Data-Augmentation-Enabled Continuous User Authentication via Passive Vibration Response

Hangcheng Cao, *Student Member, IEEE*, Hongbo Jiang, *Senior Member, IEEE*, Kehua Yang, Siyu Chen, *Student Member, IEEE*, Wenqi Wu, Jiangchuan Liu, *Fellow, IEEE*, and Schahram Dustdar, *Fellow, IEEE*

***Abstract*—Continuous identity authentication is critical for privacy protection throughout an entire user login session. In this article, we propose a continuous user authentication mechanism, namely, HandPass, which employs the vibration responses from hand biometrics and is passively activated by natural user-device interaction. Hand vibration responses are embedded in the mechanical vibration of a force-bearing body consisting of one mobile device and one user hand. A built-in accelerometer of the device can capture hand-dependent vibration signals. Considering the concealment of vibration generation and the nonreplicability of hand structure, it is difficult for attackers to counterfeit user identity. Moreover, for ensuring the robustness of authentication performance to tapping behavior interference, we construct a data augmentation module jointly leveraging a signal processing and learning-based pipeline. It can generate enough vibration responses representing hand structure biometrics under various behaviors, thereby making HandPass comprehensively understand vibration response variation. We prototype HandPass on smartphones, and extensive experiments demonstrate that HandPass can achieve satisfactory authentication accuracy.**

***Index Terms*—Continuous user authentication, data augmentation, hand vibration response.**

I. Introduction

USER authentication is one basic while critical component for privacy data protection on smartphones [1]. To be specific, authenticating users by checking identity information helps confirm that the person is who he/she claims to be. Current smartphones only authenticate users at a transient login moment. Once the authentication is successful, unlocked devices can be accessed even if their owner temporarily leaves afterwards [2]. In this case, severe security risks like an illegitimate copy of sensitive documents always happen. According to statistical data, nearly 41% of data leakage events originate from illegitimate intrusion into smartphones [3] and the leaked data includes personal activity trajectory, health records, financial information, etc. To resist such crucial security threats, the concept of continuous authentication [4] is proposed, aiming to keep tracking user identity during entire usage sessions. It can prevent illegitimate invasion from adversaries when a legitimate user is away or overwhelmed. For HandPass, one authentication process fails under the following cases, e.g., legitimate users leave or adversaries appear, then the device is automatically locked and thus privacy-sensitive information is protected. To achieve this design, one naive way is asking users to periodically authenticate themselves by inputting preset passwords or verifying biometric features. However, this active way immensely affects normal device usage and thus compromises user experience.

To alleviate the incompatibility between security and user-friendliness, recent studies propose advanced continuous authentication schemes based on user behavior features, such as motion orientation when using a smartphone [2], mouth area movement patterns during speaking [5], eye movement trajectories [6] when scanning screen contents, etc. Although these schemes achieve identity authentication in a nonintrusive way, behavioral-based schemes are discovered to be vulnerable to impersonation attacks [7]. Moreover, these schemes rely on authentication traits always being contaminated by unstable human psychological and emotional states, which may compromise user experience. Recently, some emerging approaches employ biometrics, such as skin vibration characteristics during speaking [8] and heartbeat patterns [9] for authentication. However, they require either nontrivial sensors or ingenious environment configurations for biometric feature collection. To sum up, existing continuous authentication approaches still have a long way to go before they realize efficient and user-friendly authentication via an efficient way.

In this article, we propose HandPass, a system that employs vibration responses of concealed hand biometrics passively activated by the natural user-device interactions (i.e., tapping

Manuscript received 15 November 2022; revised 2 February 2023; accepted 25 March 2023. Date of publication 19 April 2023; date of current version 8 August 2023. This work was supported in part by the National Natural Science Foundation of China under Grant U20A20181, Grant 61732017, and Grant 61902060; in part by the China Scholarship Council (CSC); in part by the Science and Technology Innovation Program of Hunan Province under Grant 2021RC4023; in part by the Open Research Fund from Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ) under Grant GML-KF-22-23; in part by the Key Research and Development Program of Hunan Province under Grant 2021WK2001; and in part by the National Social Science Foundation of China under Grant 19ZDA103. This article was presented in part at the 41st IEEE International Conference on Distributed Computing Systems (ICDCS 2021), Washington, DC, USA, Jul. 7–10, 2021. [DOI: 10.1109/ICDCS51616.2021.00101]. *(Corresponding authors: Hongbo Jiang; Kehua Yang.)*

Hangcheng Cao, Hongbo Jiang, Kehua Yang, and Siyu Chen are with the College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, Hunan, China (e-mail: hangchengcao@hnu.edu.cn; hongbojiang@hnu.edu.cn; khyang@hnu.cn; csy990406@hnu.edu.cn).

Wenqi Wu is with Shenzhen Research Institute, Shenzhen Technological Inc., Shenzhen 518000, China (e-mail: 1451767523@qq.com).

Jiangchuan Liu is with the School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada, and also with Jiangxing Intelligence Inc., Nanjing 210000, China (e-mail: jcliu@sfu.ca).

Schahram Dustdar is with the Research Division of Distributed Systems, TU Wien, 1040 Wien, Austria (e-mail: dustdar@dsg.tuwien.ac.at).

