Design and Robust Evaluation of Next Generation Node Authentication Approach

Dinh Duc Nha Nguyen, Keshav Sood, Yong Xiang, *Senior Member, IEEE,* Longxiang Gao, *Senior Member, IEEE,* Lianhua Chi, *Senior Member, IEEE,* Gurpreet Singh, and Shui Yu, *Senior Member, IEEE*

***Abstract*—The flexibility of 5G-NGNs makes them an ideal infrastructure for supporting mission-critical IoT applications that require low latency and high bandwidth. However, due to the rapid proliferation and the integration of IoTs with 5G, the threat surface has considerably expanded. Hence the security of IoT devices is a big concern. Unfortunately, IoT devices have limited resources, and the traditional security approaches (authentication and intrusion detection approaches) of cryptography do not work effectively on 5G-IoT ecosystems. Motivated from this, we leverage the distinctive RF (Radio Frequency) fingerprinting signatures of IoT devices and used them to train a Deep learning model, Mahalanobis Distance theory in addition to the Chi-square distribution theory, to authenticate the IoT nodes. Under robust scenarios we have tested the approach shows detection accuracy (99.35%) as well as significant amount of reduction in model's training time as these two metrics are one of the primary key performance indicators (KPIs). In order to evaluate the effectiveness of the proposed method in real-time scenarios, we tested the proposed solution with a real RF dataset and the OSM-MANO 5G platform. The model underwent formal verification using the Tamarin Prover tool, and the proposal was also compared with recent research works.**

***Index Terms*—Authentication, Physical layer security, RF fingerprinting, IoT, 5G.**

I. Introduction

The integration of 5G with Internet of Things (IoTs) offers a flexible foundation for large-scale IoT applications, including smart cities, autonomous vehicles, and more. [1], [2]. A recent report estimates that around twenty billion devices will be connected to the Internet by 2025 [3]. The intergration of Next Generation Networks (e.g., 5G) and the Internet of Things (NGN-IoTs) has resulted in an expanded security threat for NGN-IoT networks and applications [4], [5], [6]. This is due to the inherent weakness of IoT devices, which have limited resources and are not fully inadequate of implementing advanced authentication protocols [7], making them an easy target for attackers. This is an attractive feature for an adversary to form IoT botnets and attempt to compromise network and user security [8]. In this case lightweight authentication solutions play a crucial role to authenticate legitimate nodes [9].

Traditional authentication methods rely on credentials and cryptographic techniques to grant network access to users. However, such methods have limitations [10], [11], particularly with regard to IoT devices, which has limited computation for the strength cryptographic algorithms. Moreover, these methods are vulnerable to impersonation attacks [12], [13] and may introduce significant latency [5], [14], which can affect the performance of real-time applications in 5G networks. Consequently, authentication of IoT nodes in next generation IoT networks remains a challenge.

By leveraging the unique attributes of physical devices, the Physical Layer Authentication (PLA) approach creates unique identifiers for each device that can be used to distinguish them from one another. In order to better comprehend, we classify the current research in authentication solution into two distinct groups: a) based on traffic analysis, and b) based on analyzing PLA signatures [15]. In IoTs, the Radio Frequency (RF) fingerprinting method leverages the physical properties of wireless devices for accurately identifying nodes [16], [17], [18]. RF signature-based methods extract wireless device features resulting from hardware random defects that occur during device manufacturing. The RF signatures are claimed to be unique and difficult to copy, so it can be used to verify the device identifiers [19] [20].

Compared to other authentication methods such as network traffic analysis and resource-limited cryptographic solutions, the use of RF features for node authentication has significant advantages [21], [22]. Our proposed methodology uses RF signatures to authenticate wireless IoT nodes in 5G-IoT networks and involves feature collection, extraction, and classification. The process of feature collection and extraction can be categorized into two distinct sections, namely transient signals and steady-state signals [23]. The RF fingerprinting technique based on transient signals obtains consistent device characteristics from the transient portion of incoming RF signals [24]. However, the primary disadvantage of this category is that it requires expensive gear to identify the features. Alternatively, RF fingerprinting techniques based on the steady-state portion of the signal collect characteristics from the modulated portion of the devices by leveraging the signals received at gateways [25], [26]. In our work, the steady-state signal approaches are selected. Then we employ statistical approaches such as Mahalanobis Distance and Chi-squared Distribution to compute the data from the steady-state

Dinh Duc Nha Nguyen, Keshav Sood and Gurpreet Singh are with Centre of Cyber Security Research and Innovation (CSRI), School of IT, Deakin University, Geelong, 3220, VIC, Australia. E-mail: nguyendinh@deakin.edu.au, keshav.sood@deakin.edu.au, zpx@deakin.edu.au; Yong Xiang is with Deakin Blockchain Innovation Lab, School of IT, Deakin University, Geelong. E-mail: yong.xiang@deakin.edu.au; Longxiang Gao is with the 1. Qilu University of Technology (Shandong Academy of Sciences); 2. Shandong Computer Science Center (National Supercomputer Center in Jinan). E-mail: longxiang.gao@deakin.edu.au; Lianhua Chi is with Computer Science & Information Technology, La Trobe University. Australia. E-mail: l.chi@latrobe.edu.au; Shui Yu is with The University of Technology Sydney, NSW, Australia. E-mail: shui.yu@uts.edu.au.

signal and identify the anomaly detection. The anomalies can be considered as the illegitimate device. This means the RF fingerprint technique and statistical approaches can be used to verify the device's authenticity. Our work utilizes physical layer security and statistical methods to authenticate a device and ensure that it has not been tampered with or modified in any way that could pose a security threat.

We acknowledge that our early work is discussed and evaluated in [27], [28]. It should be noted that [27], [28] has a significant gap that RF feature collection is only simulated using a Matlab ToolBox and this is a strong limitation of [27] and [28]. Although simulations can provide valid preliminary results, however especially with PLA, tests need to be executed with real devices, as the reality is very far from what simulations can achieve. Additionally, the vulnerability of the proposed model was not investigated in [27] and [28]. Moreover, end-to-end delay is the important metric to evaluate an authentication framework under 5G-IoT networks that was not tested in the previous work.

The following are the significant contributions of our solution.

1) We propose an authentication solution that uses radio-frequency fingerprinting and statistical analysis, which is significantly less computationally intensive than conventional methods. Our approach employs the Mahalanobis Distance theory.
2) We have used a real data set to further investigate the credibility of the approach. Secondly, we conducted robust experiments in various scenarios, including varying numbers of nodes, signal-to-noise ratios, and numbers of features, to compare our approach to recent works. The results demonstrate that our method achieves higher detection accuracy and mitigates latency problems. The key metrics we use for evaluation are average accuracy, computed resources, and the end-to-end delay.
3) To demonstrate the lightweight and the real-time efficacy, we implemented and evaluated our method using the ETSI-5G OSM-MANO and the AWS cloud platform. Additionally, we compared our approach's performance with other existing methods.
4) ”We have used a formal verification tool to verify the proposed method and to prove the security analysis against impersonate and man-in-the-middle attacks.

The organization of the paper is as follows: Section II and III provide an overview of related works and preliminaries, respectively. Our proposed method is presented in Section IV, and the theoretical analysis is conducted in Section V. We evaluate the performance of our approach in Section VI and analyze its security in Section VII. Finally, we conclude our work in Section VIII. Table I shows the list of abbreviations and Table IV provides the notations used in the paper.

TABLE I
LIST OF ABBREVIATIONS USED IN THE PAPER

Abbreviation	Definition	Abbreviation	Definition
AP	Access Point	OSM MANO	Management and Orchestration
AWS	Amazon Web Services	MDA/ML	Multiple Discriminant Analysis/Maximum Likelihood
ANN	Artificial Neural Network	MSCNN	Multisampling Convolutional Neural Network
AE	Autoencoder	NLOS	Non-line-of-sight
BLE	Bluetooth Low Energy	OTs	Operational Technologies
CFD	Carrier Frequency Differences	PSD	Power Spectral Density
CFO	Carrier Frequency Offset	PLA	Physical Layer Authentication
CPU	Central Processing Unit	PA	Power Amplifier
CSI	Channel State Information	PSD	Power Spectral Density
CD	Chi-square Distribution	PCA	Principal Component Analysis
CNN	Convolutional Neural Network	QAM	Quadrature Amplitude Modulated
DAE	Deep Autoencoders	RF	Radio Frequency
DNN	Dense Neural Network	RSSI	Radio Signal Strength Indicator
DC	Direct Current	RFO	Random Forest
DWT	Discrete Wavelet Transform	SOMs	Self-Organizing Maps
ETSI	European Telecommunications Standards Institute	SNR	Signal-to-noise ratio
HOS	Higher Order Statistical	SEI	Specific Emitter Identification
IoT	Internet of Things	SVM	Support Vector Machines
IDS	Intrusion Detection System	NGNs	The Next Generation Networks
KNN	K-nearest neighbour	TFED	Time-frequency-energy
LOS	Line-of-sight	UAVs	Unmanned Aerial Vehicles
LSVM	Linear Support Vector Machine	UN	User Node
LSTM	Long Short-Term Memory	VMD	User Node
ML	Machine learning	VNF	Virtualized Network Function
MD	Mahalanobis Distance	WIOT	Wireless Internet of Things

II. Related Work

Authentication based on RF fingerprinting is a widely recognized research area in wireless networks. Mirsky *et al.* [29] have introduced Kitsune, an approach that aims to detect attacks without the need for an effort to manually label the dataset and update the model. The method simplifies the process of labeling a dataset, utilizing an autoencoder to discover the usual pattern and identify anomalous situations. However, the approach may not be suitable for large-scale networks due to its high memory consumption. In other work, Chatterjee *et al.* [20] proposed approach involves using an Artificial Neural Network (ANN) to detect specific RF signatures based on two characteristics. Their approach achieved an accuracy of 97%.

