USING SDR PLATFORM TO EXTRACT THE RF FINGERPRINT OF THE WIRELESS DEVICES FOR DEVICE IDENTIFICATION

Ting-Yu Lin, Chia-Min Lai and Chi-Wei Chen

Institute for Information Industry, Taipei, R.O.C, Taiwan

ABSTRACT

Due to the advent of the Internet of Things era, the number of related wireless devices is increasing, making the abundant and complex information networks formed by communication between devices. Therefore, security and trust between devices a huge challenge. In the traditional identification method, there are identifiers such as hash-based message authentication code, key, and so on, often used to mark a message that the receiving end can verify it. However, this kind of identifiers is easy to tamper. Therefore, recently researchers address the idea that using RF fingerprint, also called radio frequency fingerprint, for identification. Our paper demonstrates a method that extracts properties and identifies each device. We achieved a high identification rate, 99.9% accuracy in our experiments where the devices communicate with Wi-Fi protocol. The proposed method can be used as a stand-alone identification feature, or for two-factor authentication.

KEYWORDS

Internet-of-Things (IoT), Authentication, RF fingerprint, Machine Learning (ML), Device Identification.

1. INTRODUCTION

The IoT, Internet of Things, is growing rapidly with the diverse technologies and usages. It allows data to be transmitted between wireless devices and the Internet. In such a convenience environment, a great number of devices are also increased and can be seen widely including medical devices, sensors and airplanes [1] (Figure 1). However, in the position of huge business opportunities, it is also accompanied with risk. It might result in that the information systems to be intruded, used, damaged, and modified if there is no appropriate management technology about wireless devices. In other words, the importance of information protection and security cannot be ignored anymore.

The growing number of intelligent devices will create abundant and complex information network that allow the supply chain to utilize wireless technology to realize the communication between devices. The utilization of the Internet not only help with building the connection between humans but also linking between human and objects, object and objects. For example, people make the use of smart phone to control the vehicle or intelligent appliances.

Safety aspect is the most concern issue in IoT. The application data can be personal, agriculture, industry, enterprise, health care or environmental protection. It should be well-protected in case it is stolen or tampered. For example, the application can save physical conditions, purchasing

David C. Wyld et al. (Eds): CCSEA, BIoT, DKMP, CLOUD, NLCAI, SIPRO - 2020
pp. 161-174, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101013

behavior, location, financial statements, inventory, business order, environmental monitoring and history record.

Each device in IoT can create massive data, as the result, saving, protection and analysis are the big challenges. Internet should be able to deal with high capacity and high density devices. Moreover, it should be recognized between legitimate and malicious wireless devices. Therefore,

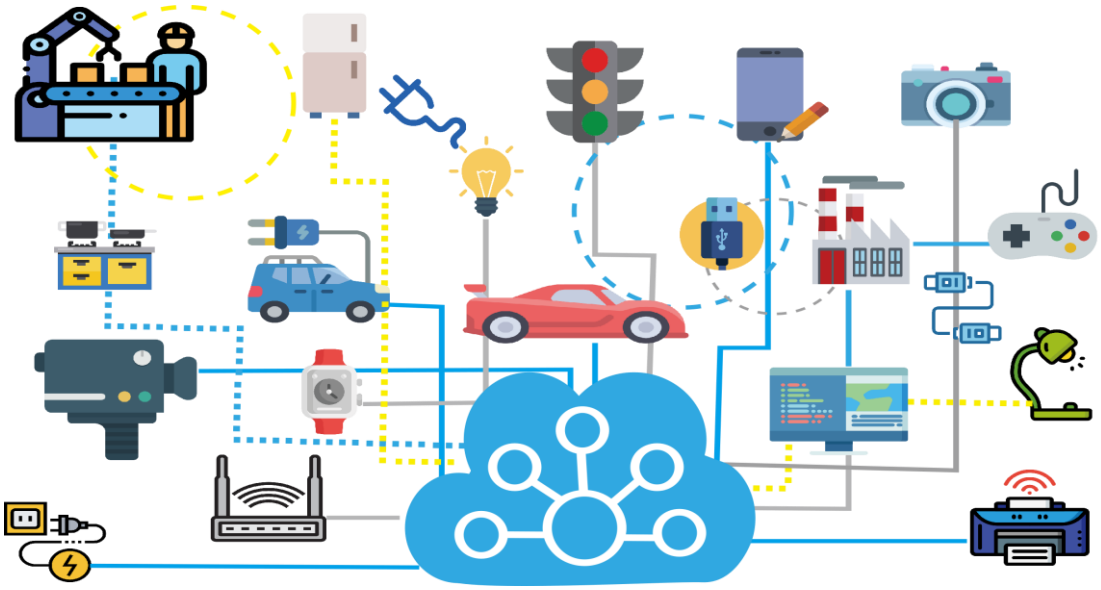

Figure 1. IoT network

how to identify different wireless devices has attracted a great attention of cybersecurity researches and related industries.

Authentication is the act of proving an identity assertion that provides the safety communication process between the users. Traditional authentication information is often mark with identifiers to verify the legitimacy of the information, such as Message Authentication Code (MAC), and Key that enable the receiver to authenticate the key as we call it Symmetric Encryption, which realize the secure data communication within the network [2]. This kind of encryption is good at handling the small number of nodes in the internet, but in the face of a large and complex internet like IoT, it would be doubt and risking if all nodes share the same key. Therefore, in the case that the information security may be insufficient, it is necessary to have a corresponding new technology to effectively improve its security.

