Authenticating Users Through Fine-Grained Channel Information

Hongbo Liu, Yan Wang, Jian Liu, Jie Yang, *Member, IEEE*, Yingying Chen, and H. Vincent Poor, *Fellow, IEEE*

Abstract—User authentication is the critical first step in detecting identity-based attacks and preventing subsequent malicious attacks. However, the increasingly dynamic mobile environments make it harder to always apply cryptographic-based methods for user authentication due to their infrastructural and key management overhead. Exploiting non-cryptographic based techniques grounded on physical layer properties to perform user authentication appears promising. In this work, the use of channel state information (CSI), which is available from off-the-shelf WiFi devices, to perform fine-grained user authentication is explored. Particularly, a user-authentication framework that can work with both stationary and mobile users is proposed. When the user is stationary, the proposed framework builds a user profile for user authentication that is resilient to the presence of a spoofer. The proposed machine learning based user-authentication techniques can distinguish between two users even when they possess similar signal fingerprints and detect the existence of a spoofer. When the user is mobile, it is proposed to detect the presence of a spoofer by examining the temporal correlation of CSI measurements. Both office building and apartment environments show that the proposed framework can filter out signal outliers and achieve higher authentication accuracy compared with existing approaches using received signal strength (RSS).

Index Terms—User authentication, channel state information, wireless networks

1 Introduction

The rapid advancement of wireless technologies has made wireless networks ubiquitous and thus network services can be accessed at anytime and anywhere. However, securing wireless networks is challenging due to the shared nature of the wireless medium, as adversaries can eavesdrop upon or intercept any wireless transmission [16]. For example, an adversary can passively monitor wireless networks to obtain valid device identities and further launch identity-based attacks, which serves as a basis for launching a variety of malicious attacks across multiple network layers [6]. Indeed, such identity-based attacks are easy to launch in WiFi networks, where the Access Points (APs) can be spoofed, resulting in Denial of Service (i.e., a rogue AP attack) [32]. Although existing cryptographic based authentication techniques (such as WiFi Protected Access and 802.11i) can protect data frames, an attacker can still spoof the 802.11 management frames [23]. In addition, the increasingly dynamic mobile environments make it harder to utilize cryptographic-based authentication, due to its infrastructural and key management overhead [2], [5], [7].

- *H. Liu is with the Department of Computer Information and Graphics Technology, Indiana University-Purdue University Indianapolis, Indianapolis, IN 46222. E-mail: hl45@iupui.edu.*
- *Y. Wang, J. Liu, and Y. Chen are with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030. E-mail: {ywang48, jliu28, yingying.chen}@stevens.edu.*
- *J. Yang is with the Department of Computer Science, Florida State University, Tallahassee, FL 32306. E-mail: jyang5@fsu.edu.*
- *H.V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544. E-mail: poor@princeton.edu.*

Manuscript received 20 Dec. 2015; revised 6 Dec. 2016; accepted 14 Jan. 2017. Date of publication 22 June 2017; date of current version 5 Jan. 2018. (Corresponding author: Hongbo Liu.)
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TMC.2017.2718540

Recently authentication based on non-cryptographic methods has been proposed to complement and enhance the existing cryptography based schemes [3], [6], [10]. For example, channel based authentication schemes use the Received Signal Strength (RSS) of wireless packets or the Channel Impulse Response (CIR) of a single frequency to generate fingerprints of the wireless channel to perform user authentication [6], [22]. The rationale behind these schemes is that both RSS and CIR present unique spatial properties due to path loss and multi-path effects. An adversary, situated at a different location from the legitimate user, will incur different RSS or CIR profiles. However, the RSS and CIR extracted from a single frequency only provide coarse-grained information about the wireless channel and thus the effectiveness of user authentication is largely limited. For example, RSS-based authentication can hardly distinguish between two users with similar signal signatures even though they may be located far away from each other [6].

In this paper, we exploit the fine-grained physical layer information made available from orthogonal frequency-division multiplexing (OFDM) to perform user authentication. The channel response from multiple subcarriers of OFDM provides detailed Channel Sate Information (CSI) [11], which can be used to detect and differentiate minute changes in the wireless channel [28]. Measuring the channel frequency response thus is an ideal candidate for achieving accurate user authentication leveraging temporal and spatial properties of the wireless channel. Specifically, we show that CSI can be utilized to accurately authenticate users with similar signal fingerprints and differentiate a legitimate user from a spoofing attacker. The detailed channel information has the power to enable user authentication at a per packet level, making it a promising utility to achieve user authentication at a much higher granularity (in both spatial and temporal domains) than existing channel-based (i.e., RSS

and CIR) approaches. Further, CIR can only be obtained on dedicated hardware (e.g., Universal Software Radio Peripherals (USRPs)), which prevents it from being widely adopted in real-world scenarios. In this work, we conduct user authentication by associating each individual user with his own wireless device, which is not accessible by other users. Each wireless device represents a distinct user in the network. Thus user authentication can be performed by examining the channel state information of the associated wireless device.

CSITE [14] applies a sliding-window based technique to CSI measurements to build the user profile for authentication purposes. CSITE assumes that the CSI collection is benign (without the presence of an identity-based attacker) when building the user profile. However, in practice an identity based attacker could be present at any time. Thus, the CSI measurements could be a mixture of readings from both the legitimate user and a spoofer, leading to a misclassification of the user profile and falsely authenticating the spoofer. To tackle such challenges arising in real-world scenarios, we study how to construct the user profile even when a spoofer is present and perform robust user authentication under various adversarial scenarios, e.g., when the legitimate user is not present but the spoofer is active. In particular, we propose a user authentication framework that can work with both stationary and mobile users. For the stationary scenario, the proposed framework builds the user profile for user authentication and is resilient to the presence of the spoofer. It involves two main components: an *Attack-resilient Profile Builder* and a *Profile Matching Authenticator*. The *Attack-resilient Profile Builder* has the capability to accurately construct the user profile of the legitimate user even when a spoofing attacker is present. We further develop a *Profile Matching Authenticator* grounded in machine-learning based techniques to perform robust per-packet user authentication in real-time based on CSI measurements. In addition, we are among the first to study the effect of different modulation and coding scheme rates on CSI to achieve accurate user authentication. For mobile scenarios, we propose to detect the presence of a spoofer by examining the temporal correlation of CSI measurements. This involves two main components: a *Correlation Analyzer* and a *Correlation-based Authenticator*. The *Correlation Analyzer* calculates the temporal correlation between adjacent CSI measurements within a coherence time period, whereas the *Correlation-based Authenticator* authenticates the mobile user in real-time based on the temporal correlation between adjacent CSI measurements.

We summarize the main contributions of our work as follows:

- We show that it is feasible to perform user authentication by utilizing CSI from OFDM even when the users possess similar signal fingerprints, making fine-grained user authentication achievable in practice.
- We develop a user authentication framework that works with both stationary and mobile users. To deal with a stationary user, the proposed framework has the capability of building a user profile under the presence of a spoofing attack and achieves higher authentication accuracy compared with existing channel based (e.g., RSS-based) methods. For mobile users, the framework performs real-time user authentication by leveraging temporal correlation of CSI measurements.
- We validate the framework by conducting real experiments in both office and apartment environments using off-the-shelf WiFi devices. Experimental results confirm that our framework is highly robust and effective in user authentication under various attacking scenarios without requiring any additional overhead on wireless devices.

The rest of the paper is organized as follows. In Section 2, we put our work in the context of related work. The attack model and our framework are described in Section 3, and the feasibility of using CSI to perform user authentication is presented in Section 4. In Section 5, we detail the proposed Attack-resilient Profile Builder based on clustering analysis. The Profile Matching Authenticator grounded on machine learning techniques is described in Section 6. Section 7 depicts the correlation-based user authentication algorithm for mobile scenarios. We discuss the experimental setup and methodology, and further present the performance evaluation results of our proposed CSI-based authentication framework in both office and apartment environments in Section 8. Finally, we conclude our work in Section 9.

2 Related Work

The traditional approach to providing user authentication is to use cryptographic-based authentication. For example, Wu et al. [30] have introduced a secure and efficient key management (SEKM) framework. SEKM builds a Public Key Infrastructure (PKI) by applying a secret sharing scheme and an underlying multicast server group. Wool [29] implements a key management mechanism with periodic key refresh and host revocation to prevent the compromise of authentication keys. The application of cryptographic authentication requires reliable key distribution, management, and maintenance mechanisms, which reduces its usability in a dynamic mobile wireless environment (i.e., due to a lack of a fixed key management infrastructure) or resource-constrained wireless networks (i.e., due to limited resources on wireless devices).

Recently non-cryptographic based authentication has drawn considerable attention [33]. In general, non-cryptographic solutions can be categorized into four groups: software based, hardware based, biometric and physical-trait based, and channel-signature based. Software based authentication basically relies on the unique characteristics of the software programs or protocols running on the devices [10], [26], whereas hardware based authentication leverages unique hardware traits such as channel-invariant radiometric signatures [3], [24] and clock skews [13], [18] to identify users. Biometric and physical-trait based authentication relies on the behavioral modalities including on-screen touch and finger movement patterns [9], [20]. Channel-signature based authentication schemes have been proposed to use either Received Signal Strength [6], [15], [31], [32], [34] or Channel Impulse Response [22], [27] to identify users. The major advantage of using channel signatures is that it exploits the naturally available random and location-distinct characteristics of the wireless channel, which are very hard to falsify, for user authentication.

