Unsupervised Detection of Sybil Attack in Wireless Networks

Yinghua Tian†, Song Chen† and Long Zhang†
†National Digital Switching System Engineering and Technological Research and Development Center (NDSC), Information Engineering University (IEU), Zhengzhou 450001, China.
*Email: tian.ying.hua@foxmail.com

Abstract. To address the problem of Sybil attacks in the wireless network, an unsupervised attack detection method using signal frequency bias distribution features is proposed. The method first estimates the signal frequency bias of the emitted signal of each wireless device and then calculates the signal frequency offset distribution characteristics, next it uses the DBSCAN clustering method to perform cluster analysis based on the distribution features. This method does not require prior learning of each device's signal features and is not affected by the changing of channel status. Experiments based on the real-world environment and commercial hardware show that the proposed method can effectively detect the Sybil attack and achieve high accuracy in identifying malicious devices.

1. Introduction
With the rapid spread of smartphones, laptops, and other smart devices in recent years, wireless local area network (WLAN) technology has become increasingly important in the modern world. A large number of wireless hotspots in cities can provide convenient network access services for both personal and business users. However, the openness of WLAN technology makes it vulnerable to cyber-attacks, which may threaten the information security of the public. How to improve WLAN security has become the focus in the field of network security. [1-3]

Sybil attack [4] is a major threat to wireless networks. Sybil attacks are introduced to denote a form of cyber-attack that hackers use a device to forge multiple virtual identities, and then use these false identities to launch attacks on other devices in the network. A hacker can forge several different client devices and connect to a target wireless hotspot by claiming them as different user identities, which can lead to a flooding attack on the wireless network and deplete network resources. [5] A hacker can also set multiple different wireless hotspots to spoof other nearby client devices. By this means, it could perform location spoofing attacks [6], hijack their Internet access, and intercept the user information [7] or block their connection to the Internet. [8] Due to the prevalence of WiFi technology, hackers can easily launch such attacks through laptops, rogue routers, and low-cost embedded devices.

Currently, WLAN security highly relies on cryptographic methods. However, there are several problems with this type of mechanism: (1)There is no validation scheme for the legality of the wireless hotspot; (2)The protocol may have security vulnerabilities; (3)As the computing capability increases, the encryption algorithm can be easily cracked; (4)Hackers can obtain the authentication password by social engineering and bypass the security scheme; (5)Hackers can exploit the unencrypted management frames to launch wireless attacks. [2]

To address these problems faced by traditional security methods, a major research direction in wireless security is detecting wireless attacks and enhancing wireless network security based on the features of raw wireless signals. [9-10] The signal features are also called as radio frequency fingerprints, which comes from the difference of hardware of the transmitter and the propagation...
process in wireless channels. It can provide a reliable way to verify the identities of wireless devices. Depending on the signal features used, these methods can be divided into two main categories: channel-based detection methods and hardware-based detection methods.

The channel feature-based detection methods mainly utilize the variation in the channel environment between illegal and legitimate devices due to their different locations for identification. In [11], a deep learning detection method based on channel measurement results is proposed to detect spoofing attacks by forging hardware identities. The research paper [12] proposed a spoofing attack detection method based on the indication of the strength of the received signal to identify malicious devices by measuring the strength of the transmitting signal of the device; Xiao Liang et al. proposed an attack detection method using the different wireless channel transmission response from the device to the receiver at different locations to detect Sybil attacks. [13-15] Since the channel-based detection methods depend on the varied wireless channel states, which means it is less adaptive to the environment changes.

Hardware-based detection methods can distinguish legitimate devices from illegitimate ones by extracting and detecting hardware features of wireless devices. Commonly used hardware features include carrier frequency offset, instantaneous signal spectrum, IQ imbalance, clock stability, phase shift, signal switching delay, etc. Researchers in [16-19] used the modulation errors caused by different devices, i.e., frequency offset, I/Q shift, and amplitude-phase error, as RF fingerprints of different devices. [20] proposed a method to identify WLAN devices by actively sending non-standard data frames and identifying them based on differences in the response of the target device. All of the above methods need to record and extract the features of legitimate devices in advance for learning, makes them hard to be applied in the real scenario.

