Real-time Identification of Rogue WiFi Connections Using Environment-Independent Physical Features

Pengfei Liu, Panlong Yang, Wen-Zhan Song, Yubo Yan, Xiang-Yang Li
University of Science and Technology of China, Hefei, China
liupf17@mail.ustc.edu.cn, plyang, yuboyan@ustc.edu.cn, wsong@uga.edu, xiangyang.li@gmail.com

***Abstract*—WiFi has become a pervasive communication medium in connecting various devices of WLAN and IoT. However, WiFi connections are vulnerable to the impersonation attack from rogue access points (AP) or devices, whose SSID and/or MAC/IP address are identical to the legitimate devices. This kind of attack is difficult to countermeasure with traditional network security mechanisms. In this paper, we present a novel security mechanism to detect and identify rogue WiFi devices or AP using *environment-independent* characteristics extracted from channel state information (CSI), and refuse their connections. We find that nonlinear phase errors of different subcarriers change with WiFi network interface cards (NIC), due to the I/Q imbalance and imperfect oscillator of each WiFi NIC. Validated by our experiments, this phase feature across subcarriers is consistent and invariant to location and external environment, and can be extracted to build an essential signature of the NIC itself. Such signature of the transmitter can be calculated in real-time by the receiver and cannot be forged by rogue devices. Extensive experiments with dozens of WiFi devices demonstrate that the proposed mechanism can reliably detect the rogue WiFi connections and prevent impersonation in various scenarios. The speed of identification is $8\times$ faster than that of the state-of-the-art solution. Moreover, the accuracy of rogue connection detection is up to 96% and false alarm rate is shown below 2%.**

I. Introduction

A. Backgrounds and Motivations

WiFi has become a pervasive communication medium in connecting various wireless devices in Local Area Networks (LAN) and Internet of Things (IoT). Specifically in IoT, the number of connected devices will be measured in billions and often use WiFi for connections due to the ease of deployment [1]. Unfortunately, many security threats have been discovered with WiFi connections. For example, the authentication protocols, such as WEP, WPA and WPA2, have been proved insecure and broken [2], [3]. Furthermore, the side channel attacks can get the phone payment password from CSI with a rogue AP [4].

Another serious threat is freeloading. In the network of government, military and financial institutions, illegal users can gain access to the Intranet with the spoofed MAC address of authenticated device. The hacker can not only steal data from the Intranet, but also form Mirai botnet among hijacked devices to attack the infrastructure [5]. Despite of the protection mechanism via password authentication, replay attacks can bypass this protection through the same legal management frame [6].

All in all, in a WiFi network, the rogue AP and client devices may forge password, SSID and/or MAC/IP address of legitimate devices, and play various impersonation and replay attacks, as illustrated in Fig. 1. A real-time detection, identification and prevention mechanism is needed to countermeasure rogue WiFi connections. However, those problems are difficult to solve with traditional network security mechanism, particularly because the attack may happen in and below data link layer.

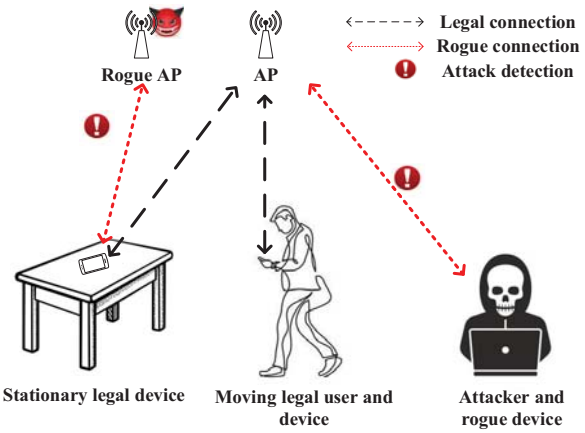


Fig. 1: Security threats from rogue AP and client device.

Thus, in this paper, we design a device fingerprint based on physical layer information and use it to identify and prevent rogue WiFi connection. Specifically, we explore hardware fingerprint-based security and countermeasure mechanism. Compared with calculating a hardware fingerprint by the receiver, transmitting it over the air brings more security issues, thus those physical unclonable function (PUF) based hardware fingerprint methods (such as PUF of microprocessor and RAM) do not apply in this context [7].

Indeed, the fingerprint shall be consistent and invariant to environment conditions. Different WiFi devices shall exhibit different signatures with very high probability, so that the chance for same signature between two devices is very small. It shall be also impossible for a hacker to generate the same signature at the receiver. Invariance means that the movement of people or objects shall not change such a signature.

B. Challenges and Contributions

The needed fingerprint method shall be based on communication/WiFi connection, and related to hardware parameters. According to the identification mechanism PARADIS pro-

posed by Brik *et al.* [8], the radio features such as frequency error, magnitude error, sync correlation and I/Q imbalance, are intrinsic with the NIC for fingerprinting. Although additional equipments were required in PARADIS and these equipments were expensive, it has inspired amounts of researchers to design identification mechanisms based on radio features such as received signal strength indicator (RSSI) and CSI. Unfortunately, the existing fingerprint methods need many samples and take long time to collect and calculate. Besides, RSSI and CSI are usually used for localization and human activity tracking due to their sensitivity to location and environment. Thus, these mechanisms only apply to static and stable environment. In this paper, we are committed to resolving the contradiction between identification speed and changing application scenarios.

In order to obtain a fingerprint that is invariant to time, location and environment, we discovered that it is possible to extract new invariant feature from the constantly changing CSI. Nonlinear phase errors of different subcarriers change with WiFi NICs, due to the I/Q imbalance and imperfect oscillator at each WiFi NIC. Benefit from the convenience of extracting CSI from any transmitters such as APs, laptops, smart watches and mobile phones without additional equipments, this phase feature across the subcarriers can be extracted to build a device fingerprint. As mentioned above, CSI reflects fine-grained characteristic of channel. It is necessary to eliminate multipath effects without restricting the movement of IoT devices. Moreover, previous identification mechanisms based on CSI require large amounts of samples, and their sampling processes consume a lot of time. How to obtain better accuracy with less samples is a problem that have to be solved.

In summary, we face the following challenges to extract the fingerprint:

- **Invariant fingerprints *vs.* changing CSI:** CSI is sensitive to the change of location and environment. The fingerprint should be independent of time, location and environment. It is difficult to extract invariant fingerprints from changing CSI.
- **Long sampling time *vs.* high identification speed:** There is a lot of noise in wireless channels. The longer the sampling time is, the less obvious the noise is and the higher the accuracy is. However, too long sampling time will make the system more vulnerable to hackers. How to balance the accuracy and speed is an inevitable problem.

