WiONE: One-Shot Learning for Environment-Robust Device-Free User Authentication via Commodity Wi-Fi in Man–Machine System

Yu Gu, *Senior Member, IEEE*, Huan Yan, Mianxiong Dong, *Member, IEEE*, Meng Wang, Xiang Zhang, Zhi Liu, *Senior Member, IEEE*, and Fuji Ren, *Senior Member, IEEE*

***Abstract*— User authentication is the first and most critical step in protecting a man–machine system from a malicious spoofer. However, security and privacy are just like the two sides of one coin, hard to see both at the same time, especially by the current mainstream credential- and biometric-based approaches. To this end, we propose WiONE, a safe and privacy-preserving user authentication system leveraging the ubiquitous Wi-Fi infrastructure by exploring "*how you behave*" rather than "*who you are*". The key idea is to apply deep learning to user physical behavior captured by Wi-Fi channel state information (CSI) to identify legitimate users while rejecting spoofers. The design of WiONE faces two challenges, namely, how to capture the subtle behavior, such as a keystroke on CSI, and how to mitigate the heavy environment-specific training required by deep learning. For the former, we design a behavior enhancement model based on the Rician fading to highlight the behavior-induced information by suppressing the behavior-unrelated information on channel response. For the latter, we develop a behavior characterization method tailored for the prototypical networks to facilitate the extraction of the domain-independent behavioral features and enable one-shot recognition of a new user in a new environment. Numerous experiments are conducted in several real-world environments, and the results show that WiONE outperforms its state-of-the-art rivals in authentication performance with much less training effort.**

***Index Terms*— Channel state information (CSI), cross-domain learning, deep learning, one-shot learning, Rician fading, user authentication, Wi-Fi.**

Manuscript received December 4, 2020; revised January 21, 2021; accepted January 24, 2021. Date of publication March 24, 2021; date of current version May 28, 2021. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB0803403, in part by the National Natural Science Foundation of China (NSFC) under Grant 61772169, in part by the Fundamental Research Funds for the Central Universities under JZ2018YXQN0121, and in part by the JSPS KAKENHI under Grant JP20F20080. *(Corresponding author: Zhi Liu.)*

Yu Gu, Huan Yan, Meng Wang, and Xiang Zhang are with the School of Computer and Information, Hefei University of Technology, Hefei 230006, China (e-mail: hfut_bruce@hfut.edu.cn; yanhuan@mail.hfut.edu.cn; mengw512@mail.hfut.edu.cn; zhangxiang@mail.hfut.edu.cn).

Mianxiong Dong is with the Muroran Institute of Technology, Muroran 0508585, Japan (e-mail: mx.dong@csse.muroran-it.ac.jp).

Zhi Liu is with the Department of Computer and Network Engineering, The University of Electro-Communications, Tokyo 182-8585, Japan (e-mail: liu@ieee.org).

Fuji Ren is with the Department of Information Science and Intelligent Systems, Tokushima University, Tokushima 770-8501, Japan (e-mail: ren@is.tokushima-u.ac.jp).

Digital Object Identifier 10.1109/TCSS.2021.3056654

I. Introduction

AUTHENTICATION, the act of proving the identity of other people, remains a great challenge [1]–[3]. On the one hand, we can recognize an acquaintance through his biometrics, such as the face, voice, or even behavior metrics, such as gait. On the other hand, we may need a token to identify a stranger. As we step into the informatics society, a similar challenge arises: *how can a computer recognize its legitimate users while rejecting spoofers?* A human being is born with five senses, i.e., sight, hearing, smell, taste, and touch. However, a computer, especially a primitive one, lacks such abilities to capture biometrics. Therefore, a token, also known as a credential, is assigned to each user for authentication in current man–computer systems, such as automated teller machines (ATMs), access control systems (ACSs), or desktop computers.

The blossom of computer technology dramatically increases the number of accounts possessed by one user in recent years. A study in 2006 involving a million users found that each user had about 25 accounts on average. This number reached 80 in 2018 [4], posing a constantly repeated question in our daily life: *what is the password?* Consequently, 61% users use one unified password for all accounts [5], i.e., putting all the eggs in one basket. It will significantly increase the risk of all accounts being breached. The leak of highly sensitive privacy data, such as financial and social information, could cause huge economic loss to both individuals and society or even raise social disasters, e.g., the WikiLeaks [6]. The average data breach exposes 25 575 sensitive consumer records and carries a total cost of $3.92 million on a global scale [7]. Secure user authentication, thus, becomes a must in modern society.

Therefore, there is a recent trend of exploiting biometrics, such as fingerprint and facial recognition, for a more secure way of user authentication. Unlike credential that establishes user authentication based on *"what he possesses"*, the biometric recognition more relies on *"who he is"* and, thus, constitutes a more reliable and efficient approach. However, it requires the computer to be equipped with specialized devices, such as fingerprint sensors or facial cameras to collect biological data, raising the cost and availability issues

and offsetting its merits in real-world applications. Moreover, the privacy issue becomes another major concern recently.

In this article, we ask the question: can we design a user authentication approach for human–machine systems, such as ATMs, ACSs, or desktop computers that: 1) *employs a unified password for all users instead of a credential for each user*; 2) *contactlessly recognizes legitimate users and spoofers without deploying any specialized devices*; and 3) *robustly works in a wide range of environments?* To this end, we propose WiONE, a credential-free and device-free user authentication system leveraging one-shot learning on the ubiquitous Wi-Fi signals. The key idea is to perform deep learning on the personalized behavioral information of keystrokes derived by the Wi-Fi channel state information (CSI) to identify users while rejecting spoofers. WiONE faces two fundamental design challenges at the signal-level and data-level, namely, how to capture a subtle human behavior like keystrokes using CSI and how to mitigate the environment-specific training effort of deep learning.

For the signal-level challenge, we propose a behavior enhancement model based on the Rician fading to highlight the behavior-induced information by suppressing the behavior-irrelevant information. More specifically, the sensitivity of Wi-Fi CSI over human motion plays an important role in recovering a subtle behavior, such as keystrokes. Unlike our rivals mostly relying on the empirical study for tuning the antenna layout, we propose a Rician-K factor-based theoretical model to enhance the sensing granularity. The key is to highlight the motion-induced information by suppressing the motion-unrelated information on channel response. We amplify the ratio of the non-line-of-sight (NLOS) component corresponding to motions along with signal propagation over channel response via weakening the line-of-sight (LOS) part. Both theoretic analysis and experimental results have verified that the proposed model can significantly improve the sensitivity of CSI over behaviors with little implementation cost.

For the data-level challenge, we design a behavior learning framework based on the prototypical networks to facilitate the extraction of the domain-independent behavioral features and enable one-shot user authentication in a new environment. In particular, deep learning usually suffers from several inherent drawbacks, such as heavy environment-specific training. To mitigate this issue, one-shot learning has emerged recently as a promising approach [8]. Its key insight is that we could leverage knowledge derived from the previously trained environments for a new environment to enable one-shot recognition. In other words, only one sample at the new environment is enough to extract discriminate features for recognition via bridging between the new environment and previously trained environments. Therefore, we design a learning framework leveraging the state-of-the-art one-shot learning method, i.e., prototypical networks [9], to recognize legitimate users while rejecting spoofers via the enhanced CSI.

To evaluate the performance of WiONE, we prototype it with commodity Wi-Fi devices and conduct extensive experiments in real-world environments. Empirical results show that WiONE achieves on average 94.3% authentication accuracy in real environments. We also verify that WiONE is robust to various system settings and environmental conditions, such as password, positions of the keyboard, training size, and experimental sites. Moreover, some attractive features of WiONE have been revealed. For instance, counterintuitively, a simple password works better than a complex one since users often pause to think or make wrong inputs for the latter.

The main contributions of this article are summarized as follows.

1) We propose WiONE, a behavioral biometrics-based user authentication scheme leveraging the prototypical networks with enhanced CSI to enable one-shot recognition of a new user in a new environment.
2) We design a behavior enhancement model based on the Rician fading to highlight the behavior-induced information and suppress the behavior-irrelevant information on channel response. To the best of our knowledge, it is the first quantitative model on leveraging the Rician fading for improving the sensitivity of channel response to human behavior.
3) We develop a behavior learning framework leveraging the prototypical networks to facilitate the extraction of the domain-independent behavioral features and enable one-shot user authentication in a new environment.
4) We conduct numerous comparative experiments in several real-world environments with some state-of-the-art Wi-Fi sensing systems to evaluate the performance of WiONE. The empirical results show that WiONE outperforms its rivals in authentication performance with much less training effort.

