Hindawi
Wireless Communications and Mobile Computing
Volume 2018, Article ID 6039878, 10 pages
https://doi.org/10.1155/2018/6039878

Research Article

Physical-Layer Channel Authentication for 5G via Machine Learning Algorithm

**Songlin Chen,[1] Hong Wen,[1] Jinsong Wu,[2] Jie Chen,[1] Wenjie Liu,[1]
Lin Hu,[3] and Yi Chen[1]**

[1] *The National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China*
[2] *Department of Electrical Engineering, Universidad de Chile, Santiago 833-0072, Chile*
[3] *Chongqing Key Laboratory of Mobile Communication Technology, Chong Qing University of Post & Telecommunication of China, Chongqing, China*

Correspondence should be addressed to Hong Wen; wcdma_2000@hotmail.com

Received 26 January 2018; Accepted 19 September 2018; Published 2 October 2018

Academic Editor: Vicente Casares-Giner

By utilizing the radio channel information to detect spoofing attacks, channel based physical layer (PHY-layer) enhanced authentication can be exploited in light-weight securing 5G wireless communications. One major obstacle in the application of the PHY-layer authentication is its detection rate. In this paper, a novel authentication method is developed to detect spoofing attacks without a special test threshold while a trained model is used to determine whether the user is legal or illegal. Unlike the threshold test PHY-layer authentication method, the proposed AdaBoost based PHY-layer authentication algorithm increases the authentication rate with one-dimensional test statistic feature. In addition, a two-dimensional test statistic features authentication model is presented for further improvement of detection rate. To evaluate the feasibility of our algorithm, we implement the PHY-layer spoofing detectors in multiple-input multiple-output (MIMO) system over universal software radio peripherals (USRP). Extensive experiences show that the proposed methods yield the high performance without compromising the computing complexity.

1. Introduction

5G mobile communication system puts forward the requirements that are high-speed, high efficiency, and high security under three typical application scenarios: enhanced Mobile Broadband (eMBB), Large-Scale Internet of Things (IoT), and ultra Reliable & Low-latency Connections (uRLLC) [1, 2]. The specific application scenarios that enhance the need for mobile broadband including high-traffic and high-density wireless networks are densely used in indoors or urban areas, in which large-area signals of wireless mobile networks are continuously covered in rural areas. Meanwhile, 5G involves the interconnection and communication between a large number of machines and equipment, which is a necessary condition for the operation of IoT [3]. Many mobile devices access the wireless network at the same time, which results in heavy burden of authentication computing in the wireless network. Therefore, lightweight access methods are required for intensive application scenarios of 5G wireless communication networks.

In response to this need, scholars have successively carried out researches on light-weight security measures based on computational cryptography [4, 5]. However, it is still very difficult to use the cipher algorithm that meets the resource-constrained application scenarios such as wireless mobile terminals, IoT, and sensor networks. Therefore, there is a need to find new technologies to construct the lightweight security scheme. In the last decade, the research of PHY-layer security technology has brought new vitality to the wireless mobile communication industry [6–10]. The physical layer of the characteristics is difficult to be counterfeit, which can provide high level security with low cost to overcome the lack of the cipher based security technologies. Consequently, physical layer characteristics which can be used to improve the security of wireless communications have been widely concerned for researchers.

Several PHY-authentication techniques are proposed. In [11–17], the received signal strength (RSS) and channel impulse response (CIR), as well as channel state information (CSI), are utilized to detect identity-based attacks in wireless networks, such as man-in-the middle and denial-of-service (DoS) attacks. The work [18] presents a PHY-authentication framework that can be adapted for multicarrier transmission. In order to detect Sybil attacks, [19, 20] present a PHY-authentication protocol that combines with high-layer authentication based on the channel response decorrelations rapidly in space, and channel-based detection of Sybil attacks in wireless networks is implemented. In [21], Peng Hao et al. developed a practical authentication scheme by monitoring and analyzing the packet error rate (PER) and received signal strength indicator (RSSI) at the same time to enhance the spoofing attack detection capability. In [22–24], the authors analysed the spatial decorrelation property of the channel response and validated the efficacy of the channel-based authentication for spoofing detection in MIMO system by the comparison between channel information "difference" of two or several frames.

However, in above-mentioned works, artificial thresholds are needed to detect spoofing attack. In fact, threshold range cannot be accurately confirmed, resulting in spoofing detection with low precision. In this paper, a machine learning based PHY-layer authentication is developed, which provides an intelligent decision method instead of a one-dimension test threshold. Specifically, Adaboost [25, 26] based algorithm with one-dimensional feature is employed to detect spoofing attacks. To enhance authentication performance, the two-dimensional feature is carried out. The major contributions of this paper are summarized as follows:

(1) An AdaBoost based PHY-layer authentication algorithm is proposed to increase the authentication rate.

(2) The authentication model based on two-dimensional feature is established, which has a stronger performance for cheating detection than the one-dimensional authentication method.

(3) The proposed PHY-layer channel authentication scheme is implemented in a real world environment, based on MIMO-OFDM systems. The simulation results show that the detection rate is greatly increased.

The rest of this paper is organized as follows. Section 2 describes system model and problem formulation. Our proposed algorithm for PHY-layer authentication is presented in Section 3. The system experiment and simulation results are presented in Section 4. In Section 5, we conclude this paper.

2. System Model

In this section, we provide a system model of physical layer authentication and hypothesis testing.

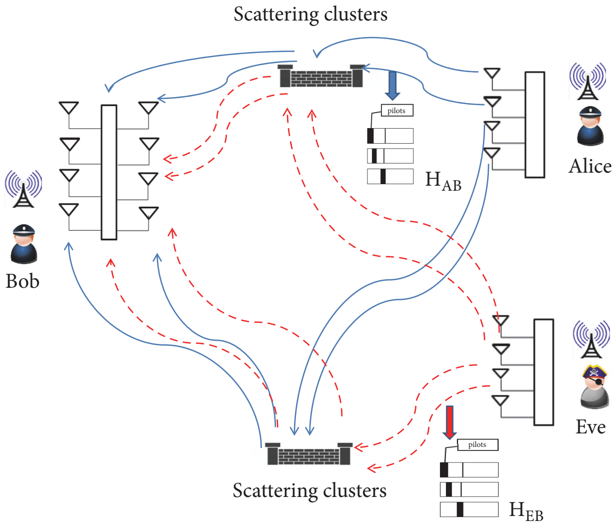


Figure 1: Alice-Bob-Eve model in MIMO system.

