Towards Environment-independent Behavior-based User Authentication Using WiFi

Cong Shi*, Jian Liu†, Nick Borodinov‡, Bruno Leao‡, Yingying Chen*
*Rutgers University, New Brunswick, NJ 08901, USA
†University of Tennessee, Knoxville, TN 37996, USA
‡Siemens, Princeton, NJ 08540, USA
Email: *{cs1421, yingche}@scarletmail.rutgers.edu, †jliu@utk.edu, ‡{nikolay.borodinov, bruno.leao}@siemens.com

***Abstract*—With the increasing prevalence of smart mobile and Internet of things (IoT) environments, user authentication has become a critical component for not only preventing unauthorized access to security-sensitive systems but also providing customized services for individual users. Unlike traditional approaches relying on tedious passwords or specialized biometric/wearable sensors, this paper presents a device-free user authentication via daily human behavioral patterns captured by existing WiFi infrastructures. Specifically, our system exploits readily available channel state information (CSI) in WiFi signals to capture unique behavioral biometrics residing in the user's daily activities, without requiring any dedicated sensors or wearable device attachment. To build such a system, one major challenge is that wireless signals always carry substantial information that is specific to the user's location and surrounding environment, rendering the trained model less effective when being applied to the data collected in a new location or environment. This issue could lead to significant authentication errors and may quickly ruin the whole system in practice. To disentangle the behavioral biometrics for practical environment-independent user authentication, we propose an end-to-end deep-learning based approach with domain adaptation techniques to remove the environment- and location-specific information contained in the collected WiFi measurements. Extensive experiments in a residential apartment and an office with various scales of user location variations and environmental changes demonstrate the effectiveness and generalizability of the proposed authentication system.**

I. Introduction

With the increasing deployment of smart and IoT environments, private resources are linked to a huge number of spatially distributed intelligent devices (e.g., smart hub, voice assistant, laptop) and many of them store private information or run security-sensitive applications. Unauthorized access to such restricted devices puts user privacy and property at high risk. For example, an unauthorized user may obtain private information (e.g., classified file, credit card information, healthcare data) stored on a laptop [1] or disarm the home security system via a smart speaker [2]. Thus, user authentication has become a critical component in assuring user security in smart spaces. Furthermore, smart environments have a growing trend of exploring the ability to identify users and deliver personalized services for improving the utility of the space, such as recommending music channels, supporting online shopping, and controlling home appliances. User identification/authentication is serving as an inseparable step in enabling personalized and convenient use of emerging services in smart environments.

Traditionally, passwords and physiological biometrics such as fingerprint and facial information are widely used to authenticate users. They either require the user to remember secret details or need deployment of dedicated biometric sensors. A new trend is to verify users through behavioral biometrics. For instance, gestures on smart glasses' touchpad (e.g., taps or swipes) could be utilized to continuously authenticate users [3]. In another instance, gait patterns derived from mobile devices can be utilized to identify walking people [4]. Furthermore, unique behavioral hallmarks captured through wrist-worn sensors are explored to identify users when they are operating home appliances. This trend has the ability to provide continuous user authentication and also brings convenience to users. But these new solutions require the user to wear devices, such as wearable sensors and smart glasses. To further advance user authentication based on user's behaviors, in this work, we aim to develop a device-free system that verifies the user identity through daily behaviors at smart spaces without requiring the user to wear any devices or deploy any dedicated sensors. We would like to utilize WiFi signals generated by the regular communications among smart/IoT devices to capture inherited behavioral characteristics to facilitate identification/authentication. Particularly, we aim to derive a solution that could be environment-dependent (e.g., robust to the changes of furniture/appliance placement).

In the past decade, smart environments have incorporated WiFi technologies to provide increasing wireless interconnections among appliances and mobile devices (e.g., voice assistants, smart TVs, smart refrigerators, smart glasses). This increased popularity of wireless technologies produces WiFi signals that cover almost every corner of the smart spaces. In addition to the prevalence usage of communication, users' unique behaviors could be embedded inside the wireless signals through their daily behaviors (e.g., interacting with smart home appliances, operating on a laptop, entering a corporate building). It is thus natural to explore wireless signals ubiquitous in smart environments to derive inherited behavioral characteristics to authenticate and identify users in a convenient and device-free manner. Recent years have witnessed the emergence of user identification using WiFi [5]–[9]. Many of these solutions rely on gait patterns derived from WiFi signals to identify walking users [5]–[7]. They usually require the user to walk through well-designed paths and are

thus limited to recognizing walking people, which might not be practical in many scenarios. Our research group obtained the initial success of identifying the user through daily activities [8] or finger gestures [9]. These initial studies showed the success of using WiFi to perform user authentication during daily activities. However, they did not address the problem in real-world scenarios with dynamics, such as small location changes during activities and physical environmental dynamics (e.g., placement changes of furniture/appliances, movements of WiFi/smart IoT devices).

In this work, we take one step further to design and develop user identification/authentication using WiFi signals that can work under various scales of environmental dynamics. In particular, we address two important problems: **(i) Small-scale Location Variations:** A user usually conducts the same behavior in proximity with small location variations; and **(ii) Large-scale Environmental Changes:** The status of a physical environment (e.g., the placements of furniture and home appliances) could vary from day to day in practical scenarios. Propagation of wireless signals can be impacted by the environmental changes, and WiFi signal, especially the channel state information (CSI), is very sensitive to small-scale user location changes. Thus, both the small-scale location variations and large-scale environmental changes could lead to changes of fluctuation pattern in the fine-grained WiFi measurements, thereby causing signal profile mismatches during user authentication. This profile inconsistency problem usually requires the system to collect massive training data to cover all the possible locations and environmental status and retrain the model, making the system hard to be deployed in practice.