It is impossible that two different devices are exactly the same. Because there are some uncontrollable random physical changes in the process of producing, resulting in some slight differences between transmitters. These differences existed in randomness and uniqueness, which establish the foundation for non-replicable. So we can use these features as the fingerprint of the transmitters, called RF fingerprint.

Recently, there are researches pointing out that inconsistency between hardware can efficiently identify different devices, enhancing the security of wireless communication devices. It has aroused our interest and started to study related technologies. In order to explore the performance of the device identification by RF fingerprint, we extracted the features after receiving the signal transmitted by the wireless devices in the shielding box by SDR platform, and then used the machine learning suite like XGBoost to train the classification model through the recorded data.

Finally, the model can identify devices for subsequent new data. The accuracy rate of device identification is 99.97% in Experiment A. We only used power spectral density (PSD) as an RF fingerprint, the accuracy rate of identification is 99.94% in Experiment B. In experiment C, we tended to investigate that what would happen if we switch the original receiver to another one.

The identification rate did drop significantly. Per the result, we supposed that RF fingerprint is relative, related to transmitter (Tx) and receiver (Rx). Detailed experimental results are presented in the fifth chapter. The contribution of this paper has two main parts:

- We proposed a system that can extract modulation-base and transient-based properties from signals and distinguish right devices from others. Our system achieves high accuracy rate, 99.97% and 99.94% using modulation-based and transient-based features respectively.
- We had observed that RF fingerprint existed between Tx and Rx is relative.

By developing the device identification technology, in addition to the device control and device resolution in a specific field, it can also be regarded as a way of authentication to improve security requirements.

The remainder of the paper is organized as follows. Section 2 presents the recently researches in RF fingerprint and machine learning (ML). Section 3 give an introduction of feature extraction and classification model. We show the experiments and finding in section 4. We conclude our paper and set the goal in the future in section 5.

2. RELATED WORK

2.1. Communication Process

We can utilize the inherent digital signal processing to affect the RF characteristics of the signal transmitter, and the RF fingerprint of the signal received and stored by the receiver. We use the transmitting and receiving process of basic modulation signal as an example to explain what deviations may occur in the entire communication process (Figure 2).

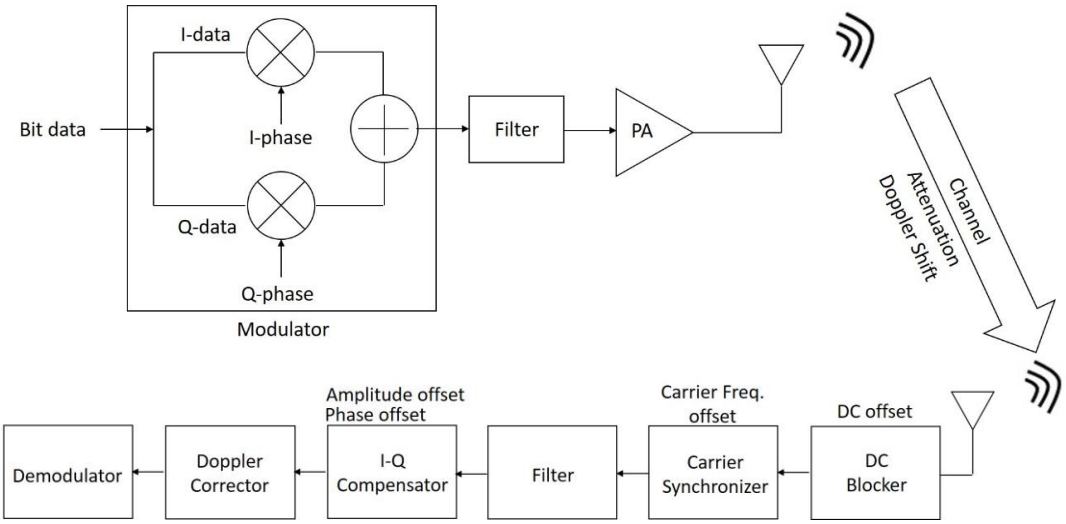

Figure 2. The transmitting and receiving process of basic modulation signal

First, the I-Q Imbalance is the amplitude and phase mismatch between the two paths of in-phase and quadrature signals [3]. And [4] pointed out that the non-ideal deviation caused by I-Q are also the magnitude of the carrier feedthrough signal and the angle error between I-Q components. DC offset is the mean amplitude of the time domain signal [5]. If it is not eliminated, it will cause the offset of the symbol position in the constellation [6]. The carrier frequency offset (CFO) means that the carrier frequency between transmitter (Tx) and receiver (Rx) is not synchronized. By IEEE 802.11 WLAN standards, the range of deviation is strictly limited [7]. The attenuation refers to the fact that when a signal propagates in space, a part of the energy is converted into heat or absorbed by the transmission medium, resulting in weakened signal strength [8]. Or because the signal collides with the object during the propagation process, such as reflected, refracted, and diffracted, the signal strength is weakened. The Signal-to-Noise Ratio (SNR or S/N) is usually used to compare signal strength and background noise strength. It is defined as signal power and the ratio of noise power [9]. There are related research results between SNR and device identification in [10]. When there is relative motion between the signal source and the receiver, the wave path-difference is generated due to the change of the propagation path. The frequency of the transmitted signal is inconsistent with the frequency of the received signal. This phenomenon is called the Doppler effect, and the deviation between the transmitted frequency and the received frequency is called the Doppler shift [11].

