

Article

An Identity Authentication Method of a MIoT Device Based on Radio Frequency (RF) Fingerprint Technology

Qiao Tian ^{1,2}, Yun Lin ² , Xinghao Guo ², Jin Wang ^{3,4,*} , Osama AlFarraj ⁵  and Amr Tolba ^{5,6} 

¹ College of Computer Science and Technology, Harbin Engineering University, Harbin 150001, China; tianqiao@hrbeu.edu.cn

² College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China; linyun@hrbeu.edu.cn (Y.L.); s317080023@hrbeu.edu.cn (X.G.)

³ Hunan Provincial Key Laboratory of Intelligent Processing of Big Data on Transportation, School of Computer & Communication Engineering, Changsha University of Science & Technology, Changsha 410004, China

⁴ School of Information Science and Engineering, Fujian University of Technology, Fujian 350118, China

⁵ Computer Science Department, Community College, King Saud University, Riyadh 11437, Saudi Arabia; oalfarraj@ksu.edu.sa (O.A.); atolba@ksu.edu.sa (A.T.)

⁶ Mathematics and Computer Science Department, Faculty of Science, Menoufia University, Shebin-El-kom 32511, Egypt

* Correspondence: jinwang@csust.edu.cn

Received: 24 December 2019; Accepted: 19 February 2020; Published: 22 February 2020



Abstract: With the continuous development of science and engineering technology, our society has entered the era of the mobile Internet of Things (MIoT). MIoT refers to the combination of advanced manufacturing technologies with the Internet of Things (IoT) to create a flexible digital manufacturing ecosystem. The wireless communication technology in the Internet of Things is a bridge between mobile devices. Therefore, the introduction of machine learning (ML) algorithms into MIoT wireless communication has become a research direction of concern. However, the traditional key-based wireless communication method demonstrates security problems and cannot meet the security requirements of the MIoT. Based on the research on the communication of the physical layer and the support vector data description (SVDD) algorithm, this paper establishes a radio frequency fingerprint (RFF or RF fingerprint) authentication model for a communication device. The communication device in the MIoT is accurately and efficiently identified by extracting the radio frequency fingerprint of the communication signal. In the simulation experiment, this paper introduces the neighborhood component analysis (NCA) method and the SVDD method to establish a communication device authentication model. At a signal-to-noise ratio (SNR) of 15 dB, the authentic devices authentication success rate (ASR) and the rogue devices detection success rate (RSR) are both 90%.

Keywords: RF fingerprint; identity authentication; mobile Internet of Things; feature extraction

1. Introduction

The Internet of Things is the “internet connected to everything”, which is an extended and expanded network based on the Internet. It combines various information sensor devices with the internet and forms a huge network to realize the interconnection of people, machines and things at any time and any place [1]. In the 21st century, the internet, new energy, new materials, and biotechnology are forming huge industrial capacities and markets very quickly, and this will elevate the entire social

production system to a new level and promote a new industrial revolution. In 1999, American Auto-ID first put forward the concept of the “Internet of Things”, which is mainly based on item coding, RFID technology, and the internet [2]. Some experts and scholars have put forward the concept of the mobile Internet of Things (MIoT) on the basis of integrating mobile devices and Internet of Things. In the MIoT’s communication mode, wireless communication is very important.

In the field of MIoT research, many scholars have focused on improving the transmission performance of the network [3–5], and this article focuses on the security of the network. There are many ways to improve network security, including radio frequency (RF) fingerprinting technology. A RF fingerprint refers to the difference of the transmitters due to the production, processing, and debugging. The signal received from the transmitters can be used to extract the difference and to realize the individual identification of wireless devices. The process of extracting signal differences is called RF fingerprint extraction [6]. RF fingerprint authentication technology aims at distinguishing authorized transmitters of multifarious users based on the unique feature from their radio frequency signals at the physical layer [7]. The core is that only authorized users can be allowed to intervene in the network, which improves the network security to a certain extent. RF fingerprint techniques have been used in many systems such as intrusion detection systems, radar systems, satellite communication systems, Internet of Things systems, and network security in 4G and 5G networks.

Artificial intelligence is a concept that began to be widely discussed in 1956. With the continuous development of artificial intelligence technology, machine learning and deep learning gradually occupy all aspects of people’s lives. In the field of communication, the development of artificial intelligence is also very rapid, including wireless sensor networks [8–11], network resource allocation [12], modulation signal recognition [13–15] communication equipment individual recognition [16], abnormal information identification and detection [17,18], and others. Many scholars have conducted studies in these aspects. This paper studies the use of artificial intelligence algorithms to solve identity authentication problems in the MIoT. Generally, the process of RFF technology contains 4 primary procedures: signal processing and acquisition, features extraction from the gained signals, features matching with dataset of fingerprint reference, and allocating the best matching collection for those features. In the final phase, if there is no appropriate class, signals under check ought to see as malicious signals of illegal transmitters and banned from network access.

The main purpose of the research in this paper is to establish the radio frequency authentication model and the radio frequency fingerprint database through the research on radio frequency fingerprint technology, to effectively distinguish between authentic devices and rogue devices in wireless communication. Based on our recent research on the application of the RF fingerprint authentication model to MIoT, we arrange the structure of the paper as follows: in Section 2, we systematically summarize and classify the research on the RF fingerprint authentication model and feature extraction algorithm from the past few years. Then, in Section 3, we list the principles and details of some methods used in this paper. In Section 4, we conduct simulation experiments on mobile device authentication models based on measured data and evaluate their performance. Finally, we summarize the whole thesis in Section 5.

2. Related Work

Based on the research of RF fingerprint related fields, this paper focuses on two important aspects: (1) the RF fingerprint feature extraction of wireless devices; (2) RF fingerprint authentication. The radio frequency fingerprint feature extraction is to perform feature transformation and feature extraction on the identifiable part of the communication signal, and convert the communication signal into a radio frequency fingerprint feature that can be identified.

2.1. RF Fingerprint Feature Extraction

The following is a summary of the recent work on RF fingerprint feature extraction methods for wireless devices:

Existing methods for the RF fingerprint extraction contain numerical feature extraction established on signal constraints, based on signal transform domain, based on nonlinear characteristic of transmitter and based on the methods used in image processing. The feature extraction method based on the numerical features of the signal refers to finding a linear or nonlinear conversion to reveal the essential structure of the preprocessed signal, and project the imaginative signal into the unique feature space, this can decrease the dimension of the source signal, and reduce the problem of over fitting with classifiers. The common signal factors consist of time domain parameters, high order spectral parameters, domain frequency parameters, high order moments parameters, and so on. In collected works [19,20], researchers identify the transient signals amplitude by using IEEE 802.11 and Bluetooth devices. The detection rate average can extent to 95%, however the complexity time is relatively high. In [21], the I/Q modulation domain imbalance is used as a RF fingerprint. The idea is constructed on the authentication of the 2 parameters assumption test and the probability ratio test.

The final simulation result proved the correctness of the idea. In the year of 2015, the researcher of [22] take out the information phase of the baseband signal by straining the signal with alike model from the similar manufacturer, and then used it as the RF fingerprint. The results display that information phase be able to use to categorize devices, but the classification performance will change owing to the difference of the channel distance. The accuracy average of classification in distance with short range is about 99.6%, however the accuracy average of classification reduced to the level 81.9% after the channel distance becomes elongated. The entropy represents the internal chaos degree in a system. Whenever the chaos increased, the entropy gets a higher value. The characteristics of entropy are frequently used features in RF fingerprint. For instance, the authors of [23] proposed recognition method of a new fingerprint established on multi-dimensional permutation entropy. According to this experiment, the transceiver distance is fixed to 10 meters, thus the signal can be propagated in the channel of short-wave line of sight. Experimental results show that the method is efficient.