To tackle these challenges, we designed a filter to seek out the regular phases from the CSIs that change randomly. Then, we combined the phase errors caused by oscillator and I/Q imbalance after studying the components of NIC. Based on normalization, we achieved an invariant fingerprint to location and environment.

All in all, we make the following contributions:

- We propose a novel phase-based device identification mechanism, which extracts hardware fingerprints from CSI. These fingerprints are independent of locations and robust to environments.
- We design a convenient approach for deployment in commercial-off-the-shelf (COTS) wireless device without any additional devices or system modifications.
- Extensive experiments with dozens of devices have been performed in various scenarios. The results show that our approach can achieve 96% rogue detection accuracy and is $8\times$ faster than the state-of-the-art solution.

The rest of this paper is organized as follows: Section II presents the preliminaries and observations of CSI-based fingerprinting. In Section III, we describe the algorithm and system design of the proposed mechanism in details. Implementation and evaluation are presented in Section IV and Section V. Section VI discusses the related works. We finally conclude our work in Section VII.

II. Preliminaries and Observations

In this section, we present the basics of CSI and phase error, and describe observations from some preliminary tests.

A. CSI and Phase Error

CSI: The NIC continuously captures CSI, which indicates signal strength and phase information of each OFDM symbol. It includes the combined effect of scattering, fading, multipath and power decay on the signal propagation path from transmitter to receiver. Recently, WiFi technologies commonly use Multi-input Multi-output (MIMO), thus channel between each transmitter-receiver (Tx-Rx) antenna pair consists of multiple subcarriers [9]. The CSI of Tx-Rx pair i, expressed through the channel matrix $\mathbf{H}_i$, is trained by selected subcarriers. Let $\mathbf{X}_i$ be the transmitted signal, the received signal $\mathbf{Y}_i$ can be expressed as

$$\mathbf{Y}_i = \mathbf{H}_i\mathbf{X}_i + \mathcal{N}_i, \tag{1}$$

where $\mathcal{N}_i$ is the noise matrix. By modifying the driver of COTS NICs such as Intel5300 and AR9580 [10], CSI can be easily obtained. A group of sampled *channel frequency response* (CFR) is used to express this channel state [11]:

$$\mathbf{H}_i = \sum_{k\in\boldsymbol{K}} \|h_k\| \cdot e^{-j\cdot\angle h_k} \quad , \tag{2}$$

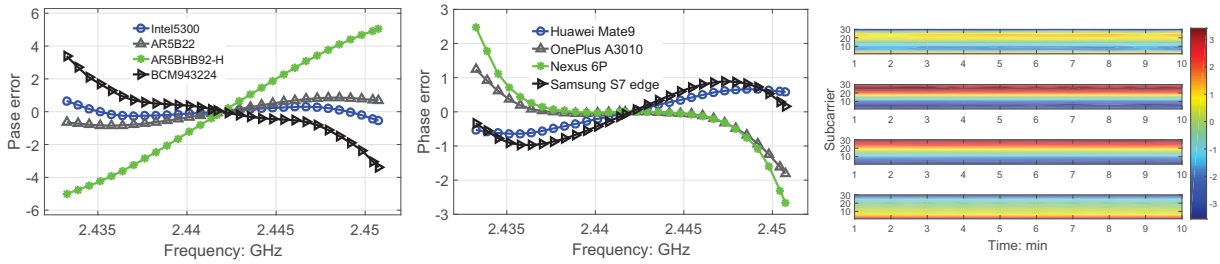
where $\|h_k\|$ and $\angle h_k$ denote the amplitude and phase of subcarrier k. $\mathbf{K}$ contains the subcarrier index. There are total of 30 subcarriers are measured, *i.e.*, $\mathbf{K} = [-28,-26,...,-2,-1,1,3,...,27,28]$, for Intel 5300 NIC, working at 2.4 GHz ISM band with 20 MHz bandwidth [12]. The raw CFR estimated in Intel5300 is recorded as the I/Q signal. Each column of $\mathbf{H}$ is a subcarrier's CFR of a pair of Tx-Rx antenna, which is given by

$$\mathbf{H}_i = \left[\mathbf{H}_i^{(-28)}\mathbf{H}_i^{(-26)}\cdots\mathbf{H}_i^{(k)}\right]^T k \in \mathbf{K}, \tag{3}$$

$$\mathbf{H}_i^{(k)} = I_i^{(k)} + jQ_i^{(k)}. \tag{4}$$

The phase $\angle h_k$ can be calculated from the I/Q component:

$$\mathbf{\Phi} = \tan^{-1}(\frac{Q}{I}). \tag{5}$$



(a) Phase errors of four different NICs. (b) Phase errors of four different mobile phones. (c) Phase errors measured at different time.

Fig. 2: Phase errors of NICs and mobile phones.

Phase Error: In wireless communication, the measured phases at the receiver are different from that at the transmitter due to both hardware design errors and the signal transmission environment (*e.g.*, location, distance and obstacle). According to [12], for a particular pair of transmitter and receiver, the phases of subcarriers $\mathbf{\Phi}$ measured at the receiver can be expressed as

$$\mathbf{\Phi} = \boldsymbol{\varphi} + \boldsymbol{\omega} + \boldsymbol{\theta} + \boldsymbol{\psi} + \boldsymbol{\epsilon}, \tag{6}$$

where $\boldsymbol{\varphi}$ denotes the phases of signal at the transmitter, $\boldsymbol{\omega}$, $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ represent the phase offset due to frame detection delay (FDD), sampling frequency offset (SFO) and time of flight (TOF) respectively. The last element $\boldsymbol{\epsilon}$ denotes error caused by I/Q imbalance. The effect of FDD and SFO of the same frame are related to subcarrier sequence $\mathbf{K}$, which can be expressed by the following equations:

$$\boldsymbol{\omega} = 2\pi\alpha \cdot \mathbf{K}, \boldsymbol{\theta} = 2\pi\beta\mathbf{K}, \tag{7}$$

where α and β are constant depending on FDD and SFO. The TOF $\boldsymbol{\psi}$ is related to subcarrier frequency:

$$\boldsymbol{\psi} = 2\pi t_f \mathbf{F}, \tag{8}$$

where t_f is time of fly and affected by location. $\mathbf{F}$ is the set of subcarrier frequencies. As $\mathbf{F}$ can be denoted by center frequency f_c and subcarrier sequence $\mathbf{K}$, *i.e.*, $\mathbf{F} = f_c \cdot \vec{I} + w\mathbf{K}$ the effect caused by ToF $\boldsymbol{\psi}$ can be rewritten as