In contrast to previous efforts that employ steady-state signals for RF fingerprinting, the study in [31] employs transient signals. By utilizing variational mode decomposition (VMD) and higher order statistical (HOS) features extracted from Bluetooth transient signals, the study proposes a new approach for classifying Bluetooth devices. The classification is performed using Linear Support Vector Machine (LSVM) which exhibits superior classification accuracy under conditions of low signal-to-noise ratio and limited number of features. The approach achieves a detection accuracy of 98.8%. Another work is Slora [32] which presents an RF fingerprinting-based authentication method that utilizes two RF characteristics, CFO and link signatures. The classification model is SVM, and the combination of the two characteristics enhances the detection accuracy. SLoRa, which exclusively employs the RF fingerprinting method for LoRa communications, achieves an accuracy of around 97%. The detection and classification of

TABLE II
OVERVIEW OF EXISTING WORKS

Year/Ref.	Project aim/overview	Model Used
2018 [29]	Kitsune: To minimize the amount of manual labeling required, Autoencoder is employed to differentiate between normal and abnormal patterns.	Autoencoder
2019 [20]	RF-PUF: A distinctive feature is determined based on RF characteristics and classified using an ANN method.	ANN
2019 [23]	RF fingerprinting features are classified using SVM after dimension reduction with PCA.	PCA + SVM
2019 [30]	ZigBee devices are classified based on their region-of-interest characteristic using MSCNN.	MSCNN
2019 [31]	LSVM is used to classify transient signals in RF fingerprinting for authentication purposes.	LSVM
2020 [32]	SLORA: Two RF characteristics and SVM-based categorization are utilized in Slora to prevent impersonation attempts.	SVM
2020 [33]	Wi-Fi and Bluetooth protocols are fingerprinted using RF fingerprinting for UAV detection and classification.	KNN
2020 [34]	Enhance the efficiency of the fingerprint classification scheme through the extraction of features of the region of interest.	Lightweight CNN
2020 [35]	By identifying the region of interest characteristic, the performance of a fingerprint classification scheme can be improved.	CNN
2020 [36]	Recognition effectiveness can be improved by identifying the optimal subset of features and using KNN for classification.	KNN
2020 [37]	Anomaly detection and attack classification are employed in the two-stage H2ID attack detection method.	Deep Autoencoder
2020 [38]	A dataset is provided for RF fingerprinting, and a CNN-based algorithm is used for evaluation.	CNN
2020 [39]	ORACLE is a deep-learning-based novel method to identify a unique radio from a pool of devices using CNNs.	CNN
2021 [40]	The raw sample's waveform domain is classified for device identification using deep learning.	DNN
2021 [41]	Nelder-Mead simplex-based channel estimator is employed to mitigate noise under Rayleigh fading conditions.	MDA/ML
2021 [25]	The impact of RFFI is extensively simulated and tested for narrowband transmitter-receiver impairments using models.	CNN
2021 [42]	Real-time performance is improved, and pose uncertainty can be detected by enhancing the kNN algorithm.	KNN
2021 [43]	Using Machine Learning methods, RSSI-based fingerprinting can be augmented and classified.	RF
2022 [44]	An unsupervised algorithm is used to train an artificial neural network matrix to generate self-organizing maps.	ANN
2022 [45]	The beam pattern properties of mmWave-enabled devices are used for spoofing detection, offering a security approach.	AE
2022 [46]	A Bluetooth fingerprint localization technique based on Bi-LSTM networks is introduced, using deep learning.	LSTM

unmanned aerial vehicles (UAVs) using RF fingerprinting on Wi-Fi and Bluetooth technologies were investigated by the authors in [33]. The authors employed a two-step process that involved naive Bayes based on Markov models for RF signal extraction and KNN for classification. They evaluated the accuracy of their approach on 15 different UAVs with various SNRs and compared it to five other machine learning models. Their method attained a detection accuracy of 98.13%. Additionally, the classification method's performance was improved using a Lightweight CNN model [34], which utilized the ROI during the classification phase. The initial step involves the pre-processing of raw pictures and the extraction of ROI patterns.

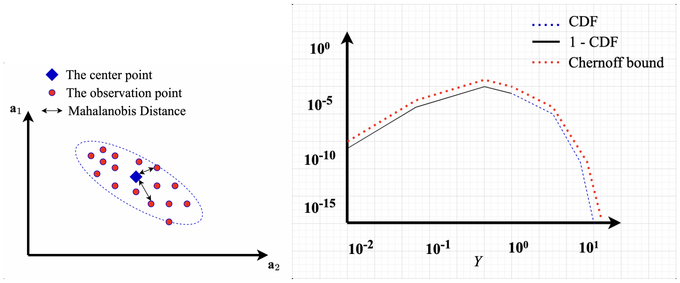
The use of RF fingerprint techniques and machine learning models have raised attention in the access control research. The work [35] provides a technique based on convolutional neural networks. They achieve highest accuracy at 99.70% under various epoch settings. Another work [36] employs K-Nearest Neighbor and assessed its resilience under various SNR settings. Their work reaches the highest detection accuracy at 97.86%. Although both [35] & [36] achieves a high detection accuracy, the dataset is generated by a simulated method and it is a strong limitation of these works. In a separate work [37], authors proposed H2ID to improve the effectiveness of attack detection. The paper provides two phases for identifying an attack: the anomaly detection phase employs a lightweight solution based on Deep Autoencoder (AE), and the attack classification phase employs the openset classification methodology. The primary advantage of the technology is its efficiency performance for IoT applications.

Furthermore, to fulfil the demand for a dataset for RF fingerprinting research, a large dataset was collected, comprising 20 devices with matching RF circuitry [38]. The study analyzed the effect of the wireless channel on CNN-based fingerprinting, showing that the wireless channel significantly impacts categorisation accuracy. Additionally, as the number of devices increases significantly, so does their accuracy. The author in [39] introduced a new system called ORACLE, which employs convolutional neural networks to identify a particular radio from a vast array of devices by training the network on the hardware limitations imposed by radio circuits on physical-layer I/Q samples. Li *et al.* [40] recently proposed a novel method that combines RF fingerprinting with deep learning in the waveform domain to achieve high device identification accuracy of about 99%. They haved captured images of the waveform from raw samples and use Dense Neural Network (DNN) for classification. In another work [41], researchers explored the use of Specific Emitter Identification (SEI) for RF fingerprinting, proposed an approach based on the Nelder-Mead (N-M) simplex-based channel estimator to overcome noise influence under Rayleigh fading conditions.

The authors in [42] propose an indoor localization technique based on RSSI fingerprints that does not require pose knowledge. They use a modified kNN algorithm and machine learning techniques to enhance real-time performance and tackle pose uncertainty. The method reduces the computational cost by employing a kNN algorithm to limit the searching region. In [43], the RF fingerprinting method and machine learning classification are employed to recognize the position of the user node for the device's authenticity. They have used the RSSI as the RF fingerprint features in the method. By incorporating fingerprinting data, the method has tried to improve accuracy, reduce computing costs and ensure framework precision. They tested with some machine learning models and the Random Forest shows the best performance with a test accuracy of 96%. The work in [44] presents a novel unsupervised machine learning approach for fast RF fingerprinting of LoRa modulated chirps. The approach involves training an artificial neural network matrix using an unsupervised machine learning technique that generates self-organizing maps (SOMs) for each authenticated transmitter and a potential rogue node. The proposed work has been tested and training time is the only metric to be discussed. To evaluate an intrusion detection, accuracy detection should be evaluated and this is the main limitation of the work. Recently, in [45], the authors have suggested a security approach for NGIoT networks that can identify wireless spoofing attacks by utilizing the distinct beam pattern characteristics of mmWave-capable devices. The proposed method detected fraudulent devices with 98.6% accuracy in their testing. Hu *et al.* [46] present a novel Bluetooth fingerprint-based localization technique that employs a bidirectional long short-term memory (Bi-LSTM) network, which differs from conventional localization algorithms that use machine learning techniques like k-nearest neighbor, random forest, and support vector machine. By utilizing deep learning, the proposed algorithm can fully learn from the localization data, leading to improved localization accuracy.