On behalf of weak dissimilarities among like devices, transmitter hardware internal noise and nonlinearities be able to produce spurious modules at the signal receiving. Most of all signal modules are non-Gaussian and non-stable, consequently numerical analysis approaches for frequency domain and time domain parameters might no longer be appropriate. Therefore, scholars gradually use methods of signal processing for analyzing signals, convert them into assured transform domains, and then process and analyze them. Including analysis of wavelet, analysis of time frequency, fractal features, transformations of empirical mode decomposition (EMD), and the transformations of intrinsic time decomposition (ITD), etc. In the late of 1998, Huang et al. [24] introduced a method of data analysis constructed on EMD, which be able for generating a set of intrinsic mode functions(IMF). The transformation of Hilbert is able to use for deriving instantaneous frequency and local energy from IMF for obtaining a complete distribution of the energy frequency time. But, the authors in [25] point out that EMD has some shortcomings. For instance, the process to obtain the IMF in EMD is ineffective and there are serious boundary effects in it. The more importantly though is that the process of EMD produces new modules which do not present in the source signal. In the year 2012, Klein [26] utilized a dual-tree complex wavelet transform (DT-CWT) feature extracted from a non-transient preamble response of an OFDM established on 802.11a signal for identifying four devices with same model by Cisco with dissimilar serial numbers. The accuracy of classification be able to reach 80% when the SNR is lower than 20 dB.

Like many technologies, RF fingerprinting technology is derived from military technology, which can be traced back to the identification of enemies and radar in World War II. The main enemy judgment is made by directly comparing the waveform of the received signal with the waveform map that has been registered by our radar. However, with the increase of equipment and the improvement of production processes, it is impractical to directly compare the signal waveforms. As early as 1995, Choe and Toonstra began to study the characteristics of the extracted device communication signals to detect illegally operating VHF FM transmitters [27,28]. Subsequently, a large number of

RF fingerprinting technologies began to be researched, and various RF fingerprint extraction and authentication methods emerged.

In 2018, Peng et al. designed a hybrid and adaptive classification scheme adjusting to the environment conditions, and carried out extensive experiments to evaluate the performance [29]. They constructed a testbed using a universal software radio peripheral platform as the receiver and 54 ZigBee nodes as the candidate devices to be classified, which is the most ZigBee devices ever tested. The classification error rate is as low as 0.048 in the LOS scenario, and 0.1105, even when a different receiver is used for classification, 18 months after the training. In 2019, Wang et al. selected different characteristics of RF fingerprints and compared the identification accuracy of Zigbee devices with five classification algorithms [30]. The experimental research shows that the highest identification accuracy reached approximately 100% by using multi-features of frequency offset, IQ offset, and circle offset based on the neural network algorithm under a high SNR.

2.2. RF Fingerprint Authentication

In the research of RF fingerprint authentication, device ID verification is an important research content [31–33]. As a device identity detection for a declared ID number, it can be applied to RF fingerprint authentication. This one-to-one verification is computationally intensive and is ideal for lightweight authentication devices. Device ID verification can correctly provide network access to authorized users, as well as network access requests from malicious devices. Therefore, this research has received extensive attention from more and more scholars.

Authorization of network devices has been a serious problem when it comes to access to infrastructure. Dubendorfer, based on the research of Zigbee “hacker” tools, found that the existing anti-attack methods have considerable security risks [31]. First, he collected the transient signals of Zigbee devices and used the RF-DNA fingerprints of seven authorized devices to perform multiple discriminant analysis (MDA) training and identifies these devices. The device’s network access rights are then determined by verifying the device’s claimed identity to filter the rogue device. For authorized devices, a test statistic based on the hypothetical multivariate Gaussian (MVG) likelihood values is used. At SNR = 5 dB, the detection rate of the rogue device is 85%; at SNR = 10 dB, the detection rate exceeds 90%.

In 2012, Cobb proposed that device identification and verification was done by passively monitoring and utilizing the inherent characteristics of IC unintentional RF transmissions, without any modifications to the device being analyzed [32]. He used multiple discriminant analysis to train the recognition system and reduce the data dimension, and used a linear Bayesian classifier for device ID verification. Then, by comparing the Bayesian posterior probability with a specific threshold, it is verified whether the identity of the device is consistent with the classification result. The identification and verification simulation of this study consisted of 40 devices of the same model. At 10 dB, the average verification rate reached 99% and the test error rate was less than 0.05%.

With the popularity of Zigbee equipment in home automation, transportation and industrial control systems, its safety has also received more and more attention. In 2015, Patel analyzed the past Fisher-based multi-discriminant analysis and maximum likelihood (MDA-ML) classification and verification process in detail, and pointed out its problem: When the distribution of RFF does not meet the Gaussian normal condition, the performance of MDA-ML will be reduced [33].

This paper proposes introducing nonparametric random forest and multi-class AdaBoost integrated classifiers into classification and authentication process of devices. In performance testing, this paper used four authorized ZigBee devices for classifier training. Nine rogue devices that have not been seen before have been introduced to measure the performance of the classifier. At 10 dB, the error classification probability is less than 10%.

Unauthorized network access and fraud attacks have been the main research task of information technology security in wireless network communications. In 2015, Reising attempted to solve this problem using RF fingerprint technology to enhance WAP security [34]. This paper proposes the

detection of malicious devices posing as authorized devices by means of dimensionality reduction analysis (DRA) and device ID authentication. Moreover, in recent years, research on the application of RF fingerprint technology in other fields has begun to emerge [35–37], which has enriched the application range of RF fingerprint technology and removed some obstacles for subsequent engineering applications.

In reference [38], a lightweight terminal identity authentication scheme is proposed. The authors use a method based on radio frequency fingerprint identification (RFFID) to authenticate terminal devices with limited computing capabilities in the Internet of Things. In real scenarios, the wavelet features used by this method have a high recognition rate, and prove the effectiveness of this authentication scheme.

ZigBee devices are widely used in the field of IoT. However, more and more scholars pay attention to the authentication security of ZigBee AD hoc network's decentralized architecture. Reference [39] applied the RF-Distinct Native Attribute (RF-DNA) method to device authentication at low SNR. Multipath channel and interference from other devices are considered in the research. The authors used nonparametric random forest and multi-class AdaBoost to authenticate ZigBee devices.

Previous work has focused on using RFF technology to authenticate mobile devices [29–35]. This authentication method requires the device to be authenticated to provide a claimed identity in advance. The authentication model will compare the RFF of the device with the RFF of the device that claims the identity. This comparison results in a judgment as to whether the device to be authenticated is legal. However, the method in this paper does not require the device to be authenticated to provide a claimed identity. It only needs to obtain the RFF of the device to be authenticated, and the authentication model in this paper can handle the legality of the identity of the authentication device to make a decision. This enables the method in this paper to make a decision without other prior information, which further enhances the security of the RFF authentication algorithm.