II. Related Work

WiONE essentially is a hybrid research of two areas, i.e., user authentication and Wi-Fi-based gesture recognition. In this section, we will introduce the existing work on both topics to outline the motivation of our research.

A. User Authentication

For thousands of years, people recognize each other through human body characteristics, such as face and voice. However, it remains a great challenge to enable a computer to recognize people, i.e., user authentication. Current research on this topic can be roughly divided into two categories, i.e., credential-based and biometric-based.

1) Credential-Based User Authentication: The credential was first introduced in 1967 on the first ATM invented by John Shepherd-Barron for a Barclays Bank in London [10]. The carbon 14 mark of the inserted check, no plastic card back then, would be detected and matched to a six-digit credential for user authentication. With the fast spread of ATMs over the world, it becomes the most prevailing technique for user authentication. Later, some varieties of credentials are proposed for different scenarios. For instance, a graphical password composed of a sequence of user-generated pictures was used instead of digit numbers [11]. Recently, as wearable devices become more and more popular, Hutchins *et al.* [2]

developed Beat-PIN, a user authentication system for wearable devices equipped with touch sensors using different beats/rhythms of tapping rather than numbers or letters.

However, the credential can be easily leaked to others and, thus, is particularly vulnerable to malicious attackers. Therefore, biometric-based authentications using fingerprint, iris, and face have risen recently as efficient alternative solutions.

2) Biometric-Based Solution: Biometric-based user authentication leverages unique and distinctive physiological features, such as fingerprint [12], face [13], behavioral features, such as gait [14], or keystroke [15] to recognize legitimate users and deny illegal access to the rendered systems.

Archeological evidence reveals that fingerprints have been used as a form of identification since 7000 years ago [16]. Both the immutability and the uniqueness properties have determined fingerprint as one of the most reliable sources for people identification [17].

While physiological features are usually used for one-time authentication, behavioral characteristics can be explored for continuous user authentication. Gaines *et al*. [3] were among the first to apply keystroke dynamics for continuous authentication. They defined the digraph and ternary graph to represent keystroke characteristics. Later, Kenneth Revett [18] combined keystroke dynamics with machine learning to distinguish legitimate users from attackers. Zheng *et al*. [19] exploited mouse dynamics for user authentication.

B. Wi-Fi-Based Gesture Recognition

With the fast expansion of wireless communications, Wi-Fi becomes the most pervasive communication technology in indoor environments. Recently, there is a trend of exploring Wi-Fi for ubiquitous device-free gesture recognition [20]–[23].

Wang *et al.* [24] proposed WiFall that detects the fall of a human subject in an indoor environment using CSI values. Zhang *et al*. [25] proposed SMARS that exploits ambient radio signals to recognize sleep stages and assess sleep quality. Ali *et al*. [26] proposed Wikey that captures user keystrokes by extracting unique patterns from the time series of CSI profiles. Abdelnasser *et al.* [27] proposed WiGest that leverages changes in the Wi-Fi signal strength to sense in-air hand gestures around the user's mobile device. Recently, Meng *et al.* [28] pushed the research one step further by proposing WindTalker. WindTalker exploits the strong correlation between CSI fluctuations and keystrokes to infer a user's number input on mobile devices.

Furthermore, deep learning-based methods are also widely used in Wi-Fi-based gesture recognition. Specifically, Li *et al.* [29] proposed the WiHF system that uses a dual-task deep neural network with split and splicing schemes for gesture recognition and user authentication. Yang *et al.* [30] proposed a novel deep Siamese representation learning architecture for one-shot gesture recognition. Han *et al.* [32] proposed a fine-grained deep adaptation network-based gesture recognition scheme. The system first uses a generative adversarial network (GAN)-based method [31] for data augmentation and then uses domain adaptation based on the multikernel maximum mean discrepancy scheme for cross-domain gesture recognition [32].

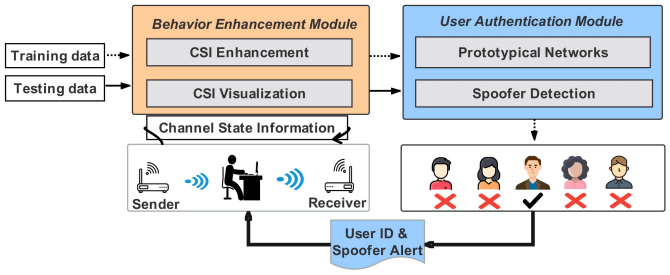


Fig. 1. System overview.

Recently, Kong *et al.* [33] proposed a pioneering work FingerPass for continuous user authentication via finger gestures in smart homes using commodity Wi-Fi. Though we target a different scenario, FingerPass constitutes a major inspiration for our work. Compared to the phase profile used in Fingerpass, we explore the amplitude profile, which is more resilient to environmental interferences. Moreover, we build a theoretical model that can significantly improve the sensing granularity via blocking the direct signal propagation components based on the Ricean distribution. This model can be seamlessly integrated into similar Wi-Fi-based systems with little implementation cost. Finally, though we both rely on deep learning for user authentication, we design a novel mapping scheme to preserve the behavioral features in the energy and spectrum domains through visualizing the channel data into time-series images. This technique allows us to better preserve the behavioral information.

III. System Design

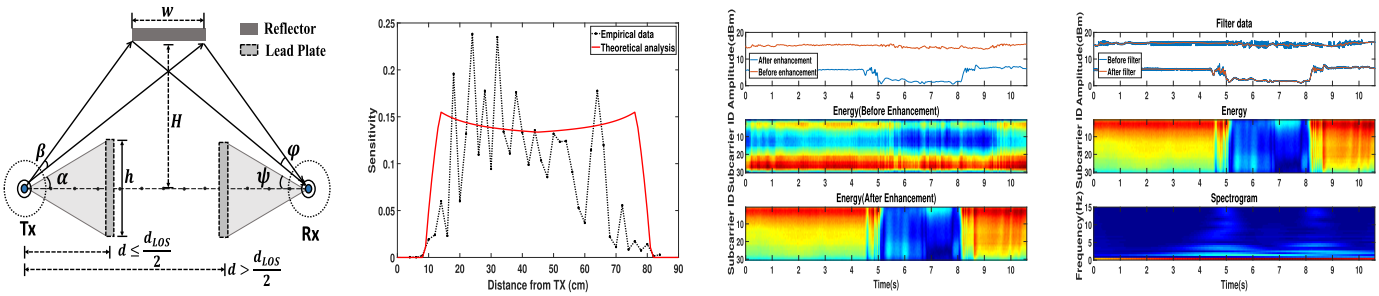
A. System Overview

Fig. 1 shows an overview of our WiONE system, which consists of two major modules: behavior enhancement module (BEM) and user authentication module (UAM). The first module focuses on enhancing the CSI to improve its sensitivity to human subtle behaviors, such as keystrokes, while the second module concentrates on one-shot user authentication via constructing the prototypical networks on the enhanced CSI data.

B. Behavior Enhancement Module

In this module, we first introduce how to collect the CSI where signal attenuation may reflect human behaviors. Then, we leverage the Ricean fading to enhance CSI for better capturing the subtle behaviors. The collected CSI representing the variation of channel response induced by behaviors is preprocessed to reduce the background noise and data dimension. Finally, we design a visualization mechanism to map the enhanced CSI data on all subcarriers into a series of heatmap images, which allows us to use state-of-the-art learning methods, such as the one-shot learning for user authentication. The definitions of variables in the BEM are summarized in Table I.

1) CSI Collection: CSI describes the signal's attenuation on its propagation paths, such as scattering, multipath fading



(a) (b) (c) (d)

Fig. 2. (a) Top view of the experimental setup. (b) Example validating our theoretical analysis with actual measurements. (c) CSI enhancement. (d) CSI filtering and visualization.

TABLE I
DEFINITIONS OF VARIABLES IN THE BEM

Variable	Definition
δ	System sensitivity to the human behavior
d	The distance between T_x and lead plate
α	The angle of occlusion area (T_x side)
β	The angle of reflection area (T_x side)
ψ	The angle of occlusion area (R_x side)
ϕ	The angle of reflection area (R_x side)
w	The maximum width of the human body
h	The shielding width of the lead barrier
H	Vertical distance between the person and the LOS path
P_0	Initial power of the transmitter
P_T	The power actually scattered into the environment
P_d	The power of dynamic component of received signal
P_R	Received power

or shadowing fading, and power decay over distance. In the frequency domain, it can be characterized as

$$\vec{Y} = \vec{H} \cdot \vec{X} + \vec{N} \tag{1}$$

where $\vec{Y}$ and $\vec{X}$ are the received and transmitted signal vectors, respectively. $\vec{N}$ is the additive white Gaussian noise, and $\vec{H}$ is the channel matrix representing CSI information.

