Device-Free Secure Interaction with Hand Gestures in WiFi-enabled IoT Environment

Yanchao Zhao, *Member, IEEE,* Ran Gao, Shangqing Liu, *Student Member, IEEE,* Lei Xie, *Member, IEEE,* Jie Wu, *Fellow, IEEE,* Huawei Tu, *Member, IEEE,* Bing Chen, *Member, IEEE,*

***Abstract*—Recent research advancement of wireless sensing technology has made device-free interaction in WiFi-enabled the IoT environment possible. Although gesture-based interaction with such a smart environment greatly improves usability, it also introduces many security problems such as shoulder surfing attacks. By spoofing the gestures of legitimate users, the attacker could easily access private information or services and cause even worse consequences. A secure interaction mechanism for this environment is required to prevent attackers without compromising the usability, while the limited recognition ability and low robustness of WiFi sensing make this target extremely challenging. To this end, we propose a secure interaction mechanism called SiWi, which provides the ability to resist shoulder surfing attacks without compromising the usability by using just WiFi signals. SiWi innovates in a concurrent interaction/authentication framework with only three elemental gestures (push, swing, and wave) and four types of identity-related imperceptible/hidden features (time distribution, direction, angle, and distance). HMM, and Fresnel Model-based algorithms are used to recognize the gestures and extract hidden features robustly and efficiently. Extensive experiments in a real implemented system were conducted to investigate the effectiveness of the proposed secure interaction system. The results show that our system can achieve an average accuracy of 93% to identify legitimate users and 97% to resist the spoofer.**

***Index Terms*—Wireless Sensing, Device-Free Sensing, Channel State Information, Secure Interaction, Gesture Recognition**

I. Introduction

A. Motivations

As a device-free sensing scheme, sensing based on Channel State Information (CSI) of WiFi has drawn considerable research attentions recently [1], [2], [3], [4]. Such a sensing scheme has many advantages such as low-cost (software updated without specified hardware), easy deployment (pervasiveness of WiFi devices) and non-LOS(Light of Sight) sensing. Therefore, it has been envisioned to be a key enabling technology in so-called WiFi-enabled IoT environment (e.g. smart home or offices) [5]. It offloads sensing functions from users' devices to surrounding infrastructures and allows users to device-freely receive same services as wearable-device sensing [6], [7]. Numerous applications (e.g. tracking and sleep monitoring) have been developed with the realization of WiFi-enabled IoT environment.

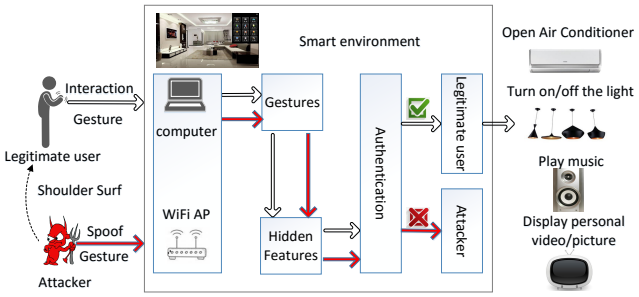


Fig. 1. Scenario of Secure Interaction with WiFi-enabled IoT environment

With advances in gesture recognition [8], [9], [10], [11], [12], [13], WiFi-enabled IoT environment could further enable people to interact and trigger personalized services or access private information [14]. As illustrated in Fig. 1, people can interact with the environment device-freely with gestures and control appliances or trigger services (e.g. control the air condition and display personal video or picture.) at their willing. Although such systems enrich personal interaction usability and improve interaction efficiency, security has become a major concern for these systems. For example, in Fig. 1, the attacker could shoulder-surf an input gesture and spoof the gesture to the IoT system, so as to access private services or information illegally. Thus, it is indispensable to require secure interaction mechanisms for the smart environment.

B. Limitation of State-of-the-Art

The primary requirement of such secure interaction mechanisms is to conduct user authentication during the interaction process. User authentication in computing systems traditionally depends on three factors: what you have (e.g. a hardware token), what you are (e.g. a fingerprint) and what you know (e.g. a password) [15]. Among all existing authentication methods, gestures belong to the category of what-you-know methods. They have advantages over other

Yanchao Zhao, Ran Gao, Shangqing Liu, and Bing Chen are with college of Computer Sciences and Technology, Nanjing University of Aeronautics and Astronautics; Collaborative Innovation Center of Novel Software Technology and Industrialization, Nanjing, China.

Lei Xie is with State Key Laboratory of Novel Software Technology, Nanjing University, Nanjing, China.

Jie Wu is with Center for Networked Computing, Department of Computer and Information Sciences, Temple University, USA;

Huawei Tu is with Department of Computer Science and Information Technology, La Trobe University, Melbourne, VIC, Australia

Yanchao Zhao and Huawei Tu are the corresponding authors. Email:yczhao@nuaa.edu.cn, H.Tu@latrobe.edu.au;

This work are supported in part by National Key R&D Program of China under Grant 2017YFB0802300, in part by National Natural Science Foundation of China (No. 61872174, 61832008, 61602242), in part by Natural Science Foundation of Jiangsu Province(No. BK20200065) in part by NSF grants CNS 1824440, CNS 1828363, CNS 1757533.

authentication methods, because gestures are highly customizable, easier to remember, and more secure potentially. Clark and *et al.* [16] have conducted a comparable study of gesture based authentication methods with different devices, including cameras, smart-phones, wearable sensors and WiFi. Their study did not fully consider recent advancement of WiFi-based activity recognition, thus underestimated the potential of WiFi-based gesture authentication. As a classic gesture based authentication work, the GEAT [17] used not only gestures but also the features like finger velocity to perform authentication on smartphones. This work was based on touch screen and the features or gestures may not be suitable for WiFi-enabled IoT.

So far, only limited works focus on device-free identification or authentication with WiFi. Some identified a person with gait features [18], [19]. Such identification approaches have two limitations. First, the identification process requires a relatively long walking distance to gather sufficient features, which restricts its usability in a large open indoor environment. Second, some services of the smart environment should only be triggered pro-actively by the user specified activities, but gait-based approaches are not flexible enough to do so. Shi *et al.* [20] also adopted the idea of extracting physical features of gestures to perform user authentication. However, the proposed approach was based on the neural network to extract the features, which highly depended on training datasets and required unacceptable labor efforts to collect labeled data. Such a cost hinders this scheme applicable for secure interaction.

C. Proposed Approach

To address the above problems, we propose a system called *SiWi (Secure Interaction via WiFi Signal)*, which is a device-free gesture based scheme considering the Shoulder Surfing Attack [17] as the major threat. Our idea is based on the observation that when users interact with WiFi based IoT system, the inherent features (e.g. time duration, distance and speed) related to the gestures have a strong correlation with the identity of the users but can hardly be perceived by the attacker. Besides, according to previous studies (e.g. [16]), gesture combination could serve as an authentication "password/PIN". As such, we proposed a two phase secure interaction scheme by combining "what-you-know" gestures and "what-you-are" features, where SiWi utilizes the CSI of WiFi signals to recognize and extract the gesture features. At the first phase, the sequence of gestures is used to examine whether the user knows the "password/PIN". If it passed, the extracted "what-you-are" features are used to perform the second phase authentication to exclude spoofers.

D. Technical Challenges and Solutions

To realize the full promises of SiWi, there are still several technical challenge:

1) The Security vs. Usability: Achieving high security and usability are two major goals when designing an interactive system. Usability refers to how easily a user interacts with a system while security refers to the ability of a system to resist attacks. For our system design, we need to strike a balance between the two goals. Complicated gestures are more likely to resist shoulder surfing attacks, but this complicated combination may lead to low usability. Therefore, we introduced three elemental gestures (push, swing and wave) for customizing gesture combinations. Empirical study has shown that these three gestures and their combinations included enough hidden features for user authentication, and some hidden features (e.g. speed and direction) could add versatilities to these gestures and thus generate more combinations for interaction and improve the security.