$$\boldsymbol{\psi} = 2\pi f_c t_f \vec{I} + 2\pi t_f w\mathbf{K} = \mathbf{Z} + 2\pi t_f w\mathbf{K}, \tag{9}$$

where w is the frequency difference between two continuous subcarriers and equal to 312.5kHz according to IEEE 802.11g standard. $2\pi f_c t_f \vec{I}$ is independent of subcarrier index $\mathbf{K}$. In the same frame, t_f is a constant. Thus, we can use $\mathbf{Z}$ to replace the first part. According to Eq. (7)and (9), the received signal can be re-described as

$$\begin{aligned} \mathbf{\Phi} &= \boldsymbol{\varphi} + 2\pi(\alpha + \beta + t_f w)\mathbf{K} + \mathbf{Z} + \boldsymbol{\epsilon} \\ &= \boldsymbol{\varphi} + 2\pi\lambda \cdot \mathbf{K} + \mathbf{Z} + \boldsymbol{\epsilon}. \end{aligned} \tag{10}$$

For a specific frame, λ is also a constant and is the sum of α, β and $t_f w$. So, the phase error $\boldsymbol{\epsilon}$ caused by I/Q imbalance can be estimated by the equation below:

$$\boldsymbol{\epsilon} = \mathbf{\Phi} - \boldsymbol{\varphi} - \mathbf{Z} - 2\pi\lambda\mathbf{K}. \tag{11}$$

B. Preliminary Tests and Observations

We performed preliminary tests to validate that phase error can be used as the needed fingerprint. An industrial personal computer (IPC) with an Intel5300 NIC is set as a WiFi AP and a WiFi device is connected to it. We first obtain the original CSI without multipath effect, by using a RF cable of 50cm and an attenuator of 30dB to connect the NICs of transmitter and receiver directly. Four NICs are installed on one device in turn. Fig. 2a shows the phase errors of different NICs. In this test, we set λ in Eq. (11) to $200ns$. To confirm that the phase errors are independent of time, the AP samples the CSI of NICs connected through a RF cable and measures phase errors for 10 minutes. Fig. 2c shows the change of phase error across these tested NICs. It is obvious that phase errors are invariant to time. Finally, in order to verify that such feature also exists on mobile devices, we tested four smartphones when AP and phones are all static. As shown in Fig. 2b, phase errors of these four devices are distinguishable. We further measured the CSIs when the smartphone was moving. Phases of four consecutive samples are shown in Fig. 5a. There are random changes of phase in dynamic environment.

From above tests, we have the following preliminary observations:

- Phase error is a feature related to hardware difference which is common to smartdevices;
- If the transmission channel is stable, phase errors will be invariant to time.

These observations give us confidence to use this feature to identify WiFi devices and AP. However, smart devices will not always stay static in realistic application scenarios. Multipath owing to environment change will affect the transmission channel and invariance of phase. If we want to use this feature as fingerprint, how to remove the impact of environment is inevitable. Next section, we will describe how to obtain steady phase error from changing CSIs in the complex and dynamic environment.

III. System Design

The proposed system has two most important functions. Firstly, a fingerprint library of legitimate device shall be established after authentication. Secondly, when a rogue device or AP tries to connect with the other one, our system must detect and reject it accurately and quickly. To establish the fingerprint

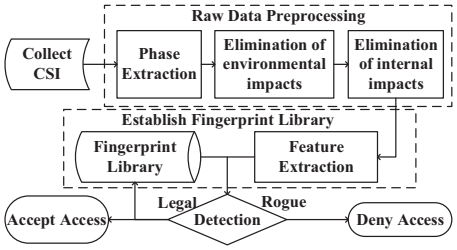


Fig. 3: System structure.

library, we should estimate the phase error. However, raw CSI in the dynamic environment is fluctuating. We plot phases of continuous CSIs when the transmitter is moving. As shown in Fig. 5a, the change of phase between adjacent subcarriers is significant.

From the above analysis, our system contains three main parts: raw data preprocessing, establishment of fingerprint library, and validation of legitimacy. We assume that there is a set of legitimate devices for us to initialize the fingerprint library.

A. Raw Data Preprocessing

Phase Extraction: As described in Section II, the raw CSI is expressed by CFR like Eq. (2) and phases can be measured from it according to Eq. (5). However, the calculated phases are distributed between $-\pi$ and π. It results in ambiguity of the relationship of subcarriers. In order to analyze the relationship between subcarriers, we should recover the real phase. From Fig. 4, we can see that raw phases distributed at bandwidth are mapped into $[-\pi, \pi]$ and shaped like a saw. So we just need to reduce the phase by $q \cdot 2\pi$ when the value of phase increases sharply. Here, it is impossible and unnecessary to know the value of q. We have to splice abnormal parts for coherent phases. We define ${}_{\triangle}\mathbf{\Phi} = \mathbf{C} \cdot \mathbf{\Phi}$ as the phase differences between adjacent subcarriers, where

$$\mathbf{C} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}_{29\times 30} \tag{12}$$

Due to the mapping phase into $[-\pi, \pi]$, there are some ${}_{\triangle}\mathbf{\Phi}_k > \pi$. By reducing 2π, we get the real phase difference of adjacent subcarriers. Eq. (13) shows the process of ${}_{\triangle}\mathbf{\Phi}_k$.

$${}_{\triangle}\mathbf{\Phi}_k = {}_{\triangle}\ \mathbf{\Phi}_k - 2\pi, \quad if \quad {}_{\triangle}\mathbf{\Phi}_k > \pi \tag{13}$$

After these operations, we update all phases by taking the phase of first subcarrier as reference. The blue line in Fig. 4 denote the unwrapped phases, it indicates that phases across subcarriers are approximately linear. However, according to the discovery of Zhu *et al.* [12], there are non-linear phase errors. It will be introduced in the next subsection. In the rest of this subsection, we describe how to eliminate the influence of the environment on the phase.

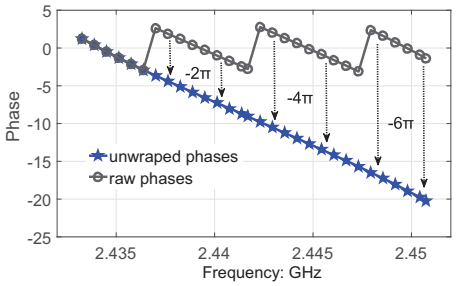


Fig. 4: Phase unwrapping.