2.2. RF fingerprint

RF fingerprint is like that in a conversation between people, the listener can identify a speaker by inherent variations and characteristics of the voice. RF fingerprint can automatically identify different wireless devices in the field by extracted the time domain and frequency domain properties of the signal during operation. The following paragraph will introduce which signal features are extracted and what conclusions are reached from the existing literature.

According to [12], they use a SDR platform for RF fingerprint extraction of Wi-Fi devices. The main extracted features are Scrambling Seed, sampling frequency offset, carrier frequency offset, and Frame Transient. The conclusion of the paper says that the results indicate that identifying Wi-Fi devices is possible (the accuracy rate of identification spanning 44%-50%). And [13] used the SDR platform to perform RF fingerprint extraction on ZigBee devices. The main extracted features are differential constellation trace figure (DCTF), carrier frequency offset (CFO), modulation offset, and I-Q offset. The paper says that the features remain stable over a long time. That is to say, these features can be long-lasting and difficult to change, so this phenomenon can be effectively regarded as a feature of the device. The power spectral density coefficients used in [14] that considered as a signal feature. The conclusion of the paper points out that the accuracy rate of identification is closely related to the receiver, and the high-end receivers will have better results. In addition, The power spectral density coefficients as the signal characteristic and analyzed the effect of SNR on the accuracy rate of identification in [10]. The paper pointed out that RF fingerprint would be related to the receiver used that may affect the accuracy rate of identification. At last, [15] used PSD as RF fingerprint for device identification, but the paper also explores the different distances between Tx and Rx and the effects of line-of-sight and non-line-of-sight on identification. The conclusion is that the identification performance will be worsened due to the increase of the distance, the main reason is the influence of multipath channel. While [16], [17] and [18] focus on the calculation of CFOs using a combination with different preambles.

2.3. Machine Learning

To put it simply, machine learning is defined an objective function about data. Then, when learning by the algorithm of the training model on the machine, the function is continuously

optimized during the training process to achieve the objective function and improve the performance of the algorithm. The reason why machine learning will be used is mainly because some data cannot be discriminated and classified manually. It can rely on automatic learning to obtain the characteristics of the data. Besides, because of a large amount of data analysis and statistics, the results have certain reliability. For example, [10] used the Multi-Layer Perceptron (MLP) neural network for identification. And [13] used the K-means clustering method for classification. Of course, there are other papers that use different classification models for analysis based on the purpose of the experiment. It is said that ML is indeed a reliable data analysis tool widely used.

3. METHOD

3.1. System Architecture

The architecture of the RF fingerprint system is shown in Figure 3:

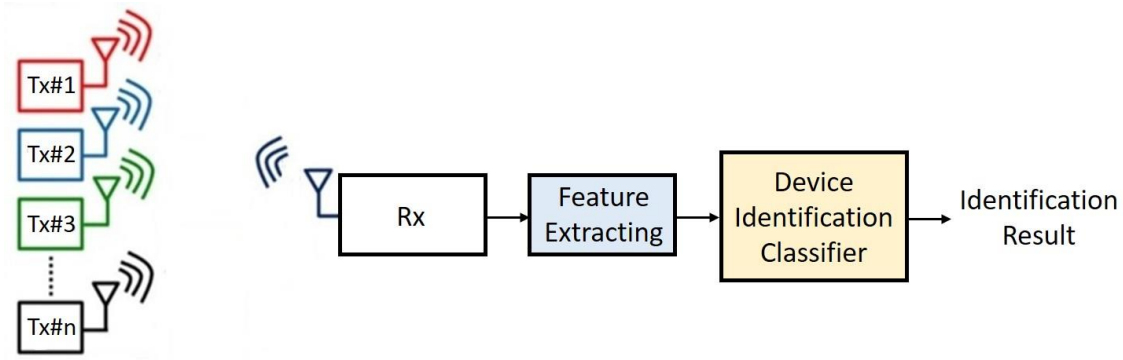

Figure 3. The architecture of the RF fingerprint system

We used the USRP B210 Software Defined Radio (SDR) Kit to receive signals transmitted by the wireless devices at the receiving end. After extracting multiple features from the received signal frames, these features are processed in the Device Identification Module. Finally, the classifier will give an identification result.

The RF fingerprint system can be used as a stand-alone physical-layer security, or for multi-factor authentication combined with other layers in the Open System Interconnection (OSI) model for better security. Additionally, it does not require additional feature extraction hardware. This allows the system can be built at low cost but robust.

3.2. Feature Extracting

According to the IEEE 802 standard, 802.11a/g uses Orthogonal Frequency-Division Multiplexing (OFDM) technology as a modulation technology for Wireless LAN (WLAN) systems [17]. This paper takes 802.11a/g as an example to further explore the results of RF fingerprint and device identification generated by related equipment.