Digital Object Identifier 10.1109/JIOT.2023.3264274

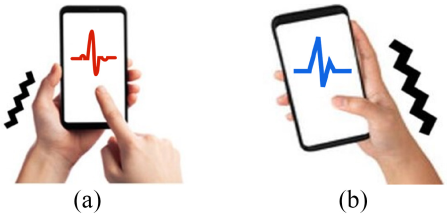


Fig. 1. Two widely used hand-holding gestures when operating smartphones. (a) Two-hand operation, denoted as G_1. (b) One-hand operation, denoted as G_2.

on a screen) for continuous user authentication. The key insight behind HandPass is that one user's hand owns a unique structure consisting of palm, wrist, and fingers, thus can generate user-specific vibration responses stimulated by an external force. Hand vibration responses (HVRs) are triggered immediately after user on-screen operations and embodied in the mechanical vibrations of the hand and the smartphone. Inspired by this observation, given typical usage scenarios as illustrated in Fig. 1, a user holds the smartphone with one hand and interacts with it by tapping on the screen. HandPass runs as a daemon in the background and it can be instantly activated by user-device interactions at run time if needed. Then, it exploits the HVRs captured by the built-in accelerometer for identity authentication in real time. The goals of our authentication mechanism are summarized as follows.

1) *Security:* Each hand has a complex and distinctive physical structure [10], thus HVRs derived from it are unique across users. Moreover, the vibration generation process is concealed and hence difficult to counterfeit.
2) *User-Friendliness:* It is essential for HandPass to avoid wearing any extra sensors or performing dedicated operations for authentication. HandPass verifies user identity in the background without compromising user experience.
3) *Efficiency:* The built-in accelerometer employed for data collection does not bring extra cost as it is always kept on by default in smartphones. The data size of each vibration response is small, thereby in low computation resource consumption in data processing modules.

To achieve the three goals mentioned above, several technical challenges require to be well addressed. First, user body movements/hand motions take noise into original vibration signals collected by a built-in accelerometer and hence weaken the performance of the following data analysis modules. Second, in addition to hand structure, user behaviors like tapping position and force also affect vibration signal forming, thus HVRs triggered by distinct behaviors of the same user may be inconsistent; this case takes a higher ratio of mistakenly rejecting legitimate users, if HandPass directly compares feature similarity of a newly inputted HVR with random registered profiles without judging tapping behaviors. Third, so as to ensure user experience, HandPass only asks users for tapping on the device screen within 1 min to complete identity profile registration; however, such a few HVRs indeed cannot adequately represent hand structure traits, especially when user behavior varies. As a result, this case inevitably makes the authentication performance lack stability.

To address these challenges, we propose three pertinent solutions. Extra noise first is divided into two types according to the variation pattern of frequency components [11], namely, steady and unsteady types. For steady one, we analyze noise frequency distribution under multiple scenarios and remove its components by a high-pass filter; nevertheless, unsteady type is extremely disruptive and is hard to separate from original HVRs; a gradient boosting decision tree (GBDT) [12] model then is utilized to recognize the unsteady type and eliminate it from original vibration signal segments to avoid spending extra computation resources of further processing. Second, we propose a nearest centroid-vector method to infer the tapping position and then determine the force of each vibration segment by analyzing vibration generation mechanism, to search for the registered HVR with similar behaviors. Last but not least, we explore vibration envelope variation patterns when each HVR-related behavior factor changes; finally, a traditional signal processing and customized conditional variational autoencoder (CVAE) [13] are jointly leveraged to generate newly "unseen" HVRs under various user behaviors and thereby fully representing hand biometrics. In HandPass, our main contributions are as follows.

1) We propose a continuous user authentication system named HandPass, leveraging concealed hand vibration biometrics triggered by user-device interaction.
2) We explore the theoretical model to describe the process of HVR generation and propagation, hence its uniqueness across individuals.
3) We design a series of effective strategies to solve the issues when implementing HandPass, such as removing steady/unsteady noises by analyzing their frequency components and employing a CVAE-based model to generate HVRs under distinct tapping behaviors.
4) We validate the effectiveness of HandPass through extensive experiments by recruiting 94 volunteers. The results present that HandPass can achieve 98.36% authentication accuracy.

The remainder of this article is organized as follows. Sections II and III, respectively, present the related works of user authentication and the preliminary background of hand vibration mechanism. Sections IV and V give the system design and tapping behavior judgment of HandPass. The feature extraction and data augmentation are, respectively, described in Sections VI and VII. Subsequently, we construct authentication models of multiple types to fully judge user identity in Section VIII. The implementation and performance evaluation of HandPass are, respectively, given in Sections IX and X. Finally, we discuss and conclude this work in Sections XI and XII.

II. Related Work

In this section, we review the literature about user authentication deployed on smartphones, while discussing the differences between HandPass and them.

One-Time User Authentication: Personal identification number (PIN) [14] is a widely adopted and traditional user authentication mechanism. However, password input can easily be stolen by shoulder surfing [15]. To overcome the awkward situation of PIN-related methods' vulnerability, plentiful studies have proposed biometrics-based authentication schemes such as fingerprint [16], voiceprint [17], [18], [19], and face recognition [20]. Nevertheless, these works still have quite a few records of being successfully compromised [21], [22]. Although the newly proposed SmileAuth [23] achieves satisfactory authentication accuracy, it consumes additional power and computing resources on wearable devices. Last but not the least, the aforementioned biometrics-based schemes cannot be compatible with the continuous authentication mechanism due to needing frequently active identity verification.