2) Uniqueness of what-you-are features and their feasibility with WiFi: Although the CSI of WiFi has been proved to be capable of fetching identity-related features from the gait [18], [19], we still need to address what features of the body gestures are closely related to the identity and can be robustly extracted by the WiFi based IoT system. Hence, we first identify the unique features of the gestures, including *Inter-gesture interval/Gesture Duration, Distance/Arm Length, Direction, and Swing Angle/Range* through experimental data analysis. We then proposed a robust segmentation algorithm and a Fresnel Zone model based algorithm to extract these features robustly.

3) Interaction & Authentication Simultaneously: In the smart environment, interaction gestures require immediate response, thus leaving very limited time for our recognition and authentication process. How to find the low cost and real time recognition and extraction algorithm is a very challenge task. In this paper, we mainly design the SiWi composed of a high efficient signal processing, gesture detection and segmentation algorithm, a HMM-based recognition algorithm and a Fresnel zone modeling based extraction algorithm, all of which are efficient enough to complete required task in real-time.

4) What-you-know vs. What-you-have: We have to find a trade off between the authentication based on explicit and implicit features, while ensuring the accuracy. A reasonable identification method of the integrated features needs to be designed. Apparently, these two factors need to play different roles in the authentication. Moreover, extracting two types features incur different amount of cost. How to balance the identification accuracy and the cost and the response time need a careful consideration. Only three elemental gestures are enough to form specified gestures.

E. System Model and Threat Model

For simplicity purposes, our system only considers the case of a single user interacting with the smart environment in a device-free sensing manner. The orientation of the operation is arbitrary in a close range to the signal receiving devices.

In the environment, an attacker could witness the gestures by the user through either remote videos or in-site sneaking. With the sneaking results, we assume the attacker can launch shoulder surfing attacks by spying the owner when s/he performs gestures.

F. Contributions

In summary, we make following contributions in our paper.

- We proposed a gesture based secure interaction scheme for the WiFi-enabled IoT system. This scheme resulted in a good tradeoff between usability (forming enough combinations with three elemental gestures) and security (preventing the shoulder surfing attacks to a large extent).
- We conducted a 4-days empirical study with 10 volunteers to identify a set of features that represent intrinsic attributes of gesture-based interaction in the WiFi-enabled IoT system.
- We proposed a set of algorithms to extract the features effectively and reliably from gesture-based interaction with CSI of WiFi signal.
- We implemented and evaluated SiWi in a real deployment. The results showed that our system can achieve an average accuracy of 93% to identify legitimate users and 97% to resist spoofers.

II. Related Work

A. Gesture based Authentication

Passwords, as a traditional authentication method, require users to remember either certain secure texts or graphical patterns. Such an authentication system solely relies on the knowledge of the password and thus easily suffers from password stolen or shoulder surfing. To overcome this weakness, researchers seek for behavioral features for authentication, such as using gestures. Liu*et al.*[17] proposed GEAT, a gesture based user authentication scheme for securely unlocking touch screen devices. GEAT applied gesture features (such as finger velocity, devices acceleration and stroke time) to user authentication and achieved an average equal error rate of 0.5% with 3 gestures. Luca *et al.* [21] used the time of drawing a password pattern on touch screen phones for authentication and their scheme achieved an accuracy of 55%. Sae-Bae *et al.* [22] performed user authentication based on the time of articulating five-finger gestures. Li et al. [23] exploited five basic movements (tapping, sliding up, down, right and left) on the touchscreen and their related combinations as the user's behavioral pattern features, to perform authentication. This gesture combination contains user's distinct behavioral characteristics. Shrestha [24] introduced a Wave-to-Access approach using timestamps, duration for hand wave gesture to prevent malware attack on smartphones. Yang [25] implemented an approach using sampling interval and acceleration along the x, y & z behind the hand wave gesture for locking / unlocking purposes. However, these studies are based on smartphones.

Using physiological biometrics to authenticate is a novel way. Jungpil [26] proposed a novel user identification system based on the bio signal analysis of arm movement (3-axis accelerometer & 3-axis gyroscope) and electromyography (EMG) signal using Myo armband as a wearable user authentication system in P2P systems. A limitation of these approaches is that they require special hardware and are not applicable to daily usage.

B. WiFi based Gesture Recognition

With Halperin [27] simply modifying the Intel 5300 802.11 Network Interface Card (NIC), a finer grained channel estimation of CSI values could be obtained for activity recognition. The key intuition of WiFi sensing is that performing certain activities in a unique formation and direction could generate a unique pattern in the time series of CSI values. Prior work on WiFi sensing can be classified into two main categories: activity recognition, and user authentication and privacy protection.

1) Activity Recognition: CSI measurements have been widely used for large scale activity and small scale motion recognition in single human environment [28], [29], [30]. For large scale activity recognition, E-eyes [31] used CSI histograms as fingerprints to separate different daily activities. CARM [8] proposed CSI-speed model and CSI-activity model to quantify the correlation between CSI values and the specific human activity. WiDir[32] used the Fresnel zone model from a physical level to estimate the moving direction. For small scale motion recognition, WiFinger [33] used WIFi signals to recognize a set of finger-grained gestures and achieved up to 90.4% average classification accuracy for 9 digits finger-grained gestures from American Sign Language(ASL). WiGest [34] identified different signal change primitives from mutually independent gesture families with no modifications and no training for gesture recognition. WiGest detected the basic gestures with an accuracy of 87.5% with a single AP only and the accuracy increased to 96% using three overheard APs. In the latest work, Wang [35] proposed a system (QGesture) to measure the movement distance and direction of human hands. Based on LEVD algorithm, QGesture achieved an average accuracy of 3cm in the measurement of movement distance and more than 95% accuracy in the movement direction detection in the one-dimensional case. Furthermore, in the two-dimensional case, QGesture also had a state of the art performance.

2) User Authentication and Privacy Protection: CSI characterizes the state information of wireless channel, which can be used to authenticate users and protect privacy. Liu *et al.* [36] proposed a framework to build user profile resilient to the presence of spoofer. WiWho [18] and WifiU [19] used WiFi devices to capture fine-grained gait patterns for human identification. Wobly [37] encoded individuals' gaits and room configuration information by WiFi signals to protect privacy. FingerPass [38] uses the channel state information of WiFi signal to achieve continuous authentication by finger gestures. The system divides authentication into two stages: login and interaction. In the login stage, a deep learning-based approach is developed to extract behavioral characteristics of finger gestures for highly accurate user identification which highly depended on training datasets and required unacceptable labor efforts to collect labeled data.Some others introduce the acoustic signal into this area. Zhou et.al. [39] propose an innovative cracking method on Android Phone with only acoustic signals. This method greatly extends the reach of acoustic signal and its application in mobile devices.

With above system model and threat model, we are going to study how to achieve accurate interaction with the users while ensuring the security of such system to resist above attacks.

III. Empirical Study

In this section, we mainly answer these questions: 1) which gestures could be used as *what-you-know* gestures? and 2) which features of these gestures could be used as *what-you-are* features? We answer these questions through empirical studies. The algorithm detail of recognition and extraction will be presented later in section IV.

Fig. 2. The real deployment

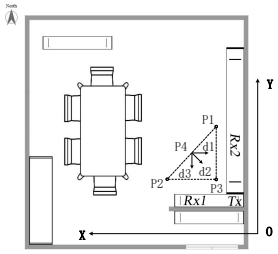


Fig. 3. Data acquisition environment

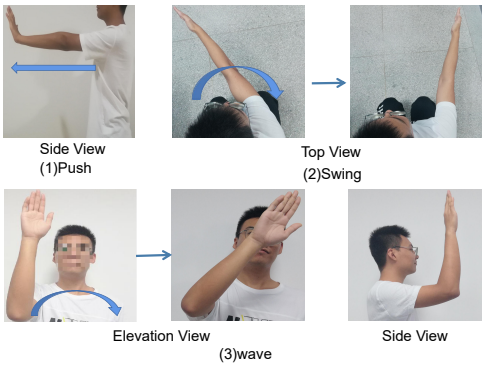


Fig. 4. Gestures used in our scheme for authentication.

