Research Article

An Ensemble Learning Method for Wireless Multimedia Device Identification

Zhen Zhang,¹ Yibing Li,¹ Chao Wang,¹ Meiyu Wang,¹ Ya Tu,¹ and Jin Wang ²

¹College of Information and Communication Engineering, Harbin Engineering University, Harbin, 150001, China
²School of Computer & Communication Engineering, Changsha University of Science & Technology, Changsha 410114, China

Correspondence should be addressed to Jin Wang; jinwang@csust.edu.cn

Received 27 May 2018; Revised 9 August 2018; Accepted 27 September 2018; Published 15 October 2018

Guest Editor: Weizhi Meng

Copyright © 2018 Zhen Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In the last decade, wireless multimedia device is widely used in many fields, which leads to efficiency improvement, reliability, security, and economic benefits in our daily life. However, with the rapid development of new technologies, the wireless multimedia data transmission security is confronted with a series of new threats and challenges. In physical layer, Radio Frequency Fingerprinting (RFF) is a unique characteristic of IoT devices themselves, which can difficultly be tampered. The wireless multimedia device identification via Radio Frequency Fingerprinting (RFF) extracted from radio signals is a physical-layer method for data transmission security. Just as people’s unique fingerprinting, different Internet of Things (IoT) devices exhibit different RFF which can be used for identification and authentication. In this paper, a wireless multimedia device identification system based on Ensemble Learning is proposed. The key technologies such as signal detection, RFF extraction, and classification model are discussed. According to the theoretical modeling and experiment validation, the reliability and the differentiability of the RFFs are evaluated and the classification results are shown under the real wireless multimedia device environments.

1. Introduction

With the rapid development of the wireless multimedia technologies, the security of wireless multimedia data transmission becomes increasingly more and more important. The design of efficient identification and authentication algorithms among different wireless multimedia devices has become an urgent subject.

As is well known, the Internet of Things [1] and the wireless sensor network (WSN) [2–4] are important carriers of multimedia data transmission; they will lead to improved efficiency, reliability, security, and economic benefits in our daily life [5]. At the same time, wireless multimedia technology and application have been widely studied by the researchers all over the world [6, 7]. However, because of the opening of transmitting channels, compared with the traditional wired network, the wireless network is more vulnerable to large-scale malicious attacks. Now, many existing networks are unprotected against a lot of different malicious attacks [8–10]; meanwhile the security of them is confronted with a series of new threats and challenges [11–13]. Traditional methods for protecting the security of wireless network are usually based on bit-level security protocol. However, there are usually loopholes in the actual wireless network security protocols [14]. For example, the IEEE 802.11 Wireless LAN’s (WLAN) wired equivalent encryption protocol (WEP) is easily attacked by statistical analysis of [15]; although it has been upgraded to WPA and WPA2, its password can be restored, and there are still a variety of security problems [16]. The Radio Frequency Fingerprinting (RFF) is an inherent characteristic of wireless multimedia devices, which can hardly be tampered. In recent years, RFF extraction and identification methods for wireless multimedia devices have been widely studied [17–20].