To address these two issues encountered in practice, we develop a deep-learning-based user recognition model together with an unsupervised domain discriminator, which are built upon both labeled data (i.e., WiFi signals containing user behaviors with user-identity labels) and unlabeled data (i.e., user behavior data without labels), to mitigate the impact of varying location and environmental changes and achieve reliable user authentication. This developed deep learning model could be utilized in a new environment when working with the collected user behaviors (i.e., unlabeled data) to perform user authentication and thus achieving environment-independency.

To implement such an environment-independent WiFi-based user authentication system, we extract information from multiple links of CSI measurements. Particularly, we exploit amplitude (i.e., from every single link) and relative amplitude (i.e., between every two links), which are affected by user behaviors and thus have the potential to capture unique behavioral characteristics. Our system then performs normalization to mitigate the impacts caused by ambient and hardware noises. Both time-domain and frequency-domain representations (i.e., spectrogram) of these amplitude and relative amplitude data will be then fed into a convolutional neural network (CNN) to learn features for characterizing both user's identity and the performed activities. By training the CNN with an adversarial loss, we will make the extracted features discriminative among different users while being transferable in different domains (i.e., user locations and environmental status). We highlight our contributions as follows:

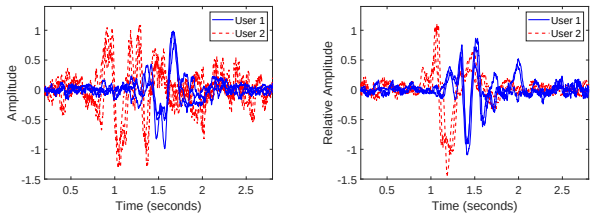
- We develop a domain adaptation technique to remove unpredictable environment-specific factors from the learned representations to achieve more practical user authentication.
- By examining the fine-grained channel state information (CSI) of WiFi signals, we find the combination of amplitude and relative amplitude between two links is effective for capturing unique individual behavioral biometrics.
- Our domain adaptation strategy is unsupervised and can derive environment-independent convolutional neural network (CNN) models for both activity recognition and user identification. The derived model is also resilient to spoofing attacks.
- Extensive experiments in a residential apartment and an office over four months demonstrate the effectiveness (e.g., over 87% user identification accuracy) and generabizability of the proposed system under various degrees of location variations and environmental changes.

II. Related Work

Traditional password or PIN number based schemes rely on using memorized secret details to confirm the user's identity. Such schemes solely based on the knowledge of the secrets instead of the user. They are vulnerable to password theft. Recent user authentication solutions utilize physiological biometrics to authenticate users, such as fingerprint, facial information, and iris. These solution need the deployment of dedicated biometric hardware or sensors, which incur extra cost.

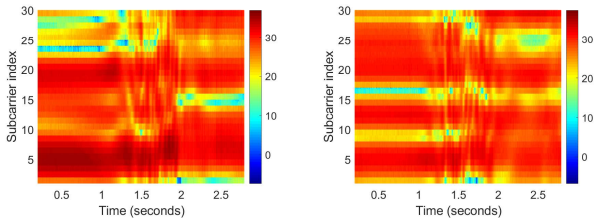
A new trend of user authentication is to explore behavioral biometrics [3], [10], [11]. Some research studies attempt to continuously authenticate users through unique typing/taping behaviors on smart devices, such as gestures on smart glasses' touchpad [3], taping on devices' touch screen [10], or keystroke dynamics [11]. Although these approaches do not require the user to remember a password, they only work when a user is operating on a device. Furthermore, Ranjan and Whitehouse [12] exploit wrist-worn sensors to capture unique behavioral hallmarks when users are operating home appliances.

The prevalent WiFi technologies opens up a new direction for researchers to explore WiFi signals to perform user identification/authentication. Many of these works use WiFi signals to capture human walking gait patterns and identify a small group of users in a shared space. For example, WiFi-ID [5], WiWho [6], and WifiU [13] extract statistic features from CSI variations in either the time or frequency domain to identify a person's walking steps and gait patterns. Instead of using handcraft features, WiAU [7] exploits a ResNet-based model to learn deep features from CSI to characterize human gait patterns and utilizes a transfer learning algorithm to alleviate the impacts of environmental variations. These approaches do not require the user to wear any devices. Our research group has shown the initial success of using WiFi signals to authenticate users through either daily activities [8]



(a) Amplitude (b) Relative Amplitude

Fig. 1: CSI amplitude and relative amplitude capturing two users's behaviors in a daily activity (i.e., picking up a remote controller).



(a) Environmental Status 1 (b) Environmental Status 2

Fig. 2: CSI amplitude of 30 subcarriers for the same user behavior (i.e., picking up a remote controller) collected in different environmental status, in which the positions of a desk and two chairs are different.

or finger gestures [9]. These approaches show reasonable user identification accuracy and confirm the advantages of using WiFi signals to authenticate user. However, they rely on stable environment status and do not address the problems of location variations and environmental changes in practical scenarios. In comparison, we propose an unsupervised domain adaptation discriminator ground on deep learning to simultaneously address these two problems. Different from WiAU [7] which requires using labeled data in adapting the deep features to new environments, our approach is unsupervised and does not require any labeled data for the adaptation. With such an unsupervised approach, our system can achieve adaptive and practical user authentication based on WiFi.

III. Preliminaries

WiFi signals generated by ubiquitous electronic devices (e.g., voice assistant, smart refrigerator, and laptop) can be easily impacted by the surrounding people's moving behaviors during their daily activities. The channel station information (CSI) available in the existing WiFi protocol (e.g., 802.11n) can describe the properties of signal propagation (e.g., scattering, fading, power decay, and multi-path) and thus can be used to derive unique individual behavioral characteristics for user authentication.