Continuous User Authentication: To reduce the security risks of a one-time approach, several recent studies propose an innovatively continuous authentication concept based on human behavioral features. Human–computer interaction behaviors, such as touch trajectory [2], tapping pressure [24], and hand geometry [25] have been leveraged for representing user identity. VoiceLive [26] leverages the characteristic of the arrival time of voice signal captured at a microphone to obtain a user's unique oral cavity motion during speaking for authentication. LipPass [5] extracts behavioral features of user speaking using the built-in microphone and speaker in a smartphone to inject acoustic signals and receive echos. Keystroke-based methods [27] focus on exploring the unique style of user's typing behaviors on screen. In addition, users' eyes show distinct movement patterns under the same visual stimulus, which can be utilized for distinguishing users [28]. Although these works state the feasibility of using behavioral features for continuous authentication, they are all vulnerable to resisting impersonation attacks [7]. Moreover, their performances are dependent on behavioral biometrics that can be easily affected by psychological and subjective states.

Recently, emerging works employ unique physiological biometrics for continuous authentication. For instance, vibration features of human skin during speaking [8] and heart movement traits [9] are utilized for tracking user identity during interaction sessions. FingerPass [29] relying on the channel state information of surrounding WiFi signals continuously authenticates users through user-specific finger gestures in a target space. However, these systems require either wearing dedicated sensors or a prepared environment setting for feature collection, which inevitably compromises user experiences.

Vibration-Based Authentication: An increasing amount of effort are devoted to exploring vibration-based user authentication. Chen et al. design VibID [30] using unique arm vibration signals stimulated by one external motor to distinguish users. Moreover, they propose secure device pairing [31] by the resonant properties. Furthermore, VibWrite [32] and Velody [10] apply vibration responses of finger and hand for user authentication. However, the above methods are all equipped with customized motors and data receivers, and thus cannot be implemented on smartphones. Besides, Taprint [33] proposes a tapping vibrometry for authentication on a smart wristband. It employs the unique vibration characteristics caused by active tapping on specific hand positions, especially finger knuckles, to identify a user. HandKey [34] is only suitable for nonmobile application scenarios like knocking on a door. Xu et al. [35] exploited a built-in motor to actively generate vibration signals and employed user-dependent responses when fingers tap on the screen. TouchPass is the most relevant work to our HandPass. However, it requires an extra activation operation to generate vibration signals and therefore is not suitable for frequently executed continuous authentication.

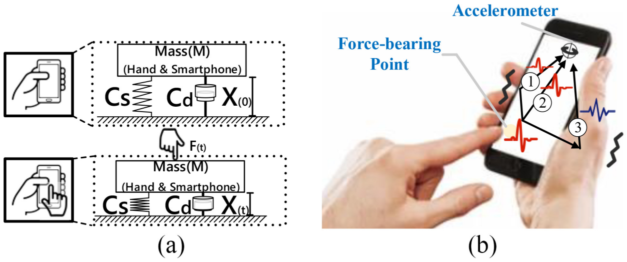


Fig. 2. Illustration of vibration generation and propagation when one hand taps on a smartphone. (a) Vibration generation. (b) Vibration propagation.

Different from existing works, HandPass employs HVRs containing concealed biometrics captured by a built-in accelerometer for continuous tracking user identity. The biometrics is passively activated by natural user-device interactions without a need for an external vibration source (e.g., a motor), hence ensuring user-friendliness. Moreover, HandPass leverages a signal processing and learning-based data augmentation approach to generate enough unseen HVRs, thereby comprehensively representing user hand biometrics under distinct tapping behaviors.

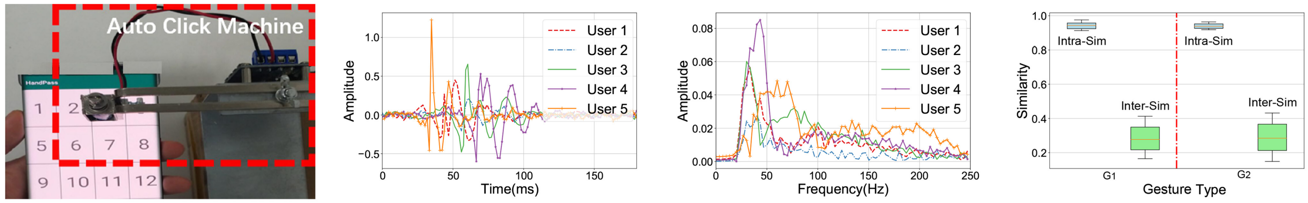
III. Preliminary

In this section, we first explore the theory model behind vibration response, including signal generation, nonlinearity derivation, and propagation. Subsequently, a feasibility study utilizing data collected in real environments is conducted; its result indicates that HVRs are unique across users while consistent for the same user, which motivates our research in this work.

A. Hand Vibration Response Mechanism

When an external tapping force acts on a hand, the force-bearing point first generates vibration waves and meanwhile spreads outwards. Finally, the hand is forced to vibrate and its amplitude continuously attenuates until dribbles away.

Vibration Generation: Each hand is composed of three main parts (namely fingers, palm, and wrist), which owns a unique structure jointly determined by its shape, muscular tissue, and skeleton [36], [37]. When one user holds and taps on a smartphone, forces transmit to the hand and then drive it to output unique vibration waves. Subsequently, a mass-spring-damper model according to [32] and [33] further describing the process of vibration generation. HandPass considers a smartphone and hand as a force-bearing body for facilitating theoretical analysis. As depicted in Fig. 2(a), the structure of one force-bearing body with a initial height $X_{(0)}$ is represented by three critical parameters: 1) mass M; 2) spring constant C_s; and 3) damper coefficient C_d. When utilizing a finger tapping on the screen, the height of the body changes with time t and is denoted



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

Fig. 3. Employing the auto click machine to ensure identical tapping forces (a), collected HVRs of five users depicted in time (b) and frequency (c) domains, respectively, and finally comparing HVR feature similarity of all users (d).