3. Method

3.1. NCA Feature Selection

Neighborhood component analysis (NCA) is a non-parametric method for selecting features with the goal of maximizing prediction accuracy of regression and classification algorithms. The important of this method is to catch positive definite matrix H connected to the spatial transformation matrix, which can be gotten by characterizing differentiable objective function of H also, by methods for iterative strategies. One of the advantages of this algorithm is that the categories number K can be well-defined by a function f (scalar constants determination). Thus, the problem of model selection can be solved by this algorithm. To describe the transformation matrix H , initially we explain the objective function f that illustrates the accuracy of classification in the target transformation matrix, and attempt to determine that H^* make best use of this objective function.

$$H^* = \arg \max_H f(H) \quad (1)$$

Once classifying a single data point, we want to consider the k -nearest neighbors (k -NN) controlled by a given separation metric, and acquire the sample class according to the labeled category of the k neighbors. In the target transformation space, we don't utilize the method of left-sort to find k nearest neighbors for each sample point, however the entire data set is considered as a random nearest neighbor in the target space. Moreover, squared Euclidean distance function has been utilized to describe the distance among a data point and additional data in the target transformation space. The function can be defined as:

$$p_{ij} = \begin{cases} \frac{e^{-\|Hx_i - Hx_j\|^2}}{\sum_k e^{-\|Hx_i - Hx_j\|^2}}, & \text{if } j \neq i \\ 0, & \text{if } j = i \end{cases} \quad (2)$$

The classification accuracy of the input point i is the classification accuracy of the nearest neighbor set C_i adjacent to it: $p_i = \sum_j^n p_{ij}$, where p_{ij} denotes the probability that j is the nearest neighbor of i . The objective function, defined by the global data set as the nearest neighbor classification method of random nearest neighbors, is defined as follows:

$$f(H) = \sum_i \sum_{j \in C_i} p_{ij} = \sum_i p_i \quad (3)$$

The objective function can be better chosen as:

$$\frac{\partial f}{\partial H} = -2H \sum_i \sum_{C_i} p_{ij} \left(x_{ij} x_{ij}^T - \sum_k p_{ik} x_{ik} x_{ik}^T \right) \quad (4)$$

The continuous gradient descent algorithm is used here.

3.2. Support Vector Data Description

SVDD (support vector data description) is established based on statistical learning theory, inherits its advantages and develops continuously, and has a very complete theoretical foundation and basis. In 2004, Tax and Duin conducted further expansion and more complete research on SVDD, and obtained SVDD without negative samples and with negative samples [40]. Wei et al. introduced the algorithm of SVDD with Markov distance as a measure to replace the traditional Euclidean distance and establish the hyperellipsoid to solve the problem that the hypersphere could not cover some training samples well in the process of fault diagnosis [41]. Zhang et al. proposed a fault diagnosis framework of analog circuits based on a single classifier, and introduced "or" combined results into the test samples to solve the fault in the overlapping region of test samples [42]. In recent years, some research on SVDD algorithm has been published [43,44], and further reasonable improvements have been made to the SVDD algorithm.

The basic idea of SVDD is: Given a data set $X = \{x_1, x_2, \dots, x_n\}$ containing n sample points, the goal of SVDD algorithm is to find a minimum circle with a as the center and R as the radius; the circle can contain all or as many sample points in X as possible. The point where the distance from the training data to the center of the circle is equal to the radius is called the support vector, so the optimization problem can be described as:

$$\text{s.t. } (x_i - a)(x_i - a)^T \leq R^2 + \xi_i \quad (5)$$

where a is the center of the circle; R is the radius; $\xi_i \geq 0$ is the slack variable; and $C > 0$ is the penalty factor, which is used to achieve a comprehensive adjustment between the size of the circle and the number of samples included. The geometric model of SVDD is shown in Figure 1. The black dots in the graph are the data samples in the set X .

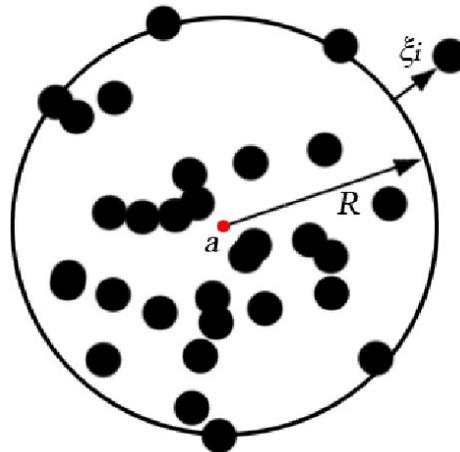


Figure 1. A support vector data description geometric model.

The above optimization problem can be solved by the Lagrange multiplier method to construct the Lagrange equation:

$$L(R, a, \alpha_i, \zeta_i) = R^2 + C \sum_{i=1}^n \zeta_i - \sum_{i=1}^n \alpha_i (R^2 + \zeta_i - (x_i^2 - 2ax_i + a^2)) - \sum_{i=1}^n \gamma_i \zeta_i \quad (6)$$

From the above equation:

$$W = \min_{\alpha} \sum_{i=1}^n \alpha_i (x_i \cdot x_i) - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j (x_i \cdot x_j) \quad (7)$$

s.t. $\sum_{i=1}^n \alpha_i = 1$

$$0 \leq \alpha_i \leq C (\forall i = 1, 2, \dots, n)$$

When the input space is non-circular, the kernel function is introduced to improve the applicability of the algorithm. We find a suitable mapping φ to map the input sample x_i to a high-dimensional feature space $\varphi(x_i)$, and find a hypersphere in the high-dimensional space to surround as many points in the input space as possible. Therefore, the inner product $(x_i \cdot x_j)$ in the above equation can be replaced by the kernel function $k(x_i \cdot x_j)$, and the Gaussian kernel function is selected in this paper. At this point, Equation (8) can be converted into a Lagrange dual problem:

$$W = \max_{\alpha} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j k(x_i \cdot x_j) - 1 \quad (8)$$

s.t. $\sum_{i=1}^n \alpha_i = 1$

$$0 \leq \alpha_i \leq C (\forall i = 1, 2, \dots, n)$$

Equation (9) is a typical quadratic optimization problem, and its decision function is:

$$f(x_i) = (\|\varphi(x_i) - a\|^2 - R^2) \quad (9)$$

According to the above equation, when $f(x_i) = -1$, x_i is classified as normal data point. When $f(x_i) = 1$, x_i are classified as outlier data points.

3.3. Whale Swarm Optimization Algorithm

As the parameters of penalty parameter C and kernel parameter g play a very critical role in the performance of the SVDD model, we need to find suitable parameters to make the performance of the SVDD model better. This paper uses the whale optimization algorithm to optimize the parameters C and g in the SVDD model. Mirjalili proposed the whale swarm optimization algorithm (WOA) in 2016 [45]. This algorithm was inspired by humpback whales using a “spiral bubble net” strategy for hunting. The position of each humpback whale represents a feasible solution. The algorithm has the advantages of less adjustment parameters, simple operation, and strong local optimal ability.

WOA is a mathematical model of humpback whales based on three behaviors: surrounding prey, hunting behavior, and random hunting behavior.

(1) Surround the prey. According to the optimized model established by Mirjalili et al., after finding prey, humpback whales can quickly surround the prey and constantly update the position. The mathematical expression of the position update is

$$\vec{D} = \left| \vec{C}\vec{X}^*(t) - \vec{X}(t) \right| \quad (10)$$

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A}\vec{D} \quad (11)$$

where t is the current iteration number; \vec{X}^* is the best position space for the current whale population; \vec{X} is the location space of individuals; \vec{A} , \vec{C} are coefficients, and \vec{D} is the distance between the current individual and the target.

The calculation method of \vec{A} and \vec{C} is

$$\vec{A} = 2\vec{a}\vec{r} - \vec{a} \quad (12)$$

$$\vec{C} = 2\vec{r} \quad (13)$$

$$\vec{a} = 2 - 2j/M \quad (14)$$

where a is the vector that decreases linearly from 2 to 0 during iteration; R is a random number between 0 and 1; and M is the maximum number of iterations.

(2) Hunting behavior. Humpback whales hunt in a spiral motion.

$$\vec{D}' = \left| \vec{X}^*(t) - \vec{X}(t) \right| \quad (15)$$

$$\vec{X}(t+1) = \vec{D}'e^{bl}(\cos 2\pi l) + \vec{X}^*(t) \quad (16)$$

where b is a constant used to define the helical shape; l is a random number between -1 and 1 , and \vec{D}' is the distance between the best individual and the target.