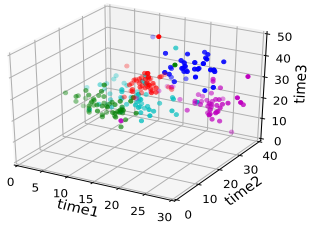


Fig. 5. The time distribution of different people conducting two gestures

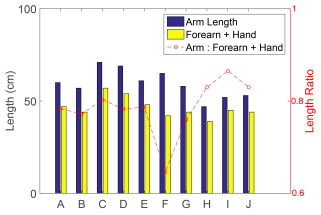


Fig. 6. Relation with volunteers' identity and arm length

A. Data Collection

We used MiWiFi as the signal transmitter, and two Intel NUC with Intel Link 5300 WiFi NIC as the signal receivers (Fig. 2). Their floor plan is shown in Fig. 3. The distance between the transmitter and the receiver was 1m, and the connections between the sender and the receiver made a right angle.

Ten volunteers with different heights and arm lengths took part in our experiments, and the data were collected from each volunteer over 4 different dates. Each volunteer was required to carry out 5 different gestures (push, swing, wave, sit and walk). Worth to mention that, we chose these gestures mainly due to the fact that they can be recognized effectively and stably in wireless sensing system. In addition, we can also extract biometric features for recognition. For the convenience of measurement, the position and moving direction are on/along the vertical line and tangent line of Fresnel zone. Volunteers needed to perform each gesture in 4 locations (P1, P2, P3, P4 in Fig. 3). Note that, we have marked the x-axis and y-axis in the figure, where they are along the line Tx-Rx1 (x-axis) and line TX-Rx2(y-axis). In each position, each participant performed the gestures by facing 3 different directions (d1, d2, d3) and repeated them 20 times. The volunteers also walked from P4 to the other three points along with their CSI recorded. In total, we obtained 28800 samples.

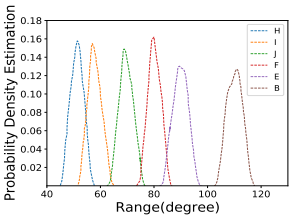


Fig. 7. The distribution of the swing angle/range for 6 volunteers

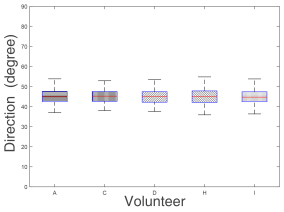


Fig. 8. The statistic of push direction in 45 degree of different volunteers

B. Data Analysis

1) What-you-know Gestures: In order to test the stability of recognizing these gestures, we used the algorithm proposed in IV to perform gesture recognition. As shown in Fig. 15, the recognition of these activities is obvious and stable. Worth to mention that, sitting and walking are not suitable for interaction. Primarily, they are too common in ordinary life; thus, the system will not be able to distinguish between intended input and casual behaviors. In addition, even with proper designed preamble action or sequence, sitting/walking need body movements rather than hand/arm movements. Users will not chose them for interaction in WiFi enabled smart environments. These three gestures in Fig. 4 represent the gestures using hand and arms towards three orthogonal directions. Our following analysis also shows that they contain enough identity-related features. Given the above three factors, we chose the first three gestures as the elemental what-you-know gestures.

2) What-you-are features: Through numerous of data collection and experiment, we examined multiple features which could be reliably extracted with algorithms and have inherent relation with user identity. In this section, we empirically examine four types of gesture features: *Inter-gesture interval/Gesture Duration, Distance/Arm Length, Swing Angle/Range and Push Direction*. The results are as follows.

- **Inter-gesture interval/Gesture Duration**: We divided continuous time data into active data and stationary data, and extract the duration of each segment in time domain of signal. Through our data analysis, we found that when

performing the same gesture combination, the same person will have the similar time distribution, but different people may have different time distributions. Fig. 5 shows an example of the time distribution of pushing hand, interval and swinging hand. The dots of different colors indicate the time distribution of five different participants. The interval distribution from the same person is concentrated, while the one from different people is relatively independent. Hence, the time series distribution can be deemed as hidden features for authentication.

- **Distance/Arm Length**: Intuitively, when performing a gesture, arm's moving range or distance should be closely related to arm length. Participants with different heights should have unique moving range or distance for the same gesture. We conduct data analysis based on this straightforward fact.

  As shown in Fig. 4, when people push their hand, the movement distance is their arm length, that is equal to the sum of the length of the forearm and the length of upper arm. When people swing their arm, the movement distance depends on the arm length and the angle of swinging. When people wave their arm, the movement distance what we can measure is the projective length of hand and forearm in a horizontal two dimensional plane. It depends on the angle of waving and the sum of the length of hand and the length of forearm. The above conclusion shows that the movement distance is related to two parts of human body, the first is the sum length of forearm and big arm, the second is the sum length of forearm and hand. Fig. 6 shows the statistic of the real measured arm length and the forearm length from each volunteers. In the figure, 'A', 'B',...'J' represent different volunteers. (Unless otherwise stated below, the capital letters in the drawing notes indicate different volunteers). Interestingly, the proportions of the forearm and the big arm (the red dash line in the figure) are different for different people. Thus, the distance/arm length could serve as the identity-related features and the extraction of this feature is feasible.
- **Swing Angle**: Another what-you-are feature is the swing angle for the swing gesture, (Fig. 4). Such a feature depicts the range when people swing their arm. Although people may lead to different ranges/angles when they perform this gesture, the distribution is fairly stable. This can be verified through statistical experiments (Fig. 7). Also, according to this figure, it can be found that for different people, their angle distributions for their gestures are different. Thus, the swing angle could reflect intrinsic characteristics of swing their arm. Besides, the swing angle could also serve as a what-you-know element. The extraction algorithm for the swing angle is described in subsection IV-D.
- **Push Direction:** Push gestures not only indicate the arm length but also provide another hidden feature of push direction. We used the proposed algorithm in subsection IV-D to extract the push direction. The statistical results are shown in Fig. 8, from which we can tell that the push direction does not vary too much among participants. However, this feature could be used as what-you-know feature and add versatility to the push gesture, thus generating more interaction gesture combinations.

C. Result Summary

According to the analysis results, we can draw the following conclusions. First, the gestures (push, swing, wave) are perceptible, which can be used as what-you-know elements for secure interaction. In addition, the arm length/forearm length and swing angle revealed in different gestures could be used as what-you-are elements, as they show close correlation with user identity. Second, what-you-know gestures and what-you-are features are not fully exclusive. The interval and duration of the gestures could be controlled by the user intentionally (what-you-know) or just as their custom (what-you-are). The same is true with the push direction. Third, not all what-you-know features are perceptible. The direction, time interval and the angle of the swing could all be used as what-you-know components of the input gesture combination, and these intentional features could also be used as the hidden information to resist shoulder surfing attack. Fourth, we could extract different hidden features from different gestures. For example, the swing angle could only be extracted from the swing gesture, and the forearm length could only be extracted from the wave gesture. Thus, intuitively, it is recommended to come up with a combination consist of all gestures in various sequences to improve the level of security.

In the next section, we will describe how to recognize the gestures and extract the hidden/imperceptible features in a robust and efficient way.

IV. System Design

A. System Overview

According to the analysis above, we design our system as illustrated in Fig. 9, which mainly consists of four parts.

In the first part, SiWi obtains the raw CSI waveform by commercial wireless devices (e.g. Intel 5300NIC). Due to the hardware deficiency and signal interference, raw CSI data commonly have noises. Hence, we firstly use ButterWorth filter to remove high frequency noises. For noises within band, we use Principal Component Analysis (PCA) to process further. Then we choose Discrete Wavelet Transform (DWT) to transform the time series data into the frequency spectrum to obtain the detailed features of the gesture. We also design a novel algorithm to correctly split each gesture. The duration of each gesture (T_{act}), the duration of two adjacent gestures (T_{inter}), the arm length and the gesture angle are features for identity authentication.

In the second part, SiWi utilizes Hidden Markov Model (HMM) to distinguish the gestures. In our system, HMM distinguishes each gesture (i.e. wave, push hand, swing) correctly. The recognized sequence of gestures is compared with the user profile to check whether the two match. However, the spoofer can imitate these gesture combinations easily. Hence, this is just the first phase for identity authentication in our system and we need to rely on hidden features behind these gestures for further authentication.

