# RF Fingerprinting-Based IoT Node Authentication Using Mahalanobis Distance Correlation Theory

Dinh Duc Nha Nguyen, Keshav Sood, Mohammad Reza Nosouhi, Yong Xiang, *Senior Member, IEEE*, Longxiang Gao, *Senior Member, IEEE*, and Lianhua Chi, *Member, IEEE*

***Abstract*—The wireless Internet of Things (IoT) node authentication approaches also used Radio Frequency (RF) fingerprinting or physical unclonable features (PUF) of IoT devices for node authentication. Machine learning based models play vital role in these approaches. In this letter, we introduce an effective and novel IoT node authentication approach using Mahalanobis Distance correlation and Chi-square distribution theories. Further, it has lower computational time to determine node authenticity. The comparative results of our proposal with three recent machine learning based approaches and PUF based approaches are promising which validates the effectiveness and the novelty of our proposal.**

***Index Terms*—IoT authentication, radio frequency (RF) fingerprinting, Mahalanobis distance, next generation networks.**

## I. Introduction and Background

THE INTERNET of Things (IoT) devices are resource constrained and are vulnerable to be exploited by adversaries [1]. To protect the networks from adversaries, node authentication is indeed essential and is the first line of defense. In this line, firstly, cryptography-based methods are widely being used by industries for device authentication. Unfortunately, IoT applications are highly constrained in computational resources, hence they are not fully effective to utilize even many modern cryptographic methods for security purposes, let alone the traditional cryptography solutions [1]. Secondly, researchers leverage *Radio frequency (RF) fingerprinting* or Physically-Unclonable-Functions (PUFs) based approaches [2], [3], [4] which extracts the physical characteristics of devices from RF signals. These RF fingerprinting or signatures are the hardware random imperfections caused at the time of radio circuitry fabrication of devices. It is arguable that the physical characteristics are impossible to mimic by adversaries, so the PUF based methods are effective for security purposes [2], [3], [5]. The PUFs based approaches also use machine learning models [1], [6]. These methods have certain limitations, for example heterogeneity in the data sets acts as a barrier in the performance of these models such as accuracy, latency, etc. These issues eventually affect the performance of security applications [7], [8].

We note that in large scale IoT networks, the cryptography-based approaches take much time to authenticate a device [1] and the machine learning based methods also suffer from dataset size [8], [9] and other issues (see [1] and see references therein) which affects directly to the accuracy of authentication methods using RF fingerprinting. Overall, this leads to the concern of the stability (substantial level of accuracy [7]) of RF fingerprinting approaches based on machine learning methods [1], [2], [5]. This emphasized the need of investigating effective solutions without having complex cryptography and machine learning parts involved. We do not argue to replace the existing solutions based on the aforementioned categories rather we encourage, and we take initiative to investigate the solution that can be used as multi-factor authentication approach in specific scenarios (such as operational technologies) where high trust is required.

Motivated from this, in this letter we propose a novel methodology for wireless IoT node authentication. Although we also leverage the RF fingerprinting approach given that the PUF features are impossible to mimic, instead of complex and well known machine learning models we use Mahalanobis Distance and Chi-square distribution approaches to detect and classify the legitimate nodes among the non-legitimate ones. In contrast to the machine learning approaches, our method provides a higher and stable detection accuracy as well as reduces the authentication time. We also show that the proposal is not much affected w.r.t the changes in environmental conditions in which the device operates. We show that at varying signal to noise ratio (SNR) the approach still gives a stable accuracy. We also provide comparisons between the proposed method and other recent RF fingerprinting approaches to demonstrate the effectiveness of our solution.

The followings are our contributions.

1) We propose an effective methodology for anomaly detection (and node authentication) using RF fingerprinting in IoT networks. We use Mahalanobis Distance theory to authenticate IoT nodes. We are among the early ones applying Mahalanobis Distance in RF fingerprinting.
2) We have proved that the approach is effective under different scenarios such as the accuracy and node detection time is very stable w.r.t changes in the number of IoT nodes and SNR values.