The frame structure of the IEEE 802.11a/g standard is shown in the following figure 4:

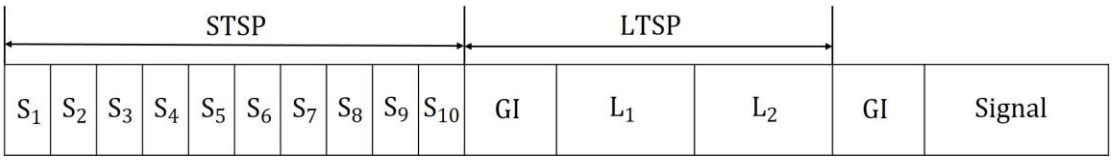

Figure 4. The frame structure of the IEEE 802.11a/g standard

GI refers to the guard interval, the function is to avoid interference between different blocks. The preamble is divided into two parts: STSP is Short Training Sequence Preamble and LTSP is Long Training Sequence Preamble [17]. The function of the preamble is to enable the receiving end to detect the starting position of the frame from the received signal, thereby deciphering the data bits.

The state of the received signal is related to the extracted features. We extracted features from modulation-based received signals and transient-based received signals, respectively. The detailed feature extraction method will be explained in the following.

3.2.1. Modulation-Based

First, we extracted the features from the received frequency-domain signal. If the up-conversion of the signal in transmitter (Tx) and the down-conversion of the signal in receiver (Rx) are inconsistent, in other words, the carrier frequency is not synchronized, it will cause carrier frequency offset (CFO). And then that will cause Inter-Carrier Interference (ICI) effect, which affects the performance of the OFDM system. If the sampling rate between the RF front ends of Tx and Rx is not synchronized, the sampling frequency offset (SFO) is caused.

The CFO is usually calculated and compensated by the symbols of STSP and LTSP to help the system to synchronize. If the system cannot achieve synchronization, the received signal may not be demodulated subsequently. Because the signal may be affected by multipath effect or delay during transmission, S_1 and S_2 are susceptible to interference from the delayed signal. Therefore, it is not recommended to include S_1 and S_2 in the calculation to estimate a better compensation.

Regarding the calculation method of CFO, this paper uses the Moose algorithm [19] to calculate the principle based on the periodicity of the training sequence. In the 802.11a/g system, the STSP symbol with two adjacent length is N_t, the relationship between the n-th sample of the previous group and the (n+N_t)-th sample of the latter group in the time domain and the frequency domain. As shown in the following formula, where the CFO is represented by ϵ

$$\mathrm{y}[\mathrm{n}+N_t]=\mathrm{y}[\mathrm{n}]e^{\frac{j2\pi N_t\epsilon}{N_t}} \overset{F.T}{\longrightarrow} Y[n+N_t]=Y[n]e^{2\pi\epsilon} \tag{1}$$

So the CFO estimated in the frequency domain

$$\epsilon=\frac{1}{2\pi}\angle\left(\frac{\sum_{n=0}^{N_t-1} I_m\{y_1^*[n]y_2[n+N_t]\}}{\sum_{n=0}^{N_t-1} R_e\{y_1^*[n]y_2[n+N_t]\}}\right) \tag{2}$$

Although the CFO has been calculated and compensated in the receiver, in order to obtain a more accurate CFO, we will calculate the residual CFO by LTSP. Then combined the two to estimate the CFO of the OFDM system. In addition, CFO may be time-varying, so it must be tracked continuously.

As for the calculation of SFO, it is estimated at the receiving end by using the sliding window method [20] to find the beginning of the data symbol, which is expressed mathematically as

$$\delta = arg\ \min \sum_{i=\delta}^{N_t-1+\delta} J_{SFO} \tag{3}$$

Where J_{SFO} is the cost function of the estimated SFO, $J_{SFO} = |y[n+i] - y[n+N+i]|$.

The amplitude and the phase imbalances are represented by A and φ, respectively. The outputs of the in-phase and the quadrature paths are denoted as $y_I(t)$ and $y_Q(t)$, respectively. If $y(t)$ is the ideal reception signal, the received signal affected by I-Q Imbalance is

$$\begin{aligned} \hat{y}(t) &= y_I(t) + y_Q(t) \\ &= R_e\{y(t)\} + jI_m\{Ae^{i\varphi}y(t)\} \end{aligned} \tag{4}$$

If implemented it in SDR, the in-phase and quadrature signals of the baseband are sent to the computer for calculation. Ideally, the in-phase and quadrature are $y_I(t) = \cos(\omega_0 t)$ and $y_Q(t) = \sin(\omega_0 t)$ respectively. Where ω_0 is the baseband signal. After the RF signal is down-converted to the baseband, the baseband signal affected by I-Q Imbalance [21] is

$$\begin{aligned} \widehat{y_I}(t) &= \alpha\cos(\omega_0 t) + \widehat{\beta_I} \\ \widehat{y_Q}(t) &= \sin(\omega_0 t + \varphi) + \widehat{\beta_Q} \end{aligned} \tag{5}$$