as $X_{(t)}$. Learned from the Hooke [38] and Newton's second laws [39], we formalize

$$F_{(t)} = Ma(t) + C_s\big(X_{(0)} - X_{(t)}\big) + C_d V_t \tag{1}$$

where $F_{(t)}$ is the external force of finger tapping, $a(t)$ is the acceleration, $X_{(0)} - X_{(t)}$ is the vertical displacement denoted as $\Delta x(t)$, and V_t is the movement speed. To simplify the formula for analysis, we replace $a(t)$ and V_t in (1) with displacement $\Delta x(t)$

$$F_{(t)} = M\frac{d^2\Delta x(t)}{dt^2} + C_s\Delta x(t) + C_d\frac{d\Delta x(t)}{dt}. \tag{2}$$

According to (2), when the finger exerts the same force $F_{(t)}$ on the force-bearing body, the displacement $\Delta x(t)$ can be determined by parameters M, C$_s$, and C$_d$. So the vibration wave generation (i.e., vertical displacement) depends on the structure of the force-bearing body (i.e., hand)

$$h : \big[F_{(t)}, M, C_s, C_d\big] \rightarrow [\Delta x(t)] \tag{3}$$

where h is a mapping function between vibration response and hand structure under force $F_{(t)}$.

Nonlinearity Derivation: The human hand owning complex geometry and structure is regarded as a nonlinear response medium [10]. Thus, inputting vibration signals [i.e., $\Delta x(t)$] can generate new harmonics and intermodulation frequency components embedded in hand responses. Here, we introduce a nonlinear model to describe the nonlinearity derivation of HVRs in the frequency domain. The original vibration response $\Delta x(t)$ is decomposed into multiple signals of different frequencies

$$\Delta x(t) = \sum_{i=1}^{N} A_i \sin(2\pi f_i) \tag{4}$$

where f_i and A_i, respectively, express the frequency and amplitude of the ith signal component. We then denote the frequency components derived by nonlinearity as S_{new}

$$S_{\text{new}} = \text{Har}(\Delta x(t)) + \text{Inter}(\Delta x(t)) \tag{5}$$

$$S_{\text{new}} = \sum_{k=1}^{K} B_k \sin(2\pi f_k). \tag{6}$$

Har($\Delta x(t)$) is the harmonics (i.e., $n \cdot f_i$) and Inter($\Delta x(t)$) is the intermodulation (i.e., $f_1 + f_2, f_2 - f_1, \ldots, f_n - f_{n-1}$) of the original signals. B_k is the amplitude of the signal of frequency f_k. The new derived frequency components and corresponding amplitudes depend on the hand structure.

Vibration Propagation: The external force $F_{(t)}$ disappears as a finger stops tapping screen, thus the force-bearing body enters into a free vibration state. The amplitudes of vibration signals continually attenuate when propagating from the force-bearing point to the accelerometer as presented in Fig. 2(b). Subsequently, we model a vibration attenuation function to describe this process as follows:

$$y = S_{\text{new}} e^{-aD_p} \tag{7}$$

where a expresses the attenuation coefficient depending on the medium structure and D_p is vibration propagation distance. As illustrated in Fig. 2(b), vibration waves generated at the force-bearing point propagates through the medium along multiple paths, such as direct (e.g., *path 2*) and reflected ones (e.g., *paths 1 and 3*). Note that, the vibration wave along *path 3* propagates through not just the smartphone but also the hand to the accelerometer. The actual propagation is a complex process determined by the medium structure and therefore the propagation paths are difficult to be predicted. We just display three representative paths to easily understand the vibration propagation process. To sum up, identical vibration sources S_{new} produce unique HVRs responses even if triggered by the same force when distinct hands hold one smartphone.

B. Study Case Employing HVRs for User Authentication

We subsequently verify the feasibility of utilizing HVRs to distinguish user identity. Before data collection, we tell users about the goal of this study. All users hold the smartphone OPPO Reno 6 with two gestures as depicted in Fig. 1. Now, we explore the distinguishability of HVRs among users, triggered by exactly the same external force. To ensure the inputting force is identical, we employ an auto-click machine to tap on the screen as Fig. 3(a). The built-in accelerometer of the smartphone collects vibration data in real time. In the following, we randomly select five users and, respectively, present their averaging HVRs (using gesture G_1) in both time and frequency domains in Fig. 3(b) and (c). These vibration signals are unique in terms of envelopes and amplitudes. According to the analysis in Section III-A, these differences in the time domain are caused by the unique mass, damping coefficient, and attenuation coefficient of each hand. Moreover, due to the differences in nonlinear responses of hands, the HVRs' amplitudes of any two users in the frequency domain are also dramatically different in Fig. 3(c). To fully understand the HVR uniqueness among all users, we obtain their fine-grained response features extracted in Section VI-C and

