


# The individual identification method of wireless device based on dimensionality reduction and machine learning

Yun Lin<sup>1</sup> · Xiaolei Zhu<sup>1</sup> · Zhigao Zheng<sup>2</sup>  · Zheng Dou<sup>1</sup> · Ruolin Zhou<sup>3</sup>

Published online: 15 December 2017

© Springer Science+Business Media, LLC, part of Springer Nature 2017

**Abstract** The access security of wireless devices is a serious challenge in present wireless network security. Radio frequency (RF) fingerprint recognition technology as an important non-password authentication technology attracts more and more attention, because of its full use of radio frequency characteristics that cannot be imitated to achieve certification. In this paper, a RF fingerprint identification method based on dimensional reduction and machine learning is proposed as a component of intrusion detection for resolving authentication security issues. We compare three kinds of dimensional reduction methods, which are the traditional PCA, RPCA and KPCA. And we take random forests, support vector machine, artificial neural network and grey correlation analysis into consideration to make decisions on the dimensional reduction data. Finally, we obtain the recognition system with the best performance.

---

✉ Zhigao Zheng  
zhengzhigao@hust.edu.cn

Yun Lin  
linyund\_phd@hrbeu.edu.cn

Xiaolei Zhu  
zhuxiaolei@hrbeu.edu.cn

Zheng Dou  
douzheng@hrbeu.edu.cn

Ruolin Zhou  
ruolin.zhou@wne.edu

- <sup>1</sup> College of Information and Communication Engineering, Harbin Engineering University, Harbin, China
- <sup>2</sup> School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China
- <sup>3</sup> Department of Electrical and Computer Engineering, Western New England University, Springfield, MA, USA

Using a combination of RPCA and random forests, we achieve 90% classification accuracy is achieved at  $\text{SNR} \geq 10\text{dB}$  when reduced dimension is 76. The proposed method improves wireless device authentication and improves security protection due to the introduction of RF fingerprinting.

**Keywords** Authentication security · RF fingerprint · Dimensional reduction · Machine learning

## 1 Introduction

Nowadays, due to the rapidly development of wireless communication, wireless network plays important roles in different fields, such as the communication of device to device [11], and the privacy and security issues. However, compared to the traditional wired network, the open wireless networks provide greater opportunity for intruders which can attack wireless networks through a few leaked messages. Traditional security methods for wireless networks are usually implemented in the higher level of the OSI model [6, 9, 38] through the data link layer, network layer or bit-level mechanism. But this mechanism often has a lot of defects, mainly in bit-level mechanism. It is simple to copy device identification information; in this case, unauthorized network can easily obtain authorized access [34, 50]. Therefore, in recent years, the physical layer-based security mechanisms have been widely studied to improve the security of wireless networks [17, 43].

MAC address spoofing is an example of intrusion. Rogue devices will attack when authenticated users initiate access to wireless devices [12]. The intuition behind this is that the authenticated users will be transmit 802.11 management frame through a wireless access, the access point will extract MAC address of authentication device from the 802.11 management frame and make comparison with the existing authorized MAC address list to determine whether it can access. In the process of signal transmission, some rogue devices may use some tools to decipher transmission encryption and obtain the MAC address of the authenticated device. Once the rogue device gets the authenticated MAC address, it can attack an access point of wireless receive device or steal information by faking identity [12, 46].

In the past 10 years, the fingerprint extraction and recognition method of RF wireless communication equipment has been attracting extensive attention at home and abroad. It enables to prevent access from cloning rogue devices and enhance network access security [8]. This method extracts the radio frequency fingerprint of the equipment by analysing the communication signals of wireless devices. As the fact that each person has different fingerprints, each wireless device also has a different RF fingerprint which means hardware differences, such as oscillator and delay circuits. And these differences in the hardware will be reflected on signal, manifesting into the differences of a transceiver's phase or frequency [29]. Although the differences of hardware are minute, it has been proved that these differences can be used for device authentication. The method of extracting device hardware characteristics based on communication signals is called "radio frequency fingerprint extraction", and radio frequency fingerprint, which is called "radio frequency fingerprint identification", is

used to identify different wireless devices. The method, RF fingerprint extraction and recognition of wireless communication equipment, is implemented in the physical layer, so it can operate independently, assist and enhance the recognition mechanism of the traditional wireless network; in this case, it provides wireless networks with higher safety performance [18].

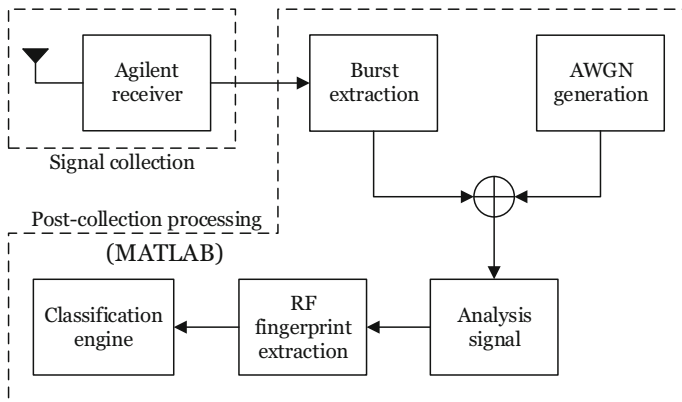
The method to identify equipment based on communication signal was proposed by Choe et al. in the early 1995, based on which Hall et al. [13] proposed the concept of RF fingerprint in 2003 in the literature. Hall et al. identified Bluetooth communication devices through Bluetooth RF fingerprint extraction of communication signals. Toonstra et al. explicitly proposed that using unique “fingerprint” produced by wireless transmitter transient signals to identify device. In the next 10 years, people focus on the research on the transient signal. Until 2008, Kennedy et al. proposed radio frequency fingerprinting based on steady-state signals for the first time. In recent years, most researches focused on the extraction and recognition of radio frequency fingerprints based on steady-state signals, and some achievements have also been reached [18,28]; for example researches used RF fingerprint on semiconductor manufacturing [35], Cloned HF RFID Proximity Cards [51], Bluetooth signals [45] and so on.