Our observation: The effectiveness of a machine learning model is directly linked to the size of the dataset it has been trained on, which creates a challenge for small to medium-sized IoT applications and networks [1], [47]. Additionally, the process of training a machine learning-based model is time-consuming, which can be problematic for mobile IoT devices that frequently join or leave the network, resulting in fluctuations in the amount of available data. This issue can affect both model training and node authentication process [4], [5], adding to the computational resources needed for making and enforcing authentication decisions, resulting in increased network latency. These issues are particularly problematic for 5G-IoT networks, which require low latency processing and accurate authentication systems that balance precision with detection time. We note that these challenges go against the fast processing needs of 5G-IoT networks and the accuracy demands of authentication systems. This trade-off between precision and detection time has a major impact on mission-critical applications, especially in certain scenarios or industrial sectors such as operational technologies (OTs) like micro-grids and smart power plants.

Previous studies have expressed concerns regarding the use of computational resources, the size of the training data, and the stability of detection accuracy. However, none of these studies have tested their suggested methods on a 5G test bed, nor have they provided evaluation metrics, such as model training time, node authentication time, and resources requirements, which are critical to gaining insight into the deployment challenges of such schemes on next-generation network platforms. Furthermore, the trade-off between accuracy and model training time in relation to frequent network scale changes have not been detail discussed and tested. These factors are essential to the success of any NG-IoT authentication solution.



(a) (b)

Fig. 1. (a) The Mahalanobis Distance. (b) The use of Chernoff bound (k=10).

This research presents a framework for addressing these challenges, which is compatible with the ETSI-NFV design standard for 5G networks and can be used in both small and large 5G-based IoT networks. Specifically, this study proposes the use of Mahalanobis Distance correlation theory with RF fingerprints for authenticating 5G-IoT nodes, which has not been previously proposed in the literature.

III. Preliminaries

A. Mahalanobis Distance

The Mahalanobis Distance (MD) is a distance measure that calculates the distance between two points in multidimensional space based on the mean and covariance matrix of a distribution. The use of a multivariate normal distribution enables us to compare data distributions of train and test samples, which is a common practice in pattern recognition research. The covariance matrix determines the shape of the data distribution in feature space, and in two-dimensional space (assuming two variables are measured), the persistent zone of Mahalanobis distance surrounding the central point forms an ellipsoid or hyperellipsoid in higher dimensions, see in 1(a). Unlike the Euclidean distance, the Mahalanobis distance can eliminate false information if the variables are correlated by transforming them into uncorrelated variables, computing their variance as a constant number, and calculating the Euclidean distance based on this.

B. Chi-square distribution

In probability theory, the Chi-square Distribution (CD) is a type of continuous probability distribution that is commonly used to determine the sum of independent squared random variables. It is used to calculate confidence intervals for the variance and standard deviation of a population and to check how well a sample fits to a specific distribution. The Mahalanobis Distance (MD) is a subset of the Chi-square distribution, which is used to estimate the distance between a

This article has been accepted for publication in IEEE Transactions on Dependable and Secure Computing. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TDSC.2024.3373778

point and the mean of a multivariate distribution. The number of degrees of freedom in MD is equivalent to the number of dimensions in the dataset.

Mathematically, the Chi-squared Distribution (also χ^2-distribution) is the distribution of the total number of n squares. Assume that $a_1, a_2, ...a_n$ are independent variables, hence the calculation of the sum of squares of independent variables can be done as follows::

$$P = \sum_{j=1}^{n} {a_j}^2 \qquad (1)$$

the chi-square distribution with n is denoted as

$$P \sim \chi^2(n) \quad \text{or} \quad P \sim \chi^2_n \qquad (2)$$

The following notation is used to refer to the Cumulative Distribution Function (CDF) of the Chi-square distribution:

$$G(a;n) = \frac{\gamma(\frac{n}{2}, \frac{a}{2})}{\Gamma(\frac{n}{2})} \qquad (3)$$

where the lower incomplete gamma function is denoted by $\gamma(d{,}y)$ and the regularised gamma function is denoted by $Q(d{,}y)$.

The Chernoff bound places constraints on the sums of the tail distributions of independent random variables that are exponentially diminishing in severity. In other words, it provides a representation of the cumulative distribution function's (CDF) shape. This pertains to situations where $0 < x < 1$, which includes instances where the CDF would be less than 50 percent.

$$G(xn;n) \leq (xe^{(1-x)})^{\frac{n}{2}} \qquad (4)$$

Similar to the previous case, the tail bound for situations in which xn is greater than one is (see Fig. 1 (b)).

$$1 - G(zn;n) \leq (ze^{(1-x)})^{\frac{n}{2}} \qquad (5)$$

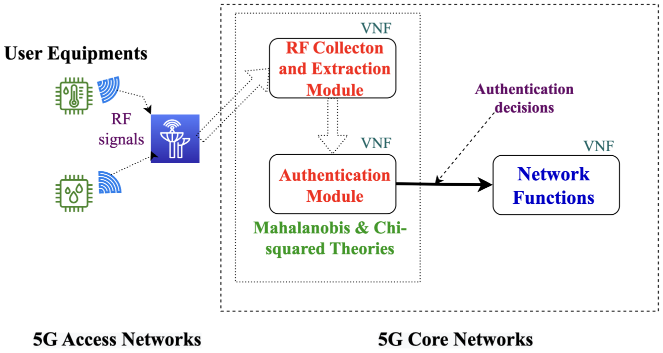


Fig. 2. The proposed approach's architecture presented at a high level.

IV. The Proposed Framework

The proposed approach is illustrated in Fig. 2. The approach is decomposed into two individual modules. It is assumed that the modules will be placed at the access control functions of the 5G network, taking the form of virtualized network functions (VNFs). Deploying them as VNFs provides more flexibility to incorporate them into the ETSI NFV's standard architecture.

Module 1: RF Feature Collection and Extraction. This module is responsible for collecting RF feature data of diverse devices and sending them to the next module. In order to convert the raw RF signals that are transmitted by IoTs' transmitters into RF feature data, this module collaborates with receivers (gateways or base stations) that are located within 5G Access Networks. In our method, we assumed that the data is collected from IoT devices at the gateways. These features are then transmitted to the backend server for further analysis using the Mahalanobis and Chi-squared distribution theories. During the course of IoT device's regular connections, the RF properties of the IoT nodes are derived from their RF signals. A large number of devices have their anomaly detection handled by the backend server, which in our case is the authentication module. The following radio frequency characteristics are utilised in the implementation of the suggested method, which can be seen in Table III.

TABLE III
The RF characteristics employed in our experiments

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

The first feature is Carrier Frequency Offset. A local oscillator can be found in each and every radio frequency component of a wireless device. The inherently variable characteristics of the local oscillator are caused by the local oscillator's manufacturing process, which is not faulty. It causes the radio frequency transmitters in any device to have a different frequency offset. The approach of RF fingerprinting can make use of the one-of-a-kind frequency offset as a distinguishing feature [20]. Because of this, the RF Collection and Extraction module, provided it has a high-quality reference clock, is able to take advantage of the frequency offset.

In addition to this, before the radio frequency signals in wireless communication are transmitted, they are modulated, and then they are demodulated at the receiving end. In-phase (I) and quadrature are the two primary stages that make up the process (Q). Nevertheless, the varying amplitude and phase of the components of a signal are constructed as a result of the defective analogue circuitry caused by during the manufacturing process. The depiction of the constellations reveals the unbalance quite clearly. The mismatch between I and Q can be identified, and it is distinct for each individual device [20], [48].

Accordingly, the RF Collection and Extraction module is able to compute the I-Q mismatch values by using the signals that it has received. In addition, every gadget that operates on radio frequency possesses a digital circuit that has the clock system. However, due to the different manufacturing process, the hardware of the clock in the transmitter and the receiver are not identical to one another. Therefore, the time mismatching occurs between the two source clocks. This peculiar occurrence is known as the clock skew. There is a one-to-one correspondence between the skew of any two separate clocks and the ability to identify it. As a result, RF Collection and Extraction, when combined with a reference clock of a high quality, are able to calculate clock skew as an RF feature.