2.1. MIMO Three Parts System Model. As shown in Figure 1, our analysis is based on an Alice-Bob-Eve model in MIMO system, where Alice and Bob are legitimate users equipped with N_T and N_R antennas, respectively. Eve with N_T antennas attempts to spoof Alice by using her identity. They are assumed to be located in spatially separated positions. In order to address this spoofing detection, Bob tracks the uniqueness of wireless channel responses to discriminate between legitimate signals from Alice and illegitimate signals from Eve. That is a physical layer authentication. The detailed physical layer authentication process is as follows: Signals with the pilots which can be used to estimate the channel response of the corresponding transmitter are transmitted over the wireless multipath channel to the receiver. The i-th transmission data contains N_f-frames, while each frame consists of N_s OFDM symbols.

Bob is assumed to obtain the Alice-Bob channel information for any frame index $k > 1$, $\widehat{H}_k^{AB}$, and save it which extracted by the channel estimation. After a while, when Bob receives the next data frame, the $k+1$th data frame, $\widehat{H}_{k+1}^{AB}$, which is extracted and estimated by Bob the unknown channel response information. Bob compares $\widehat{H}_{k+1}^{AB}$ with the channel of Alice, $\widehat{H}_k^{AB}$, to determine whether the corresponding signal is actually send by Alice.

If the values of $\widehat{H}_k^{AB}$ and $\widehat{H}_{k+1}^{AB}$ are approaching, Bob considers the sender's identity as valid and stores it. On the contrary, Bob determines that the sender's identity is invalid and directly abandons the data frame.

Channel information is detected by the channel estimation algorithm, denoted by $\widehat{H}_k^{AB}$ and $\widehat{H}_{k+1}^{AB}$. Each data frame contains N_s OFDM symbols. Thus, the channel information is given by

$$\widehat{H}_k^{AB} = \left[\widehat{H}_{k,1}^{AB}, \widehat{H}_{k,2}^{AB}, \ldots, \widehat{H}_{k,N_s}^{AB}\right] \tag{1}$$

where $\widehat{H}_{k,x}^{AB}$ $(x = 1, 2, \ldots, N_s)$ denotes the x-th OFDM symbol of channel information.

2.2. Hypothesis Testing. A binary hypothesis testing is performed to determine the identity authentication in the continuous data frames. Let the receiver Bob verify that the kth data frame originates from the legitimate sender Alice, and the extracted channel information is H_k^{AB}; the sender of the $k+1$ th data frame is still unknown and the channel information is H_{k+1}^{AB}: the null hypothesis H_0 indicates that the packet is indeed sent by the Alice. The alternative hypothesis H1 is that the real client of the packet is not Alice. The spoofing detection builds the hypothesis test given by

$$\begin{aligned} H_0 &: H_{k+1}^{AB} \longrightarrow H_k^{AB} \\ H_1 &: H_{k+1}^{AB} \longrightarrow H_k^{AB} \end{aligned} \tag{2}$$

where all elements of N_k and N_{k+1} are i.i.d. complex Gaussian noise samples $CN(0, \delta^2)$. Therefore, if channel information for hypothesis testing is directly used, the need of considering the impact of noise variables will increase the certification complexity. To this end, since N_k and N_{k+1} are with the same statistical characteristics, the "difference" of channel information can eliminate the influence of noise variables. The physical layer authentication translates into the comparison between the "difference" of the channel information and the set threshold. Equation (2) can be expressed as

$$\begin{aligned} H_0 &: \mathit{diff}\left(H_{k+1}^{AB}, H_k^{AB}\right) < \eta \\ H_1 &: \mathit{diff}\left(H_{k+1}^{AB}, H_k^{AB}\right) > \eta \end{aligned} \tag{3}$$

where $\mathit{diff}(A, B)$ denotes the calculating result of the difference between A and B and η is the test threshold.

The null hypothesis, $\boldsymbol{H}_0$, is that the identity is legitimate and Bob accepts this hypothesis if the test statistic he computes, $\mathit{diff}(A, B)$, is below some threshold η. Otherwise, Bob accepts the alternative hypothesis, $\boldsymbol{H}_1$, that the identity is illegitimate. The channel response "difference" is recorded as T, and (3) can be also written as

$$T = \mathit{diff}\left(H_{k+1}^{AB}, H_k^{AB}\right) \underset{< H_0}{\overset{> H_1}{}} \eta \tag{4}$$

As shown in (4), the physical layer authentication is actually a comparison between channel information "difference" and authentication threshold. Thus, the difference between channel information and authentication threshold is the key of physical layer authentication. The test statistics can measure the similarity of channel information and calculate the channel information difference. In this paper, we use two kinds of test statistic T_A and T_B, respectively. In particular, assuming Bob obtains two consecutive frame channel response of $\widehat{H}_{k-1,x}^{AB}$ and $\widehat{H}_{k,x}^{AB}$, respectively, from Alice. We build test statistics of T_A and T_B based on the two frames for the purpose of discrimination identity of Alice or Eve. Subsequently, Bob acquires the k+1th frame channel information as $\widehat{H}_{k+1,x}^{AB}$.

The test statistics are calculated as

$$T_A(k) = \left| \frac{\mathit{diff}\left(\widehat{H}_{k+1,x}^{AB} - \widehat{H}_{k,x}^{AB}\right)}{\mathit{diff}\left(\widehat{H}_{k,x}^{AB} - \widehat{H}_{k-1,x}^{AB}\right)} \right| = \left| \frac{\sum_{x=1}^{N_s} \sum_{m=1}^{N} \sum_{n=1}^{N} \left| \widehat{H}_{k+1,x}^{AB}(m,n) - \widehat{H}_{k,x}^{AB}(m,n)\, e^{j\widehat{\theta}(m,n)} \right|}{\sum_{x=1}^{N_s} \sum_{m=1}^{N} \sum_{n=1}^{N} \left| \widehat{H}_{k,x}^{AB}(m,n) - \widehat{H}_{k-1,x}^{AB}(m,n)\, e^{j\widehat{\theta}(m,n)} \right|} \right| \underset{< H_0}{\overset{> H_1}{}} \eta_A, \tag{5}$$

where $\widehat{\theta}(m,n)$ is the phase offset and can be denoted by

$$\widehat{\theta}(m,n) = \arg\left(\widehat{H}_{k,x}^{AB}(m,n) \left[H_{k+1,x}^{XB}(m,n)\right]^*\right) \tag{6}$$

From (5), T_A can be taken as the difference of the subcarrier amplitude, which avoids the effect of $\widehat{\theta}(m,n)$.

