





Research Article

Wasserstein Metric-Based Location Spoofing Attack Detection in WiFi Positioning Systems

Yinghua Tian ^{1,2}, Nae Zheng ^{1,2}, Xiang Chen ², and Liuyang Gao ^{1,2}

¹National Digital Switching System Engineering and Technological Research and Development Center (NDSC), Information Engineering University (IEU), Zhengzhou 450001, China

²State Key Laboratory of Complex Electromagnetic Environment Effects on Electronics and Information System (CEMEE), Luoyang 471032, China

Correspondence should be addressed to Nae Zheng; 13837122426@163.com

Received 22 September 2020; Revised 1 March 2021; Accepted 23 March 2021; Published 7 April 2021

Academic Editor: Jinwei Wang

Copyright © 2021 Yinghua Tian et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

WiFi positioning systems (WPS) have been introduced as parts of 5G location services (LCS) to provide fast positioning results of user devices in urban areas. However, they are prominently threatened by location spoofing attacks. To end this, we present a Wasserstein metric-based attack detection scheme to counter the location spoofing attacks in the WPS. The Wasserstein metric is used to measure the similarity of each two hotspots by their signal's frequency offset distribution features. Then, we apply the clustering method to find the fake hotspots which are generated by the same device. When applied with WPS, the proposed method can prevent location spoofing by filtering out the fake hotspots set by attackers. We set up experimental tests by commercial WiFi devices, which show that our method can detect fake devices with 99% accuracy. Finally, the real-world test shows our method can effectively secure the positioning results against location spoofing attacks.

1. Introduction

Driven by the demands of location-based services (LBS) and the Internet of Things (IoT), 3GPP Release 16 has introduced a variety of positioning technologies as supplements to the cellular-based positioning method in the 5G location services (LCS) [1]. As shown in Figure 1, the hybrid LCS architecture integrates global navigation satellite systems (GNSS) and WiFi positioning systems (WPS), to offer a positioning result of high accuracy, availability, and reliability. Applications such as autonomous driving, unmanned aerial vehicles, and massive IoT tracking will benefit from the improvement of LCS.

In the architecture of LCS, GNSS can provide the most accurate position for the user's mobile device in the open area, but it suffers from the poor visibility of satellites in the urban area and high power consumption [2]. On the contrary, WPS can provide fast and relatively less accurate positioning results in the indoor area where other methods are inadequate due to multipath or signal blockage. But the

introduction of those methods also imports potential threat vectors from those technologies. For now, although the design of 5G LCS has evaluated the credibility of the positioning results based on if the network access is trusted. But it did not propose specific schemes to secure the positioning results from those methods, which may damage its ability to serve the high-level applications. Due to the opening nature of WiFi technology, the WPS is the weakest part of LCS in the cases of both trusted and untrusted access.

WPS uses massively deployed WiFi access points (AP) in urban areas as location anchors. The implementation of WPS is shown in Figure 2. In the first stage, WPS service providers collect information on APs in urban areas and build a hotspot database that matches APs to their actual geolocation. In the second stage, when the user needs to obtain the current location, the mobile device gathers the APs' information in the surrounding environment and sends them to WPS service providers. 5G LCS requests the basic service set identifier (BSSID), which usually is the MAC address of nearby APs to locate the mobile device. The

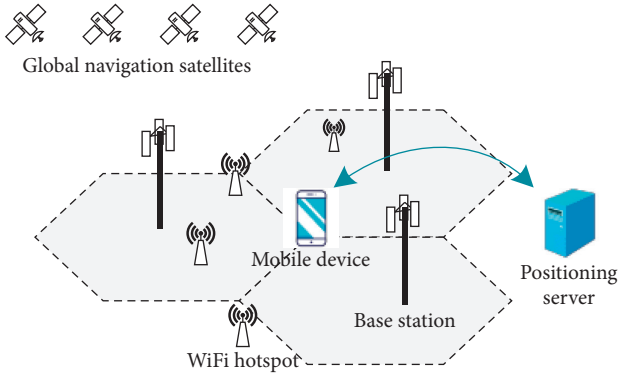


FIGURE 1: Positioning methods in the 5G location services.

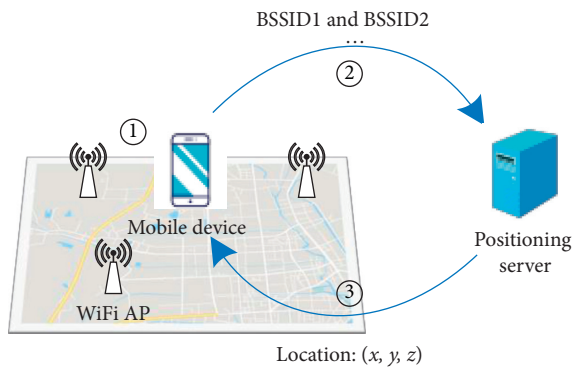


FIGURE 2: Procedure of WiFi positioning.

received signal strength indication (RSSI) and service set identifier (SSID) may also be used for positioning. Finally, WPS calculates the positioning result by querying the geolocation corresponding to the APs.

In the process of positioning by WPS, the mobile device cannot verify the identity of the wireless hotspots, which makes WPS vulnerable to location spoofing attacks. The attacker can convince the WPS to output a fake position that is far away from the actual location of the mobile device. This spoofing attack technique leads to various kinds of security threats and user privacy leakage.

In 2009, researchers first proposed the location spoofing attacks method to WPS, which aims at widely used smartphones and commercial WPS services such as Google Maps and Skyhook [3]. They deceived the positioning results of smartphones by transmitting fake WiFi hotspot beacons. In the study by Matte et al. [4], researchers combined the WiFi location spoofing method with the social network information leakage to obtain the user identity of nearby smartphones, which poses a great threat to personal privacy. Reference [5] proposed a method of using fake WiFi hotspots in specific places to recognize the MAC address of a specific user's device and track the activities. Since location-based services have been integrated into a wide range of mobile applications (apps) on the smartphone, the location spoofing attacks in WPS may directly threaten the usability of those applications and the user's information security. As shown in Table 1, the location spoofing attacks against WPS have effects on various types of mobile

apps. Hackers can leverage this spoofing method to perform attacks such as hijacking the positioning results of navigation apps, injecting a wrong location to the online order system, pushing misleading information through the location-based advertising system, and obtaining the nearby user's identity in social network apps.

Figure 3 presents a proof-of-concept WiFi location spoofing attack. In the experiment, the attacker manipulates the positioning result of the Google Maps on a smartphone by simply setting up a rogue WiFi access point and sending fake WiFi beacons. Due to the open nature of the WiFi technique, the hacker can easily perform this kind of attack by the commercially off-the-shelf (COTS) devices. Figure 4 shows several WiFi hacking devices that can be used to launch location spoofing attacks against WPS.

As the WPS was introduced into 5G location systems, the location spoofing attack in WPS also becomes a potential threat to the 5G LCS. The above discussion motivates us to propose a novel method to tackle location spoofing attacks. The proposed method is implied on mobile devices, and it can secure the positioning results of both the WPS subsystem in 5G LCS and the standalone WPS.