The Wi-Fi spectrum is divided into multiple orthogonal subcarriers. For each subcarrier, the channel frequency response (CFR) can be expressed as

$$H_i = L_i + j \cdot Q_i = |H_i| e^{j \angle H_i} \tag{2}$$

where i is the subcarrier index. L_i and Q_i are the CFR. $|H_i|$ and $\angle H_i$ are the amplitude and phase of the ith subcarrier, respectively. Thus, the dimension of CFR matrix H is $N_t \times N_r \times N_s \times T$. N_t, N_r, and N_s are the numbers of transmitting antennas, receiving antennas, and subcarriers, respectively. In our WiONE system, $N_t = 1$, $N_r = 3$, and $N_s = 30$, and the CSI matrix is described as follows:

$$\begin{bmatrix} \text{CSI}_{1,1} & \text{CSI}_{1,2} & \dots & \text{CSI}_{1,30} \\ \text{CSI}_{2,1} & \text{CSI}_{2,2} & \dots & \text{CSI}_{2,30} \\ \text{CSI}_{3,1} & \text{CSI}_{3,2} & \dots & \text{CSI}_{3,30} \end{bmatrix}. \tag{3}$$

2) CSI Enhancement: In the system, a person acts like a reflector to Wi-Fi signals (both 2.4 and 5 GHz), generating the NLOS signal propagation paths. Therefore, his/her movements, such as typing, will cause signal attenuation on these NLOS components. This phenomenon lays the foundation for Wi-Fi sensing. However, a centimeter-level human behavior, such as keystroke, is difficult to capture via CSI due to the environmental noise. To this end, we design a CSI enhancement model leveraging the Rician fading (Rician-K factor) by amplifying the ratio of the NLOS component corresponding to the behavior via weakening the LOS component, e.g., deploying a flat lead plate in front of the transmitting antenna, as shown in Fig. 2(a).

More specifically, the Rician-K factor [34] is defined as the ratio of the power of the LOS path to the power of the NLOS path in the received signal, which is a useful indicator for measuring the quality of the communication link. In theory, the baseband in-phase/quadrature-phase (I/Q) of the received signal $x(t)$ is expressed as

$$x(t) = \underbrace{\sqrt{\frac{K\Omega}{K+1}}}_{A:=} e^{j(2\pi f_D \cos(\theta_0)t + \phi_0)} + \underbrace{\sqrt{\frac{\Omega}{K+1}}}_{\sigma:=} h(t) \tag{4}$$

where K represents the Rician factor. θ_0 and ϕ_0 represent the Angle of Arrival (AoA) and phase of LOS, respectively. Ω and f_D represent the total received power and maximum Doppler frequency, respectively. $h(t)$ is the diffuse component given by the sum of multipath components, forming a complex Gaussian process. For K, we can prove that if $K > 1$, the system sensitivity to human behaviors increases as K decreases. Please refer to the Appendix for details.

This conclusion verifies our idea of CSI enhancement through weakening the LOS components (i.e., decreasing K), but it naturally brings a critical question: *where to deploy the lead plate to get the best performance?*

To answer this question, we study the variation of system sensitivity to the location of the plate from the energy point of the view. More specifically, as shown in Fig. 2(a), the system's ability to perceive tiny motions depends on the ratio of the power of the CSI's dynamic component to the total power of the received signal. The larger the ratio is, the stronger the system's sensing capabilities are. Thus, the system's sensitivity to human behaviors can be defined as

$$\delta = \frac{P_d}{P_R} \tag{5}$$

where P_d and P_R represent the power of the CSI's dynamic component and the total power of the received signal, respectively. Assume that the main lob of the transmitting

antenna (T_x) is almost completely covered by the lead plate placed on the LOS path, and the NLOS components spread into the environment can be regarded as evenly distributed. The power of the transmitted signal scattered in the indoor environment can be expressed as

$$P_T = P_0 \cdot \left(1 - \frac{\alpha}{\pi}\right) \tag{6}$$

where α, P_0, and P_T represent the angle of the area covered by the plate, the power of T_x, and the power actually scattered into the environment, respectively [as shown in Fig. 2(a)]. During the first scattering process, the power in the reflection area of the human body can be simplified as

$$P_d = P_0 \cdot \frac{\beta}{2\pi} \tag{7}$$

where β is the angle of the reflection area of the human body. Taking R_x as the observation point, when the distance between the lead plate and T_x is less than $d_{\mathrm{LOS}}/2$, the indoor scattering effect makes the distribution of the received signal on the horizontal plane roughly even due to the long distance between the lead plate and the R_x. Therefore, the received power can be approximated as

$$P_R \approx P_T. \tag{8}$$

Thus, the system sensitivity can be expressed as

$$\delta = \frac{P_d}{P_R} = \frac{\beta}{2(\pi - \alpha)}. \tag{9}$$

As the distance between the lead plate and T_x increases, the angle of the reflection area of the human body will change due to being blocked, and β can be written as

$$\beta = \begin{cases} 0, & 0 < d \le \dfrac{h}{2\tan\alpha_1} \\ \alpha_1 - \alpha, & \dfrac{h}{2\tan\alpha_1} < d \le \dfrac{h}{2\tan\alpha_2} \\ \alpha_1 - \alpha_2, & \dfrac{h}{2\tan\alpha_2} < d \le \dfrac{d_{\mathrm{LOS}}}{2} \end{cases} \tag{10}$$

where α_1 is represented as $\arctan(2H/(d_{\mathrm{LOS}} - w))$, α_2 is represented as $\arctan(2H/(d_{\mathrm{LOS}} + w))$, and α is the angle of the occluded area, which can be formed as

$$\alpha = \arctan\left(\frac{h}{2d}\right). \tag{11}$$

When the distance between T_x and the lead barrier is greater than $d_{\mathrm{LOS}}/2$, the shielding effect on the transmitter is very small and can be ignored, and thus, the transmitting power can be approximated as

$$P_T \approx P_0. \tag{12}$$

From the perspective of the receiver, the power corresponding to the reflection area of the human body is

$$P_d = P_T \cdot \frac{\phi}{2(\pi - \psi)} \tag{13}$$

where ϕ represents the angle between R_x and the blocking area, and ψ represents the angle between R_x and the reflection area. The system sensitivity can be expressed as

$$\delta = \frac{P_d}{P_R} = \frac{\phi}{2(\pi - \psi)}. \tag{14}$$

Similar to the effect of blocking the transmitted signal, ϕ will change as the lead barrier moves

$$\phi = \begin{cases} \alpha_1 - \alpha_2, & \dfrac{d_{\mathrm{LOS}}}{2} < d \le d_{\mathrm{LOS}} - \dfrac{h}{2\tan\alpha_2} \\ \alpha_1 - \psi, & \left(d_{\mathrm{LOS}} - \dfrac{h}{2\tan\alpha_2}\right) < d \le \left(d_{\mathrm{LOS}} - \dfrac{h}{2\tan\alpha_1}\right) \\ 0, & \left(d_{\mathrm{LOS}} - \dfrac{h}{2\tan\alpha_1}\right) < d \le d_{\mathrm{LOS}} \end{cases} \tag{15}$$

where ψ is the angle of the occluded area and can be calculated as

$$\psi = \arctan\left(\frac{h}{2(d_{\mathrm{LOS}} - d)}\right). \tag{16}$$

As mentioned above, when the lead plate is deployed on the LOS path, the system sensitivity will increase. As shown in Fig. 2(a), we quantitatively analyze the sensitivity of the lead plate at different positions of the LOS path (please refer to Section IV-A for the detailed experimental settings). Fig. 2(b) shows how the system sensitivity changes in theory given by (9) and (14) and the actual measured system sensitivity. Considering that our theoretical analysis is simplified in free space, experimental results are consistent with our CSI enhancement model. Therefore, in order to enhance the sensitivity of the system and highlight the behavior-induced information, we placed the lead plate at a distance of 20 to 30 cm from T_x when collecting data.

Fig. 2(c) shows such an example that records a volunteer inputting the same keyword "$yxx960919$" twice with and without CSI enhancement (please refer to Section IV for the detailed experimental settings). The top subfigure compares the amplitude profiles of subcarrier #16 before (red line) and after enhancement (blue line), where the sensitivity of CSI to the keystrokes in terms of the amplitude fluctuation has been improved by 3.89 times after enhancement. We further visualize the energy profiles of all subcarriers in the middle and bottom subfigures for the original CSI and enhanced CSI. It is quite interesting to see that the color change corresponding to the keystrokes becomes much sharper after enhancement, indicating the outstanding effect of our model in practice.