Where $\alpha = 1/A$ and φ are the amplitude and phase errors caused by the aforementioned I-Q Imbalance. $\widehat{\beta_I}$ and $\widehat{\beta_Q}$ are the DC bias of the residual in-phase and the quadrature path after down-converting, respectively. After deducting the corresponding DC bias estimator from the in-phase and the quadrature signal, then substituting by $\sin(\omega_0 t + \varphi) = \sin(\omega_0 t)\cos(\varphi) + \cos(\omega_0 t)\sin(\varphi)$, the baseband signal has the following matrix form

$$\begin{bmatrix} \widehat{y_I}(t) \\ \widehat{y_Q}(t) \end{bmatrix} = \begin{bmatrix} \alpha & 0 \\ \sin(\varphi) & \cos(\varphi) \end{bmatrix} \begin{bmatrix} y_I(t) \\ y_Q(t) \end{bmatrix} \tag{6}$$

The amplitude offset α and the phase offset φ can be calculated as follows

$$\begin{aligned} < y_I(t) \cdot y_I(t) > &= \alpha^2 < \cos^2(\omega_0 t) > = \frac{\alpha^2}{2} \\ \rightarrow \alpha &= \sqrt{2\langle y_I(t) \cdot y_Q(t) \rangle} \\ < y_I(t) \cdot y_Q(t) > &= \frac{\alpha^2}{2}\sin(\varphi) \end{aligned} \tag{7}$$

$$\rightarrow \varphi = \sin^{-1}\left((\alpha^2/2)\langle y_I(t) \cdot y_Q(t)\rangle\right) \quad (8)$$

In summary, the features that we extracted from modulation-based signal are CFO, SFO, amplitude offset, and phase offset.

3.2.2. Transient-Based

We also extract the LTSP from the received time-domain signal, and then calculate the power spectral density (PSD) after Fast Fourier Transform (FFT) and take nature log of these data [22]. We regard Logarithmic PSD as an RF fingerprint. Assuming LTSP is $x(m)$, after 64-FFT conversion, $X(k)$ is *k*-th discrete Fourier coefficient of signal $x(m)$ that can be obtained as follows

$$X(k) = \frac{1}{64}\sum_{m=1}^{64} x(m) exp\left[\frac{-2\pi j}{64}(m-1)(k-1)\right] \quad (9)$$

Then, we can calculate the Logarithmic PSD as the following mathematical formula

$$\Psi(k) = 10 \cdot log_{10}|X(k)|^2 \quad (10)$$

3.3. Classifier

In the software library, there is a tree-based tool called XGBoost [23]. It is a powerful classifier formed by assembling many decision tree models, supported in many programming languages and operating systems. Besides, it uses a number of ways to prevent overfitting when classifying, and supporting parallel computing [24]. Therefore, it is widely used in various fields, such as research competition and industry.

3.4. Evaluation

In order to verify the performance of the RF fingerprint system, we must rely on a reliable method for data analysis. As for how to evaluate the performance of the trained classification model, it is generally used as a performance index with verification indexes. According to our experiments, the main purpose is device identification. Therefore, we used classification metrics as our performance evaluation. Classification can be divided into binary case and multiclass case. The confusion matrix [25] is a table that is often used to show the performance of a classifier on a set of validate data for predicted results. Finally, we can calculate the performance index from the confusion matrix. The performance index we used is the accuracy rate, which represents the proportion of data that our classifier can correctly classify.

3.5. Experiments Setup

In the experimental process, we have prepared two computers, a wireless access point (AP) and a receiver. One of the computers will be connected to the Wi-Fi device, then transmitting signals after associating to the AP, which was regarded as the transmitting end. The other computer was connected to the receiver USRP B210 as the SDR platform. After receiving the signal, the SDR platform can extract RF fingerprint by the algorithm such as carrier frequency offset (CFO), sampling frequency offset (SFO), amplitude offset, phase offset, and power spectral density (PSD) from the signal frame.

In order to verify the feasibility of device identification by RF fingerprint, we carried out experiments with 9 Wi-Fi devices, including three brands: ASUS, Panda, and TOTO-Link, each of which contains 3 devices of the same model. We collected similar numbers of features and designed three types of experiments to identify devices by XGBoost. The detailed description of the experiments are as follows:

- Experiment A: To verify that different brands of Wi-Fi devices will produce different RF fingerprint. We obtained data by transmitting and receiving pairings between 9 Wi-Fi devices and a fixed Rx. We trained the multiclass classification model to try to classify and observe whether 9 pairs can be effectively classified.
- Experiment B: To probe if PSD can be considered as an RF fingerprint or not. We selected one of the wireless devices of three different brands, and extracted the PSD value of LTSP from each received signal frame. Then these data we were classified by training the multiclass classification model to analyze the identification performance.
- Experiment C: Increasing Rxs as a variation factor and testing whether different Rx would affect RF fingerprint. We obtained data from 9 Wi-Fi devices corresponding to 3 different Rxs. We tried to explore if RF fingerprint received by different Rxs is similar or not. For example, if we used the 9 sets of paired data transmitted and received by Rx #1 to train the multiclass classification model, 9 pairs of paired data of Rx #2 or Rx #3 can be classified or not.

4. EXPERIMENT AND RESULT

The experimental process is shown in Figure 5. After continuously collecting tens of thousands of data included five kinds of features in the shielding box and the SNR is about 20dB, the features were input into the classifier for classification and identification. We used XGBoost as classifier, the two main parameters are the depth and the n_estimators of Decision Tree, with values of 3 and 300, respectively. The experimental results were showed in following subsections.