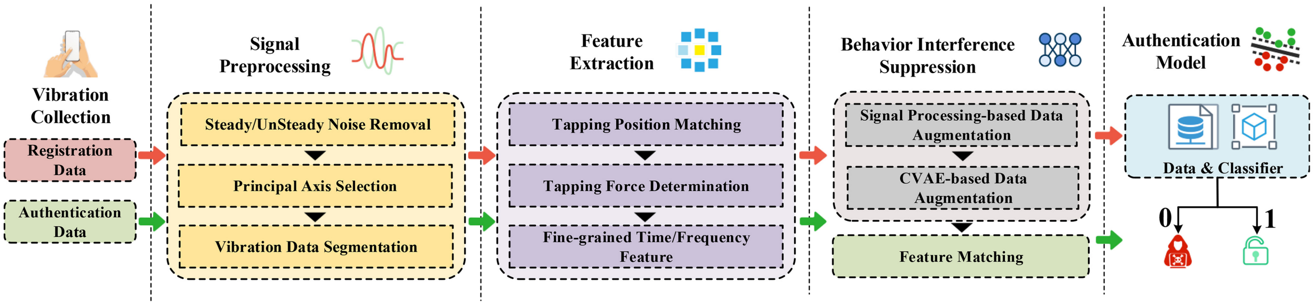

Vibration Collection
Registration Data
Authentication Data
Signal Preprocessing
Steady/UnSteady Noise Removal
Principal Axis Selection
Vibration Data Segmentation
Feature Extraction
Tapping Position Matching
Tapping Force Determination
Fine-grained Time/Frequency Feature
Behavior Interference Suppression
Signal Processing-based Data Augmentation
CVAE-based Data Augmentation
Feature Matching
Authentication Model
Data & Classifier
0
1

Fig. 4. System architecture of HandPass, consisting of five modules: vibration collection, signal preprocessing, feature extraction, behavior interference suppression, and authentication model.

then utilize cosine similarity [40] to calculate their consistency. Cosine similarity is a widely used metric to measure the consistency between two samples. Finally, the vibration signal similarity distribution of the same user (Intrasim) and across users (Intersim) are depicted in Fig. 3(d). Intrasim is significantly larger than Intersim with an average interval of more than 0.65. This study demonstrates that HVRs are distinguishable across users while consistent for identical users. Motivated by this observation, we further employ the HVRs to achieve continuous user authentication on smartphones.

IV. System Overview

As illustrated in Fig. 4, HandPass consists of five major modules achieving the goal from original vibration signal collection to identity authentication. In the registration phase, a user taps on the screen of mobile devices multiple times at different positions with habitual hand-holding gestures and forces; meanwhile, the accelerometer captures vibration signals behind the scenes. In *Signal Preprocessing*, HandPass removes steady/unsteady noise caused by body movements and segments vibration data into frames, each frame representing one tapping event. Nevertheless, even for identical users, he/she generates also inconsistent HVRs as forces and positions change; thus, our system needs to determine the above two factors of data segments in *Feature Extraction* for accurate feature matching in authentication phase. In the following, HandPass extracts fine-grained features from time and frequency domains to represent unique hand structure. Finally, we leverage the registration profile of one user to construct our authentication model and save it in a database.

During an authentication phase, HandPass conducts a similar processing flow with the registration one in addition to adding *Behavior Interference Suppression* module. As mentioned above, for ensuring user experience, HandPass allows users to input a few HVRs to register legitimate identity information. However, HVRs show diversity as related factors like handholding gesture and tapping force change, and thus such a few registered HVRs are not enough to represent the characteristics of user hand structures. To solve this problem, HandPass first utilizes linear signal envelope stretch and amplitude warping to generate new HVRs as a single factor changes. Furthermore, a customized CVAE-based model aims to complete nonlinear data augmentation when multiple factors simultaneously change. Finally, HandPass extracts features of HVRs and matches the registration data in the database which has the most similar tapping position and force with newly inputting HVRs. Our authentication model judges the legitimacy of each HVR; if the feature similarity is larger than the predefined threshold, user authentication is successful; otherwise, the authentication fails.

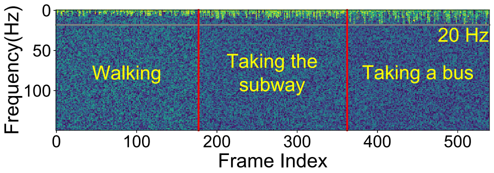


Fig. 5. Noise frequency distribution induced by human body movements in three common scenarios (without unsteady noise event arising).

V. Signal Preprocessing

To leverage HVRs for continuous user authentication on smartphones, HandPass first needs to remove the interference from noise induced by human body movements. Subsequently, collected signals are segmented into frames corresponding to tapping events, for further data processing.

A. Steady/Unsteady Noise Removal

Users being in dynamic states like walking and taking public transport when utilizing smartphones, inevitably introduces external noise polluting original HVRs. These noises are always divided into two types, i.e., steady and unsteady, referring to the stability of signal component distribution [41]; the former means that current noise keeps stable frequency components that can be inferred by historical data, while another type is highly unpredictable because of its mutability. To be specific, as one user moves with keeping consistent motion (e.g., speed and orientation), the frequency distribution of introduced noise is steady in the temporal dimension. We present collected vibration signals of body movements caused by daily activities without tapping the screen in Fig. 5. It displays that the frequency distribution of noise keeps stable and a high-pass filter with a cutoff frequency of 20 Hz can readily filter out it. Nevertheless, some emergencies, such as brake and sharp turning cause instantaneous frequency pattern variations in vibration signals (called unsteady noise), which always

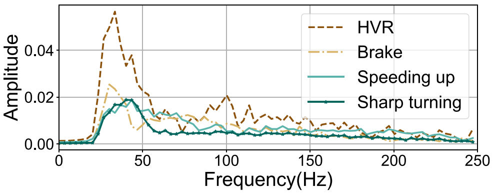


Fig. 6. Frequency components of unsteady noise events.