Next, we extract the hidden features with the help of Fresnel zone formed by WiFi signal propagation. We can calculate the

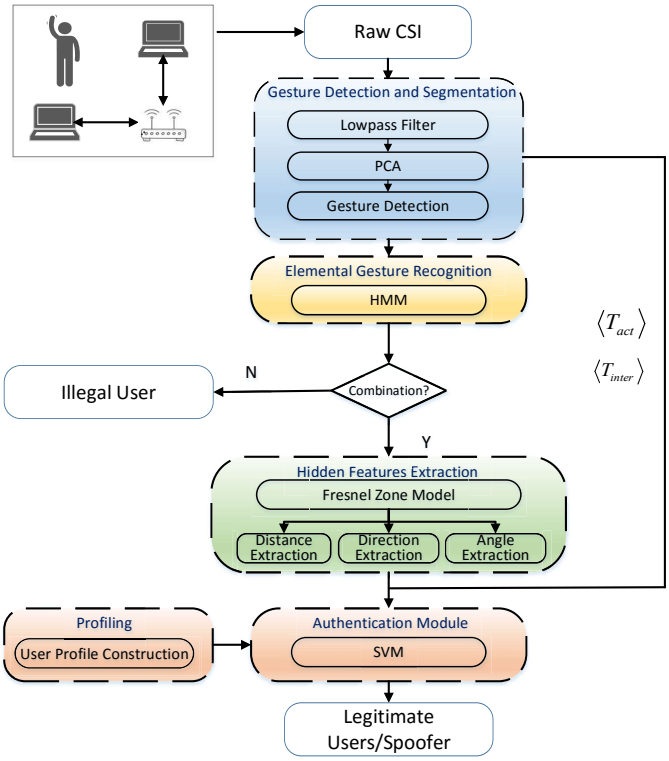


Fig. 9. The framework of SiWi

distance, push direction and swing angle from 2-Dimension Fresnel zone model. These hidden features are relatively stable and as proposed in the last section, the diversity among different people is obvious. Hence, the extracted features are used for authentication of the next step.

The last part is the authentication module which trains a robust model with these hidden features (arm length, angle, T_{act}, T_{inter} and direction), we utilize traditional Support Vector Machine (SVM) to generate the model. Meanwhile, we can determine user identification by comparing the similarity of features between the testing sample and user profile.

As illustrated in Fig. 10, SiWi extracts features from the activity elements and interval elements. The activity elements contain the activity type, activity direction, activity distance and the time cost of the activity. The interval element means the intervals between a sequence of activities. We can distinguish the types of activities by HMM and the directions and distances can be calculated by Fresnel zone. The time cost of each activity and intervals between a sequence of activities can be measured by the process of gesture detection and segmentation. The features such as direction, distance, time cost are correlated with user's shape and habits. However, the activity type is training dependent, which means that we need to collect samples to build a HMM model to distinguish different activities.

B. Gesture Detection and Segmentation

1) Wave Filter: As shown in Fig. 11, the original CSI amplitude features extracted from commodity WiFi NICs are noisy owing to the fast fading characteristics of the wireless

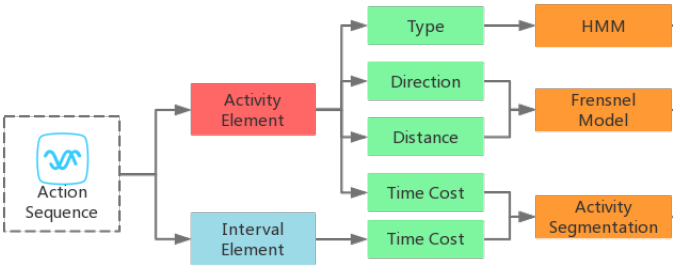


Fig. 10. The activity classification of SiWi

channel and the electromagnetic interference in the environment. As the frequency of the variations in CSI stream due to gestures is within 1-60Hz, we first remove high frequency noises through low-pass Filter. We set the cutoff frequency of 60Hz to remove the noises. However, the noise within band can not be eliminated. Hence, we have to do further processing for low frequency noises. Worth to mention that, the ButterWorth filter could be the bottleneck that hinder the whole system to response in realtime, if the sampling rate exceed 800hz. For high sampling rate case, it it recommend to use average filter.

To remove low frequency noises and combing CSI streams, we apply principal component analysis (PCA) and discrete wavelet transform (DWT) to do further processing. Details of PCA noise removable for CSI of WiFi signal could be found in [8]. From Fig. 11, we can see that the noises are removed.

The CSI waveform after ButterWorth Filter can be expressed as $H(t) = [H(1), H(2), ..., H(t)]$ and $H(t)$ is denoted as the following equation:

$$\begin{pmatrix} H_{1,1} & H_{1,2} & \cdots & H_{1,30} \\ H_{2,1} & H_{2,2} & \cdots & H_{2,30} \\ \vdots & \vdots & \vdots & \vdots \\ H_{6,1} & H_{6,2} & \cdots & H_{6,30} \end{pmatrix} \quad (1)$$

where $H_{i,j}$ represents the CSI waveform at j_{th} sub-carrier in the i_{th} Tx-Rx pair. We apply PCA by the following three

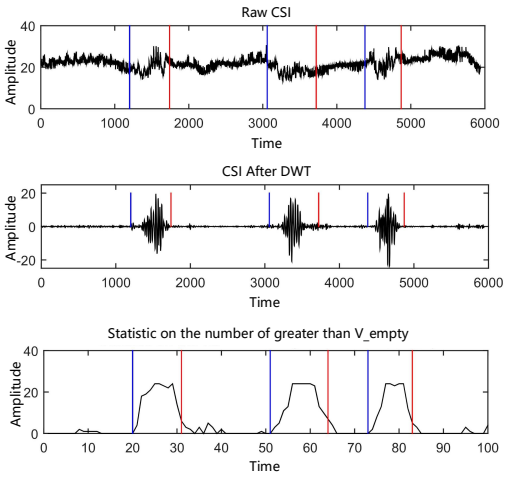
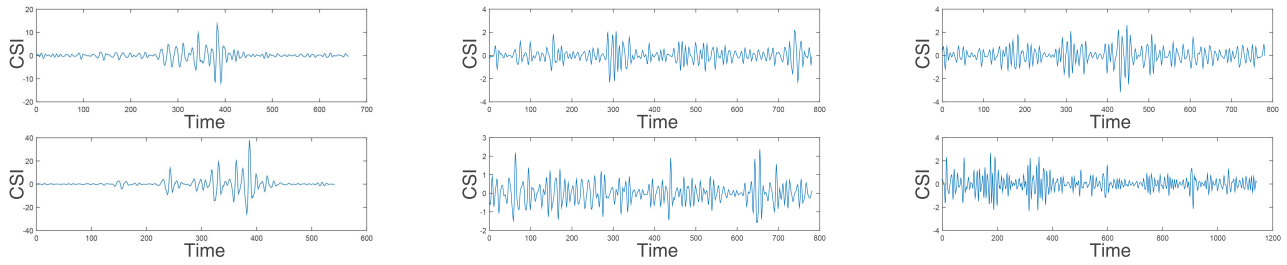


Fig. 11. Activity Detection and Segmentation



(a) Push (b) Swing (c) Wave

Fig. 12. DWT Data of Different Gestures

steps:

- *DC component removal:* we first subtract the corresponding constant offsets to remove the Direct Current(DC) component for every subcarrier. The DC component can be calculated through a long term averaging over that subcarrier.
- *Eigenvector decomposition:* SiWi gets the eigenvector q_i through decomposition of the correlation matrix M as $M = H^T \times H$.
- *Principal Components:* In this step, we construct the principal components using the equation $h_i = H \times q_i$, where q_i and h_i are the i^{th} eigenvector and the i^{th} principal components respectively.

SiWi extracts the 2^{rd} to 11^{th} components and discards the first principal component. After PCA, we apply DWT to transfer the time series data into frequency spectrum. SiWi selects the Daubechies D4 wavelet to decompose the PCA components into ten levels. The frequency features can be expressed as matrix D:

$$\begin{pmatrix} D_{t_{1,1}} & \cdots & D_{1,30} \\ \vdots & D_{t_{i,j}} & \vdots \\ D_{t_{n,1}} & \cdots & D_{t_{n,m}} \end{pmatrix} \tag{2}$$

where $D_{t_{i,j}}$ represents the result of i^{th} dimensional PCA and j^{th} level frequency spectrum values.