Some methods of RF fingerprint recognition are accomplished by instantaneous amplitude [44], while some other works use instantaneous amplitude, frequency and phase as the original feature [2,7,31,32]. In the continued feature extraction, some transform domains are used, such as wavelet transform and Hilbert transform, while some dimensionality reduction methods are used, such as principal component analysis, or direct selection of available features [48]. There is also using the singular values of axial integrated Wigner bispectrum as the features [40,52], and signal amplitude ranking [49], and multidimension permutation entropy [10].

In the procedure of RF fingerprint extraction and identification, the acquired radio frequency signals are affected by multipath channels compared with the sending signal. Thus, extraction should consider two parts: the characteristics of the channel based and the characteristics of the transmitter based. Channel-based features, which are called channel fingerprinting, characterize the wireless channel response as well as the surrounding environment. While the transmitter-based features, which are called device fingerprints, mainly represent the radio frequency characteristics of the transmitter analogue circuits. The main research in this paper is the latter [18].

In this paper, we propose an intrusion detection method based on RF fingerprint identification. The device authentication of some known communication radio stations via RF fingerprint identification is an important component of wireless network authentication security. Firstly, we extract transient starting signal of the radio and utilize Hilbert transform to obtain the primary feature. Then, the fingerprint feature dimension is reduced using principal component analysis, and we also compare the different dimensional reduction method from three different angles and choose the best reduction method. Finally, machine learning is used to achieve the verdict. After simulation, we propose the final complete model, including the combination of dimension reduction method and classifier in the specific dimension, which can meet the needs of the application.

The remainder of this paper is organized as follows: Sect. 2 contains the signal collection, and several kinds of dimensional reduction methods and classifiers are



**Fig. 1** RF fingerprint extraction and identification

introduced briefly. Section 3 gives the experimental simulation, the dimensional reduction method and classifier for the samples are discussed, and a recognition model with high efficiency of identification is obtained. Section 4 summarizes the full paper, and for the last section, the outlook for the future work is exhibited.

## 2 Theories and methods

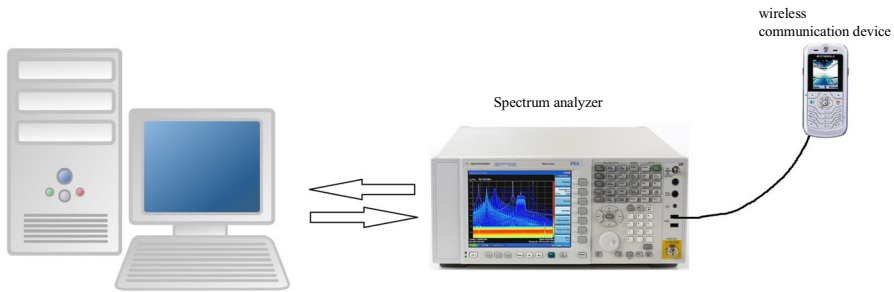
In this paper, we focus on wireless communication devices identification of security intrusion detection. Figure 1 shows the flow chart of signal collection and post-collection processing. First, we collect the signals through a receiver and extract the transient signals by energy threshold detection. The transient signal is artificially added with AWGN noise to simulate channel interference. After Hilbert transform, the dimensional reduction method is used to extract the RF fingerprint features further. Finally, the classifier gives decision results. The simulation in this paper is implemented in MATLAB 2014a.

### 2.1 Signal collection

The instantaneous RF signal is transmitted from ten different wireless communication devices, the Motorola interphone. We collect all signal by using an Agilent receiver. To ensure that there is no noise interference to the collected signals, we attach the wireless communication devices and spectrum analyser directly with the line as shown in Fig. 2. In later experiments, noise is added using simulation software artificially.

### 2.2 RF feature extraction

Hilbert transform is widely used in signal processing field; it can obtain the analytic representation of signals so that the amplitude remains unchanged and the phase shifts [22].



**Fig. 2** The scheme of signal collection

Given a real-time signal  $x(n)$ , its Hilbert transform is defined as:

$$\hat{x}(n) = \frac{2}{\pi} \sum_{k=0}^{N-1} \frac{x(n-2k-1)}{2k-1} \quad (1)$$

It is obvious that  $\hat{x}(n)$  is a linear function of  $x(n)$ . The phase of the signal after the Hilbert transform gets  $j(+\pi/2)$  delay for the positive frequency and  $j(+\pi/2)$  ahead for negative frequencies. The Hilbert transform is the harmonic conjugate of the original function  $x(n)$  [1]; it can transform a real signal to an analytic signal [20].

### 2.2.1 PCA

The classical principal component analysis (PCA) is arguably one of the most important tools in high-dimensional data analysis. Not only in the data processing, but also in dimensionality reduction, PCA has a wide range of applications [30].

For a measured matrix,  $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$ , where  $n_1$  is the number of training samples, while  $n_2$  is the dimension of training samples as well as the number of variables. PCA makes the derived variables capture maximal variance over seeking the linear combinations of the primitive variables. We calculate PCA with the data matrix singular value decomposition (SVD).

PCA makes the greatest variance when mapping the original variables to a new coordinate system, and the projection with the largest component is called the first principal component, and on the second coordinate, the second greatest variance appears, and so on. Finally, multiply the first  $l$  largest singular values and corresponding singular vectors, we can get a truncated score matrix. After construction, the size of the transformed data matrix is  $n_1$ -by- $l$ .

### 2.2.2 RPCA

The robust principal component analysis (RPCA) [5, 37] is the improved method of PCA through a matrix decomposition. RPCA decomposes a matrix  $\mathbf{M} \in \mathbb{R}^{(n_1 \times n_2)}$  into the matrix  $\mathbf{L} \in \mathbb{R}^{(n_1 \times n_2)}$  with low rank and a matrix  $\mathbf{S} \in \mathbb{R}^{(n_1 \times n_2)}$  in sparse by solving

a convex program, the principal component pursuit [41]:

$$\min_{\mathbf{L}, \mathbf{S}} \|\mathbf{L}\|_* + \lambda \|\mathbf{S}\|_1 \quad \text{subject to } \mathbf{M} = \mathbf{L} + \mathbf{S} \quad (2)$$

where  $\mathbf{M}$  is the observation matrix,  $\|\cdot\|_*$  is the norm of the matrix and  $\|\cdot\|_1$  represents the  $L_1$ -norm of a matrix, and  $\lambda$  is a tuning parameter, as shown in [5]. We can estimate  $\lambda$  value by  $\lambda = \frac{1}{\sqrt{\max(n_1, n_2)}}$  and then fine tune it further. The RPCA algorithm can extract the useful information except noise of the original data to find robust low-rank estimation [4].