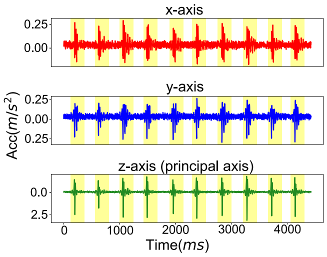


Fig. 7. Vibration data collected by the triaxial accelerometer.

trigger the authentication mechanism and cause event misjudgment. Fig. 6 displays common unsteady noise events, while their frequency components easily overlap and pollute with legitimate HVRs. To ensure the availability of input signals, HandPass chooses to abandon data segments corresponding to unsteady noise data. Subsequently, HandPass constructs a GBDT model applying to the vibration signal type classification. GBDT is widely used because of its high accuracy and stability in the small-size sample classification task [42]. Daily unsteady noise data segments, such as brake, speed-up/down, making a sharp turning, and driving on nonflat roads, are considered as negative samples; original HVRs are positive ones. We collect more than two thousand samples from real environments for training the GBDT model and each class accounts for 50%. Finally, the optimized parameter combination is discovered by a grid search way [43] that is a solution to choose a set of optimal parameters for a learning algorithm. HandPass unitizes 80% data for training and the remaining part for testing, finally obtaining a satisfactory accuracy up to 99.47%. Once the label of one vibration data segment is 0, it is regarded as an invalid signal and does not be further processed.

B. Principal Axis Selection and Data Segmentation

Vibration data sensed by the triaxial accelerometer is projected on three axes, namely, *x*/*y*/*z*-axis. As depicted in Fig. 7, the signal amplitudes of three components are distinct and rely on the smartphone rotation angle. In this case, the *z*-axis owns the largest vibration component and hence with the highest signal-noise-ratio, denoted as the *principal axis*. Compared with the other two ones, the vibration signal belonging to the principal axis has a higher amplitude fluctuation during a user-device interaction period. Thus, it is more suitable for being used to tapping event detection and data segmentation. To determine the principal axis, HandPass calculates the data variances of three axes, respectively, in the time dimension. The variance presents the amplitude fluctuation of each axis and the principal axis exhibits the largest variance.

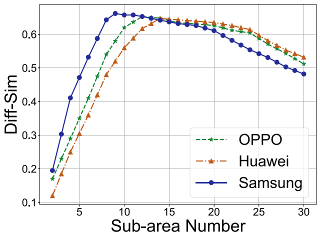


Fig. 8. Similarity differences as setting the number of subareas as distinct values.

For further feature extraction, it is needed to segment vibration signals into each tapping event. HandPass applies a threshold-based sliding window approach to determine boundary points corresponding to each event at the principal axis. As illustrated in Fig. 7, the vibration data caused by tapping events at the principal axis is marked with a yellow shadow, which has larger amplitudes than nontapping periods. According to our experiment analysis, we set the window size as 200 samples and the sliding step is set as 50 samples. We detect tapping events based on the following two conditions: 1) the absolute amplitude of the 1st sample and 2) the averaging amplitude of all samples in the window are three times as many as samples in nontapping windows. The goal of verifying the averaging amplitude is to avoid event misjudgment caused by value jitter from hardware defects. Note that, the threshold used to detect tapping events is empirically calculated from user registration data, i.e., the averaging value of all samples belonging to tapping events.

VI. Feature Extraction

To ensure the behavior consistency of matched registered profiles and newly captured HVRs, HandPass first determines the tapping position and force of each HVR. Then, we extract fine-grained features from time and frequency domains, thereby effectively representing user identity.

A. Tapping Area Partition

Even for identical users, tapping on distinct positions of a screen still generates distinct HVRs, due to differences in force-bearing points and vibration propagation paths. For collecting enough vibration response data to represent user identity in a registration stage, the most intuitive way is to ask users to tap on all positions of a screen. Obviously, this way requires deep and cumbersome user involvement, which is impractical for any application scenario. Therefore, HandPass attempts to divide each screen into multiple virtual subareas by gray lines as depicted in Fig. 3(a); users only need to tap on each subarea a few times to complete HVR registration. The rationality behind this solution is that HVRs belonging to identical subareas keep high similarity, thus a few ones can represent their characteristics. Nevertheless, setting the number of subareas (NS) is a balance problem that should be further studied. To be specific, setting it with a large value, the user still needs to input vast registered HVRs; a small NS leads to low HVR similarity and consistency even if in the same subarea. Thus, an ideal case should ensure that the

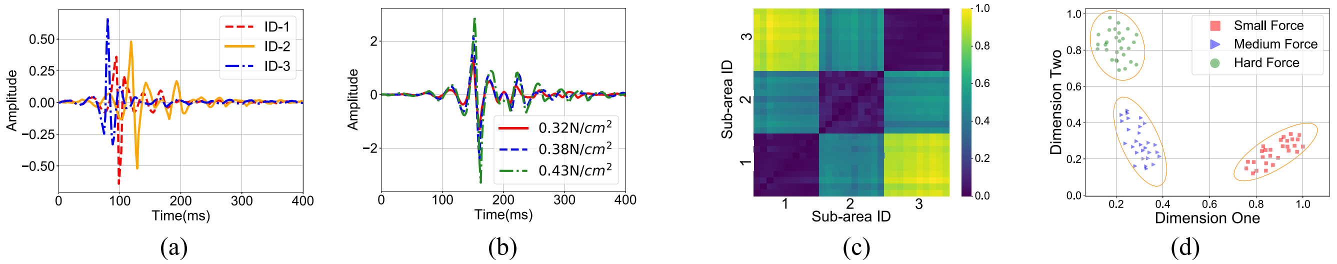


Fig. 9. Tapping position and force determination, first presenting theirs impacts on HVR envelopes in (a) and (b), then further presenting similarity variation by DTW in (c) and t-SNE in (d). (a) HVRs from three tapping positions. (b) HVRs from three tapping forces. (c) DTW distances of HVRs from three tapping subareas. (d) HVRs separation from three tapping force ranges by t-SNE.