2) Activity Detection: The core of gesture detection is to compare the waveform of existent gesture with the empty room to get the start and end point. The algorithm is presented as follows:

- **Step 1:** To detect the start and end of an gesture, SiWi needs the benchmark value of CSI when the room is empty which denotes as H_{empty}. We also need to do the wave filter for these values and get the mean variance V_{empty}.
- **Step 2:** For gesture waveforms, we divide the data into T/t time slices according to the time window size t. For the i^{th} time slice, we get the variance V as the basis for judgment. We perform a statistic on the number of V greater than V_{empty}. If the number is greater than the threshold, the time slice is regarded as the beginning of the gesture.

Fig. 11 shows the waveform of pushing hand continuously for three times. The blue vertical line represents the start of each gesture, the red vertical line represents the gesture ended and we can see that our algorithm correctly splitting each gesture. Meanwhile, we can get the time interval T_{inter} between two adjacent gestures. Note that, in our system three different gestures are used. The start and the end of the gestures may not be the same. This may cause the user performing an extra movement before conducting next gesture. We find that these extra movement could be detect and split in segmentation stage. To alleviate such interference, this extra movement segment will be recognized as an incomplete push/pull. We only deem this incomplete push/pull as a gesture interval and merge it with its neighboring intervals.

C. Recognition of Elemental Gestures

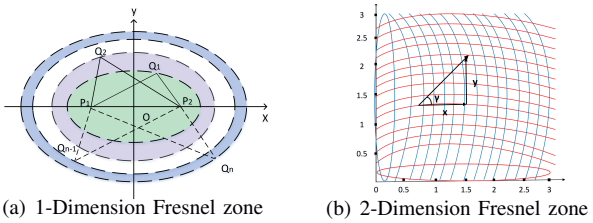
The two typical wave forms of each gesture after DWT are illustrated in Fig. 12. To distinguish these three types of gestures in this form, SiWi extracts the average energy and variance in each time slice as features and uses Hidden Markov Model(HMM) to build a CSI gesture recognition model. As human gestures generally comprise different states, for example push hand can be divided into start, acceleration, deceleration and stop. These states correspond to the concept of states in HMM. When people are in the relatively slow movement state, most CSI energy is on the low frequency components. However, when entering into the acceleration state, the energy is distributed in high frequency components. By looking at these transitions between different states, we can infer different types of gestures. Hidden Markov Model is a special kind of Bayesian network. The variable Y_t denotes t^{th} node in the network and each node has possible states. For variable $Y_1 \cdots Y_T$, we have

$$P(Y_1 \cdots Y_T) = P(Y_T|Y_{T-1})P(Y_{T-1}|Y_{T-2}) \cdots P(Y_1) \tag{3}$$

The probability distribution of the state at t moment only depends on the state at $t-1$, which is called transition probability. The characteristics of the transition probability make HMM perform well on time series problems. For the gestures which are not included in the three gestures (wave, push, swing), we set the probability threshold α. When the probability is less than α, the gesture can be considered as unknown gesture.

D. Extraction of Hidden Gesture features

We utilize HMM to distinguish different types of gestures such as wave or push hand, however these gestures are not enough for authentication. The spoofer may imitate the gesture combination. Hence, we need to extract hidden features from



(a) 1-Dimension Fresnel zone (b) 2-Dimension Fresnel zone

Fig. 13. Fresnel Zone Model

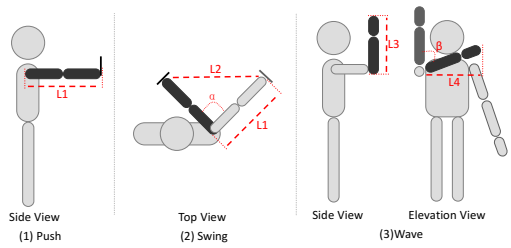


Fig. 14. Gestures and the hidden features

these activities. Although Shi et al. [20] have proposed the concept to extract the physical features to perform identification, the proposed approach is based on the neural network to extract the features which is highly depended on the training set and require unacceptable labor efforts to collect the labeled data. To this end, we proposed a high level feature extraction approach to extract the hidden features that are closely related to the authentication.

We mainly rely on the Fresnel Zone model [32] to extract the features, we can get these hidden features with the help of Fresnel model.

1) Modeling Activity with Fresnel zone: In the context of radio propagation, Fresnel zones refer to the concentric ellipses with foci in the transceivers. Assume P_1 and P_2 are transmitter and receiver respectively as shown in Fig. 13(a), for a given wavelength λ, the Fresnel zones containing n ellipses which can be constructed by the following equation:

$$|P_1Q_1| + |P_2Q_n| - |P_1P_2| = n\lambda/2 \tag{4}$$

where Q_n is the point in the n_{th} ellipse. According to the equation when the object is located in the odd number Fresnel zone, the reflected signal would enhance the receiving signal at P_2. However when the object is located at the even number Fresnel zone, the reflected signal and LOS signal would cancel each other. This means when an object moves from the $1st$ Fresnel zone to the nth Fresnel zone, the received signal would present peaks or valleys. Hence, we can calculate the number of passed Fresnel zone by observing the peaks and valleys of the waveform. Empirically, the object moves through one Fresnel zone at the distance of half wavelength. Above this, we can get the distance of moving object through 1-dimension Fresnel zone model. However, the angle cannot be obtained from 1-dimension Fresnel zone model. As the angle can be extracted if we have the same features from an orthogonal dimension, we set two receivers to form an orthogonal Fresnel zone to solve this problem. The 2-dimension Fresnel zone is illustrated in Fig. 13(b).

2) Inter-Gesture Interval/Gesture Duration: Regarding the Inter-gesture interval/gesture duration, assuming that we have a gesture combination $\langle act_1, inter_1, \cdots, act_n, inter_n \rangle$, then we can obtain the time series interval and gesture duration distribution denoted as $\langle T_{act_1}, T_{inter_1}, \cdots, T_{act_n}, T_{inter_n} \rangle$ in the process of gesture segmentation. Worth to mention that the sampling rate of our data is 1200 per second and we select 60 as the length of time window, thus the sampling gratuity of the measured time is 0.05 second.

3) Arm Length: As shown in Fig. 14, the push length is the sum length of forearm and upper arm. From Fig. 13(b), we find that the push length can be projected on the X axis and Y axis and the projection distance is the sum of quantities that pass through Fresnel zone. Hence, We can get the sum length of forearm and upper arm which is denoted as L_1 from pushing hand.

$$L_1 = \sqrt{x^2 + y^2} \tag{5}$$

For wave, the sum length of hand and forearm is denoted as L_3 and the projective distance is denoted as L_4. Hence, L_3 and L_4 satisfy the following equation:

$$L_4 = L_3 \sin\beta, \tag{6}$$

where β is the angle of wave. By 2-Dimension Fresnel zone, we can easily get L_4, but L_3 is unknown. From the equation 6, we can easily find when β equals $90°$, L_3 and L_4 is equal. However, in the actual experimental environment, we cannot guarantee the angle equals $90°$ exactly. Since L_3 is between 40cm to 60cm for most people which means generating 1cm distance error requires at least $11°$. This indicates the errors due to user behavior are negligible. Hence we can get L_3.

Need to mention that, when one may conduct incomplete gestures, or different gesture combination and consequence may lead to subtle different in different gestures. In our solution, we require the user to conduct the gestures in full stretch, so that we can measure the arm length without mislead information.

4) Move Range/Angle: We can extract the angle from swing gesture. Denote L_2 as the projective distance in 2-Dimension Fresnel zone and the angle α of swing can be expressed as the following equation:

$$\alpha = 2\arcsin\frac{L_2}{2L_1} \tag{7}$$

From the equation 7, we find that L_1 is the sum of the length of forearm and upper arm extracted from push hand and L_2 can be calculated by 2-Dimension Fresnel zone, hence we can easily get the angle.