The identification and authentication wireless multimedia device based on RFF is an important physical-layer method for wireless multimedia security [21–23], which has been widely used in intrusion detection [24], access control [25], wormhole detection [26], and cloning detection [27]. RFF is extracted from radio signals from wireless multimedia devices, which is a unique characteristic of wireless multimedia devices themselves and can difficultly be tampered. In
physical layer, RFF is just as people’s unique fingerprinting; different wireless multimedia devices exhibit different RFF which can be used for identification and authentication. As is well known, RFF is derived hardware imperfection of wireless multimedia device, which can be observed and extracted. With the development of machine learning and the emergence of a large number of new technologies [28], new methods about RFF have been put forward continuously in recent years. Han J et al. [29] propose a physical-layer identification and authentication system for ultrahigh frequency (UHF) passive tags, which is called GenePrint. The classification accuracy of the passive tags is higher than 99.68%. Furthermore, GenePrint can effectively defend against the replay attack. Huang G et al. [30] propose a novel Specific Emitter Identification (SEI) method based on nonlinear characteristics. The permutation entropy is calculated as the radio signal fingerprint for identifying the unique transmitter. Furthermore, the technology of bispectrum and stray parameter are used for the comparison with the new method, which indicates that the proposed method has a better performance in the classification of the wireless network cards. The PHY-based security based on Time Domain (TD) is also studied extensively in the recent years. Donald R et al. [31] demonstrate the performance of Dimensional Reduction Analysis (DRA) using discrete Gabor-Transform features, which are extracted from the Wi-Fi and WiMAX signals. Jia Y et al. [32] attempt to simultaneously find a low-rank representation matrix of original data and the optimal classifier parameter, which can be used to improve the performance of radiometric identification. Experiments indicate that the new method not only has a higher accurate classification and identification rate, but also has better robustness against noise.

In this paper, the structure of wireless multimedia device identification is proposed. Firstly, the main components of this structure are presented. Secondly, the key technologies such as signal collection, RFF generation, and classification model are discussed. Thirdly, according to the theoretical modeling and experiment validation, the differentiability of the RFFs is extracted. The classification result is shown under the real wireless multimedia device environments. Finally, the advantages and disadvantages of the proposed algorithm and its future prospects are described.

2. General View

In physical layer, the classification and identification of wireless multimedia device include four entities, which are shown in Figure 1: Acquisition Signal Module: acquiring the radio signals from wireless multimedia devices; Burst Extraction Module: detecting and intercepting the start of the turn-on transient; Signal Analysis Module: obtaining identification-relevant information from the radio signals; Fingerprint Generation Module: reducing assist information and generating the RFF; Classifier Module: a classifier for comparing RFF and requesting the identification of the comparison results [33, 34]. Furthermore, in order to better verify the identification performance of wireless multimedia device under different signals to noise ratio (SNR), the Additive White Gaussian Noise (AWGN) module and the data condition module are used in the experiments.

2.1. Data Set Definitions. As shown in Figure 2, the signals in this paper are mainly collected by Agilent oscilloscope, and 10 wireless multimedia devices of the same model and manufacturer are used for this research. In order to dislodge the influence of channel environment, wireless multimedia devices and receiving devices are connected directly with a cable (that is to say, ignoring the influence of multipath, time delay, and clutter in signal transmission process).

The sampling rate of the receiver is 40 MHz; each of the turn-on transients contains 159901 sampling points. For obtaining the complete turn-on transients, the Variance Trajectory (VT) [35] algorithm and Bayesian Change Detection (BCD) [36] are used for transient point detection and interception. The original dataset contains 500 transients from 10 devices, each of which contain 50 samples. According to the ratio of 2:3, the 500 signals are divided into training samples and test samples. At the same time, in order to
2.2. Signal Analysis

2.2.1. Time Domain RF-DNA Fingerprinting. For TD fingerprinting, Radio Frequency-Distinct Native Attribute (RF-DNA) can be generated by the instantaneous amplitude, frequency, and phase response of radio signal’s subsequence [37]. The unique features are obtained by the standard deviation (δ), variance (δ²), skewness (γ), and kurtosis (κ) from N_R + 1 subsequence of the original signal, where N_R represents the number of the subsequence. The statistics can be arranged as follows:

\[ F_{R_i} = \begin{bmatrix} \delta_{R_i}, \delta_{R_i}^2, \gamma_{R_i}, \kappa_{R_i} \end{bmatrix}_{1 \times 4} \]

(1)

Where \( i = 1, 2, \ldots, N_R + 1 \). Then, \( F_{R_i} \) can be used to generate the final TD fingerprint:

\[ F^C = \begin{bmatrix} F_{R_1}, F_{R_2}, \ldots, F_{R_{N_R+1}} \end{bmatrix}_{4 \times (N_R+1)} \]

(2)

Where \( C \) refers to the signal’s instantaneous parameter, including \( \{a(n)\}, \{\varphi(n)\} \) and \( \{f(n)\} \).