In addition, the mixer process that takes place between the local oscillator and the RF input in wireless signal transmission creates an undesired DC (Direct current) that is referred to as the DC offset. It is possible to construct the receiver demodulator in such a way that it may adjust for the offset [49]. This helps to alleviate the problem. Therefore, the RF Collection and Extraction module has the ability to ascertain the offset value as an RF feature. In a nutshell, the initial module of the proposed technique is able to extract RF features from raw signals transmitted by transmitters and pass on the values of those RF features to the subsequent module, which is called the Authentication Module.

We make the assumption that the method of feature extraction that is carried out in our module is congruent with the procedures described in [20], [48]. For example, the amplitude and phase mismatch that exists between the in-phase (I) and quadrature (Q) components of the transmitted signal is one of the characteristics that distinguishes one transmitter from another. In the case of a 16-quadrature amplitude modulated (QAM) transmission, these imbalances have the potential to impact the constellation diagram at the Rx. The constellation is also affected by other transmitter variables, such as power-amplifier (PA) back-off and gain variations. The outlying symbols of the constellation are more likely to be impacted by compressive nonlinearity than their more central counterparts.

As a direct consequence of this, the amplitude and phase information for each sign in the constellation needs to be retrieved. Together, these traits and the frequency mistakes produced by each transmitter have the potential to enable one-of-a-kind identification of each transmitter inside the network. As a result, the receiver is able to take the imbalance value and use it as the RF feature. Another illustration of this would be the frequency offset. The transmitter is responsible for the generation of a wide variety of distinctive carrier frequencies due to the presence of several local oscillators within its circuitry. The receiver is able to calculate the frequency offset because it use a reference clock of a very high quality. The offset value is one of a kind and can be implemented as a feature of the RF signal.

Module 2: Authentication. To determine if a node is legitimate or not, the Authentication Module classifies the RF features which is collected from previous module by a statistic theory, particularly the Mahalanobis Distance and Chi-Squared Distribution. The MD is preferred over other distance measures because it has better classification performance [50], [51], [52]. Algorithm 1 outlines the theoretical model used in our approach. When a new node joins the network, the framework computes the Mahalanobis distance between the node and the dataset's mean (represented by the blue diamond shape) to determine its legitimacy, as shown in Figure 1.

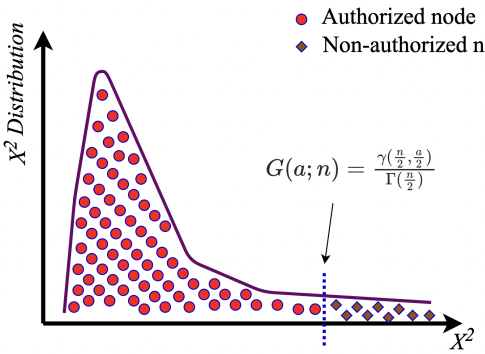


Fig. 3. Mahalanobis Distance and Chi-squared Distribution are employed to detect authentic nodes.

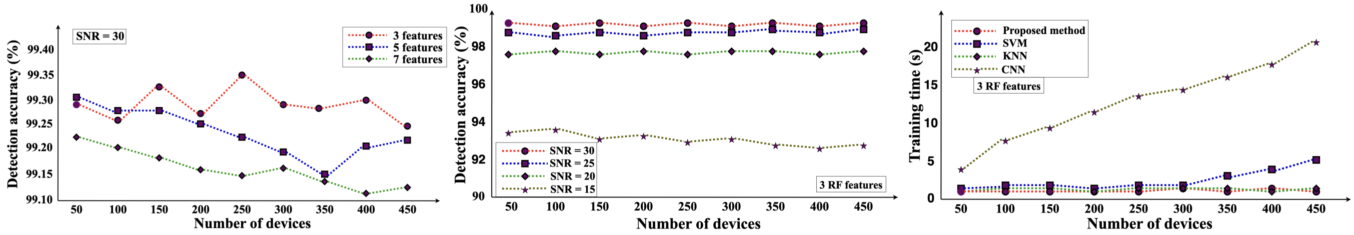
TABLE IV
KEY NOTATIONS IN THE PAPER

Notation	Explanation
Md	the Mahalanobis distance
a	RF feature vector
b	the mean vector
σ	the standard deviation
$d(a,b)$	the Euclidean Distance between vector a and b
Oi	the observed frequency of the i^{th} event
Ei	the expected frequency of the i^{th} event
A	set of observation
i	Mean values vector of the variables
T	Vector in the transposed form
D	the covariance matrix
n	count of RF features
p_{values}	the probability of observing a test statistic
χ^2	the Chi-squared distribution
P	the sum of independent variables squares
Z	the Chernoff bounds
C^{-1}	the Inverse Covariance matrix of independent variables
p_{cutoff}	cut-off value
γ	the lower incomplete gamma function
Γ $G(a;n)$	the gamma function the cumulative distribution Function

The cut-off value (refer to Fig 3 is determined by the CDF, which separates the rejection region (illegitimate nodes) and the sampling distribution. This value is determined by using equation 5, which can be interpreted as "when comparing a node's median distance to the dataset's mean, the legitimacy of the node is validated based on whether or not it is lower than the cut-off value". There are many different algorithms that may be used to measure this, however the Mahalanobis Distance achieves better results when it comes to classification tasks [50], [51], [52]. In other words, the proposed framework performs the same classification tasks as state-of-the-art approaches which employs the machine learning algorithms. However, it significantly improves classification efficiency by reducing the performance metrics such as end-to-end delay, memory and computation overhead.

V. Theoretical Analysis

We tested our model's ability to classify nodes accurately using RF signatures from wireless signals. This method detects



(a)

(b)

(c)

Fig. 4. The result of experiments with the synthet dataset (a) The average detection accuracy of our approach by varying the number of features. (b) The detection accuracy of our method on average is evaluated for different values of SNR and the number of devices. (c) A comparison is made between the proposed method and different machine learning models.

```
Algorithm 1: The process flow of the proposed scheme
Phase 1:
Data: raw data (unprocessed RF signals)
        $n$ (the amount of attributes)
        $a_i$ (the location of the $j$th legitimate IoT device)
        $m$ (the amount of IoT devices)
Outcomes: dataset : $dataset_{n \times j}$
Phase 2:
Data: $dataset_{n \times j}$
        The mean: $l = \sum_{o=1}^{k} a_i/k$
        Cut-off value: $p_{cutoff} = \frac{\gamma(\frac{n}{2},\frac{a}{2})}{\Gamma(\frac{n}{2})}$
        Mahalanobis Distance: $Md_j = (\sqrt{(a-l)^T.C(l)^{-1}.(a-l)}$
if $Md_j <= p_{cutoff}$:
        authentication = TRUE
else :
        authentication = FALSE
Outcomes: authentication
```

unauthorized devices and uses the Mahalanobis distance and chi-squared distribution for classification.

The Mahalanobis distance measures data point deviations from a threshold. Points beyond this threshold are seen as anomalies. This threshold is set using the chi-squared distribution, which compares observed and expected event frequencies. In our approach, this helps distinguish between legitimate and non-legitimate devices. The chi-squared value is derived from the differences between observed and expected frequencies.

$$\chi^2 = \sum (Oi - Ei)^2/Ei \tag{6}$$

where Oi is the observed frequency of the $i-th$ event, Ei is the expected frequency of the $i-th$ event, and $\sum$ denotes the sum over all events.

The chi-squared distribution helps determine the likelihood of a value within its function. A small probability suggests a notable difference between observed and expected frequencies, signaling an anomaly. Using Mahalanobis Distance and Chi-squared Distribution with a set threshold, devices are classified as legitimate or not. Our focus is on the mathematical model used by our proposed algorithm. Given observations $A^n = a_1, a_2, ..., a_k$ in set **A**, where k is the feature count and i is the average, we can compute MD.

$$Md = \sqrt{(a-i)^T.C^{-1}.(a-i)} \tag{7}$$

The inverse covariance matrix C and transpose vector T are used in this scenario. The expression $(a–i)$ represents the distance of the vector from its mean. To standardize the data, the covariance matrix is used, which is a variant of the standard formula for multidimensional data normalization $(x = (a - i)/\sigma)$ (where μ denotes the mean and σ is the standard deviation). The difference between the raw score and the mean is indicated by the absolute value of the variable x. The equation shows that there is an inverse relationship between covariance and distance.