Two consecutive data frames, $\widehat{H}_{k,x}^{AB}$ and $\widehat{H}_{k+1,x}^{AB}$, represent measurement errors in the phase of the channel response. Each channel response value consists of N_s frequency domain channel matrix, which is OFDM symbol of N dimensional square matrix and n denotes the mth row and n denotes the column element phase offset.

$$T_B(k) = \left| \frac{\mathit{diff}\left(\widehat{H}_{k+1,x}^{AB} - \widehat{H}_{k,x}^{AB}\right)}{\mathit{diff}\left(\widehat{H}_{k,x}^{AB} - \widehat{H}_{k-1,x}^{AB}\right)} \right| = \left| \frac{\sum_{x=1}^{N_s} \sum_{m=1}^{N} \sum_{n=1}^{N} \left| \widehat{H}_{k+1,x}^{XB}(m,n) - \widehat{H}_{k,x}^{AB}(m,n) \right|^2}{\sum_{x=1}^{N_s} \sum_{m=1}^{N} \sum_{n=1}^{N} \left| \widehat{H}_{k,x}^{AB}(m,n) - \widehat{H}_{k-1,x}^{AB}(m,n) \right|^2} \right| \underset{< H_0}{\overset{> H_1}{}} \eta_A \tag{7}$$

where T_B is the test statistic based on amplitude and phase information. We use T_A and T_B as the one-dimensional test statistic, respectively, for detecting spoofing attack. Unfortunately, it is hard to find the best threshold for achieving high accuracy authentication detection rate. To tackle this problem, we propose a learning algorithm based on AdaBoost to achieve physical layer authentication, in which T_A and T_B are used as training features.

3. Physical Authentication with AdaBoost Algorithm

In this section, we propose a learning algorithm based on AdaBoost for physical authentication.

3.1. AdaBoost Algorithm. AdaBoost is the abbreviation of adaptive boosting and developed by Yoav Freund [24] and is

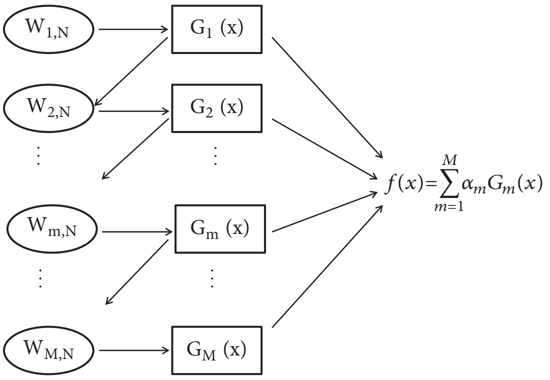


Figure 2: AdaBoost algorithm.

the most widely used form of boosting algorithm. Boosting is a powerful technique combined with base classifiers [25] to produce a form of committee whose performance can be significantly better than other base classifiers. The principal of AdaBoost algorithm is that this algorithm improves its performance by the iterative algorithm, which is adaptive in the sense that subsequent weak classifiers, called as learners, are adjusted to improve those instances misclassified by previous classifiers. AdaBoost can be seen as a particular method of training a boosted classifier. A boost classifier is a classifier as follows:

$$f(x) = \sum_{m=1}^{M} \alpha_m G_m(x) \tag{8}$$

where each $G_m(x)$ is a weak classifier that takes x as input and returns a value y_m indicating the class of x. The weak classifiers, each of classifiers is trained by using a weighted coefficient $w_{m,i}$ from the data set where the weighting coefficient associated depending on the performance of the weak classifiers such as decision tree (support vector machine) SVM, are trained in sequence. More specially, data points which are misclassified by one of the weak classifiers are being given greater weight, which are used to train the next weak classifier. As illustrated in Figure 2, once all the classifiers have been trained until there are no misclassified data points, then their final model is generated via a weight majority voting scheme.

3.2. Physical Authentication with AdaBoost Algorithm. The physical authentication with AdaBoost algorithm is proposed for detection spoofing. The performance chart of the algorithm is illustrated in Figure 3. Bob collects the channel matrix, $\widehat{\mathbf{H}}_1^{AB}$, which obtained by channel estimation using the pilot from Alice and records it. When Bob receives the next data frame from the Alice, the Bob collects channel information, $\widehat{\mathbf{H}}_2^{AB}$. Similarly, Bob collects continuous N-frames channel information from Alice and stores as $\widehat{\mathbf{H}}^{AB} = [\widehat{\mathbf{H}}_1^{AB}, \widehat{\mathbf{H}}_2^{AB}, \dots, \widehat{\mathbf{H}}_N^{AB}]$. In the same time an Eve sends the data frames to the Bob and claims that he is Alice. In practical communication scenarios, we do not know where and who Eves are. But in proposed scheme Eves are needed to be test training purpose. Therefore, one or several Eve nodes are set for this purpose. Bob continuously extracts the continuous N frames channel information from Eve and stores as $\widehat{\mathbf{H}}^{EB} = [\widehat{\mathbf{H}}_1^{EB}, \widehat{\mathbf{H}}_2^{EB}, \dots, \widehat{\mathbf{H}}_N^{EB}]$.