2. Related Works

To counter the spoofing attacks in the WiFi positioning system, researchers have developed a variety of verification and attack detection methods to secure positioning results. Reference [6] proposes the Location Validation System (LVS) to tackle location spoofing attacks. LVS verifies the positioning result by testing whether multiple users in the same area can communicate with each other. At the same time, to deal with collusion attacks by multiple attackers, a scoring algorithm of each node is proposed to solve the problem. This method relies on collaborative sensing from multiple users and is not suitable for scenarios with low user density. In their work [7], Ye et al. utilize the unique tags with the time and location attributes extracted from the frames of a hotspot to verify its' validity. The implementation of this method needs to introduce a credible witness device nearby. Reference [8] mainly uses multiple wireless APs as sensors to collect the Received Signal Strength Indicator (RSSI) of the user devices and detects the attack based on the correlation between the RSSI and the devices. Reference [9] aims at detecting spoofing attackers in wireless networks through RSSI and then locating the position of attackers. This method requires the cooperation of multiple devices at different locations. But RSSI is not a reliable feature and the attacker can easily bypass these RSSI-based detection methods by simply changing the signal transmission power.

Another group of detection methods is based on RF fingerprint (RFF). RFF refers to the unique signal characteristics of wireless communication devices due to the variability in their hardware and the propagation channel. As a result, it can be used to verify the identity of a device. Commonly used RFF includes the following: carrier frequency offset, channel status information (CSI), power spectrum, IQ imbalance, clock stability, signal phase offset, the delay between transmission and reception, etc. Wang

TABLE 1: Mobile applications vulnerable to the WiFi location spoofing attacks.

Application type	Name	Functions vulnerable to the WiFi location spoofing attacks			
		Positioning	Online ordering	Message pushing	Sharing location
Map and navigation	Google Maps	✓	—	—	—
	MapQuest	✓	—	—	—
	Baidu Map	✓	✓	—	—
Location review and online ordering	Foursquare	✓	—	✓	—
	Groupon	✓	✓	✓	—
	Yelp	✓	✓	✓	—
	Trip Advisor	✓	✓	✓	—
Social media	Twitter	✓	—	✓	✓
	TikTok	✓	—	✓	—
	Sina Weibo	✓	—	—	✓
	WeChat	✓	✓	✓	✓

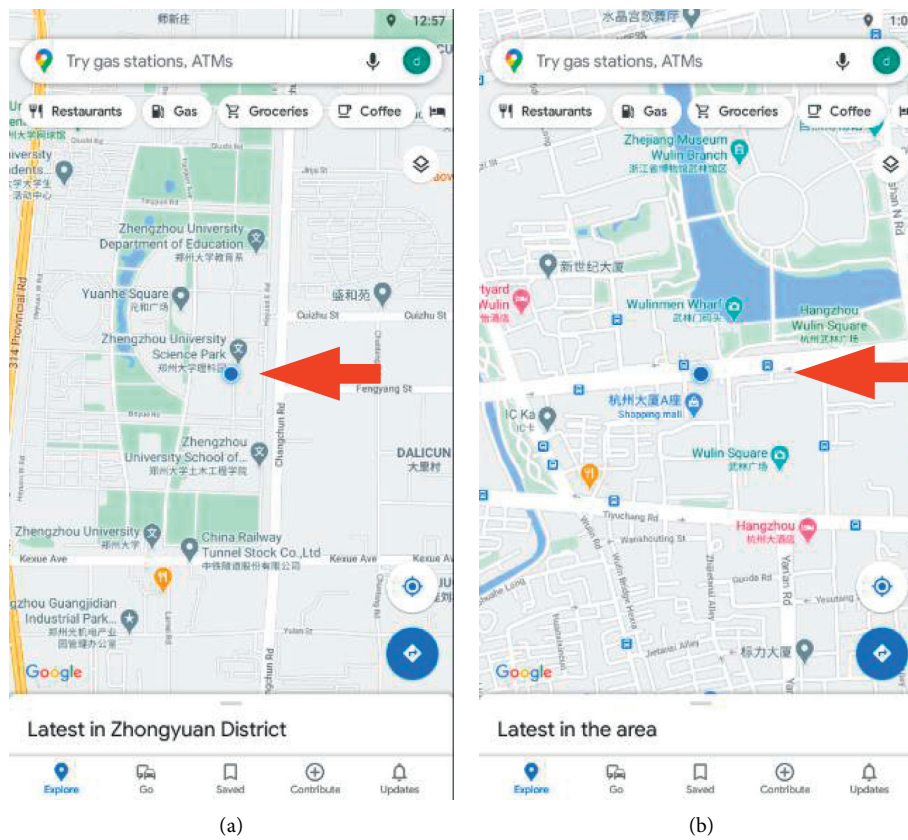


FIGURE 3: Positioning results of Google Maps on Android before and during the spoofing attack. (a) The real position is in the Zhengzhou University, Henan Province, China. (b) The positioning result is hijacked to Hangzhou, Zhejiang Province, China, which is about 800 km away from the real geolocation.

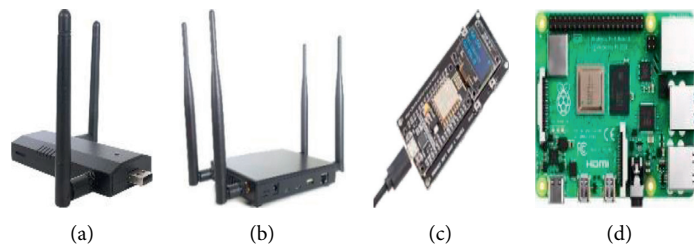


FIGURE 4: WiFi hacking devices. (a) Pocket-size WiFi attack tool. (b) Rogue WiFi access point. (c) Card-size WiFi attack tool based on ESP8266 chip. (d) WiFi attack tool based on RaspberryPi.

et al. [10] leveraged the channel status to detect spoofing attacks in wireless networks. Reference [11] identifies wireless devices by using the spectrum features of transmitters. Reference [12] combines the phase differences and the frequency differences to enhance wireless security. Recently, researchers introduced machine learning methods into RFF recognition, which can provide better performance compared with traditional methods [13–16].

3. Problem Model

In this section, we first introduce the current model of spoofing attacks against WiFi positioning systems as the background. Then, we describe the attack detection model to deal with this problem. The detailed implementation of our method will be introduced in Section 4.

3.1. Attack Model. The process of spoofing attack against WPS is shown in Figure 5. In the preparation stage, the attacker needs to collect information such as the SSID and MAC address of WiFi access points (AP) in the target location. This process can be achieved by querying the public WiFi geolocation database such as WIGLE [17] or collecting information in the real world. In the attack stage, the attacker uses WiFi attack devices to send forged WiFi beacon frames. Those beacon frames claim multiple fake APs which do not exist in the environment. When the user's mobile device received the forged WiFi frames, it sends a list of nearby APs which contains fake data injected by the attacker to the positioning server. Finally, the server receives the AP list and responds a false positioning result to the mobile device. The attacker can manipulate the positioning result of the mobile device in real time by modifying the fake beacon frames.