(3) Search for prey. The model of the whale group is

$$\vec{D} = \left| \vec{C}\vec{X}_{\text{rand}} - \vec{X}(t) \right| \quad (17)$$

$$\vec{X}(t+1) = \vec{X}_{\text{rand}} - \vec{A}\vec{D} \quad (18)$$

where \vec{X}_{rand} is the position vector of the randomly selected whale population.

Based on the SVDD model and WOA parameter optimization algorithm, this paper proposes an integrated SVDD model. The specific operation steps of this model are as follows: During the training phase of integrated SVDD model, users need to store the samples of the training set separately according to categories, and convert the original signals into RFF features through RFF feature extraction and a selection module. The next step is to input the extracted RFF into the SVDD single-category certification model, which also requires separate training according to the

category, that is, the RFF characteristics of each type of sample are trained into a SVDD single-category certification model. The training portion of the integrated SVDD model is completed by training multiple SVDD single-category certification models and combining them.

During the integrated SVDD model test phase, we used the same RFF feature extraction and selection method as the training set to convert the communication signals of the test equipment into RFF features. The second step requires the help of the nearest neighbor finder. By inputting the RFF of the test device into the nearest neighbor finder, we obtain the RFF cluster of A certain type of training sample closest to the RFF of the device in the training set (assuming such a training sample is category A). Finally, we input the RFF into the SVDD single-category authentication model corresponding to category A. As the output of the SVDD single-category authentication model is binary, we can determine whether the test device belongs to category A based on the output. As category A is the training sample, the devices in the training set in the RFF authentication model are all registered legal devices. Therefore, if the test device has passed the identity authentication of the SVDD single-category authentication model corresponding to category A, it is legitimate device. Otherwise, it is a rogue device.

4. Experiment

4.1. Experimental Environment and Experimental Devices

The content of the previous chapter introduced some basic theories of RFF authentication systems. Our actual MIoT operation usually involves the following scenario: When a mobile device or user wants to access the MIoT to complete practical tasks, it needs to complete the authentication operation first. The authentication system needs to verify whether this device is a registered device in the background database, complete normal authentication operation for the registered device, and terminate its authentication operation for the unregistered rogue device.

As the methods and strategies proposed in this paper are based on the physical layer, it means that we can only obtain the subtle fingerprint characteristics from the received signal of the device to determine its identity. For the registered device in the background database, we will record its signal fingerprint characteristics by receiving the signal of the device several times. For unregistered rogue devices, we do not know their signal characteristics in advance. The purpose of this section of the experiment is to establish a complete set of registration and authentication systems by modeling the signals of real mobile devices. We perform normal authentication operations on registered authentic devices and prevent unregistered rogue devices from logging in through the authentication system.

In the experiment of this section, the authors prepared 10 wireless devices of the same model as experimental devices and numbered them uniformly (#1–#10). The wireless devices' model is the Motorola walkie-talkie A12. The experimental setup is shown in Figure 2. We use cables to connect the Agilent oscilloscope to the wireless device to collect its transient signal. Then add Gaussian white noise manually. In order to efficiently obtain the subtle differences between the signals of different experimental device, we extracted the instantaneous amplitude envelope of the signal using a Hilbert transform and performed the 50:1 sampling process. This method can reduce the computation burden of the authentication system without affecting the experimental results.



Figure 2. The experimental setup.

The authentication system is divided into two parts: the registration part and the authentication part. During the experiment, the authors divide the devices into two groups, one is the authentic devices (number of devices is 8), and the other is the rogue devices (number of devices is 2). Among them, the authentic devices are devices registered with the server, and the rogue devices are devices that are not registered with the server. During the registration phase, the oscilloscope receives the signals of the authentic devices multiple times, performs RF fingerprint feature extraction on the signals, and stores the extracted RF fingerprint features according to the device categories. We then use the RF fingerprint training of the authentic devices to train the authentication system model. In the authentication phase, we not only use authentic devices for device login and authentication, but also use unregistered rogue devices to try to log in to the system.

The above content introduces the experimental equipment and experimental procedures. Next, we need to know the performance indicators of the authentication model. There are two main performance indicators: authentic devices authentication success rate (ASR) and rogue devices detection success rate (RSR). ASR refers to the rate that the authentication system recognizes it as the correct legal identity when the authentic devices log in to the system. For example, when the #1 device authenticates, the system recognizes it as the #1 device. RSR refers to the rate that the system can detect it as an unregistered device when rogue devices log in to the system. The results of the next experiment are also based on the performance of these two indicators.

4.2. Simulation Analysis

In this paper, the signals of 10 wireless devices were collected as research samples. First, we intercepted the original signal to obtain the power-on transient signals of 10 wireless devices. Regarding the interception of transient signals, this paper uses the `findchangepts` function in Matlab 2019a to complete the purpose of change point detection. Figure 3 shows the power-on transient signals of four of these devices. We can see that the difference between the transient signals of the four devices is not obvious. Then, we extracted the amplitude envelope of the transient signal through the Hilbert transform. The principal component analysis (PCA) method was used to reduce the dimension of amplitude envelope to obtain the signal features. Finally, the NCA feature selection method was used to screen the signal features obtained in the previous step to obtain the RFF features. In this paper, the NCA feature selection algorithm was used to reduce the dimension of amplitude envelopes. Features with weight w greater than 1 in the NCA algorithm are selected to form the RFF feature. The RFF feature generation process is shown in Figure 4.

In this section, the dimensionality reduction processing of communication signals is divided into two steps: First, this paper chooses 11-dimensional features that contain 95% of the energy of the amplitude envelope of signal as the first step of dimensionality reduction processing. The energy proportion after dimensionality reduction of the original signal with PCA is shown in Table 1. Subsequently, the NCA method is used to select the features of the signal, and the 7-dimensional features with an NCA score greater than 1 are selected as the final RF fingerprint. This is the second step of the dimensionality reduction process.

Table 1. Energy proportion of the feature dimensions after principal component analysis (PCA) dimensionality reduction.

Energy Proportion	85%	90%	95%
Feature dimension	6	8	12

The RFF training SVDD model of authentic devices was also used, and the training data included the communication signals (transient signals) of eight devices. In this paper, the SVDD model was established for eight devices under six SNRs (0 dB, 5 dB, 10 dB, 15 dB, 20 dB, and 25 dB).

The following is a graphical example of how we can use the SVDD model to determine the legality of the device. Only a visualization of the SVDD model for device #2 is shown here. As shown in Figure 5, the x-axis represents the serial number of the sample points. Since this example is the signal of device #2, its serial number range is 31–60 (a total of 10 devices in the test set, and each device has 30 sample points). The y-axis represents the distance between the sample point and the center of the SVDD hypersphere, and the horizontal line in the figure is the decision threshold (i.e., the radius of the hypersphere). The radius of the hypersphere is determined by the WOA algorithm during model training. The discriminant threshold is determined by the supersphere radius.

It can be seen that under the condition of 5 dB, the distance between most positive samples and the center of the sphere is still below the decision threshold, but some of them fall above the decision threshold. Moreover, since the sample of device #4 is relatively close to the sample of device #2, the sample of device #4 is also relatively close to the decision threshold. It can be seen that at SNR = 10 dB, although a small number of positive samples are above the judgment threshold, the center of SVDD is far away from the sample points of other devices, which reduces the possibility of misjudgment in the model. When the SNR increases gradually, the performance of the SVDD model becomes better and better because the sample points are more concentrated in the feature space.