The data set is preprocessed by Bob. Firstly, Bob calculates the value of data set, $\widehat{\mathbf{H}}^{AB}$, $\widehat{\mathbf{H}}^{EB}$. Secondly, Bob calculates the test statistics based on test statistics T_A, T_B as

$$T_A^{XB}(k) = \left| \frac{diff\left(\widehat{H}_{k+1,x}^{XB} - \widehat{H}_{k,x}^{XB}\right)}{diff\left(\widehat{H}_{k,x}^{XB} - \widehat{H}_{k-1,x}^{XB}\right)} \right| = \left| \frac{\sum_{x=1}^{N_s} \sum_{m=1}^{N} \sum_{n=1}^{N} \left| \widehat{H}_{k+1,x}^{XB}(m,n) - \widehat{H}_{k,x}^{XB}(m,n) e^{j\widehat{\theta}(m,n)} \right|}{\sum_{x=1}^{N_s} \sum_{m=1}^{N} \sum_{n=1}^{N} \left| \widehat{H}_{k,x}^{XB}(m,n) - \widehat{H}_{k-1,x}^{XB}(m,n) e^{j\widehat{\theta}(m,n)} \right|} \right| \begin{matrix} > H_1 \\ < H_0 \end{matrix} \eta_A^X, \tag{9}$$

$$T_B^{XB}(k) = \left| \frac{diff\left(\widehat{H}_{k+1,x}^{XB} - \widehat{H}_{k,x}^{XB}\right)}{diff\left(\widehat{H}_{k,x}^{XB} - \widehat{H}_{k-1,x}^{XB}\right)} \right| = \left| \frac{\sum_{x=1}^{N_s} \sum_{m=1}^{N} \sum_{n=1}^{N} \left| \widehat{H}_{k+1,x}^{XB}(m,n) - \widehat{H}_{k,x}^{XB}(m,n) \right|^2}{\sum_{x=1}^{N_s} \sum_{m=1}^{N} \sum_{n=1}^{N} \left| \widehat{H}_{k,x}^{XB}(m,n) - \widehat{H}_{k-1,x}^{XB}(m,n) \right|^2} \right| \begin{matrix} > H_1 \\ < H_0 \end{matrix} \eta_B^X \tag{10}$$

Finally, Bob generates training data set of two categories. The first one is

$$T_A^{AB} = \{x_1, \dots, x_i, \dots, x_N, y_A\}, \tag{11a}$$

$$T_B^{AB} = \{x_1, \dots, x_i, \dots, x_N, y_A\}, \tag{11b}$$

where $x_i \in T_A^{AB}(k)$ or $x_i \in T_B^{AB}(k)$, $y_A = +1$, by substituting $\widehat{\mathbf{H}}^{AB}$, into (9), (10), yields T_A^{AB}, T_B^{AB}, and the value of y_A represents that the transmitter is the legal transmitter from Alice. And the second training set is

$$T_A^{EB} = \left\{x_1^E, \dots, x_i^E, \dots, x_N^E, y_B^E\right\} \tag{12a}$$

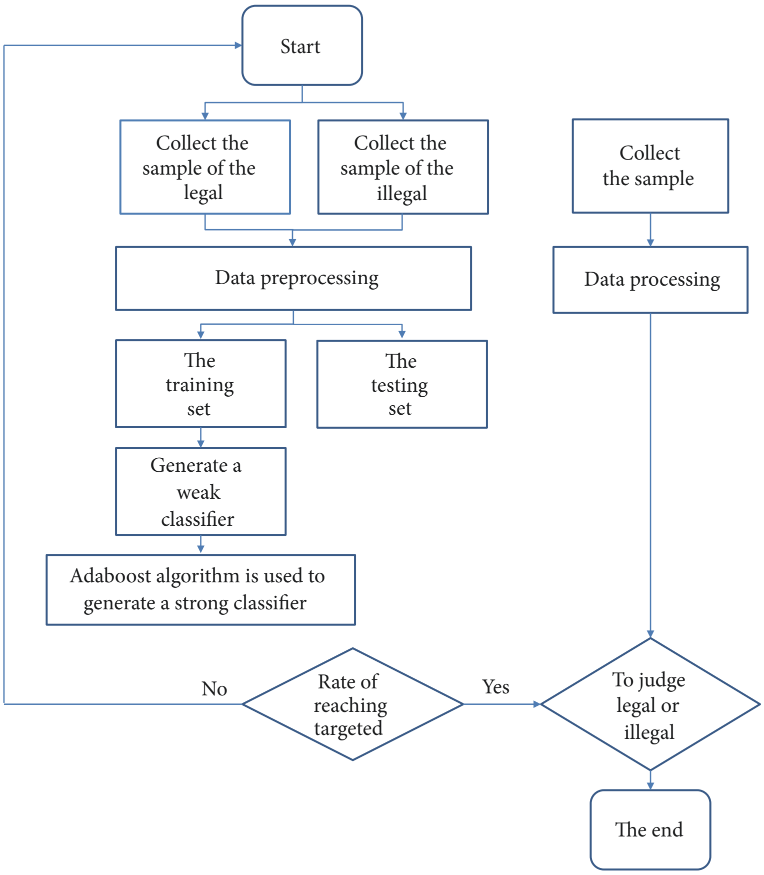

Start
Collect the sample of the legal
Collect the sample of the illegal
Collect the sample
Data preprocessing
Data processing
The training set
The testing set
Generate a weak classifier
Adaboost algorithm is used to generate a strong classifier
No
Rate of reaching targeted
Yes
To judge legal or illegal
The end

Figure 3: Physical authentication with AdaBoost algorithm.

$$T_B^{EB} = \left\{x_1^E, \ldots, x_i^E, \ldots, x_N^E, y_B^E\right\} \tag{12b}$$

where $x_i^E \in T_A{}^{EB}(k)$ or $x_i^E \in T_B{}^{EB}(k)$, $y_B^E = -1$, by substituting $\widehat{\mathbf{H}}^{EB}$, into (9) and (10), yields T_A^{EB} and T_B^{EB}, and the value of y_i represents that the transmitter is the illegal transmitter from Eve. Bob uses the two classification training data set $T_A^{AB}, T_B^{AB}, T_A^{EB}$, and T_B^{EB} as input training set.

