Random Matrix Theory Based Radio Frequency Fingerprinting Identification of WiFi Signal

You Jiang*, Jian Xiong*, *Senior Member, IEEE*, Junqi Wu‡, Bo Liu§, *Senior Member, IEEE*

*Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China
‡Student Innovation Center, Shanghai Jiao Tong University, Shanghai, China
§School of Computer Science, University of Technology Sydney, Ultimo, NSW, Australia
*Email: {2022jy, xjarrow}@sjtu.edu.cn

***Abstract*—This paper introduces a method of Radio Frequency Fingerprint (RFF) recognition based on Random Matrix Theory (RMT). RFF identification is a physical layer method to identify devices that exchanging messages in the network. The extracted subtle features of the transmitters working as the biometric fingerprint can uniquely represent the identity of the devices. However, the differences in RFF become extremely subtle when transmitters are the same type of industrial standard equipment and it is very difficult to identify the devices under this scenario. In the proposed method, a random matrix of the I/Q (In-phase/Quadrature) data of bit-similar Universal Software Radio Peripheral (USRP) devices is constructed firstly, and the extracted features are sent to the Convolutional Neural Networks (CNN) network for recognition. Experiments are carried out under Signal-to-Noise Ratio (SNR)∈[-10,20]dB between different datasets. It demonstrates a 0.98 accuracy in identification, outweighting the traditional classification method.**

***Index Terms*—RF fingerprint, Random Matrix Theory, Convolutional Neural Network, Raw I/Q**

I. INTRODUCTION

Radio frequency fingerprint (RFF) is a promising method for physical layer security in wireless networks. RFF is based on the use of hardware originated imperfections of the transmitting devices. Due to the inherent difficulty in replicating the hardware features of devices, RFF offers high levels of uniqueness and security. It can be extracted either from the transient or the steady state regions of the transmitted signal. In practical communication networks, user anomalous behavior may be manifested through information such as spatial, temporal, and frequency characteristics of signals emitted by devices. Research on machine learning-based RFF analysis and recognition techniques can identify the source of interference, providing essential grounds for further interference identification, avoidance, and suppression strategies. RFF recognition utilizes waveform-level defects imposed by Radio Frequency (RF) circuits to obtain a "fingerprint" of wireless devices. And this is hard for adversarial devices to imitate. Analyzing RFF assists in the identification of both legal and illegal devices, offering a means of physical layer authentication. However, industrial and aerospace devices often exhibit good RF characteristics, resulting in minimal RF differences between devices and adding significant complexity to RFF recognition.

This work was supported in part by the National Natural Science Foundation of China under Grant 62071294; in part by the Shanghai Key Laboratory of Digital Media Processing and Transmission; in part by the 111 Project under Grant B07022.

While deep learning technology has made significant advancements, researchers have also introduced it to the RF fingerprint recognition field [1], [2]. The methods of RFF identification using deep learning methods can be roughly divided into two categories: those based on expert features and deep learning, and those based on I/Q data and deep learning. These two approaches are illustrated in Fig. 1. In the Feature Extraction-based method: Raw I/Q (In-phase/Quadrature) data is firstly sent into the Pre-processing module. Then feature extraction module will extract the RFF of the device. Finally, the identification results will be given out by Deep Learning Method module. In the Raw I/Q-based method, feature extraction is omitted compared with method one. Processed I/Q data will be directly fed into a deep learning neural network for learning and classification.

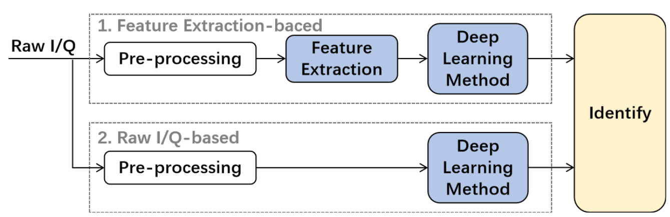


Fig. 1: RF fingerprint recognition methods based on deep learning

In method one, the Feature Extraction module consists of commonly used feature extraction methods like Continuous Wavelet Transform (CWT), Hilbert-Huang Transform (HHT) [3], Differential Constellation Trace Figure (DCTF) [4], and Short-Time Fourier Transform (STFT) spectrograms [5]. When prior information about signal protocols is known,we can design artificial features based on specific domain knowledge to maximize individual differences. This often achieves better recognition accuracy in specific scenario. However, for non-cooperative devices, preprocessing operations based on prior knowledge about the signal such as carrier frequency offset compensation, phase compensation, and time synchronization can barely be carried out. The preprocessing employed by Shen et al. [6] and Gritsenko et al. [7], are not applicable. In such cases, almost only raw I/Q data are available when prior knowledge is unaccessible.