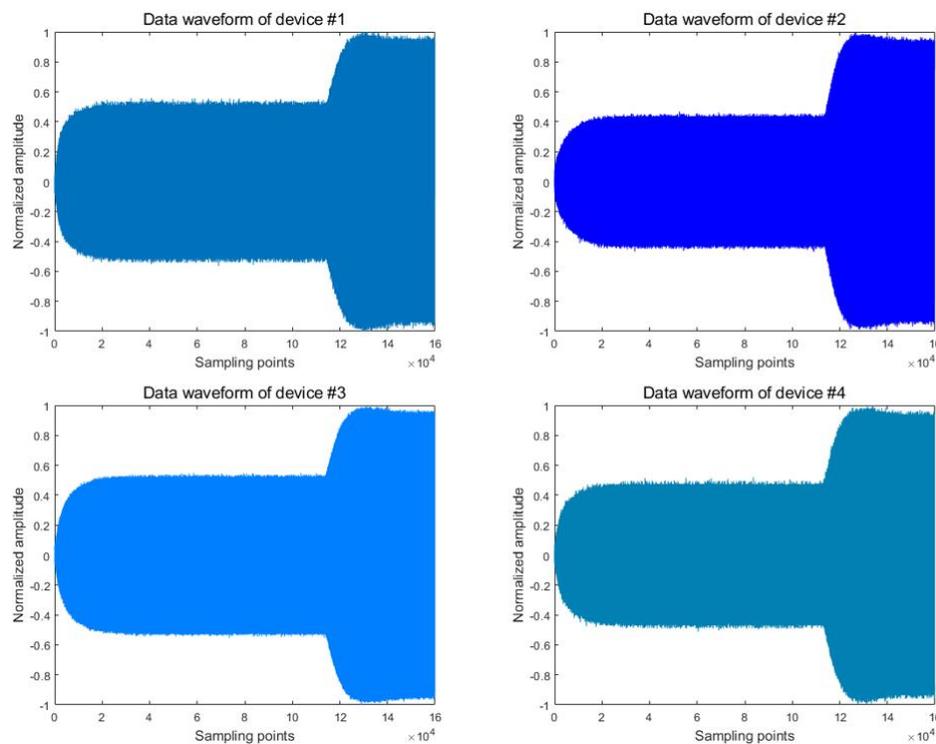


Figure 3. Transient signal waveforms of four wireless devices.

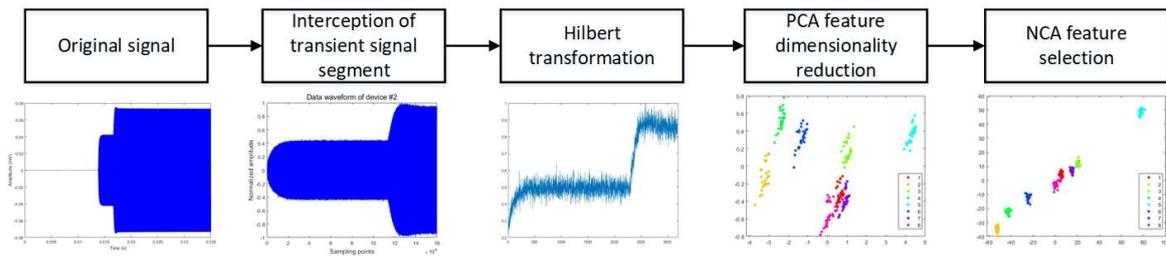


Figure 4. The radio frequency fingerprint (RFF) feature generation process. Neighborhood component analysis (NCA).

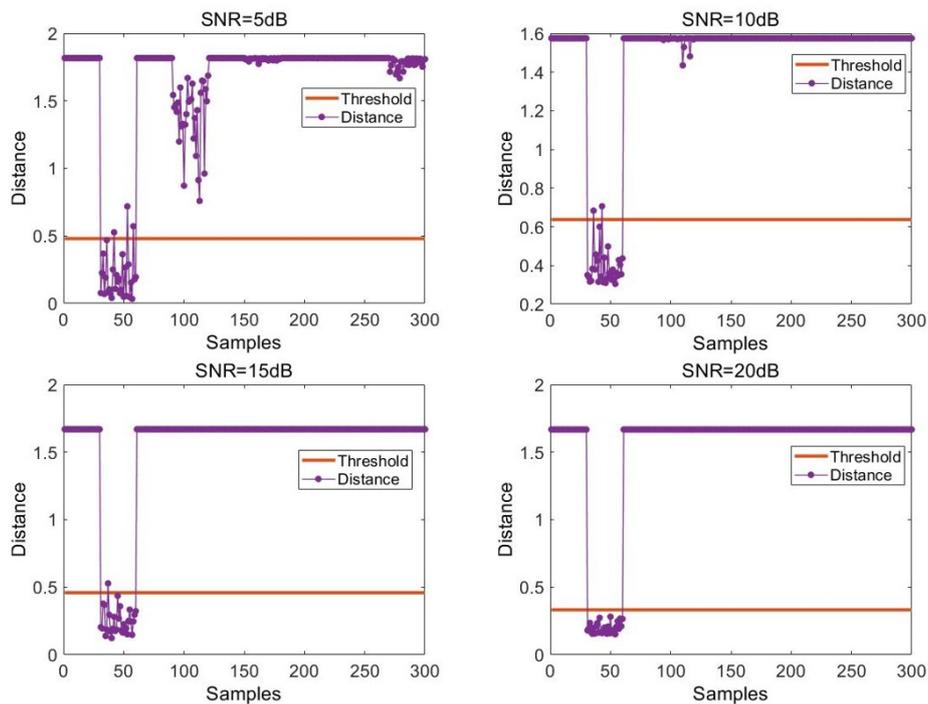


Figure 5. Schematic diagram of the distance between the sample point and the hypersphere. Signal-to-noise ratio (SNR).

With these examples, we introduced how a single category of SVDD model works. However, as authentic devices often contain many categories, we need to build a multi-category device authentication model. The authentication process is shown in Figure 6. And the specific form of the RFF authentication model is shown in Figure 7. The authentication model is divided into three parts: RFF generator, K nearest neighbor finder, and SVDD discriminator. The RFF generator is used to convert transient signals into RFF features. The K-nearest neighbor finder is responsible for assigning the sample to be authenticated to the SVDD model of a specific device. It finds the category of the K samples closest to the sample to be certified in the RFF feature space and determines which SVDD model the sample to be certified is assigned to. The SVDD discriminator is composed of N SVDD models (N is the number of authentic devices), and each authentic device establishes an SVDD model. It provides a deny option to the system to deny illegal access to those rogue devices. The above three parts together constitute the authentication model. Its tasks include two aspects: 1. Denying illegal access by rogue devices; 2. Accurately identify authentic devices.

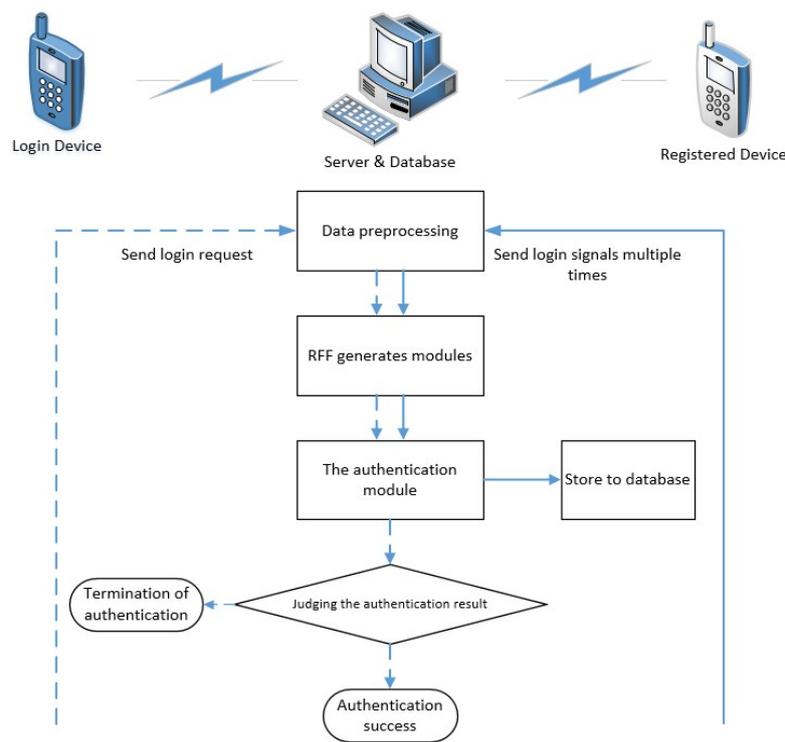


Figure 6. The process of the RFF authentication.