Spoofing detection is essentially a two-classification problem, which is considered to be solved through AdaBoost algorithm. The training data is made up of a bunch of sample points. Each sample point comprises input sample x_i and label y_i where $y_i \in \{-1, 1\}$. Each sample point is given an associated weight parameter $w_{m,i}$, m means m-th training, and i means the number of sample points, which is initially set $1/i$ for all sample points. We suppose that we have a procedure available for training a weak classifier using weighted sample points. At each iteration of the training process, AdaBoost trains a new weak classifier by using the sample points in which the weighting coefficients are adjusted according to the performance of the previously trained weak classifier, so as to give greater weight to the misclassified data points, in which the classification error rate e_m is used to evaluate misclassified data set D_m

$$e_m = P\left(G_m\left(x_i\right) \neq y_i\right) = \sum_{i=1}^{2t} w_{mi} I\left(G_m\left(x_i\right) \neq y_i\right) \tag{13}$$

Then the coefficient α_m of G_m is calculate as

$$\alpha_m = \frac{1}{2} \log \frac{1 - e_m}{e_m} \tag{14}$$

Finally, we generate a final model that different weight is being given to different weak classifiers in (8). The AdaBoost algorithm is given as in Algorithm 2, in which the point of the training data can be doubled by combining with the one-dimension test statistics T_A and T_B together and become a new two-dimensional features authentication model for spoofing detection. Therefore, in the AdaBoost algorithm, the input training data set T is following two optional sets:

(1) One-dimension test statistics training data set:

$$\begin{aligned} T &= \left\{T_A^{AB}, T_A^{EB}\right\} \\ \text{or } T &= \left\{T_B^{AB}, T_B^{EB}\right\} \end{aligned} \tag{15}$$

(2) Two-dimension test statistics training data set:

$$T = \left\{\left(T_A^{AB}, T_B^{AB}\right), \left(T_A^{EB}, T_B^{EB}\right)\right\} \tag{16}$$

```
Input:
The channel information of legal transmitter or illgal transmitter:
Process:
1: Bob calculates the value of data set $\widehat{\mathbf{H}}^{AB}$ and $\widehat{\mathbf{H}}^{EB}$ from Alice and simulated Eve:
    $\widehat{\mathbf{H}}^{AB} = \left[\widehat{\mathbf{H}}_1^{AB}, \widehat{\mathbf{H}}_2^{AB}, \dots, \widehat{\mathbf{H}}_N^{AB}\right]$
    $\widehat{\mathbf{H}}^{EB} = \left[\widehat{\mathbf{H}}_1^{EB}, \widehat{\mathbf{H}}_2^{EB}, \dots, \widehat{\mathbf{H}}_N^{EB}\right]$
2: The data set are preprocessed by Bob:
3: The data set are divided into two parts, and the one is training data set and the other is testing data set:
4: Use training data set to get the weak classifier:
5: Use the Adaboost algorithm to generate a strong classifer:
6: The testing data set is used to verify whether the claasifier can achieve the target detection rate, otherwise it will return to
    the first step:
7: The final classifier is the authenticaton decision model, which can judge whether the new packets are legitimate or illegal:
    End
```

Algorithm 1: Physical authentication.

```
Input:
training data set $T$:
Process:
1: Initialize the weight distribution of the sample points:
    $D_1 = (w_{11}, \dots, w_{1i}, \dots, w_{1,2t}), \quad w_{1i} = \frac{1}{2t}, \; i = 1, 2, \dots, 2t$
2: for $m = 1$ to $M$ do, $m$ means $m$-th training
3: Use the training data set of $D_m$ to learn and get the weak classifier:
    $G_m(x) : x_i \longrightarrow \{-1, +1\}$
4: Calculate the classification error rate of $D_m$ on the training data set:
    $e_m = P(G_m(x_i) \neq y_i) = \sum_{i=1}^{2t} w_{mi} I\left(G_m\left(x_i\right) \neq y_i\right)$
5: Calculate the coefficient of $G_m$:
    $\alpha_m = \frac{1}{2} \log \frac{1 - e_m}{e_m}$
6: Update the weight distribution of the training data set:
    $D_{m+1} = (w_{m+1,1}, \dots, w_{m+1,i}, \dots, w_{m+1,2t}),$
    $w_{m+1,i} = \frac{w_{mi}}{Z_m} \exp\left(-\alpha_m y_i G_m\left(x_i\right)\right), \quad i = 1, 2, \dots, 2t$
    $Z_m = \sum_{i=1}^{2t} w_{mi} \exp\left(-\alpha_m y_i G_m\left(x_i\right)\right)$
7: Construct a linear combination of weak classifiers:
    $f(x) = \sum_{i=1}^{M} \alpha_m G_m(x)$
  End for
return: $G(x) = \operatorname{sign}(f(x))$
```

Algorithm 2: AdaBoost.

4. Experimental Verification

In this section, we will describe the system setup and the test process of measuring the Algorithm 1 for detecting Alice and Eve.

4.1. System Setup. We consider the spoofing detection of a receiver called Bob, the legal transmitter called Alice, and the spoofing node called Eve. They were placed in three separate locations in a room, surrounded by many other devices such as printers, desktops, and other types of equipment as shown in Figure 4. There are scattering and refraction phenomena in the room due to the presence of obstacles in the wireless channel from Alice to Bob and Eve to Bob. As shown in Figure 5, we set up experimental platform which implemented on USRPs, and experiments were performed in an indoor environment. Bob is equipped with an 8∗8 MIMO system, Alice is equipped with a 2∗2 MIMO system, and the spoofing node called Eve is equipped with a 2∗2 MIMO system. The signals are sent over 2 antennas each at center frequency 3.5GHz with bandwidth 2MHz.

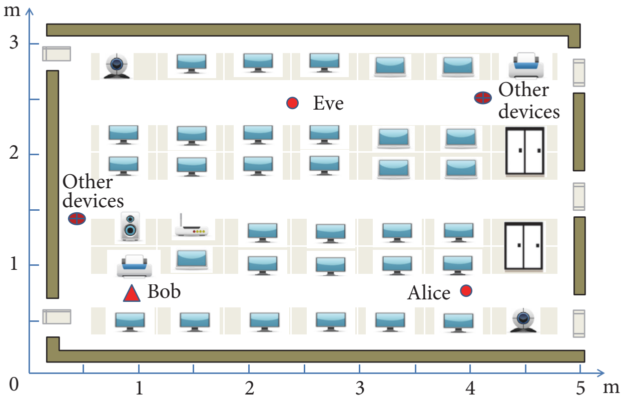


Figure 4: The experiments consisted of Alice, Bob, and Eve.

Figure 5: Real MIMO communication platform consisted of Alice, Bob, and Eve.

4.2. Experiment. In the experiment, the following steps are taken.

Step 1. Bob extracts channel information from Alice and Eve by the existing channel estimation mechanisms, respectively.

Step 2. Bob preprocesses the dataset according to (5), (7), (9), and (10) while the threshold is between [0, 1] (normalization).