From the equation 7, we have:

$$Md_j = \sqrt{(a_o - i)^T.C^{-1}.(a_o - i)} = f_D(a_j) \tag{8}$$

With the n number of features used in the proposed solution, we have the Chi-Square Distribution denoted as $\chi_n^2(Md_j)$. If the n is a positive integer, so we can calculate the Gamma function as below:

$$\Gamma(\frac{n}{2}) = (\frac{n}{2} - 1)! \tag{9}$$

In case the $n/2$ is a complex numbers with a positive real part, we compute the Gamma function as:

$$\Gamma(\frac{n}{2}) = \int_0^\infty a^{\frac{n}{2}-1} e^{-a}\, da \tag{10}$$

Now, assume γ is the incomplete gamma function, we have:

$$\gamma(d, a) = \int_0^a t^{d-1}\, e^{-y}\, dy. \tag{11}$$

From the equation(3), the CDF $\gamma(\frac{n}{2}; \frac{a}{2})$ can be computed as:

$$\gamma(\frac{n}{2}; \frac{a}{2}) = \int_0^{\frac{a}{2}} t^{\frac{a}{2}-1}\, e^{-y}\, dy. \tag{12}$$

The cut-off value is determined by the CDF, which separates the exclusion zone (illegitimate nodes) from the legitimate nodes (see Fig. 3). If the MD of a node relative to the mean of the dataset is lower than the cut-off value, then the node in question is regarded to be valid. The proposed method's high-level workflow is depicted in Algorithm 1. In summary, the chi-squared distribution provides a statistical framework for detecting anomalies in data sets by comparing the observed and expected frequencies of events. By calculating the chi-squared statistic and comparing it to a threshold based on the chi-squared distribution, anomalies can be identified. The anomalies means that the fake RF signatures can be detected. As discussed previously, the RF signatures are nearly impossible to mimic and can be represented as device identifiers.

Therefore, the device identifier can be verified to determine the genuine device and mitigate the impersonate attack.

VI. Performance evaluation

In this section of the article, we have evaluated our methodology by using both synthetic and real dataset in our evaluations.

A. Evaluation using synthetic data set

Firstly to generate the dataset, we employ the MATLAB simulink package known as the Wireless Waveform Generator toolbox. We use 450 testing devices, and for each device, the frequency, amplitude, and phase have been subtly altered so that we can simulate the non-idealities of RF features properties. For each device we collect one hundred RF signal data. The RF features that we used in our study are mentioned in table III (including their mean and standard deviation measurements). The processed dataset can be accessed through the GitHub link provided below [1].

It can be seen in Fig. 4(a), that at three features the model gives the highest level of detection accuracy (99.35%). It is clear that there is not a significant reduction in accuracy when there is an increase in the number of features (from three to seven). We gradually added more devices from 50 to 450 to investigate the impact of varying SNR levels and evaluate the detection accuracy across SNR values from 15dB to 30dB. Our experiment utilized three features: CFO, Amplitude Mismatch, and Phase Offset. The results are illustrated in Fig. 4(b). While the number of devices increases, the detection accuracy remains stable, but alterations in SNR values affect it. Our solution's training duration was calculated utilizing MD and Chi-square distribution theories and compared to that of other machine learning models. We found that our approach outperforms state-of-the-art methods, as shown in Fig.4(c) and TableV. Table VI demonstrates the confusion matrix.

Table V provides a comparison of our approach with the most recent research. Our method achieved the highest detection accuracy of 99.35%. Overall, our findings indicate that the proposed approach is comparable to those of previous works. To determine the node authentication time, we incorporated training time as a performance metric in addition to accuracy versus the number of IoT nodes. We evaluated the training time required for different number of transmitter using the same test settings (three RF characteristics and a dataset with a signal-to-noise ratio of 30) and compared it to other RF fingerprinting methods. The results are presented in Fig. 4 (c). In the setting of the next generation of 5G Internet of Things, in which the Internet of Things devices frequently added or removed from the network, the measure is crucial. The lesser training time of the model is highly needed as it impacts the time for IoT node authentication in large-scale mobile networks.

B. Evaluation using real RF fingerprinting data set

Furthermore, we have validated our approach using the real data-set collected by the Institute for the Wireless Internet of Things (WIOT) [2]. The transmitter was configured to emit Wi-Fi signals from 20 Software Defined Radios (SDRs) provided by National Instruments and powered by Gnuradio. On the receiver side, a second Gnuradio is used to capture the I/Q signal values. Using a single receiver, the data is captured. The data is received as unprocessed I/Q samples prior to the Wi-Fi demodulation operation. Analyzing RF fingerprinting using such a wide variety of I/Q samples allow researchers to determine the impact of each stage of communication/processing on the accuracy of fingerprinting based authentication schemes. In this dataset, the authors employed the "Arena Wireless Different Antenna" setup from Day 1. BPSK modulation was used, along with the same IEEE 802.11 a/g (2.432 GHz and 20 MS/s) transmission parameters (Wi-Fi with 20 MS/s and 2.432 GHz), and the Ettus VERT2450 antenna was used. Our experiment involved converting the I/Q raw data into actual values and collecting one hundred thousand data points from each device. Therefore, the dataset has a $20 \times 100,000$ matrix.

TABLE V
A comparison was made between our method and the existing approaches

Reference and ML Model	MDA/ML [41]	CNN [30]	SVM [32]	KNN [33]	LSVM [31]	Proposed method
Highest detection accuracy (%)	95%	97%	97%	98.13%	98.8%	99.35%

TABLE VI
High classification metrics were obtained with the proposed method for 450 devices and SNR=30

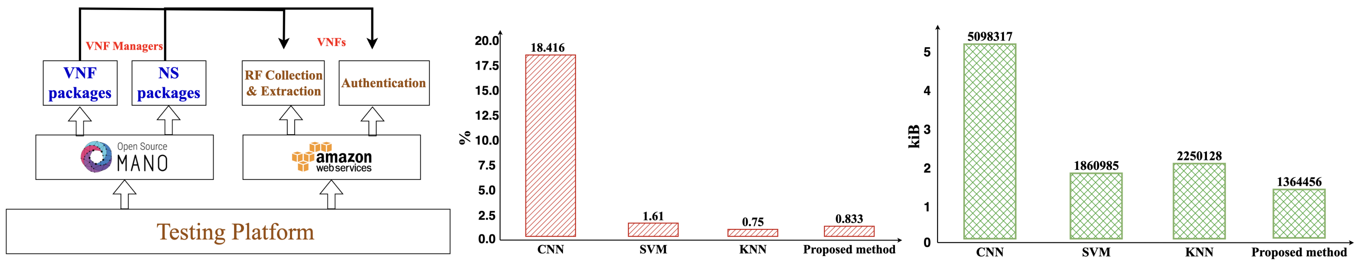
Metric	Proposed method	CNN	SVM	KNN
Accuracy score	99.35%	98.4%	91.38%	93.31%
F1-score	99.67%	98.42%	95.46%	96.53%
Precision	99.36%	98.40%	99.69%	96.21%
Recall	99.99%	98.46%	91.58%	96.86%

Using this real data set, we have compared our work with existing nine other works. The metrics in the evaluations are: confusion matrix, CPU Utilization, training time and the detection time. In terms of the detection accuracy which is the most important metric for the access control system, our method outperforms other existing works. It can be explained that in specific scenarios that statistical method can outperform the machine learning in terms of classifications [53], [54], [55]. Moreover, in our method, we employ the Mahalanobis Distance which has some advantages than other distance techniques in classification task [50], [51], [52]. The proposed method not only achieves the highest recognition accuracy at 99.2% but also better in comparison to the other methods w.r.t the confusion matrix parameters. The details of the confusion matrix is given in the Table VII. Further, the proposed method has the CPU Utilization at 8.6% and this has shown a better result in the evaluations.

The Autoencoder (AE) used by authors of [45] consumes the most CPU during testing scenario. It is proven that our method takes less CPU resources for the authentication process. Moreover, the LSTM models [46] takes 26.123 seconds to finish the training process while our proposed method only needs 0.0809 seconds for the whole training process. Besides,

[1] https://github.com/ndducnha/mahalanobis_dataset

[2] https://repository.library.northeastern.edu/collections/neu:gm80kf51n

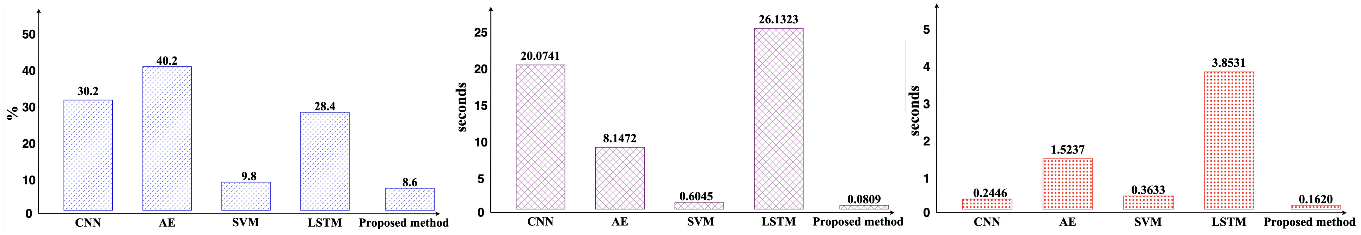


(a) (b) (c)

Fig. 5. (a) Overview of 5G ETSI-MANO platform in the testing platform. (b) CPU utilization comparisons (c) Memory usage comparisons.