Manuscript received September 21, 2021; revised January 31, 2022; accepted April 8, 2022. Date of publication April 15, 2022; date of current version May 25, 2022. The associate editor coordinating the review of this article and approving it for publication was C. Wang. *(Corresponding author: Dinh Duc Nha Nguyen.)*

Dinh Duc Nha Nguyen, Keshav Sood, and Mohammad Reza Nosouhi are with the Centre of Cyber Security Research Innovation, Deakin University, Geelong, VIC 3220, Australia (e-mail: nguyendinh@deakin.edu.au; keshav.sood@deakin.edu.au; m.nosouhi@deakin.edu.au).

Yong Xiang is with the Deakin Blockchain Innovation Laboratory, School of Information Technology, Deakin University, Geelong, VIC 3220, Australia (e-mail: yong.xiang@deakin.edu.au).

Longxiang Gao is with the Shandong Academy of Sciences, Qilu University of Technology, Jinan 250316, China, and also with the Shandong Computer Science Center, National Supercomputer Center in Jinan, Jinan 250101, China (e-mail: longxiang.gao@deakin.edu.au).

Lianhua Chi is with the Department of Computer Science and Information Technology, La Trobe University, Melbourne, VIC 3086, Australia (e-mail: l.chi@latrobe.edu.au).

Digital Object Identifier 10.1109/LNET.2022.3167665

2576-3156 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Deakin University. Downloaded on September 17,2024 at 21:59:43 UTC from IEEE Xplore. Restrictions apply.

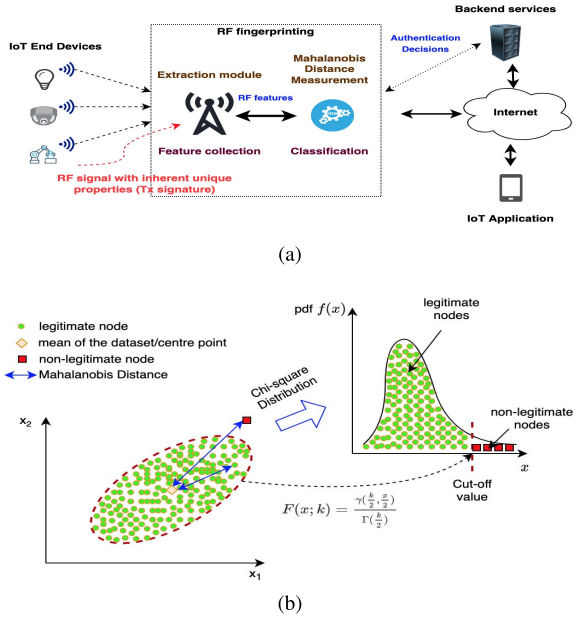


Fig. 1. (a) A high level framework of the proposed approach. (b) Overview of Mahalanobis Distance and Chi-squared Distribution.

3) The comparison of our approach with three recent approaches using machine learning models is shown. The results are promising which has demonstrated the effectiveness of our solution.

*Benefits:* The benefits of our method are as follows: (1) The method can be implemented on any existing systems without requiring additional resources as the computation to determine node's legitimacy is shifted to the IoT gateway rather than on the IoT devices itself. (2) Our method outperforms other machine learning approaches; this enables RF fingerprinting methods to be applied in authentication systems where latency is a sensitive feature for real-time IoT applications. (3) Our approach has a stable detection accuracy of authentication despite the size changes of the dataset. It is suitable for the dynamic and flexible IoT networks.

## II. Proposed Approach

A high level framework of the proposed solution is shown in Fig. 1(a). The scheme has two main phases: a) feature collection phase and b) classification phase. The IoT gateway or base station receives the RF signals/data generated by the IoT device/s (transmitters) in the network. Then the correlation among these features is computed using Mahalanobis Distance (MD) and Chi-square Distribution based theoretical approaches to determine the legitimacy of the IoT nodes.