3) CSI Visualization: In this part, we propose a CSI visualization method to map the enhanced time-series CSI to images, which allows us to leverage the cutting-edge learning frameworks, such as the prototypical networks for efficient user authentication.

The signal fluctuations caused by human behaviors usually are at the low-frequency band, while the background noises induced by hardware and environment tend to focus on the high-frequency band. Therefore, we utilize the Butterworth low-pass filter to eliminate the high-frequency noises while retaining the user behavioral information. The Butterworth filter removes out-of-band noise (stopband) to a large extent while ensuring that the information in the passband is not significantly distorted. In WiONE, we set the cutoff frequency to 0.1884 according to our empirical studies.

After denoising, the behavior-irrelevant information in the CSI has been filtered. Then, we propose a visualization method

to map the CSI amplitude time-series profiles to images. More specifically, as mentioned in Section III-B1, for a wireless transmission process with N_t transmitting antennas, N_r receiving antennas, K subcarriers, and T received packets, directly processing the original CSI $H \in R^{N_t \times N_r \times K \times T}$ in the prototypical networks bears the risk of performance deterioration due to its large volume. To this end, we visualize the channel response only in the energy and spectral domains to reduce the data dimension. The energy image EN of a wireless link can be created with channels as the Y-axis and time instants as the X-axis, as follows:

$$\mathrm{EN} = [\mathrm{EN}_1, \ldots, \mathrm{EN}_t, \ldots, \mathrm{EN}_T] \tag{17}$$

where EN_t represents the amplitude value of all subcarriers at time t, i.e.,

$$\mathrm{EN}_t = \left[\mathrm{EN}_t{}^1, \ldots, \mathrm{EN}_t{}^k, \ldots, \mathrm{EN}_t{}^K\right]. \tag{18}$$

Each user's keystroke speed is different, and the frequency of CSI time-domain caused by finger and hand movements is also different. Therefore, the frequency component in the spectrum image is also very important. We use the principal component analysis (PCA) to select the most representative or primary components of all CSI time-domain data and then convert it into a spectrogram by the discrete wavelet transform (DWT).

Fig. 2(d) demonstrates the same example being denoised and visualized after CSI enhancement. The top subfigure shows the effect of denoising on subcarrier #16, where the signal blurs caused by the impulse noises have been removed, while the signal distortions caused by keystrokes have been preserved. The middle subfigure records the energy profiles where the input of the password has been vividly captured. The bottom subfigure describes the corresponding event in the frequency domain.

C. User Authentication Module

In the BEM, we generate images in the energy and spectral domains containing the user behaviors from the original CSI through enhancement, denoising, and visualization. In this module, we design a deep learning framework effectively handling these images to explore behavioral information for user authentication and spoofer detection.

Deep learning relies heavily on abundant labeled instances when training the network. The human annotation and the scarcity of samples in certain rare categories significantly limit the availability of the network. For user authentication, such a training burden would neutralize users' willingness and offset any potential merits the system can bring. To fill in the gap, we propose a new learning framework based on the prototypical networks, which only requires few labeled samples in each class, and simulate various feature distributions from the previously trained environment to the new environment through novel training strategies. Moreover, we also design a spoofer detection mechanism embedded in the authentication process, i.e., verifying whether the login user is a spoofer.

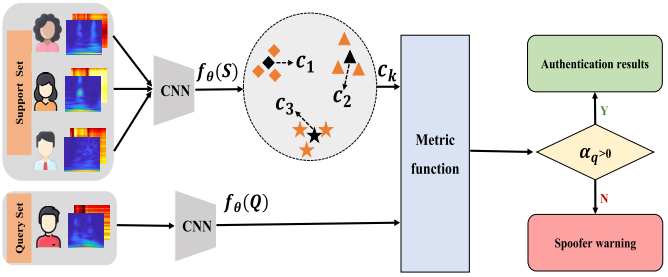


Fig. 3. Illustration of the proposed one-shot learning framework. Three samples of each class in the support set are selected to extract features (i.e., the triangles, rectangles, and five-pointed stars marked with yellow marks in the figure), and prototypes c_k (i.e., the black marks in the middle of the features in the figure) is calculated. Then, the distance between the extracted features of the query set sample and the prototypes is measured for user authentication and spoofer detection.

1) Architecture of Prototypical Networks: The one-shot learning problem is typically characterized as N_w way (number of categories) and one shot (number of labeled examples for each category). Fig. 3 shows an example of the UAM on the three-way one-shot user authentication task. Episode-based training is the most effective way to use the training set to mimic one-shot learning [35]. In each training iteration, each episode randomly selects C categories and N_s samples of each category as the support set and then randomly selects N_q samples from the remaining C categories as the query set. One-shot learning means that there is only one sample per category in the support set. This support/query set split in the training set is designed to simulate the support/query set that will be encountered at test time. A model trained from the support/query set in the training set can effectively realize user authentication using only one sample for each user from the test set. It should be noted that the label space of the training set disjoints the test set. Our goal is to use the knowledge obtained from the support set and the transferable knowledge obtained from the training set to identify the label of each sample in the query set.

UAM generally contains a feature encoder, a prototype construction layer, and a metric layer, as illustrated in Fig. 3. Let (x, y) stand for the support set pairs, and $x = \{EN, SP\}$ is the input energy and spectrum image with a size of $H \times W \times 6$, where H and W represent the width and height of the image, respectively. y is the output label for the corresponding user. Then, the support set S with N samples can be written as

$$S = \{(x_i, y_i)\}_{i=1}^{N}. \tag{19}$$

The convolutional neural network (CNN) is utilized as the feature encoder. CNN mainly includes four convolutional blocks, each of which contains a convolutional layer, a batch normalization layer, a rectified linear unit (ReLU) layer, and a max-pooling layer. Let θ be the set of CNN parameters. Given the support set S, we can obtain the CSI feature representations of the user's keystrokes as follows:

$$F = f_\theta(S). \tag{20}$$

The sample feature dimensions are converted through CNN: $R^D \rightarrow R^M$. The prototypical networks use the feature

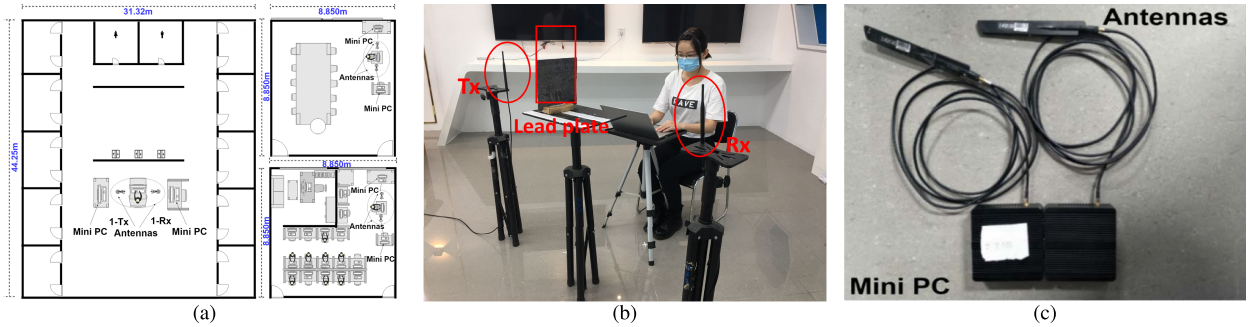


Fig. 4. (a) Our prototype system in three different experimental sites: the left is the lobby, the upper right corner is the meeting room, and the lower right corner is the laboratory. (b) Layout of the prototype system in the lobby site. (c) Mini-PC and antenna equipment.

representation of the samples in each category to solve the mean vector as the prototype of the network, which is an M-dimensional representation $c_k \in \mathbb{R}^M$, as follows:

$$c_k = \frac{1}{|S_k|} \sum_{(x_i, y_i) \in S_k} f_\theta(x_i) \tag{21}$$

where S_k represents the set of samples marked as class k. The prototypical networks then measure the distance between any sample in the query set and the prototype of each category to predict the category of the sample in the query set. The prediction process is defined as

$$\hat{y} = \arg\min_k \; d(f_\theta(q), c_k) \tag{22}$$

where d is a metric function (i.e., squared Euclidean distance). In the actual training process, we often add $softmax$ layer to the distance to generate the distribution of the query set, so prototypical networks produce a distribution over classes for a query point q based on a $softmax$ over distances to the prototypes in the embedding space, i.e.,

$$p_\theta(y = k|q) = \frac{\exp(-l(f_\theta(q), c_k))}{\sum_{c_{k'}} \exp(-l(f_\theta(q), c_{k'}))}. \tag{23}$$

The network only requires using stochastic gradient descent (SGD) algorithm to learn by minimizing the negative log-probability of the real class k, and the loss function is defined as follows:

$$J(\theta) = -\log p_\theta(y = k|q). \tag{24}$$

The one-shot learning training strategy is often directly used as in [35] and [36]. However, when the test environment and training environment are quite different, the performance of user authentication will be significantly degenerated. Based on episode training, we propose a novel two-step training strategy that can be carried out in a new test environment.