TABLE VII
A COMPARISON OF THE CLASSIFICATION METRICS OF THE PROPOSED APPROACH AND OTHER WORKS WITH THE REAL DATASET

Metric	CNN [38]	ANN [44]	AE [45]	SVM [32]	LSTM [46]	CNN [39]	CNN [25]	KNN [42]	RF [43]	Proposed method
Accuracy score	86.84%	90.72%	91.52%	92.12%	93.21%	93.38%	94.39%	93.48%	95%	99.2%
F1-score	85.42%	94.96%	95.56%	95.89%	96.24%	93.07%	97.07%	96.63%	97.44%	99.58%
Precision	85.03%	95.09%	94.66%	95.16%	94.98%	94.96%	97.71%	95%	95%	99.16%
Recall	79.05%	94.42%	94.96%	96.65%	98.08%	90.5%	99.36%	98.25%	95%	99.16%



(a) (b) (c)

Fig. 6. Using a real dataset, we assess the proposed method against various machine learning models in (a) CPU Ultilization. (b) Training time. (c) Detection time.

end-to-end delay is the important metric in the next Generation Networks and we also conducted experiments to show the comparisons. The CNN [38] and SVM [32] require 0.2446 and 0.3633 seconds respectively to provide the authentication decisions. However, our proposal only needs 0.1620 seconds to make the authentication decision. All metric comparisons are given in the Fig 6. Therefore, for the testing with the real dataset, it is confirmed that our proposed method outperforms other existing works.

C. Proof of Concept (PoC) of the Real-time deployment of our approach

In real-life deployment, the 5G platform is often implemented using virtualized network functions (VNFs) [56]. In this project we employ OSM-MANO [57] which is developed by ETSI Industry Specification Group for Network Functions Virtualization standards for the purpose of evaluating the technique's performance on a real-world 5G network. The 5G platform can be deployed on the Cloud platforms. There are many enterprise cloud services which can be used to implement the 5G OSM-MANO. In this experiment, we choose Amazon Web Services (AWS). A high level experimental set up illustration is shown in Fig. 5(a). The classification module, which serves as the Authentication VNF, is implemented on the AWS platform and is integrated with OSM version 10, using Amazon EC2. The EC2 provides 4 virtual CPUs and 8GB of RAM, while OSM is connected to the Virtualized Infrastructure represented by AWS. A VNF with 8 virtual CPUs and 16GB of RAM is deployed by OSM, both of which are powered by Intel Xeon processors. The study employed Python 3.6.9, Keras 2.8.0, and TensorFlow 1.14.0. The evaluation findings are detailed in Fig. 5 (b) and (c).

In order to demonstrate the efficacy of our strategy in real time, we have carried out further evaluation to analyse the CPU and Memory Consumption. In this experiment we compare our model with SVM, KNN, and CNN, we employ 450 devices with an SNR of 30dB. As indicated in Figure 5, our approach achieved a detection accuracy of 98.4%, which is higher compared to 91.38% and 93.31% for SVM and KNN, respectively. Our approach uses the least amount of RAM compared to other works. In comparison to other models, our approach and KNN make significantly less use of the computer's central processing unit (CPU). The distinction between the CPU metrics of our approach and KNN is minor, with only 0.83% and 0.75% deviation, respectively.

VII. FORMAL VERIFICATION

To demonstrate the proof of concept for the proposed method, we have validated it using the formal analysis technique. Formal analysis is a popular method for testing the security aspects of a protocol [58], in this verification, we

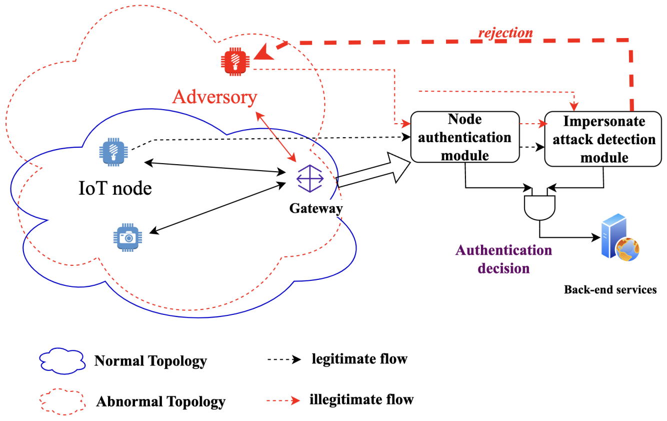


Fig. 7. The high-level of architecture of the impersonation attack framework.

use the Tamarin prover [59]. The Tamarin prover is a state-of-the-art tool designed for the formal analysis of security protocols. With its interactive theorem prover, Tamarin facilitates the analysis of security properties of protocols against a broad spectrum of threat models. Moreover, the Tamarin prover offers a robust framework for the formal verification of security protocols. By defining function symbols, rules, and lemmas, users can model real-world protocols and verify their security properties against various threat models. The utilization of Tamarin ensures that the protocols are not only theoretically sound but also practically secure against potential adversaries. In our analysis, we use two lemmas: `no_impersonation` to prove the prevention of impersonation attacks and `no_mitm` to ensure that the proposed method can prevent man-in-the-middle attacks.

In the Tamarin prover, protocols are defined using a multiset rewriting system. This system facilitates the definition of rules that describe the transmission, reception, and processing of messages. Each rule can have preconditions (the left-hand side) and postconditions (the right-hand side), representing the state before and after the rule's application. The full code of the Tamarin prover that we use can be found below:

```
theory Next_Generation_Node_Authentication

begin

// Function symbols
functions: Authenticate/1, ComputeMD/1, ComputeChiSquared
     /1, Decision/2

// Rules
rule send_request:
[Fr(T),Fr(R),Fr(~rf_signature), !Sent(T,R,~rf_signature)]
--[Send_Request(T,R,~rf_signature)]->
[Out(<T,R,Authenticate(~rf_signature)>), !Sent(T,R,~
     rf_signature)]

rule compute_values:
[In(<T,R,Authenticate(~rf_signature)>),Fr(~rf_signature)]
--[Compute_Values(R,T,~rf_signature)]->
[Out(<R,T,ComputeMD(~rf_signature)>),Out(<R,T,
     ComputeChiSquared(~rf_signature)>)]

rule make_decision:
[In(<R,T,ComputeMD(~rf_signature)>),In(<R,T,
     ComputeChiSquared(~rf_signature)>),Fr(~rf_signature)]
--[Make_Decision(R,T,~rf_signature)]->
[Out(<R,T,Decision(ComputeMD(~rf_signature),
     ComputeChiSquared(~rf_signature))>)]

// Lemmas
lemma no_impersonation:
"All R T sig #i. Send_Request(T, R, sig)@i ==> Ex #j. #j <
     #i & Send_Request(T, R, sig)@j"

lemma no_mitm:
"All R T1 T2 sig #i #j. Send_Request(T1, R, sig)@i &
     Send_Request(T2, R, sig)@j ==> T1 = T2"

end
```

Firstly, we introduce the function symbols which we use in our analysis:

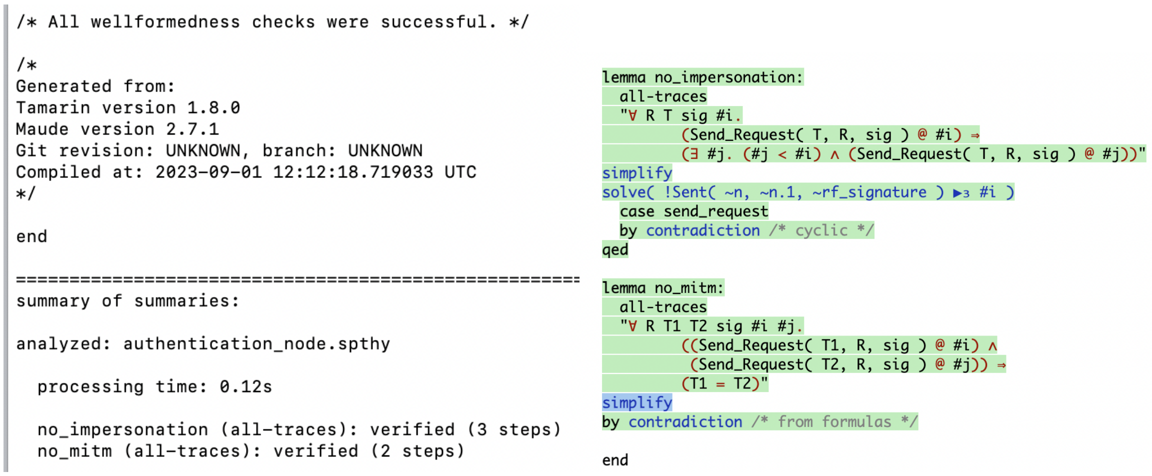
- `Authenticate`: Represents the authentication process.
- `ComputeMD`: Represents the computation of the Mahalanobis Distance calculation.
- `ComputeChiSquared`: Represents a chi-squared computation.
- `Decision`: Represents an authentication decision making based on the results of `ComputeMD` and `ComputeChiSquared`.