### A. Feature Collection Phase

The base station or IoT gateway receives raw RF signals from IoT transmitters which is filtered out to keep only the key RF features to be used for device authentication. We notice that there is a threat that dynamic channel conditions affect the RF fingerprinting extraction processes [2], [10]. Therefore ,to maintain the stability of RF approaches, the channel features must be estimated and compensated. For instance, the work in [2] used a root-raised cosine filter and a Doppler corrector to mitigate the effect of noise/attenuation in the communication medium. In our work we consider that the feature extraction process is similar to the approaches discussed in [2] and [3]. Moreover, we tested our solution at different SNRs to demonstrate the impact w.r.t the changes in environmental and channel conditions.

### B. Classification Phase

In this phase we use the MD and Chi-square Distribution theoretical approaches to accurately determine whether a new IoT node is a legitimate or not. To better comprehend this phase, we have given Fig. 1(b). We emphasize that although these theories are well known (in pattern recognition and other medical domains [11], [12], [13]), however, to the best of our understanding, they have not been used in RF fingerprint based anomaly detection in IoT networks and related research works. As seen from Fig. 1(a), we employ the MD for classifying the incoming IoT nodes into two categories, legitimate and non-legitimate.

**Mahalanobis Distance** is a measure of the distance between a variable $x$ and a distribution which is calculated by a mean and the covariance matrix. This theory is often utilized by pattern recognition researchers to measure the similarity between the data distribution of train and test samples. Hence MD is postulated to be a multivariate normal distribution. As seen in Fig. 1(b), the region of persistent MD around the center point constructs an ellipse in two - dimensional space (assuming 2 variables are measured).

In Fig. 1(b), if a new node enters in the network (red dot outside the MD region), the MD calculates the distance between a new node and the mean of the dataset (from the centred point shown as an orange dot).

The formula to compute MD is as below:

$$D = \sqrt{(x-y)^T.C^{-1}.(x-y)} \tag{1}$$

Here, $(x - y)$ is the distance of the vector from the mean. We then divide this by the covariance matrix (or multiply by the inverse of the covariance matrix).

From the equation, it can be seen that the distance has an inverse relationship with covariance. Furthermore, the covariance reveals the correlation of the variables in dataset.

**Chi-square distribution** results in the continuous distribution of the total of squared random variables in case the variables are independent. It demonstrates the confidences encompassing the variance and standard deviation of a point to a normal distribution. Furthermore, it is also employed to evaluate how good sample data shape to the actual population. From the equation (1), it can be seen that the MD is a subbranch of Chi-square distribution with $k$ degrees of freedom ($k$ is a number of dimensions of the dataset). In such cases, the conversion to Chi-square $p$-values (probability of observing a test statistic) serves to recode the MD to a 0-1 scale. In general (see right side of Fig. 1(b), the $p$-value reflects the probability of seeing a MD value as large or larger than the actual MD value, $p$ values close to 0 reflect high MD values and hence they are very dissimilar (legitimate nodes) to the ideal combination of new variables. Mathematically, the Chi-squared Distribution (also $\chi^2$-distribution) is the distribution of a total of the squares of $k$. Suppose $x_1, x_2, \ldots, x_k$ are

Authorized licensed use limited to: Deakin University. Downloaded on September 17,2024 at 21:59:43 UTC from IEEE Xplore. Restrictions apply.

TABLE I
RF FEATURES USED IN OUR EXPERIMENTS

| No. | Features | Mean | Standard Deviation |
|---|---|---|---|
| 1 | Carrier Frequency Offset(CFO) | 2.4 GHz | 48 kHz |
| 2 | Amplitude Mismatch (In-Phase) | 0 dB | 3 dB |
| 3 | Amplitude Mismatch (Quadrature) | 0 dB | 3 dB |
| 4 | Phase Offset (In-Phase) | $0^o$ | $10^o$ |
| 5 | Phase Offset (Quadrature) | $0^o$ | $10^o$ |
| 6 | Clock skew | 0 ns | 40 ns |
| 7 | DC offset | 0 V | 1 V |

independent variables, so the total of their squares as:

$$Q = \sum_{i=1}^{k} {x_i}^2 \tag{2}$$

so, the chi-square distribution with *k* is denoted as

$$Q \sim \chi^2(k) \quad \text{or} \quad Q \sim \chi_k^2 \tag{3}$$

The probability density function is defined as:

$$f(x;k) = \frac{x^{\frac{k}{2}-1} e^{-\frac{x}{2}}}{2^{\frac{k}{2}} \Gamma(\frac{k}{2})}, x > 0 \tag{4}$$