Specifically, the CSI measurements are complex values that characterize how WiFi signals propagate from the transmitter to the receiver at certain subcarrier frequencies. The CSI amplitude describes the combined attenuation effects of the WiFi signals propagating through different paths in the environment. Human behaviors can alter the propagation path of the WiFi signals, resulting in distinctive characteristics on the CSI amplitude. We thus propose to extract representative features from CSI amplitude to capture unique human behavioral characteristics. To extract extra information of the user's behaviors over multiple OFDM antennas, we propose to utilize *relative amplitude* derived from every two OFDM WiFi links. It captures the gain difference between the two links and is more stable under the changing environmental status. Considering two WiFi links l_1 and l_2, the relative amplitude at the k_{th} subcarrier can be represented as:

$$\hat{H}_k^{l_1 \leftrightarrow l_2} = |H_k^{l_1}(H_k^{l_2})^*|, \quad (1)$$

where $|\cdot|$ represents the absolute value and $(\cdot)^*$ denotes the complex conjugate. The relative amplitude can avoid ever-changing gain offsets shifted by the hardware control mechanism and the ambient radio frequency interference.

Figure 1 (a) and (b) show the CSI amplitude and relative amplitude information of a subcarrier over time when two users were picking up a remote controller (3 rounds for each) respectively in a residential apartment. We observe that both the amplitude and relative amplitude exhibit distinctive fluctuation patterns between these two users, which confirms their capability in capturing unique user behavioral characteristics. Additionally, compared to amplitude, the relative amplitude shows different variation trends for the same user's behavior, demonstrating its capability on providing extra information on unique behavioral characteristics.

IV. System Design

A. Challenges

Reliability of CSI Measurements. CSI is susceptible to ambient radio frequency interference (e.g., from neighboring WiFi devices) and channel condition changes. It is thus necessary to convert the CSI readings into a reliable form to mitigate the impacts of such channel distortions.

Small-scale Location Variations. In real-world scenarios, people may perform the same activity at slightly different locations every time. It is thus highly desired to make the extracted features invariant to such small-scale location variations for robust user authentication.

Large-scale Environmental Changes. The variations of environmental status (e.g., placement changes of furniture/home devices) could alter the multipath environment and distort the frequency-selective fading patterns of CSI [14], thereby resulting in different CSI fluctuation patterns. We demonstrate this with two sets of CSI associated with the same user behavior (i.e., picking up a remote controller) collected in two environmental status, in which the positions of a desk and two chairs are different. From Figure 2, we can observe that the two sets of CSI are significantly different, making it very challenging to derive a reliable authentication model.

B. Attack Model

Random Attack. An adversary does not have prior information on the activities used by the legitimate user for authentication. To pass the authentication, the adversary attempts to conduct random activities to create similar impacts on CSI measurements as the legitimate user.

Mimic Attack. An adversary has observed how the legitimate user performed activities during the authentication process multiple times by peeping (e.g., through videotaping). The adversary tries to pass the authentication by imitating the performed activities of the legitimate user with the same environmental status when the user enrolled the system.

C. System Overview

We design an environment-independent system that exploits behavioral features derived from WiFi signals for user authentication. As illustrated in Figure 3, our system takes as input time-series CSI measurements from WiFi links between smart IoT devices (e.g., voice assistants, smart TVs) for data preprocessing. Our system then examines the frequency components of CSI amplitude to determine the CSI segment containing user behavior. To capture the user's unique behavioral characteristics, our system also extracts relative amplitude information at each OFDM subcarrier, which captures the relative channel response between two WiFi links and is more stable. The amplitude and relative amplitude information are then calibrated with Z-score normalization to suppress the impacts of channel condition changes. Moreover, our system converts CSI amplitude and relative amplitude into frequency domain representations (i.e., STFT holograms) to characterize the moving speeds of different body parts.

After data preprocessing, we feed both standardized CSI amplitude/relative amplitude and STFT holograms into *Environment-independent User Authentication Model* for user authentication and activity recognition. Particularly, our system uses two CNN models with 3 convolutional layers to extract time and frequency domain features, which characterize both human identity and activity uniqueness. The time domain and frequency domain features together characterize behavioral characteristics such as gesture preferences, walking gait patterns, and movement speeds of torso and leg. Based on the extracted features, our system utilizes a user recognizer, a 2-layer fully-connected neural network, to learn non-linear biometric abstractions that amplify the user's unique characteristics and are robust to small-scale behavior variations. To deliver personalized customized services (e.g., recommending music channels, controlling home appliances), an activity recognizer based on a 2-layer fully-connected network is used to identify the user's activity.

To make the CNN models environment-independent, our system employs an unsupervised domain discriminator to remove the domain-specific information from the time and frequency domain features. The domain discriminator is optimized predict the domain (i.e., user location and environmental status) of both labeled data and unlabeled data. It seems to contradict with our goal of extracting domain-independent features. However, by simultaneously training the domain discriminator and the CNN models with an adversarial loss, the CNN models can gradually learn to extract features that are indistinguishable by the domain discriminator. At the same time, the domain discriminator also increases its capability in predicting domain labels. This two-player game eventually renders the extracted features invariant under different domains.

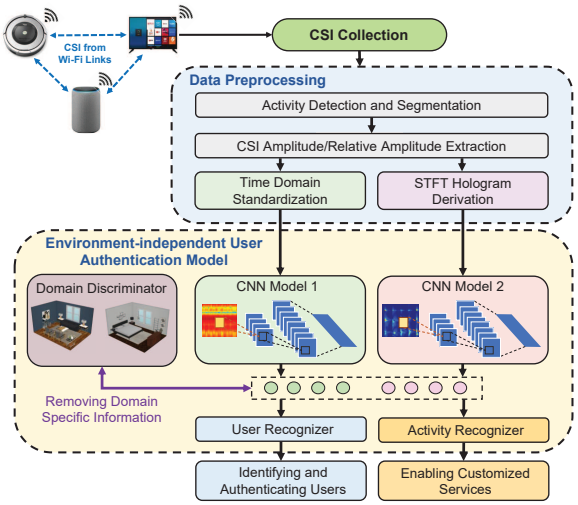


Fig. 3: Architecture of the proposed system on extracting environment-invariant behavioral features.