In this paper, we propose an unsupervised detection method to detect the possible Sybil attacks in the environment. The proposed method differs from existing methods in the following aspects: (1)The method adopts an unsupervised learning scheme and does not require data to be collected in advance for training; (2)Unlike the detection method based on channel features, the method is independent of the channel state, so it is adaptive with the change of device's location; (3)The method does not bring additional communication overhead; (4)Compared with the existing detection method based on hardware features, the method in this paper only uses the distribution of signal frequency offset, makes it relatively easy to be integrated with existing systems.

2. Problem Model
This paper discusses the Sybil attack in which hackers use a small number of wireless devices to impersonate multiple different fake identities. As shown in figure 1, we refer to the devices set by the hacker as Sybil devices and other devices of the same type in the environment as legitimate devices; finally, we denote the attacked target in the environment as the target device. In the Sybil attack scenario, the hacker launches an attack on the target device by forging multiple false identities. In this scenario, the target device communicates with \((L+F)\) devices present in the environment; \(L\) of them are legitimate devices and the remaining \(F\) devices are Sybil devices forged by the attacker.

![Figure 1. Problem model of the Sybil attack.](image-url)
The main problem to be solved in this problem model is how to detect Sybil attacks and identify the Sybil devices from all the \((L + F)\) devices. In the Sybil attack scenario, since signal frames of Sybil devices are created by the same device of the attacker, therefore they will share similar signal features which provide the possibility to detect the Sybil attack. The proposed method detects the Sybil attack in the surrounding area and distinguishes the Sybil devices from other legitimate devices by this principle. Then we can secure the target device by simply block those Sybil devices.

3. Unsupervised Detection of Sybil attack

For each wireless device, due to device differences in the manufacturing process of its transmitter, particularly errors in the radio frequency oscillator, the transmitted signal may end up with a frequency offset compared with the standard signal. Due to the influence of factors such as local oscillation drift and the nonlinear characteristics of the RF amplifier, the frequency offset of the emitted signal is not fixed, but will form a random distribution. Based on the statistical analysis of the signal frequency offset distribution, the distribution feature can be used to verify the identity of each device.

3.1. Estimation of the Signal Frequency Offset

The most commonly used WLAN technology standards, including IEEE802.11a, IEEE802.11g, and IEEE802.11n are using the Orthogonal Frequency Division Multiplexing (OFDM) waveform in their physical layer. In these protocols, each physical frame of the wireless signal contains a preamble sequence, which can be used for signal detection and frequency synchronization. The structure of the preamble is shown in figure 2, which consists of a short training sequence and a long training sequence. The short training sequence contains 10 short training symbols (STS) with a duration of 0.8\(\mu\)s; the long training sequence contains 2 long training symbols (LTS) with a duration of 3.2\(\mu\)s. The two training sequences are separated by a guard interval (GI) of 1.6\(\mu\)s.

![Figure 2. Structure of the preamble sequence.](image)

In this paper, the frequency offset estimation is done by the delay correlation method, which is based on the periodic repetition of the short training sequence. [21] The digital sampling \(s[n]\) of the short training sequence is obtained by IQ sampling the signal at a sampling rate of 20MHz. When there is a carrier frequency deviation \(\Delta f\) between the transmitter and the receiver, the frequency offset causes the sampling values of the STS to rotate in phase. In this case, the following relationships exist in the sampled signal of the short training sequence:

\[
s[n + 16] = s[n]e^{\frac{2\pi \times 16 \times \Delta f}{f_s}}
\]  

(1)

In equation (1), \(f_s\) is the sampling frequency and \(s[n + 16]\) is the signal sampled 0.8\(\mu\)s after \(s[n]\). Therefore, an estimate of the carrier frequency offset can be made as follows.

\[
\hat{f} = \frac{f_s}{32\pi} \arg \left( \sum_{n=0}^{6} s[n] \ast \overline{s[n + 16]} \right)
\]

(2)

where \(\hat{f}\) is the estimation of carrier frequency offset and \(\arg(\ast)\) denotes the operation to calculate the phase of a complex.
3.2. Extraction of the Frequency Offset Distribution Characteristics

Due to the possible oscillation drift of the transmitter and the nonlinear characteristics of the RF amplifier, the frequency offset will behave as a random distribution. This distribution can be described by the mean and variance of the frequency offset records. The frequency offset of $i$th frames emitted by the $k$th wireless device in the acquisition environment is denoted as $\hat{f}_{k,i}$, where $i = 1, 2, \cdots I$.

