Practical User Authentication Leveraging Channel State Information (CSI)

Hongbo Liu
Department of CIT
Indiana University Purdue University Indianapolis (IUPUI)
Indianapolis, IN, USA
hl45@iupui.edu

Yang Wang
Department of ECE
Stevens Institute of Technology
Hoboken, NJ, USA
ywang84@stevens.edu

Jian Liu
Department of ECE
Stevens Institute of Technology
Hoboken, NJ, USA
jliu28@stevens.edu

Jie Yang
Department of CSE
Oakland University
Rochester, MI, USA
yang@oakland.edu

Yingying Chen
Department of ECE
Stevens Institute of Technology
Hoboken, NJ, USA
yingying.chen@stevens.edu

ABSTRACT

User authentication is the critical first step to detect identity-based attacks and prevent subsequent malicious attacks. However, the increasingly dynamic mobile environments make it harder to always apply the 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, we explore to use channel state information (CSI), which is available from off-the-shelf WiFi devices, to conduct fine-grained user authentication. We propose an user-authentication framework that has the capability to build the user profile resilient to the presence of the spoofer. Our machine learning based user-authentication techniques can distinguish two users even when they possess similar signal fingerprints and detect the existence of the spoofer. Our experiments in both office building and apartment environments show that our framework can filter out the signal outliers and achieve higher authentication accuracy compared with existing approaches using received signal strength (RSS).

Categories and Subject Descriptors

C.2 [**COMPUTER-COMMUNICATION NETWORKS**]: General—*Security and protection*

General Terms

Design, Experimentation, Measurement, Performance

ASIA CCS'14, June 4–6, 2014, Kyoto, Japan.

http://dx.doi.org/10.1145/2590296.2590321.

Keywords

Channel state information, user authentication, wireless networks

1. INTRODUCTION

The rapid advancement of wireless technologies are making the wireless networks ubiquitous and people can access network services at anytime and anywhere. However, securing wireless networks is challenging due to the shared nature of wireless medium as adversaries can eavesdrop 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 (AP) can be spoofed, resulting in Denial of Service (i.e., rogue AP attack) [30]. 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 [22]. 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].

Recently authentication based on non-cryptographic methods are proposed to compliment and enhance the existing cryptography based schemes [6, 9, 3]. For example, the 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, 21]. The rational behind these schemes is that both RSS and CIR present unique spatial properties due to path loss and multi-path effects. An adversary, resided 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 of the wireless channel and thus the effectiveness of user authentication is largely limited. For example, the RSS-based authentication can hardly distinguish 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) [10], which could become an ideal candidate for achieving accurate user authentication. Specifically, we show that CSI can be utilized to accurately authenticate users with similar signal fingerprints and discriminate the legitimate user from a spoofing attacker. The detailed channel information has the power to enable user authentication at 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. 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 could 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 purpose. They assume the CSI collection is benign (without the presence of an identity-based attacker) when building the user profile. However, in practice the 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 the spoofer, leading to a misclassification of the user profile and falsely authenticate the spoofer. To tackle such challenges raised from 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 framework consisting of two main components: *Attack-resilient Profile Builder* and *Profile Matching Authenticator*. The *Attack-resilient Profile Builder* has the capability to accurately construct the user profile of the legitimate user even when the spoofing attacker is present. We further develop a *Profile Matching Authenticator* grounded on 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 to CSI to achieve accurate user authentication.

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 the fine-grained user authentication achievable in practice.
- We develop an user authentication framework that has the capability to build the user profile under the presence of the spoofing attack and achieves higher authentication accuracy compared with existing channel based (e.g., RSS-based) methods.
- 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 the related studies. The attack model and our framework overview 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. 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 7. Finally, we conclude our work in Section 8.

2. RELATED WORK

The traditional approach to provide user authentication is to use cryptographic-based authentication. For example, Wu et al. [28] 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 [27] 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 reduce its usability in a dynamic mobile wireless environment (i.e., lacks of a fixed key management infrastructure) or resource-constrained wireless networks (i.e., limited resources on wireless devices).

Recently non-cryptographic based authentication has drawn considerable attention [31]. 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 [25, 9], whereas hardware based authentication leverages the unique hardware traits such as channel-invariant radiometric [3, 23] and clock skews [18, 13] to identify users. Biometric and physical-trait based authentication relies on the behavioral modalities including on-screen touch and finger movement patterns [8, 20]. And channel-signature based authentication schemes are proposed to use either Received Signal Strength (RSS) [29, 6, 30, 32, 15] or Channel Impulse Response (CIR) [26, 21] 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 is very hard to falsify, for user authentication.