For channel based user authentication using RSS, a series of approaches [6], [31], [32] have been proposed to detect identity-based attacks, determine the number of attackers when multiple adversaries are masquerading as the same node identity, and localize the adversaries. Reciprocal Channel Variation-based Identification (RCVI) [34] exploits

the reciprocity of RSS variance to decide if all packets come from a single or more than one sender. Ensemble [15] leverages a user's growing collection of trusted devices that analyze variations in RSS to determine whether the pairing devices are in physical proximity to each other. It is important to note that although RSS is available on current wireless devices, RSS is known to be sensitive to multipath effects and affected by the transmission power level. As a result, a legitimate user may be mistakenly regarded as a malicious user due to the inherent RSS variation. Different from RSS which is readily available in the existing wireless infrastructure, CIR is usually extracted from specialized devices such as Field-Programmable Gate Arrays (FPGAs) [27] and USRPs [22], which limit its practicality in real-world scenarios.

Different from the previous work, we propose to use Channel State Information, fine-grained channel information readily available from current commercial hardware (i.e., 802.11 a/g/n devices), which represents both amplitude and phase for each subcarrier on the 802.11 a/g/n OFDM system. Exploiting CSI has the potential to achieve much higher granularity (in both spatial and temporal dimensions) for user authentication than applying existing channel based (i.e., RSS and CIR) authentication methods. The work most related to ours is CSITE [14], which utilizes CSI magnitude measurements averaged over time to generate profiles for legitimate users. Jiang et al. [14] assumes the CSI collected over time is benign and there is no identity-based spoofing attack present when building the profile. However, in practice a spoofing attack could be present at any time. Thus, the profiles built under such attacks may not represent legitimate users and may lead to false authentication of malicious users. In our work, we develop an Attack-resilient Profile Builder, which has the ability to detect the presence of spoofing attacks when building profiles for legitimate users. Furthermore, we study the effect of different modulation and coding scheme rates on CSI to achieve a higher accuracy of user authentication under both single antenna and multiple antenna cases.

3 Attack Model and System Overview

In this section we first introduce the attack model we consider in this work. We then present the flow of our proposed CSI-based user-authentication framework.

3.1 Attack Model

User authentication is a technique of confirming the identity of a user. Based on the user authentication result, a system can determine whether a user is allowed to access certain restricted services, such as restricted access of certain web sites and enterprise data retrieval [25]. User authentication is particularly challenging in wireless networks as it is very hard, if not impossible, to physically confirm the truth of a user's identity due to the open nature of the wireless medium. In our user-authentication framework, we focus on the identity-based attack, in which an adversary can collect a legitimate user's identity and then masquerade as the legitimate user to pass the user authentication process [6]. The identity-based attack is very harmful as, once having passed the user authentication, the adversary can gain certain access privileges and further launch a variety of malicious attacks. For example, an adversary could easily obtain the Medium Access Control (MAC) address of a legitimate WiFi device by passively monitoring the wireless traffic and then impersonate the legitimate device by changing its MAC address. Another example is that by masquerading as an authorized wireless AP or an authorized client, an attacker could launch a variety of attacks including session hijacking, denial-of-service (DoS) attacks, or falsely advertising services to wireless clients [32].

In this work, we assume that an identity-based attack can be present at any time. That is, unlike the previous work, which only considers the presence of such an attack during the authentication phase, we take the viewpoint that identity-based attacks could be present at any time in real-world scenarios even when building profiles for legitimate users. Once such an attack is present in the network, the adversary spoofs the legitimate user's device identity (e.g., a WiFi device's MAC address) to send out packets. Once the attacker obtains the legitimate user's device identity, it can access the network with or without the presence of the legitimate user. Furthermore, the spoofer can be either stationary or mobile, whereas the legitimate device is mostly placed at a fixed position but could be moved from one location to another (e.g., the user walks from an office to a meeting room). The movement of the device can be detected using existing techniques [4], [17], [19] (e.g., examining the variance of the wireless signal). We assume that the attacker does not have the capability to capture and replay the CSI, and thus the attacker cannot alter or jam signals transmitted by the legitimate user.

3.2 System Overview

We consider both stationary and mobile users. The mobility of a user can be detected by using existing techniques [4], [17], [19]. For a stationary user, the basic idea is to profile the user by exploiting the readily available fine-grained CSI extracted from orthogonal frequency-division multiplexing based wireless networks, such as 802.11 a/g/n networks. CSI reveals the wireless channel response depicting the amplitudes and phases of every OFDM subcarrier. In general, CSI measurements from each user present a unique pattern corresponding to the wireless communication channel. Such CSI patterns can be extracted and utilized to uniquely identify each user. If the observed wireless packet (from a wireless device) contains a different CSI pattern from the legitimate profile, the network will raise an alert indicating a possible identity-based attack and the user authentication fails on that device. Particularly, when both a legitimate user and spoofer are present, the proposed system should be able to know which one is the legitimate user given its pre-constructed CSI profile. For mobile users, we propose to detect the presence of a spoofer by examining the temporal correlation of CSI measurements. The adjacent CSI measurements from the same user within a coherence time show strong correlation, while CSI measurements from different users are uncorrelated due to spatial and temporal diversities. Such a correlation pattern could be used to determine whether an identity-based attack exists. In mobile scenarios, we note that when both a legitimate user and spoofer are present, the proposed system can detect the presence of the spoofer but cannot differentiate which CSI measurements are from the legitimate user.

As noted above, our proposed user authentication framework, depicted in Fig. 1, consists of four components covering both the stationary and mobile scenarios: Attack-resilient Profile Builder, Profile Matching Authenticator, Temporal Correlation Analyzer and Correlation-based Detector. The first two components deal with stationary users, and the last two are

Fig. 1. Overview of the CSI-based user authentication framework.

for mobile users. The network implementing this framework will keep monitoring the wireless traffic and examining CSI measurements from each packet based on the device's identity.

3.2.1 Components for Stationary User Authentication

Attack-Resilient Profile Builder. The novelty of our profile builder is that it has the capability of building the actual user profile under the presence of a spoofer. When building the user profile for a specific user identity (ID), the presence of the spoofer will cause the CSI measurements collected from this ID to contain the mixture signal information from both the legitimate user and spoofer. As a result, the profile built under such a scenario is thus undermined by the spoofer, leading to mistakenly authenticating the spoofer or denying the legitimate user's access. We assume that the legitimate user communicates first in the network to build its CSI profile. Even if the spoofer learns the legitimate user's ID quickly and comes online right after the legitimate user communicates, there will be two CSI clusters formed. Next, our profile builder employing clustering analysis separates the CSI measurements of the legitimate user from the ones of the spoofer and determines the presence of the spoofer. Thus, the CSI profile will not be built in this case, and the legitimacy of the user profile construction can thus be ensured.

Furthermore, when the legitimate user moves from a current location to another, e.g., from an office to a meeting room, our framework can adaptively rebuild the user's profile. This rebuilding procedure can be user triggered or triggered after detecting the user movement based on existing techniques [4], [17], [19].

Profile Matching Authenticator. This component examines the real-time CSI measurements per packet from a device ID and performs user authentication by performing user profile matching. It is grounded in machine-learning based techniques and raises an alert if the profile matching fails. Our authenticator aims to achieve fine-grained user authentication as it can work at a per packet level – authenticating each packet of the device ID. It is capable of authenticating different users even when they possess similar signal fingerprints due to the complex environments arising in real systems. The authenticator works well under both single-antenna and multiple-antennas cases.

3.2.2 Components for Mobile User Authentication

Temporal Correlation Analyzer. In this component, we aim to analyze the correlation of the CSI measurements collected from the same device ID. Particularly, the Pearson correlation coefficient is used to indicate the correlation between any two adjacent CSI measurements. Further, we filter out neighboring CSI measurements that are not within a coherence time period due to various factors (i.e., traffic collisions, varying transmission rates, etc.).

Correlation-Based Detector. Unlike stationary users, mobile users do not have a fixed CSI profile to match for authentication. Instead, we examine the correlation between two adjacent CSI measurements for user authentication. In particular, the temporal correlation between adjacent CSI measurements of the same user within a coherence time period should be very high. However, the correlation should be low (i.e., uncorrelated) if two adjacent CSI measurements come from two different users (i.e., in the presence of spoofing attack) due to spatial diversity of wireless channels. Therefore, we examine the correlation between any two adjacent CSI measurements within a particular segment (i.e., time window) for user authentication. If the correlation of these adjacent CSI measurements falls below a predefined threshold, the presence of a spoofing attack is declared. This threshold is determined empirically in our experiments.

4 Feasibility Study of CSI-Based User Authentication

In this section, we first provide the background on CSI measured from OFDM subcarriers. We then discuss the feasibility of using CSI for user authentication. We next present our data pre-processing techniques applied to CSI measurements for more reliable user authentication.

4.1 Preliminaries

Our authentication framework exploits the CSI measured from OFDM subcarriers, a reliable and fine-grained description of channel characteristics. OFDM is widely used in wireless communication systems to improve the communication performance by utilizing the frequency diversity of wireless channels. For example, OFDM is used in popular wireless networks including IEEE $802.11a/g/n$, WiMAX, $4G$ and Digital Subscriber Line (DSL). OFDM is a method of encoding data streams on multiple carrier frequencies. In particular, Data in OFDM is split into multiple streams, which are coded and modulated respectively into different subcarriers. The frequency of each subcarrier is designed to be orthogonal to other subcarriers, so that the interference during transmission is minimized. For example, for the OFDM employed by the $802.11a/g/n$ physical layer, a relatively wideband

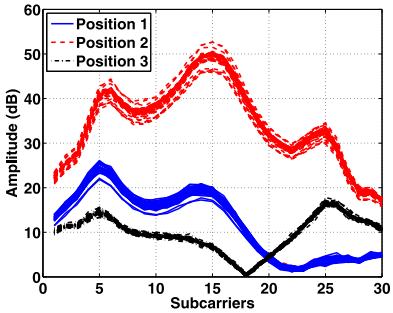


Fig. 2. Example of channel state information for OFDM that is collected at three different positions.

channel (or carrier) with 20 or 40 MHz is partitioned into 54 or 108 subcarriers for data transmission, so that each subcarrier can be used as a narrowband channel. This inspires us to exploit the channel state information extracted from OFDM subcarriers for user authentication, which can provide a fine granularity of the channel information and has the potential to achieve higher accuracy for user authentication in practice. Fig. 2 depicts the amplitude of typical channel state information across 30 subcarrier groups at three different positions. For each position, the CSI of 50 packets is measured from an Intel WiFi 5,300 card in a laptop.