First, we calculate the mean and variance for each device’s frequency offset records.

$$\text{Mean}(\hat{f})_k = \frac{1}{I} \sum_{i=1}^{I} \hat{f}_{k,i}$$

$$\text{Var}(\hat{f})_k = \frac{1}{I} \sum_{i=1}^{I} [\hat{f}_{k,i} - \text{Mean}(\hat{f})_k]^2$$

Then, the mean and variance of the frequency bias records are combined to obtain the frequency offset feature for the $k$th device.

$$\vec{v}_k = \{\text{Mean}(\hat{f})_k, \text{Var}(\hat{f})_k\}$$

The frequency offset feature of each device can be viewed as a point on a planar space. The relative position of these points can show the hardware similarity between each pair of devices. Figure 3 shows the feature points of all devices in a Sybil attack scenario. The horizontal axis is the statistical mean of the estimated frequency offset and the vertical axis is the statistical variance.

![Feature points of wireless devices in a Sybil attack scenario.](image)

There are multiple legitimate wireless devices in the scenario, which are scattered in different positions in the space because they have different frequency offset distribution characteristics. At the same time, the Sybil attack generates multiple fake devices, but the signals of these fake devices are actually emitted from the same device and thus they have similar features. As marked in figure 3, there is an aggregation of the feature points. The densely clustered points represent the Sybil devices generated by the attacker. We can detect Sybil attack based on this behavior.

3.3. Identify Sybil Devices by DBSCAN

Based on the frequency offset feature of each device, the following problems to be solved are: (1) how to distinguish whether there are signals in the environment with similar frequency offset features; (2) how to indentify the Sybil device.

In this paper, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) method [22] is used to discover Sybil attacks based on the aggregation behavior of the Sybil device’s feature points.

Compared to other clustering algorithms, DBSCAN is better suited for this scenario as it does not need to set the number of clusters. The process of identifying the Sybil device by DBSCAN is:
1. Set the minimum number of clusters as \( \text{min}\_\text{pts} \) and the clustering radius as \( \varepsilon \).
2. Find the points in the \( \varepsilon \)-neighborhood of every feature point and identify the core points with more than \( \text{min}\_\text{pts} \) neighbors.
3. Find the connected components of core points, ignoring non-core points.
4. Assign each non-core point to a nearby cluster if the cluster is an \( \varepsilon \)-neighborhood, otherwise assign it to noise.
5. Mark the remaining noise points as legitimate devices and the ones that have been clustered as Sybil devices.

3.4. Overall of the Attack Detection Method

In summary, combining the extraction method of the frequency offset features of wireless devices and the DBSCAN algorithm, the overall process of our Sybil attack detection method is summarized in figure 4. The method is unsupervised since it doesn’t need a training stage.

![Figure 4. Process of the attack detection method.](image)

Firstly, the signal emitted by each wireless device to be detected is collected to estimate its signal frequency offset; after that, we calculate the statistical distribution features by the signal frequency offset records of each device; then the DBSCAN clustering algorithm is used to analyze the feature points of all devices; finally, if we find more than one feature point in a cluster, it will send the Sybil attack warning, and all the devices in the cluster will be marked as the Sybil devices.

4. Experiments and Results

To evaluate the effectiveness of the proposed method, we build a Sybil attack simulation scenario and implement a prototype attack detection system based on the software-defined radio (SDR) platform to verify the proposed method in the real-world environment.