For the channel based user authentication using RSS, a series of approaches [29, 6, 30] have been proposed to detect identity-based attacks, determine the number of attackers when multiple adversaries masquerading as the same node identity, and localize the adversaries. Reciprocal Channel Variation-based Identification (RCVI) [32] 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 the current wireless devices, RSS is known to be sensitive to the multipath effects and affected by the transmission power level. As a

result, a legitimate user may be mistakenly regarded as the malicious user due to the inherent RSS variance. Different from RSS which is readily available in the existing wireless infrastructure, CIR is usually extracted from the specialized devices such as Field-Programmable Gate Array (FPGA) [26] and Universal Software Radio Peripheral (USRP) [21], which limit its practical usage in real-world scenarios.

Different from the previous work, we propose to use Channel State Information (CSI), a readily available fine-grained channel information from the current commercial hardware (i.e., 802.11 a/g/n devices), which represents both amplitude and phrase 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) for user authentication than applying existing channel based (i.e., RSS and CIR) authentication methods. The most related work to us is CSITE [14], which utilizes CSI magnitude measurements averaged over time to generate profiles for legitimate users. They assume the CSI collected over time is benign and there is no identity-based spoofing attack present when building the profile. However, in practice the spoofing attack could present at any time. Thus, the profiles built under such attacks cannot represent legitimate users and may lead to authenticate malicious users falsely. 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 to 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 [24]. 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 masquerades as the legitimate user to pass the user authentication process [6]. The identity-based attack is very harmful as once passing 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 Media Access Control (MAC) address of a legitimate WiFi device by passively monitoring the wireless traffic and then impersonate as the legitimate device by changing its MAC address. Another example is that by masquerading as an authorized wireless access point (AP) or an authorized client, an attacker could launch a variety of attacks including session hijacking, denial-of-service (DoS) attacks, or falsely advertise services to wireless clients [30].

In this work, we consider the identity-based attack can be present at any time. That is, different from the previous work, which only considers the presence of such an attack during the authentication phase, we take the view point that the 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., 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 static 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 one office room to a meeting room). The movement of the device can be detected using the existing techniques [19, 17, 4] (e.g., examining the variance of the wireless signal). In addition, we assume the attacker does not have the capability to capture and replay the CSI, thus the attacker cannot alter or jam the signals.

3.2 System Overview

Our basic idea is to profile the user by exploiting the readily available fine-grained CSI extracted from orthogonal frequency-division multiplexing (OFDM) 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 identity) contains a different CSI pattern from the legitimate profile, the network will raise an alert indicating possible identity-based attack and fails the user authentication on the specific device identity.

Our proposed user authentication framework, as depicted in Figure 1, consists of two main components: Attack-resilient Profile Builder and Profile Matching Authenticator. The network implementing this framework will keep monitoring the wireless traffic and examining CSI measurements from each packet based on the device's identity.

Attack-Resilient Profile Builder: The novelty of our profile builder is that it has the capability to build the actual user profile under the presence of the spoofer when building the user profile. 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 containing 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 authenticate the spoofer or deny the legitimate user's access. Our profile builder employing clustering analysis can separate the CSI measurements (from the legitimate user) from the ones (from the spoofer) and determine the presence of the spoofer. It can thus ensure the legitimacy of the user profile construction.

Furthermore, when the legitimate user moves from current location to another, e.g., from 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 [19, 17, 4].

Profile Matching Authenticator: This component examines the real-time CSI measurements per packet from a device ID and performs user authentication by perform-

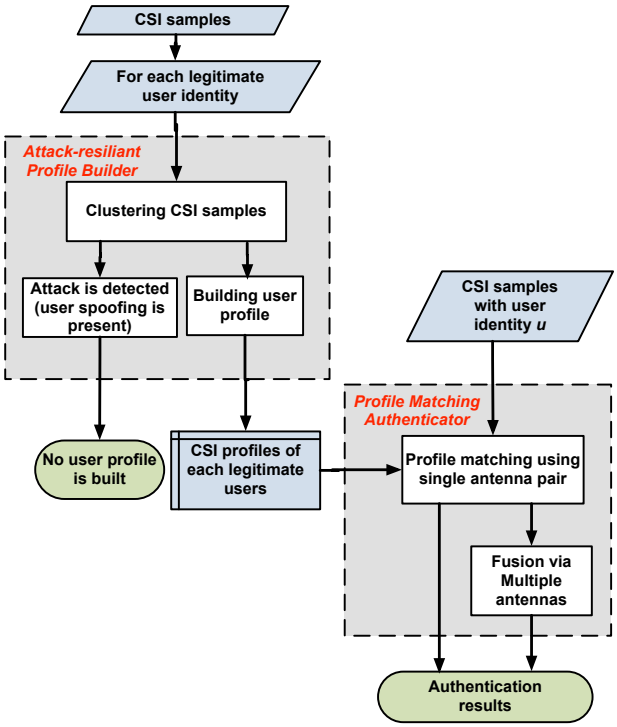


Figure 1: Overview of the CSI-based user authentication framework.