In the first step, we randomly select the support set and query set from the training environment and the test environment. Note that the data from the test environment are not from unfamiliar users. In the second step, we authenticate unclassified users in the test environment by using the trained model in the first step. Let T be the distribution of user identities, which can be denoted as $T_{\text{tre}} \cup T_{\text{tee}}$. Here, T_{tre} represents the data distribution from the training environment, and T_{tee} represents the data distribution from the test environment (in this case, the training is the registered user data, not the unfamiliar users' data). In each training iteration, an episode is composed of a set randomly sampled according to the category in T, and then, the support set S and the query set Q are obtained through sampling. The one-shot learning is used to simulate the support/query set encountered during testing through this support/query set split design. The optimization problem of the one-shot learning is expressed as

$$\arg\min_\theta \; E_{S,Q\sim T}\left[\sum_{(x,y)\in Q} \log P_\theta(y|x, S)\right]. \tag{25}$$

2) User Authentication and Spoofer Detection: To detect legitimate user and spoofer, we set a variable TH to describe the Euclidean distance between the extracted feature vectors of a registered user and a spoofer

$$\text{TH}(q) = \max_{k\in 1,\ldots,K} \; l(f_\theta(q), c_k) - \gamma \min_{k\in 1,\ldots,K} \; l(f_\theta(q), c_k) \tag{26}$$

where γ represents a control weight, which is set to 3 in our system based on the empirical results. If $\text{TH}(q)$ is greater than 0, the maximum similarity and the minimum similarity between the unclassified user and the registered users are very different. In this case, the unclassified user can be identified as the user with the closest metric distance. If $\text{TH}(q)$ is less than 0, it means that the maximum similarity and the minimum similarity between the unclassified user and the registered user are close, and then, the attempted login user is identified as a spoofer.

IV. Evaluation

In this section, we implement WiONE with low-cost commodity Wi-Fi devices and conduct extensive experiments at three real-world environments, i.e., laboratory, meeting room, and lobby, to evaluate the performance of the proposed system.

A. Experimental Settings

We implement a fully WiONE functional prototype to evaluate the performance of the proposed system with commercial off-the-shelf devices. It consists of two mini-PCs (2.16-GHz CPU, 4-GB RAM, and 240-GB SSD) mounted with NIC 5300 as the sender and receiver, as shown in Fig. 4(c). The sender

emits the 802.11n Wi-Fi signal at 1000 Hz in the injection model, while the receiver keeps recording the CSI data via the Linux 802.11n CSI tool in the monitor mode. The distance between transmitting and receiving antennas is set to 60 cm. We deploy a $2.6 \times 20 \times 0.1$ cm^3 lead plate 20 cm in front of the transmitting antenna to highlight the signal distortions caused by keystrokes by blocking the LOS components. Since WiONE intends to leverage the existing Wi-Fi infrastructure, we assume that the hardware (including its position) is fixed during the experiment.

For CSI visualization, we resize the energy image and the spectrum image to 224×224. For the prototypical networks, we use a 64-filter 3×3 convolution and a 2×2 max-pooling in each convolution block of the feature encoder. We use the same feature encoder for feature extraction of samples in the support set and query set. We train the networks via SGD using the Adam algorithm with the initial learning rate of 0.0001 and cut the learning rate in half every 100 episodes. All the models are trained on a single NVIDIA RTX 2080 Ti for 40 epochs. We implement the proposed framework using MATLAB and Pytorch.

We evaluate the performance of WiONE in three real environments: 1) a lobby of 13.62×44.25 m^2 without any furniture; 2) a meeting room of the same size full of furniture but no students; and 3) our laboratory of 8.85×8.85 m^2 full of students and furniture. Fig. 4(a) and (b) shows the layout of three real environments. There is no restriction on the presence of other people during experiments. We have recruited 19 participants (nine females) from our university, whose age, weight, and height range from 19 to 34, 38 to 95 kg, and 1.58 to 1.82 m, respectively. Though all of them are familiar with different keyboards, we still ask them to practice on a 104-key Logitech K120 keyboard before the formal experiment.

WiONE is a credential-free authentication system, and therefore, all users are assigned a unified password to log in. Considering the average length of a password is 9.6 characters [37], we set the password of WiPass as "yxx960919," which consists of the initials and birth-date of one female participant. All participants find it easy to remember and easy to input. Later, we will study the impact of passwords on system performance. We split 19 participants into two groups, i.e., legitimate users (16) and spoofers (3). Participants in both groups know the password, and they are asked to login via the keyboard 30 times in the three different environments. For each site, we set the training size to 20. We also compare WiONE with two other state-of-the-art one-shot learning methods i.e., MAML [38] and MatNet [35]. Please note that we use "one-shot" and "five-shot" to indicate using one and five samples, respectively, for each user from the support set. WiONE aims to provide an accurate and reliable user authentication service. Therefore, we define the following evaluation metrics similar to [33]: 1) authentication accuracy (AA): the probability that a user can be correctly recognized and authenticated and 2) false accept rate (FAR): the probability that a spoofer being authenticated as a legitimate user.

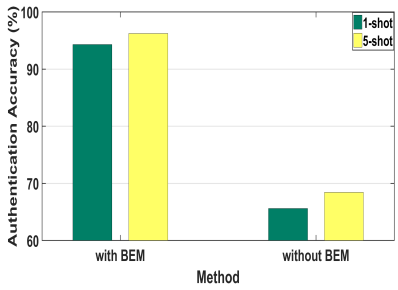


Fig. 5. Impact of BEM in the laboratory environment.

B. Overall Performance

Fig. 5 shows one-shot recognition results over all users in terms of AA in the lobby, meeting room, and laboratory, respectively. On average, WiONE achieves 94.67% AA and 0% FAR over the three experimental environments using only one training sample. The results indicate that WiONE is very effective in user authentication and spoofer detection.

On the one hand, we notice that WiONE is environment-robust. For example, if WiONE is trained in the meeting room first (95.2%), it achieves 91% AA and 94.1% using only one training sample from the laboratory and lobby, respectively. In other words, the average performance degeneration is only 2.65%. On the other hand, we surprisingly find that no matter where WiONE is trained and tested, it can always yield 0% FAR. More specifically, WiONE can always identify a spoofer no matter what the training/test combination of environments is.

C. Impact of CSI Enhancement Model

The impact of BEM on the performance of the proposed WiONE is investigated in this section to demonstrate the importance of BEM for the whole proposed scheme. Fig. 5 presents how well BEM can improve the authentication accuracy compared to the case without using it. We can also see that BEM can achieve the highest authentication accuracy in both one-shot setting and five-shot setting. This is because the proposed BEM can find the best blocking position so that the system has the highest sensitivity. We can find the reasons for the performance improvement from the BEM principle. BEM is a behavior enhancement model based on the Rician fading theory. It mainly finds the best blocking position so that amplifying the ratio of the NLOS component corresponding to the behavior via weakening the LOS component (i.e., the Rician-K factor is the smallest). Therefore, WiONE can highlight the behavior-induced information by suppressing the behavior-irrelevant information on channel response.

D. Performance Comparison for Different Methods

In Fig. 6, we list the accuracy of our approach and other one-shot learning methods for comparison. Note that we only use the original episode-based training strategy in different environments because here we only verify the user authentication performance of WiONE in a separate environment

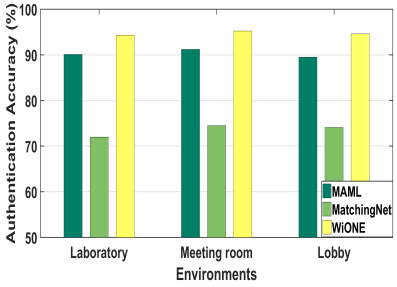


Fig. 6. Comparison with two other state-of-the-art rivals.

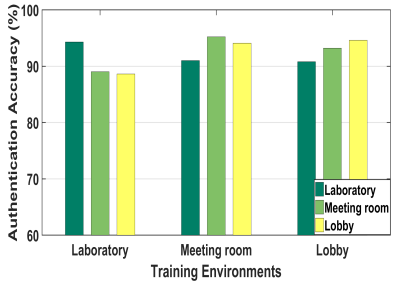


Fig. 7. Overall performance of WiONE.

for "one-shot." From the results, we can see that WiONE has the best user authentication performance compared to MatchingNet and MAML. Moreover, it can be observed that the authentication accuracy difference of WiONE in the three home environments is very small and all have good authentication results, which indicates that WiONE is robust to different indoor environments (note that both the training set and the test set come from the same environment).