similarity difference value (Diff-Sim) between Intrasim and Intersim keeps enough large, while NS is as small as possible. We make OPPO Reno 6, Huawei Mate30 Pro, and Samsung Galaxy S8 smartphones with common screen sizes (i.e., 6.43, 6.53, and 5.8 In) as examples, and explore the dependency relationship between Diff-Sim and NS value. We divide the screen into subareas with multiple numbers, then collect corresponding HVRs while calculating their similarity as shown in Fig. 8. These devices all present a similar Diff-Sim changing pattern consisting of two states: increasing first and then decreasing. It is surprising to discover that the Diff-Sim displays a decreasing trend when NS is greater than 12 (making OPPO Reno 6 an example). By analyzing the similarity variation under distinct subarea sizes, we conclude that if NS is too large, HVRs across subareas still maintain high similarity, resulting in larger Intersims and hence small Diff-Sim. To sum up, smartphones with distinct screen sizes should be set with an appropriate NS, e.g., 12 for OPPO Reno 6 and 9 for Samsung Galaxy S8.

B. Tapping Position and Force Determination

In the following, we present HVR patterns when a user taps on the screen on three positions (i.e., in ID-1, ID-2, and ID-3 subareas) in Fig. 9(a) while using distinct forces (i.e., 0.32 N/cm^2, 0.38 N/cm^2, and 0.43 N/cm^2) in Fig. 9(b). Obviously, even if from identical users, HVRs still change since the above two factors affect both vibration envelopes and amplitudes. To ensure the successful rate of authenticating legitimate users, HandPass first should estimate tapping positions and forces before the similarity calculation of paired HVRs.

Tapping Position Matching: We ask one user to tap on each subarea 10 times and the default number of subareas is set as 12. Then the dynamic time warping (DTW) [44] distance is utilized to measure the consistency of any two HVRs. DTW is a widely used similarity judgment metric and a small DTW distance indicates that HVRs are similar. As depicted in Fig. 9(c), the similarity of identical subareas HVRs is higher than that across distinct ones. Since vibration signals keep high consistency in identical subareas, we only need to judge the subarea of HVRs belonging to first, for matching the compared ones. The intuitive idea of inferring the tapping subarea is calculating the DTW distances between the newly captured HVR and all registered ones. Nevertheless, we discover that the smallest DTW value between two HVRs means they belong to the same subarea with the highest probability, while the centroid vector [45] is always utilized to represent the characteristics of data in the same class (subarea). Therefore, considering the computational cost, we can calculate the DTW distances between the newly captured HVR and the centroid vector of each subarea. The centroid vector can be obtained as follows:

$$c_k = \sum_{i=1}^{T_s} \mu_{ki} / T_s \tag{8}$$

where c_k is the centroid vector and T_s is the number of HVRs collected in the kth subarea. μ_{ki} is the ith hand response. Because of HVRs consisting of distinct numbers of samples, we employ linear interpolation [46] to make them keep identical lengths before centroid vector calculation. During the authentication phase, we calculate the DTW distances between the newly captured HVR and all the centroid vectors of the registration data. The tapping subarea (i.e., position) of newly captured HVR is determined as the centroid vector which has the smallest DTW distance with it. By only calculating the distance with respect to the centroid vectors, the computational cost of DTW calculation in HandPass is reduced from $O(NT_s)$ to $O(N)$, where N is the number of subareas.

Force Strength Determination: To explore the effect of tapping forces imposed on HVRs, one user taps on the screen with small, medium, and hard forces, respectively, while keeping other behavior parameters identical. We first apply the t-SNE method [47] to original HVRs for visualization, which makes similar samples cluster in closing positions. As displayed in Fig. 9(d), HVRs corresponding to distinct forces has larger differences than those under similar forces. Therefore, we need to first estimate the force of HVRs to achieve accurate data matching. As described in (2) and (7), the damping and attenuation coefficients of each hand are fixed. The total vibration amplitude a_t and duration time d_t of one HVR increase with force f_s. This theoretical basis motivates us to search for a binary linear relation [48] of f_s, a_t, and d_t for tapping force estimation

$$f_s = w_1 \cdot a_t + w_2 \cdot d_t \tag{9}$$

where w_1 and w_2 are the weights to balance two force-dependent parts. We utilize stochastic gradient descent (SGD) [49] to estimate the weights, which alternately updates

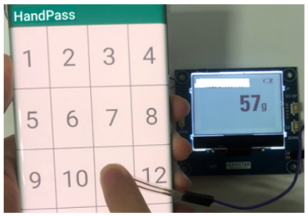


Fig. 10. Film pressure sensor.

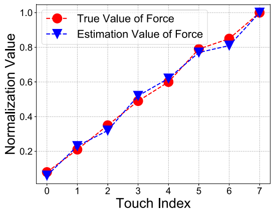


Fig. 11. Accuracy of force strength estimation in HandPass.

them according to the difference value (i.e., v_d) of the estimated force f_s and the true force strength. We first initialize w_1 and w_2 with random values and then update them as follows:

$$w^{n+1} = w^n + \gamma \cdot {v_d}^n \tag{10}$$

where w^n denotes either w_1 or w_2 and ${v_d}^n$ denotes the value difference at the nth parameter update cycle. This iteration process stops until the error is less than 0.01 in HandPass. Note that the ground truth of force is measured by a film pressure sensor (Fig. 10) with an accuracy of 1 g, while the sum vibration amplitude and duration time can be calculated from registered HVRs.