The deep learning modules in both methods primarily utilize various common neural network models, such as Convolutional Neural Networks (CNN), RNN(Recurrent Neural Network), Transformer, FCNN(Fully Convolutional Networks), etc. CNN models are the most widely and early applied neural network structures in RFF recognition. Researchers have made improvements by incorporating techniques such as adding residual networks [8], attention mechanisms [9], etc. In [9], data augmentation is applied to enhance data richness, achieving good recognition results. In most cases, method two has higher requirements for network modeling than method one.

This paper proposes a feature extraction method based on random matrices to deal with Raw I/Q data. First, a random matrix is constructed for the input IQ data. Then, the RFF is generated by random matrix. Finally, a concatenated network of CNN and MLP(Multilayer Perceptron) is employed for device classification. To the best of our knowledge, it is the first time to introduce random matrix theory into the RF fingerprint recognition field.

The remainder of the article is organized as follows.The construction of the I/Q matrix based on Random Matrix Theory will be introduced in Section II. The network and experiment design as well as simulation results are presented in Section III, and the the paper is summarized in Section IV.

II. METHODOLOGY

In this section, the relevant background knowledge of Random Matrix Theory will be firstly introduced in II-A, then the way how to construct the Wishart-Laguerre matrix of I/Q signal will be shown in II-B.

A. Introduction of Random Matrix Theory

Random Matrix Theory (RMT) is originated from the statistical mechanics analysis in physics. Its essence lies in the analysis of matrix eigenvalue spectra constructed through random variables. From a high-dimensional perspective, it reflects the overall characteristics of a large number of random variables. The Wishart-Laguerre matrix(W-L for short) is a type of random matrix introduced in the context of RMT [10]. Specifically, it is a member of the Wishart ensemble of random matrices, which are used to model the covariance matrices of random vectors. Symbols appear in the following formula are explained in TableI. The joint distribution of W-L matrix is given by equation (1):

$$\rho(x_1, x_2, ..., x_j) = \frac{1}{Z_{N,\beta}} e^{-\frac{1}{2}\sum_{i=1}^{N} x_i^2} \prod_{j<k} |x_j - x_k|^\beta \quad (1)$$

where

$$Z_{N,\beta} = (2\pi)^{\frac{N}{2}} \prod_{j=1}^{N} \frac{\Gamma(1 + j\beta/2)}{\Gamma(1 + \beta/2)} \quad (2)$$

is a normalization constant, enforcing $\int_{\mathbb{R}^N} dx\rho(x_1, ..., x_N) = 1$ and $\beta = 1, 2, 4$ is called the *Dyson index*. Henceforth, $dx = \prod_{j=1}^{N} dx_j$. Note that the eigenvalues are considered to be unordered here. However, when N and M tend to positive infinity, the edge distribution of its eigenvalues will conforms to the description of equations (3) and (4):

$$\rho(x) = \frac{1}{\beta N} \rho_{MP}(\frac{x}{\beta N}) \quad (3)$$

$$\rho_{MP}(y) = \frac{1}{2\pi y} \sqrt{(y - \zeta_-)(\zeta_+ - y)} \quad (4)$$

Such a distribution also can be observed when N, M is large enough. On the whole, this equation can reflect the eigenvalue distribution of the W-L matrix constructed with uncorrelated variables. $\zeta_- = (1 - c^{-\frac{1}{2}})^2$ and $\zeta_+ = (1 + c^{-\frac{1}{2}})^2$ in equation (3) and (4) are respectively the soft lower bound and soft upper bound of the eigenvalues of an uncorrelated random matrix, where $c = N/M$. Researches in [11], [12] show that the distribution of eigenvalues conforms to certain rules and the distribution of W-L matrix can reflect the relevance of the input data. If the elements of the W-L matrix are the I/Q data of transmitter devices, the characteristics extracted by the W-L matrix can reflect the characteristics of the transmitters. And these characteristics can be viewed as the RFF of the transmitters.