The Cumulative Distribution Function (CDF) of Chi-square distribution is denoted as:

$$F(x;k) = \frac{\gamma(\frac{k}{2}, \frac{x}{2})}{\Gamma(\frac{k}{2})} \tag{5}$$

The cut-off value (see Fig. 1(b)) is determined by the CDF which separates the rejection region (illegitimate nodes) and the sampling distribution (legitimate nodes). A node is confirmed as legitimate if its MD to the mean of the dataset is less than the cut-off value.

## III. PERFORMANCE EVALUATION

We evaluate our method by conducting experiments under various scenarios. We use the Wireless Waveform Generator toolbox of MATLAB to generate a dataset (matrix vector is 500 x 100). We have 500 testing devices, for each device we have slightly changed the frequency, amplitude and phase to stimulate the nonideality of RF features. Each device provides 100 RF signal data. We only select features which are location independent in order to enable the approach effective in IoT mobility applications. The RF features (as well as their mean and standard deviation values) we used for our experiments are described in Table I. The experiments are conducted on a non-GPU desktop computer with CPU Intel Core i7 and 16GB RAM using Python 3.8.3, Keras 2.4.3 with TensorFlow 2.4.1.

Firstly, we evaluate the performance evaluation of our approach w.r.t different number of RF features. In Fig. 2(a), the highest detection accuracy we have obtained is with three features. Although even if we increase the number of features (from 3 to 7), we do not see significant drop in the accuracy. To analyse the impact of varying SNR on the performance of our approach, we gradually increased the number of devices from 50 to 500 to determine the detection accuracy under various SNR values (from 15dB to 30dB). In this test we used

TABLE II
KEY NOTATIONS

| Notation | Explanation |
|---|---|
| $D$ | the Mahalanobis distance |
| $x$ | RF feature vector |
| $y$ | vector of mean values of the variables |
| $T$ | transpose vector |
| $C$ | the covariance matrix |
| $k$ | the number of RF features |
| $v_{thr}$ | cut-off value |
| $\gamma$ | the lower incomplete gamma function |
| $\Gamma$ | the gamma function |
| $f(x;k)$ | the probability density function |
| $F(x;k)$ | the cumulative distribution Function |

TABLE III
CLASSIFICATION METRICS OF OUR SCHEME AT SNR=30 AND THE NUMBER OF DEVICES ARE 300

| Metric | Value |
|---|---|
| F1-score | 0.987 |
| Balanced Accuracy | 0.983 |
| Sensitivity | 99.18% |
| Specificity | 96.75% |

TABLE IV
COMPARISON OF OUR METHOD WITH EXISTING APPROACHES

| Reference and ML Model | MDA/ML [10] | CNN [5] | LSVM [14] | Proposed method |
|---|---|---|---|---|
| Highest detection accuracy (%) | 95% | 97% | 98.8% | 99.28% |

three features: CFO, Amplitude Mismatch, and Phase Offset. The results of the tests are demonstrated in the Fig. 2(b). The detection accuracy is stable. As the SNR increases, it can be observed that the detection accuracy improves. This is a natural finding, as the SNR quantifies the influence of noise signals on the original signals. We have also conducted further experiments to evaluate classification metrics, the results are shown in Table III. These values show the effectiveness of our solution.

To determine the node authentication time of the proposed approach, as the network scale increases, we use detection time as a metric to evaluate the effectiveness of our method. We compare the detection time of our approach at various transmitter counts with other RF fingerprinting approaches based on machine learning in the same scenario (we also used three features: CFO, Amplitude Mismatch and Phase Offset, but decreased the number of devices from 10 to 100 due to the resources and time constraints). The result is shown in Fig. 2(c). It shows that our method outperforms the other machine learning approaches. This makes sense because the machine learning algorithms require a high processing time for the classification [7] while our method with effortless distance calculations is time efficient. This has significant advantage in scenarios where real-time detection is a key constraint.