4.2 Feasibility Study

To be able to use CSI for user authentication, the measured CSI from different devices should satisfy the *uniqueness* requirement. That is, the CSI measured at different devices situated at different locations should be distinct, while the CSI collected from different packets emitted by the same device should be similar, if not identical. We observe in Fig. 2 that the amplitude of CSI at different subcarriers is different due to frequency diversity. Furthermore, the CSI shapes from these three devices at different locations are distinct. This is because the CSI is the reflection of the complicated wireless channel and is affected by the wireless environment due to reflection, refraction, shadowing, etc. The CSI decorrelates with location rapidly. If two users are located at different locations, their CSI profiles should differ significantly. Additionally, we observe that the CSI of multiple packets from the same device at a fixed location exhibits the same trend, which indicates that a unique profile could be built for each user and can serve as the basis for user authentication.

Note that compared to the RSS, which only provides overall received power for each packet, CSI provides fine-grained channel information, i.e., channel responses on multiple subcarriers. Therefore, instead of deploying multiple landmarks or monitors to collect multi-dimensional RSS readings for user authentication purposes, a single monitor can provide multi-dimensional channel state information sufficient for user authentication. Furthermore, since the widely adopted IEEE 802.11 n standard [1] already defines a mechanism to exchange detailed CSI between a pair of wireless devices, employing CSI as a unique means for user authentication will not involve extra communication cost for existing WiFi networks.

Temporal Correlation. We conduct experiments to study the temporal correlation between CSI measurements. The CSI measurements are simultaneously collected from two mobile users, u and u', which move at the normal speed in an indoor environment. We observe that the adjacent CSI measurements from the same device exhibit high correlation with each other. As shown in Fig. 4a, generally the correlation coefficients between two CSI measurements $C_u(k)$ and $C_u(k')$ within a coherence time (e.g., $\|k' - k\| \leq 0.025$ sec) are close to 1. On the other hand, the CSI measurements, $C_u(k)$ and $C_{u'}(k')$ from two different devices, even if the time difference $\|k' - k\|$ is less than the coherence time, have correlation coefficients spanning a large range due to spatial diversity as shown in Fig. 4b. Based on the above observations, the temporal correlation between two adjacent CSI measurements can be used to detect a mobile spoofer and thereby authenticate a mobile user.

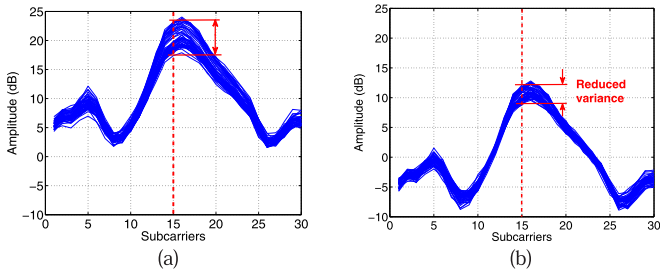


Fig. 3. CSI samples before and after data processing.

Data Preprocessing. In our study, we observe that the mean amplitude value of CSI measurements may shift over time. We call such a mean value shift as a *temporal bias*, and it will result in inaccurate CSI profile construction for user authentication. Therefore, our framework develops a data preprocessing strategy to cope with CSI samples to mitigate the effects caused by such temporal bias.

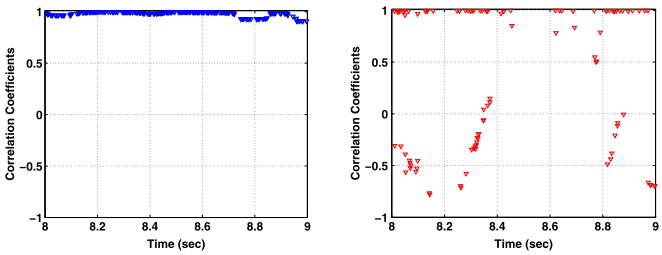
In particular, we observe a shift in the amplitude of a specific subcarrier due to the interference presented at either transmitter or receiver. Fig. 3a plots a typical curve of the CSI sample in a packet and many curves are collected over time. It shows that the amplitude of each subcarrier in CSI samples varies over time. Our data preprocessing strategy adjusts the mean value of the CSI sample (from a specific packet) to zero. This helps to reduce the overall variance of CSI measurements across subcarriers before performing user authentication. To illustrate, we denote the raw CSI sample per packet from a particular user u as a k-dimensional vector C_u, and the preprocessed CSI sample can be obtained as

$$C_u = C_u - 1_{1\times K}\frac{1}{K}\sum_{k=1}^{K} C_u(k), \tag{1}$$

where K is the number of subcarriers within a single CSI sample, and $1_{1\times K}$ is a K-length all-one vector. After applying the data preprocessing strategy, the updated CSI samples will have smaller variance and reduced amplitude bias on each subcarrier as shown in Fig. 3b. In addition, the wireless devices in an 802.11 n network are usually equipped with multiple-antennas. Thus, the CSI samples collected from each channel between the transmitting antenna i and receiving antenna j of two communicating devices will go through the pre-process as shown in Equation (1), where C_u will be replaced by $C_u^{i,j}$.

5 Attack-Resilient User Profile Builder

In this section, we describe the attack-resilient profile builder which employs clustering analysis on CSI measurements to determine whether the network environment is benign or a spoofer is present when constructing the user profile.



(a) Single user. (b) 2 Users of the same ID.

Fig. 4. Temporal correlation coefficients of consecutive CSI measurements.

5.1 Basic Idea

Since the spoofing attack could be present at any time, we need to determine whether a spoofer is present when constructing a user profile. Our attack-resilient profile builder aims to ensure the legitimacy of the user profiles even under a malicious wireless environment. The rationale behind our attack-resilient profile builder is that the CSI measurements of each device present unique spatial characteristics: the CSI has strong spatial correlation with the device's location. Although the wireless channel may fluctuate over time, the CSI of wireless packets from one device at a fixed location should be clustered together in the multi-dimensional signal space constructed by CSI measurements. For example, the 30 subcarriers obtained in our experiments can form a 30-dimensional CSI space, and the amplitudes of the CSI from multiple packets in Position 1 are clustered together (i.e., has a constant shape) in CSI space as shown in Fig. 2. Furthermore, the CSI measurements of the wireless packets collected from another device situated at a different location (Position 2) should form a different cluster in the CSI space. Thus, when the environment is benign, the CSI measurements from a particular device should be clustered together and form one cluster in the CSI space, while under a spoofing attack, the spoofer utilizes the same device identity as the legitimate user to transmit packets, and the CSI readings of the device identity are a mixture of readings from both the legitimate user and the spoofer, resulting in more than one CSI cluster.

To determine whether the network environment is benign, our framework applies clustering analysis to partition the CSI from one device identity into two clusters. Under normal conditions without spoofing, the distance between the partitioned two CSI clusters should be small since there is basically only one cluster from a single device at a physical location. However, under a spoofing attack, there is more than one device at different physical locations claiming the same device identity. As a result, more than one CSI cluster will be formed in the CSI space. Therefore, the distance between two partitioned clusters will be large as the cluster centers are derived from the different CSI clusters associated with different locations in physical space. Therefore, by examining the distance between the two partitioned CSI clusters, any presence of the spoofing attack can be determined when building user profiles. The flow of the Attack-resilient Profile Builder is shown in Fig. 5. Only when there is no spoofing attack present, will the profile of the legitimate user be built.

5.2 Algorithm Description

5.2.1 Modulation and Coding Scheme Study

WiFi devices usually use a fixed range of modulation and coding scheme (MCS) for data transmission. We find in our

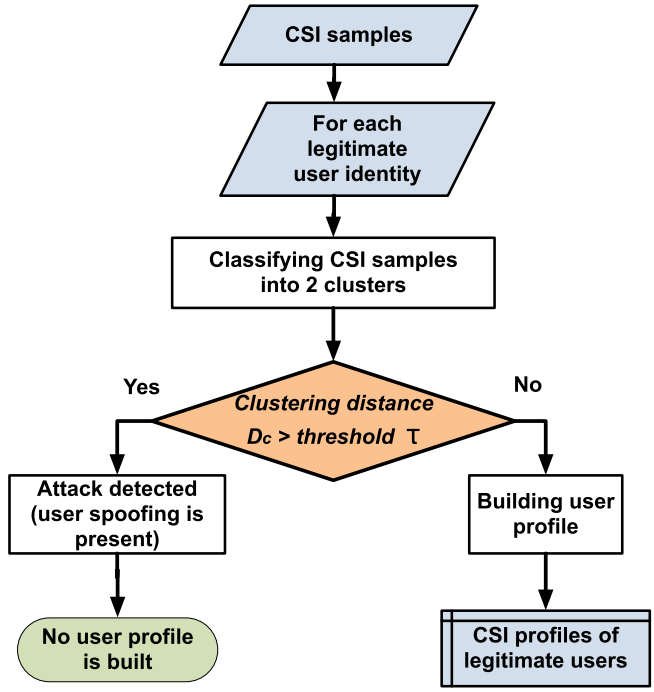
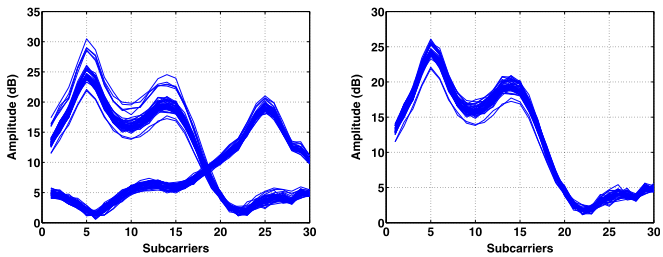


Fig. 5. Work flow for the attack-resilient profile builder.