ing user profile matching. It is grounded on the 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 per packet level - authenticating each packet of the device ID. It is capable to authenticate different users even when they possess similar signal fingerprints due to the complex environment setup in real-life. The authenticator works well under both single antenna as well as multiple-antennas cases (using data fusion).

4. FEASIBILITY STUDY OF CSI-BASED USER AUTHENTICATION

In this section, we first provide the background of 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 Preliminary

Our authentication framework exploits the CSI measured from OFDM subcarriers, a reliable and fine-grained description of channel characteristics, for user authentication. OFDM technique has been extensively 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 802.11$a/g/n$, 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

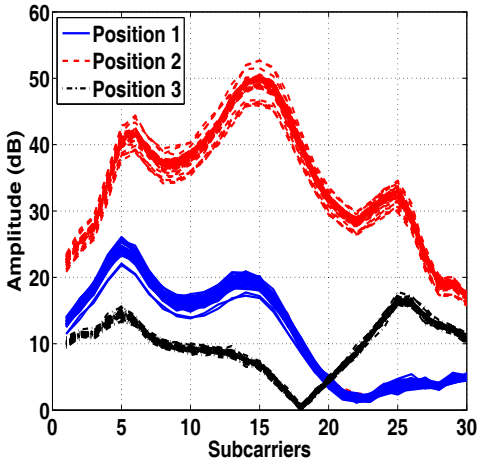


Figure 2: Example for channel state information of OFDM that is collected at three different positions.

frequency of each subcarrier is designed to be orthogonal to each other, so that the interference during transmission is minimized. For example, for the OFDM employed by the 802.11$a/g/n$ physical layer, a relatively wideband 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 (CSI) extracted from OFDM subcarriers for user authentication, which can provide a finer granularity of the channel information and has the potential to achieve higher accuracy for user authentication in practice. Figure 2 depicts the amplitude of channel state information across 30 subcarrier groups at three different positions. For each position, the CSI of 50 packets are measured from an Intel WiFi 5300 card in a laptop [11].

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 resided 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 Figure 2 that the amplitude of CSI at different subcarriers is different due to frequency diversity. Furthermore, the CSI shape 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, the profile of CSI should differ significantly. Additionally, we observe that the CSI of multiple packets from the same device at a fixed location exhibit the same trend, which indicates that an unique profile could be built for each user and serves 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 reading for user authentication purpose, a single monitor can provide

multi-dimensional channel state information sufficient for user authentication. Furthermore, since the widely adopted IEEE 802.11n standard[1] already defines a mechanism to exchange detailed CSI between a pair of wireless devices, employing CSI as an unique means for user authentication will not involve extra communication cost for the prevalent WiFi networks.

Data Preprocessing: In our study, we observe that the mean amplitude value of CSI measurement may shift over time. We call such a mean value shift as *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 on the amplitude of a specific subcarrier due to the interference presented at either transmitter or receiver. Figure 3 (a) plots the 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 across the subcarriers on CSI measurements 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:

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

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 Figure 3(b). In addition, the wireless devices in the 802.11n 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 $C_u^{i,j}$.

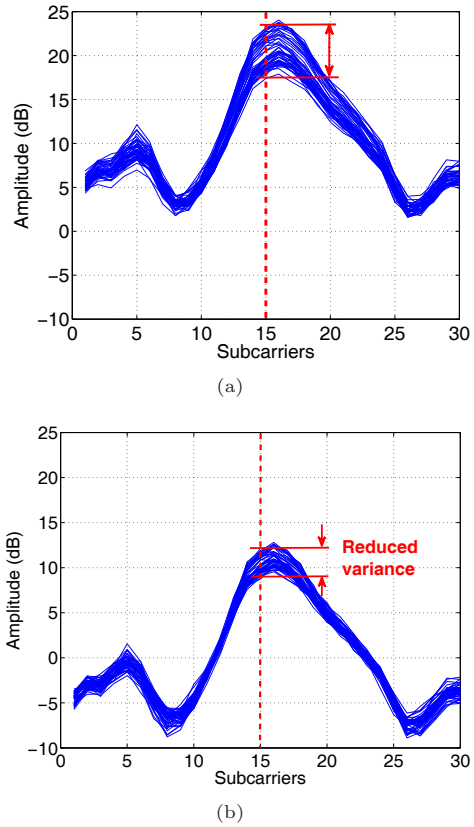


Figure 3: CSI samples before and after data processing.