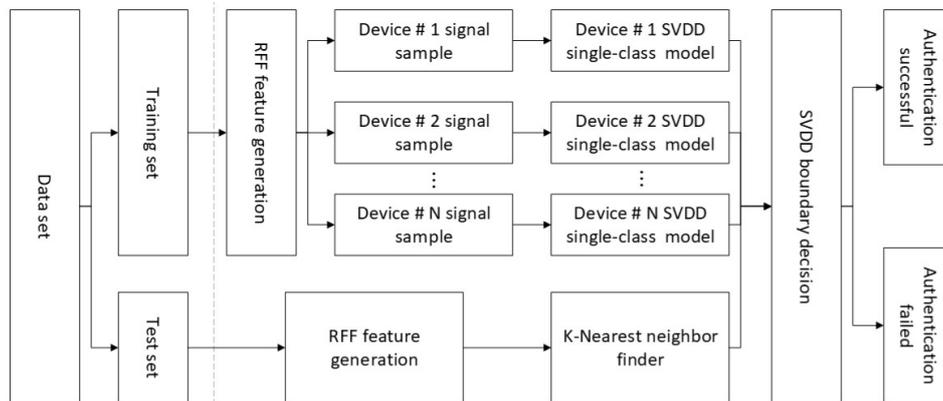


Figure 7. The specific form of the RFF authentication model. Support vector data description (SVDD).

In order to ensure the reliability of the experimental results, and avoid experimental results being affected by a specific grouping of equipment, a total of five independent experiments were performed. Each group of experiments randomly selected two devices as rogue devices and the remaining eight devices as authentic devices. Finally, the average result of five independent experiments was taken as the experimental result. Through the experimental results, the authentication ability of this method to each authentic device and the detection ability of this method to each rogue device were analyzed. Table 2 shows the authentication success rate of eight authentic devices when using the SVDD authentication model.

Considering the completeness and conciseness of the experimental results, we only listed the maximum and minimum ASR in each group of experiments, as this can show both the best case of SVDD model for each group of data authentication and the worst case of each group of data authentication. This does not take up a great deal of space. When the SNR is lower than 5 dB, the difference between the best and worst results of each group of experiments is about 8%. With the increase of SNR, the distribution of features becomes more concentrated, and the gap between the best

and worst results of each group of experiments becomes smaller and smaller. At a signal-to-noise ratio of 20 dB, the worst-case ASR was over 95%.

The comparison method selected in this paper is based on the sample average distance authentication model and the posterior probability SVM. The ASR curve of authentic devices is shown in Figure 8. We can see that the ASR of the three methods exceeds 40% when the SNR is greater than 1 dB. As the SNR increases, the performance of the three methods begins to improve significantly. During the 5–10 dB period, the ASR of the method of this paper achieved a rapid growth. This is because, in this interval, the feature distribution changes rapidly, and the characteristics of the similar sample points are accelerated. The SVDD model also performs better for training samples in the sample point set. It can be seen that when the SNR exceeds 15 dB, the ASR of this method exceeds 90%, which further proves the effectiveness of the proposed method.

Table 2. Authentic devices authentication success rate (ASR) for the SVDD model.

Experiment	ASR_{\min} and ASR_{\max}	SNR					
		0 dB	5 dB	10 dB	15 dB	20 dB	25 dB
1	ASR_{\min}	0.48	0.63	0.77	0.88	0.95	0.99
	ASR_{\max}	0.55	0.69	0.85	0.95	1.00	1.00
2	ASR_{\min}	0.41	0.61	0.78	0.86	0.94	0.98
	ASR_{\max}	0.49	0.7	0.84	0.92	0.99	1.00
3	ASR_{\min}	0.5	0.66	0.78	0.9	0.95	0.99
	ASR_{\max}	0.6	0.71	0.82	0.94	0.98	1.00
4	ASR_{\min}	0.48	0.67	0.79	0.92	0.97	1
	ASR_{\max}	0.51	0.71	0.83	0.96	1.00	1.00
5	ASR_{\min}	0.48	0.65	0.74	0.88	0.96	0.99
	ASR_{\max}	0.56	0.72	0.82	0.94	1.00	1.00

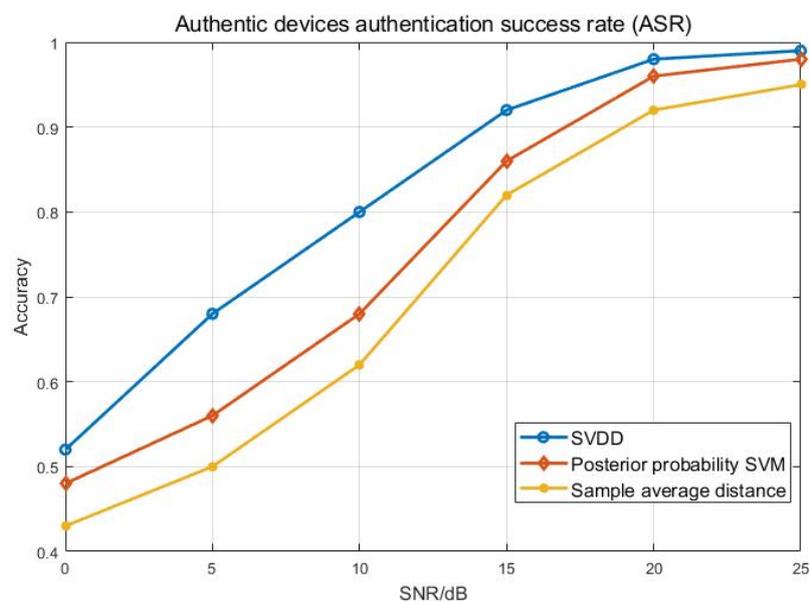


Figure 8. Authentic devices authentication success rate.

In the authentication phase, we not only completed the correct authentication of the authentic devices, but also the detection and judgment of the rogue devices. We use rogue devices to attempt to log in to the authentication system, and then the authentication system gives the authentication results. In the specific operation, we will judge according to the distance of the SVDD model output in the authentication system. We use the hypersphere radius as a threshold, and for devices below the

threshold, the system terminates its authentication operation. Table 3 shows the detection success rate of two rogue devices when the SVDD authentication model is used. It can be seen that the detection capability of the SVDD model for rogue devices is basically the same, with an average difference of 1–2% for five SNR.

The effect of the authentication system successfully detecting rogue devices is shown in Figure 9. We can see that our method is usually ten percent ahead of the comparison method. The performance difference is especially obvious in the case of low SNR. With the improvement of SNR, the gap starts to narrow. This is because the distribution of RFF characteristics of different categories becomes sparse in the case of high SNR. When the SNR is 15 dB, the RSR of all three methods reaches 90%. It is not difficult to find that the method in this paper has better authentication performance and detection performance than the comparison method, which is largely due to the advantage of the SVDD algorithm in data description—making full use of the sample tag information.

Table 3. Rogue devices detection success rate (RSR) for the SVDD model.