5) Push Direction: In the 2-Dimension Fresnel zone, we can get the push direction as hidden features. As shown in Fig. 13(b), we can get the x component and the y component by 2-Dimension Fresnel zone. Hence, push direction can be expressed as the following equation:

$$\gamma = arctan\frac{y}{x} \tag{8}$$

E. Authentication Module

Through the above process, we can recognize the type of gestures (Wave, Swing, Push) through HMM which are used as the interaction gestures. For push hand gesture, we can get the sum length of forearm and upper arm and the direction of push hand. The angle and the sum length of forearm and hand are extracted from swing and wave. We can also get the time distribution from our gesture detection algorithm. These features can be seen as hidden features. For these features, we need to build users profiles. Suppose we have a gesture combination $\langle Push, Swing, Wave \rangle$, the feature vector for training can be expressed as $F = [T_{act_p}, T_{inter_p}, T_{act_s}, T_{inter_s}, T_{act_w}, T_{inter_w}, L_1, \gamma, \alpha, L_3]$. Before the interaction process, the users could conduct several times of the preferred gesture combinations for training and attach them to certain functions. We use traditional Support Vector Machine(SVM) with Gaussian kernel function to train the input samples. The output model and the gesture combination are set to be the user profile.

In addition, to tackle the shoulder surfing attack, we utilize one-class support vector machine to detect the spoofer. The spoofer is someone who does not exist in user profiles and tries to imitate the witnessed legitimate user's gesture combinations. For each legitimate user in the users profiles, we construct the one-class model using the feature vector F. Then, we get a score denoted as S which compares the similarity between the features denoted as F_{test} extracted from testing samples and the support vectors F_{user} of the user in the profiles. The equation can be expressed as the following equation:

$$S = \sum_{n=1}^{N} k(F_{test}, F_{user}) \tag{9}$$

where N is the number of legitimate users and $k()$ is the Gaussian kernel function. When S is larger, it implies the testing sample has less distance to the user in the profiles. We can set a threshold η to discover the possible spoofer and if S is less than η, it implies the testing sample is from spoofer.

V. Experimental Evaluation

A. Experimental Setup

The deployment of the system is mainly divided into two parts. We use MiWiFi as the signal transmitter and two Intel NUCs with Intel Link 5300 WiFi NIC as the signal receivers. The deployment is shown in Fig. 3 and has been described in subsection III-A.

Our experiments were conducted in the 5GHz frequency band with 20MHz bandwidth channels. On one hand, the band of 5GHz has a shorter wavelength (6cm in 5Ghz while 12cm in 2.4Ghz), which leads to better movement speed resolution. On the other hand, 5GHz has lower coherence rate than 2.4GHz according to the standard of IEEE 802.11a and IEEE 802.11b. The devices in 5Ghz is also much less than the ones in 2.4Ghz, thus leading to better SNR and CSI. All the experiments were conducted with 10 volunteers in a Lab. Unlike the data analysis experiment in section III-A, in these experiment, the participants are required to perform predefined gesture combinations in different locations. The gestures are performed at their own preferences except that the pushing has to be full stretching in order to measure the arm length. The collected data are processed to give following results.

Fig. 15. The result of gesture recognition).

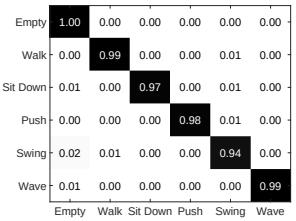


Fig. 16. The limitations of activity recognition

B. Evaluation Metrics

- Confusion Matrix: Each column in the confusion matrix indicates the ground truth of an identity/gesture and each row represents the classified identity/gesture in our system. Each entry in the matrix represents the percentage of correctly classified identify/gesture. Need to mention that, in order to keep enough details, we used the ceiling function to preserve the accuracy of the two decimal points. This may lead to the sum of the rows of the matrix exceed 100%.
- Authentication/Recognition Accuracy: The percentage of the human identity/gesture/features correctly recognized by our system over the total samples.
- Receiver Operating Characteristic (ROC): ROC curve shows the trade-off between the False Positive Rate and True Positive Rate under different values of threshold. The more the ROC curve hugs the point (0, 1), the better the performance is. The minimum distance between the point (0, 1) of ROC space and any point on ROC curve gives the optimal threshold.
- Area Under The Curve (AUC): The area under the ROC curve shows the performance of the learner, and represents the probability that the predicted positive cases are ahead of the negative ones. The larger the area is, the better performance of the learner is.
- Equal Error Rate (EER): This is used to depict the Shoulder Surfing attacker detection accuracy. It is the rate that the spoofer has successfully been treated as a legitimate user.

C. Gesture Recognition Accuracy

1) Generally Accuracy: The recognition accuracy for empty, walk, sit down, push, swing and wave is greater than 97%. Fig. 15, indicates that we can easily distinguish these gestures with each other and other interference activities and gestures. The control group only by accurately identifying these kinds of gestures can we extract hidden features from them and this is also the basis for our system to operate correctly.

2) Impact of Location: To evaluate the impact of the location to our recognition system, we mainly use the data in different location (P1, P2, P3, P4) to verify. The gestures

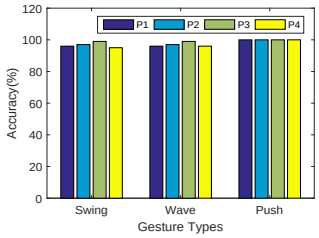


Fig. 17. Gesture Recognition Accuracy with vary Location.

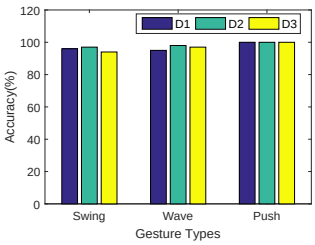


Fig. 18. Gesture Recognition Accuracy with vary Directions.

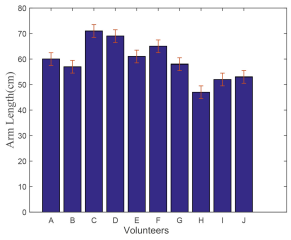


Fig. 19. The accuracy of arm length extraction

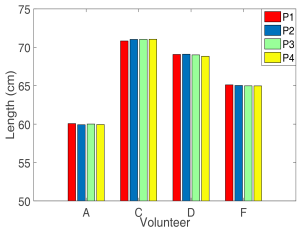


Fig. 20. The accuracy of arm length extraction vs. Location (P1-P4 stand for different location).

in this experiment include swing, wave and push, which is the candidate gestures of our authentication system. All the gestures are performed when facing towards the transmitter (D2 in Fig. 2). The results are summarized in Fig. 17, from which we can see that the location only slightly changed the recognition accuracy. When the location is distant from the Tx-Rx pair, it indeed impact the recognition accuracy[8]. But it could be mitigated by increasing the deployment density.

3) Impact of Direction: To evaluate the impact of the facing direction to our recognition system, we mainly use the data facing different location(D1, D2, D3) to verify. The gestures in this experiment also include swing, wave and push, all of which are performed in location P3. The results are summarized in Fig. 18, from which we can see that the location accuracy merely changed with the vary of direction. With the summarize of the results from Fig. 17 and Fig. 18, the gestures can be recognized effectively and stably. This is why we chose these gestures as part of our system.

D. The limitations of activity recognition model

To find whether we can directly use the type of activities for user identification, we conduct a simple experiment. Specifically, we collected the one third volunteers' activity samples for training and the others samples for the rest volunteers for testing. Basically, if only the gesture recognition could have variance in terms of identity, our experiments will results in low recognition accuracy. However, as is illustrated in Fig. 16, the model for the first third volunteers can also have a good performance on the remaining volunteers' samples. What is more, the average accuracy is near to 98%, which means activity type is not sufficient for identification and we need to use more detailed information such as interval, activity distance for identification. The types of activities have a lower degree of division on people identification.

E. Feature Extraction Accuracy

1) Arm length: The mean absolute error of the estimated arm length is within 3cm. We measure the arm length error size from all 10 volunteers which shown in Fig. 19. The mean absolute error is within 3cm which means the arm length we obtain from the Fresnel zone is relatively precise.