Since the transient signals generated by radio stations have certain structural information inside, we can assume that this is a low-rank signal because it is linearly related to each row or column. On the contrary, since the transmission process contains noise, and the noise is sparse, we can express noise as a component of a sparse matrix. By RPCA, we expect a low-rank matrix  $\mathbf{L}$  which contain signals information and a sparse matrix  $\mathbf{S}$  containing noise [16]. Finally, we implement the traditional PCA dimensionality reduction for low-rank matrix  $\mathbf{L}$ .

### 2.2.3 KPCA

The basic idea of kernel principal component analysis (KPCA) [36] is also the improved method of PCA. The raw input data  $\mathbf{M}$  is mapped into a new space  $F$  using a kernel functions, and finally, the traditional PCA is implemented in  $F$  [19].

Generally speaking, PCA can be performed greatly in a set of linear variations of observation, while badly in nonlinear variations. However, for KPCA, it can handle nonlinear data. According to the Cover's theorem, nonlinear samples data get linear structure after mapping to high-dimensional space. We call the high-dimensional space of this mapping as feature space ( $F$ ), which can be represented by a kernel function. The kernel function can realize the mapping from nonlinear space to linear space. Finally, KPCA implements traditional PCA algorithm in feature space [23].

## 2.3 Classifier

### 2.3.1 Random forest

Random forest (RndF) was first proposed by Breiman [3]. Random forest is a kind of machine learning; it is an integrated classifier made up of many decision trees. The random vectors used as input have the same distribution and are independent to each other. The final output of the random forest is determined by the results of all decision trees.

Training process of random forest is the training of each decision tree; the training of each decision tree is independent to each other. So the training of random forest can be realized by parallel processing, which will greatly improve the efficiency of generating model. A random forest can be obtained by combining all the  $n_{\text{tree}}$  decision trees trained in the same way [26].

Random forests can exhibit excellent performance in identification mainly because of two stochastic processes: the selection of sample subspaces and the splitting process. As we know, the sampling of the samples is random, the samples in each subspace can be repeated, and samples in different subspaces can also be duplicated. Thus, in the former process, some samples will not be selected, and these samples are called “out-of-bag” (OOB) samples [33]. As for the latter one, decision tree selects the characters randomly and obtains the character subspace.

### 2.3.2 Support vector machines

Support vector machine (SVM) is a kind of machine learning method, which is related to statistical learning theory, VC dimension theory and structural risk minimization principle. SVM has mature theoretical foundation and outstanding application effect [42]. It shows many unique advantages in solving small sample problems and largely overcomes the problems of “Curse of dimensionality” and “over learning”.

The principle of SVM is to find an optimal hyperplane in the sample space to classify, so that the hyperplane can maximize the blank area on both sides of the hyperplane when guaranteeing the accuracy of classification. For the linear separable problem, SVM directly uses the hyperplane to separate the training data in the current feature space. Theoretically, SVM can obtain the optimal classification results in this case [39]. For linear non-separable problems, SVM implements classification by introducing a kernel function. In general, SVM maps training data to a higher-dimensional space via a kernel function, and then, hyperplane separation is performed. By choosing different kernel functions, different high-dimensional spaces can be mapped. The hyperplane in the higher-dimensional space is more complex [42].

### 2.3.3 Artificial neural network

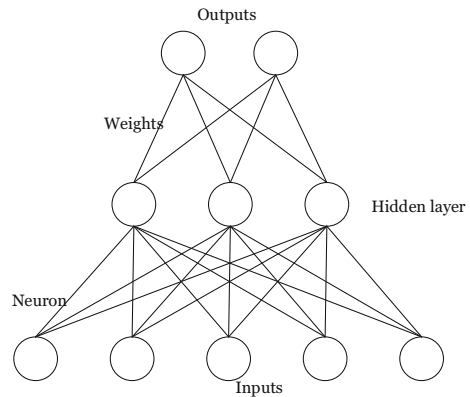
The primary purpose of artificial neural network is to mimic the structure of the human brain and establish a complete system similar to human brain. For convenience, artificial neural networks are generally referred to as neural networks. The neural network has many advantages, it can learn by itself, and the network is very stable. It also shows excellent ability in recognition problem. The neural networks improve the classification performance by changing the node weight, just like brain.

Generally, a neural network consists of three layers, an input layer, a hidden layers and a final output layer. Figure 3 shows a typical architecture, where lines connecting neurons are also shown [47]. The neural network gets outstanding classification ability through training, and this network is difficult to describe with specific formula [15].

### 2.3.4 Grey relational analysis

Grey relational analysis (GRA), which was proposed to get the degree of relation between discreet sequences from imprecise and incomplete information [14], is a part of grey systems. This method is suitable for dealing some classification problems because of its simple calculation [21].

**Fig. 3** Typical architecture of a neural network



The mathematics of GRA is derived from space theory (Deng, 1988). GRA calculates the relation between two discrete sequences to make a distinction [27]. GRA is based on a clustering approach in which the training samples are classified into several groups according to the relations in features and then analyses uncertain relations between signal to be recognized and all the other signals in a given system [25].

GRA contains three major steps. Firstly, comparability sequences are obtained by quantifying the grey relation. Then, in order to get a ideal target sequence, the reference sequence is chosen. Finally, calculate the degree of grey relation and grey relational grade based on comparing the comparability sequence with the reference sequence [21,24].

### 3 Results

#### 3.1 Dataset definitions

For this research, the dataset is composed of the real part of the Hilbert transform envelope of the device transient signal. In order to reduce the number of sampling points, we take a point at each 50 points as a variable, so that each observation consists of 3187 variables. All observations constitute the datasets, and they are divided into the training datasets and testing datasets. The original datasets which are sampled from the authorized devices are contain 500 observations from 10 devices, and each device generate 50 observations without noise, as described in Sect. 2.1. Among all the observations, 300 observations compose the training datasets, and 200 observations compose the test datasets. For more accurate recognition, we add white Gaussian noise (WGN) for many times on each original observation. Training datasets and testing datasets are both constructed with 6000 observations at each SNR. For both training and testing datasets, SNR step is set as 2.5 dB (the range is 0–20 dB). The sizes of datasets are summarized in Table 1, where  $N_{\text{test}}$  is the number of testing datasets observations,  $N_{\text{training}}$  is the number of training datasets observations and  $N_{\text{variables}}$  is the length of each variable.