4.1. Experimental Setup

Figure 5 shows the experimental Sybil attack detection system built to verify the proposed algorithm. We use several routers, RaspberryPi 3b embedded devices, and a laptop with wireless NIC to simulate the public wireless environment and use a RaspberryPi 3b as the attack device to launch the fake wireless hotspot Sybil attack. The goal of this attack is to convince the target device to believe that there are a large number of wireless hotspots in the surrounding environment by constantly sending forged beacon frames. This attack can be used to spoof WiFi-based location systems, steal user information, or block user’s access to the Internet. Since wireless hotspots use a one-way broadcast communication mechanism to send beacon frames, which can’t be encrypted, the existing security mechanisms in the WiFi protocol cannot defend against this attack.
In this experiment, we build a prototype Sybil attack detection system based on SDR. The system uses a USRP (Universal Software Radio Peripheral) B210 SDR device to collect the wireless signal. The detection performance of the proposed method is analyzed by comparing the detection results with the list of devices set up in the test scenario.

4.2. Performance Evaluation
In the evaluation stage, we use the standard statistical indicator to check the detection performance of the proposed method. We compare the detection results across different scenarios, and count the number of fake hotspots correctly detected as True Positive (TP), the number of legitimate hotspots incorrectly detected as False Positive (FP), the number of legitimate hotspots detected correctly as True Negative (TN), and the number of fake hotspots incorrectly detected as False Negative (FP). Detection Accuracy is used as an overall performance metric to evaluate our method, which is defined in equation (5).

\[
\text{Accuracy} = \frac{TN + TP}{TN + TP + FN + FP}
\]  

(5)

We build several different test scenarios and collect frequency offset data from WiFi devices. Different sets of devices are selected to build 400 test scenarios, of which 200 scenarios simulate normal conditions and 200 scenarios are with the presence of Sybil attack. For the attack scenarios, an additional RaspberryPi 3b device is used to launch the Sybil attack.

Figure 6 shows the impact of different clustering parameters on the detection performance. The horizontal axis of the figure is the \( min \_ pts \) and the vertical axis is the detection accuracy. In each experiment, 200 frequency offset samples are collected for each wireless device. As shown in the figure, when the \( min \_ pts \) is set to 3 and the radius \( \epsilon \) is set to 0.04, the best detection accuracy (95.5%) is achieved.

Figure 7 shows the impact of the number of samples used on the performance when the optimal clustering parameter is used. It can be seen that the accuracy of the proposed method increases with the number of samples collected. In the case of 100 samples collected, more than 95% detection accuracy can be obtained. When the number of collected samples is increased to 140, a detection accuracy of 95.5% is achieved.
For comparison, under similar conditions, the detection accuracy of the channel feature and deep learning-based detection method proposed in [11] is 95% for spoofing attacks on wireless networks, and the signal phase error-based detection method proposed in [19] is 96%, both of which are supervised classification methods. According to the IEEE802.11 standard, the default beacon frame transmitting rate of WiFi hotspots is 10 Hz. Considering that the wireless hotspot also generates other data traffic, which means the proposed method can collect enough signal samples in less than 10 seconds for effective attack detection.

5. Conclusion

In this paper, an unsupervised attack detection method based on the clustering of signal frequency offset distribution features is proposed for Sybil attacks in WLAN systems. The method uses the DBSCAN algorithm to cluster the signal frequency offset distribution characteristics of wireless devices to detect the Sybil attack. The proposed method is evaluated in real-world experiments and the results show that the proposed method can achieve a 95.5% detection accuracy against the Sybil attack. Compared to the existing studies, the proposed method does not require additional communication overhead and spectrum resources; it does not need a training phase to learn the signal features of legitimate devices in advance. In further works, we will introduce other signal features of wireless devices into our method to achieve better performance.