3.2. Attack Detection Model. Figure 6 shows a general detection model to counter the spoofing attack. There are four roles in the attack detection model: legitimate APs, fake APs, the user's mobile device, and the positioning server. Among them, legitimate APs are the real wireless devices in the environment and can provide correct location information. Fake APs set by the attacker intends to mislead the target mobile device to a false positioning result. In order to counter the WiFi location spoofing attack, this article intends to design an attack detection system to tell whether an AP is a legitimate device or a fake device based on the RFF of their WiFi signals. Then, the mobile device can filter out the fake APs and exclude the fake APs before sending the positioning request. Finally, the mobile device can obtain the correct positioning result from the positioning server.

4. Spoofing Attack Detection Based on Wasserstein Metric

4.1. Basic Idea. Based on the principle of the RF fingerprint, we know that the signals emitted by different wireless devices carry different features. In the spoofing attack scenario, since WiFi signal frames of fake APs are created by the same

device of the attacker, therefore, they will share similar signal features which provide the possibility to detect spoofing attacks.

In this paper, we utilize the frequency offset as the signal feature to identify devices. The frequency offset is the difference between the carrier frequency of the transmitted signal and the ideal signal. Different transmitters usually have different frequency offsets. Compared with other RF fingerprints, it can be obtained from the carrier synchronization stage [18] of the WiFi receiver and does not need additional hardware, which gives it advantages in implementation and cost.

Figure 7(a) shows the frequency offsets of WiFi signal frames sent by three wireless APs within 10 seconds. Figure 7(b) shows the kernel density estimation (KDE) of the signal frequency offsets of each device. It can be seen that, for different devices, their statistical distributions are different. Figure 7(c) shows the signal frequency offsets of WiFi frames sent by the attack device which claims three fake identities; Figure 7(d) shows the KDE of these three fake devices. As shown in Figure 7(d), although these signal frames claim to come from different devices, they share a similar distribution of frequency offsets, which indicates that they are transmitted by the same device.

Based on previous analysis, the basic idea of our detection method is to find if there are WiFi APs that have similar frequency offset distributions. To accomplish this work, we combine the Wasserstein metric and the clustering algorithm in our method. The Wasserstein metric is a practical way to compare the probability distributions of two random processes; for instance, random signal frequency offsets of a WiFi AP in this paper. The clustering algorithm is applied to find similar APs in the point of Wasserstein metric. Finally, the APs in the cluster are classified as fake devices set by the attacker.

4.2. Wasserstein Metric. The Wasserstein metric is a distance function defined to measure between probability distributions. It is also called Wasserstein distance. This concept was introduced in the optimal transport problem, where distribution is viewed as a pile of earth and the earthmover needs to move the mass to turn one pile into the other [19]. The Wasserstein metric is the minimum cost, which is the amount of earth times the distance it needs to be moved. For the problem of this paper, the expression for Wasserstein distance is

$$W(P, Q) = \inf_{\gamma \sim \Pi(P, Q)} E_{(x, y) \sim \gamma} \|x - y\|, \quad (1)$$

where P and Q are the frequency offset probability distributions of the signals sent by two wireless APs and $\Pi(P, Q)$ is the set of all possible joint distributions when P and Q are combined. A sample $(x, y) \sim \gamma$ is taken from the joint distribution γ , the distance between them $\|x - y\|$ is calculated, and the distance expectation $E_{(x, y) \sim \gamma} \|x - y\|$ under that joint distribution is calculated based on it. $\inf(\cdot)$ in the equation represents the minimum value taken, i.e., the lower

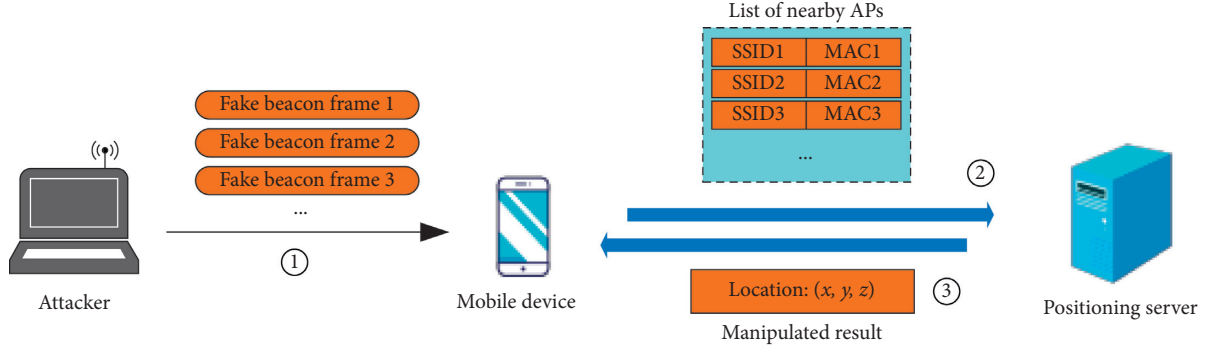


FIGURE 5: Spoofing attack model in the WiFi positioning system.

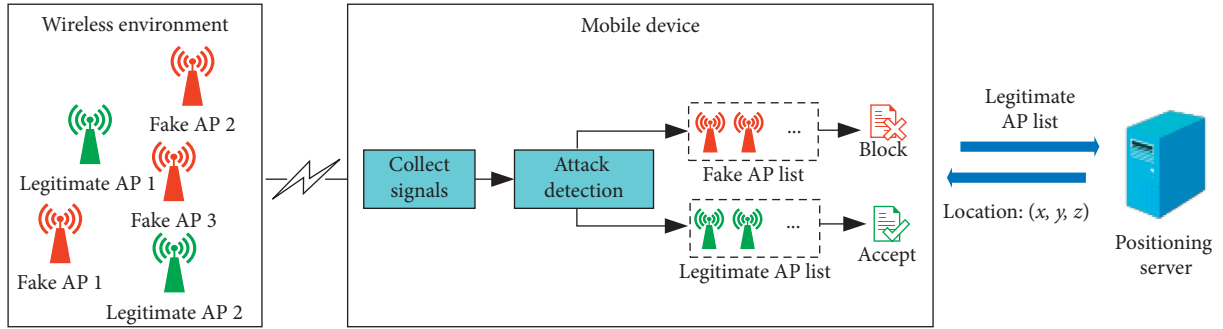


FIGURE 6: Attack detection model in the WiFi positioning system.

bound under all possible joint distributions, defined as the Wasserstein metric.

In the case of one-dimensional distributions, the closed-form solution of the Wasserstein metric is

$$W_p(P, Q) = \left(\int_0^1 |F^{-1}(z) - G^{-1}(z)|^p dz \right)^{1/p}, \quad (2)$$

where $F(z)$ and $G(z)$ are the cumulative distributions' function of the two hotspots' signal frequency offset distributions, i.e., the integral of the probability density function.