4.1. Experiment A

Table 1. Confusion matrix

	A1	**A2**	**A3**	**P1**	**P2**	**P3**	**T1**	**T2**	**T3**
A1	1720	1	0	0	0	0	0	0	0
A2	1	1653	0	0	0	0	0	0	0
A3	0	0	1498	0	0	0	0	0	0
P1	0	0	0	2214	0	0	0	0	0
P2	0	0	0	0	2304	0	0	0	0
P3	0	0	0	1	0	2117	0	0	0
T1	0	0	0	0	0	0	2428	0	0
T2	0	0	0	0	0	0	0	2306	1
T3	0	0	0	0	0	0	0	1	1488

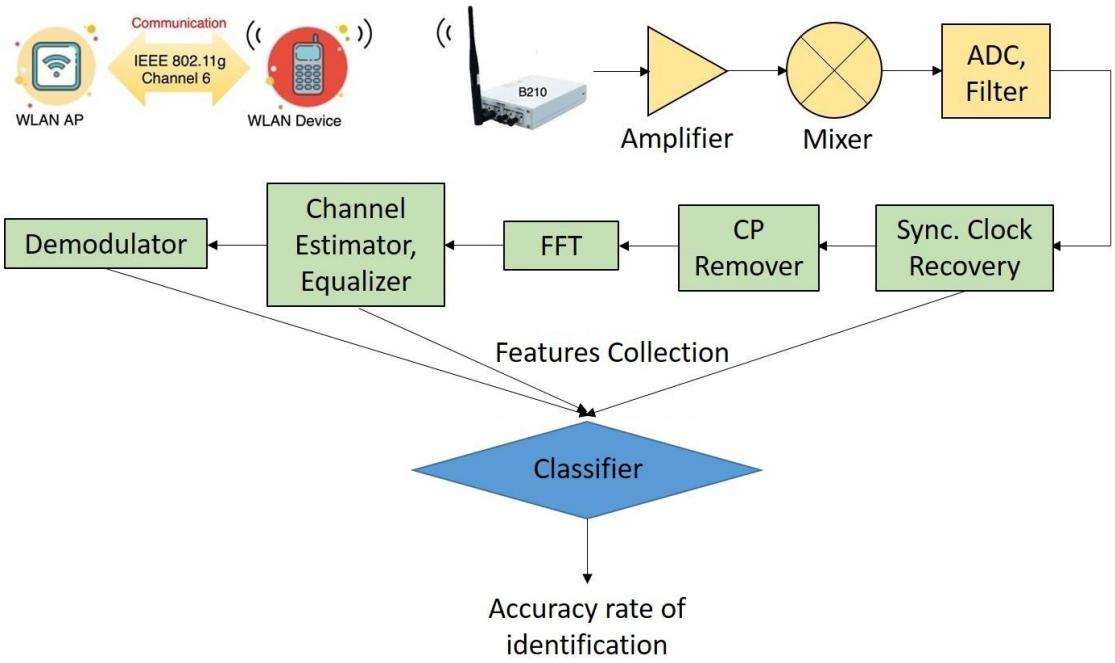

Figure 5. Experimental process

In experiment A, in order to verify that different brands of Wi-Fi devices will produce different RF fingerprint, we obtained features by transmitting and receiving pairings between 9 Wi-Fi devices and a fixed receiver in the shielding box. After confirming that settings were ready, we could start experimenting. Totally we gathered 177,253 samples. 90% of the samples were used for training the multiclass classification model, and 10% of the samples were used for verification. After training and testing, we could get the accuracy rate of identification is 99.97% (Table 1). This shows that we can effectively classify RF fingerprint caused by different transmitters.

4.2. Experiment B

In experiment B, we did an experiment with power spectral density (PSD) to probe if PSD can be considered as an RF fingerprint. We selected one of the wireless devices of three different brands. After the PSD values of the LTSP in each received signal frame were extracted through the above experimental process. The drawing of datasets was as shown in Figure 6.

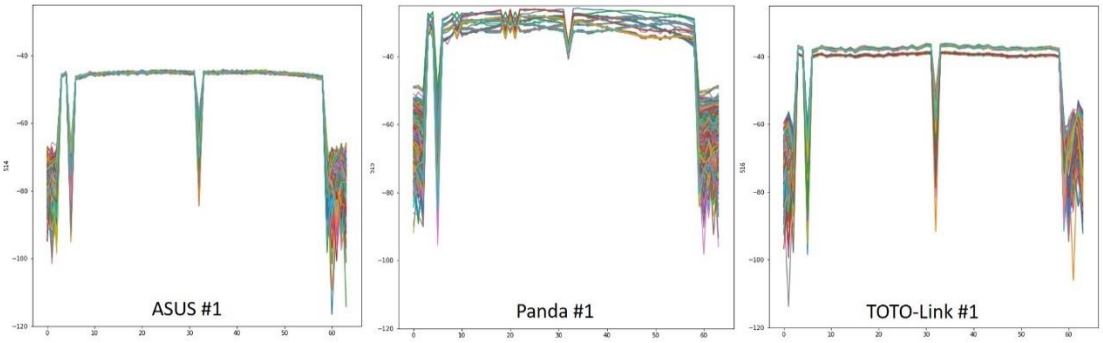

Figure 6. The PSD caused by different Wi-Fi devices

From the results, we found that the three PSDs were significantly different under the same experimental conditions. Therefore, PSD does have the opportunity to be considered an RF fingerprint.