(a) (b)

Fig. 6. CSI samples before and after filtering based on MCS rate.

experiments that the modulation and coding scheme occasionally changes to a different one and then switches back due to the variation of the wireless channel. And the occasionally changed modulation and coding scheme creates outliers in the CSI measurements. Thus, our framework first performs outlier filtering based on the modulation and coding scheme used for packet transmission before conducting clustering analysis. In particular, MCS is a specification of the high-throughput (HT) physical layer (PHY) parameter in the 802.11 n standard [1]. It contains the information of the modulation order (e.g., BPSK, QPSK, 16-QAM, 64-QAM), the forward error correction (FEC) coding rate, etc. Each 802.11 n packet header (in the 2.4 GHz band) contains a 16-bit MCS, which can be extracted together with the CSI sample of each packet.

Fig. 6a shows the raw CSI measurements for a wireless device with two clusters formed in our experiments. Under such cases, the MCS rate is changing according to the channel conditions, and we can observe CSI samples resulting from different MCS rates. For these cases we find the MCS values are greater than 263, different from most of the other test cases in both the lab and apartment environments. We thus filter out CSI for the packets with MCS values greater than 263, which corresponds to a single spatial stream with transmission rate 60 Mbps [1]. After filtering out these outliers, the CSI coming from the rest of the

packets exhibits a similar shape (i.e., forms one cluster in the CSI space) as shown in Fig. 6b.

5.2.2 Clustering Analysis

We utilize the K-means algorithm to partition the filtered CSI measurements from the device identity u into two clusters. The K-means algorithm is one of the most popular iterative descent clustering methods [12]. The squared Euclidean distance is chosen as the dissimilarity measure. If there are S CSI samples from the device u, the K-means clustering algorithm partitions S CSI samples into K disjoint subsets L_k containing S_k sample points so as to minimize the sum-of-squares criterion

$$J_{min} = \sum_{k=1}^{K} \sum_{C_{u,s} \in L_k} \|C_{u,s} - \mu_{\mathbf{k}}\|^2, \tag{2}$$

where $C_{u,s}$ is a CSI vector representing the CSI value for the sth packet and $\mu_{\mathbf{k}}$ is the geometric centroid of the sample points for L_k in CSI space. In our cluster analysis, we choose $K = 2$. We further choose the distance between two centroids as the test statistic $\mathbf{T}$ for identity-based attack detection,

$$D_c = ||\mu_{\mathbf{k}} - \mu_{\mathbf{k}'}||, \tag{3}$$

with $k, k' \in \{1, 2\}$.

Under normal conditions in a benign network environment, the distance between the centroids from the K-means cluster analysis in CSI space should be close to each other, because there is only one cluster from a single device u at a physical location. However, when a spoofer is present, there is more than one device situated at different physical locations, claiming the same device identity. The distance between two partitioned CSI clusters thus will be large. Through the analysis above, we show that the clustering method has the capability of detecting the presence of the spoofer by applying the threshold τ to D_c as follows:

$$\begin{cases} D_c > \tau & \text{attacker is present;} \\ D_c \leq \tau & \text{normal condition.} \end{cases} \tag{4}$$

5.2.3 User Profile Building

If the CSI samples are collected in a benign environment, the framework deposits the pre-processed CSI samples, C_u, as the profile for user u for future profile matching based authentication. We note that the user profile only requires a small number of packets, e.g., less than 100 packets.

If the user moves from one location to another (e.g., walks from an office to a meeting room), the user authentication framework will adaptively rebuild the user's profile. The following are two possible ways to update the user's profile: 1) the user can actively trigger the profile updating after moving to a new place; or 2) the profile updating can be triggered by detecting the user movement using existing techniques.

6 User Authentication Leveraging Profile Matching

In this section, we present our profile matching authenticator, which uses machine-learning based methods for packet-level user authentication.

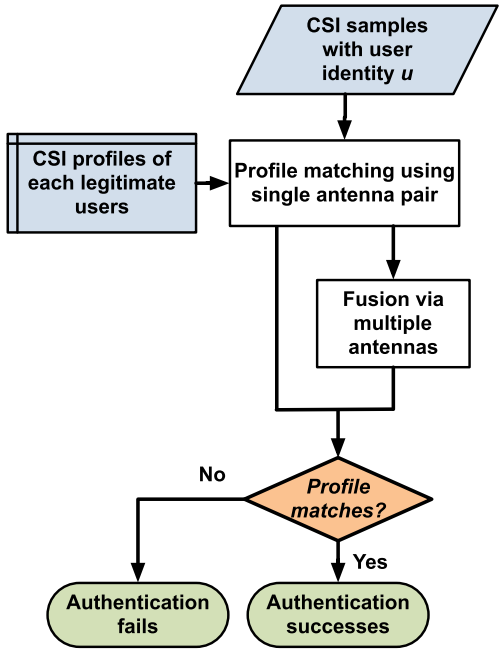


Fig. 7. Work flow for the profile matching authenticator.

6.1 Basic Idea

The basic idea of our profile matching authenticator is to use machine learning to determine whether the CSI measurement for the incoming packet with the user identity u matches the profile constructed by the profile builder. If the incoming CSI sample matches the user profile, the corresponding packet can be authenticated successfully as being from user u. Otherwise, the user authentication fails. Fig. 7 illustrates the work flow of our profile matching authenticator. In particular, the profile matching scheme works at the packet level, which minimizes the latency of the authentication process. In addition, the packet-level authentication can also be used to monitor and count the number of packets injected by an attacker.

6.2 Approach Description

We next present the profile matching method using the CSI samples from a single antenna. We then present the profile matching using CSI samples from multiple antennas to improve the performance of user authentication.

6.2.1 Profile Matching Using a Single Antenna Pair

We perform the profile matching via the support vector machine (SVM) technique, which is a computationally efficient way of learning good separating hyperplanes in a high dimensional feature space. The CSI samples are used as features in SVM to perform profile matching for each user. We first study the case using a single antenna pair for profile matching.

In this study, we consider the profile matching as a two-class pattern classification problem. The CSI sample C_u with user identity u denotes the data to be classified, where $u = 1, \ldots, U$ (with U the total number of legitimate users), and y denotes its class ($y \in \{-1, 1\}$). We use $\{(C_{u,s}, y_{u,s}), s = 1, \ldots, S\}$ to denote a set of CSI samples associated with the user identity u. The challenge is how to construct a decision function $f(C_u)$ that correctly classifies the input CSI data, which could be different from all the constructed profiles.

If the constructed CSI user profiles are linearly separable, we can represent them with a linear function in the following form:

$$f(C_u) = w^T C_u + b, \tag{5}$$

such that $f(C_{u,s}) \geq 0$ for $y_{u,s} = 1$ and $f(C_{u,s}) \leq 0$ for $y_{u,s} = -1$, where w and b represent the hyperplane $f(C_u) = 0$ separating the two classes.

We seek such a hyperplane that maximizes the separating margins between the two classes. In particular, this hyperplane can be found by minimizing the following cost function:

$$\begin{aligned} &\min J(w,\xi) = \frac{1}{2}\|w\|^2 + \Gamma \sum_{s=1}^{S} \xi_{u,s}, \\ &s.t., y_{u,s}(w^T \Phi(C_{u,s}) + b) \geq 1 - \xi_{u,s}, \xi_{u,s} \geq 0, s = 1, \ldots, S, \end{aligned} \tag{6}$$

where $\Phi(\cdot)$ is a nonlinear operator mapping the CSI profile C_u to a higher dimensional space, Γ indicates the significance of the constraint violations with respect to the distance between the points and the hyperplane and ξ is a slack variable vector.

The mapping between the input CSI samples $C_{u,s'}$ and user profile $C_{u,s}$ is constructed in the form of a kernel function $Kernel(\cdot,\cdot)$, such as $Kernel(C_{u,s}, C_{u,s'}) = \Phi^T(C_{u,s})\Phi(C_{u,s'})$. Particularly, we choose a polynomial kernel as the mapping function and the problem in Equation (6) can be expressed as

$$\begin{aligned} &\max_{\alpha_s} \left\{ \sum_{s=1}^{S} \alpha_s - \frac{1}{2} \sum_{s=1}^{S} \sum_{s'=1}^{l} \alpha_u (y_{u,s} y_{u,s'} Kernel(C_{u,s}, C_{u,s'})) \alpha_{s'} \right\}, \\ &s.t., \alpha_s \geq 0, \sum_{s=1}^{S} \alpha_s y_{u,s} = 0, \end{aligned} \tag{7}$$

where α_s are Lagrange multipliers associated with Equation (7). Thus, the profile matching classifier for input CSI sample $C_{u,s'}$ is derived as

$$f(C_{u,s'}) = sign\left(\sum_{s=1}^{S} (\alpha_s y_{u,s} Kernel(C_{u,s'}, C_{u,s}) + b) \right), \tag{8}$$

and the authentication result is determined as

$$f(C_{u,s'}) = \begin{cases} 1 & success \\ -1 & failure. \end{cases} \tag{9}$$

6.2.2 Fusion via Multiple Antennas Pairs

When multiple antennas are available, we can further improve the performance of the user authentication accuracy. For example, we can employ a simple majority voting process to combine the independent profile matching results from different antenna pairs. Denoting the input CSI samples with user identity u between the transmitting antenna i and receiving antenna j by $C^{i,j}_{u,s'}$, all the independent results from different antenna pairs comprise the voting set $\Omega = \{f(C^{i,j}_{u,s'}), 1 \leq i \leq I, 1 \leq j \leq J\}$, where I and J are the numbers of transmitting and receiving antennas respectively. Finally, the authentication result is given by

$$f'(C_{u,s'}) = sign\left(\sum_{i=1}^{I} \sum_{j=1}^{J} f(C^{i,j}_{u,s'}) \right). \tag{10}$$

If $f'(C_{u,s'}) = 1$, the authentication successes; otherwise it fails.