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 the spoofer is present when constructing the user profile.

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 relational behind our attack-resilient profile builder is that the CSI measurements of each device presents 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 the 50 packets in Position 1 are clustered together (i.e., has a constant shape) in CSI space as shown in Figure 2. Furthermore, the CSI measurements of the wireless packets collected from another device resided 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 identity should be clustered together and form one cluster in the CSI space, while under the 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 the mixture readings from both 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 devices at different physical locations claiming the same device identity. As a result, more than one CSI clusters 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

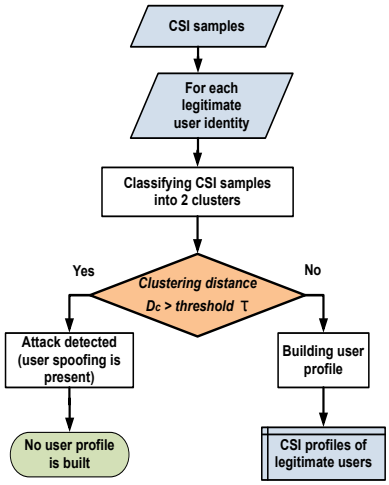


Figure 4: Work Flow for the Attack-Resilient Profile Builder.

when building user profiles. The flow of the Attack-resilient Profile Builder is shown in Figure 4. Only when there is no spoofing attack present, the profile of the legitimate user will 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 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 802.11n 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 802.11n packet header (at 2.4GHz band) contains a 16-bit MCS, which can be extracted together with the CSI sample of each packet.

Figure 5(a) 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 condition, 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 testing cases in both the lab and apartment environments. We thus filter out CSI for the packets with MCS value greater than 263, which corresponds to single spatial stream with transmission rate 60Mbps [1]. After filtering out the outliers, the CSI coming from the rest of the packets exhibit the similar shape (i.e., form one cluster in the CSI space) as shown in Figure 5(b).

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:

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

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,

$$D_c = ||\mu_{\mathbf{k}} - \mu_{\mathbf{k}'}|| \quad (3)$$

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 devices residing 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

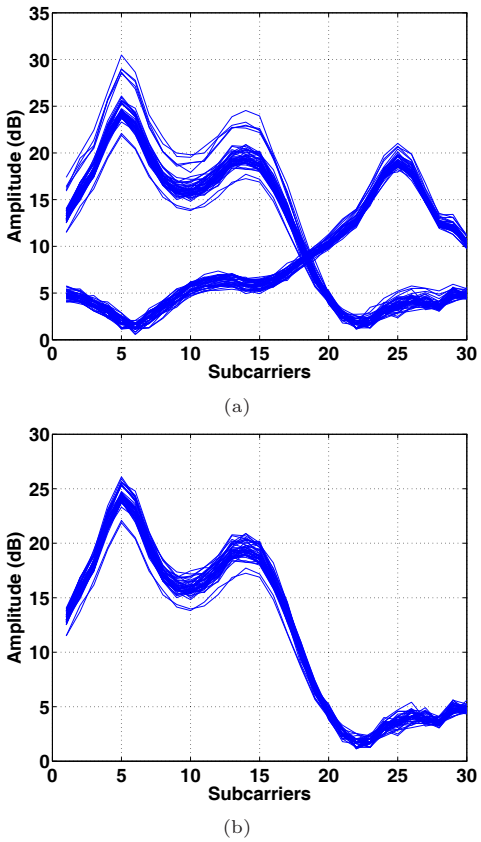


Figure 5: CSI samples before and after filtering based on MCS rate.

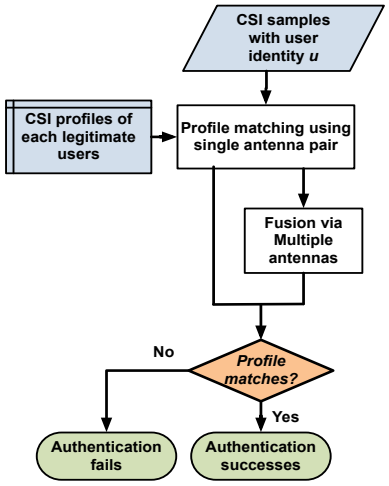


Figure 6: Work Flow for the Profile Matching Authenticator.

method has the capability of detecting the presence of the spoofer by applying the threshold τ to the D_c as following:

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

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, i.e. less than 100 packets.