In practice, we cannot obtain the closed form of $F(z)$ and $G(z)$, but can estimate it by random sampling. Assuming that samples of random distribution P are obtained by collecting the frame from a device and are ranked from small to large as X_1, X_2, \dots, X_n and another n incremental sample results Y_1, Y_2, \dots, Y_n for the frequency offset distribution Q , then

$$\begin{aligned} \hat{F}(x) &= \frac{1}{n} \sum_{i=1}^n \varepsilon(x - X_i), \\ \hat{G}(x) &= \frac{1}{n} \sum_{i=1}^n \varepsilon(x - Y_i), \end{aligned} \quad (3)$$

where $\varepsilon(x)$ is the unit step function and $\hat{F}(x)$ and $\hat{G}(x)$ are probability accumulation function estimates of the frequency bias distributions P and Q . In the case of $p = 1$, we can deduce the estimation of $W_p(P, Q)$ as

$$\begin{aligned} \hat{W}_{p=1}(P, Q) &= \left(\int_0^1 |\hat{F}^{-1}(z) - \hat{G}^{-1}(z)|^p dz \right)^{1/p}, \\ &= \int_0^1 |\hat{F}^{-1}(z) - \hat{G}^{-1}(z)| dz, \\ &= \int_{-\infty}^{\infty} |\hat{F}(x) - \hat{G}(x)| dx, \\ &= \int_{-\infty}^{\infty} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon(x - X_i) - \frac{1}{n} \sum_{i=1}^n \varepsilon(x - Y_i) \right| dx, \\ &= \frac{1}{n} \int_{-\infty}^{\infty} \sum_{i=1}^n |\varepsilon(x - X_i) - \varepsilon(x - Y_i)| dx, \\ &= \frac{1}{n} \int_{-\infty}^{\infty} \sum_{i=1}^n |X_i - Y_i| dx, \\ &= \frac{1}{n} \sum_{i=1}^n |X_i - Y_i|. \end{aligned} \quad (4)$$

Thus, we can estimate the Wasserstein metric between two APs' by their frequency offset samples:

$$\hat{W}_{p=1}(P, Q) = \frac{1}{n} \sum_{i=1}^n |X_i - Y_i|. \quad (5)$$

Based on the above analysis, it is possible to measure the similarity between multiple APs by calculating the Wasserstein metric between the frequency offset distributions of their emitting signals. Figure 8(a) shows the Wasserstein metric matrix of signals from a set of legitimate Wi-Fi APs.

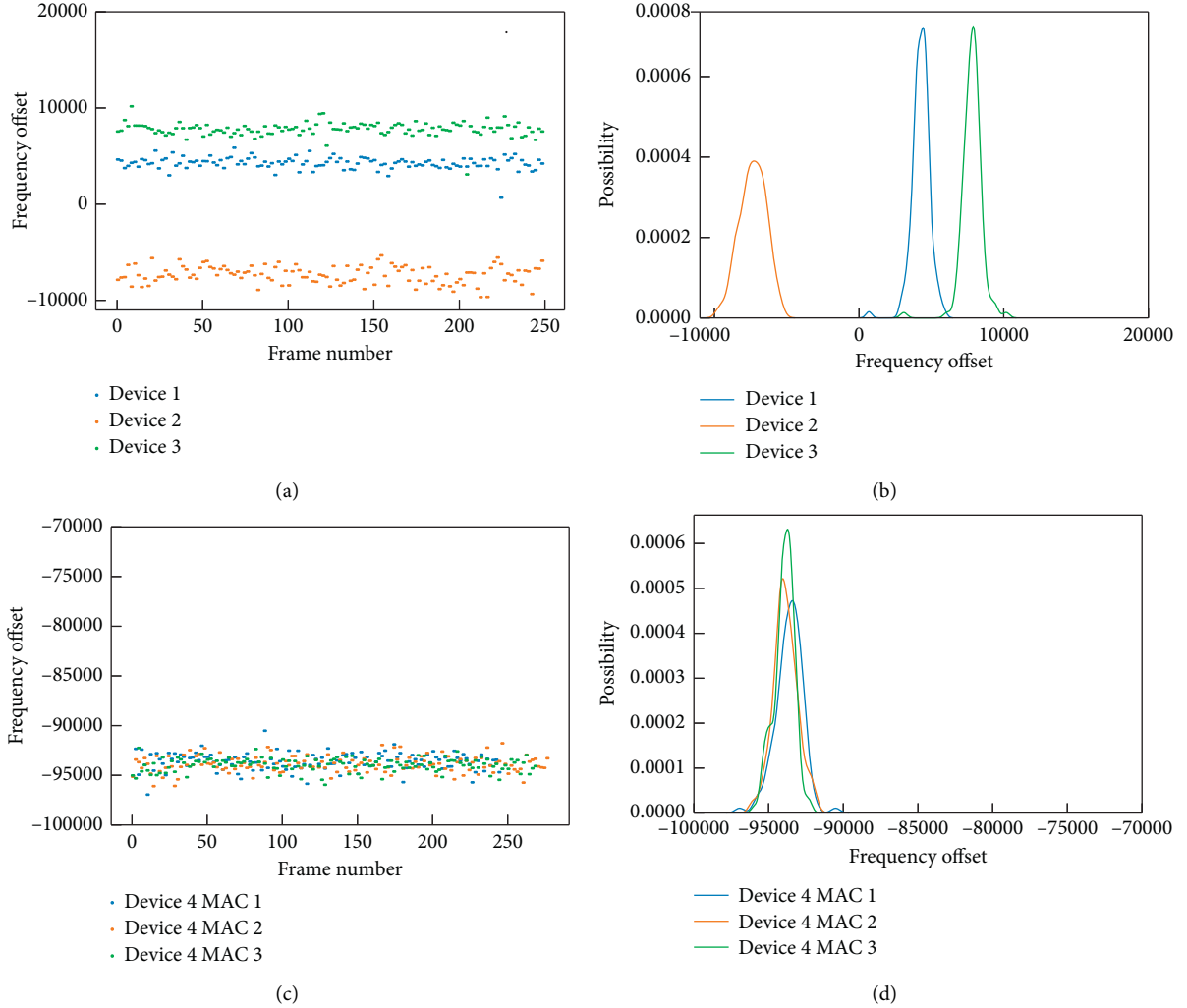


FIGURE 7: Frequency offsets and their distributions. (a) Frequency offsets of WiFi signal frames sent by three wireless APs. (b) Frequency offsets' distribution of the signal frequency offsets of the legitimate devices. (c) Frequency offsets of WiFi frames sent by the attack device. (d) Frequency offsets' distribution of the fake devices.

Each element in the matrix stands for the value of Wasserstein metric for a pair of devices. Figure 8(b) is the Wasserstein distance matrix of signals from a set of Wi-Fi APs under attack, wherein device nos. 9 ~ 16 are the fake APs generated by an attack device.

4.3. Distribution Clustering. Given the obtained Wasserstein metric of the frequency offset distribution between each device, the questions that need to be addressed are (1) determine if there are signals in the environment with similar frequency offset distributions and (2) distinguish the fake APs from legitimate APs. As we have no a priori knowledge of legitimate APs, it leads us to a typical distribution clustering problem.

In this paper, we use the density-based spatial clustering of applications with noise (DBSCAN) method to cluster the frequency offset distribution of signals.

DBSCAN was proposed by Ester et al. in 1996, which is one of the most commonly used clustering methods [20]. The DBSCAN method is suited to solve this clustering problem for it does not need to specify the number of clusters and is robust to a large number of out-of-cluster noise points, which are the legitimate APs in this paper.