Experiment	Rogue Devices	SNR					
		0 dB	5 dB	10 dB	15 dB	20 dB	25 dB
1	#9	0.58	0.72	0.89	0.93	0.97	0.99
	#10	0.59	0.75	0.90	0.96	0.99	1.00
2	#2	0.64	0.76	0.92	0.95	0.98	0.99
	#6	0.61	0.70	0.86	0.91	0.96	0.99
3	#7	0.58	0.70	0.87	0.92	0.97	0.99
	#8	0.56	0.67	0.89	0.94	0.96	1.00
4	#1	0.65	0.76	0.92	0.96	0.98	0.99
	#4	0.58	0.67	0.87	0.91	0.97	1.00
5	#3	0.61	0.73	0.90	0.95	0.98	1.00
	#5	0.60	0.69	0.86	0.92	0.97	0.99

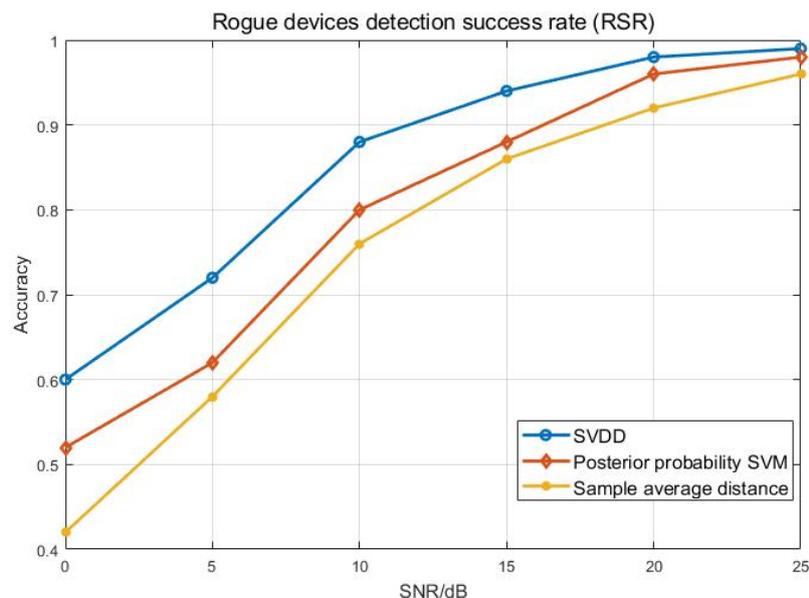


Figure 9. Rogue device detection success rate.

5. Conclusions

This paper introduces the theoretical basis and experimental demonstration of an RFF authentication model based on the SVDD algorithm. First, the NCA feature selection algorithm

and SVDD algorithm are introduced. Their theory and application method are introduced in detail. Then we apply these two algorithms to the RFF authentication model. We have outlined the operation of the authentication model. The model is evaluated with real devices. Experimental results show that the RFF authentication model is reliable. From the above simulation experiments, it can be seen that the application of the RFF authentication model to the identity authentication of the MIIoT device can achieve better performance, and due to its work in the physical layer, the security is also greatly guaranteed. In other words, the RFF authentication model can well meet the security performance requirements of a device in the MIIoT.

Due to the time and space limitations, this paper only discusses the recognition method based on specific channel conditions. In the future, we are trying to use a functional model to describe the unique physical layer differences of the devices. Furthermore, the channel influence is separated from the function model, so that this technology can be used in a scenario where the channels are dynamic.

Author Contributions: Conceptualization, Y.L.; Data curation, Q.T., X.G., O.A. and A.T.; Formal analysis, Q.T., Y.L., X.G., O.A. and A.T.; Funding acquisition, Y.L. and O.A.; Methodology, Q.T. and X.G.; Project administration, Q.T., Y.L., X.G. and J.W.; Writing—original draft, Q.T. and X.G.; Writing—review and editing, Q.T., Y.L., X.G., J.W., O.A. and A.T. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: This work is supported by the National Natural Science Foundation of China (61771154) and the Fundamental Research Funds for the Central Universities (HEUCFG201830). This work is funded by the Researchers Supporting Project No. (RSP-2019/102) King Saud University, Riyadh, Saudi Arabia.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Atzori, L.; Iera, A.; Morabito, G. The Internet of Things: A survey. *Comput. Netw.* **2010**, *54*, 2787–2805. [[CrossRef](#)]
- Welbourne, E.; Battle, L.; Cole, G.; Gould, K.; Rector, K.; Raymer, S. Building the Internet of Things using RFID: The RFID ecosystem experience. *IEEE Internet Comput.* **2009**, *13*, 48–55. [[CrossRef](#)]
- Wang, J.; Gu, X.; Liu, W.; Sangaiah, A.K.; Kim, H.J. An Empower Hamilton Loop based Data Collection Algorithm with Mobile Agent for WSNs. *Hum. Centric Comput. Inf. Sci.* **2019**, *9*, 1–14. [[CrossRef](#)]
- Li, W.; Chen, Z.; Gao, X.; Liu, W.; Wang, J. Multimodel Framework for Indoor Localization Under Mobile Edge Computing Environment. *IEEE Internet Things J.* **2019**, *6*, 4844–4853. [[CrossRef](#)]
- Wang, J.; Gao, Y.; Yin, X.; Li, F.; Kim, H.J. An Enhanced PEGASIS Algorithm with Mobile Sink Support for Wireless Sensor Networks. *Wirel. Commun. Mob. Comput.* **2018**, *2018*, 9472075. [[CrossRef](#)]
- Adam, C.P.; Dennis, L.G. Wireless Device Identification Based on RF Oscillator Imperfections. *IEEE Trans. Inf. Forensics Secur.* **2015**, *10*, 2492–2501.
- Gungor, O.; Koksal, C.E. On the basic limits of RF-fingerprint-based authentication. *IEEE Trans. Inf. Theory* **2016**, *62*, 4523–4543. [[CrossRef](#)]
- Wang, J.; Gao, Y.; Liu, W.; Sangaiah, A.K.; Kim, H.J. An Intelligent Data Gathering Schema with Data Fusion Supported for Mobile Sink in WSNs. *Int. J. Distrib. Sens. Netw.* **2019**, *15*, 23–30. [[CrossRef](#)]
- Wang, J.; Gao, Y.; Wang, K.; Sangaiah, A.K.; Lim, S.J. An Affinity Propagation-Based Self-Adaptive Clustering Method for Wireless Sensor Networks. *Sensors* **2019**, *19*, 2579. [[CrossRef](#)]
- Lin, Y.; Tao, H.; Tu, Y.; Liu, T. A Node Self-Localization Algorithm with A Mobile Anchor Node in Underwater Acoustic Sensor Networks. *IEEE Access* **2019**, *7*, 43773–43780. [[CrossRef](#)]
- Wang, J.; Gao, Y.; Liu, W.; Sangaiah, A.K.; Kim, H.J. Energy Efficient Routing Algorithm with Mobile Sink Support for Wireless Sensor Networks. *Sensors* **2019**, *19*, 1494. [[CrossRef](#)] [[PubMed](#)]
- Dou, Z.; Si, G.; Lin, Y.; Wang, M. An Adaptive Resource Allocation Model with Anti-jamming in IoT Network. *IEEE Access* **2019**, *4*, 93250–93258. [[CrossRef](#)]
- Tu, Y.; Lin, Y.; Wang, J.; Kim, J.U. Semi-Supervised Learning with Generative Adversarial Networks on Digital Signal Modulation Classification. *CMC Comput. Mater. Continua* **2018**, *55*, 243–254.
- Wang, H.; Guo, L.; Dou, Z.; Lin, Y. A New Method of Cognitive Signal Recognition Based on Hybrid Information Entropy and D-S Evidence Theory. *Mob. Netw. Appl.* **2018**, *4*, 677–685. [[CrossRef](#)]

