A. Abbasi, A. Goshvarpour, Investigating the effect of traditional persian music on ecg signals in young women using wavelet transform and neural networks. Anatolia J. Cardiol. **17**(5), 398–403 (2017)
18. M. Lukacs, P. Collins, M. Temple, Classification performance using 'RF-DNA' fingerprinting of ultra-wideband noise waveforms. Electron. Lett. **51**(10), 787–789 (2015)
19. B. Hammer, T. Villmann, Generalized relevance learning vector quantization. Neural Netw. **15**(8), 1059–1068 (2002)
20. N. Hu, Y. D. Yao, in *IEEE International Conference on Communications*. Identification of legacy radios in a cognitive radio network using a radio frequency fingerprinting based method (IEEE, 2012), pp. 1597–1602. https://doi.org/10.1109/icc.2012.6364436
21. P. Scanlon, I. O. Kennedy, Y. Liu, Feature extraction approaches to RF fingerprinting for device identification in femtocells. Bell Labs Tech. J. **15**(3), 141–151 (2010)
22. S. Chen, F. Xie, Y. Chen, H. Song, H. Wen, in *IEEE International Symposium on Electromagnetic Compatibility*. Identification of wireless transceiver devices using radio frequency (RF) fingerprinting based on STFT analysis to enhance authentication security (IEEE, 2017), pp. 1–5. https://doi.org/10.1109/emc-b.2017.8260381
23. R. W. Klein, M. A. Temple, M. J. Mendenhall, Application of wavelet-based RF fingerprinting to enhance wireless network security. J. Commun. Netw. **11**(6), 544–555 (2012)
24. M. K. D. Williams, S. A. Munns, M. A. Temple, M. J. Mendenhall, in *International Conference on Network and System Security*. RF-DNA fingerprinting for airport WiMax communications security (IEEE, 2010), pp. 32–39. https://doi.org/10.1109/nss.2010.21
25. C. Zhao, L. Huang, L. Hu, Y. Yao, in *International Conference on Computer Science & Education*. Transient fingerprint feature extraction for WLAN cards based on polynomial fitting (IEEE, 2011), pp. 1099–1102. https://doi.org/10.1109/iccse.2011.6028826
26. J. Hall, M. Barbeau, E. Kranakis, in *IASTED International Multi-Conference on Wireless and Optical Communications*. Detection of transient in radio frequency fingerprinting using signal phase, (2003)
27. T. Debnath, M. M. Hasan, T. Biswas, in *International Conference on Electrical and Computer Engineering*. Analysis of ECG signal and classification of heart abnormalities using artificial neural network (IEEE, 2017), pp. 353–356. https://doi.org/10.1109/icece.2016.7853929
28. M. Cheng, W. J. Sori, F. Jiang, A. Khan, S. Liu, in *IEEE International Conference on Computational Science and Engineering*. Recurrent neural network based classification of ECG signal features for obstruction of sleep apnea detection (IEEE, 2017), pp. 199–202. https://doi.org/10.1109/cse-euc.2017.220

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.