We utilize (5) as the distance function for the DBSCAN algorithm to cluster the signal samples from all devices collected. The APs in the same cluster mean they are similar in the meaning of Wasserstein metric. Ultimately, the wireless devices that are successfully clustered are assigned as fake APs, while the remaining wireless APs that are not in the clustering are legitimate devices. Algorithm 1 is shown as follows.

An example of the DBSCAN clustering result is illustrated in Figure 9. As shown in the figure, all the fake APs are in the cluster, and the legitimate APs are marked as noise points.

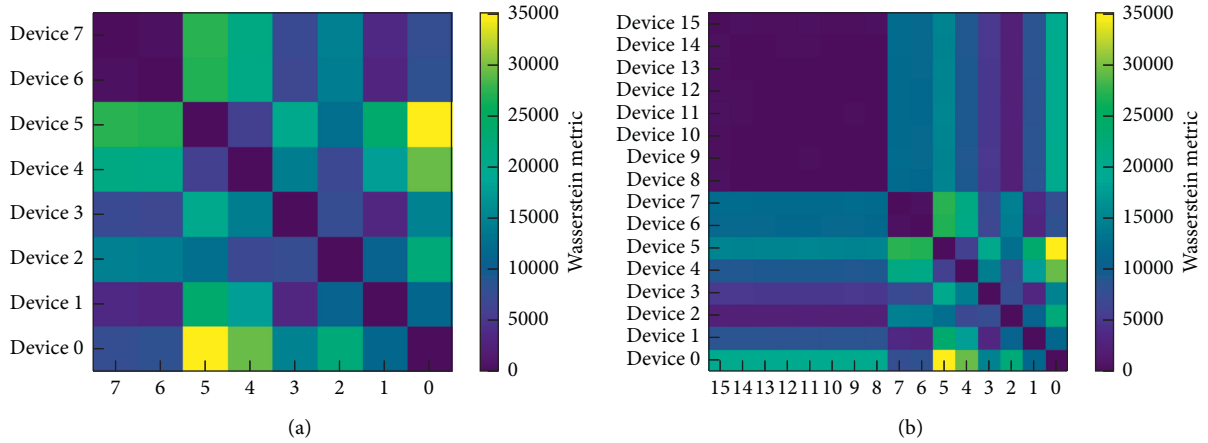


FIGURE 8: Matrix of the Wasserstein metric. (a) Wasserstein metric matrix of signals from a set of legitimate WiFi APs. (b) Wasserstein metric matrix of signals from both legitimate and fake WiFi APs.

4.4. Overall of the Attack Detection Method. Finally, Figure 10 shows the overall view of the proposed attack detection scheme integrated with the Wasserstein metric and DBSCAN algorithm. Our method first collects the beacon signals of WiFi APs in the environment and estimates their frequency offset, next calculates the Wasserstein metric between each pair of the frequency offset distributions, and then uses the DBSCAN algorithm to find similar devices in the meaning of Wasserstein metric. If there are 2 or more APs in the same cluster, it means there are spoofing attacks in the environment. Finally, we mark the APs in the cluster as fake APs and add them to the blacklist. Those detected fake APs can be blocked in the WPS which ensures that users can acquire the correct positioning result from the positioning server.

5. Experiments and Results

This section describes the experimental testing of the proposed spoofing attack detection method. We build a simulation scenario of spoofing attacks by commercial WiFi devices, implement a prototype attack detection system based on the USRP device, and evaluate the performance of our method.

5.1. Experimental Setup. The hardware devices used in our experimental system are shown in Figure 11. The left side of the photo includes three commercial outdoor wireless APs, a home wireless router, and several RaspberryPi 3B board computers, which are used to build the wireless environment of WPS scenarios. Another RaspberryPi 3B board computer is installed with the MDK3 wireless attack tool to simulate the WPS spoofing attack by hackers. To the right of the image are the USRP B210 device and a laptop used to implement our attack detection method.

Figure 12 shows the architecture of the location spoofing detection experimental system. Firstly, the WiFi signal is collected by the USRP B210 device. Then, we use the GNU Radio-based software receiver [21] to demodulate the WiFi signal. The spoofing attack detection method proposed is parallelly

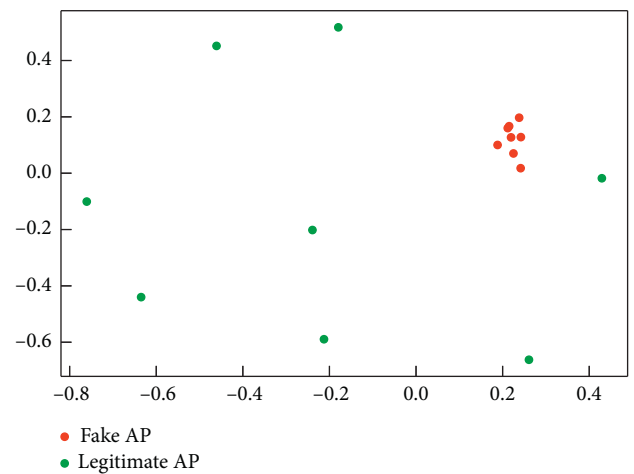


FIGURE 9: The clustering result presented by multidimensional scaling visualization.

integrated with the software receiver. It is fed with frequency offset records and outputs the MAC addresses of fake APs.

In order to test the performance of the proposed method, we use the system to collect signal records from multiple APs and use different combinations of devices to construct 400 test scenarios, 200 of which simulate normal conditions and 200 of which are with the presence of attackers. Each scenario contains at least 10 APs. For each AP, 400 samples of WiFi signals were collected. In the evaluation stage, the signals are randomly sampled from the datasets to ensure the stability of test results.

5.2. Evaluation Metrics. In each round of tests, we first get the confusion matrix of the test dataset as shown in Table 2, where TP is true positive, which refers to the number of correctly detected fake WiFi APs. The FN means false negative, and TN/FP stands for the counts of true negative and false positive.

With the definition above, to evaluate the performance of our method, we adopt standard statistical metrics such as accuracy, precision, and recall, and these metrics are defined as follows:

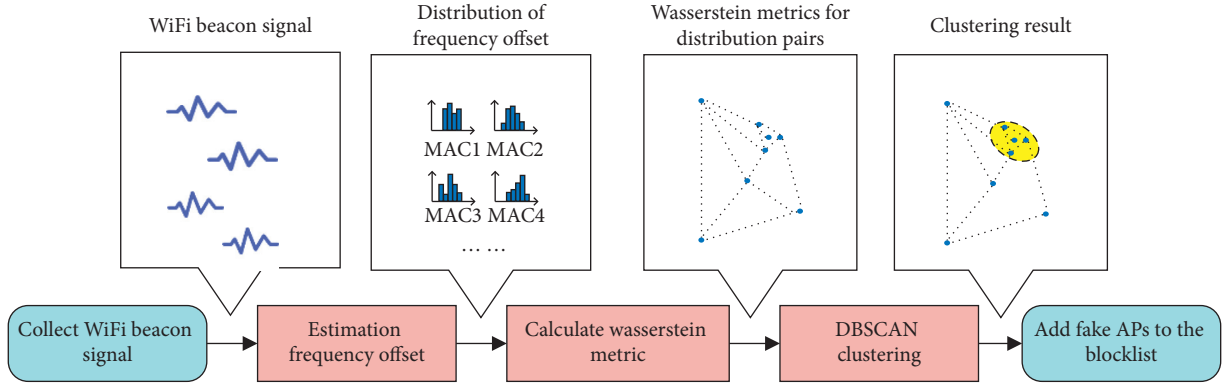


FIGURE 10: Flow diagram of the proposed detection method.