Next, under the same process and conditions, we performed PSD experiments on 9 Wi-Fi wireless devices. The datasets were classified using a classifier to analyze the identification performance. The results are as Table 2:

Table 2. Confusion matrix

	A1	**A2**	**A3**	**P1**	**P2**	**P3**	**T1**	**T2**	**T3**
A1	1713	8	0	0	0	0	0	0	0
A2	0	1653	0	0	0	0	0	0	0
A3	0	0	1498	0	0	0	0	0	0
P1	0	0	0	2214	0	0	0	0	0
P2	0	0	0	0	2304	0	0	0	0
P3	0	0	0	1	0	2117	0	0	0
T1	0	0	0	0	0	0	2428	0	0
T2	0	0	0	0	0	0	0	2307	0
T3	0	0	0	0	0	0	0	1	1488

The accuracy rate of identification is 99.94%. According to the results, if PSD was only used as an RF fingerprint, it could be performed effectively to device identification.

4.3. Experiment C

In experiment C, in order to further verify if different Rx is one of the factors affecting pairing RF fingerprint, we increased the Rx as a variation factor and testing whether different Rx would affect RF fingerprint. We obtained features after transmitting and receiving pairs of 9 Wi-Fi devices and 3 Rxs of the same brand and the same model. Totally we gathered 503,227 samples. 90% of the samples were used for training the multiclass classification model, and 10% of the samples were used for verification. Then we fellow the two steps below to gradually confirm the experimental goal.

First, we need to evaluate the accuracy rate of the model that trained by the dataset received by a certain Rx. So we trained the datasets received by the 3 Rxs respectively and Table 3 showed the results.

Table 3. Performance of the device identification

	The accuracy rate of identification
Rx #1 \ Dataset #1 \ Model #1	99.97%
Rx #2 \ Dataset #2 \ Model #2	99.97%
Rx #3 \ Dataset #3 \ Model #3	99.95%

Second, we used datasets collected from different receivers to validate the models' performance. For example, we used the dataset received by Rx #1, and let it to train the multiclass classification model. In step one, we could get the accuracy of identification from Dataset #1 to Model #1 is 99.97%. But when we validate the classifier with the dataset from Rx #2 and Rx #3, the accuracy rate of identification significant drop off to 69.17% and 36.78%, respectively. Therefore, the fingerprint model training with dataset form Rx #1 could not effectively identify the samples from Rx #2 or Rx #3. It could be seen that samples from different Rx did affect RF

fingerprint model. The same conclusion was obtained when the same experiment was repeated for Rx #2 and Rx #3. All performances of the device identification were organized in Table 4

Table 4. Performance of the device identification

	Dataset #1 \ Rx #1	**Dataset #2 \ Rx #2**	**Dataset #3 \ Rx #3**
Model #1 \ Dataset #1	99.97%	**69.17%**	**36.78%**
Model #2 \ Dataset #2	**59.97%**	99.97%	**71.04%**
Model #3 \ Dataset #3	**59.91%**	**74.19%**	99.95%

Based on the results, we found that for the same Tx, RF fingerprint with different Rx generating pairs were significantly different. The conclusion is that the existence of RF fingerprint is relative to Tx and Rx.

Parameter comparison in the confusion matrix can be referred to Table 5

Table 5. Parameter Comparison Table

A1 : ASUS #1	T1 : TOTO-Link #1
A2 : ASUS #2	T2 : TOTO-Link #2
A3 : ASUS #3	T3 : TOTO-Link #3
P1 : Panda #1	Rx1 : Receiver #1
P2 : Panda #2	Rx2 : Receiver #2
P3 : Panda #3	Rx3 : Receiver #3

5. CONCLUSION

In this paper, we introduced the concept about RF fingerprint, and then analyze and did experiments to discuss the feasibility of using RF fingerprint for device identification in the IoT network. We implemented a low-cost SDR platform to measure RF signals transmitted by Wi-Fi devices, and extracted RF fingerprint from signals. Then we used these features to distinguish 9 transmitters with machine learning model as classifier. The accuracy rate of identification is 99.97%. Besides, we use only power spectral density as an RF fingerprint to identify wireless devices. The accuracy rate of identification is 99.94%. Finally, we regarded the receiver as a factor affecting the RF fingerprint, and explored whether the RF fingerprint received by one receiver can be used and compared to another receiver. The results showed that RF fingerprint is relative to transmitter and receiver. It indicates that the RF fingerprint cannot be directly shared between different receivers. By developing RF fingerprint system in physical layer, if the security mechanism of other layers in the OSI model are combined, the information security of the user can be effectively improved. In the future, we will continue to study the transferability of the receiver, and try to resolve the relativity of RF fingerprint existed in the transmitter and receiver to make the RF fingerprint system more widely applicable to the deployment in actual scenes.