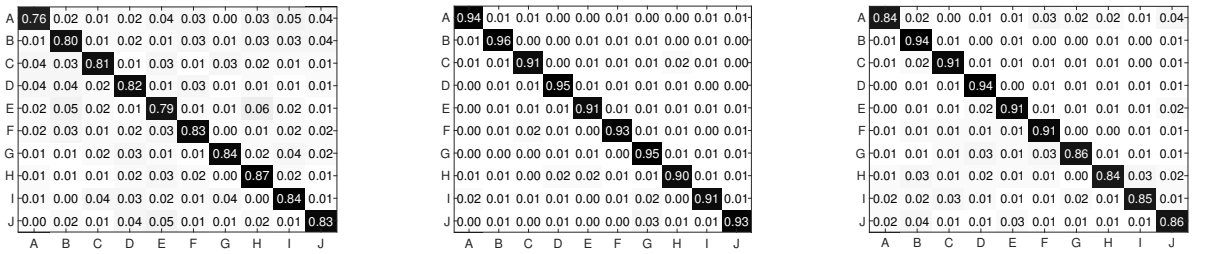
2) Angle/Direction: This variance of estimated angle is at most $15°\pm$*, while the average of the samples approximate the real value.*

The angle of swing or the push direction is derived through the modeling and extraction algorithm based on Fresnel zone based. Our experiment was conducted with 5 volunteers standing at the location P1 and pushing towards the router, which refers to the zero point of the plain and the push angle of 45 degree. Totally 50 samples were collected from each volunteer. The statistical results are shown in Fig. 8, indicating that the average of the angle is correctly extracted by our algorithm. However, the variance of the extracted results is still observable. This variance is at most $15°\pm$, which is acceptable for the deriving of the forearm length. But for identification, it has to work with other features together. By combing the results from Fig.8 and Fig. 19, we can also find an interesting result that the shorter the arm is, the larger the estimation variance of the arm length is.

3) Location Dependency: The location does not affect the accuracy of arm length estimation within certain range. The estimation of arm length is the base for estimation of angle/direction. Thus, we use the result of the arm length to depict the location dependency of the feature extraction. The experiments are taken by 4 volunteers pushing their hand in 4 different locations in Fig.3. The results are shown in Fig. 20 with the average arm length from the samples. We can see that the measured arm length vary in negligible amount (with in 3 cm). Thus, our system perform very robust in terms of location change for the feature extractions.

F. Authentication Accuracy

The EER values of 3 and 6 gestures are 14.5% and 9.5% respectively and the average authentication accuracy approaches 92.9% among ten volunteers. In this experiment, we require the volunteers to perform the motion sequences of push-swing-wave and push-swing-swing-wave-wave-push respectively. As shown in 21(a) and 21(b), in the first instance, the mean accuracy is above 82% and in the next instance, the mean accuracy approaches 92.9% among ten volunteers. Hence, we find that the accuracy is improved largely when the complexity of gesture sequence improved which conform to the actual situation. Fig. 21(c) plots the accuracy without the features of time distribution when the volunteers perform the same complex gesture sequence. The accuracy drops to 89% indicates that time distributions for different people vary greatly owing to the behavioral habit and this feature is indispensable for our secure interaction mechanism. Also, we use average authentication accuracy and ROC curve to evaluate the performance of the authentication system. First, we draw the ROC curves of three gestures and six gestures as shown in Fig. 22. It can be seen from the figure that the



(a) p-s-w (b) p-s-s-w-w-p (c) p-s-s-w-w-p without time distribution

Fig. 21. Authentication Result with different Gesture Combination

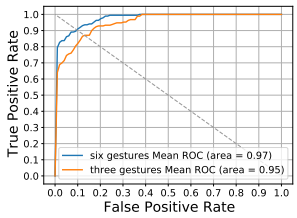


Fig. 22. The ROC of with different Gesture Combination

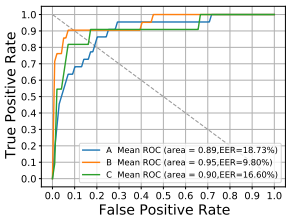


Fig. 23. The three gestures ROC of Shoulder surfing attack

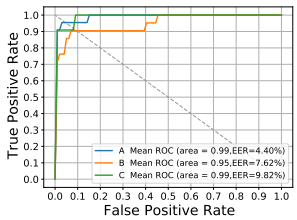


Fig. 24. The six gestures ROC of Shoulder surfing attack

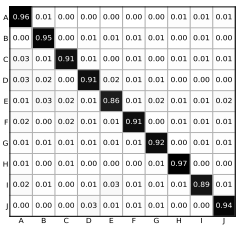


Fig. 25. The impact of the scale of dataset

AUC values of three actions and six actions are greater than 0.95 and the EER values of 3 and 6 gestures are 14.5% and 9.5% respectively, which indicates a good performance of our certification system.

G. System Response Time Evaluation

We evaluate the response time of the system. The program ran in a computer with i7 8700 CPU and 16G DDR3 RAM. All the codes are in form of MATLAB code. The time consumption of action recognition, hidden feature extraction and SVM is shown in Table I. In the time consumption of whole system, the time consumption of SVM is about 0.00085s, which can be ignored given hidden feature extraction is very time-consuming. The total time cost of the whole system is 2.08s. Worth to mention that, the above results were measured when the program was coded in MATLAB. The system delay could be greatly improved if all the code transformed into C/C++. Specifically, several the experimental reports regarding the execution efficiency of C/C++ and MATLAB, show that the performance difference between recursive code C++ and MATLAB is more than 500 times. In the normal running mode, the loop statement of MATLAB is much slower than that of C++ by about 4-10 times. Thus, if our system is implemented with C++, the system response time is will be 5-20 times less, which means 200ms or less with high probability. In such response time delay, the simultaneous interaction and authentication will be possible.

TABLE I
SYSTEM RESPONSE TIME

feature extraction	action recognition	SVM	Total Time(s)
1.71($\pm$0.27)	0.37($\pm$0.13)	0	2.08($\pm$0.4)

H. Shoulder Surfing Attack

The average EER of the attackers with 3 and 6 gestures combinations is 15.4% and 7.3% respectively. To examine the security of SiWi, we come up with a scenario that three volunteers act as a legitimate user performing 3 and 6 gestures combinations and the other 9 volunteers act as the attackers. The attackers are allowed to watch the video as many times as they wanted and then are requested to perform the sequence. In Fig. 23 and 24, we can see that the average EER of two group of gesture combinations are 15.4% and 7.3% respectively and the ROC curve hugs the point (0,1) more, which indicates the system has better performance on the spoofing detection.

I. The impact of dataset

SiWi attempts to use training dataset to construct a robust model for resisting shoulder surfing attack, the scale of training set is an important indicator of the generality of the model. We treat the samples of first three volunteers as training set and the others are test set. Meanwhile, we reduce the samples in each activity to 7 to check the performance. From the Fig. 25, we find that the average accuracy maintains around 90%. Hence, although we reduce samples, we can still get a state-of-art performance.

VI. CONCLUSION

We proposed a gesture-based secure interaction scheme, called SiWi, for the WiFi-enabled IoT environment, which could resist shoulder surfing attack. It is mainly realized by the robust recognition of the combination of the interactive gestures and the hidden/imperceptible features related to the user's identity. From extensive data collection and empirical study,

we identified three elemental gestures (push, swing, wave), which is easy to conduct and robust to recognize, and four hidden/imperceptible features (Inter-gesture interval/Gesture Duration, Distance/Arm Length, Swing Angle/Range and Push Direction), which have strong correlation with people's identity. We also proposed corresponding efficient and robust algorithms for gesture recognition and feature extraction. Experiments in real deployed systems were conducted with 10 volunteers, and the results showed that the authentication accuracy could reach 92.9% on average, while a small combination of our proposed gestures is strong enough (in percentage of 97%) to resist the shoulder surfing attacks.

REFERENCES

[1] A. Q. Mohammed and F. Li, "Wiger: Wifi-based gesture recognition system," *Isprs International Journal of Geo Information*, vol. 5, no. 6, p. 92, 2016.

[2] C. Wu, F. Zhang, Y. Fan, and K. J. R. Liu, "Rf-based inertial measurement," in *Proc. of ACM SIGCOMM*, 2019.