6. References

[1] Cho Y, Jo J and Jeong C 2018 A study on vulnerability analysis and security plan through public WiFi attack Proc. Conf. of the Korean Institute of Information and Communication Sciences (The Korea Institute of Information and Communication Engineering) pp 493–496
[2] Peng H 2012 WiFi network information security analysis research 2nd Int. Conf. on Consumer Electronics, Communications and Networks (IEEE) pp 2243–2245
[3] Zou Y, Zhu J, Wang X and Hanzo L 2016 A survey on wireless security: Technical challenges, recent advances, and future trends Proceedings of the IEEE (IEEE) pp 1727–1765
[4] Douceur J R 2002 The sybil attack Int. workshop on peer-to-peer systems (Springer) pp 251–260
[5] Newsome J, Shi E, Song D and Perrig A 2004 The sybil attack in sensor networks: analysis & defenses 3rd Int. Sym. on information processing in sensor networks (IEEE) pp 259–268
[6] Tippenhauer N O, Rasmussen K B, Pöpper C and Čapkun S 2009 Attacks on public WLAN-based positioning systems Proc. of the 7th Int. Conf. on Mobile systems, applications, and services (ACM) pp 29–40
[7] Cunche M 2014 I know your MAC address: Targeted tracking of individual using WiFi Journal of Computer Virology and Hacking Techniques 10 (Springer) 219–227
[8] Dondyk E and Zou C C 2013 Denial of convenience attack to smartphones using a fake Wi-Fi access point IEEE 10th Consumer Communications and Networking Conf. (IEEE) pp 164–170

**Figure 6.** Detection performance vs different parameters.

**Figure 7.** Detection performance vs the number of samples.
[9] Ureten O and Serinken N 2007 Wireless security through RF fingerprinting Canadian Journal of Electrical and Computer Engineering 32 (IEEE) pp 27–33

[10] Xu Q, Zheng R, Saad W and Han Z 2015 Device fingerprinting in wireless networks: Challenges and opportunities IEEE Communications Surveys & Tutorials 18 (IEEE) pp 94–104

[11] Jiang P, Wu H, Wang C and Xin C 2018 Virtual MAC spoofing detection through deep learning 2018 IEEE Int. Conf. on Communications (IEEE) pp 1–6

[12] Chen Y, Trappe W and Martin R P 2007 Detecting and localizing wireless spoofing attacks 4th Annual IEEE Communications Society Conf. on Sensor, Mesh and Ad Hoc Communications and Networks (IEEE) pp 193–202

[13] Xiao L, Greenstein L J, Mandayam N B and Trappe W 2009 Channel-based spoofing detection in frequency-selective rayleigh channels IEEE Transactions on Wireless Communications 8 (IEEE) pp 5948–56

[14] Liu H, Wang Y, Liu J, Yang J and Chen Y 2014 Practical user authentication leveraging channel state information Proc. of the 9th ACM Sym. on Information, Computer and Communications Security (New York: ACM) pp 389-400

[15] Wang C, Zhu L, Gong L, Zhao Z, Yang L, Liu Z and Cheng X 2018 Accurate sybil attack detection based on fine-grained physical channel information Sensors 18 p 878

[16] Suski II W C, Temple M A, Mendenhall M J and Mills R F 2008 Using spectral fingerprints to improve wireless network security Global Telecommunications Conf. (IEEE) pp 1–5

[17] Brik V, Banerjee S, Gruteser M and Oh S 2008 Wireless device identification with radiometric signatures Proc. of the 14th ACM Int. Conf. on Mobile computing and networking (ACM) pp 116–127

[18] Nguyen N T, Zheng G, Han Z and Zheng R 2011 Device fingerprinting to enhance wireless security using nonparametric bayesian method IEEE INFOCOM (IEEE) pp 1404–1412

[19] Liu P, Yang P, Song W-Z, Yan Y and Li X-Y 2019 Real-time identification of rogue WiFi connections using environment-independent physical features IEEE INFOCOM (IEEE) pp 190–198

[20] Bratus S, Cornelius C, Kotz D and Peebles D 2008 Active behavioral fingerprinting of wireless devices Proc. of the first ACM conference on Wireless network security (ACM) pp 56–61

[21] Sourour E, El-Ghoroury H and McNeill D 2004 Frequency offset estimation and correction in the IEEE 802.11a WLAN IEEE 60th Vehicular Technology Conf. (IEEE) pp 4923-4927

[22] Ester M, Kriegel H-P, Sander J and Xu X 1996 A density-based algorithm for discovering clusters in large spatial databases with noise Proc. of the Second Int. Conf. on Knowledge Discovery and Data Mining (AAAI Press) pp 226–31

[23] IEEE 802.11 Working Group. Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications: higher-speed physical layer extension in the 2.4 GHz band, (IEEE, 1999).