If the user moves from one location to another (e.g., walks from his office to a meeting room), the user authentication framework will adaptively rebuild the user's profile. Following are possible two ways to update the user's profile: 1) the user can actively trigger the profile updating after he moves to a new place; 2) the profile updating can be triggered by detecting the user movement using existing techniques operating on wireless signals [19, 17, 4].

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.

6.1 Basic Idea

The basic idea of our profile matching authenticator is using machine learning to determine whether the CSI measurement for the incoming packet with the user identity u matches the profile constructed at the profile builder. If the incoming CSI sample matches the user profile, the corresponding packet can be authenticated successfully as from the user u. Otherwise, the user authentication fails. Figure 6 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 the 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 the 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 $u = 1, \cdots, U$ (with U as total number of legitimate users), and let scalar y denote its class ($y \in \{-1, 1\}$). We use $\{(C_{u,s}, y_{u,s}), s = 1, ..., S\}$ 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:

$$f(C_u) = w^T C_u + b \quad (5)$$

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 two classes.

We seek to find 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:

$$\min J(w, \xi) = \frac{1}{2}\|w\|^2 + \Gamma \sum_{s=1}^{S} \xi_{u,s} \quad (6)$$

subject to the following constraints:

$$y_{u,s}(w^T \Phi(C_{u,s}) + b) \geq 1 - \xi_{u,s}, \xi_{u,s} \geq 0, s = 1, \cdots, S, \quad (7)$$

where $\Phi(\cdot)$ is a non linear 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 the 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:

$$\max_{\alpha_s} \{\sum_{s=1}^{S} \alpha_s - \frac{1}{2} \sum_{s=1}^{S} \sum_{s'=1}^{l} \alpha_u (y_{u,s} y_{u,s'} Kernel(C_{u,s}, C_{u,s'})) \alpha_{s'}\} \quad (8)$$

subject to the constraints:

$$\alpha_s \geq 0, \sum_{s=1}^{S} \alpha_s y_{u,s} = 0, \quad (9)$$

where α_s is the Lagrange multipliers associated with equation 7. Thus, the profile matching classifier for input CSI

sample $C_{u,s'}$ is derived as:

$$f(C_{u,s'}) = sign(\sum_{s=1}^{S}(\alpha_s y_{u,s} Kernel(C_{u,s'}, C_{u,s}) + b)). \quad (10)$$

And the authentication result is determined as:

$$f(C_{u,s'}) = \begin{cases} 1 & success \\ -1 & failure. \end{cases} \quad (11)$$

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. Assume that the input CSI samples with user identity u between the transmitting antenna i and receiving antenna j are represented as $C^{i,j}_{u,s'}$, all the independent results from different antenna pairs consist of the voting set $\Omega = \{f(C^{i,j}_{u,s'}), 1 \le i \le I, 1 \le j \le J\}$, where I and J are the numbers of transmitting and receiving antennas respectively. Finally, the authentication result is given by:

$$f'(C_{u,s'}) = sign(\sum_{i=1}^{I}\sum_{j=1}^{J} f(C^{i,j}_{u,s'})). \quad (12)$$

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

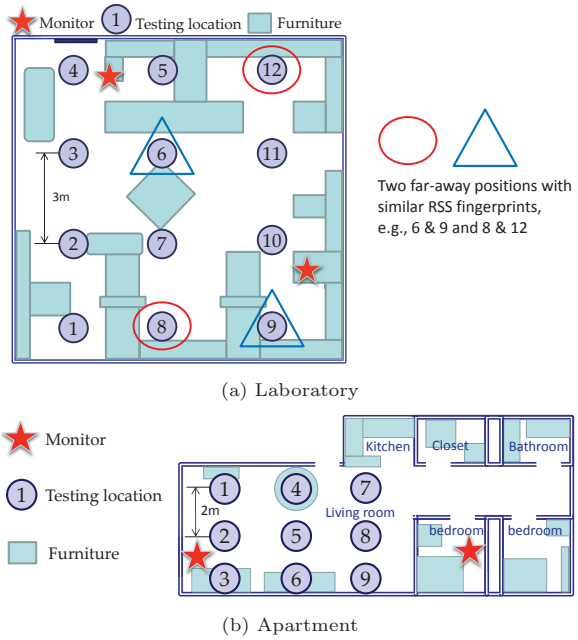
7. PERFORMANCE EVALUATION

In this section, we present the 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.