To verify the proposed novel training strategy when the test environment and the training environment are different, we select 20 samples per user in the existing environment and ten samples per user in the training set in the test environment as the training set. After the model is trained, the test set in the test environment is used for user authentication. In this article, when the environment is not the same, we assume that there are six verified users in the test environment. Fig. 7 shows the user authentication result using the novel training strategy. We only use the WiONE and one-shot settings to verify the proposed novel training strategy. The horizontal axis represents the environment of the training set, and each color block corresponds to the test set from the test environment. On the one hand, in Fig. 7, we can see that the recognition accuracy decreases when the training environment and the test environment are inconsistent. On the other hand, Fig. 7 shows that the recognition accuracy in the meeting room and the lobby is higher than that in the laboratory. This is because students carry out their normal activities so that the experimental environment is more complicated. When performing user authentication in the meeting room and the lobby, the environment is particularly empty, and there is no interference from surrounding people. When the training environment is a laboratory and the test environment is a meeting room or lobby, the user authentication accuracy drops

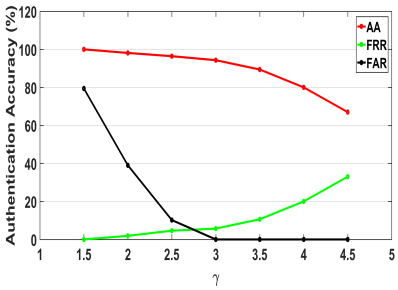


Fig. 8. Impact of γ under one-shot setting in the laboratory.

by 5.4% at the highest, and vice versa, the user authentication accuracy drops by 4.2% at the highest.

E. Performance Analysis for Spoofer Detection

To explore whether WiONE could achieve satisfactory performance in the presence of a spoofer, we first analyze why spoofer affects the performance of the user authentication system. Usually, after the network model is trained, the model can only be tested among existing users, and it will be forced to assign an identity to the spoofer. This type of spoofer attack on the user authentication system is very harmful. Once the user authentication is passed, the spoofer can obtain certain access privileges and initiate various malicious attacks. At this time, spoofer detection is an important step to improve the robustness of the user authentication system. Hence, we conduct experiments to explore the performance of WiONE in the presence of a spoofer. The rationale behind our spoofer detection is that, when there is a spoofer, there is little difference in the characteristic metric distance between the spoofer and the registered users. Therefore, we define a threshold to control the metric distance between spoofer behavior characteristics and real user behavior characteristics. If TH(q) is greater than 0, there is a huge difference between the maximum similarity and the minimum similarity between the attempted login user and the registered user. At this time, the attempted login user can be identified as the user with the closest metric distance. If TH(q) is less than 0, it means that the difference between the maximum similarity and the minimum similarity between the attempted login user and the registered user is very small, and then, the attempted login user is identified as a spoofer.

In this section, we use AA, FAR, and false reject rate (FRR) to evaluate the performance of the system during spoofer detection [39]. Note that, during the experiment, we designate three users in the test set as spoofers, while the remaining six users are not spoofers. In Fig. 8, we illustrate the authentication performance under different thresholds γ. It can be seen from the figure that, as the threshold increases from 1.5 to 4.5, the authentication accuracy gradually decreases, while the FAR gradually increases. Since the AA and FRR add up to 1, the change in the FRR is opposite to the authentication accuracy. The reason is that, when γ is small, the system has small limits on the metric distance between the query set and the support set. Therefore, when a spoofer attempts to log in to the system, an identity will be assigned to the

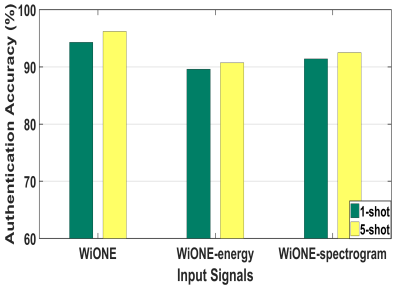


Fig. 9. Impact of different kinds of images.

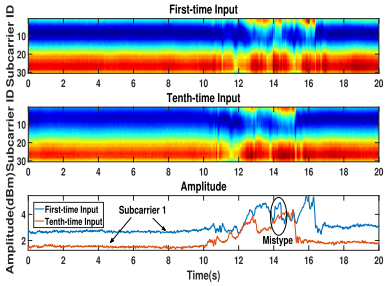


Fig. 10. How a user adapts to a complex password.

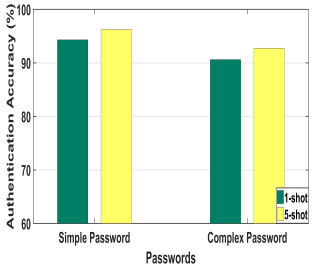


Fig. 11. Impact of different password sequences on the authentication accuracy.

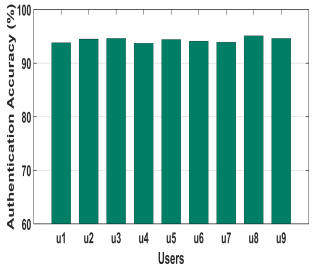


Fig. 12. Authentication accuracy for different users in the laboratory environment.

spoofer compulsorily, resulting in that incorrect authentication and FAR will be very high. On the contrary, when γ is large, the system has strong limits on the metric distance between the query set and the support set. Therefore, the system will not compulsorily identify the spoofer as to the user with the smallest metric distance. To make the best balance between AA and FAR, we set $\gamma = 3$ for WiONE during the evaluation.

F. Impact of Different Input Signals on WiONE

This experiment studies how the type of input signals impacts the performance (i.e., energy image and spectrogram). From Section III-B3, we know that the energy image contains rich information about the user's keyboard actions because it contains all the channel information in a wireless link. The spectrogram contains rich frequency domain information with the user's keyboard actions. Therefore, we explore the joint input of energy images and spectrogram to promote user authentication accuracy. Fig. 9 illustrates the average recognition accuracy with different input signals in the first indoor environment (i.e., laboratory). We use WiONE-energy to indicate that the system only uses the energy image as input signal and WiONE-spectrogram to indicate that the system only uses the spectrogram as the input signal. We can find that no matter whether the input signal is an energy image or a spectrogram, the joint input signals produce the highest recognition accuracy. Furthermore, the features extracted from the energy image only indicate the amplitude changes of all CSI time-domain data in a wireless link and cannot indicate the frequency change information in the frequency domain. At this time, the spectrogram makes up for this deficiency, and more essential features related to the user's actions are extracted through the joint input of the energy image and the spectrogram.

G. Impact of Different Passwords

Users have different typing habits for different password sequences, and users are less familiar with passwords with high complexity than passwords with low complexity. Moreover, for the same password sequence, the user becomes more and more proficient as the length of the password increases. In WiONE, a simple password cannot fully characterize the effectiveness of the user authentication system. Thus, we conduct a comparative experiment of different password sequences, and we consider a very complicated password sequence (i.e., abc!@&123) to observe the changes in CSI after different input times. Fig. 10 shows the change of CSI when the user enters the complex password for the first time and the complex password for the tenth time. The top subfigure shows the energy image of the user's first password input, the middle subfigure shows the energy image of the user's tenth password input, and the bottom subfigure shows the amplitude image of the user's first and tenth passwords. We can find that, when the user enters the password for the first time, it is not very proficient, and there is a pause during this period, but there is no pause when the password is entered for the tenth time. To more clearly verify the validity of WiONE for complex password sequences, we list the system's recognition accuracy for complex passwords and simple password sequences. Fig. 11 shows that WiONE achieves the best performance when the password sequence is simple compared to complex password sequences.

H. Impact of Different Users

We evaluate the performance of WiONE under different users in the laboratory environment. Fig. 12 shows the

authentication accuracy of WiONE under different users. The participants have various heights, weights, and somatotypes. Note that the users also have different strengths on the keyboard. It can be observed from Fig. 12 that the authentication accuracy of different users is not much different. In the one-shot setting, the highest authentication accuracy is 1.4% larger than the lowest authentication accuracy; these results show that WiONE has strong robustness in the actual environment.

V. Discussion and Future Work

We discuss several important issues of WiONE, including the advanced invasion resistance and the continuous user protection.