V. Human Behavior Detection and Data Segmentation

To authenticate users through their daily behaviors, it is necessary to first detect the presence of human activities and precisely segment the corresponding CSI measurements. In practice, CSI measurements are usually susceptible to the interferences of ambient radio frequency signals and hardware control mechanisms, which lead to high-frequency distortions in CSI and cause false detections. To circumvent this issue, we utilize time-frequency analysis for detecting human behaviors that mainly reside at a low frequency range [13]. Particularly, for each subcarrier, we calculate a spectrogram by applying Fast Fourier Transform to a sliding frame. Furthermore, we accumulate spectrograms across all the subcarriers on the link between the main antenna pair (i.e., 1st antenna in both transmitter and receiver) to ensure reliable behavior detection.

Figure 4 (a) and (b) show the accumulated spectrogram and the time-series CSI amplitude for three consecutive activities (i.e., walking to a seat, sitting down, fetching a document on a table). We observe that the accumulated spectrogram of the user behaviors exhibits high energy in low frequencies. We are thus motivated to use spectrogram energy below $100Hz$ to segment user behaviors. For each frame of the spectrogram, we calculate the average energy below $100Hz$. We then use a threshold-based approach to detect the starting point of an activity and then search for the ending pointing with the same threshold. Figure 4 (c) shows that we can correctly locate the starting/ending points of the behaviors, which validates the effectiveness of our threshold-based approach.

VI. Preprocessing in Time and Frequency Domains

A. Time-domain Standardization

Human behaviors can alter CSI amplitude/relative amplitude and produce distinctive time-series patterns. However, the channel condition changes can alter the gain offset of the wireless link, shifting the value distributions of amplitude/relative

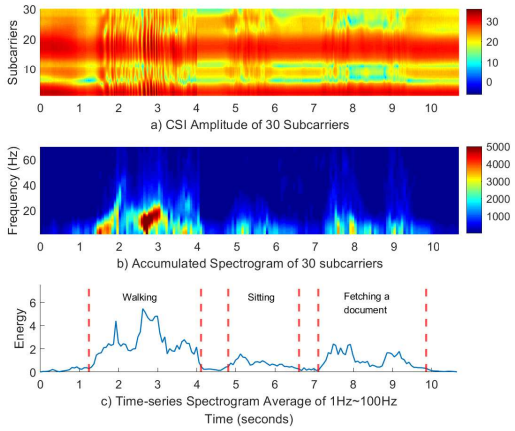


Fig. 4: Illustration of CSI amplitude of 30 subcarriers, accumulated spectrogram, and detected starting and ending points of three consecutive activities (i.e., walking to a seat, sitting down, and fetching a document on a table).

amplitude. To remove the unpredictable gain offset in each segment, we exploit Z-score normalization for data calibration:

$$H'_k = \frac{H_k - \mu_k}{\sigma_k}, \tag{2}$$

where H_k is the segmented data of either amplitude or relative amplitude from the kth subcarrier. μ_k and σ_k are the mean and the standard deviation, respectively. Such a standardization process can also increase the stability of the training process and improve the system's performance.

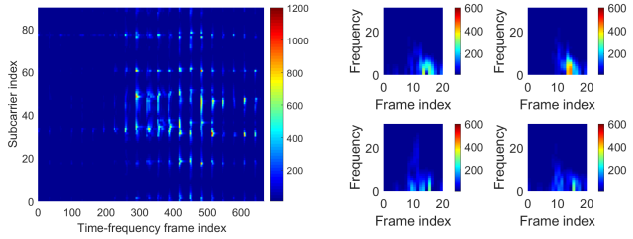
B. STFT Hologram Derivation

In addition to time-series data, we calculate spectrograms of both amplitude and relative amplitude to characterize the motion speeds of user behaviors [13]. To feed the high-dimension 3D spectrogram matrices (i.e., subcarrier, time, and frequency dimensions) to standard CNN with 2D kernels, we convert the spectrogram matrices into a 2D form. Particularly, we first flatten the spectrogram of each subcarrier into a 1D array and then stack the 1D arrays of all subcarriers to form a 2D matrix. We refer to this matrix as STFT hologram as it integrates information across time, frequencies, and subcarriers. The STFT hologram still preserves the time-series patterns in the spectrogram. We illustrate this with an example STFT hologram in Figure 5 (a), which is associated with a behavior of sitting down on a chair. The spectrograms are derived from CSI amplitude of 90 subcarriers (i.e., from 3 MIMO antenna pairs). As shown in Figure 5 (b), the $9 \sim 17$ frames of the spectrograms are associated with the $288 \sim 544$ frames of the STFT hologram, with the elements across the frequency dimension flattened. By transforming a 3D spectrogram matrix into such a 2D form, a standard CNN with 2D kernel can effectively learn features in the frequency domain.

VII. Environment-Independent User Authentication Model

A. Model Overview

To enable adaptive and practical user authentication, we develop a deep learning model and an unsupervised domain



(a) STFT hologram of CSI amplitude (b) Spectrograms of four subcarriers

Fig. 5: A STFT hologram of 90 subcarriers from 3 OFDM WiFi links (i.e., 30 subcarriers for each link) and the corresponding spectrograms of the subcarrier 1, 20, 40, 60.

adaptation strategy to learn domain-invariant features from the standardized CSI amplitude/relative amplitude and STFT hologram. Figure 6 illustrates the model architecture. The deep learning model takes both labeled data X and unlabeled data X' as input. The input data are first mapped into a set of low-rank behavioral features Z by using the feature extractor, which consists of two CNN models to process time-domain data (i.e., amplitude/relative amplitude information) and frequency-domain data (i.e., STFT holograms), respectively. Based on the extracted features, a user recognizer (i.e., a fully-connected neural network) can predict the user's identity $\hat{Y}_u$. In addition, an activity recognizer is used to obtain the activity types (i.e., $\hat{Y}_a$) of all the input data. To remove the domain-specific information entangled in Z, a domain discriminator is trained to predict the domain label $\hat{Y}_d$ (i.e., user location or environmental status), which seems to contradict with our objective of deriving domain-invariant features. However, by using an adversarial loss, the CNN models are guided to derive features that indistinguishable by the adversarial network, while at the same time, maximize the performance of the user recognizer and the activity recognizer. Through this minimax game, the derived deep learning model can finally extract domain-independent features that characterize both identity and activity uniqueness. Besides adapting to the location and environment changes, our domain adaptation strategy may be extended to address domain variations caused by the movement or replacement of WiFi devices. We leave the detailed study of this case to our future work.