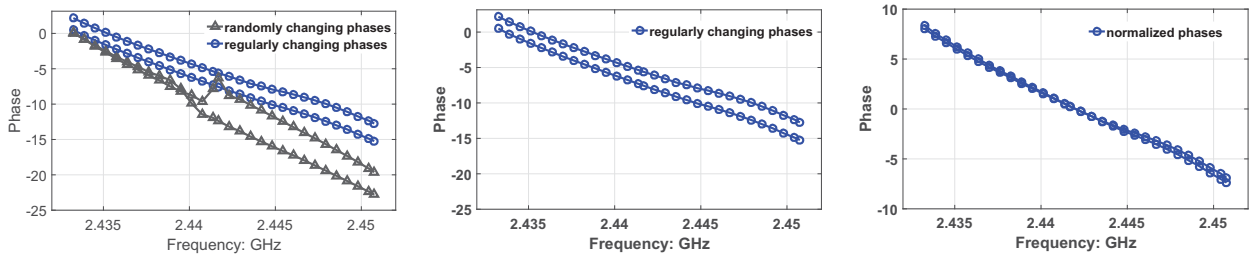
Elimination of environmental impacts: In the preliminary tests of our design in Section II, we have introduced the impact of environment to phases of subcarriers. To further study the impact of environment on the phases of subcarriers, we need to construct a scenario that can produce significant multipath changes. To the best of our ability, we find a room with area of $2.4m \times 2.4m$ and put all devices in it. To simulate changes in the environment, we make the device and people at the following three kinds of movement states:

- Static: The AP, mobile phone and people are all stationary;
- Dynamic 1: The AP and mobile phone are both static, and four people keep walking around the mobile phone;
- Dynamic 2: The AP is static, and a person with the mobile phone keeps walking.

To study the changes of phases, we define $\nabla\mathbf{\Phi}$ as the phase gradients:

$$\nabla\mathbf{\Phi} = \left[\frac{\mathbf{\Phi}_2 - \mathbf{\Phi}_1}{\mathbf{K}_2 - \mathbf{K}_1} \quad \cdots \quad \frac{\mathbf{\Phi}_{i+1} - \mathbf{\Phi}_i}{\mathbf{K}_{i+1} - \mathbf{K}_i} \quad \cdots \right]. \tag{14}$$

Fig. 6a shows the phases and gradients in above three kinds of movement states. Obviously, when all people and devices are stationary, the phase changes from one subcarrier to another smoothly. The dark gray lines in Fig. 5a indicate that there are randomly phase changes in dynamic environment. If we want to obtain the phase characteristics affected by physical components, it is necessary to purify the sampling CSI phases. After further analysis of the obtained data, we find that there are also phases of some sampled CSIs changing smoothly. The blue lines in Fig. 5a validate our observation. A feasible method is to filter out the frames which contain steep phase changes. Compared with static environment, gradients in dynamic environment change randomly. In other words, $\mathbf{\Phi}$ in static environment is more concentrated than that in dynamic. As shown in Fig. 6a, there are steep variance of phases in dynamic environment. We plot the cumulative distribution of variance on Fig. 6b. The variance of phase gradients in static environment is significantly smaller than that in dynamic environment. Our purpose is to seek out these CSIs whose gradients have smaller variance when the external environment is changing. From above comparison, we design a gradient filter to seek out phases with little gradient variance. Algorithm 1 describes the filtration process. Phases used for fingerprint extraction are filtered as shown in Fig. 5. Obviously, phases of frames after filtering change smoothly

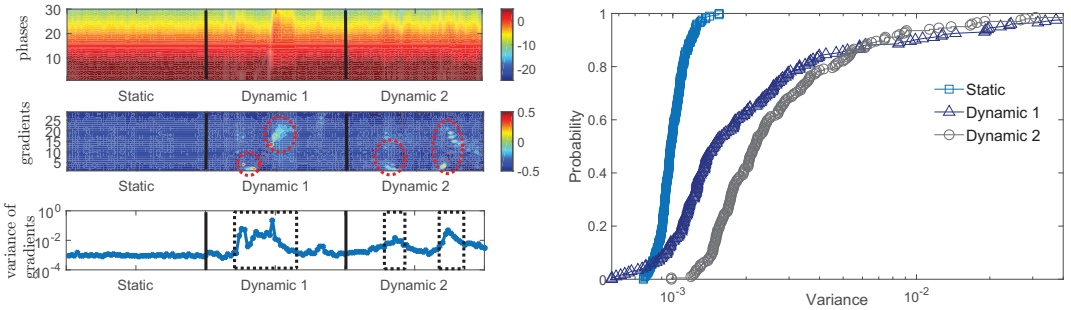


(a) Raw phases of different frames in dynamic scene.

(b) Phases after filtering based on variance of gradients.

(c) Normalized phases based on phase of center frequency.

Fig. 5: Phase filtering and normalization for fingerprint extraction.



(a) Phases and gradients in different scenes.

(b) CDF of gradients' variance in static/dynamic environment.

Fig. 6: Phases in three different states.

Algorithm 1 LPF based on variance of gradients

Input: frames' phases of subcarriers $\mathbf{P} = \{\mathbf{\Phi}^1, \mathbf{\Phi}^2, \cdots\}$, subcarrier sequence $\mathbf{K}$
Output: phases for fingerprinting $\mathbf{P}_f$

$len \leftarrow length(\mathbf{P})$;
for $i \leftarrow 1; i <= len; i++$ **do**
  for $j \leftarrow 1; j < 30; i++$ **do**
    $r_j^i \leftarrow (\mathbf{\Phi}_{j+1}^i - \mathbf{\Phi}_j^i)/(\mathbf{K}_{j+1} - \mathbf{K}_j)$;
  end for
  $\nabla\mathbf{\Phi}^i = [r_1^i, r_2^i, \cdots, r_{29}^i]$;
  if $var(\nabla\mathbf{\Phi}^i) < \varepsilon_{min}$ **then**
    $\mathbf{P}_f.add(\mathbf{\Phi}^i)$;
  end if
end for
return $\mathbf{P}_f$

between adjacent subcarriers.