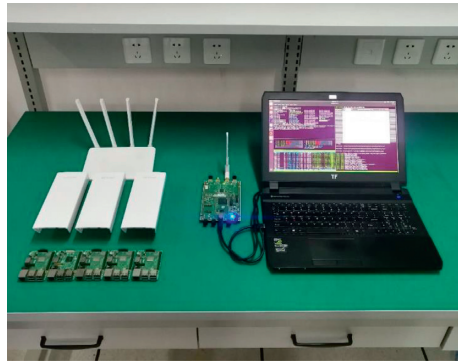


FIGURE 11: Hardware devices in the experimental system.

$$\begin{aligned}
 \text{Accuracy} &= \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}}, \\
 \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\
 \text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}.
 \end{aligned} \tag{6}$$

5.3. Performance of Identification

5.3.1. Impact of Various Parameters. The first experiment aimed to find the optimized parameters of the proposed algorithm. Two parameters, min_pts and ϵ , directly affect the detection performance of the proposed algorithm. We change the parameters to test the impact of those parameters on the performance metrics in the 400 scenarios. The test results are shown in Figure 13, where the best accuracy (99%) is achieved when $\text{min_pts} = 3$ and $\epsilon = 320$; meanwhile, precision and recall are also high. This set of optimized parameters will be used in the subsequent tests.

5.3.2. Impact of Sample Numbers. Figure 14 gives the impact of a varying number of samples on the detection performance. The more the signal samples collected the more improved is the identification performance. The proposed

method performs poorly when the number of samples is less than 25. As the number of samples increases, the performance shows an upward trend. When using 40 samples, the accuracy, precision, and recall surpass 95%, and when using more than 80 samples, the method achieves 99% accuracy and ensures high recall and precision. In general, WiFi hotspots broadcast beacon frames at a rate of 10 Hz, so the algorithm can collect 100 samples of data from each AP hotspot within 10 seconds, which ensures the effective detection of fake APs.

5.4. Real World Experiment. The purpose of this experiment was to test the efficacy of our method combined with the traditional WPS on countering location spoofing attacks in the real-world environment.

Figure 15(a) shows the physical location (Zhengzhou East Railway Station, Henan, China) of our test area which is filled with public WiFi hotspots. First, we launched the location spoofing attack by setting up fake wireless APs with MAC addresses that are located far (Hangzhou Tower, Zhejiang, China) from the actual physical location. Then, we collect WiFi signals of the test area during the spoofing attack.

At present, the 5G standard does not prescribe the detailed implementation of WPS in LCS, and we use the Google Maps geolocation service as an alternative solution to acquiring the positioning result. Table 3 shows the WiFi AP list collected

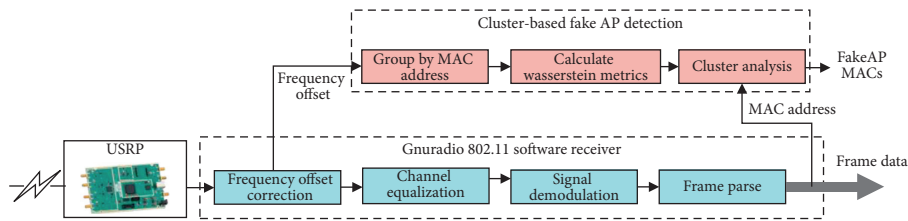


FIGURE 12: The architecture of the location spoofing attack detection system.

- (1) Collect all WiFi APs in the environment as the points in the dataset
- (2) Estimate the Wasserstein metric between each pair of points in the dataset
- (3) Find the points in the ϵ -neighborhood of every point, and identify the core points with more than min_pts neighbors
- (4) Find the connected components of core points on the neighbor graph, ignoring all noncore points
- (5) Assign each noncore point to a nearby cluster if the cluster is an ϵ - neighborhood, otherwise assign it to noise
- (6) Mark the remaining noise points in the dataset as legitimate APs and the ones that have been clustered as fake APs

ALGORITHM 1: Density-based spatial clustering of applications with noise.

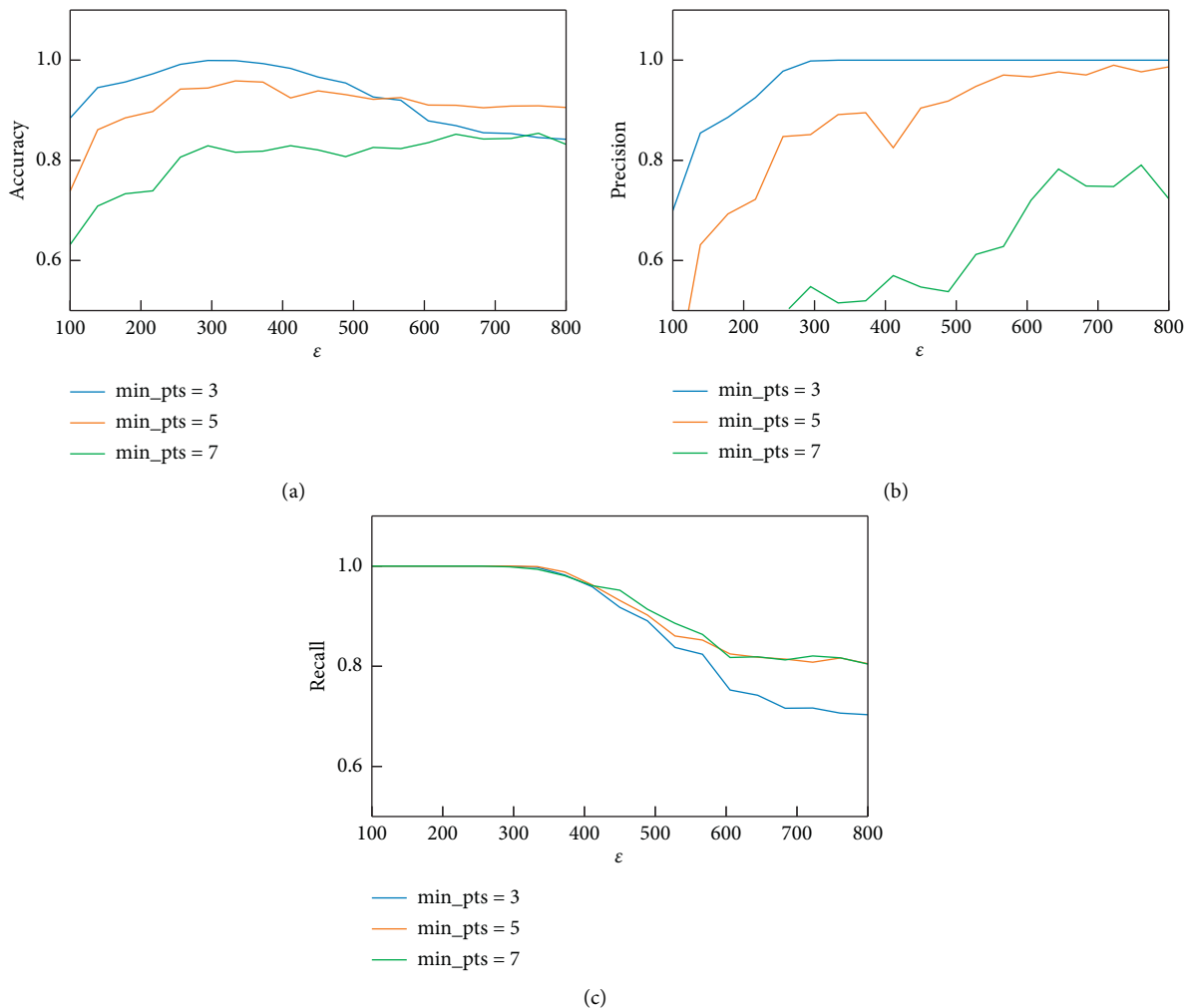


FIGURE 13: Detection performance with different ϵ and min_pts .