We also compare our approach with other recent RF fingerprinting approaches to further justify that the proposal outperforms the recent existing approaches. The work given in [10] extracts physical characteristics from waveform coloration as RF features in the presence of a Rayleigh fading channel. The method employs a neural network-based classifier, known as Multiple Discriminant Analysis/Maximum

Authorized licensed use limited to: Deakin University. Downloaded on September 17,2024 at 21:59:43 UTC from IEEE Xplore. Restrictions apply.

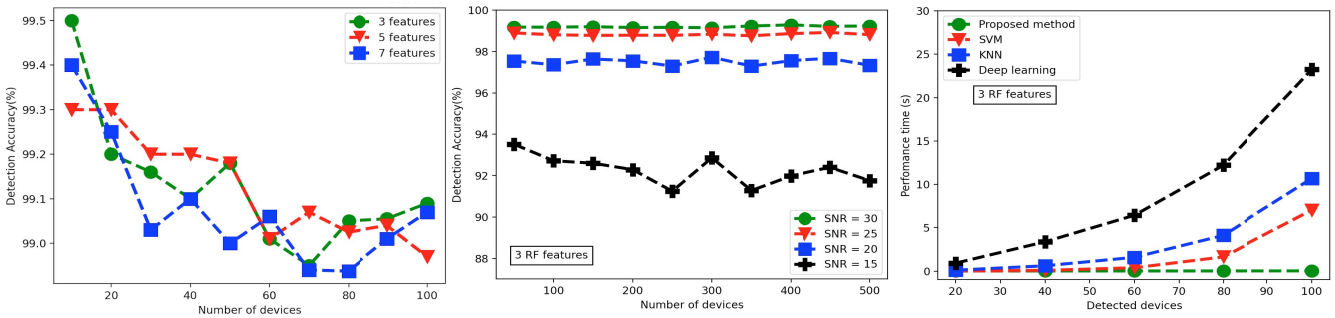


Fig. 2. (a) Average detection accuracy of our scheme at different number of features. (b) Average detection accuracy of our scheme at different number of devices and different values of SNR. (c) Comparison of the proposed method with different machine learning models.

Likelihood (MDA/ML). In [5] authors proposed a scheme for device identification and authentication by extracting RF fingerprints from the ROI (region of interest). The scheme puts extraction features into a multisampling convolutional neural network (MSCNN) for classification and detects ZigBee devices. The approach in [14] also focuses on RF fingerprinting methods to identify radio devices. After collecting the RF features, it employs Linear Support Vector Machine (LSVM) for classification and identifying devices. Very recently, SLoRa [15] proposes a node authentication method by leveraging CFOs and link signatures. The approach has shown good performance and is effective to detect the spoofing attacks in the scenarios that the attackers mimic the victim's CFO. However, the method focuses only on specific RF fingerprinting features (CFO and link signature) for particularly LoRa communications. Moreover, SLoRa cannot be applied to mobile scenarios because both LoRa gateways and nodes must be at relatively fixed positions. In contrast, our solution has a wider coverage in wireless communications (e.g., Wi-Fi and ZigBee) than SLoRa. Furthermore, we only select features which are location independent, hence our approach is effective in mobility scenarios.

From our experiments and from Table IV, we show that the proposed approach has highest detection accuracy comparisons between these works and our method. We emphasized that we use the simple approach over the complex models and have better performance evaluations.

## IV. Conclusion and Future Work

We proposed a novel IoT node authentication methodology using RF fingerprinting and Mahalanobis Distance measurement. Our method is simple and outperforms over the existing complex RF fingerprinting approaches. In comparison to the existing methods our method achieves higher detection accuracy, lesser computational time, and has shown stable results at varying SNR values. We note that the RF fingerprinting features are impossible to mimic by attackers, so the PUF based methods are effective for security purposes. On the other side, the adversarial machine learning approaches of the PUF based schemes are not fully threat prone, therefore, the security evaluation of any PUF based approach in itself is also another branch of study. Further, the dataset simulated by MATLAB is used in many RF studies and the results are encouraged, the real dataset needs to be evaluated in the future work.