TABLE I: Symbol Explanation

symbol	explanation
ρ	the edge distribution of eigenvalues
Γ	Gamma function, $\Gamma(z) = \int_0^\infty x^{z-1}e^{-x}dx$
i, j, k	positive integer:i,j,k=1,2,3,...
$Z_{N,\beta}$	normalization constant
M	rows of the matrix
N	columns of the matrix
c	define $c = N/M$
x_i	element in the matrix
β	Dyson index(electable value:1,2,4)
ζ_-	soft lower bound
ζ_+	soft upper bound
H	matrix composed of random variables
H^*	the conjugate transpose of H
I	In-phase signal
Q	Quadrature signal

B. The W-L matrix construction of I/Q signal

The whole process of generating RFF by can be described in Fig.2. The two crucial parts are the construction of W-L matrix and the generation of eigenvalue sequence.

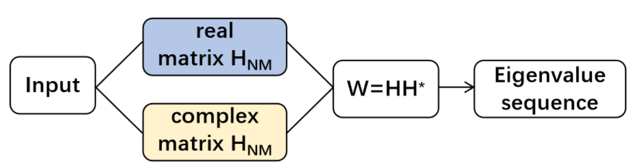


Fig. 2: W-L matrix construction frame

The W-L matrix can be constructed as follow:

$$W = HH^* \quad (5)$$

where $H_{N\times M}$ is a matrix composed of random variables, H^* is the conjugate transpose of H. It's easy to prove that $W_{N\times N}$ is a conjugate symmetric semi-definite matrix with N real eigenvalues.

There are two ways of constructing the H matrix:

(1) For a random variable X conforming to $N\ (0,1)$, the sample of X can form a real matrix H, and the W-L matrix obtained by the construction of real matrix H has a *Dyson index* $\beta = 1$.

(2) For independent samples of random variables X, Y conforming to $N\ (0,1)$, the complex matrix H can be formed with elements $H_{MN} = (X + jY)/2$. The W-L matrix obtained by the construction of complex matrix H has a Dyson exponent *Dyson index* $\beta = 2$.

From the above construction, it can be deduced that the element W_{ij} of the W matrix is actually the cross-correlation of the I_{th} row vector and the J_{th} row vector in the H matrix.

In the Eigenvalue sequence part, the N eigenvalues of matrix W are firstly calculated. It can be deduced from Equation (3) that the eigenvalue scale is proportional to the size of βN. In order to improve the robustness, the eigenvalues are firstly normalized to βN . Then the eigenvalue sequence will be obtained after all these operation.

C. W-L matrix of I/Q data

There is no requirement on which part of the transmission the I/Q data is taken from in this method. It can be the transient part, the steady part of the transmission, or the whole part of the transmission. Let $S = \{s_1, s_2, \ldots, s_K\}$ to be the collection of I/Q data of K transmitters. Since the length of the data entered into the H matrix is $N \times M$, the I/Q data of each device should be cut into segments with length of $N \times M$. The i_{th} segment of device k can be expressed as follow:

$$s_{k_i} = \begin{bmatrix} I_1 & I_2 & \cdots & I_{N\times M} \\ Q_1 & Q_2 & \cdots & Q_{N\times M} \end{bmatrix} \tag{6}$$

Since the second method of constructing W-L matrix can preserve the phase information to the greatest extend, a complex W-L matrix is constructed in this paper.Then the H matrix can be presented as follow:

$$H = \begin{bmatrix} I_{11} + jQ_{11} & I_{12} + jQ_{12} & \cdots & I_{1M} + jQ_{1M} \\ I_{21} + jQ_{21} & I_{22} + jQ_{22} & \cdots & I_{2M} + jQ_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ I_{N1} + jQ_{N1} & I_{N2} + jQ_{N2} & \cdots & I_{NM} + jQ_{NM} \end{bmatrix} \tag{7}$$

After the construction of H matrix of I/Q data, the W-L matrix can be easily constructed according to Equation (5). Then the eigenvalue of matrix W will be calculated and processed according to the instructions shown in II-B. The eigenvalues which can be seen as the RFF of the device can be finally obtained.