In Tamarin Prover, ”rules” define the behavior of the protocol. Each rule has a name, preconditions, actions, and postconditions. For instance, the `send_request` rule describes the process of sending a request. In our model, we define three primary rules to represent the authentication process using RF fingerprinting characteristics. The send_request rule captures the initiation phase, where a transmitter T sends an authentication request to a receiver R with a unique RF signature. This rule ensures that each request is unique and hasn't been sent before. Upon receiving this request, the compute_values rule is triggered. Here, the receiver R computes two significant values: the Mahalanobis Distance and the Chi-square distribution, both derived from the received RF signature. These computations serve as the foundation for the subsequent decision-making process. Finally, the make_decision rule encapsulates the decision-making phase. Based on the previously computed values, the receiver R determines the legitimacy of the transmitter T and sends out a decision. This decision is a culmination of the computed Mahalanobis Distance and Chi-square distribution values, ensuring a robust authentication process for IoT nodes.

Lemmas state properties or assertions about the protocol that we aim to prove. In our verification process, we have two lemmas:

1) `no_impersonation`: This rule is our main guard against fake requests. It says that if a device `T` sends a message to another device `R` with a certain signature at a certain time, there should be an earlier time when the same message was sent. This helps us spot and stop any fake or copied messages.
2) `no_mitm`: This lemma asserts that if two terminals, `T1` and `T2`, send a request to the same receiver `R` with the same signature `sig`, then `T1` and `T2` must be identical, ensuring no man-in-the-middle attacks.

The results of the formal verification can be found in Fig. 8. The results clearly prove that the proposed method has passed the formal verification analysis, protecting the authentication decision-making process from impersonation attacks and man-in-the-middle attacks.



```
/* All wellformedness checks were successful. */

/*
Generated from:
Tamarin version 1.8.0
Maude version 2.7.1
Git revision: UNKNOWN, branch: UNKNOWN
Compiled at: 2023-09-01 12:12:18.719033 UTC
*/

end

=====================================================
summary of summaries:

analyzed: authentication_node.spthy

  processing time: 0.12s

  no_impersonation (all-traces): verified (3 steps)
  no_mitm (all-traces): verified (2 steps)
```

(a)

```
lemma no_impersonation:
  all-traces
  "∀ R T sig #i.
        (Send_Request( T, R, sig ) @ #i) ⇒
        (∃ #j. (#j < #i) ∧ (Send_Request( T, R, sig ) @ #j))"
simplify
solve( !Sent( ~n, ~n.1, ~rf_signature ) ▶₃ #i )
  case send_request
  by contradiction /* cyclic */
qed

lemma no_mitm:
  all-traces
  "∀ R T1 T2 sig #i #j.
        ((Send_Request( T1, R, sig ) @ #i) ∧
         (Send_Request( T2, R, sig ) @ #j)) ⇒
        (T1 = T2)"
simplify
by contradiction /* from formulas */

end
```

(b)

Fig. 8. The results of formal verification are: (a) Summary of analysis: The lemmas ”no_impersonation” and ”no_mitm” have been successfully verified. (b) The lemmas turned green (indicating successful verification) in interactive mode.

VIII. Conclusion and Future work

This paper introduces a new approach for authenticating IoT nodes using RF fingerprinting characteristics. In contrast with other existing RF fingerprinting methods, we proposed the use of the Mahalanobis Distance and Chi-square distribution theories to determine the legitimate and non-legitimate devices. The proposed scheme has been tested with both synthetic and real dataset and the testing experiments are implemented into the ETSI-NFV 5G standard architecture. The testing results show that our approach outperforms other existing works in terms of detection accuracy, resource computations and the latency. The threat model has been discussed and the work also analyzed the proof of concept as well as the security analysis by formal verification tool. In the future work, we will expand the testing to cover the impact of the environmental changes to the proposed framework. We also plan to evaluate the proposed method under various scenarios such as location changing of observed devices. We note that having a unique identifier for each IoT device can be useful for certain security mechanisms such as authentication and identification. However, it is not a complete solution to all security problems. To cover more security aspects, the future work will focus on the integration of the proposed framework with the trust model to enhance the security level.

References

[1] 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 Network*, vol. 35, no. 2, pp. 34–39, 2021.

[2] M. Alshaikhli, T. Elfouly, O. Elharrouss, A. Mohamed, and N. Ottakath, “Evolution of internet of things from blockchain to iota: A survey,” *IEEE Access*, vol. 10, pp. 844–866, 2022.

[3] Y. Sasaki, “A survey on iot big data analytic systems: Current and future,” *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 1024–1036, 2022.

[4] M. A. Al-Garadi, A. Mohamed, A. K. Al-Ali, X. Du, I. Ali, and M. Guizani, “A survey of machine and deep learning methods for internet of things (iot) security,” *IEEE Communications Surveys Tutorials*, vol. 22, no. 3, pp. 1646–1685, 2020.

[5] K. Sood, S. Yu, D. D. N. Nguyen, Y. Xiang, B. Feng, and X. Zhang, “A tutorial on next generation heterogeneous iot networks and node authentication,” *IEEE Internet of Things Magazine*, vol. 4, no. 4, pp. 120–126, 2021.

[6] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, “On multi-access edge computing: A survey of the emerging 5g network edge cloud architecture and orchestration,” *IEEE Communications Surveys Tutorials*, vol. 19, no. 3, pp. 1657–1681, 2017.

[7] A. Fotovvat, G. M. E. Rahman, S. S. Vedaei, and K. A. Wahid, “Comparative performance analysis of lightweight cryptography algorithms for iot sensor nodes,” *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 8279–8290, 2021.

[8] P. Gope, B. Sikdar, and O. Millwood, “A scalable protocol level approach to prevent machine learning attacks on puf-based authentication mechanisms for internet-of-medical-things,” *IEEE Transactions on Industrial Informatics*, pp. 1–1, 2021.

[9] J. Qiu, Z. Tian, C. Du, Q. Zuo, S. Su, and B. Fang, “A survey on access control in the age of internet of things,” *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 4682–4696, 2020.

[10] V. Hassija, V. Chamola, V. Saxena, D. Jain, P. Goyal, and B. Sikdar, “A survey on iot security: Application areas, security threats, and solution architectures,” *IEEE Access*, vol. 7, pp. 82 721–82 743, 2019.

[11] F. Hussain, R. Hussain, S. A. Hassan, and E. Hossain, “Machine learning in iot security: Current solutions and future challenges,” *IEEE Communications Surveys Tutorials*, vol. 22, no. 3, pp. 1686–1721, 2020.

[12] B. Bera, A. K. Das, S. Garg, M. Jalil Piran, and M. S. Hossain, “Access control protocol for battlefield surveillance in drone-assisted iot environment,” *IEEE Internet of Things Journal*, vol. 9, no. 4, pp. 2708–2721, 2022.

[13] A. K. Das, B. Bera, M. Wazid, S. S. Jamal, and Y. Park, “On the security of a secure and lightweight authentication scheme for next generation iot infrastructure,” *IEEE Access*, vol. 9, pp. 71 856–71 867, 2021.

[14] T. Alladi, V. Venkatesh, V. Chamola, and N. Chaturvedi, “Drone-map: A novel authentication scheme for drone-assisted 5g networks,” in *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2021, pp. 1–6.

[15] K. Sood, D. d. n. Nguyen, M. R. Nosouhi, N. Kumar, F. Jiang, M. Chowdhury, and R. Doss, “Performance evaluation of a novel intrusion detection system in next generation networks,” *IEEE Transactions on Network and Service Management*, pp. 1–1, 2023.

[16] N. Xie, H. Tan, L. Huang, and A. X. Liu, “Physical-layer authentication in wirelessly powered communication networks,” *IEEE/ACM Transactions on Networking*, vol. 29, no. 4, pp. 1827–1840, 2021.