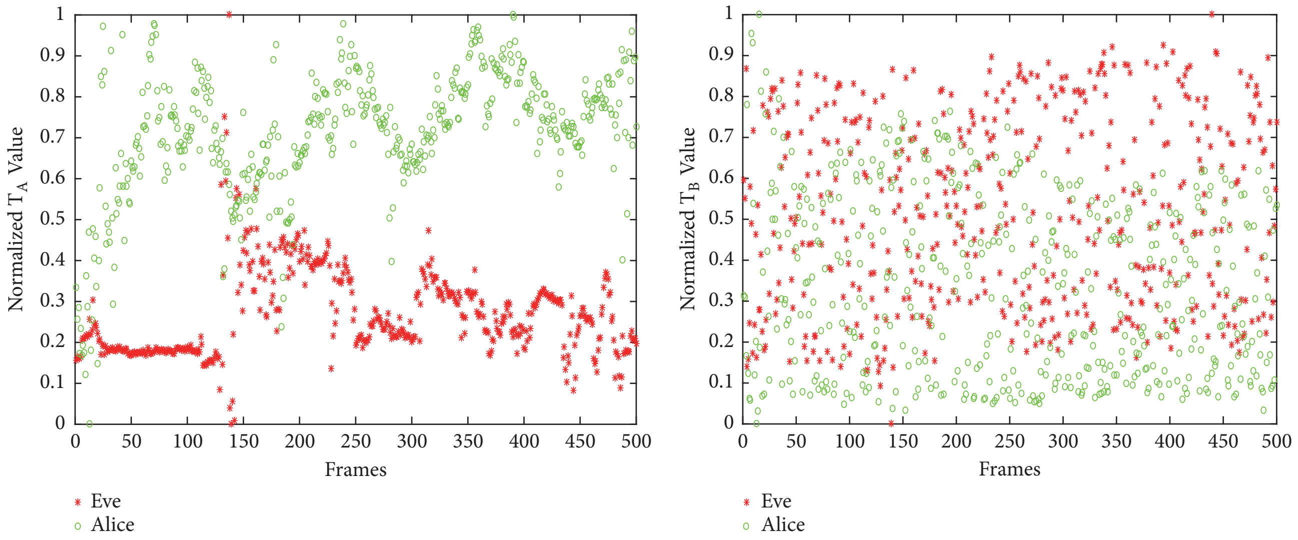
Step 3. Bob generates a training data set of two classifications according to (11a), (11b), (12a), and (12b).

Step 4. The two classification training data set T is generated according to (15) or (16).

Step 5. Bob is trained to generate a strong classifier based on the training data set of two classifications by using AdaBoost algorithm under Matlab program.

Step 6. Bob uses a strong classifier to judge the test set and obtain the authentication detection rate.

In the experiment, we consider that the collection frames are five hundred frames and the value of test statistic was normalized between 0 and 1. The test statistic T_A of channel information of the Alice and Bob as a function of frames is shown in Figure 6(a), in which the red points is $T_A(k)$ in (5) and green points is $T_A^E(k)$ in (9). As can be seen, there is the overlapped area. Meanwhile, from Figure 6(b), the overlapped area is large, when we chose the test statistic T_B of channel information in which the red points is $T_A(k)$ in (7) and green points is $T_A^E(k)$ in (10). It is clearly shown that it is difficult to acquire the best manual test threshold for the accuracy of authentication. Moreover, we use T_A, T_B, and the number of frames, respectively, to draw a three-dimensional plot. As shown in



(a) Normalized T_A of Alice and Eve

(b) Normalized T_B of Alice and Eve

FIGURE 6: Normalized T_A and T_B value of the legal transmitter Alice and the spoofing node Eve for spoofing detection with center frequency 3.5GHz with bandwidth 2MHz.

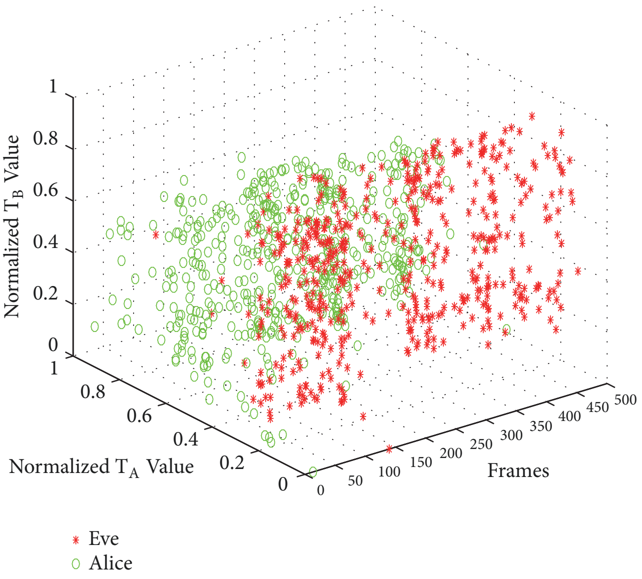


FIGURE 7: Normalized T_A value and T_B value of the legal transmitter Alice and the spoofing node Eve drawing three dimensional plot.

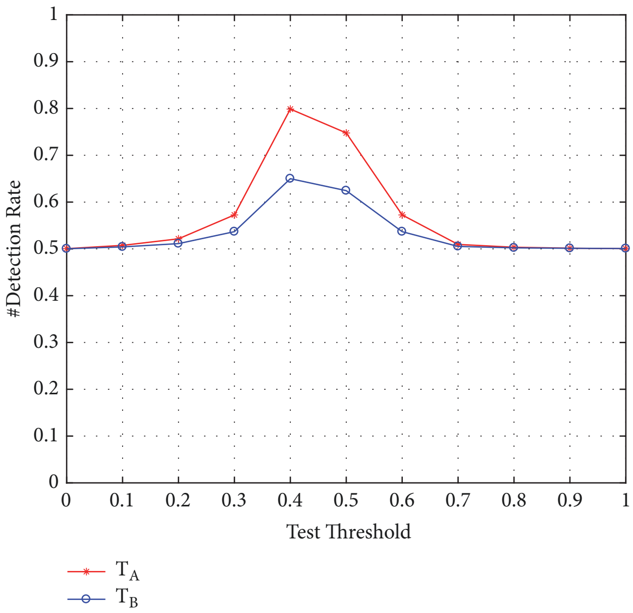


FIGURE 8: Correct classification rate of T_A and T_B.