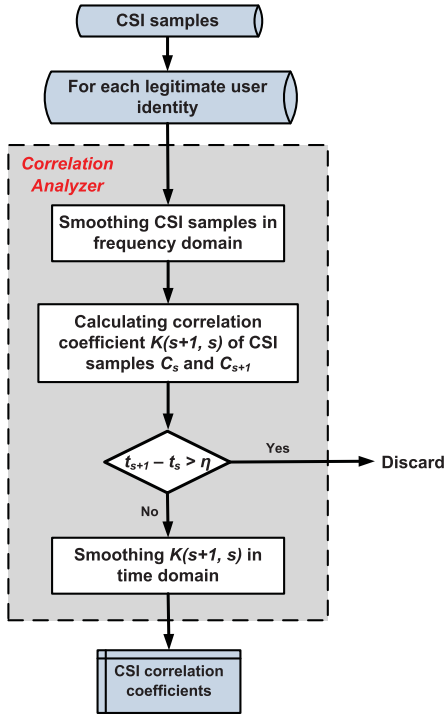


Fig. 8. Work flow for the temporal correlation analyzer.

7 Correlation-Based Mobile User Authentication

In this section, we present the proposed correlation-based method to authenticate mobile users.

7.1 Basic Idea

The basic idea to authenticate a mobile user is to examine the temporal correlation of adjacent CSI measurements from the user. The rationale behind this is that CSI measurements for the same user within a coherence time period should be high, whereas the correlation is low if the CSI measurements come from different mobile users. We next illustrate the work flow of the proposed authentication method.

7.2 Algorithm Description

The proposed mobile user authentication algorithm consists of two major components: a temporal correlation analyzer and a correlation-based detector. The details of each component are explained as follows:

Temporal Correlation Analyzer. In this component, we first derive the correlation coefficients between any two adjacent CSI measurements. The work flow of the temporal correlation analyzer is shown in Fig. 8. Before calculating temporal correlation coefficients, we smooth the CSI measurements in the frequency domain to mitigate the impact of interference and channel noise. Given a particular CSI measurement C_s, which is a vector representing the channel response of 30 carriers, the smoothing operation in the frequency domain is achieved as follows:

$$\mathcal{C}_s = \frac{1}{N} \sum_{max(0, f - \lfloor \frac{N}{2} \rfloor)}^{min(30, f + \lfloor \frac{N}{2} \rfloor)} C_s(f), \tag{11}$$

where N is the length of the smoothing window in the frequency domain. After smoothing, the random channel noise

over different subcarrier frequencies would be reduced from the original CSI measurement C_s.

Given the sequence of smoothed CSI measurements $\{\mathcal{C}_s\}, 0 \leq s \leq S$, where s indicates the index of CSI measurement and S is the number of CSI measurements, the temporal correlation coefficient $K(s+1, s)$ between two adjacent CSI measurements C_{s+1} and C_s is defined as follows:

$$K(s+1,s) = \frac{\mathcal{C}_{s+1}\mathcal{C}_s^T}{\|\mathcal{C}_{s+1}\|\|\mathcal{C}_s\|}, \tag{12}$$

where the superscript T denotes transpose.

Further, for regular wireless traffic, two adjacent CSI measurements may not be within a coherence period. The long time interval between two adjacent CSI measurements could result in the low correlation for a mobile user due to temporal and spatial diversity. We thus need to filter out these CSI measurements before performing user authentication. To facilitate such filtering, the correlation coefficients defined in Equation (11) is multiplied by a de-correlation factor η. Then, the updated temporal correlation coefficient is defined as

$$K'(s+1,s) = \frac{\mathcal{C}_{s+1}\mathcal{C}_s^T}{\|\mathcal{C}_{s+1}\|\|\mathcal{C}_s\|} \times \eta$$
$$\eta = \begin{cases} 1, & 0 < t_{s+1} - t_s \leq 0.025\,\text{sec} \\ Null, & o.w., \end{cases}$$

where s indicates the index of the CSI measurement, t_s denotes the time stamp of C_s, and η is the de-correlation factor indicating whether the correlation coefficient should be discarded. Particularly in our experiments, we adopt the de-correlation factor with 0.025 sec as the coherence time, which is a typical value arising in multipath propagation environments [8].

If the correlation coefficients are derived from two adjacent CSI samples spanning beyond a coherence period, they will be discarded; otherwise, they are forwarded to the correlation-based detector to perform mobile user authentication.

However, the CSI measurements may also suffer from occasional rapid movement or strong interference, which would destroy the correlation between adjacent CSI measurements. In order to derive reliable correlation coefficients, temporal smoothing is used to alleviate such effects, i.e.,

$$\mathcal{K}(s+1,s) = \frac{1}{w} \sum_{max(0,s-\lfloor \frac{w}{2} \rfloor)}^{min(L,s+\lfloor \frac{w}{2} \rfloor)} K'(s+1,s), \tag{13}$$

where w is the length of the temporal smoothing window size.

Correlation-Based Detector. As shown in Fig. 9, the correlation-based detector authenticates the mobile user leveraging the temporal correlation coefficients obtained from the correlation analyzer. Given the correlation coefficients $K'(s+1, s)$ of L CSI measurements, we divide them into several segments, where each has length L_k. For each segment, given a threshold ϵ, if there are m correlation coefficients $K'(s+1, s) \geq \epsilon$, where m is derived from empirical study, we declare the presence of attackers, otherwise the authentication succeeds.

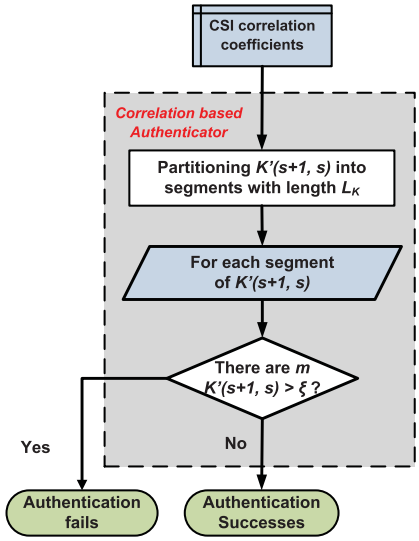


Fig. 9. Work flow for the correlation-based detector.

8 Performance Evaluation

In this section, we present performance evaluation of the proposed CSI-based user authentication framework in two types of real environments, laboratory and apartment. We show that the CSI-based authentication framework is resilient to attacks, and outperforms existing RSS-based authentication methods.

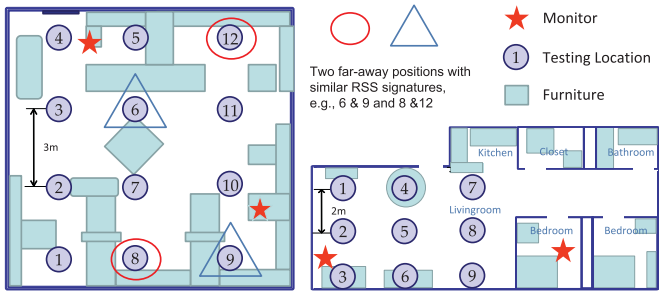
8.1 Experimental Setup

We conduct experiments in an 802.11n WiFi network with two laptops (i.e., Lenovo T500 and T61) serving as monitors that collect the wireless packets. These two laptops run Ubuntu 10.04 LTS with the 2.6.36 kernel and are equipped with Intel WiFi Link 5,300 wireless cards. Both Intel wireless cards' drivers are able to collect CSI information from each packet [1]. A commercial wireless access point, Linksys $E2500$, is sending out packets that can be captured by these two monitors. We use the *ping* command on two laptops to simulate the authentication packets continuously transmitted over the network. The packet rate is set to 10 packets/second. For each packet, we extract CSI for 30 subcarrier groups, which are evenly distributed in the 56 subcarriers of a 20 MHz channel. We also record the RSS value of each packet for comparison.

We conduct experiments in two indoor environments, i.e., a laboratory and an apartment. The laboratory represents the typical office environment, which has an office cubicle and pieces of furniture that create complex multipath effects in a large room. The apartment, on the other hand, represents the typical home environment with small rooms and simple furniture. The size of these two environments are $11\,\text{m} \times 12\,\text{m}$ and $11\,\text{m} \times 6\,\text{m}$, respectively. The experimental setups in these two environments are shown in Fig. 10. The numbered circles in the figures are the positions used to collect CSI data for evaluating our user authentication framework for the stationary case, and the two red stars represent two network monitors.

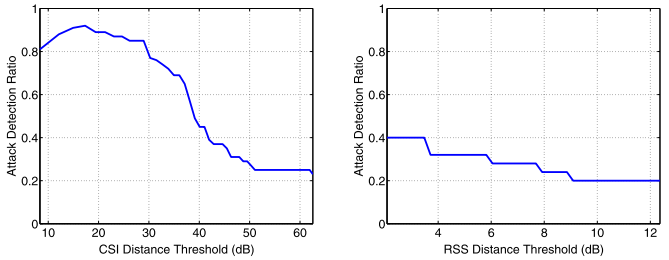
8.2 Experimental Methodology

In the experiments, we collect 400 packets at each location, and both CSI and RSS values of each packet are recorded. When using RSS measurements for user authentication,



(a) Laboratory (b) Apartment

Fig. 10. Experimental setups for (a) Laboratory and (b) Apartment.