## References

[1] P. Gope, B. Sikdar, and O. Millwood, "A scalable protocol level approach to prevent machine learning attacks on physically unclonable function based authentication mechanisms for Internet of Medical Things," *IEEE Trans. Ind. Informat.*, vol. 18, no. 3, pp. 1971–1980, Mar. 2022.

[2] 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 Things J.*, vol. 6, no. 1, pp. 388–398, Feb. 2019.

[3] N. Soltanieh, Y. Norouzi, Y. Yang, and N. C. Karmakar, "A review of radio frequency fingerprinting techniques," *IEEE J. Radio Freq. Identif.*, vol. 4, no. 3, pp. 222–233, Sep. 2020.

[4] H. Thapliyal and S. P. Mohanty, "Physical unclonable function (PUF)-based sustainable cybersecurity," *IEEE Consum. Electron. Mag.*, vol. 10, no. 4, pp. 79–80, Jul. 2021.

[5] J. Yu, A. Hu, G. Li, and L. Peng, "A robust RF fingerprinting approach using multisampling convolutional neural network," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6786–6799, Aug. 2019.

[6] L. Cui *et al.*, "Security and privacy-enhanced federated learning for anomaly detection in IoT infrastructures," *IEEE Trans. Ind. Informat.*, vol. 18, no. 5, pp. 3492–3500, May 2022.

[7] W. Jian, Y. Zhou, and H. Liu, "Lightweight convolutional neural network based on singularity ROI for fingerprint classification," *IEEE Access*, vol. 8, pp. 54554–54563, 2020.

[8] K. Sood, K. K. Karmakar, V. Varadharajen, N. Kumar, Y. Xiang, and S. Yu, "Plug-in over plug-in evaluation in heterogeneous 5G enabled networks and beyond," *IEEE Netw.*, vol. 35, no. 2, pp. 34–39, Mar./Apr. 2021.

[9] X.-Y. Liu and X. Wang, "Real-time indoor localization for smartphones using tensor-generative adversarial nets," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 8, pp. 3433–3443, Aug. 2021.

[10] M. Fadul, D. Reising, T. D. Loveless, and A. Ofoli, "Nelder-mead simplex channel estimation for the RF-DNA fingerprinting of OFDM transmitters under rayleigh fading conditions," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 2381–2396, 2021.

[11] L. Friedman and O. V. Komogortsev, "Assessment of the effectiveness of seven biometric feature normalization techniques," *IEEE Trans. Inf. Forensics Security*, vol. 14, pp. 2528–2536, 2019.

[12] J. Jin, Y. Zhu, Y. Zhang, D. Zhang, and Z. Zhang, "Micrometeoroid and orbital debris impact detection and location based on FBG sensor network using combined artificial neural network and Mahalanobis distance method," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, Jun. 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9462134

[13] G. Gallego, C. Cuevas, R. Mohedano, and N. García, "On the Mahalanobis distance classification criterion for multidimensional normal distributions," *IEEE Trans. Signal Process.*, vol. 61, no. 17, pp. 4387–4396, Sep. 2013.

[14] A. Aghnaiya, A. M. Ali, and A. Kara, "Variational mode decomposition-based radio frequency fingerprinting of Bluetooth devices," *IEEE Access*, vol. 7, pp. 144054–144058, 2019.

[15] X. Wang, L. Kong, Z. Wu, L. Cheng, C. Xu, and G. Chen, "SLoRa: Towards secure LoRa communications with fine-grained physical layer features," in *Proc. 18th Conf. Embedded Netw. Sens. Syst.*, 2020, pp. 258–270.

Authorized licensed use limited to: Deakin University. Downloaded on September 17,2024 at 21:59:43 UTC from IEEE Xplore. Restrictions apply.