A. Advanced Invasion Resistance

If the trained model is directly authenticated for users, when a spoofer tries to log in to the system, the model will force the authenticator to appoint a user category. In WiONE, we set a threshold to control the metric distance between spoofer behavior characteristics and real user behavior characteristics. Besides the direct invasion, there may be skilled spoofers who can obtain data that already exists in the user authentication system (i.e., support set). In this case, the spoofer can perform advanced invasion by injecting the behavior characteristics collected from a login operation of a registered user. To prevent such an advanced attack, WiONE sets a threshold to control the minimum metric distance between the same user in the query set and the support set because, in actual scenarios, the behavior characteristics of the same user's keyboard tying cannot be 0. In the future, we will focus on the security issues of the support set, which can prevent data from being stolen by criminals and interfere with the user authentication system.

B. Continuous Authentication Protection

After the user is successfully authenticated, sometimes, she/he may forget to log out of the system. At this time, the spoofer may steal user information midway. In this work, we do not explicitly study continuous authentication protection in this context; we note that the continuous authentication protection concerns are worth exploring in the future.

VI. Conclusion

In this article, we propose a behavioral biometrics-based user authentication scheme, WiONE, which leverages the prototypical networks with enhanced CSI to enable one-shot recognition of a new user in a new environment. WiONE first designs a behavior enhancement model based on the Rician fading to outstand the behavior-induced information on channel response. To the best of our knowledge, it is the first quantitative model on leveraging the Rician fading for improving the sensitivity of channel response to human behavior. Then, we develop a learning framework based on the prototypical networks to extract the domain-independent behavior features and enable one-shot user authentication in a new environment. Our numerous comparative experiments demonstrate that WiONE can accurately authenticate users and detect spoofers.

Appendix
Rician-K Factor

In wireless communications, when there is an LOS between the transmitter and the receiver, the received signal can be written as the sum of a complex exponential and a narrowband Gaussian process, which is known as the "LOS component." The ratio of the powers of the LOS component to the whole received power is the Rician factor, which measures the relative strength of the LOS and represents the link quality. The baseband I/Q representation of the received signal can be modeled as [34]

$$x(t) = \sqrt{\frac{K\Omega}{K+1}} e^{j(2\Pi f_D \cos(\theta_0)t)+\phi_0} + \sqrt{\frac{\Omega}{K+1}} h(t) \quad (27)$$

where K is the Ricean Factor, Ω denotes the total received power, and θ_0 and ϕ_0 are the AoA and phase of the LOS, respectively. f_D is the maximum Doppler frequency, and $h(t)$ is the diffuse components.

In our system, considering that the antenna layout is fixed, i.e., $f_D = 0$, we can simplify (27) to

$$x(t) = \sqrt{\frac{K\Omega}{K+1}} e^{\phi_0} + \sqrt{\frac{\Omega}{K+1}} h(t). \quad (28)$$

Since we only measure the effect of the factor K on the received signal, it can be written as follows:

$$x(t) = \sqrt{\frac{K}{K+1}} + \sqrt{\frac{1}{K+1}}. \quad (29)$$

The LOS components and the part of NLOS components that are not affected by gestures are static. Other components are dynamic. Therefore, we can define H_s and H_d as follows (if omitting transmit power):

$$H_s = \sqrt{\frac{K}{K+1}} + \sqrt{\frac{1}{K+1}} \cdot \rho \quad (30)$$

$$H_d = \sqrt{\frac{1}{K+1}} \cdot (1-\rho) \quad (31)$$

where ρ is the proportion of static paths in the NLOS component.

We define $f(K) = |H_s| \cdot |H_d|$ to indicate the system sensitivity to the behavior. Assume that all NLOS components belong to the dynamic paths, we have the conclusion that, if $K > 1$, $f(K)$ increases as K decreases. The proof is as follows.

The received signal consists of both static and dynamic paths; thus, the receiving signal has a time-varying amplitude [40]

$$|H_{f,\theta}|^2 = |H_s(f)|^2 + |H_d(f)|^2 + 2|H_s(f)||H_d(f)|\cos\theta. \quad (32)$$

Combining (6), we can get the following equation:

$$\begin{aligned} |H|^2 &= |H_s|^2 + |H_d|^2 + 2|H_s||H_d|\cos\theta \\ &= \frac{K+\rho^2+2\sqrt{K}\rho\cos\alpha}{K+1} + \frac{(1-\rho)^2}{K+1} \\ &\quad + \frac{2(1-\rho)\sqrt{K+\rho^2+2\sqrt{K}\rho\cos\alpha}}{K+1}\cos\theta \end{aligned}$$

where α is the phase difference of the LOS component to the NLOS component in the static paths. It can be seen that the factors affecting the range of waveform fluctuation caused by the motions are K and ρ.

We define $f(K) = |H_s| \cdot |H_d|$ to indicate the system sensitivity to the gesture. Assuming that all NLOS components belong to the dynamic paths, i.e., $\rho = 0$ and $\alpha = (pi/2)$, we get the following equations:

$$f(K) = |H_s||H_d| = \frac{\sqrt{K}}{K+1} \tag{33}$$

$$f'(K) = \frac{1-K}{2\sqrt{K}(1+K)^2}. \tag{34}$$

When $K > 1$, $f(K)$ increases as K decreases. In practice, we could simply deploy an obstacle, such as a flat steel plate in front of the transmitting antenna to bound off the LOS signal to increase $f(K)$.

REFERENCES

[1] M. Tao, K. Ota, M. Dong, and Z. Qian, "AccessAuth: Capacity-aware security access authentication in federated-IoT-enabled V2G networks," *J. Parallel Distrib. Comput.*, vol. 118, pp. 107–117, Aug. 2018.

[2] B. Hutchins, A. Reddy, W. Jin, M. Zhou, M. Li, and L. Yang, "*BeatPIN*: A user authentication mechanism for wearable devices through secret beats," in *Proc. Asia Conf. Comput. Commun. Secur.*, 2018, pp. 101–115.

[3] R. S. Gaines, W. Lisowski, S. J. Press, and N. Shapiro, "Authentication by keystroke timing: Some preliminary results," RAND Corp., Santa Monica, CA, USA, Tech. Rep. R-2526-NSF, 1980.

[4] A. Hanamsagar, S. S. Woo, C. Kanich, and J. Mirkovic, "Leveraging semantic transformation to investigate password habits and their causes," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, p. 570.

[5] LastPass. (2017). *The Password Expose*. [Online]. Available: https://lp-cdn.lastpass.com/lporcamedia/document-library/lastpass/pdf/en/LastPass-Enterprise-The-Password-Expose-Ebook-v2.pdf

[6] WikiLeaks. (2019). *Wikileaks*. [Online]. Available: https://wikileaks.org

[7] Data-Breach. (2019). *Data-Breach*. [Online]. Available: https://www.ibm.com/security/data-breach

[8] Y. Wang, Q. Yao, J. T. Kwok, and L. M. Ni, "Generalizing from a few examples: A survey on few-shot learning," *ACM Comput. Surv.*, vol. 53, no. 3, p. 63, Jun. 2020.

[9] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2017, pp. 4077–4087. [Online]. Available: http://papers.nips.cc/paper/6996-prototypical-networks-for-few-shot-learning.pdf

[10] Wikipedia. (2019). *Personal Identification Number*. [Online]. Available: https://en.wikipedia.org/wiki/Personal_identification_number

[11] T. Takada and H. Koike, "Awase-E: Image-based authentication for mobile phones using user's favorite images," in *Proc. Int. Conf. Mobile Human-Comput. Interact.* Berlin, Germany: Springer, 2003, pp. 347–351.

[12] A. Jain, L. Hong, and R. Bolle, "On-line fingerprint verification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 4, pp. 302–314, Apr. 1997.

[13] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proc. IEEE*, vol. 83, no. 5, pp. 705–741, May 1995.

[14] D. Gafurov, E. Snekkenes, and P. Bours, "Spoof attacks on gait authentication system," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 3, pp. 491–502, Sep. 2007.

[15] L. C. F. Araujo, L. H. R. Sucupira, M. G. Lizarraga, L. L. Ling, and J. B. T. Yabu-Uti, "User authentication through typing biometrics features," *IEEE Trans. Signal Process.*, vol. 53, no. 2, pp. 851–855, Feb. 2005.

[16] A. K. Jain, R. Bolle, and S. Pankanti, *Biometrics: Personal Identification in Networked Society*, vol. 479. Springer, 2006.

[17] A. K. Datta, *Advances in Fingerprint Technology*. Boca Raton, FL, USA: CRC Press, 2001.

[18] K. Revett, "A bioinformatics based approach to behavioural biometrics," in *Proc. Frontiers Converg. Bioscience Inf. Technol.*, 2007, pp. 665–670.

[19] N. Zheng, A. Paloski, and H. Wang, "An efficient user verification system via mouse movements," in *Proc. 18th ACM Conf. Comput. Commun. Secur. (CCS)*, 2011, pp. 139–150.