Elimination of internal impacts: As shown in Fig. 5, the phases of subcarrier -28 are different among sampled CSIs. The reason is that phases are affected by not only environment but also internal components. Phases may be interfered even if there are minor changes in oscillator and amplifier components, by the factors such as temperature, humidity and electric current. Phases of subcarriers can be decomposed as true phase φ_I and component errors $\mathbf{e}$, φ:

$$\varphi = \varphi_I + \mathbf{e}. \tag{15}$$

Here, component errors are different between subcarriers and $\mathbf{e}$ is a vector, $\mathbf{e} = [e_{-28}, e_{-26}, \cdots, e_k, \cdots], k \in \mathbf{K}$. Then the received signal can be rewritten as

$$\mathbf{\Phi} = 2\pi\lambda \cdot \mathbf{K} + \varphi_I + \mathbf{Z} + \mathbf{e} + \boldsymbol{\epsilon} = 2\pi\lambda \cdot \mathbf{K} + \mathbf{Z}^* + \mathbf{E}, \tag{16}$$

where errors from components and I/Q imbalance are combined as $\mathbf{E}$. $\mathbf{Z}^*$ contains the true phase and constant $\mathbf{Z}$. As introduced in Section II, a pair of mirror subcarriers -1 and 1 are sampled. To remove the impact of FDD, SFO and TOF, we sum phases of these two subcarriers as the following equstion:

$$\begin{aligned}\mathbf{\Phi}_1 + \mathbf{\Phi}_{-1} &= 2\pi \cdot \lambda \cdot (-1+1) + 2 \cdot \mathbf{Z}^* + \mathbf{E}_{-1} + \mathbf{E}_1 \\ &= 2 \cdot \mathbf{Z}^* + \mathbf{E}_{-1} + \mathbf{E}_1.\end{aligned} \tag{17}$$

Here, we subtract $\mathbf{Z}^*$ from the phases of each received frame for phase normalization. From preliminary tests in Section II-B, $\mathbf{E}_{-1}$ and $\mathbf{E}_1$ are almost opposite numbers. $\mathbf{Z}^*$ can be calculated by Eq. (18) approximately. After normalization, phases of filtered CSIs are shown in Fig. 5c.

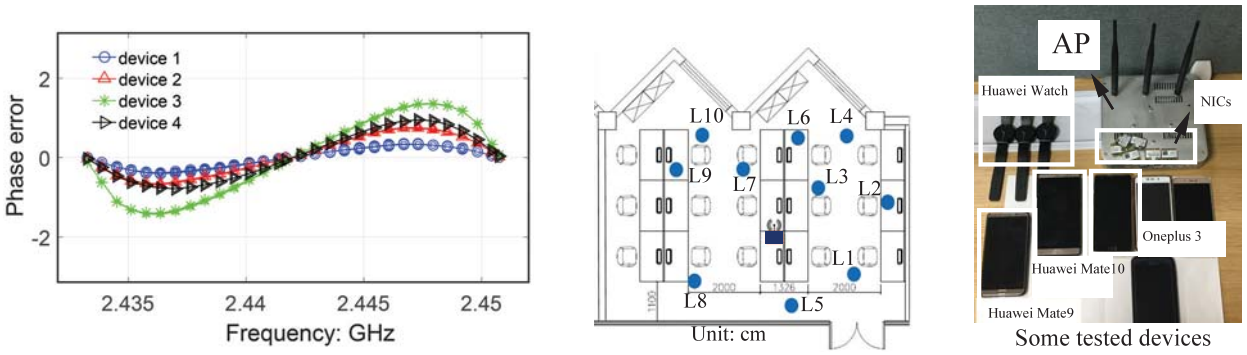
$$\mathbf{Z}^* \approx \frac{\mathbf{\Phi}_1 + \mathbf{\Phi}_{-1}}{2} \tag{18}$$

B. Establishment of Fingerprint Library

Now we see the hope of getting phase errors introduced in Section II-A. Different from the phase errors mentioned earlier, our phase errors are measured in a real scene where the signal is transmitted through the air. Our phase errors can be calculated by the following equation:

$$\mathbf{E} = \mathbf{\Phi} - (2\pi\lambda \cdot \mathbf{K} + \mathbf{Z}^*) \tag{19}$$

where $\mathbf{E}$ contains phase errors from not only I/Q imbalance but also imperfect components of transmitter. The parameter λ is related to FDD, SFO and TOF and will change across sampled CSIs. In order to obtain a steady fingerpeint, we use λ which makes $\mathbf{E}_{-28}$ and $\mathbf{E}_{28}$ equal to zero. Then we can use E_k as the fingerprint of device. Fig. 7a shows the phase errors of four moving smartphones. For each device, we draw the fingerprints extracted six times. It is obvious that



(a) Fingerprint of Four Different Devices.

(b) Experiment site: one access point and ten tested locations of devices.

(c) Hardware: AP and some tested devices.

Fig. 7: Fingerprint of four devices and Implement of Experiment.

the fingerprints are different across devices. The fingerprints extracted six times for the same device are also approximately coincident. Here, it is worthy to emphasize that, we will not apply a fitting curve to fingerprint. Although Zhu *et al.* have fitted the errors due to I/Q imbalance in [12], as shown in Fig. 2a, we find that BCM943224's errors do not match the model Zhu *et al.* proposed. It is because **E** consists of errors caused by imperfect oscillator of NIC. Thus, we use **E** as fingerprint directly and build mapping relations between MAC address and fingerprint ($MAC \mapsto fingerprint$).

C. Attack Detection

For rogue WiFi connection, there are two main types: rogue AP and rogue device. The main idea of detecting rogue AP is to verify the legitimacy of fingerprint after matching SSID and MAC address. When it comes to detecting rogue devices, we assume that there is a white list of authorized devices. Rogue device may use IP/MAC address spoofing or replay attacks to impersonate legitimate devices. When the device sends data to an AP, the AP will extract fingerprint from CSIs at once. Then, the established fingerprint library will be used to check legitimacy which depends on whether the MAC address is match the fingerprint stored in the library.

IV. Experiment Design and Evaluation

Hardware Implementation: In the experiment, we build our system based on industrial personal computers (IPCs) with Intel 5300 WiFi chips. Mobile communication terminals include smartphones and smartwatches, *e.g.*, Huawei, Nexus, SamSung and Huawei Watch. All data is collected from IPCs at 2.4GHz OFDM wireless signal.

Software Implementation: We modified the driver of the network card in IPCs and installed *Linux 802.11n CSI Tool* [13] to collect CSI for fingerprinting. To collect CSI at AP, we also install *Hostapd* in IPC to make the network card work in master mode as an AP. Due to the limitation of *Linux 802.11n CSI Tool*, we set the AP to be non-encrypted. After training, we employ our fingerprint library in the IPC for attack detection and identification.