(a) CSI-based (b) RSS-based

Fig. 11. Attack-resilient profile builder: Attack detect ratio versus cluster distance threshold when a spoofer is present.

we employ the RSS values collected from two network monitors as the two-dimensional feature vector for clustering and profile matching, while our proposed CSI-based authentication framework only uses the CSI measurements from one network monitor to perform user authentication.

To evaluate the performance of our proposed framework dealing with a stationary user, we examine two main attacking scenarios: 1) In the first attacking scenario, both the legitimate user and the attacker are present at the same time in the network. 2) In the second attacking scenario, after the attacker obtains the legitimate user's identity, only the attacker is active in the network. In order to obtain the statistical results, we choose all possible point pairs in both experimental environments and treat one point as the position of the legitimate user and the other point as the position of the attacker. We run the proposed framework through all possible combinations of point pairs. There are a total of 66 pairs for the laboratory environment and 36 pairs for the apartment environment. The experimental results are presented in the following sections for the attack resilient profile builder and profile matching authenticator.

For mobile user authentication, two data transmission patterns, including non-burst and burst, are studied. Particularly, non-burst data transmission means multiple users (i.e., two users in this work) are alternatively transmitting with similar low data rate at the same time, while burst data transmission means one individual user would successively transmit massive data packets at a high data rate during a short period. The two transmission patterns comprehensively depict the traffic pattern in real wireless environments, particularly, non-burst data transmission is for light traffic (e.g., link maintenance packets transmission) while burst data transmission is for heavy traffic (e.g., multimedia transmission). For each transmission pattern, we also examine two attacking scenarios: 1) In the first attacking scenario, both the legitimate user and the attacker are moving. 2) In the second attacking scenario, the legitimate user is moving but the attacker remains stationary. In both scenarios, the legitimate user and attacker are continuously transmitting signals with the same device ID.

8.3 Metrics

In order to evaluate the performance of our proposed user authentication framework for stationary users, we define the following two metrics: attack detection ratio and authentication accuracy.

Attack Detection Ratio (During Profile Building). We define the attack detection ratio $\bar{R}$ as the number of correctly detected spoofing attacks over the total number of experiments with spoofing attacks. The spoofing attacks presented when building the user profile belong to the attacking scenario 1. Given a total number of P attacking cases the attack detection ratio can be written as

$$\bar{R} = \frac{1}{P}\sum_{p=1}^{P} H_p s.t. \;\; H_p = \begin{cases} 0 & D_c \leq \tau \\ 1 & D_c > \tau, \end{cases} \tag{14}$$

where D_c is the distance between two centroids of clusters formed in the profile builder, and τ is the threshold used for spoofing attack detection.

Authentication Accuracy (During User Authentication). We define the authentication accuracy A_p as the number of correctly classified packets over the total number of packets collected in the pth attacking run. The attacks could belong to either the attacking scenario 1 or 2. We use $N_{u,p}$ to denote the number of packets that are sent by a legitimate user u and are correctly determined as being from the user u by our system. Similarly, we use $N'_{u,p}$ to denote the number of packets sent by the adversary using the identity of the legitimate user u and are correctly determined as not being from user u. We then define the authentication accuracy for the pth experimental run as $A_p = \frac{N_{u,p}+N'_{u,p}}{N_{a,p}}$, where $N_{a,p}$ is the total number of packets received with user identity u, and $N_{u,p} + N'_{u,p} \leq N_{a,p}$.

We further define the *average authentication accuracy* and *worst authentication accuracy* as shown below to evaluate the general and worst-case performance.

- *Average authentication accuracy:* Given P tested cases, the average authentication accuracy is given as $A_{avg} = \frac{1}{P}\sum_{p=1}^{P} A_p$,
- *Worst authentication accuracy:* The worst authentication accuracy chooses A_p from the attacking case with the smallest number of $N_{u,p}$ and $N'_{u,p}$ as $A_{worst} = \min_p A_p$.

8.4 Evaluation Results

8.4.1 Attack Detection Study During Profile Building for a Stationary User

We first compare the effectiveness of our Attack-resilient Profile Builder when determining the presence of a spoofer (during profile building) using CSI to that using RSS. We examine the attack detection ratio by varying the threshold τ. As shown in Fig. 11, the results show that the averaged detection ratio for the proposed CSI based approach achieves 0.92 with the optimal distance threshold 17 dB in Fig. 11a, while the maximum detection ratio for the RSS-based method is only 0.4 with distance threshold 2 dB as shown in Fig. 11b. This observation indicates that our

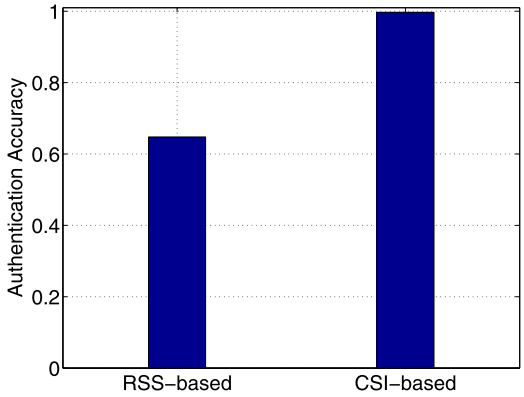


Fig. 12. Performance of the profile matching authenticator: Authentication accuracy when two users possess similar RSS fingerprints.

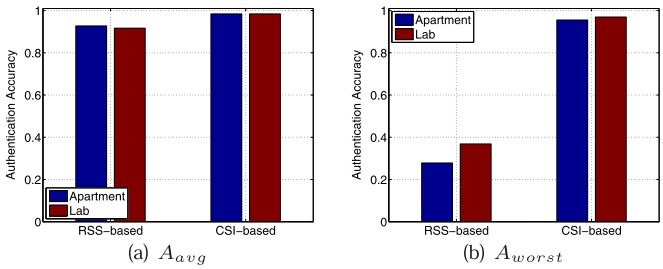


Fig. 13. User authentication accuracy comparison between CSI-based and RSS-based methods.

profile builder can effectively determine whether the network environment is benign or a spoofer is present when building the user profiles.

8.4.2 Authentication Accuracy Study for a Stationary User

Discriminating Two Far-Away Users with Similar RSS Fingerprints. Due to the irregularity of wireless signal propagation, two geographically distant users may share similar RSS signatures. For example, in Fig. 10a, the two positions 6 and 9 are about $6-7$ m away from each other, but their RSS fingerprints obtained from our network monitor look similar; positions 8 and 12 present the same signal phenomenon. This makes RSS-based user authentication schemes suffer poor performance when two legitimate users (but physically separated) present similar signal fingerprints. In particular, we observe that the authentication accuracy for the RSS-based method degrades to only around 0.64 as shown in Fig. 12. However, our proposed CSI-based method could still achieve the authentication accuracy close to 1. These results confirm that CSI measurements provide fine-grained information on differentiating users, even when their RSS measurements are similar.

Comparison with RSS-Based Methods. We next study the overall performance of our CSI based user authentication method. Fig. 13 shows the comparison of the authentication accuracy when using CSI-based and RSS-based methods in the two different environments (i.e., a laboratory and an apartment). We note that the RSS-based method relies on RSS values collected from two network monitors to perform user authentication, while our proposed CSI-based authentication framework only uses the CSI measurement from one antenna at one network monitor. We observe that our proposed CSI-based method outperforms the RSS-based

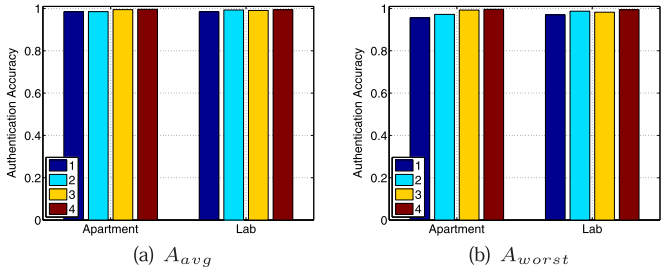


Fig. 14. CSI-based user authentication accuracy when involving single and multiple antennas.

method in both experimental environments. Specifically, Fig. 13a shows that the average authentication accuracy for CSI-based method is very high (above 0.984), and the RSS-based method has a lower authentication accuracy (i.e., 0.92). Furthermore, we see that the worst authentication accuracy for RSS-based method reduces to around 0.27 and 0.36 in the apartment and laboratory environments respectively, whereas our CSI-based method maintains high authentication accuracy over 0.95 as presented in Fig. 13b. These observations strongly indicate the robustness of our CSI-based user authentication framework even when only a single antenna is used on WiFi devices.

Impact of Single/Multiple Antennas. We further examine the performance when employing measurements from multiple antennas. We expect that using measurements from multiple antennas can provide better reliability for user authentication. Fig. 14 shows that both the average and worst authentication accuracy exhibit an increasing trend when more antennas are used. In particular, the authentication accuracy of using a single antenna in the apartment and laboratory environments is over 0.95. When the number of antenna pairs (i.e., a set of transmitting and receiving antennas) increases from 1 to 4, the average authentication accuracy in both laboratory and apartment further improves, and the worst authentication accuracy improves even more. We also observe that when using three antenna pairs in the laboratory environment the authentication accuracy has a slight drop when comparing to that of using two antenna pairs. This is because although current commodity wireless devices are usually equipped with multiple antennas, the main antennas usually have better quality of signal reception. Therefore, including the CSI samples from the main antennas (i.e., using one or two antenna pairs in our experiments) results in better stability of user authentication.