B. Feature Extractor

The feature extractor consists of two CNN models that learn a set of behavioral features to characterize both human identity and activity uniqueness. As illustrated in Figure 7, the CNN models consist of a 3-layer stacked CNN. In each layer of the CNNs, a convolutional layer with 2D kernels is used to calculate 2D feature maps that characterize behavioral uniqueness of different activities/users. In addition, a batch normalization layer is used to calibrate the input data, aiming to mitigate small-scale input variations, and a dropout layer is utilized to prevent over-fitting. The 2D feature maps are then flattened and compressed with three fully-connected layers. Particularly, given input CSI amplitude/relative amplitude X_t

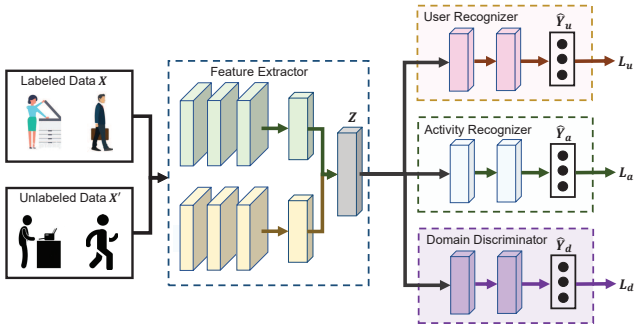


Fig. 6: Domain adaptation strategy to derive the environment-independent user authentication model.

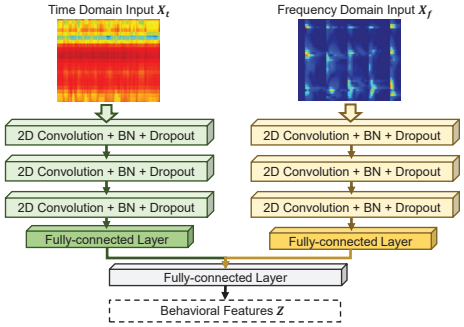


Fig. 7: Architecture of the deep learning-based feature extractor to extract behavioral features by taking inputs from both the time and frequency domains.

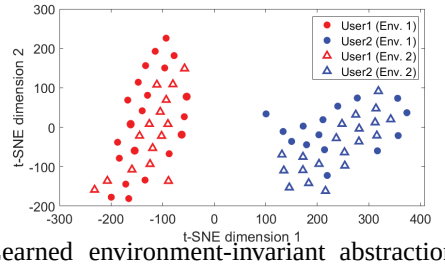


Fig. 8: Learned environment-invariant abstractions of two users (marked in red and blue) picking up a remote controller under two environmental statuses (circle and triangle markers).

and STFT holograms X_f, the CNN models map the input into behavioral features Z as follows:

$$Z = CNN(X_t, X_f, \Theta), \tag{3}$$

where Θ represents a set of learnable parameters (i.e., weights and biases) in the feature extractor. The activations are all Leaky ReLU.

C. User Recognizer and Activity Recognizer

Based on the extracted behavioral features Z, two fully-connected neural networks are used as classifiers to perform user identification and activity recognition. Both the user and the activity recognizers have the same architecture so we omit the subscripts u and a in the symbols for simplicity. The fully-connected networks further derive non-linear feature abstractions to characterize the behavioral biometrics/activity patterns. Based on the abstractions, a SoftMax layer is used to predict the user identities or activity types. Given the input feature Z, the mapping function is defined as:

$$\hat{Y} = G(Z; \Phi), \tag{4}$$

where $G(\cdot)$ represents the mapping function and Φ is the a of learnable parameters in the neural network.

To train the deep learning model (i.e., feature extractor and user/activity recognizer), we use both labeled and unlabeled data. For the labeled data, we calculate the cross-entropy loss between the predictions $\hat{Y}$ and the ground truth labels Y. For unlabeled data, we also calculate the entropy of predictions $\hat{Y}'$ as the loss, which reduces the uncertainty when predicting on the unlabeled data. The loss function is defined as:

$$L = L_{cls}(Y, \hat{Y}) + H(\hat{Y}'), \tag{5}$$

where $L_{cls}(\cdot, \cdot)$ represents the cross-entropy loss function. $H(\cdot)$ denotes the entropy. L could be either L_u or L_a. Note that our system simultaneously optimizes the user and the activity recognizers so that the behavioral features Z characterize both identity and activity uniqueness.