7.1 Experimental Setup

We conduct experiments in a 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 5300 wireless card. Both Intel wireless cards' drivers we installed are able to collect CSI information from frames transmitted in HT rate [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 [11]. 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 office cubicle and many 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 $11m \times 12m$ and $11m \times 6m$, respectively. The experimental setups in these two environments are shown in Figure 7. The numbered circles in the figures are the positions used to collect



(a) Laboratory

(b) Apartment

Figure 7: Experimental setups in (a) Laboratory and (b) Apartment.

CSI data for evaluating our user authentication framework, and the two red stars represent two network monitors.

7.2 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, 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, 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 laboratory environment and 36 pairs for apartment environment. The experimental results are presented in the following sections for the attack resilient profile builder and profile matching authenticator.

7.3 Metrics

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

Attack detection ratio (during profile building): We define the attack detection ratio R as the number of correctly detecting the presence of 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:

$$\bar{R} = \frac{1}{P}\sum_{p=1}^{P} H_p$$
$$s.t. \quad H_p = \begin{cases} 0 & D_c \leq \tau \\ 1 & D_c > \tau, \end{cases} \tag{13}$$

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 p^{th} 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 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 from the user u. We then define the authentication accuracy for the p^{th} experimental run as:

$$A_p = \frac{N_{u,p} + N'_{u,p}}{N_{a,p}}, \tag{14}$$

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 testing cases, the average authentication accuracy is given as:

$$A_{avg} = \frac{1}{P}\sum_{p=1}^{P} A_p. \tag{15}$$

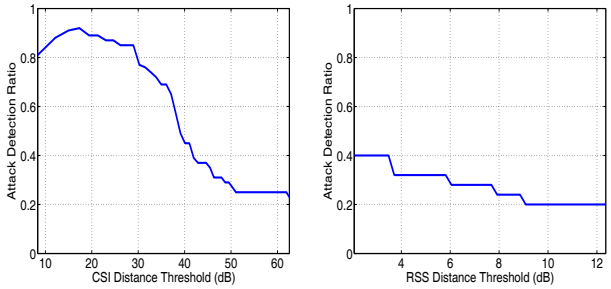
- **Worst authentication accuracy:** The worst authentication accuracy chooses A_p from the attacking case with the smallest number of $N_{u,p}$ and $N'_{u,p}$:

$$A_{worst} = \min_{p} A_p. \tag{16}$$

7.4 Evaluation Results

7.4.1 Attack Detection Study During Profile Building

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 Figure 8, the results show that the averaged detection ratio for the proposed CSI based approach achieves 0.92 with the optimal distance threshold $17dB$ in Figure 8 (a), while the maximum detection ratio for the RSS-based method is only 0.4 with distance threshold $2dB$ as shown in Figure 8 (b). This observation indicates that our profile builder can effectively determine whether the network environment is benign or a spoofer is present when building the user profiles.



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

Figure 8: Attack-resilient Profile Builder: Attack detect ratio versus cluster distance threshold when a spoofer is present.

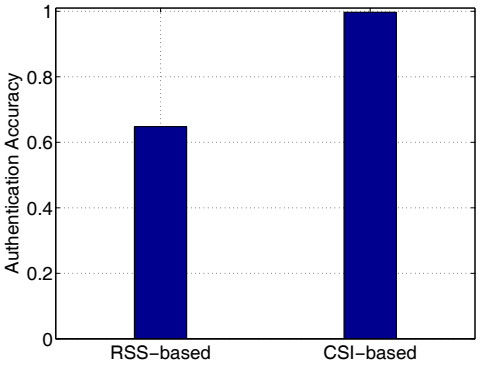
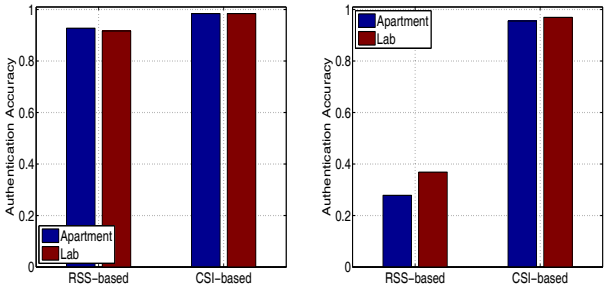


Figure 9: Performance of the Profile Matching Authenticator: Authentication accuracy when two users possess similar RSS fingerprints.