Impact of User Profile Size. Finally, we study how the number of packets (i.e., user profile size) employed to build the user profile affects the performance of our framework. We vary the size of the user profile from 1 to 200 packets, and the corresponding average authentication accuracy is shown in Fig. 15. When the size of the user profile increases, the authentication accuracy increases and then maintains a high level (i.e., over 0.95). We note that even if the profile of each user contains only one CSI sample, the authentication accuracy is still over 0.91. These results demonstrate that our profile builder is highly effective in our proposed framework.

8.4.3 Authentication Performance for a Mobile User

To evaluate the authentication performance for a mobile user, we study both non-burst and burst data transmission patterns. Note that the data rate for non-burst data transmission is

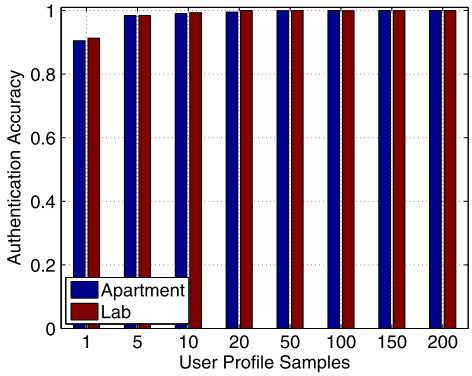
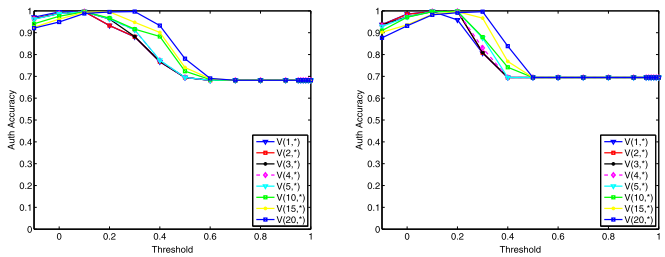


Fig. 15. Impact of user profile size on CSI-based user authentication accuracy.



(a) Scenario 1 (b) Scenario 2

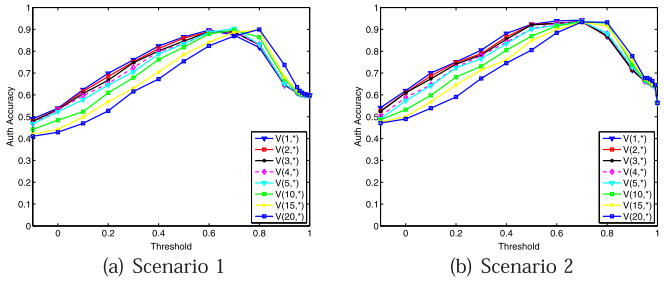
Fig. 16. Mobile user authentication accuracy for non-burst transmission.

TABLE 1
Authentication Performance for Non-Burst Data Transmission

Threshold	Scenario 1		Scenario 2	
	DR	FPR	DR	FPR
0.1	98.3%	0	97.4%	0%
0.2	99.3%	0	98.8%	0%
0.3	100%	0.1%	99.5%	0%

10 pkt/sec, while burst data transmission is 100 pkt/sec. Given the fixed data transmission rate, the authentication accuracy maintains when varying the de-correlation factor threshold. Each round of the authentication process takes about three seconds. For each data transmission pattern, we also examine two attacking scenarios as introduced in Section 8.1.

Non-Burst Data Transmission Study. First we study the authentication performance under non-burst data transmission. In order to evaluate the authentication performance under the impact of different parameters, we vary the correlation coefficient threshold from -0.1 to 1. Further, we evaluate the performance by varying the parameter m from 1 to 20. Note that we require m correlation coefficients of adjacent CSI measurements to exceed the threshold for user authentication. The total number of packets used for each round of user authentication is about 30. The authentication accuracy for Scenarios 1 and 2 are shown in Figs. 16a and 16b respectively, where $V(m, *)$ indicates the authentication accuracy curve under the requirement of at least m correlation coefficients of adjacent CSI measurements beyond different correlation coefficient thresholds. For both scenarios, as the correlation coefficient threshold increases, the authentication accuracy increases first and then decreases. Particularly, given $m = 20$, the authentication accuracy first goes up to as high as 99.7 percent with the threshold at around



(a) Scenario 1 (b) Scenario 2

Fig. 17. Mobile user authentication accuracy for burst transmission.

TABLE 2
Authentication Performance for Burst Data Transmission

Threshold	Scenario 1		Scenario 2	
	DR	FPR	DR	FPR
0.6	79.5%	0	70.6%	0%
0.7	88.3%	0	80.5%	3.6%
0.8	93.6%	7.2%	92.2%	13.4%

0.4, and then decreases to around 68.1 percent as the threshold reaches 1 in Scenario 1. A similar trend is also observed in Scenario 2 (i.e., given $m = 20$, the authentication accuracy reaches a maximum of 99.6 percent at the threshold $\epsilon = 0.3$, and then falls to 69.5 percent as the threshold increases to 1). Further, we also observe that the proposed approach achieves high authentication accuracy (i.e., over 99 percent) under all values of m ranging from 1 to 20.

In addition, to provide a more comprehensive view of authentication performance of the proposed approach, we also summarize the detection ratio and false positive ratio in Table 1 with $m = 20$. For both scenarios, given appropriate threshold values, the detection ratio would approach 100 percent while the false positive ratio is close to 0 percent. The above results confirm that the proposed framework is highly effective in authenticating mobile users with non-burst data transmission.

Burst Transmission Study. Next we examine the authentication performance for burst data transmission. The authentication accuracy for burst data transmission also follows the same trend as non-burst data transmission as the threshold value increases. Different from non-burst data transmission, burst data transmission involves many long CSI sequences from a single user, which leads to high correlation coefficients over a long period. So, the number of low correlation coefficients is fewer than that in non-burst data transmission. Therefore the thresholds to achieve maximum authentication accuracy should be higher. Consequently, a higher threshold would also falsely determine some low correlation coefficients from a single user as attacks. As shown in Figs. 17a and 17b, given $m = 20$, the maximum authentication accuracy is 93.25 and 90 percent for Scenarios 1 and 2 respectively when the threshold is 0.8. Therefore, the overall mobile user authentication performance under burst data transmission is slightly worse than that of non-burst data transmission. The detection ratio and false positive ratio for burst transmission are presented in Table 2, which confirm good authentication performance for both attacking scenarios. Particularly, as the threshold increases from 0.6 to 0.8, the detection ratio increases from 79.5 and 70.6 percent to 93.6 and 92.2 percent for Scenarios 1 and 2

respectively, while the false positive ratio increases from 0 and 7.2 percent to 0 and 13.4 percent for Scenarios 1 and 2 respectively.

9 Conclusion

In this paper, we have proposed to utilize channel state information to perform practical user authentication in wireless networks. The fine-grained channel information revealed in CSI has the potential to perform accurate user authentication. We have proposed a CSI-based user authentication framework that can work with both stationary and mobile users. Particularly, stationary user authentication includes an Attack-resilient User Profile Builder and Profile Matching Authenticator. The Attack-resilient Profile Builder employs clustering analysis to intelligently determine whether the network environment is benign without the presence of an identity-based attack when constructing the profile for the legitimate user. The Profile Matching Authenticator performs packet level user authentication grounded on Support Vector Machine techniques. It has the capability to distinguish between two users even when they possess similar signal fingerprints. Mobile user authentication leverages the temporal correlation of the wireless channel and includes a Temporal Correlation Analyzer and Correlation-based Detector. The Temporal Correlation Analyzer derives the temporal correlation coefficients between consecutive CSI measurements after filtering, whereas the Correlation-based Detector performs authentication based on an empirical threshold. Our extensive experiments conducted in both laboratory and apartment environments confirm the feasibility of exploiting CSI to perform accurate user authentication. The evaluation results show that the proposed CSI-based approach is highly effective for both stationary and mobile users.

Acknowledgments

A preliminary version of this work was presented in "Practical User Authentication Leveraging Channel State Information" [21] at the 9th ACM Symposium on Information, Computer and Communications Security (ASIA CCS) 2014. This work was supported in part by the National Science Foundation under grant numbers CNS1514436, CNS-1318748, and the CNS-1464092 and the Army Research Office under grant number W911NF-13-1-0288.