D. Domain Discriminator

We aim to use domain adversarial training [15] to derive a mapping shared under different environmental statuses or in different locations. The key component of the domain adaptation technique is a domain discriminator that is used in the training process to force the feature extractor to derive domain-independent features. Particularly, the domain discriminator consists of 2 fully-connected layers using Leaky ReLUs as the activation functions. By taking the behavioral features Z as input, the domain discriminator acquires the environmental status/location label as:

$$\hat{Y}_d = G_d(Z, \Omega), \tag{6}$$

where $G_d(\cdot)$ represents the mapping function and $\hat{Y}_d$ is the domain label. $\hat{Y}_d$ represents either the environmental status and the location label based on the domain adaptation task. Ω is a set of trainable parameters in the adversarial network. To train the adversarial network for recognizing the domain, we define the domain loss as:

$$L_d = L_{cls}(Y_d, \hat{Y}_d), \tag{7}$$

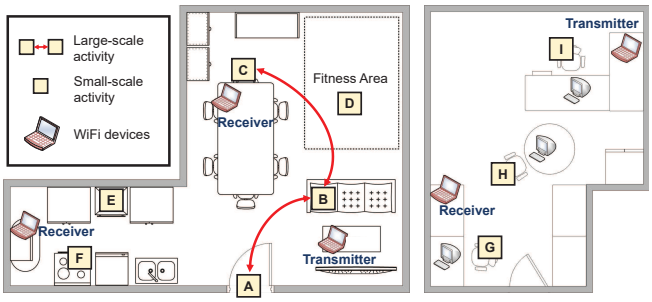
where Y_d is the set of domain labels, which can be passively collected.

E. Unsupervised Domain Adversarial Training

The objective of the domain discriminator seems to contradict with our goal of location- and environment-independent user authentication and activity recognition. But with a carefully designed loss function, we can use the domain discriminator to guide the feature extractor on learning domain-invariant features. The key is a negative factor $-\lambda$ applied to the domain loss so that the feature extractor is trained to maximize the loss of the domain discriminator. We define the adversarial loss for optimizing the feature extractor as:

$$L_f = L_u + \alpha L_a - \lambda L_d, \tag{8}$$

where L_u, L_a, and L_d are the user loss, activity loss, and domain loss, respectively. α and λ are the weighting parameters. Particularly, λ controls the trade-off between the transferability and the distinctiveness of the learned features. During the adversarial training process, we take turns to update Θ, $\{\Phi_u, \Phi_a\}$, and Ω.



(a) Residential Apartment (b) Office

Fig. 9: Experimental setups and the illustration of the activities in an residential apartment and an office.

To examine the proposed domain adaptation strategy, we conduct a preliminary experiment by asking two volunteers to perform the same activity (i.e., picking up a remote controller) under two environmental statuses, i.e., Env. 1 and Env. 2, where the positions of a desk and two chairs are different. We visualize the user abstractions (i.e., outputs of the last layer in the user recognizer excluding the SoftMax layer) on a 2D space with t-SNE as shown in Figure 8. Note that the ground truth labels for Env. 2 are not used for training. We can observe that the abstractions of the two users can form two clearly separate clusters. Furthermore, for different environmental statues, the abstractions of the same user fall into the same cluster, showing that the abstractions are environment-independent. These results validate the effectiveness of the proposed domain adaptation strategy.

VIII. Performance Evaluation

A. Experimental Setup and Methodology

Devices and Network. We use two commercial laptops (i.e., Dell E6430) to emulate IoT devices in smart environments. One laptop is used as the transmitter and the other laptop is used as the receiver. Both laptops are equipped with 3 MINI PCI-E internal antennas and an Intel 5300 WiFi NIC which internally tracks and reports CSI at 30 subcarriers [16]. Our system can also be extended to use a wider bandwidth (e.g., 802.11ac, 802.11ax) and more WiFi devices to achieve enhanced performance. We extract the CSI amplitude on the 9 OFDM links between the transmitter and the receiver and calculate the relative amplitude between every two links. The WiFi packet transmission rate is set to 1000 $pkts/s$ for extracting fine-grained frequency domain features.

Data Collection. Experiments are conducted in a residential apartment and an office with the size of $33ft \times 17ft$ and $21ft \times 12ft$, respectively. Figure 9 illustrates the positions of the two laptops emulating IoT devices. For the residential apartment, we place the receiver on two locations to collect behavior data in the living room and the kitchen. A total of 10 representative activities (20 rounds for each) are performed by 10 and 5 users in these two scenarios. The details of these activities are shown in Table I.

Specifically, to evaluate the environmental independency of our approach, for both residential apartment and campus office, we collect data under 3 different environments in which the furniture placements are different. For the residential apartment, we change the positions of 1 sofa, 1 microwave oven, 3 cabinets, and 5 chairs. While for the office, we move 3 desks, 3 chairs, and some books placed on the desks. Each piece of furniture is moved at least $3ft$ for emulating large-scale environmental changes. Particularly, when collecting data in the office, we ask 5 users to perform 3 activities (i.e., G, H, I) with location variations. The users conduct each activity at 4 different proximate locations at least one foot away from each other. In total, we collect 4,079 behavior segments performed by 10 users in the apartment and 3,513 behavior segments performed by 5 users in the office.

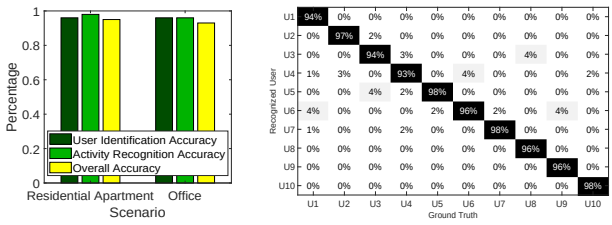
TABLE I: Detailed activities performed.

Code	Activity	Code	Activity
A→B	Walking (trajectory 1)	E	Operating on the oven
B→C	Walking (trajectory 2)	F	Using the stove
B	Picking up a remote control	G	Sitting in a seat
C	Sitting in a chair	H	Stretching the body
D	Exercising	I	Typing on a keyboard

We separate the collected data into source dataset (i.e., labeled data) and target dataset (i.e., unlabeled data), with the locations/environments referred as the source and target locations/environments. We use N_s : N_t to present the number of locations/environments involved in the source (i.e, N_s) and the target datasets (i.e, N_t) when presenting the results. Half of the target dataset is used for unsupervised training and the other half is used for testing. We refer the environments associated with the source dataset as the source locations/environments and that corresponding to the target dataset as the target locations/environments.