III. SIMULATION AND DISSCUSION

This section shows the simulation process, including the dataset used in this paper, CNN model as well as the experiment results.

A. Dataset

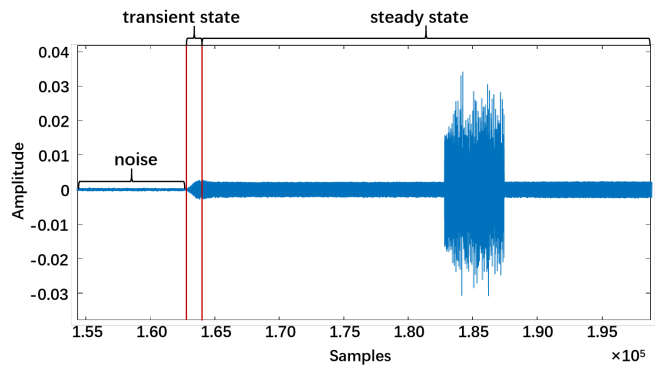


Fig. 3: Signal segment of one transmitter

These transmissions are received by a single NI USRP N210 receiver, which recorded the In-phase and Quadrature (I/Q) data of the transmitted signals. Each transmitter underwent five transmission sessions. The data is stored in cf32 type. Each sample ranges from 800 to 1000MB in size after transmitting into binary format. For this study, we select the first ten devices from the Arena Testbed "in the wild"/Wired dataset. An exceprt of the signal is presented in Fig. 3, including both noise segment and the transient as well as the steady-state portions of the signal. And the dataset of each transmitter emcompasses over 1500 such segments.

Two datasets are prepared, including different devices: Dataset1={device1,device2,device3,device6,device9}; Dataset2={device1,device2,device3,device4,device6}. These are all Raw I/Q data without any data augmentation. The setting of the Matrix parameters also affects the subsequent recognition accuracy, and the best parameter setting are obtained through many experiments.

TABLE II: Parameter Table

Parameter	Value
M	512
N	512
β	1

And the setting of the parameters to generate W-L matrix is shown in Table. II

B. CNN Model

A CNN model comprising 3 convolutional layers and 3 fully connected linear layers is constructed in this paper. The structure and detailed parameters of the CNN network are illustrated in Fig. 4. Each convolutional layer consists of a convolutional part followed by a max-pooling part, with LeakyReLU employed as the activation function. As depicted in the figure, the outputs from the LeakyReLU function are greater than or equal to 0.

The features extracted after the convolution process are then fed into the MLP. The MLP used in this experiment consists of three linear layers. Subsequently, classification results (representing the device name) are obtained after passing through the MLP.

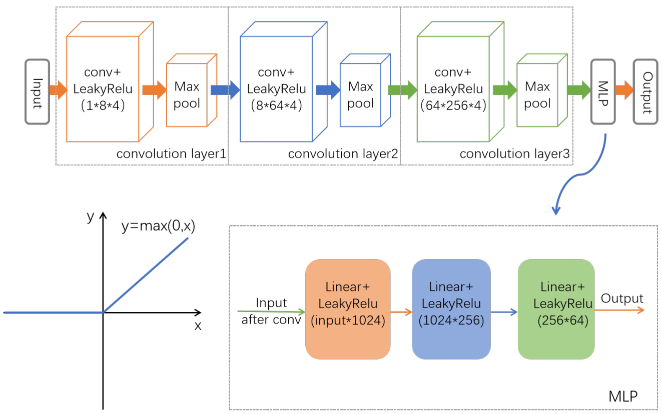


Fig. 4: CNN network frame

C. Experiments and Results

The effectiveness of the RMT in RFF identification has been validated in the following experiments.

1) Classification Accuracy between unprocessed data and processed data: Unprocessed data from Dataset1 is initially fed into our classifier. Then data processed through RMT method is put in the same classifier. The identification confusion matrices are shown in Fig. 5. The identification accuracy of Raw I/Q data without the processing of our method is 0.35 while the accuracy increases to 0.98 on Dataset1 and Dataset2 with the RMT method.