[3] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "Spotfi:decimeter level localization using wifi," in *Proc.of ACM SIGCOMM*, 2015, pp. 269-282.

[4] C. Han, K. Wu, Y. Wang, and L. M. Ni, "Wifall: Device-free fall detection by wireless networks," *IEEE Transactions on Mobile Computing*, vol. 16, pp. 271–279, 2017.

[5] H. Jiang, C. Cai, X. Ma, Y. Yang, and J. Liu, "Smart home based on wifi sensing: A survey," *IEEE Access*, pp. 1–1, 2018.

[6] L. Xie, X. Dong, W. Wang, and D. Huang, "Meta-activity recognition: A wearable approach for logic cognition-based activity sensing," in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, pp. 1–9, 2017.

[7] S. Zhang, X. Liu, Y. Liu, B. Ding, S. Guo, and J. Wang, "Accurate Respiration Monitoring for Mobile Users with Commercial RFID Devices," *IEEE Journal on Selected Areas in Communications*, 2020.

[8] W. Wang, A. X. Liu, M. Shahzad, K. Ling, and S. Lu, "Understanding and modeling of wifi signal based human activity recognition," in *Proc. of 21st Annual International Conference on Mobile Computing and Networking (Mobicom)*, pp. 65–76, 2015.

[9] L. Guo, X. Wen, Z. Lu, X. Shen, and Z. Han, "Wiroi: Spatial region of interest human sensing with commodity wifi," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, 2019.

[10] Q. Bu, G. Yang, X. Ming, T. Zhang, and J. Zhang, "Deep transfer learning for gesture recognition with wifi signals," *Personal and Ubiquitous Computing*, no. 3, 2020.

[11] T. Zhang, T. Song, D. Chen, T. Zhang, and J. Zhuang, "Wigrus: A wifi-based gesture recognition system using software defined radio," *IEEE Access*, vol. PP, no. 99, pp. 1–1, 2019.

[12] X. Liu, J. Yin, S. Zhang, B. Xiao, and B. Ou, "Time-Efficient Target Tags Information Collection in Large-scale RFID Systems," *IEEE Transactions on Mobile computing*, 2020.

[13] X. Liu, J. Yin, Y. Liu, S. Zhang, S. Guo, and K. Wang, "Vital signs monitoring with RFID: opportunities and challenges," *IEEE Network*, vol. 33, no. 4, pp. 126–132, 2019.

[14] L. Sun, S. Sen, D. Koutsonikolas, and K.-H. Kim, "Widraw: Enabling hands-free drawing in the air on commodity wifi devices," in *Proc. of 21st Annual International Conference on Mobile Computing and Networking (Mobicom)*, pp. 77–89, 2015.

[15] J. Brainard, A. Juels, R. L. Rivest, M. Szydlo, and M. Yung, "Fourth-factor authentication: somebody you know," in *Proc. of 13th ACM conference on Computer and communications security (CCS)*, pp. 168–178, 2006.

[16] G. D. Clark and J. Lindqvist, "Engineering gesture-based authentication systems," *IEEE Pervasive Computing*, vol. 14, no. 1, pp. 18–25, 2015.

[17] M. Shahzad, A. X. Liu, and A. Samuel, "Secure unlocking of mobile touch screen devices by simple gestures: you can see it but you can not do it," in *Proc. of 19th annual international conference on Mobile computing and networking (Mobicom)*, pp. 39–50, 2013.

[18] Y. Zeng, P. H. Pathak, and P. Mohapatra, "Wiwho: Wifi-based person identification in smart spaces," in *Proc. of 15th International Conference on Information Processing in Sensor Networks (IPSN)*, pp. 1–12, 2016.

[19] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using wifi signals," in *Proc. of 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*, pp. 363–373, 2016.

[20] C. Shi, J. Liu, H. Liu, and Y. Chen, "Smart user authentication through actuation of daily activities leveraging wifi-enabled iot," in *Proc. of 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing (Mobihoc)*, pp. 1–12, 2017.

[21] A. D. Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann, "Touch me once and i know it is you!: implicit authentication based on touch screen patterns," in *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 987–996, 2012.

[22] N. Sae-Bae, K. Ahmed, K. Isbister, and N. Memon, "Biometric-rich gestures: a novel approach to authentication on multi-touch device," in *Proc. of the SIGCHI Conference on Human Factors in Computing Systems (SIGCHI)*, pp. 977–986, 2012.

[23] L. Li, X. Zhao, and G. Xue, "Unobservable re-authentication for smartphones," *Ndss*, 2013.

[24] B. Shrestha, N. Saxena, and J. Harrison, "Wave-to-access: Protecting sensitive mobile device services via a hand waving gesture," in *Cryptology and Network Security*, pp. 199–217, 2013.

[25] L. Yang, Y. Guo, X. Ding, J. Han, Y. Liu, C. Wang, and C. Hu, "Unlocking smart phone through handwaving biometrics," in *IEEE Trans. Mobile Comput*, pp. 1044–1055, 2015.

[26] J. Shin, R. Islam, A. Rahim, and H.-J. Mun, "Arm movement activity based user authentication in p2p systems," *Peer-to-Peer Networking and Applications*, no. 3, 2019.

[27] D. C. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Tool release: Gathering 802.11n traces with channel state information," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 1, pp. 53–53, 2011.

[28] Z. Zhou, Z. Yang, C. Wu, L. Shangguan, and Y. Liu, "Towards omnidirectional passive human detection," in *Proc.of IEEE International Conference on Computer Communications (INFOCOM)*, 2013, pp. 3057-3065.

[29] B. Chen, V. Yenamandra, and K. Srinivasan, "Tracking keystrokes using wireless signals," in *Proc. of the 13th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys)*, 2015, pp. 31-44.

[30] P. Melgarejo, X. Zhang, P. Ramanathan, and D. Chu, "Leveraging directional antenna capabilities for fine-grained gesture recognition," in *Proc. of ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (UbiComp)*, 2014, pp. 541-551.

[31] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, "E-eyes: In-home device-free activity identification using fine-grained wifi signatures," in *Proc.of 20th annual international conference on Mobile computing and networking (MobiCom)*, 2014, pp. 617-628.

[32] D. Wu, D. Zhang, C. Xu, Y. Wang, and H. Wang, "Widir: walking direction estimation using wireless signals," in *Proc. of 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (Ubicomp)*, pp. 351–362, 2016.

[33] H. Li, W. Yang, J. Wang, Y. Xu, and L. Huang, "Wifinger:talk to your smart devices with finger-grained gesture," in *Proc. of ACM Conference on Ubiquitous Computing (Ubicomp)*, pp. 250–261, 2016.

[34] H. Abdelnasser, M. Youssef, and Khaled.A.Harras, "Wigest: A ubiquitous wifi-based gesture recognition system," in *Proc. of IEEE International Conference on Computer Communications (INFOCOM)*, pp. 1472–1480, 2015.

[35] N. Yu, W. Wang, A. X. Liu, and L. Kong, "Qgesture: Quantifying gesture distance and direction with wifi signals," in *Proc. of ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (UbiComp)*, 2018.

[36] H. Liu, Y. Wang, J. Liu, J. Yang, and Y. Chen, "Practical user authentication leveraging channel state information (csi)," in *Proc.of 9th ACM symposium on Information, computer and communications security (ASIA-CCS)*, 2014, pp. 389-400.

[37] Y. Li and T. Zhu, "Gait-based wi-fi signatures for privacy-preserving," in *Proc.of 11th ACM on Asia Conference on Computer and Communications Security (ASIA CCS)*, 2016, pp. 571-582.

[38] H. Kong, L. Lu, J. Yu, Y. Chen, L. Kong, and M. Li, "Fingerpass: Finger gesture-based continuous user authentication for smart homes using commodity wifi," in *Proc. of the Twentieth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, Mobihoc '19, (New York, NY, USA), p. 201–210, Association for Computing Machinery, 2019.

[39] M. Zhou, Q. Wang, J. Yang, Q. Li, F. Xiao, Z. Wang, and X. Chen, "Patternlistener: Cracking android pattern lock using acoustic signals," in *Proc. of ACM Conference on Computer and Communications Security (CCS)*, pp. 1775–1787, ACM, 2018.