Baseline Methods. We compare our approach with a CNN model only consisting of the feature extractor and the user recognizer (i.e., described in Section VII-C) without applying domain discriminator. In addition, we build another baseline model based on transferable component analysis (TCA) [17]. TCA aims to learn a set of transferable representations based on unlabeled data from the source datasets and the target training datasets. Specifically, we extract time and frequency domain features (used in our previous work [8]) from source and target datasets and learn a set of transferable representations using TCA. Based on these representations, an support vector machine classifier is used for user identification/activity recognition on target testing datasets.

Evaluation Metrics. We define four different evaluation metrics: *user identification/activity recognition accuracy* is the percentage of predicted user identities/activities are correctly recognized among all user behaviors; *confusion matrix* visualizes the percentage of each user's behaviors being identified among all users (i.e., the correct user and the other users); *true positive rate (TPR)* is the percentage of a legitimate user's behaviors that are correctly accepted among all behaviors from the legitimate user *false positive rate (FPR)* is the percentage of the adversary's behaviors being mistakenly accepted among all behaviors of the adversary.



(a) Overall performance (b) User Identification in Apartment

Fig. 10: Performance of environment-dependent user identification and activity recognition.

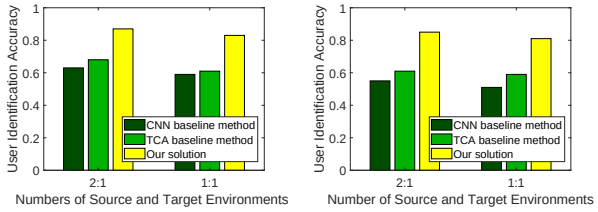
B. Performance of Environment-dependent User Identification/Activity Recognition

We first show the performance of our system on user identification and activity recognition using the randomly split training and testing dataset of roughly same size (referred to as environment-dependent). Figure 10 (a) shows that for both scenarios, our system achieves over 93% overall accuracy on simultaneously recognizing both the identity and activity of user behaviors. Figure 10 (b) gives the confusion matrix for identifying 10 users in the apartment. We can find that our system can achieve high accuracy on user identification. The results demonstrate that our system is effective in both user identification and activity recognition, showing its potential for enabling customized services.

C. Performance of Environment-independent User Identification/Activity recognition

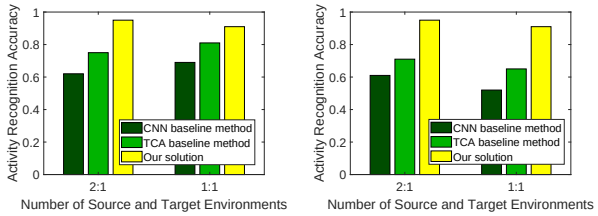
We use all source dataset and half of the target dataset for training. The other half of the target datasets are used for testing. Figure 11 (a) gives the user identification performance under large-scale environmental changes in the residential apartment. Our approach can achieve 87.3% and 83.6% user identification accuracies for two different settings. Particularly, when training with 2 source environments and testing on 1 target environment, the accuracies are 24.2% and 18.7% higher than the CNN baseline model and the TCA-based method, respectively. From Figure 11 (b), we find that our approach also has the highest user identification accuracy in the office, with 85.2% and 83.6% accuracies under the 2 settings. Furthermore, we observe that the user identification accuracies are slightly higher in the residential apartment. This is because the apartment does not have high-power WiFi infrastructures (e.g., campus-wide WiFi station) that create strong interference. The above results confirm that the proposed approach can realize environment-independent user identification.

As shown in Figure 12, for activity recognition, the proposed domain adaptation approach still outperforms the two baseline methods in both scenarios. Particularly, in the office environment, our approach is at least 22.4% and 10.2% higher than the accuracies of CNN and TCA baselines, respectively. We also find that the activity recognition accuracies are slightly higher than user identification accuracies in both scenarios, indicating that domain adaptation for the user identification task is more difficult. Overall, the proposed domain adaptation method can achieve robust activity recognition and user identification under large-scale environment variations.



(a) Residential Apartment (b) Office

Fig. 11: Performance of environment-independent user identification.



(a) Residential Apartment (b) Office

Fig. 12: Performance of environment-independent activity recognition.

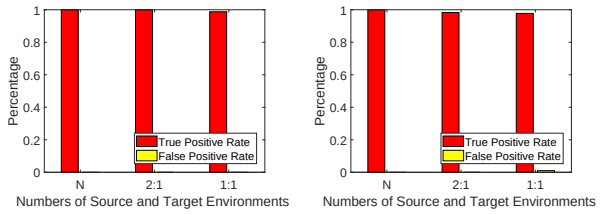
D. Performance of Environment-independent Spoofer Detection

To detect unauthorized users, we train a user recognizer (i.e., presented in Section VII-C) to differentiate a legitimate from all other users. The user recognizer then serves as a spoofer detection component. In the training phase, we randomly select 2 users as spoofers to guide the user recognizer on detecting the legitimate user. In addition, we select the data of the same set of users from the target dataset to extract environment-invariant features. During testing, we select 2 different users acting as spoofers to conduct random and mimic attacks, using their behavior data from both source and target datasets.

Figure 13 (a) and (b) show the spoofer detection performance under random attacks in the apartment and the campus office. We find that our system can achieve close to 100% TPR with a low FPR below 1% for all settings. These results confirm that the random activities of the attacker can hardly create similar behavioral biometrics as the legitimate user, and thus the system can reliably defend against random attacks. Figure 14 shows the performance of our system under mimic attacks. We can find that the system has FPRs lower than 2% under all dataset settings, with over 94% and 91% TPR for the apartment and the office, respectively. The results show that our system is effective in defending against both random attack and mimic attack. This is because even for the same activity, our approach can extract unique behavioral characteristics of the legitimate users.