2) Classification Accuracy under different SNR: Gaussian white noise of different amplitude is added to each device in Dataset1, supposing s_i is the data of each device:

$$s_i \leftarrow \{s_i + n_1, s_i + n_2, \ldots, s_i + n_O\} \quad (8)$$

where n_i represents the Gaussian white noise. In this experiment, the $SNR_j = (s_i/n_i) \in \{-10, -8, -6, -4, -2, 0, 5, 10, 15, 20\}$dB. The transmissions are normalized to ensure their power equals 1, and then different levels of Gaussian white noise are added to the received I/Q data.

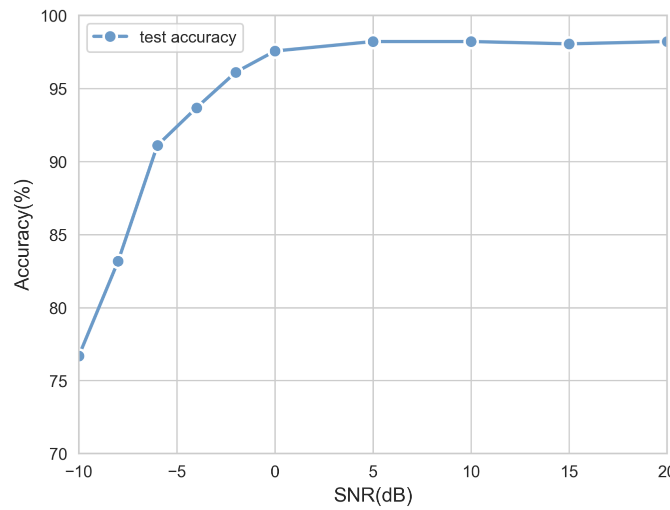


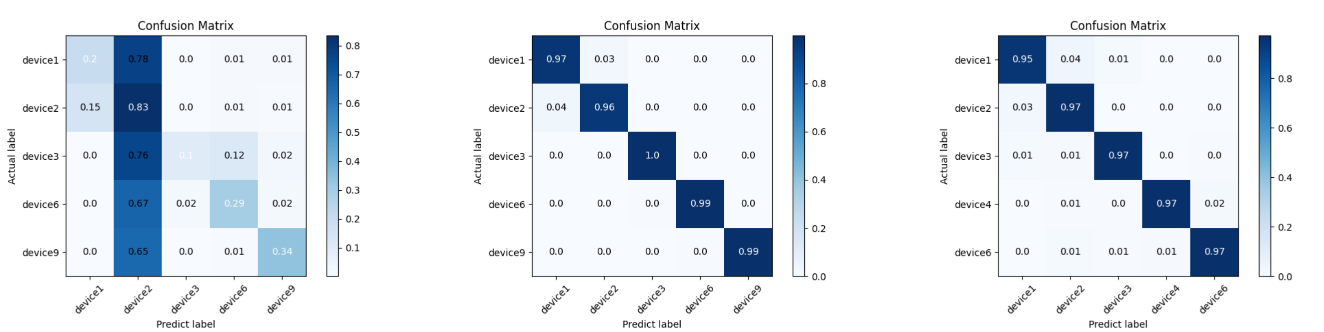
Fig. 6: Accuracy under different SNR

Fig. 6 shows the identification accuracy maintains over 0.9 when the SNR is higher than -4dB. It proofs that our method is robust to Gaussian white noise.

3) Classification Accuracy using Wsnet: This paper also use the multi-class support vector machine(SVM) as the classifier and the scattering-coefficient matrix S extracted by WSNet as the feature for classification. [13] Both methods are implemented when SNR in the range of -10dB to 20dB. Table III shows that both methods have relatively good accuracy in the case of low SNR. The method proposed in this paper demonstrates a slight superiority in scenarios with medium to high SNR, achieving an average accuracy of around 98% when SNR is higher than 5dB .