[20] H. Li, K. Ota, M. Dong, and M. Guo, "Learning human activities through Wi-Fi channel state information with multiple access points," *IEEE Commun. Mag.*, vol. 56, no. 5, pp. 124–129, May 2018.

[21] S. Lv, Y. Lu, M. Dong, X. Wang, Y. Dou, and W. Zhuang, "Qualitative action recognition by wireless radio signals in human–machine systems," *IEEE Trans. Human-Machine Syst.*, vol. 47, no. 6, pp. 789–800, Dec. 2017.

[22] Y. Cao *et al.*, "Contactless body movement recognition during sleep via WiFi signals," *IEEE Internet Things J.*, vol. 7, no. 3, pp. 2028–2037, Mar. 2020.

[23] Y. Gu, Y. Wang, Z. Liu, J. Liu, and J. Li, "SleepGuardian: An RF-based healthcare system guarding your sleep from afar," *IEEE Netw.*, vol. 34, no. 2, pp. 164–171, Mar. 2020.

[24] Y. Wang, K. Wu, and L. M. Ni, "WiFall: Device-free fall detection by wireless networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 581–594, Feb. 2017.

[25] F. Zhang *et al.*, "SMARS: Sleep monitoring via ambient radio signals," *IEEE Trans. Mobile Comput.*, vol. 20, no. 1, pp. 217–231, Jan. 2021.

[26] K. Ali, A. X. Liu, W. Wang, and M. Shahzad, "Keystroke recognition using WiFi signals," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2015, pp. 90–102.

[27] H. Abdelnasser, K. Harras, and M. Youssef, "A ubiquitous WiFi-based fine-grained gesture recognition system," *IEEE Trans. Mobile Comput.*, vol. 18, no. 11, pp. 2474–2487, Nov. 2019.

[28] Y. Meng, J. Li, H. Zhu, X. Liang, Y. Liu, and N. Ruan, "Revealing your mobile password via WiFi signals: Attacks and countermeasures," *IEEE Trans. Mobile Comput.*, vol. 19, no. 2, pp. 432–449, Feb. 2020.

[29] C. L. Li, M. Liu, and Z. Cao, "WiHF: Gesture and user recognition with WiFi," *IEEE Trans. Mobile Comput.*, early access, Jul. 15, 2020, doi: 10.1109/TMC.2020.3009561.

[30] J. Yang, H. Zou, Y. Zhou, and L. Xie, "Learning gestures from WiFi: A siamese recurrent convolutional architecture," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10763–10772, Dec. 2019.

[31] Q. Li *et al.*, "AF-DCGAN: Amplitude feature deep convolutional GAN for fingerprint construction in indoor localization systems," *IEEE Trans. Emerg. Topics Comput. Intell.*, early access, Nov. 5, 2020, doi: 10.1109/TETCI.2019.2948058.

[32] Z. Han, L. Guo, Z. Lu, X. Wen, and W. Zheng, "Deep adaptation networks based gesture recognition using commodity WiFi," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, May 2020, pp. 1–7.

[33] H. Kong, L. Lu, J. Yu, Y. Chen, L. Kong, and M. Li, "Fingerpass: Finger gesture-based continuous user authentication for smart homes using commodity WiFi," in *Proc. 20th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2019, pp. 201–210.

[34] C. Tepedelenlioglu, A. Abdi, and G. B. Giannakis, "The Ricean K factor: Estimation and performance analysis," *IEEE Trans. Wireless Commun.*, vol. 24, no. 5, pp. 799–810, May 2003.

[35] O. Vinyals *et al.*, "Matching networks for one shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 3630–3638.

[36] S. Ravi and H. Larochelle, "Optimization as a model for few-shot learning," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, Toulon, France, Apr. 2017.

[37] J. Lampe. (2014). *Beyond Password Length and Complexity*. [Online]. Available: https://resources.infosecinstitute.com/beyond-password-length-complexity

[38] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," 2017, *arXiv:1703.03400*. [Online]. Available: http://arxiv.org/abs/1703.03400

[39] X. Xu *et al.*, "TouchPass: Towards behavior-irrelevant on-touch user authentication on smartphones leveraging vibrations," in *Proc. 26th Annu. Int. Conf. Mobile Comput. Netw.*, 2020, pp. 1–13.

[40] H. Wang *et al.*, "Human respiration detection with commodity WiFi devices: Do user location and body orientation matter?" in *Proc. ACM Int. Joint Conf. Pervasive Ubiquitous Comput.*, 2016, pp. 25–36.

Yu Gu (Senior Member, IEEE) received the B.E. degree from the Special Classes for the Gifted Young, University of Science and Technology of China, Hefei, China, in 2004, and the D.E. degree from the University of Science and Technology of China in 2010.

In 2006, he was an Intern with Microsoft Research Asia, Beijing, China, for seven months. From 2007 to 2008, he was a Visiting Scholar with the University of Tsukuba, Tsukuba, Japan. From 2010 to 2012, he was a JSPS Research Fellow with the National Institute of Informatics, Tokyo, Japan. He is currently a Professor and the Dean Assistant with the School of Computer and Information, Hefei University of Technology, Hefei. His current research interests include pervasive computing and affective computing.

Dr. Gu is also a member of Association for Computing Machinery (ACM). He was a recipient of the IEEE Scalcom2009 Excellent Paper Award and the NLP-KE2017 Best Paper Award.

Huan Yan was born in Guizhou, China, in 1995. He received the B.E. degree from the Hefei University of Technology, Hefei, China, where he is currently pursuing the Ph.D. degree.

His research interests include intelligent information processing and wireless sensing and affective computing.

Mianxiong Dong (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in computer science and engineering from The University of Aizu, Aizuwakamatsu, Japan.

He is the youngest ever Vice President and Professor of the Muroran Institute of Technology, Muroran, Japan. He was a JSPS Research Fellow with the School of Computer Science and Engineering, The University of Aizu, and a Visiting Scholar with the BBCR Group, University of Waterloo, Waterloo, ON, Canada, supported by the JSPS Excellent Young Researcher Overseas Visit Program from April 2010 to August 2011.

Dr. Dong was selected as a Foreigner Research Fellow (a total of three recipients all over Japan) by the NEC C&C Foundation in 2011. He was a recipient of the IEEE TCSC Early Career Award 2016, the IEEE SCSTC Outstanding Young Researcher Award 2017, The 12th IEEE ComSoc Asia-Pacific Young Researcher Award 2017, the Funai Research Award 2018, and the NISTEP Researcher 2018 (one of only 11 people in Japan) in recognition of significant contributions in science and technology. He is also the Clarivate Analytics 2019 Highly Cited Researcher (Web of Science).

Meng Wang received the bachelor's degree from the Hefei University of Technology, Hefei, China, where she is currently pursuing the master's degree.

Her current research interests include intelligent information processing, wireless sensing, and machine learning.

Xiang Zhang was born in Anhui, China, in 1996. He received the B.E. degree from the Hefei University of Technology, Hefei, China, where he is currently pursuing the Ph.D. degree.

His research interests include intelligent information processing and wireless sensing and affective computing.

Zhi Liu (Senior Member, IEEE) received the B.E. degree from the University of Science and Technology of China, Hefei, China, and the Ph.D. degree in informatics in the National Institute of Informatics, Tokyo, Japan.

He is currently an Associate Professor with The University of Electro-Communications, Tokyo. His research interest includes video network transmission, vehicular networks, and mobile edge computing.

Dr. Liu is also an Editorial Board Member of *Wireless Networks* (Springer). He has been a Guest Editor of *Mobile Networks & Applications* (ACM/Springer), *Wireless Networks* (Springer), and *IEICE Transactions on Information and Systems*.

Fuji Ren (Senior Member, IEEE) received the B.E. and M.E. degrees from the Beijing University of Posts and Telecommunications, Beijing, China, in 1982 and 1985, respectively, and the Ph.D. degree from Hokkaido University, Sapporo, Japan, in 1991.

He is currently a Professor with the Faculty of Engineering, Tokushima University, Tokushima, Japan. His research interests include information science, artificial intelligence, language understanding and communication, and affective computing.

Dr. Ren is also a member of the Institute of Electronics, Information and Communication Engineers (IEICE), Chinese Association for Artificial Intelligence (CAAI), the Institute of Electrical Engineers of Japan (IEEJ), Information Processing Society of Japan (IPSJ), Japanese Society for Artificial Intelligence (JSAI), and the Asia-Pacific Association for Machine Translation (AAMT). He is also a fellow of the Japan Federation of Engineering Societies. He is also the President of the International Advanced Information Institute.