Figure 7, obviously, it is hard to use the traditional manual threshold method to identify the identity of data sets in the three-dimensional condition. However, machine learning algorithm based the authentication model can effectively settle this problem and a dividing curved surface can perform the identification by the AdaBoost adaptive adjustment algorithm.

4.3. Simulation Results. In this section, simulation results are provided to demonstrate the performance of the proposed authentication scheme.

As a comparison, we considered the PHY-layer spoofing detection [15] with a varied test threshold. From the Figure 8, we can see that when test threshold equals 0.4, the best authentication detection rate results of using T_A or T_B reached 79.8% and 65.4%, respectively. In addition, our proposed method which combined two test statistics T_A and T_B as a two-dimensional feature can improve the accuracy of detection. We use T_A, T_B, and the number of frames, respectively, to draw a three-dimensional plot. Figure 9 illustrates the comparison of spoofing detection among the three methods, from which we can conclude that manual threshold method based on T_A test statistics can achieve 79.8% detection rate

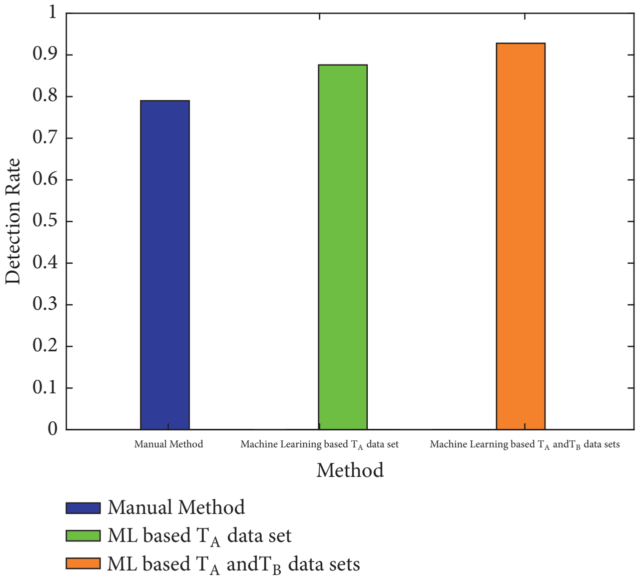


Figure 9: The simulation result with the different method of authentication scheme.

while machine learning based authentication method with T_A test statistic can acquire 87.1% detection rate and machine learning based authentication method with two-dimensional features T_A and T_B can achieve 91.3% accuracy rate with an additional 10% more computation complexity.

To sum up, the proposed authentication scheme achieves a superior performance over manual threshold strategy [15]. Based on the above observation, the proposed machine learning based authentication scheme with tow-dimensional feature not only exhibits excellent performance than manual method but also has higher authentication rate than that of the same algorithm with one-dimensional feature.

5. Conclusions

In this paper, machine learning algorithm based physical-layer channel authentication for the 5G wireless communication security is proposed. A machine learning authentication method could draw a conclusion whether the received packets are from a legitimate transmitter or from a counterfeiter by using one-dimension or two-dimensional joint features. The effectiveness of the proposed authentication scheme is validated by widely simulations. All the data used in the simulation are derived from real OFDM-MIMO communication platform, which provides a real communication environment. Moreover, the authentication results show that the novel methods provide a higher rate in detecting the spoofing attacks than those of the manual threshold based physical layer authentication schemes. The training of the classifier can be done offline. Therefore, the novel method can perform authentication fast. In addition, whether we can use more machine learning algorithms to further optimize our authentication model and find a better statistical test of large difference in channel information is issue that we need to deal with in the future.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was supported by NSFC (no. 61572114), Sichuan Sci & Tech. Achievements Transformation Project (no. 2016CC003), Sichuan Sci & Tech. Service Development Project (no. 18KJFWSF0368), Hunan Provincial Nature Science Foundation Project 2018JJ2535, Chile CONICYT FONDECYT Regular Project 1181809, and National Key R&D Program of China (2018YFB0904900 and 2018YFB0904905).

References

[1] M. Agiwal, A. Roy, and N. Saxena, “Next generation 5G wireless networks: A comprehensive survey,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 1617–1655, 2016.

[2] J. Thompson, X. Ge, and H.-C. Wu, “5G wireless communication systems: prospects and challenges [Guest Editorial],” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 62–64, 2014.

[3] C. Perera, C. H. Liu, S. Jayawardena, and M. Chen, “A survey on internet of things from industrial market perspective,” *IEEE Access*, vol. 2, pp. 1660–1679, 2014.

[4] A. Bogdanov, M. Knežević, G. Leander, D. Toz, K. Varıcı, and I. Verbauwhede, “SPONGENT: the design space of lightweight cryptographic hashing,” *IEEE Transactions on Computers*, vol. 62, no. 10, pp. 2041–2053, 2013.

[5] R. Zhang, L. Zhu, C. Xu, and Y. Yi, “An Efficient and secure RFID batch authentication protocol with group tags ownership Transfer,” *IEEE Collaboration and Internet Computing*, pp. 168–175, 2015.

[6] L. Hu, H. Wen, B. Wu et al., “Cooperative Jamming for Physical Layer Security Enhancement in Internet of Things,” *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 219–228, 2018.

[7] L. Hu, H. Wen, B. Wu, J. Tang, F. Pan, and R.-F. Liao, “Cooperative-Jamming-Aided Secrecy Enhancement in Wireless Networks with Passive Eavesdroppers,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 3, pp. 2108–2117, 2018.

[8] H. Wen, *Physical layer approaches for securing wireless communication systems*, Springer, New York, NY, USA, 2013.

[9] L. Hu, H. Wen, B. Wu, J. Tang, and F. Pan, “Adaptive Base Station Cooperation for Physical Layer Security in Two-Cell Wireless Networks,” *IEEE Access*, vol. 4, pp. 5607–5623, 2016.

[10] L. Hu, H. Wen, B. Wu, J. Tang, and F. Pan, “Adaptive Secure Transmission for Physical Layer Security in Cooperative Wireless Networks,” *IEEE Communications Letters*, vol. 21, no. 3, pp. 524–527, 2017.

[11] D. B. Faria and D. R. Cheriton, “Detecting identity-based attacks in wireless networks using signalprints,” in *Proceedings of the ACM Workshop on Wireless Security*, pp. 43–52, Los Angeles, Calif, USA, 2006.