TABLE III: Classification accuracy of RMT and WSNet

Method	SNR Level		
	-10∼0dB	1∼10dB	11∼20dB
Proposed method(RMT)	89.7%	98.2%	98.1%
Scattering-coefficient	84.9%	86.4%	89.5%



(a) Accuracy of Dataset1 without RMT (b) Accuracy of Dataset1 with RMT (c) Accuracy of Dataset2 with RMT

Fig. 5: Illustration of identification confusion matrix

IV. CONCLUSIONS

This paper presents a novel RF fingerprinting extraction method utilizing Random Matrix Theory, evaluated across various SNR. Additionally, the Wsnet+SVM method proposed in [13] is employed on the same dataset for comparison. Results show a notable improvement in classification accuracy with the RMT+CNN method, demonstrating enhancements of 4.8%, 11.9%, and 8.6% across SNR ranges of -10 to 0 dB, 1 to 10 dB, and 11 to 20 dB, respectively. The RMT-based RF fingerprint method achieves high accuracy in classifying bit-similar devices even under low SNR conditions. Moreover, this method enables satisfactory identification accuracy with a streamlined CNN model.

REFERENCES

[1] W. Lee, S. Y. Baek, and S. H. Kim, "Deep-learning-aided rf fingerprinting for nfc security," *IEEE Communications Magazine*, vol. 59, no. 5, pp. 96–101, 2021.

[2] Y. Peng, P. Liu, Y. Wang, G. Gui, B. Adebisi, and H. Gacanin, "Radio frequency fingerprint identification based on slice integration cooperation and heat constellation trace figure," *IEEE Wireless Communications Letters*, vol. 11, no. 3, pp. 543–547, 2021.

[3] Y. Pan, S. Yang, H. Peng, T. Li, and W. Wang, "Specific emitter identification based on deep residual networks," *IEEE Access*, vol. 7, pp. 54425–54434, 2019.

[4] L. Peng, J. Zhang, M. Liu, and A. Hu, "Deep learning based rf fingerprint identification using differential constellation trace figure," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 1091–1095, 2019.

[5] G. Shen, J. Zhang, A. Marshall, L. Peng, and X. Wang, "Radio frequency fingerprint identification for lora using deep learning," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 8, pp. 2604–2616, 2021.

[6] G. Shen, J. Zhang, A. Marshall, M. Valkama, and J. Cavallaro, "Radio frequency fingerprint identification for security in low-cost iot devices," in *2021 55th Asilomar Conference on Signals, Systems, and Computers*, pp. 309–313, IEEE, 2021.

[7] T. Jian, B. C. Rendon, E. Ojuba, N. Soltani, Z. Wang, K. Sankhe, A. Gritsenko, J. Dy, K. Chowdhury, and S. Ioannidis, "Deep learning for rf fingerprinting: A massive experimental study," *IEEE Internet of Things Magazine*, vol. 3, no. 1, pp. 50–57, 2020.

[8] T. Jian, B. C. Rendon, E. Ojuba, N. Soltani, Z. Wang, K. Sankhe, A. Gritsenko, J. Dy, K. Chowdhury, and S. Ioannidis, "Deep learning for rf fingerprinting: A massive experimental study," *IEEE Internet of Things Magazine*, vol. 3, no. 1, pp. 50–57, 2020.

[9] W. Zhang, W. Zhao, X. Tan, L. Shao, and C. Ran, "Adaptive rf fingerprints fusion via dual attention convolutions," *IEEE Internet of Things Journal*, vol. 9, no. 24, pp. 25181–25195, 2022.

[10] G. Livan, M. Novaes, and P. Vivo, *Value the Eigenvalue*, pp. 7–13. Cham: Springer International Publishing, 2018.

[11] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters, "Noise dressing of financial correlation matrices," *Physical review letters*, vol. 83, no. 7, p. 1467, 1999.

[12] J. Veraart, D. S. Novikov, D. Christiaens, B. Ades-Aron, J. Sijbers, and E. Fieremans, "Denoising of diffusion mri using random matrix theory," *Neuroimage*, vol. 142, pp. 394–406, 2016.

[13] X. Yan, Y. Chen, Q. Wang, H.-C. Wu, and Y. Wu, "Robust bluetooth rf-fingerprint identifier using wavelet scattering network," in *2023 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, pp. 1–5, IEEE, 2023.