**Table 1** Datasets summaries

	$N_{\text{test}}$	$N_{\text{training}}$	$N_{\text{variables}}$
Dataset	54,000	54,000	3187

### 3.2 Dimensional reduction performance using random forest

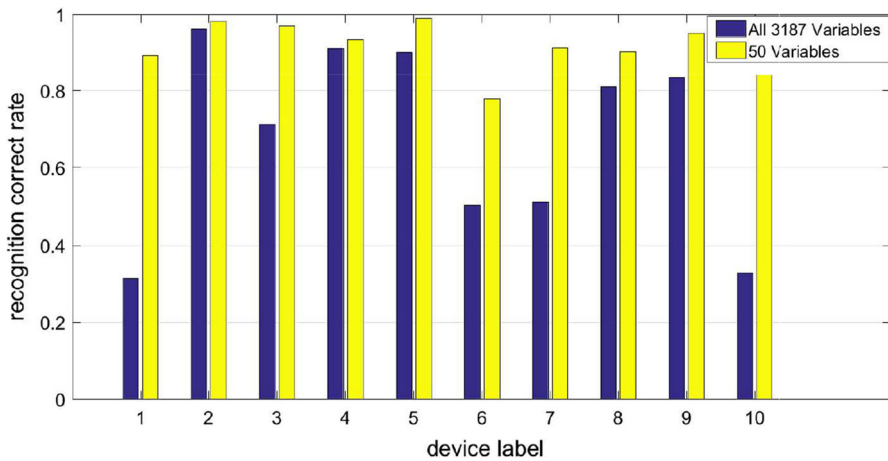
In this paper, we extract fingerprint from the feature of Hilbert transform by reducing the dimension. In order to verify the improvement of the accuracy of equipment identification via extracting, RPCA was introduced here to reduce the feature dimension, and the classifier is simply chosen random forest. The training dataset samples and testing dataset samples under the certain SNR are used to proceed the RPCA method, and we set SNR as 10 dB, since the recognition correct rate is not outstanding so that the comparative effectiveness is obvious. Contrast to the features of the Hilbert transform, the number of reduced variables is arbitrarily chosen as 50.

Figure 4 shows the recognition rate of authorized devices using the random forest to directly classify, since random forests exhibit excellent performance in the classification of some high-dimensional data. In the process, the subspace size arbitrary chosen at each node is  $m = \sqrt{\text{total number of variables}}$ , and total number of decision trees is 1000. The results indicate that for each device individually, the results of dimensionality reduction are far superior to those without dimensionality reduction. In the total recognition results of all devices, the accuracy of recognition without dimensionality reduction is 0.6795, and the recognition accuracy after dimensionality reduction is 0.9165. Theoretically, the original features without dimensionality reduction contain not only the most complete fingerprint information, but also redundancies, which can cause unnecessary interference to the classification. At the same time, samples without dimensionality reduction are correlated with each other, which may lead to misleading classification. The robust principal component analysis makes the original feature be projected in the direction of orthogonal to each other, which greatly reduces the correlation and redundancy between features. Meanwhile, the reduction in input feature can also reduce the complexity of classifier and improve the recognition accuracy.

### 3.3 The discussion of dimensional reduction method

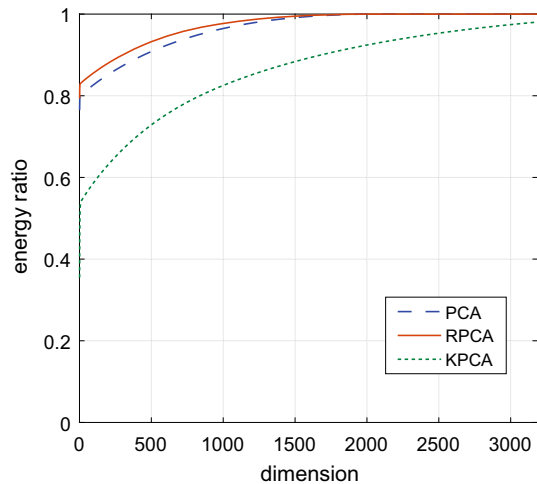
It has been proved that dimensionality reduction has a great contribution to the recognition accuracy of the equipments, so it is worth studying to choose a dimension reduction method which can achieve better results. We will compare the three dimensionality reduction methods mentioned in the second II, the traditional PCA, RPCA and KPCA.

The training datasets and the testing datasets at SNR = 20 are used. Figure 5 shows the energy ratio of the original feature curves obtained by three kinds of dimensional reduction method. The KPCA algorithm maps the raw data into the high-dimensional space firstly and then reduces the dimension in the high-dimensional space. The abscissa of KPCA curve ranges from 1 to higher dimensions. But for comparison,



**Fig. 4** The recognition rate of authorized devices using the random forest

**Fig. 5** The energy ratio of the original feature curves varying with dimensions



the figure only shows the first 3187 dimensional curve of KPCA. It is obvious that RPCA shows the best energy retention characteristics.

Table 2 shows a comparison of dimensionality when the energy reaches 80, 85, 90 and 95% of the total energy. When the reserved energy is constant, the smaller the dimension is, the better the corresponding dimensional reduction method is. From another point of view, Table 3 shows a comparison of reserved energy ratio when the dimension after reduction is 4, 10, 50, 100, 500 and 1000. When the dimension is constant, the bigger the energy ratio is, the better the corresponding dimensional reduction method is. The contents written in bold font are the best in performance, and it is obvious that RPCA is best in dimensional reduction from both tables.