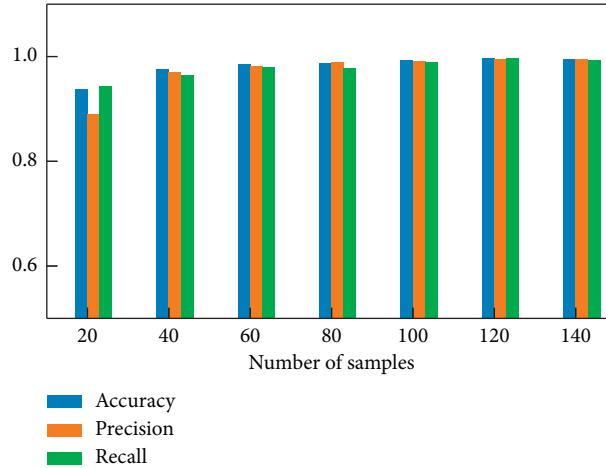


FIGURE 14: Impact of a varying number of samples on the performance.

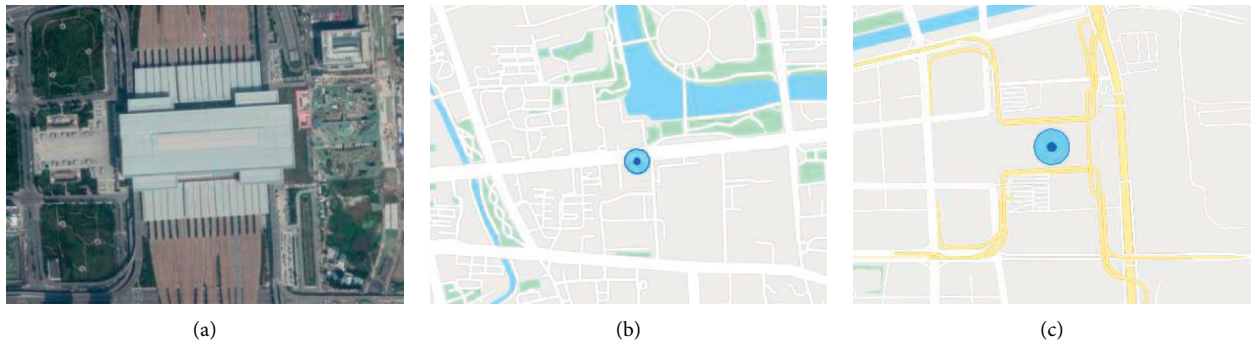


FIGURE 15: Real-world test positioning results. (a) Physical location of the test area. (b) Positioning result manipulated by spoofing attack. (c) Positioning result secured by the proposed method.

during the spoofing attack. The last column in the table shows the detection results of our method.

When the positioning service was fed with the full list of WiFi APs, the positioning result acquired was manipulated to Hangzhou Tower, as shown in Figure 15(b), which is the fake location. Then, we applied our spoofing attack detection system and excluded all the APs marked as “blocked” before actually sending the positioning request to the service. Figure 15(c) shows that the correct positioning result was acquired from the positioning service after filtering out the fake APs by our method.

6. Discussion

Our studies serve as a proof-of-concept that develops an unsupervised detector to recognize and counter the location spoofing attacks against WPS, which can be seen as a part of the generalized WiFi spoofing attack problem. Table 4 shows a comparison between our proposed method and other related works.

As shown in Table 4, our detection method provides relatively high performance and competitive features compared with existing methods. First, unlike the CSI and RSSI-based

methods, our attack detection method is independent of the communication channel parameters, so it will keep stable performance despite the change in the location of devices; secondly, this is a user-side detection method, which does not require additional communication overhead and collaborative devices. Third, our method does not need to record and learn the feature of legitimate devices in advance, which is the biggest difference compared with most other RFF-based methods.

The main innovation of our work is leveraging the distribution of a device’s signal feature instead of the one-shot feature to identify the malicious devices. Combined with the Wasserstein metric and DBSCAN algorithm, this approach can be seen as dimension-raising processing and maximize the information extracted from one signal feature. But it also means we need to make a trade-off between the performance and the delay of collecting samples. Another main limitation of the proposed method is that it still needs access to the raw signal to extract the signal frequency offset. Although all the hardware receivers have the frequency offset estimation stage to synchronize the signal, few devices can provide the interface to output this parameter, which makes it difficult to be applied on most COTS devices. Moreover, the frequency offset used in our method is affected by the Doppler effect. It could bring in

TABLE 2: Confusion matrix of the detection result.

Test predict	Fake AP	Legitimate AP
Fake AP	TP	FP
Legitimate AP	FN	TN

TABLE 3: WiFi APs' information collected in the real-world test.

Category	MAC address *	Status
Legitimate AP	00: *: *: *: *: *: *: 55:8C	Good
	00: *: *: *: *: *: *: F5:62	Good
	00: *: *: *: *: *: *: 58:35	Good
	00: *: *: *: *: *: *: 19:C4	Good
	00: *: *: *: *: *: *: 39:33	Good
	06: *: *: *: *: *: *: DA:22	Good
Fake AP	78: *: *: *: *: *: *: 7E:29	Blocked
	78: *: *: *: *: *: *: 7E:2A	Blocked
	78: *: *: *: *: *: *: 7E:2C	Blocked
	78: *: *: *: *: *: *: 7E:2E	Blocked
	80: *: *: *: *: *: *: FD:50	Blocked
	78: *: *: *: *: *: *: 7E:28	Blocked
	70: *: *: *: *: *: *: 4F:A1	Blocked
	C4: *: *: *: *: *: *: C8:6D	Blocked
E6: *: *: *: *: *: *: C0:86	Blocked	

*The MAC addresses are masked for privacy concerns.

TABLE 4: Comparison of different methods on WiFi spoofing attack detection.