Fingerprint Library Establishment: In this experiment, we use 30 WiFi devices including 24 mobile terminals and 6 APs. The mobile terminals consist of different types of smartphones and smartwatches. Table I shows the types and quantities of tested mobile terminals. When collecting raw CSI, we send ICMP messages to the connected mobile terminals and estimate CSI from the response frames. In our test in wireless communication, the shortest interval of ICMP ping is nearly 5 *ms* on average. So, we set the packet interval as 5 *ms*. In other words, we collect almost 200 frames per second containing CSIs. For each terminal or AP, we collect data that lasts ten seconds every time, *i.e.*, 2000 CSIs. Our method has advantage in dynamic environment, and it does not require the device to be stationary. We will identify the smartdevices when they are moving in Section V. However, we only let the equipment stand still for the same assumptions when comparing our method with Hua's solution which is based on carrier frequency offset (CFO) [14].

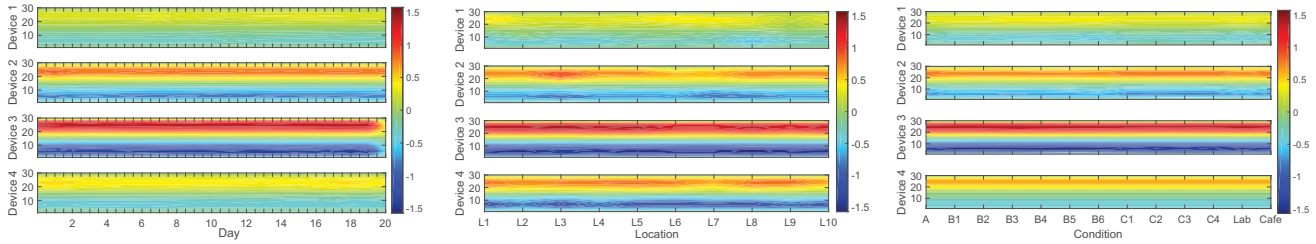
TABLE I: WiFi Terminals.

Device	Quantity	Device	Quantity
laptop with AR9580	3	Huawei Mate 9	2
laptop with AR5B22	3	Huawei Watch	3
OnePlus 3	2	Huawei Mate 10	2
Samsung S7 edge	2	Nexus 6P	2
Other devices	11		

V. System Evaluation

In this section, we will demonstrate the invariance of fingerprints to time, space, and environment, as well as the differences between different devices. Then, we will compare our method's accuracy and time cost with Hua's [14] which is the fastest solution to our knowledge. Finally, under realistic cyber-attacks, we deploy our system to detect attacks.

Time Invariance: For identification, it is necessary for fingerprints to be stable and invariant at different connection times. As our intuitive feeling, the temperature of smartdevices are changing after starting up. It is non-negligible that temperature and humidity are different every day, and CSIs may change [15]. Our fingerprints are phase errors extracted from CSI. So, it is important to ensure the stability of our fingerprints. Fig. 8a shows the phase errors at different time. We measured the CSI in a 4×15 m student office and



(a) Phase errors of four devices remain invariant across the time.

(b) Phase errors of four devices remain invariant at ten different locations.

(c) Phase errors of four devices remain invariant at dynamic environments.

Fig. 8: Invariance of fingerprints with respect to time, location and environment.

extracted fingerprints, day and night for 20 consecutive days, demonstrating the relative stability of the fingerprint over time. From our experiment, phase errors are invariant to time.

Location Invariance: As mentioned in Section II, the phases of recieved signal are sensitive to TOF. Indoor localization methods have been proposed based on that [16]. If the establishment of the device fingerprint library needs to be sampled at all locations, it will be time-consuming and labor-intensive. Fortunately, we have eliminated the effects of TOF when the fingerprint is extracted. To verify that the fingerprint is independent of the location, we set an AP at a student office and use four different smartphones to connect to this AP at different locations L1-L10 as shown in Fig. 7b. We sample the CSIs and plot the phase errors of different smartphones based on distance changes in Fig. 8b. It shows that the phase errors change negligibly with location but are significantly different among devices. It demonstrates that the fingerprints are invariant when locations are changing. In other words, we don't have to sample every location to get fingerprints.

Environment Invariance: Compared with time and location factors, influence from environment on CSI is more dramatic. Such impact can come from crowd, movement of tables and chairs and opening or closing of doors and windows [17]. The invariance of fingerprint to environment is vital because scenarios using smart devices are complex and changeable. To evaluate this property, we first measured phase errors in dynamic environment. We performed our experiments in an office with people of 4 to 5. The following conditions give the experiment details about movement of people and objects:

- **Condition A**: Four people are walking in an office, and one of them is taking along the connected smartphone;
- **Condition B**: The smartphone is deployed at a table, and we move two chairs beside this table at different locations;
- **Condition C**: The smartphone is deployed at a table, and we make the door and window open or closed.

There are six different combinations of two chairs on the edge of a table and four different opening and closing combinations of a door and a window. Fig. 8c shows that phase errors change little in the same room, although the surrounding environment is changing. After that, we tested the same pairs of AP and smartphones at different rooms in the campus such as a laboratory and a cafe. The obvious differences in interior furnishings

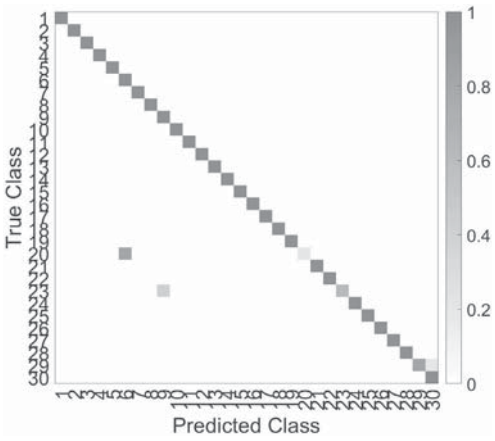


Fig. 9: Accuracy of distinction

can be regarded as dramatical changes in these scenarios. Fig. 8c shows fingerprints extracted in these scenarios. Those experiments confirmed that the proposed fingerprint based on CSI phase error is invariant to the environment.

Accuracy: As shown in Fig. 7a, the extracted fingerprints are different across devices. To create a fingerprint library, we set an AP at an office. In every device, our program samples CSIs 10 times and every time lasts for 20 seconds. Fig. 9 depicts the results of these 30 devices introduced in Section IV. Results indicate that the average accuracy of distinguishing devices ratio is 97.3%. Specially, there were not only different types of equipment, but also the same type of multiple devices. It demonstrates that our proposed fingerprint mechanism can distinguish the subtle differences between the same type of devices. It is impossible for an attacker to build an impersonation attack by using the same type device.