For further analysis, a distance measure is used to evaluate the dimensional reduction method. Intra-class distance indicates the distance of the signals from the same

**Table 2** The comparison of dimensions under different energy ratios

Energy ratio	80%	85%	90%	95%
Dimensions of PCA	6	192	448	834
Dimensions of RPCA	<b>2</b>	<b>76</b>	<b>300</b>	<b>645</b>
Dimensions of KPCA	838	1189	1682	2436

**Table 3** The comparison of energy ratios under different dimensions

Dimensions	4	10	50	100	500	1000
Energy ratio of PCA	0.7985	0.8014	0.8136	0.8274	0.9083	0.9647
Energy ratio of RPCA	<b>0.8284</b>	<b>0.8312</b>	<b>0.8431</b>	<b>0.8564</b>	<b>0.9322</b>	<b>0.9770</b>
Energy ratio of KPCA	0.5154	0.5400	0.5630	0.5878	0.8287	0.8252

Bold means the best situation

device and represents the degree of aggregation of such signals. The smaller the intra-class distance is, the higher the prediction of such signals is, the better for classification.

$$\begin{aligned}\bar{D}^2 &= \frac{1}{N} \sum_{i=1}^N \left[ \frac{1}{N-1} \sum_{j=1}^N \sum_{k=1}^n (x_{ik} - x_{jk})^2 \right] \\ &= \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1, j \neq i}^N \sum_{k=1}^n (x_{ik} - x_{jk})^2\end{aligned}\quad (3)$$

where  $x_{ik}$  is the  $k$ th variable of the  $i$ th sample,  $N$  is the number of samples with the same class and  $n$  is the number of variables.

The inter-class is the distance of two different classes. The larger the distance is, the easier to distinguish the two kinds of signals.

$$D_{i,j} = \|X^{\omega_i} - X^{\omega_j}\| \quad (4)$$

where  $X^{\omega_i} = \frac{1}{N_i} \sum_{x \in \omega_i} X$  is the mean of all samples belonging to the  $\omega_i$  class.

Since this paper totally distinguishes signals from 10 devices, there are 45 inter-class distances for all the classes. We take the average of the distances among inter-classes.

The intra-class instances and the average of inter-class instances of 10 kinds of device signals feature are shown in Table 4. However, the intra-class instance cannot directly distinguish the difference between PCA and RPCA, and we cannot obtain the overall effect only with intra-class distance nor inter-class distance. For example, the sample sets via KPCA dimensionality reduction, although it has a large inter-class distance, the intra-class distance is also large. Therefore, we propose a concept of distance ratio measure, which is the ratio between the mean value of the intra-class distance and the mean value of inter-class distance. The greater the value is, the higher the degree of comprehensive separation of the samples. We obtain the distance ratio of dimension reduced dataset, and RPCA shows the best dimensionality reduction performance.

**Table 4** The intra-class instances and the average of inter-class instances of 10 kinds of device signals feature

Device type	Intra-class distance						
	1	2	3	4	5	6	7
PCA	0.1727	<b>0.1427</b>	0.1597	0.1720	<b>0.2030</b>	<b>0.1932</b>	0.1769
RPCA	<b>0.1724</b>	0.1428	<b>0.1596</b>	<b>0.1719</b>	<b>0.2030</b>	0.1933	<b>0.1768</b>
KPCA	3.4147	1.2404	3.4804	2.7042	1.1714	3.1123	3.6608
Device type	Intra-class distance			Inter-class distance		Distance ratio	
	8	9	10				
PCA	0.1884	<b>0.1605</b>	<b>0.1803</b>	2.7568		15.7624	
RPCA	<b>0.1883</b>	<b>0.1605</b>	<b>0.1803</b>	2.7568		<b>15.7633</b>	
KPCA	4.4750	1.8943	2.8643	<b>10.6285</b>		3.7935	

Bold means the best situation

$$\text{Distance ratio} = \frac{\frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1}^N D_{i,j}}{\frac{1}{N} \sum_i \bar{D}^2} \quad (5)$$

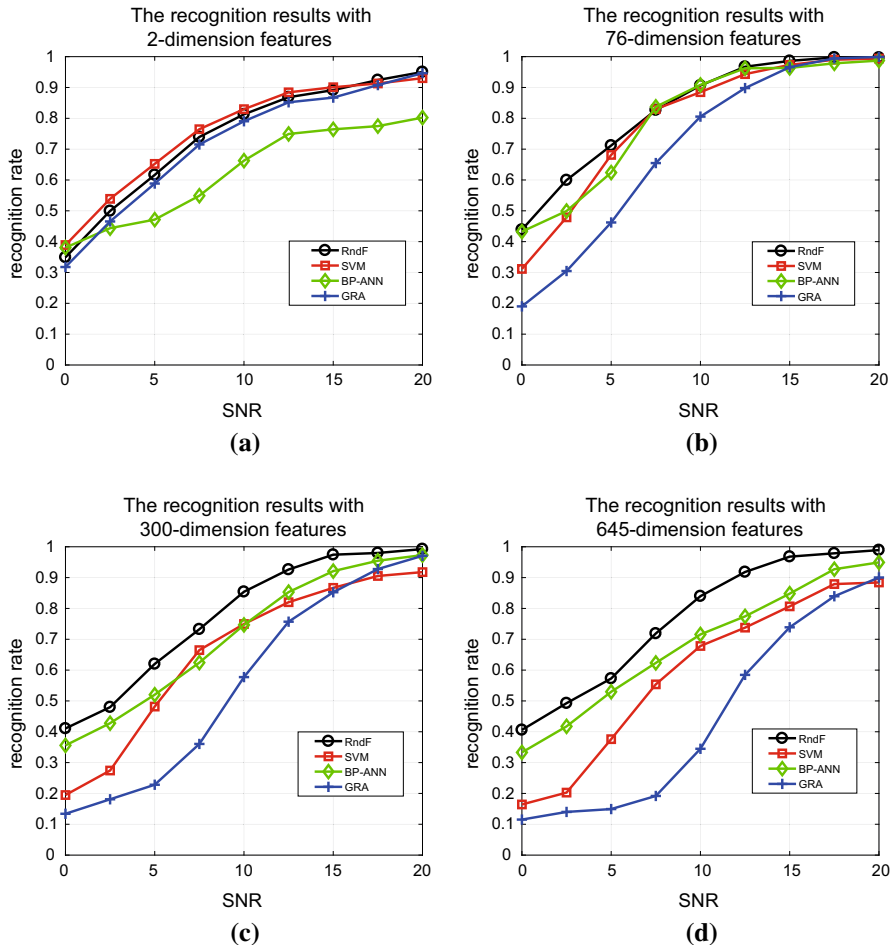
From the analysis of several points above, the KPCA method has the worst dimensional reduction effect, but we can't say that KPCA is not a good dimensionality reduction method. KPCA mainly shows a good dimensionality reduction effect on nonlinear samples, but for different samples, the internal structures are different. In the subsequent simulation experiments, we select RPCA to reduce dataset dimensions for further identification.