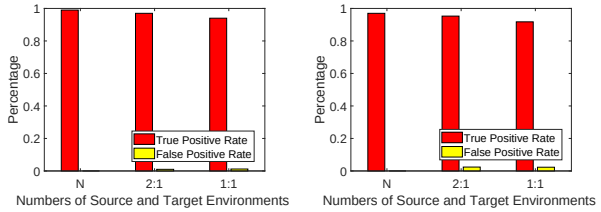
E. Performance of Location-independent User Identification/Activity Recognition

We combine all source and half of the target datasets for training. The other half of the target dataset is used for testing. From Figure 15 (a), we find that the proposed domain adaptation approach can achieve 91.3%, 84.5%, and 81.2%



(a) Apartment (b) Campus Office

Fig. 13: Performance of environment-independent spoofing detection against random attacks. N denotes the case when the attacks are launched in the source environment.



(a) Apartment (b) Campus Office

Fig. 14: Performance of environment-independent spoofing detection against mimic attacks. N denotes the case when the attacks are launched in the source environment.

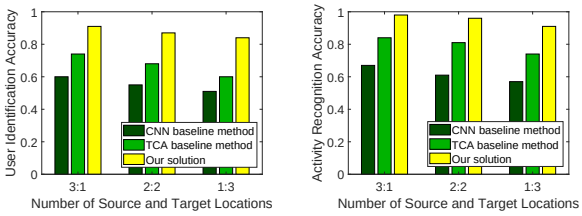
user identification accuracies given 3 source locations, which are at least 17.2% higher than the two baseline approaches. In addition, under the source-target ratio of $3:1$, our approach is 25.3% better than the TCA baseline method with the second highest accuracy. These results confirm the superiority of our approach to learn location-invariant features. Figure 15 (b) presents the performance of location-independent activity recognition. We find that our approach achieves over 91.3% under all settings, which are at least 18% higher than that of the TCA baseline method. The results show that the proposed unsupervised domain adaptation approach can effectively mitigate the small-scale location variations during daily behaviors.

IX. Conclusion

In this paper, we develop a device-free user authentication system by extracting unique behavioral characteristics captured by the CSI measurements in WiFi signals. Unlike existing WiFi-based user authentication schemes, our system aims to address two practical problems, small-scale location variations and large-scale environment changes, which lead to significant change of the CSI patterns and thereby cause profile mismatches. To realize such an environment-independent system, we design an unsupervised domain adaptation strategy to remove the location and environment-specific information entangled in CSI measurements to build an environment-independent model for user identification and activity recognition. Extensive experiments showed that the proposed system has the capability of authenticating users through daily behaviors under various scales of location variations and environmental changes.

X. Acknowledgement

This work was partially supported by the Siemens Graduate Assistantship and NSF Grant CNS1801630, CNS1820624, CNS1814590, CCF2028876.



(a) User Identification (b) Activity Recognition

Fig. 15: Performance of location-independent user identification/activity recognition.

References

[1] "A stolen laptop contained data for more than 114,000 patients at truman medical centers," 2019, bit.ly/2sCHPsn.

[2] J. S. Edu, J. M. Such, and G. Suarez-Tangil, "Smart home personal assistants: a security and privacy review," *arXiv:1903.05593*, 2019.

[3] S. Yi, Z. Qin, E. Novak, Y. Yin, and Q. Li, "Glassgesture: Exploring head gesture interface of smart glasses," in *Proceedings of International Conference on Computer Communications (IEEE INFOCOM)*, 2016, pp. 1–9.

[4] M. Muaaz and R. Mayrhofer, "Smartphone-based gait recognition: From authentication to imitation," *IEEE Transactions on Mobile Computing (IEEE TMC)*, vol. 16, no. 11, pp. 3209–3221, 2017.

[5] J. Zhang, B. Wei, W. Hu, and S. Kenhere, "Wifi-id: Human identification using wifi signal," in *Proceedings of International Conference on Distributed Computing in Sensor Systems (IEEE DCOSS)*, 2016, pp. 75–82.

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

[7] C. Lin, J. Hu, Y. Sun, F. Ma, L. Wang, and G. Wu, "Wiau: An accurate device-free authentication system with resnet," in *Proceedings of Annual IEEE International Conference on Sensing, Communication, and Networking (IEEE SECON)*, 2018, pp. 1–9.

[8] C. Shi, J. Liu, H. Liu, and Y. Chen, "Smart user authentication through actuation of daily activities leveraging wifi-enabled iot," in *Proceedings of ACM International Symposium on Mobile Ad Hoc Networking and Computing (ACM MobiHoc)*, 2017, pp. 1–10.

[9] H. Kong, L. Lu, J. Yu, Y. Chen, L. Kong, and M. Li, "Fingerpass: Finger gesture-based continuous user authentication for smart homes using commodity wifi," in *Proceedings of ACM International Symposium on Mobile Ad Hoc Networking and Computing (ACM MobiHoc)*, 2019, pp. 201–210.

[10] S. Roy, U. Roy, and D. Sinha, "Identifying soft biometric traits through typing pattern on touchscreen phone," in *Annual Convention of the Computer Society of India (Springer CERN)*, 2018, pp. 546–561.

[11] A. Acar, H. Aksu, A. S. Uluagac, and K. Akkaya, "Waca: Wearable-assisted continuous authentication," in *Proceedings of IEEE Security and Privacy Workshops (IEEE SPW)*, 2018, pp. 264–269.

[12] J. Ranjan and K. Whitehouse, "Object hallmarks: identifying object users using wearable wrist sensors," in *Proceedings of ACM International Joint Conference on Pervasive and Ubiquitous Computing (ACM UbiComp)*, 2015, pp. 51–61.

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

[14] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," *ACM SIGCOMM Computer Communication Review*, vol. 40, no. 4, pp. 159–170, 2010.

[15] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *The Journal of Machine Learning Research (JMLR)*, vol. 17, no. 1, pp. 2096–2030, 2016.

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

[17] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Transactions on Neural Networks*, vol. 22, no. 2, pp. 199–210, 2010.