[17] Y. Wang, J. Jin, Y. Li, and C. Choi, “A reliable physical layer authentication algorithm for massive iot systems,” *IEEE Access*, vol. 8, pp. 80 684–80 690, 2020.

[18] N. Zhang, X. Fang, Y. Wang, S. Wu, H. Wu, D. Kar, and H. Zhang, “Physical-layer authentication for internet of things via wfrft-based gaussian tag embedding,” *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 9001–9010, 2020.

[19] S. A. Chaudhry, K. Yahya, F. Al-Turjman, and M.-H. Yang, "A secure and reliable device access control scheme for iot based sensor cloud systems," *IEEE Access*, vol. 8, pp. 139 244–139 254, 2020.

[20] 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*, vol. 6, no. 1, pp. 388–398, 2019.

[21] W. Iqbal, H. Abbas, M. Daneshmand, B. Rauf, and Y. A. Bangash, "An in-depth analysis of iot security requirements, challenges, and their countermeasures via software-defined security," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 10 250–10 276, 2020.

[22] K. Huang, L.-X. Yang, X. Yang, Y. Xiang, and Y. Y. Tang, "A low-cost distributed denial-of-service attack architecture," *IEEE Access*, vol. 8, pp. 42 111–42 119, 2020.

[23] Y. Tu, Z. Zhang, Y. Li, C. Wang, and Y. Xiao, "Research on the internet of things device recognition based on rf-fingerprinting," *IEEE Access*, vol. 7, pp. 37 426–37 431, 2019.

[24] M. Köse, S. Taşcioğlu, and Z. Telatar, "Rf fingerprinting of iot devices based on transient energy spectrum," *IEEE Access*, vol. 7, pp. 18 715–18 726, 2019.

[25] J. Zhang, R. Woods, M. Sandell, M. Valkama, A. Marshall, and J. Cavallaro, "Radio frequency fingerprint identification for narrowband systems, modelling and classification," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 3974–3987, 2021.

[26] W. Nie, Z.-C. Han, M. Zhou, L.-B. Xie, and Q. Jiang, "Uav detection and identification based on wifi signal and rf fingerprint," *IEEE Sensors Journal*, vol. 21, no. 12, pp. 13 540–13 550, 2021.

[27] D. D. N. Nguyen, K. Sood, Y. Xiang, L. Gao, L. Chi, and S. Yu, "Towards iot node authentication mechanism in next generation networks," *IEEE Internet of Things Journal*, pp. 1–1, 2023.

[28] D. D. N. Nguyen, K. Sood, M. R. Nosouhi, Y. Xiang, L. Gao, and L. Chi, "Rf fingerprinting-based iot node authentication using mahalanobis distance correlation theory," *IEEE Networking Letters*, vol. 4, no. 2, pp. 78–81, 2022.

[29] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, "Kitsune: an ensemble of autoencoders for online network intrusion detection," *arXiv preprint arXiv:1802.09089*, 2018.

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

[31] A. Aghnaiya, A. M. Ali, and A. Kara, "Variational mode decomposition-based radio frequency fingerprinting of bluetooth devices," *IEEE Access*, vol. 7, pp. 144 054–144 058, 2019.

[32] 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 *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, 2020, pp. 258–270.

[33] M. Ezuma, F. Erden, C. Kumar Anjinappa, O. Ozdemir, and I. Guvenc, "Detection and classification of uavs using rf fingerprints in the presence of wi-fi and bluetooth interference," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 60–76, 2020.

[34] W. Jian, Y. Zhou, and H. Liu, "Lightweight convolutional neural network based on singularity roi for fingerprint classification," *IEEE Access*, vol. 8, pp. 54 554–54 563, 2020.

[35] L. Zong, C. Xu, and H. Yuan, "A rf fingerprint recognition method based on deeply convolutional neural network," in *2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC)*, 2020, pp. 1778–1781.

[36] Y. Li, Y. Lin, Z. Dou, and Y. Chen, "Research on rf fingerprint feature selection method," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, 2020, pp. 1–5.

[37] G. Bovenzi, G. Aceto, D. Ciuonzo, V. Persico, and A. Pescapé, "A hierarchical hybrid intrusion detection approach in iot scenarios," in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, 2020, pp. 1–7.

[38] A. Al-Shawabka, F. Restuccia, S. D'Oro, T. Jian, B. Costa Rendon, N. Soltani, J. Dy, S. Ioannidis, K. Chowdhury, and T. Melodia, "Exposing the fingerprint: Dissecting the impact of the wireless channel on radio fingerprinting," in *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications*, 2020, pp. 646–655.

[39] K. Sankhe, M. Belgiovine, F. Zhou, L. Angioloni, F. Restuccia, S. D'Oro, T. Melodia, S. Ioannidis, and K. Chowdhury, "No radio left behind: Radio fingerprinting through deep learning of physical-layer hardware impairments," *IEEE Transactions on Cognitive Communications and Networking*, vol. 6, no. 1, pp. 165–178, 2020.

[40] B. Li and E. Cetin, "Waveform domain deep learning approach for rf fingerprinting," in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2021, pp. 1–5.

[41] 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 Transactions on Information Forensics and Security*, vol. 16, pp. 2381–2396, 2021.

[42] D. D. Nguyen and M. Thuy Le, "Enhanced indoor localization based ble using gaussian process regression and improved weighted knn," *IEEE Access*, vol. 9, pp. 143 795–143 806, 2021.

[43] C. Jain, G. V. S. Sashank, V. N, and S. Markkandan, "Low-cost ble based indoor localization using rssi fingerprinting and machine learning," in *2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, 2021, pp. 363–367.

[44] M. Nair, T. Cappello, S. Dang, V. Kalokidou, and M. A. Beach, "Rf fingerprinting of lora transmitters using machine learning with self-organizing maps for cyber intrusion detection," in *2022 IEEE/MTT-S International Microwave Symposium - IMS 2022*, 2022, pp. 491–494.

[45] M. R. Nosouhi, K. Sood, M. Grobler, and R. Doss, "Towards spoofing resistant next generation iot networks," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 1669–1683, 2022.

[46] S. Hu, K. He, X. Yang, and S. Peng, "Bluetooth fingerprint based indoor localization using bi-lstm," in *2022 31st Wireless and Optical Communications Conference (WOCC)*, 2022, pp. 161–165.

[47] X.-Y. Liu and X. Wang, "Real-time indoor localization for smartphones using tensor-generative adversarial nets," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 8, pp. 3433–3443, 2021.

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

[49] F. Xiangning, S. Yutao, and F. Yangyang, "An efficient cmos dc offset cancellation circuit for pga of low if wireless receivers," in *2010 International Conference on Wireless Communications & Signal Processing (WCSP)*, 2010, pp. 1–5.

[50] L. Yang, Y. Li, J. Wang, and N. N. Xiong, "Fslm: An intelligent few-shot learning model based on siamese networks for iot technology," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 9717–9729, 2021.

[51] L. Friedman and O. V. Komogortsev, "Assessment of the effectiveness of seven biometric feature normalization techniques," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 10, pp. 2528–2536, 2019.

[52] D. Das and C. S. G. Lee, "A two-stage approach to few-shot learning for image recognition," *IEEE Transactions on Image Processing*, vol. 29, pp. 3336–3350, 2020.

[53] L. Senigagliesi, M. Baldi, and E. Gambi, "Statistical and machine learning-based decision techniques for physical layer authentication," in *2019 IEEE Global Communications Conference (GLOBECOM)*, 2019, pp. 1–6.

[54] M. Senigagliesi and E. Gambi, "Comparison of statistical and machine learning techniques for physical layer authentication," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1506–1521, 2021.

[55] A. Xuan, M. Yin, Y. Li, X. Chen, and Z. Ma, "A comprehensive evaluation of statistical, machine learning and deep learning models for time series prediction," in *2022 7th International Conference on Data Science and Machine Learning Applications (CDMA)*, 2022, pp. 55–60.

[56] I. Sarrigiannis, K. Ramantas, E. Kartsakli, P.-V. Mekikis, A. Antonopoulos, and C. Verikoukis, "Online vnf lifecycle management in an mec-enabled 5g iot architecture," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4183–4194, 2020.

[57] "OSM-MANO," https://osm.etsi.org//. Accessed: 28 March 2023.

[58] S. B. Ram and V. Odelu, "Security analysis of a key exchange protocol under dolev-yao threat model using tamarin prover," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, 2022, pp. 0667–0672.

[59] S. Meier, B. Schmidt, C. Cremers, and D. Basin, "The tamarin prover for the symbolic analysis of security protocols," in *Computer Aided Verification*. Springer, 2013, pp. 696–701. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-642-39799-8_48