Method	Implementation requirements			Detection rate (%)
	Collaborative devices	Priori knowledge	Access to raw signal	
Restuccia et al. [6]	Cross check with other users	✓	—	Not given
Ye et al. [7]	Cross check with positioning servers	✓	—	Not given
Faria and Cheriton [8]	RSSI with min-max match	✓	—	99.1
Yang et al. [9]	RSSI with SVM	✓	—	98
Xiao et al. [22]	CSI with generalized likelihood ratio test	✓	—	90
Liu et al. [23]	CSI with SVM	✓	—	95
Wang et al. [10]	CSI with multiple antenna positioning	—	—	98.5
Suski et al. [11]	Preamble spectrum feature	—	✓	80
Brik et al. [13]	Multiple signal features	—	✓	99
Jiang et al. [14]	CSI with deep learning	—	—	95
Our method	Frequency offset with Wasserstein metric	—	✓	99

potential performance loss as the frequency offset changes when the user is moving fast. Our future work is to utilize more signal features in the method to achieve robust performance. Another potential improvement is to introduce the innovative authentication scheme [24–26], to provide extended security protection against nontraditional threats.

7. Conclusion

In this paper, an attack detection method based on the Wasserstein metric and DBSCAN is proposed to address the problem of location spoofing attacks in WPS. The algorithm first extracts the frequency offset of the WiFi signals; then, it calculates the Wasserstein distance between each pair of APs in the environment; finally, it uses the DBSCAN algorithm to detect

the fake wireless APs that share similar frequency offset distributions. We test our method in both controlled experiments and the real-world environment. The experiments show that the proposed method achieves a high accuracy of 99% for detecting fake APs. The real-world test shows that our approach can effectively secure the positioning result against the location spoofing attack.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] “3GPP technical specification group services and system Aspects Release 16 description,” September 2020, https://www.3gpp.org/ftp/Specs/archive/21_series/21.916/21916-060.zip.
- [2] S. Tongleamnak and M. Nagai, “Simulation of GNSS availability in urban environments using a panoramic image dataset,” *International Journal of Navigation and Observation*, vol. 2017, Article ID 8047158, 12 pages, 2017.
- [3] N. O. Tippenhauer, K. B. Rasmussen, C. Pöpper, and S. Čapkun, “Attacks on public WLAN-based positioning systems,” in *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services*, pp. 29–40, ACM, Kraków Poland, June 2009.
- [4] C. Matte, J. P. Achara, and M. Cunche, “Device-to-identity linking attack using targeted wi-fi geolocation spoofing,” in *Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks*, pp. 20:1–20:6, ACM, New York, NY, USA, June 2015.
- [5] M. Cunche, “I know your MAC address: targeted tracking of individual using Wi-Fi,” *Journal of Computer Virology and Hacking Techniques*, vol. 10, no. 4, pp. 219–227, 2014.
- [6] F. Restuccia, A. Saracino, S. K. Das, and F. L. V. S. Martinelli, “A WiFi-based system to tackle Location Spoofing in location-based services,” in *Proceedings of the 17th International Symposium on A World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pp. 1–4, IEEE, Coimbra, Portugal, June 2016.
- [7] A. Ye, Q. Li, Q. Zhang, and B. Cheng, “Detection of spoofing attacks in WLAN-based positioning systems using WiFi hotspot tags,” *IEEE Access*, vol. 8, pp. 39768–39780, 2020.
- [8] D. B. Faria and D. R. Cheriton, “Detecting identity-based attacks in wireless networks using signalprints,” in *Proceedings of the 5th ACM Workshop on Wireless Security*, pp. 43–52, ACM, Los Angeles, CA, USA, September 2006.
- [9] J. Yang, Y. Chen, W. Trappe, and J. Cheng, “Detection and localization of multiple spoofing attackers in wireless networks,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, pp. 44–58, 2013.
- [10] C. Wang, L. Zhu, L. Gong et al., “Accurate sybil attack detection based on fine-grained physical channel information,” *Sensors*, vol. 18, no. 3, p. 878, 2018.
- [11] W. C. Suski II, M. A. Temple, M. J. Mendenhall, and R. F. Mills, “Using spectral fingerprints to improve wireless network security,” in *Proceedings of the Global Telecommunications Conference (GLOBECOM)*, pp. 1–5, IEEE, Orleans, LA, USA, November–December 2008.
- [12] N. T. Nguyen, G. Zheng, Z. Han, and R. Zheng, “Device fingerprinting to enhance wireless security using nonparametric Bayesian method,” in *Proceedings of the INFOCOM*, pp. 1404–1412, IEEE, Shanghai, China, April 2011.
- [13] V. Brik, S. Banerjee, M. Gruteser, and S. Oh, “Wireless device identification with radiometric signatures,” in *Proceedings of the 14th ACM International Conference on Mobile Computing and Networking*, pp. 116–127, ACM, San Francisco, CA, USA, September 2008.
- [14] P. Jiang, H. Wu, C. Wang, and C. Xin, “Virtual MAC spoofing detection through deep learning,” in *Proceedings of the IEEE International Conference on Communications (ICC)*, pp. 1–6, IEEE, Kansas City, MO, USA, May 2018.
- [15] J. Yu, A. Hu, G. Li, and L. Peng, “A robust RF fingerprinting approach using multisampling convolutional neural network,” *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6786–6799, 2019.
- [16] L. Peng, J. Zhang, M. Liu, and A. Hu, “Deep learning based RF fingerprint identification using differential constellation trace figure,” *IEEE Transactions on Vehicular Technology*, vol. 69, pp. 1091–1095, 2019.
- [17] WIGLE, September 2020, <https://www.wigle.net>.
- [18] E. Sourour, H. El-Ghoroury, and D. McNeill, “Frequency offset estimation and correction in the IEEE 802.11a WLAN,” in *Proceedings of the IEEE 60th Vehicular Technology Conference*, vol. 7, pp. 4923–4927, IEEE, 2004.
- [19] C. Villani, *Optimal Transport Old and New*, Springer Science & Business Media, Berlin, Germany, 2008.
- [20] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining*, pp. 226–231, AAAI Press, Portland, OR, USA, August 1996.
- [21] B. Bloessl, M. Segata, C. Sommer, and F. Dressler, “An IEEE 802.11 a/g/p OFDM Receiver for GNU Radio,” in *Proceedings of the Second Workshop on Software Radio Implementation Forum*, pp. 9–16, ACM, Hong Kong China, August 2013.
- [22] L. Xiao, L. Greenstein, N. Mandayam, and W. Trappe, “Channel-based spoofing detection in frequency-selective Rayleigh channels,” *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 5948–5956, 2009.
- [23] H. Liu, Y. Wang, J. Liu, J. Yang, and Y. Chen, “Practical user authentication leveraging channel state information (CSI),” in *Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security*, pp. 389–400, ACM, Kyoto Japan, June 2014.
- [24] D. Wang and P. Wang, “Two birds with one stone: two-factor authentication with security beyond conventional bound,” *IEEE Transactions on Dependable and Secure Computing*, vol. 15, pp. 708–722, 2016.
- [25] D. Wang, W. Li, and P. Wang, “Measuring two-factor authentication schemes for real-time data access in industrial wireless sensor networks,” *IEEE Transactions on Industrial Informatics*, vol. 14, no. 9, pp. 4081–4092, 2018.
- [26] S. Qiu, D. Wang, G. Xu, and K. Saru, “Practical and provably secure three-factor authentication protocol based on extended chaotic-maps for mobile lightweight devices,” *IEEE Transactions on Dependable and Secure Computing*, 2020.