7.4.2 Authentication Accuracy Study

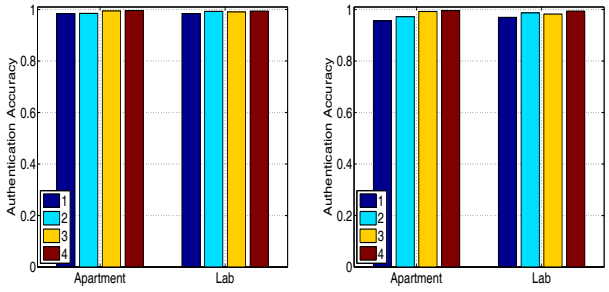
Discriminating two far-away users with similar RSS fingerprints: Due to the irregularity of wireless signal propagation, two geographical distant users may share similar RSS signatures. For example, in Figure 7 (a), two positions 6 and 9 are about $6-7m$ away from each other, but their RSS fingerprints obtained from our same 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 the similar signal fingerprints. In particular, we observe that the authentication accuracy for RSS-based method degrades to only around 0.64 as shown in Figure 9. However, our proposed CSI-based method could still achieve the authentication accuracy close to 1. The results confirm that CSI measurements provide fine-grained information on differentiating users, even when their RSS measurements are similar.

Comparison with RSS-based method: We next study the overall performance of our CSI based user authentication method. Figure 10 shows the comparison of the authentication accuracy when using CSI-based and RSS-based methods in 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



(a) A_{avg} (b) A_{worst}

Figure 10: User authentication accuracy comparison between CSI-based and RSS-based methods.



(a) A_{avg} (b) A_{worst}

Figure 11: CSI-based user authentication accuracy when involving single and multiple antennas.

from one antenna at one network monitor. We observe that our proposed CSI-based method outperforms the RSS-based method in both experimental environments. Specifically, Figure 10 (a) 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 show 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 the high authentication accuracy over 0.95 as presented in Figure 10 (b). 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 from 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. Figure 11 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 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 laboratory and apartment further improves, and the worst authentication accuracy improves even more. We also observe that when using 3 antenna pairs in the laboratory environment the authentication accuracy has a slightly drop when comparing to that of using 2 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 1 or 2 antenna pairs in our experiments) results in better stability of user authentication.

Impact from 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 user profile from 1 sample to 200 samples, and the corresponding average authentication accuracy is shown in Figure 12. When the size of user profile increases, the authentication accuracy increases and then maintains at a high level (i.e., over 0.95). We note that even if the profile of each user contains only 1 CSI sample, the authentication accuracy is still over 0.91. These results demonstrate that our profile builder is highly effective in our CSI-based user authentication framework.

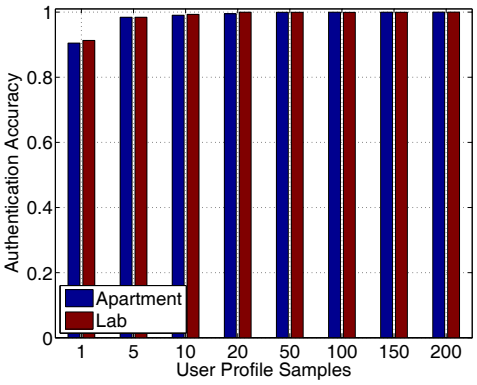


Figure 12: Impact of user profile size on CSI-based user authentication accuracy.

8. CONCLUSION

In this paper, we study utilizing channel state information (CSI) 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 propose a CSI-based user authentication framework including 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 the 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 (SVM). It has the capability to distinguish two users even when they possess the similar signal fingerprints. Our extensive experiments conducted in both lab and apartment environments confirm the feasibility of exploiting CSI to perform accurate user authentication. The evaluation results show that the CSI-based approach is highly effective as compared with methods directly applying received signal strength.

9. ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under grant numbers CNS-1318751, CNS-1318748 and Army Research Office W911NF-13-1-0288.

10. REFERENCES

[1] IEEE Std. 802.11n-2009: Enhancements for higher throughput, 2009. Available at 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 *Proceedings of the 14th ACM conference on Computer and communications security*, pages 401–410, 2007.

[3] V. Brik, S. Banerjee, M. Gruteser, and S. Oh. Wireless device identification with radiometric signatures. In *Proceedings of the 14th ACM international conference on Mobile computing and networking*, pages 116–127, 2008.

[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 Transactions on Mobile Computing*, 8(12):1663–1675, 2009.

[5] O. Cheikhrouhou, A. Koubaa, M. Boujelben, and M. Abid. A lightweight user authentication scheme for wireless sensor networks. In *IEEE/ACS International Conference on Computer Systems and Applications (AICCSA)*, pages 1–7, 2010.

[6] Y. Chen, J. Yang, W. Trappe, and R. P. Martin. Detecting and localizing identity-based attacks in wireless and sensor networks. *IEEE Transactions on Vehicular Technology*, 59(5):2418–2434, 2010.

[7] O. Delgado-Mohatar, A. FÃžster-Sabater, and J. M. Sierra. A light-weight authentication scheme for wireless sensor networks. *Ad Hoc Networks*, 9(5):727–735, 2011.