[12] M. Demirbas and Y. Song, “An RSSI-based scheme for sybil attack detection in wireless sensor networks,” in *Proceedings of*

the WoWMoM 2006: 2006 International Symposium on a World of Wireless, Mobile and Multimedia Networks, pp. 564–568, June 2006.

[13] Y. Chen, W. Trappe, and R. P. Martin, "Detecting and localizing wireless spoofing attacks," in *Proceedings of the 2007 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks, SECON*, pp. 193–202, San Diego, Calif, USA, June 2007.

[14] N. Patwari and S. K. Kasera, "Robust location distinction using temporal link signatures," in *Proceedings of the ACM International Conference on Mobile Computing and NETWORKING*, pp. 111–122, 2007.

[15] H. Wen, Y. Wang, X. Zhu, J. Li, and L. Zhou, "Physical layer assist authentication technique for smart meter system," *IET Communications*, vol. 7, no. 3, pp. 189–197, 2013.

[16] J. K. Tugnait, "Wireless user authentication via comparison of power spectral densities," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1791–1802, 2013.

[17] Z. Jiang, J. Zhao, X. Li, J. Han, and W. Xi, "Rejecting the attack: Source authentication for Wi-Fi management frames using CSI Information," in *Proceedings of the IEEE INFOCOM 2013 - IEEE Conference on Computer Communications*, pp. 2544–2552, Turin, Italy, April 2013.

[18] X. Wu and Z. Yang, "Physical-layer authentication for multicarrier transmission," *IEEE Communications Letters*, vol. 19, no. 1, pp. 74–77, 2015.

[19] L. Xiao, Y. Li, G. Han, G. Liu, and W. Zhuang, "PHY-authentication protocol for spoofing detection in wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 10037–10047, 2016.

[20] L. Xiao, Y. Li, G. Han, G. Liu, and W. Zhuang, "PHY-Layer Spoofing Detection with Reinforcement Learning in Wireless Networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 12, pp. 10037–10048, 2016.

[21] P. Hao, X. Wang, and A. Refaey, "An enhanced cross-layer authentication mechanism for wireless communications based on PER and RSSI," in *Proceedings of the 2013 13th Canadian Workshop on Information Theory, CWIT 2013*, pp. 44–48, Canada, June 2013.

[22] S. Chen et al., "Machine-to-Machine communications in ultra-dense networks—A survey," *IEEE Communications Surveys & Tutorials*, vol. 1, no. 1, 99 pages, 2017.

[23] L. Xiao, L. J. Greenstein, N. B. Mandayam, and W. Trappe, "Channel-based detection of sybil attacks in wireless networks," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 3, pp. 492–503, 2009.

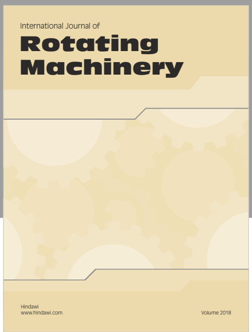
[24] H. Wen, P.-H. Ho, C. Qi, and G. Gong, "Physical layer assisted authentication for distributed ad hoc wireless sensor networks," *IET Information Security*, vol. 4, no. 4, pp. 390–396, 2010.

[25] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

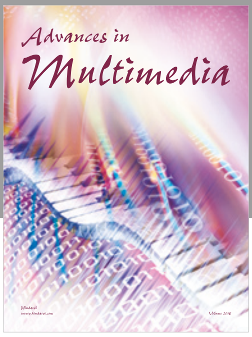
[26] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," *The Annals of Statistics*, vol. 28, no. 2, pp. 337–407, 2000.

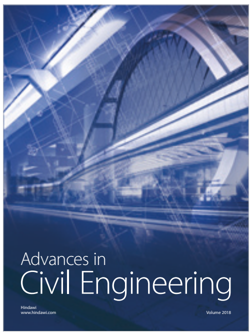
Journal of
Engineering

The Scientific
World Journal

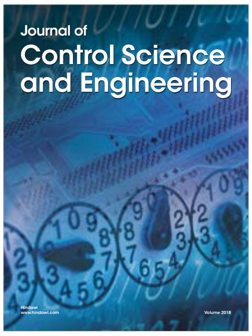

International Journal of
Rotating
Machinery

Journal of
Sensors

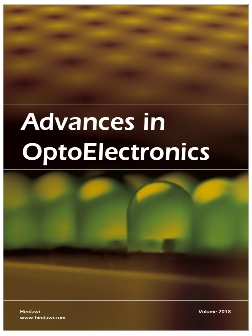

Advances in
Multimedia


Advances in
Civil Engineering

Journal of
Robotics

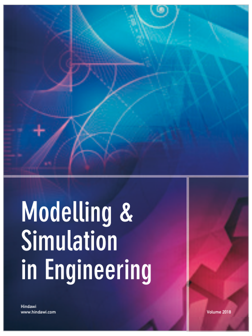

Journal of
Control Science
and Engineering

Journal of
Electrical and Computer
Engineering

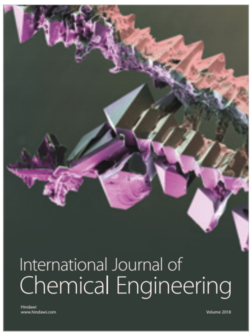

Advances in
OptoElectronics

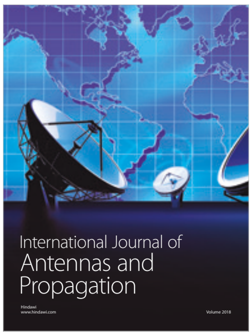
VLSI Design

International Journal of
Navigation and
Observation

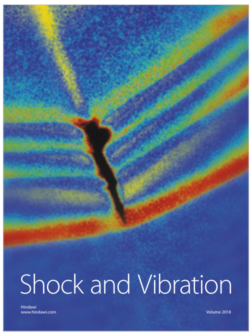

Modelling &
Simulation
in Engineering

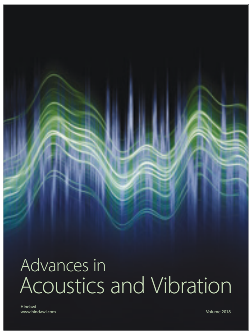
International Journal of
Aerospace
Engineering


International Journal of
Chemical Engineering


International Journal of
Antennas and
Propagation

Active and Passive
Electronic Components


Shock and Vibration


Advances in
Acoustics and Vibration