2.2.2. Wavelet Domain (WD) RF-DNA Fingerprinting. For signal analysis, the Discrete Wavelet Transform (DWT) is a very effective tool. But there are still some disadvantages. One distinct disadvantage of DWT is that it is not shift invariant. The DT-CWT is an improved method of DWT, which is used to overcome the disadvantage of DWT. The DT-CWT is commonly implemented by two real-valued filter banks, which is shown in Figure 3.

In Figure 3, the two filter banks represent two branches of Tree1 and Tree2, respectively, where the filter coefficients \( h_1(t), h_0(t), h_1'(t), \) and \( h_0'(t) \) are implemented directly as the Analysis Filters (AF) given in [38].

For real-valued input radio signals, the WD coefficients \( I_{WD}^L \) and \( Q_{WD}^L \) of Tree1 and Tree2 represent the real and imaginary components of complex coefficients [39]:

\[ S_{WD}^L(n) = I_{WD}^L(n) + jQ_{WD}^L(n) \]

(3)

Then the WD fingerprints can be generated using the similar method [40] in Section 2.2.1.

2.3. Fingerprint Generation. The Robust Principle Component Analysis (RPCA) is an improved algorithm for traditional Principal Component Analysis (PCA). Because of the serious robustness problem of traditional PCA technology, the theoretical framework of RPCA was put forward to solve this problem.

Assuming that the observation matrix \( D = R^{(M+N)} \) is originally a low-rank matrix \( A = R^{(M+N)} \), it is polluted by matrix \( E = R^{(M+N)} \) which has sparse distribution and
arbitrarily large amplitude. The RPCA tries to separate the low-rank part from the sparse part of the observation matrix $D$ and obtains the low-rank distribution matrix $A$ and the sparse distribution matrix $E$, respectively. By increasing the constraints of low rank and sparsity of matrix $A$ and matrix $E$, the sparse matrices can be computed by calculating the following convex optimization [41, 42]:

$$
\min_{A,E} \|A\| + \lambda \|E\|
$$

s.t. $D = A + E$

Where $\| \cdot \|_*$ is the kernel norm of matrix; it can also be understood as the sum of singular values in a matrix. $\lambda > 0$ is the tuning parameter [43] to balance the low-rank matrix and the sparse matrix.

In order to verify the performance, the dataset in Section 2.1 is used for simulation and evaluation. The dimensions are decreased based on RPCA analysis method. According to the contribution rate of main components, the first two principal components are selected for visualization.

### 3. Designed Classifier

#### 3.1. Adaboost Algorithm

Adaboost algorithm is an iterative algorithm. It does not need to know the prior knowledge of weak classifiers. Instead, it changes the distribution of data and combines weak classifiers to achieve classification. This method has achieved good results in practical applications.

The specific description of the Adaboost algorithm is as follows.

**Input:** $N$ labeled training samples $\{(1, c(1)), \ldots, (N, c(N))\}$, where $1, \ldots, N$ are serial numbers; $c(N)$ characterizes the different sample categories with a range of $(0, 1)$; the training sample has a distribution space and uses a weak learning algorithm; the number of iterations is set to $T$.

**Initialization:** if the sample $y_i = 0$, then make $w_{1,j} = 1/2m$; if the sample $y_i = 1$, then $w_{1,j} = 1/2l$.

**Execution:** the number of iterations is $T$, and each iteration performs the following steps.

1. **Normalised weights:**
   $$w_{i,j} \rightarrow w_{i,j}$$

2. **Calculate the current error of weight:**
   $$e_j = \sum_i w_{i,j} |h_j(x_i) - y_i|$$

Among them, $h_j$ is a weak classifier generated by the feature $j$.

3. **Screening:** the screening process is to add the smallest weak classifier in the previous step to the strong classifier we need.