### 3.4 RF fingerprint identification results

RPCA is used to reduce the dimensionality of the training datasets and testing datasets. The dimensions of samples are reduced from 3187 to 2, 76, 300 and 645 respectively, depending on the energy retention to 80, 85, 90 and 95%. The parameter settings of random forest are described in Sect. 3.2. The kernel function of SVM is radial basis function. The number of hidden layer nodes in BP neural network is set according to the empirical formula:  $m = \sqrt{N_{\text{input}} + N_{\text{output}}} + a$ , where  $N_{\text{input}}$  is the number of the input node,  $N_{\text{output}}$  is the number of output node and  $a$  is a positive integer less than 10.

#### 3.4.1 Recognition under fixed dimensions

Figure 6 shows the recognition correct rate of authorized devices when the feature dimensions are fixed. The recognition accuracy increases with the increase of SNR. The classifiers behave differently when the features of different dimensions are used as inputs. SVMs perform best when input feature is 2 dimensions, and when the input features are 76 dimensions, RndF, SVM and BP-ANN classifier work almost the same

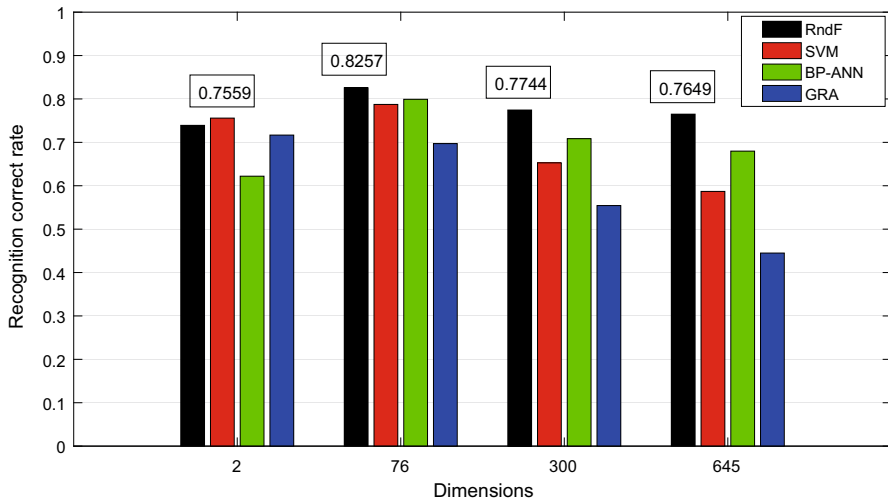


**Fig. 6** The recognition correct rate of authorized devices when the feature dimensions is fixed. **a** Two-dimensional features, **b** 76-dimensional features, **c** 300-dimensional features, **d** 645-dimensional features

effect at  $\text{SNR} \geq 7.5$  dB, while RndF classifier generally outperforms other classifiers when the feature dimensions are larger. Figure 7 shows the results of four classifiers with SNR mixed (for all ranges, 0–20 dB). When the dimensions are more than 76, the recognition rate of RndF is the highest. It is also proved that RndF classifier performs well in large dimensional problems.

### 3.4.2 Recognition under fixed classifiers

Figure 8 shows the recognition rates of different classifiers. No matter which classifier was used, the arbitrary recognition rate can reach 0.9 benchmark at 76-dimensional features, that is, the energy reservation is 85%. For this situation, RndF and BP-ANN classifier achieved at  $\text{SNR} \geq 10$  dB, and SVM and GR classifier achieved at



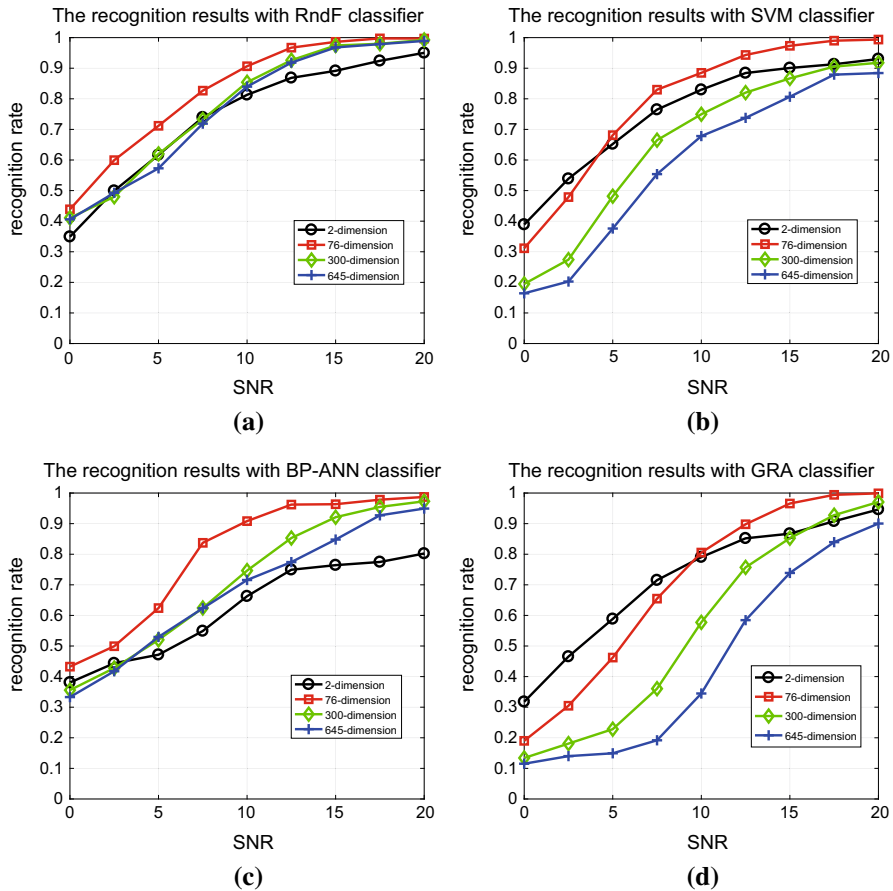
**Fig. 7** The recognition results of four classifiers without SNR fixing, where all ranges of SNR are 0–20 dB

$\text{SNR} \geq 12.5 \text{ dB}$ . Through the change of recognition rate caused by the dimension changes, we can obtain that the recognition rate is increased at the beginning and then is decreased with the increase in dimensions. This is because the fingerprint information is less when the dimension is lower, while for the higher dimensions, it will lead to some redundancy as well as the complexity of the classifier although there is more information. In all the experiments in this paper, the best recognition results are obtained by the system which is composed of 76-dimension features input and RndF classifier. Its comprehensive recognition results in the range of 0–20 dB reached 0.8257. Moreover, the system model shows the best performance in the signal recognition curve with variable SNR.