Speed: Comparing to Hua's method [14] that uses CFO as fingerprint and needs at least 6-10 seconds to sample CSIs, our method takes less than one second. We create fingerprint libraries with our method and Hua's from the same raw CSIs. Then, we use them to distinguish devices from the same CSI data. We assume that smart devices have sufficient data processing capabilities and consider the raw data collection time as the time cost of identification. Fig. 10 shows the comparison of them on accuracy and speed. Due to the measurement of phase error is independent of the number of samples, the accuracy of fingerprints extracted from phase errors can exceed 90% in one second. It is nearly $8\times$ faster

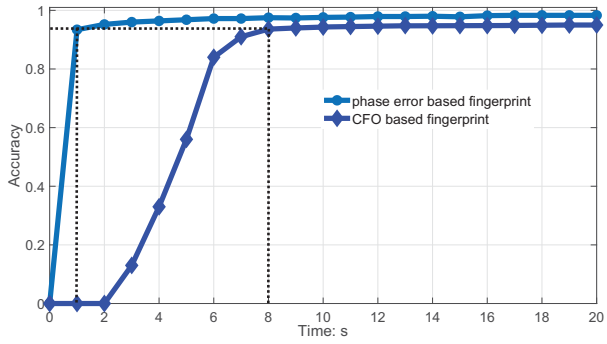


Fig. 10: Comparison of time and accuracy between phase error and CFO as device fingerprint.

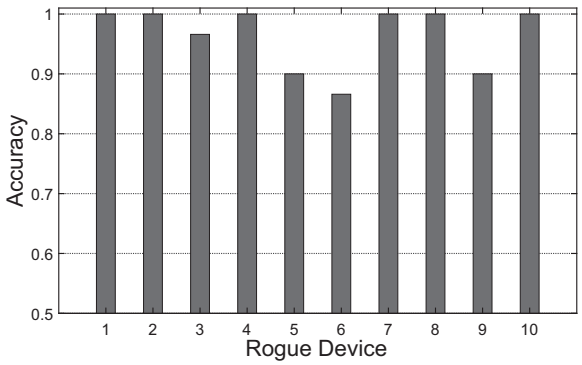


Fig. 11: Identification accuracy of rogue connections.

than the CFO based approach.

Identification and prevention of rogue connections: Given the CSI information from each packet, the receiver calculates the fingerprint of each packet. If the fingerprint does not match the elements in the fingerprint library, the AP or mobile device may drop that packet out. This essentially identify and prevent rogue sender or connection. We randomly selected one-third of the 30 devices as illegal devices and collected 1000 CSI samples per device. The CSI of the 30 devices is fingerprinted one by one and the legitimacy is detected. Fig. 11 shows detection accuracy of these ten rogue connections. Such random tests have been simulated in different scenes. We define the probability that the rogue device is detected as detection rate and the probability that the legal device is misidentified as false alarm rate. The results of detection rates and false alarm rates are displayed in detail on Table II. Although the scenes are different, the detection rates are all above 96%, while the false alarm rate are below 2%. This further demonstrates that our fingerprint mechanism is independent of the environment.

TABLE II: Identification accuracy of rogue connection in didfferent scenes

Scene	Attack detection rates	False alarm rates
student office	97.33%	1.33%
library	96.32%	1.66%
meeting room	96.67%	1.83%

The experimental results demonstrate that our system is capable of identifying and preventing rogue WiFi connections reliably in real-time.

VI. Related work

Recently, device fingerprint-based authentication methods have been proposed to enhance wireless security. Gao *et al.* estimated packet interval time in Network Layer as the AP's feature [18]. In [19], transmit rate in PHY-layer is used for distinguishing device. Although these methods of traffic analysis are effective, a large amount of data collection is required, even up to tens of GB. It takes huge time, storage and computation costs.

Some researchers turned to signal characteristics analysis. Based on the difficulty of faking CSI, Jiang *et al.* proposed a scheme by measuring the CSI of WiFi management frames to ensure the legal status of both communication parties [6]. Li *et al.* used temporal probing and multiple tone probing extracted from channel estimation to authenticate legitimate transmitter [20]. Bagci *et al.* used multiple receivers's CSI to prevent tamper attack in IoT [21]. Their methods are both based on that the location of the device is almost unchanged. Some researchers consider acoustic signal as the fingerprint [22]. However, similar to the CSI, it is sensitive to the environment and not all the devices in the IoT are equipped with microphone and speaker. Hua *et al.* [14] has proposed an algorithm to estimate CFO from CSI as the device fingerprint. Different from directly using CSI, the fingerpint based on CFO is stable and invariant to time changes. It achieves excellent performance in detecting illegal users. But all above methods will not work when the device is mobile.

Combining device motion and radio signal strength, IoT device authentication mechanism Move2Auth [23] was limited to use within 5cm. Similar to Move2Auth, Nirnimesh *et al.* also built a legal identity system through changes of radio signal strength due to movement [24]. It broke the limit of usage distance and did not need any additional equipment in the communication pairs. However, the assistance of a trusted third party is indispensable. All those schemes need 10 seconds to a few minutes to get enough samples. As Pei *et al.* [25] showed that connection set-up time longer than 5 seconds will leave a bad impression on the user. These methods remain trouble mediating contradiction between security and user experience.

Location has also been used for attack detection, Demirbas *et al.* used RSSI based location to detect sybil attack [26]. However, additional devices were also required in their method. Combining the location and CSI, Liu *et al.* authenticated users through multiple monitor [27], and devices were required to be static. Wang *et al.* used beamforming to position spoofing attacks [28]. In his method, smart antenna array is indispensable.

In summary, how to design a real-time WiFi identification and prevention mechanism in a dynamic and complex environment remains as an urgent problem. Inspired by this need and gap, we extract phase mismatches from all subcarriers between signal transmitter and receiver as fingerprints.

VII. Conclusion

This paper has demonstrated that phase error can be used as device fingerprint for real-time identification and prevention of rogue WiFi connections. It not only improves the security between the AP and the device, but also adds a layer of authentication protection between the AP relay networks. Our experiments show that such fingerprints are invariant to time, locations and environments. As a result, such a mechanism can be applied on top of traditional WLAN authentication protocols with little impact on WiFi connection time. The invariant phase error can be extracted from changing CSIs in less than a second, which is at least $8\times$ faster than that of the state-of-the-art solution. The whole identification and prevention process is non-intrusive and transparent to rogue devices. The extensive experiments on COTS APs and smartphones in different scenarios show that the accuracy of rogue connection detection is up to 96% and false alarm rate is below 2%. It demonstrates that the proposed approach is effective and reliable in various environments and setup. In future, we plan to address the problem of fingerprint migration in larger and more complex scenarios with multiple AP networks.