[8] S. Govindarajan, P. Gasti, and K. S. Balagani. Secure privacy-preserving protocols for outsourcing continuous authentication of smartphone users with touch data. *IEEE Transactions on Information Forensics and Security*, 8(1):136–148, 2013.

[9] F. Guo and T.-c. Chiueh. Sequence number-based mac address spoof detection. In *Recent Advances in Intrusion Detection*, pages 309–329, 2006.

[10] D. Halperin, W. Hu, A. Sheth, and D. Wetherall. Predictable 802.11 packet delivery from wireless channel measurements. In *ACM SIGCOMM Computer Communication Review*, volume 40, pages 159–170, 2010.

[11] D. Halperin, W. Hu, A. Sheth, and D. Wetherall. Predictable 802.11 packet delivery from wireless channel measurements. In *ACM SIGCOMM Computer Communication Review*, volume 40, pages 159–170, 2010.

[12] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning, Data Mining Inference, and Prediction*. Springer, 2001.

[13] S. Jana and S. K. Kasera. On fast and accurate detection of unauthorized wireless access points using clock skews. *IEEE Transactions on Mobile Computing*, 9(3):449–462, 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 *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, May 2013.

[15] A. Kalamandeen, A. Scannell, E. de Lara, A. Sheth, and A. LaMarca. Ensemble: cooperative proximity-based authentication. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pages 331–344, 2010.

[16] T. Karygiannis and L. Owens. Wireless network security. *NIST special publication*, 800:48, 2002.

[17] K. Kleisouris, B. Firner, R. Howard, Y. Zhang, and R. P. Martin. Detecting intra-room mobility with signal strength descriptors. In *The ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, pages 71–80, 2010.

[18] T. Kohno, A. Broido, and K. C. Claffy. Remote physical device fingerprinting. *IEEE Transactions on Dependable and Secure Computing*, 2(2):93–108, 2005.

[19] J. Krumm and E. Horvitz. Locadio: Inferring motion and location from wi-fi signal strengths. In *MobiQuitous*, pages 4–13, 2004.

[20] L. Li, X. Zhao, and G. Xue. Unobservable re-authentication for smartphones. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, February 2013.

[21] S. Mathur, R. Miller, A. Varshavsky, W. Trappe, and N. Mandayam. Proximate: proximity-based secure pairing using ambient wireless signals. In *Proceedings of the 9th international conference on Mobile systems, applications, and services*, pages 211–224, 2011.

[22] S. Mathur, W. Trappe, N. Mandayam, C. Ye, and A. Reznik. Radio-telepathy: extracting a secret key from an unauthenticated wireless channel. In *Proceedings of the 14th ACM international conference on Mobile computing and networking*, pages 128–139, 2008.

[23] N. T. Nguyen, G. Zheng, Z. Han, and R. Zheng. Device fingerprinting to enhance wireless security using nonparametric bayesian method. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, pages 1404–1412, 2011.

[24] L. O'Gorman. Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE*, 91(12):2021–2040, 2003.

[25] J. Pang, B. Greenstein, R. Gummadi, S. Seshan, and D. Wetherall. 802.11 user fingerprinting. In *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, pages 99–110, 2007.

[26] D. Shan, K. Zeng, W. Xiang, P. Richardson, and Y. Dong. Phy-cram: Physical layer challenge-response authentication mechanism for wireless networks. *IEEE Journal on Selected Areas in Communications*, 31(9):1817–1827, 2013.

[27] A. Wool. Lightweight key management for ieee 802.11 wireless lans with key refresh and host revocation. *ACM/Springer Wireless Networks*, 11(6):677–686, 2005.

[28] B. Wu, J. Wu, E. Fernandez, and S. Magliveras. Secure and efficient key management in mobile ad hoc networks. In *Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 2005.

[29] J. Yang, Y. Chen, and W. Trappe. Detecting spoofing attacks in mobile wireless environments. In *6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, pages 1–9, 2009.

[30] J. Yang, Y. Chen, W. Trappe, and J. Cheng. Detection and localization of multiple spoofing attackers in wireless networks. *IEEE Transactions on Parallel and Distributed Systems*, 24(1):44–58, 2013.

[31] K. Zeng, K. Govindan, and P. Mohapatra. Non-cryptographic authentication and identification in wireless networks. *Wireless Communications*, 17(5):56–62, 2010.

[32] K. Zeng, K. Govindan, D. Wu, and P. Mohapatra. Identity-based attack detection in mobile wireless networks. In *Proceedings of the IEEE International Conference on Computer Communications*