## 4 Conclusion

Wireless devices are widely used in civilian and military fields, and their security protection becomes critical. In this paper, RF fingerprint is used to identify the authenticated devices, which is an important component of intrusion detection. The work presented here mainly contains: (1) comparison of three kinds of dimensionality reduction methods. Three aspects of energy ratio, reduced dimension and distance ratio measure were analysed from many aspects, and the best RPCA dimensionality reduction method was got suitable for this application. (2) The recognition results of the four classifiers are compared. For all the experiments, using random forest as classifier, the recognition effect is best when the input feature is 76 dimensions. Accordingly, the best identification system model is proposed, and the system model can guarantee the real-world authentication security.

However, the method proposed in this paper still has limitations. The model in this paper is only the identification part in the process of device authentication. For some non-certified devices, the model may make a wrong judgment. So in the future, we



**Fig. 8** The recognition rate of authorized devices when the classifier is fixed **a** RndF, **b** SVM, **c** BP-ANN, **d** GRA

should add a process of judging whether it is an authentication device. Besides, as Sect. 3 described, the best recognition result increases firstly and then decreases with the increase in dimension; therefore, finding the best feature dimension is one of the key points for us to study in the future. However, the scheme may not be the same according to the different classifiers. We're going to solve the problems mentioned above and give a more suitable system model to identify the authenticated devices, including the method of RF fingerprint extraction and classifier design.

**Acknowledgements** This work is supported by the National Natural Science Foundation of China (61301095), the Key Development Program of Basic Research of China (JCKY2013604B001), the Fundamental Research Funds for the Central Universities (GK2080260148 and HEUCF1508). This paper is funded by the International Exchange Program of Harbin Engineering University for Innovation-oriented Talents Cultivation. We gratefully thank of very useful discussions of reviewers.

## Compliance with ethical standards

**Conflict of interest** All the authors declare that there is no conflict of interests regarding the publication of this article.

## References

- Benitez D, Gaydecki P, Zaidi A, Fitzpatrick A (2001) The use of the hilbert transform in ecg signal analysis. *Comput Biol Med* 31(5):399–406
- Bihl TJ, Bauer KW, Temple MA (2016) Feature selection for rf fingerprinting with multiple discriminant analysis and using zigbee device emissions. *IEEE Trans Inf Forensics Secur* 11(8):1862–1874
- Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
- Campbell-Washburn AE, Atkinson D, Nagy Z, Chan RW, Josephs O, Lythgoe MF, Ordidge RJ, Thomas DL (2016) Using the robust principal component analysis algorithm to remove rf spike artifacts from mr images. *Magn Reson Med* 75(6):2517–2525
- Candès EJ, Li X, Ma Y, Wright J (2011) Robust principal component analysis? *J ACM* 58(3):11
- Chen Y, Trappe W, Martin RP (2007) Detecting and localizing wireless spoofing attacks. In: *IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, pp 193–202
- Cobb WE, Garcia EW, Temple MA, Baldwin RO, Kim YC (2010) Physical layer identification of embedded devices using rf-dna fingerprinting. In: *Military Communications Conference, 2010-MILCOM 2010*. IEEE, pp 2168–2173
- Danev B, Luecken H, Capkun S, El Defrawy K (2010) Attacks on physical-layer identification. In: *Proceedings of the Third ACM Conference on Wireless Network Security*. ACM, pp 89–98
- Dener M (2014) Security analysis in wireless sensor networks. *Int J Distrib Sens Netw* 10(10):303,501
- Deng S, Huang Z, Wang X, Huang G (2017) Radio frequency fingerprint extraction based on multidimension permutation entropy. *Int J Antennas Propag* 2017:1–6
- Ding G, Wang J, Wu Q, Yao YD, Song F, Tsiftsis TA (2015) Cellular-base-station-assisted device-to-device communications in tv white space. *IEEE J Sel Areas Commun* 34(1):107–121
- Hall J, Barbeau M, Kranakis E (2005) Radio frequency fingerprinting for intrusion detection in wireless networks. *IEEE Trans Defendable Secur Comput* 12:1–35
- Hall J, Barbeau M, Kranakis E (2003) Detection of transient in radio frequency fingerprinting using signal phase. *Wirel Opt Commun*, pp 13–18
- Ho CY, Lin ZC (2003) Analysis and application of grey relation and anova in chemical-mechanical polishing process parameters. *Int J Adv Manuf Technol* 21(1):10–14
- Hsu KI, Gupta HV, Sorooshian S (1995) Artificial neural network modeling of the rainfall-runoff process. *Water Resour Res* 31(10):2517–2530
- Huang PS, Chen SD, Smaragdis P, Hasegawa-Johnson M (2012) Singing-voice separation from monaural recordings using robust principal component analysis. In: *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp 57–60
- Jana S, Kasera SK (2010) On fast and accurate detection of unauthorized wireless access points using clock skews. *IEEE Trans Mob Comput* 9(3):449–462
- Jia-Bao YU, Ai-Qun HU, Zhu CM, Peng LN, Jiang Y (2016) Rf fingerprinting extraction and identification of wireless communication devices. *J Cryptol Res* 3(5):433–446
- Kim KI, Jung K, Kim HJ (2002) Face recognition using kernel principal component analysis. *IEEE Signal Process Lett* 9(2):40–42
- Konar P, Chattopadhyay P (2015) Multi-class fault diagnosis of induction motor using hilbert and wavelet transform. *Appl Soft Comput* 30:341–352
- Kou G, Lu Y, Peng Y, Shi Y (2012) Evaluation of classification algorithms using mcdm and rank correlation. *Int J Inf Technol Decision Mak* 11(01):197–225
- Kuklik P, Zeemering S, Maesen B, Maessen J, Crijns HJ, Verheule S, Ganesan AN, Schotten U (2015) Reconstruction of instantaneous phase of unipolar atrial contact electrogram using a concept of sinusoidal recomposition and hilbert transform. *IEEE Trans Biomed Eng* 62(1):296–302
- Lee JM, Yoo C, Choi SW, Vanrolleghem PA, Lee IB (2004) Nonlinear process monitoring using kernel principal component analysis. *Chem Eng Sci* 59(1):223–234