Acknowledgement

This research is partially supported by National key research and development plan 2017YFB0801702, NSFC with No. 61625205, 61632010, 61772546, 61751211, 61772488, 61832010, 61520106007, Key Research Program of Frontier Sciences, CAS, No. QYZDY-SSW- JSC002, NSFC with No. NSF ECCS-1247944, and NSF CNS 1526638. NSF of Jiangsu For Distinguished Young Scientist: BK20150030. Panlong Yang and Xiang-Yang Li are the corresponding authors.

References

[1] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions," *Future generation computer systems*, vol. 29, no. 7, pp. 1645–1660, 2013.

[2] M. Vanhoef and F. Piessens, "Key reinstallation attacks: Forcing nonce reuse in wpa2," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '17. New York, NY, USA: ACM, 2017, pp. 1313–1328.

[3] M. Vanhoef, "Key reinstallation attacks breaking wpa2 by forcing nonce reuse," https://www.krackattacks.com/tools, accessed Apr 12, 2018.

[4] M. Li, Y. Meng, J. Liu, H. Zhu, X. Liang, Y. Liu, and N. Ruan, "When csi meets public wifi: Inferring your mobile phone password via wifi signals," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '16. New York, NY, USA: ACM, 2016, pp. 1068–1079.

[5] M. Antonakakis, T. April, M. Bailey, M. Bernhard, E. Bursztein, J. Cochran, Z. Durumeric, J. A. Halderman, L. Invernizzi, M. Kallitsis *et al.*, "Understanding the mirai botnet," in *USENIX Security Symposium*, 2017, pp. 1092–1110.

[6] Z. Jiang, J. Zhao, X. Y. Li, J. Han, and W. Xi, "Rejecting the attack: Source authentication for wi-fi management frames using csi information," in *2013 Proceedings IEEE INFOCOM*, April 2013, pp. 2544–2552.

[7] R. Maes, A. Van Herrewege, and I. Verbauwhede, "Pufky: A fully functional puf-based cryptographic key generator," in *International Workshop on Cryptographic Hardware and Embedded Systems*. Springer, 2012, pp. 302–319.

[8] 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*, ser. MobiCom '08. New York, NY, USA: ACM, 2008, pp. 116–127.

[9] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using wifi signals," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, ser. MobiCom '15. New York, NY, USA: ACM, 2015, pp. 90–102.

[10] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity wifi," in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. ACM, 2015, pp. 53–64.

[11] Z. Yang, Z. Zhou, and Y. Liu, "From rssi to csi: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, pp. 25:1–25:32, Dec. 2013.

[12] Y. Zhuo, H. Zhu, H. Xue, and S. Chang, "Perceiving accurate csi phases with commodity wifi devices," in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, May 2017, pp. 1–9.

[13] D. Halperin, "Linux 802.11n csi tool," http://dhalperi.github.io/linux-80211n-csitool/, accessed May 17, 2018.

[14] J. Hua, H. Sun, Z. Shen, Z. Qian, and S. Zhong, "Accurate and efficient wireless device fingerprinting using channel state information," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, April 2018, pp. 1–9.

[15] K. Ohara, T. Maekawa, and Y. Matsushita, "Detecting state changes of indoor everyday objects using wi-fi channel state information," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 88:1–88:28, Sep. 2017.

[16] R. Nandakumar, K. K. Chintalapudi, and V. N. Padmanabhan, "Centaur: Locating devices in an office environment," in *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, ser. Mobicom '12. New York, NY, USA: ACM, 2012, pp. 281–292.

[17] X. Chen, C. Ma, M. Allegue, and X. Liu, "Taming the inconsistency of wi-fi fingerprints for device-free passive indoor localization," in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, May 2017, pp. 1–9.

[18] K. Gao, C. Corbett, and R. Beyah, "A passive approach to wireless device fingerprinting," in *2010 IEEE/IFIP International Conference on Dependable Systems Networks (DSN)*, June 2010, pp. 383–392.

[19] C. L. Corbett, R. A. Beyah, and J. A. Copeland, "Passive classification of wireless nics during active scanning," *International Journal of Information Security*, vol. 7, no. 5, pp. 335–348, Oct 2008.

[20] Z. Li, W. Xu, R. Miller, and W. Trappe, "Securing wireless systems via lower layer enforcements," in *Proceedings of the 5th ACM Workshop on Wireless Security*, ser. WiSe '06. New York, NY, USA: ACM, 2006, pp. 33–42. [Online]. Available: http://doi.acm.org/10.1145/1161289.1161297

[21] I. E. Bagci, U. Roedig, I. Martinovic, M. Schulz, and M. Hollick, "Using channel state information for tamper detection in the internet of things," in *Proceedings of the 31st Annual Computer Security Applications Conference*, ser. ACSAC 2015. New York, NY, USA: ACM, 2015, pp. 131–140.

[22] P. Xie, J. Feng, Z. Cao, and J. Wang, "Genewave: Fast authentication and key agreement on commodity mobile devices," in *2017 IEEE 25th International Conference on Network Protocols*, Oct 2017, pp. 1–10.

[23] J. Zhang, Z. Wang, Z. Yang, and Q. Zhang, "Proximity based iot device authentication," in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, May 2017, pp. 1–9.

[24] N. Ghose, L. Lazos, and M. Li, "Sfire: Secret-free in-band trust establishment for cots wireless devices," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, April 2018, pp. 1–9.

[25] C. Pei, Z. Wang, Y. Zhao, Z. Wang, Y. Meng, D. Pei, Y. Peng, W. Tang, and X. Qu, "Why it takes so long to connect to a wifi access point," in *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, May 2017, pp. 1–9.

[26] M. Demirbas and Y. Song, "An rssi-based scheme for sybil attack detection in wireless sensor networks," in *2006 International Symposium on a World of Wireless, Mobile and Multimedia Networks(WoWMoM'06)*, 2006, pp. 5 pp.–570.

[27] 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*, ser. ASIA CCS '14. New York, NY, USA: ACM, 2014, pp. 389–400.

[28] T. Wang and Y. Yang, "Analysis on perfect location spoofing attacks using beamforming," in *2013 Proceedings IEEE INFOCOM*, April 2013, pp. 2778–2786.