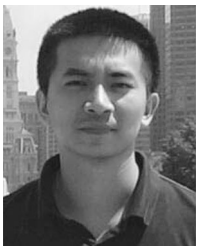
References

[1] IEEE Std. 802.11n-2009: Enhancements for higher throughput, 2009. [Online]. Available: http://www.ieee802.org
[2] B. Azimi-Sadjadi, A. Kiayias, A. Mercado, and B. Yener, "Robust key generation from signal envelopes in wireless networks," in *Proc. 14th ACM Conf. Comput. Commun. Secur.*, 2007, pp. 401–410.
[3] V. Brik, S. Banerjee, M. Gruteser, and S. Oh, "Wireless device identification with radiometric signatures," in *Proc. 14th ACM Int. Conf. Mobile Comput. Netw.*, 2008, pp. 116–127.
[4] G. Chandrasekaran, M. A. Ergin, M. Gruteser, R. P. Martin, J. Yang, and Y. Chen, "DECODE: Exploiting shadow fading to detect comoving wireless devices," *IEEE Trans. Mobile Comput.*, vol. 8, no. 12, pp. 1663–1675, Dec. 2009.
[5] O. Cheikhrouhou, A. Koubaa, M. Boujelben, and M. Abid, "A lightweight user authentication scheme for wireless sensor networks," in *Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl.*, 2010, pp. 1–7.
[6] Y. Chen, J. Yang, W. Trappe, and R. P. Martin, "Detecting and localizing identity-based attacks in wireless and sensor networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 5, pp. 2418–2434, Jun. 2010.
[7] O. Delgado-Mohatar, A. Fster-Sabater, and J. M. Sierra, "A lightweight authentication scheme for wireless sensor networks," *Ad Hoc Netw.*, vol. 9, no. 5, pp. 727–735, 2011.
[8] A. Goldsmith, *Wireless Communications*. New York, NY, USA: Cambridge Univ. Press, 2005.
[9] S. Govindarajan, P. Gasti, and K. S. Balagani, "Secure privacy-preserving protocols for outsourcing continuous authentication of smartphone users with touch data," *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 1, pp. 136–148, Sept. 2013.
[10] F. Guo and T.-C. Chiueh, "Sequence number-based MAC address spoof detection," in *Proc. Recent Advances Intrusion Detection*, 2006, pp. 309–329.
[11] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 40, pp. 159–170, 2010.
[12] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning, Data Mining Inference, and Prediction*. Berlin, Germany: Springer, 2001.
[13] S. Jana and S. K. Kasera, "On fast and accurate detection of unauthorized wireless access points using clock skews," *IEEE Trans. Mobile Comput.*, vol. 9, no. 3, pp. 449–462, Mar. 2010.
[14] 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 *Proc. IEEE Int. Conf. Comput. Commun.*, May 2013, pp. 2544–2552.
[15] A. Kalamandeen, A. Scannell, E. de Lara, A. Sheth, and A. LaMarca, "Ensemble: Cooperative proximity-based authentication," in *Proc. 8th Int. Conf. Mobile Syst. Appl. Services*, 2010, pp. 331–344.
[16] T. Karygiannis and L. Owens, "Wireless network security," *NIST Special Publication*, vol. 800, 2002, Art. no. 48.
[17] K. Kleisouris, B. Firner, R. Howard, Y. Zhang, and R. P. Martin, "Detecting intra-room mobility with signal strength descriptors," in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2010, pp. 71–80.
[18] T. Kohno, A. Broido, and K. C. Claffy, "Remote physical device fingerprinting," *IEEE Trans. Dependable Secure Comput.*, vol. 2, no. 2, pp. 93–108, Apr.–Jun. 2005.
[19] J. Krumm and E. Horvitz, "LOCADIO: Inferring motion and location from Wi-Fi signal strengths," in *Proc. 1st Annu. Int. Conf. Mobile Ubiquitous Syst.: Netw. Services*, 2004, pp. 4–13.
[20] L. Li, X. Zhao, and G. Xue, "Unobservable re-authentication for smartphones," in *Proc. 20th Netw. Distrib. Syst. Secur. Symp.*, Feb. 2013.
[21] H. Liu, Y. Wang, J. Liu, J. Yang, and Y. Chen, "Practical user authentication leveraging channel state information (CSI)," in *Proc. 9th ACM Symp. Inf. Comput. Commun. Secur.*, 2014, pp. 389–400.
[22] S. Mathur, R. Miller, A. Varshavsky, W. Trappe, and N. Mandayam, "ProxiMate: Proximity-based secure pairing using ambient wireless signals," in *Proc. 9th Int. Conf. Mobile Syst. Appl. Services*, 2011, pp. 211–224.
[23] S. Mathur, W. Trappe, N. Mandayam, C. Ye, and A. Reznik, "Radio-telepathy: Extracting a secret key from an unauthenticated wireless channel," in *Proc. 14th ACM Int. Conf. Mobile Comput. Netw.*, 2008, pp. 128–139.
[24] N. T. Nguyen, G. Zheng, Z. Han, and R. Zheng, "Device fingerprinting to enhance wireless security using nonparametric Bayesian method," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2011, pp. 1404–1412.
[25] L. O'Gorman, "Comparing passwords, tokens, and biometrics for user authentication," *Proc. IEEE*, vol. 91, no. 12, pp. 2021–2040, Dec. 2003.
[26] J. Pang, B. Greenstein, R. Gummadi, S. Seshan, and D. Wetherall, "802.11 user fingerprinting," in *Proc. 13th Annu. ACM Int. Conf. Mobile Comput. Netw.*, 2007, pp. 99–110.
[27] D. Shan, K. Zeng, W. Xiang, P. Richardson, and Y. Dong, "PHY-CRAM: Physical layer challenge-response authentication mechanism for wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 9, pp. 1817–1827, Sep. 2013.
[28] Y. Wang, J. Liu, Y. Chen, M. Gruteser, J. Yang, and H. Liu, " E-eyes: Device-free location-oriented activity identification using fine-grained WiFi signatures," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, 2014, pp. 617–628.
[29] A. Wool, "Lightweight key management for IEEE 802.11 wireless lans with key refresh and host revocation," *ACM/Springer Wireless Netw.*, vol. 11, no. 6, pp. 677–686, 2005.
[30] B. Wu, J. Wu, E. Fernandez, and S. Magliveras, "Secure and efficient key management in mobile ad hoc networks," in *Proc. 19th IEEE Int. Parallel Distrib. Process. Symp.*, 2005, Art. no. 8.

[31] J. Yang, Y. Chen, and W. Trappe, "Detecting spoofing attacks in mobile wireless environments," in *Proc. 6th Annu. IEEE Commun. Soc. Conf. Sensor Mesh ad hoc Commun. Netw.*, 2009, pp. 1–9.
[32] J. Yang, Y. Chen, W. Trappe, and J. Cheng, "Detection and localization of multiple spoofing attackers in wireless networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 1, pp. 44–58, Jan. 2013.
[33] K. Zeng, K. Govindan, and P. Mohapatra, "Non-cryptographic authentication and identification in wireless networks," *Wireless Commun.*, vol. 17, no. 5, pp. 56–62, 2010.
[34] K. Zeng, K. Govindan, D. Wu, and P. Mohapatra, "Identity-based attack detection in mobile wireless networks," in *Proc. IEEE Int. Conf. Comput. Commun.*, 2011, pp. 1880–1888.

Hongbo Liu received the PhD degree in electrical engineering from the Stevens Institute of Technology. He has been at Indiana University-Purdue University Indianapolis as an assistant professor in the Department of Computer Information and Graphics Technology since Aug. 2013. His research interests include mobile and pervasive computing, cyber security and privacy, and smart grid. He is the recipient of the Best Paper Award from ACM MobiCom 2011 and Best Paper Runner-up Award from IEEE CNS 2013.

Yan Wang received the PhD degree in electrical engineering from the Stevens Institute of Technology. He has been at SUNY Binghamton University as an assistant professor in the Department of Computer Science since Aug. 2015. His research interests include mobile and pervasive computing, smart healthcare, cyber security and privacy, and wireless networks. He is the Winner of the ACM MobiCom Student Research Competition, 2013. He was a graduate student in the Stevens Institute of Technology while working on this paper.



Jian Liu received the BE and MS degrees from the Department of Information Engineering, Wuhan University of Technology, China. He is currently working toward the PhD degree in the Department of Electrical and Computer Engineering, Stevens Institute of Technology. His current research interests include mobile computing, information security and privacy, and pervasive computing. He is currently working in the Data Analysis and Information SecuritY (DAISY) Lab with Prof. Yingying Chen.

Jie Yang received the PhD degree in computer engineering from the Stevens Institute of Technology, in 2011. He is currently an assistant professor in the Department of Computer Science, Florida State University. His research interests include cyber security and privacy, and mobile and pervasive computing, with an emphasis on network security, smartphone security and applications, security in cognitive radio and smart grid, location systems, and vehicular applications. His research is supported by the US National Science Foundation (NSF) and Army Research Office (ARO). He is the recipient of the Best Paper Award from the IEEE Conference on Communications and Network Security (CNS) 2014 and the Best Paper Award from ACM MobiCom 2011. His research has received wide press coverage including *MIT Technology Review*, *The Wall Street Journal*, NPR, CNET News, and Yahoo News. He is a member of the IEEE.

Yingying Chen received the PhD degree in computer science from Rutgers University. She is a professor in the Department of Electrical and Computer Engineering, Stevens Institute of Technology. Her research interests include cyber security and privacy, Internet of Things, smart healthcare, and mobile computing and sensing. She has published more than 100 journals and referred conference papers in these areas. Prior to joining Stevens, she was with Alcatel-Lucent, Murray Hill, New Jersey. She is the recipient of the NSF CAREER Award and Google Faculty Research Award. She also received the NJ Inventors Hall of Fame Innovator Award. She is the recipient of the Best Paper Awards from ACM AsiaCCCS 2016, IEEE CNS 2014, and ACM MobiCom 2011. She also received the IEEE Outstanding Contribution Award from the IEEE New Jersey Coast Section each year from 2005-2009. Her research has been reported in numerous media outlets including *MIT Technology Review*, CNN, Fox News Channel, *Wall Street Journal*, and National Public Radio. She serves on the editorial boards of the *IEEE Transactions on Mobile Computing*, the *IEEE Transactions on Wireless Communications*, and the *IEEE Network Magazine*.

H. Vincent Poor (S'72-M'77-SM'82-F'87) received the PhD degree in EECS from Princeton University, in 1977. From 1977 until 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990, he has been on the faculty at Princeton, where he is currently the Michael Henry Strater University professor of Electrical Engineering. During 2006 to 2016, he served as dean of Princeton's School of Engineering and Applied Science. His research interests include the areas of information theory, statistical signal processing, and stochastic analysis, and their applications in wireless networks and related fields. Among his publications in these areas is the book *Information Theoretic Security and Privacy of Information Systems* (Cambridge University Press, 2017). He is a member of the National Academy of Engineering, the National Academy of Sciences, and is a foreign member of the Royal Society. He is also a fellow of the American Academy of Arts and Sciences, the National Academy of Inventors, and other national and international academies. Recent recognition of his work includes the 2016 John Fritz Medal, the 2017 IEEE Alexander Graham Bell Medal, a Doctor of Science honoris causa from Syracuse University (2017), and Honorary Professorships at Peking University and Tsinghua University, both conferred in 2016.