4. **Update:** regenerate each weight:
   $$w_{i+1,j} = w_{i,j} \beta_t^{-e_i}$$

Where $\beta_t = e_i/(1 - e_i)$, and if the $i^{th}$ sample $x_i$ is correctly classified, then $e_i = 0$; otherwise $e_i = 1$.

#### 3.2. Gradient Boosting Decision Tree

Gradient Boosting Decision Tree (GBDT) is a combination of decision tree and Boosting method. It is also one of the integrated learning Boosting algorithms [44]. In the GBDT algorithm, the decision tree training object is not a tree but an error in the classification of the previous decision tree. This is the Boosting concept. The principle of the gradient-based decision tree algorithm is shown in Figure 4.

As can be seen from Figure 4, GBDT is a kind of linear training, so it cannot be trained in parallel. In training, the difference between the training and the actual value is the target of the second tree optimization, as shown in formula $T_2$:

$$T = T_1 + T_2 + T_3$$

From the idea of the algorithm shown in Figure 4, we can see that there is a difference between GBDT and traditional boosting. That is, the GBDT iteration is the residual or gradient descent value, and boosting is the sample data. In Figure 5 is an example of using GBDT for binary classification.

$$h_j(x) = \begin{cases} 1, & \frac{1}{2} \sum_{i=1}^r a_i h_i(x) \geq \frac{1}{2} \sum_{i=1}^r a_i \\ 0, & \text{else} \end{cases}$$

In the above formula, $a_i = \log(1/\beta_i)$.
In the process of binary classification of decision trees, the criteria for classification are to select the attribute that maximizes the classification gain, and the calculation method for split gain is as follows:

\[
G = S - S_j
\]  

(12)

In the above equation, to obtain the \(S_j\), the loss function can be replaced by a variance, as follows:

\[
S_j = \sum_{m \in L} (y_m - \mu_L)^2 + \sum_{n \in R} (y_n - \mu_R)^2
\]  

(13)

After \(S_j\) is obtained, \(S\) and \(S_j\) are sequentially expanded, respectively, as in equation (12); then the split gain \(G\) can be obtained:

\[
G = \left( \frac{\text{Sum}_L^2}{\|L\|} + \frac{\text{Sum}_R^2}{\|R\|} \right) - \frac{\text{Sum}^2}{\|Total\|}
\]  

(14)

In the above equations, \(\text{Sum}_L^2\) and \(\text{Sum}_R^2\) represent the square sum of samples on the left and right subtrees, while \(\text{Sum}^2\) is used to represent the sum of the squares of all the samples. Therefore, to make the classification work best, \(G\) gets the maximum in each iteration. It should be noted that, in the GBDT algorithm, the purpose of each decision tree training is to optimize the residual of the previous item.

4. Simulation Result

4.1. Dimension Reduction Algorithm. In general, the features extracted by the feature extraction method have certain redundancy and noise. If the identification is performed directly, it will not only increase the computational complexity of the subsequent process and the requirements for computer memory, but also affect the final classification identification effect. Therefore, it is necessary to reduce the dimensionality of original feature set. Section 2.3 introduces the dimensionality reduction algorithms for PCA and RPCA, where RPCA is an improved algorithm for PCA. In Figure 6, when SNR=20dB, the energy ratio curves of the time domain and wavelet domain characteristics were obtained by using PCA and RPCA methods, respectively.

![Figure 6: The energy ratio of original feature curves changing with dimensions.](image)

**Table 1:** The relationship between the energy ratio and the dimensionality reduction.

<table>
<thead>
<tr>
<th>TD</th>
<th>WAV</th>
</tr>
</thead>
<tbody>
<tr>
<td>85%</td>
<td>6</td>
</tr>
<tr>
<td>90%</td>
<td>13</td>
</tr>
<tr>
<td>95%</td>
<td>24</td>
</tr>
</tbody>
</table>

The energy ratio refers to the ratio of feature vector information after dimension reduction to the feature vector information without dimension reduction. It can be seen from Figure 6 that RPCA has a better dimensionality reduction effect. Under these two feature extraction methods, a higher proportion of energy can be achieved in the same dimension. Considering these advantages of RPCA, the RPCA method is chosen in this paper to reduce the original features.