15. Wang, H.; Li, J.; Guo, L.; Dou, Z.; Lin, Y.; Zhou, R. Fractal Complexity-Based Feature Extraction Algorithm of Communication Signals. *Fractals Complex Geom. Patterns Scaling Nat. Soc.* **2017**, *25*, 10–18. [[CrossRef](#)]
16. Lin, Y.; Zhu, X.; Zheng, Z.; Dou, Z.; Zhou, R. The individual identification method of wireless device based on dimensionality reduction and machine learning. *J. Supercomput.* **2019**, *75*, 3010–3027. [[CrossRef](#)]
17. Shi, Q.; Kang, J.; Wang, R.; Yi, H.; Lin, Y.; Wang, J. A Framework of Intrusion Detection System based on Bayesian Network in IoT. *Int. J. Perform. Eng.* **2018**, *14*, 2280–2288. [[CrossRef](#)]
18. Wu, R.; Chen, X.; Han, H.; Zhao, H.; Lin, Y. Abnormal Information Identification and Elimination in Cognitive Networks. *Int. J. Perform. Eng.* **2018**, *14*, 2271–2279. [[CrossRef](#)]
19. Hall, J.; Barbeau, M.; Kranakis, E. Enhancing intrusion detection in wireless networks using radio frequency fingerprinting. *Commun. Internet Inf. Technol.* **2004**, *2004*, 201–206.
20. Hall, J.; Barbeau, M.; Kranakis, E. Radio frequency fingerprinting for intrusion detection in wireless networks. *IEEE Trans. Def. Secur. Comput.* **2005**, *12*, 1–35.
21. Hao, P.; Wang, X.; Behnad, A. Relay authentication by exploiting I/Q imbalance in amplify-and-forward system. *IEEE GLOBECOM* **2014**, *12*, 613–618.
22. Knox, D.; Kunz, T. Wireless fingerprints inside a wireless sensor network. *ACM Trans. Sens. Netw. (TOSN)* **2015**, *11*, 37–52. [[CrossRef](#)]
23. Deng, S.; Huang, Z.; Wang, X. Radio frequency fingerprint extraction based on multidimension permutation entropy. *Int. J. Antennas Propag.* **2017**, *2017*, 1538728. [[CrossRef](#)]
24. Huang, N.E.; Shen, Z.; Long, S.R. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Math. Phys. Eng. Sci.* **1998**, *454*, 90–95. [[CrossRef](#)]
25. Frei, M.; Osorio, I. Intrinsic time-scale decomposition: Time-frequency-energy analysis and real-time filtering of non-stationary signals. *Proc. Math. Phys. Eng. Sci.* **2007**, *463*, 321–342. [[CrossRef](#)]
26. Klein, R.W.; Temple, M.A.; Mendenhall, M.J. Application of wavelet-based RF fingerprinting to enhance wireless network security. *J. Commun. Netw.* **2012**, *11*, 544–555. [[CrossRef](#)]
27. Polak, A.C.; Dolatshahi, S.; Goeckel, D.L. Identifying wireless users via transmitter imperfections. *IEEE J. Sel. Areas Commun.* **2011**, *29*, 1469–1479. [[CrossRef](#)]
28. Toonstra, J.; Kinsner, W. Transient analysis and genetic algorithms for classification. *Wescanex 95 Commun. Power Comput. Conf. IEEE* **1995**, *29*, 501–510.
29. Wang, J.; Zhuang, L.; Cheng, W.; Chao, X.; Wu, X.; Zhang, Z. Analysis of Classification Methods Based on Radio Frequency Fingerprint for Zigbee Devices. In *Advances in Computer Communication and Computational Sciences*; Springer: Singapore, 2019; pp. 121–132.
30. Peng, L.; Hu, A.; Zhang, J.; Jiang, Y. Design of a Hybrid RF Fingerprint Extraction and Device Classification Scheme. *IEEE Internet Things J.* **2019**, *6*, 349–360. [[CrossRef](#)]
31. Dubendorfer, C.K.; Ramsey, B.W.; Temple, M.A. An RF-DNA verification process for ZigBee networks. *Mil. Commun. Conf. IEEE* **2013**, *29*, 1–6.
32. Cobb, W.E.; Laspe, E.D.; Baldwin, R.; Temple, M.A. Intrinsic Physical-Layer Authentication of Integrated Circuits. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 14–24. [[CrossRef](#)]
33. Patel, H.J.; Temple, M.A.; Baldwin, R.O. Improving ZigBee Device Network Authentication Using Ensemble Decision Tree Classifiers With Radio Frequency Distinct Native Attribute Fingerprinting. *IEEE Trans. Reliab.* **2015**, *64*, 221–233. [[CrossRef](#)]
34. Reising, D.R.; Temple, M.A.; Jackson, J.A. Authorized and Rogue Device Discrimination Using Dimensionally Reduced RF-DNA Fingerprints. *IEEE Trans. Inf. Forensics Sec.* **2015**, *10*, 1180–1192. [[CrossRef](#)]
35. Zhang, Z.; Guo, X.; Lin, Y. Trust Management Method of D2D Communication Based on RF Fingerprint Identification. *IEEE Access* **2015**, *10*, 1180–1192. [[CrossRef](#)]
36. Mishra, D.; De, S.; Alexandropoulos, G.C.; Krishnaswamy, D. Energy-aware Mode Selection for Throughput Maximization in RF-Powered D2D Communications. In *Proceedings of the GLOBECOM 2017—2017 IEEE Global Communications Conference, Singapore, 4–8 December 2017*.
37. Chen, Y.; Wen, H.; Song, H.; Chen, S.; Xie, F.; Yang, Q.; Hu, L. Lightweight one-time password authentication scheme based on radio-frequency fingerprinting. *IET Commun.* **2018**, *12*, 1477–1484. [[CrossRef](#)]
38. Chen, S.L.; Wen, H.; Wu, J.S.; Xu, A.D.; Jiang, Y.X.; Song, H.H.; Chen, Y. Radio Frequency Fingerprint-Based Intelligent Mobile Edge Computing for Internet of Things Authentication. *Sensors* **2019**, *19*, 3610. [[CrossRef](#)]

39. Patel, H.; Temple, M.; Baldwin, R.; Ramsey, B. Application of ensemble decision tree classifiers to zig bee device network authentication using RF-DNA fingerprinting. In Proceedings of the 9th International Conference on Cyber Warfare and Security, West Lafayette, IN, USA, 24–25 March 2014; pp. 176–184.
40. Tax, D.; Duin, R. Support Vector Data Description. *Mach. Learn.* **2004**, *54*, 45–66. [[CrossRef](#)]
41. Wei, X.K.; Huang, G.B.; Li, Y.H. Mahalanobis Ellipsoidal Learning Machine for One Class Classification. In Proceedings of the IEEE 2007 International Conference on Machine Learning and Cybernetics, Hong Kong, China, 19–22 August 2007; pp. 3528–3533.
42. Zhang, Y.; Wei, X.; Jiang, H. One-class classifier based on SBT for analog circuit fault diagnosis. *Measurement* **2008**, *41*, 371–380.
43. Shi, Z.; Huang, M.; Zhao, C. Detection of LSSUAV using hash fingerprint based SVDD. In Proceedings of the 2017 IEEE International Conference on Communications, Paris, France, 21–25 May 2017; pp. 1–5.
44. Uslu, F.S.; Binol, H.; Ilarslan, M. Improving SVDD Classification Performance on Hyperspectral Images via Correlation Based Ensemble Technique. *Opt. Lasers Eng.* **2017**, *89*, 169–177. [[CrossRef](#)]
45. Mirjalili, S.; Lewis, A. The whale optimization algorithm. *Adv. Eng. Softw.* **2016**, *95*, 51–67. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).