24. Li J (2015) A new robust signal recognition approach based on holder cloud features under varying snr environment. *Ksii Trans Internet Inf Syst* 9(12):4934–4949
25. Liang RH (1999) Application of grey relation analysis to hydroelectric generation scheduling. *Int J Electr Power Energy Syst* 21(5):357–364
26. Liaw A, Wiener M et al (2002) Classification and regression by randomforest. *R News* 2(3):18–22
27. Lin SJ, Lu I, Lewis C (2007) Grey relation performance correlations among economics, energy use and carbon dioxide emission in taiwan. *Energy Policy* 35(3):1948–1955
28. Medhane DV, Sangaiah AK (2017) Search space-based multi-objective optimization evolutionary algorithm. *Comput Electr Eng* 58:126–143
29. Patel H (2015) Non-parametric feature generation for rf-fingerprinting on zigbee devices, pp 1–5
30. Peng K, Zhang M, Li Q, Lv H, Kong X, Zhang R (2016) Fiber optic perimeter detection based on principal component analysis. In: 2016 15th International Conference on Optical Communications and Networks (ICOON), Hangzhou, 2016, pp 1–3. <https://doi.org/10.1109/ICOON.2016.7875605>
31. Qiu T, Zhang Y, Qiao D, Zhang X, Wymore ML, Sangaiah AK (2017) A robust time synchronization scheme for industrial internet of things. *IEEE Trans Ind Inform* 99:1
32. Rehman SU, Sowerby KW, Coghill C (2014) Analysis of impersonation attacks on systems using rf fingerprinting and low-end receivers. *J Comput Syst Sci* 80(3):591–601
33. Rodriguez-Galiano V, Chica-Olmo M, Abarca-Hernandez F, Atkinson PM, Jeganathan C (2012) Random forest classification of mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sens Environ* 121:93–107
34. Rostami AS, Badkoobe M, Mohanna F, Keshavarz H, Hosseinabadi AAR, Sangaiah AK (2017) Survey on clustering in heterogeneous and homogeneous wireless sensor networks. *J Supercomput* 13:1–47
35. Rostami H, Blue J, Yugma C (2017) Equipment condition diagnosis and fault fingerprint extraction in semiconductor manufacturing. In: IEEE International Conference on Machine Learning and Applications. IEEE, pp 534–539
36. Schölkopf B, Smola A, Müller KR (1997) Kernel principal component analysis. In: International Conference on Artificial Neural Networks. Springer, pp 583–588
37. Shahid N, Kalofolias V, Bresson X, Bronstein M, Vandergheynst P (2016) Robust principal component analysis? In: IEEE International Conference on Computer Vision. IEEE, pp 2812–2820
38. Sheng Y, Tan K, Chen G, Kotz D, Campbell A (2008) Detecting 802.11 mac layer spoofing using received signal strength. In: INFOCOM 2008. The 27th Conference on Computer Communications. IEEE, pp 1768–1776
39. Shifei D, Binguan Q, Hongyan T et al (2011) An overview on theory and algorithm of support vector machines. *J Univ Electron Sci Technol China* 40(1):2–10
40. Sun M, Zhang L, Bao J, Yan Y (2017) Rf fingerprint extraction for GNSS anti-spoofing using axial integrated wigner bispectrum. *J Inf Secur Appl* 35:51–54
41. Tang G, Nehorai A (2011) Constrained crammer bound on robust principal component analysis. *IEEE Trans Signal Process* 59(10):5070–5076
42. Tong S, Koller D (2001) Support vector machine active learning with applications to text classification. *J Mach Learn Res* 2(Nov):45–66
43. Toonstra J, Kinsner W (1995) Transient analysis and genetic algorithms for classification. In: IEEE WESCANEX 95: Communications, Power, and Computing: Conference Proceedings. IEEE, vol 2, pp 432–437
44. Ureten O, Serinken N (2007) Wireless security through rf fingerprinting. *Can J Electr Comput Eng* 32(1):27–33
45. Uzundurukan E, Ali AM, Kara A (2017) Design of low-cost modular rf front end for rf fingerprinting of bluetooth signals. In: Signal Processing and Communications Applications Conference. IEEE, pp 1–4
46. Vishwasrao MD, Sangaiah AK (2017) Escape: effective scalable clustering approach for parallel execution of continuous position-based queries in position monitoring applications. *IEEE Trans Sustain Comput* 2(2):49–61
47. Wang SC (2003) Artificial neural network. In: Interdisciplinary computing in java programming. Springer, pp 81–100
48. Wang W, Sun Z, Piao S, Zhu B, Ren K (2016) Wireless physical-layer identification: modeling and validation. *IEEE Trans Inf Forensics Secur* 11(9):2091–2106
49. Xie FY, Cheng W, Chen Y, Wen H (2017) An optimized algorithm for the feature extraction and recognition of rf fingerprint based on signal amplitude ranking. *Cyberspace Secur* 1:85–87

50. Yuan HL (2011) Research on physical-layer authentication of wireless network based on rf fingerprinting. PhD thesis, Southeast University
51. Zhang G, Xia L, Jia S, Ji Y (2017) Identification of cloned HF RFID proximity cards based on rf fingerprinting. In: Trustcom/bigdataase/ispa, pp 292–300
52. Zhao N, Yu FR, Sun H, Li M (2016) Adaptive power allocation schemes for spectrum sharing in interference-alignment-based cognitive radio networks. *IEEE Trans Veh Technol* 65(5):3700–3714. <https://doi.org/10.1109/TVT.2015.2440428>