Table 1 shows the reduced dimension of several typical energy ratios in the use of time and wavelet domain feature extraction methods and the use of RPCA.

4.2. Classification and Identification. Experiments mainly selected the two ensemble learning classifiers described in
Section 3, Adaboost and GBDT. In order to verify the performance of classifier, k-Nearest Neighbor (KNN) and Grey relational analysis (GRA) classifier are used for comparison. The ensemble classifier depth is 5, learning rate is 0.3, and number is 50. Before classification and identification, it is necessary to determine the specific feature dimensions. According to the results in Table 1, we selected the dimensions corresponding to the energy proportions of 85%, 90%, and 95% compared with the original features, respectively. Among them, for the time domain feature set, since the energy proportion has exceeded 95% when it falls to three dimensions, the experiment is selected in 3 dimensions; for the feature set of the wavelet domain, 6, 13, 24, and 200 dimensional features are selected for testing.

From Figure 7, in the range of 0∼35dB SNR, the identification rate increases with SNR. The four classifiers have the same performance; meanwhile the identification rate of different signals is basically stable at 10dB; the identification rate is up to 95% in this condition.

We can obtain the same results as above for the classification and identification results of wavelet domain RF-DNA Fingerprinting in high SNR in Figure 8. However, the performance of time domain RF-DNA Fingerprinting is better in low SNR. Meanwhile, the performance of GBDT and Adaboost can achieve a higher identification rate compared with KNN and GRA, especially when the dimension of feature is high. At the same time, compared with the experimental results in Figure 7, it can be seen that time domain RF-DNA fingerprint features have better performance at low SNR. This is mainly because the data set in this paper is formed by the turn-on transient signal of wireless devices. It has a distinct envelope feature, and the time domain RF-DNA fingerprint is mainly based on the instantaneous amplitude phase and frequency characteristics, so it is easier to distinguish different devices. In order to verify the effect of dimension reduction after RPCA dimension reduction on overall identification rate, this article compares the overall identification rate of signals in the range of 0∼35dB SNR in different dimensions, as shown in Figure 9.

As can be seen from Figure 9, with the increase of feature dimensions, the average identification rate of the signal shows a trend of increasing first and then decreasing. This is because when the input dimension is too low, the information carried by the feature is too small, and when the input dimension is too high, although there is more comprehensive original feature information, it will also increase noise and redundancy, which will lead to a reduction of identification rate.

5. Conclusion

With the popularity and development of wireless networks, the security of wireless networks has gradually become a research hotspot. Radio frequency fingerprinting technology is a kind of wireless network security technology based on the characteristics of physical layer. It can be used for the identification of most existing wireless multimedia devices, and it has a wide range of application scenarios. RF-DNA Fingerprinting technology is a brand-new RF Fingerprinting method developed in recent years, and it has good device classification identification effect. This paper proposes an RF-DNA Fingerprinting system based on ensemble learning and uses the RF-DNA Fingerprint feature based on time domain and wavelet domain to verify the classification and identification performance. In order to reduce the computational overhead and redundant information of the original features, in this paper, RPCA is introduced to reduce the dimensionality of original features. The experimental results show that it has good classification and identification performance. When the SNR is greater than 10 dB, the GBDT classifier is used to
Figure 8: The identification rate of authorized devices under different dimensions. (a) 6-dimensional features, (b) 13-dimensional features, (c) 24-dimensional features, (d) 200-dimensional features.

Figure 9: The identification results of four classifiers with SNR = 0 dB~20 dB.
classify the RF-DNA features in the time domain and wavelet domain, and the identification rate can reach over 95%, which can realize the effective identification of different equipment types.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References