REFERENCES

[1] S. Singh and N. Singh, "Internet of Things (IoT): Security Challenges, Business Opportunities & Reference Architecture for E-commerce," *Green Computing and Internet of Thing (ICGCIoT), 2015 International Conference on*, 2015.

[2] M. A. Muhal, X. Luo, Z. Mahmood and A. Ullah, "Physical Unclonable Function Based Authentication Scheme for Smart Devices in Internet of Things," *2018 IEEE International Conference on Smart Internet of Things (SmartIoT)*, 2018.

[3] T. D. Vo-Huu, T. D. Vo-Huu and G. Noubir, "SWiFi: An Open Source SDR for Wi-Fi Networks High Order Modulation Analysis," 2008.
[4] Z. Zhuang, X. Ji and Y. Liu, "FBSleuth: Fake Base Station Forensics via Radio Frequency Fingerprinting," *AsiaCCS 2018*, 2018.
[5] DC bias (https://en.wikipedia.org/wiki/DC_bias).
[6] B. Chatterjee, D. Das, S. Maity and S. Sen, "RF-PUF: Enhancing IoT Security through Authentication of Wireless Nodes using In-situ Machine Learning," *IEEE Internet of Things Journal*, 2018.
[7] Carrier frequency offset (https://en.wikipedia.org/wiki/Carrier_frequency_offset)
[8] Attenuation (https://baike.baidu.com/item/%E8%A1%B0%E5%87%8F)
[9] Signal-to-noise ratio (https://en.wikipedia.org/wiki/Signal-to-noise_ratio)
[10] S. U. Rehman, K. W. Sowerby, S. Alam and I. Ardekani, "Portability of an RF Fingerprint of a Wireless Transmitter," *2014 IEEE Conference on Communications and Network Security*, 2014.
[11] Doppler effect (https://en.wikipedia.org/wiki/Doppler_effect)
[12] T. D. Vo-Huu, T. D. Vo-Huu and G. Noubir, "Fingerprinting Wi-Fi Devices Using Software Defined Radios," 2016.
[13] L. Peng, A. Hu, J. Zhang, Y. Jiang, J. Yu and Y. Yan, "Design of a Hybrid RF Fingerprint Extraction and Device Classification Scheme," *IEEE Internet of Things Journal*, 2018.
[14] S. U. Rehman, K. W. Sowerby and C. Coghill, "Analysis of impersonation attacks on systems using RF fingerprinting and low-end receivers," *Journal of Computer and System Sciences*, 2013.
[15] W. Wang, Z. Sun, S. Piao, B. Zhu and K. Ren, "Wireless Physical-Layer Identification: Modeling and Validation," *IEEE Transactions on Information Forensics and Security*, 2016.
[16] Y. Zhuang and Y. Wan, "LS-based Joint Estimation of Carrier Frequency Offset, I/Q Imbalance and DC Offset for OFDM-based WLANs," *Proceedings 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC)*, 2013.
[17] H. Zou and Y. Wan, "A Novel Subspace-Based Carrier Frequency Offset Estimator for OFDM-Based WLANs With DC Offset," *2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*, 2012.
[18] Qi Cheng, "Joint Estimation of Carrier and Sampling Frequency Offsets Using OFDM WLAN Preamble," *2015 15th International Symposium on Communications and Information Technologies (ISCIT)*, 2015.
[19] P. H. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Trans. Comm., vol. 42, Issue 10, pp. 2908-2914*, Oct. 1994.
[20] Z. Zhang, L. Ge, F. Tian, F. Zeng and G. Xuan, "Effects and Estimation Techniques of Symbol Time Offset and Carrier Frequency Offset in OFDM System: Simulation and Analysis," *2014 7th International Congress on Image and Signal Processing*, 2014.
[21] S. Ellingson, "Correcting I-Q imbalance in direct conversion receivers," http://argus. naapo.org/~rchilders/swe_argus_pubs/iqbal.pdf, 2003.
[22] T. Ohtsuji, T. Takeuchi, T. Soma and M. Kitsunezuka, "Noise-tolerant, Deep-learning-based Radio Identification with Logarithmic Power Spectrum," *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, 2019.
[23] XGBoost (https://zh.wikipedia.org/wiki/XGBoost)
[24] The principle of XGBoost (https://kknews.cc/zh-tw/news/grejk5m.html)
[25] Evaluation (https://medium.com/@chih.sheng.huang821/machine_learning-statistical_methods-model_evaluation-verification index-b03825ff0814.)

AUTHORS

Ting-Yu Lin, Engineer

A graduate of the Communications Engineering at Yuan Ze University in Taiwan. Research interests include cyber security, communications engineering, etc. Recent research works are mainly focusing on wireless devices identification and security by RF fingerprint.

Chia-Min, Sena, Lai, Engineer

A Ph.D. candidate at the Department of Computer Science and Information Engineering at National Taiwan University of Science and Technology in Taiwan. Research interests include signal fingerprint, network log analysis, malware analysis, artificial intelligence. Recent research works are mainly focusing on using deep learning techniques to explore the security issues.

Chih-Wei Chen, Engineer

A graduate of the Institute of Electronics at National Chiao Tung University in Taiwan. Research interests include (elliptic curve) cyber security, malware analysis, Reverse engineering, etc. Recent research works are mainly focusing on radio frequency identification and security.