We randomly choose HVRs from ten users and calculate the average values of w_1 and w_2 (denoted as $\overline{w_1}$ and $\overline{w_2}$, respectively). Then, we choose HVRs generated from eight different forces from the other five users and calculate the average estimated forces using (9). Fig. 11 indicates that the force estimates match the ground truths well and thus this proposed estimation way is effective. With obtaining the HVR's tapping position and force, the newly inputting HVR can be matched with the registration ones for further similarity comparison.

C. Features in Time and Frequency Domains

Extracting effective features to represent user identity is critical for improving authentication performance. In HandPass, we employ the customized features of vibration signals and common statistics ones in both time and frequency domains.

Customized Features: We first explore the customized features to represent hand structure-dependent vibration envelope characteristics. As depicted in Fig. 12, making the partial HVR as an example, it is divided into three (N) vibration cycles ($\{\upsilon_c^1, \upsilon_c^2, \ldots, \upsilon_c^N\}$) by searching its peaks ($\{\rho^1, \rho^2, \ldots, \rho^N\}$) and valleys ($\{\nu^1, \nu^2, \ldots, \nu^{N+1}\}$). The duration time of nth cycle are denoted as two parts, which are τ_n^1 (from ν^n to ρ^n) and τ_n^2 (from ρ^n to ν^{n+1}). Each cycle displays a reciprocating motion track, that is, a hand starts to move under external forces until elastic resistance counteracts this force, then gradually gets back to the original state. As described in Section III-A, the hand structure parameters (i.e., C_s, C_d, and M) jointly determine vibration amplitude and duration, hence forming unique HVRs across users. In the following, we extract features of the three types from cycles while each of them consists of one/multiple elements.

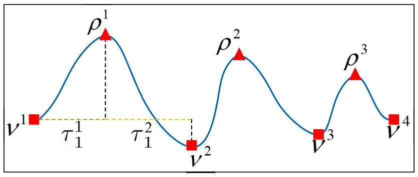


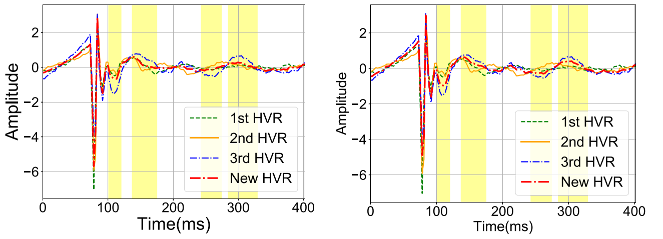
Fig. 12. Illustration of customized feature extraction.

1) *Time Duration:* Taking the first circle's duration as a benchmark, we obtains relative duration variations of other ones as $\{([\tau_1^1 + \tau_1^2]/[\tau_1^1 + \tau_1^2]), ([\tau_2^1 + \tau_2^2]/[\tau_1^1 + \tau_1^2]), \ldots, ([\tau_N^1 + \tau_N^2]/[\tau_1^1 + \tau_1^2])\}$; then calculating the time duration ratios ($\{(\tau_1^1/\tau_1^2), (\tau_2^1/\tau_2^2), \ldots, (\tau_N^1/\tau_N^2)\}$) between rising and falling edges of circles to characterize their temporal differences of reciprocating motion.
2) *Velocity:* We obtain the mean velocity of nth cycle, regarded as $\overline{\upsilon}_c^n$; furthermore, the relative mean velocity changing ratio is $\{(\overline{\upsilon}_c^1/\overline{\upsilon}_c^1), (\overline{\upsilon}_c^2/\overline{\upsilon}_c^1), \ldots, (\overline{\upsilon}_c^N/\overline{\upsilon}_c^1)\}$.
3) *Distance:* Each cycle's moving distance can be calculated by accumulating the elements in υ_c and thereby obtaining relative distance variations of all cycles, that is, $\{(\upsilon_c^{1d}/\upsilon_c^1), (\upsilon_c^{2d}/\upsilon_c^1), \ldots, (\upsilon_c^{Nd}/\upsilon_c^1)\}$.

In HandPass, for ensuring time-domain feature stability, we select five circles with the highest *signal noise ratio* to extract features representing user identity. As depicted in Fig. 3(c), HVRs from distinct users in frequency domain shows unique spectrum distribution. We denote the amplitudes of HVR in the range of 0–250 Hz with the 1 Hz step as $\{\vartheta_1, \vartheta_2, \ldots, \vartheta_{250}\}$.

1) *Energy Increase:* We count the energy increasing pattern to represent holistic spectrum distribution, which is
$$\left\{\frac{\sum_{i=1}^{1} \vartheta_i}{\sum_{i=1}^{250} \vartheta_i}, \frac{\sum_{i=1}^{2} \vartheta_i}{\sum_{i=1}^{250} \vartheta_i}, \ldots, \frac{\sum_{i=1}^{250} \vartheta_i}{\sum_{i=1}^{250} \vartheta_i}\right\}.$$
2) *Energy Variation in Frequency Bands:* The frequency band is divided into 25 equal parts and each one spans 10 Hz, hence obtaining 25 spectrum data from one HVR, denoting as $\{\gamma_{1,} \gamma_{2,}, \ldots, \gamma_{25}\}$. The spectrum energy variation is $\{^{\gamma_1}/_{\sum_{i=1}^{25} \gamma_i}, ^{\gamma_2}/_{\sum_{i=1}^{25} \gamma_i}, \ldots, ^{\gamma_{25}}/_{\sum_{i=1}^{25} \gamma_i}\}$.

Statistics Features: The statistic ones reflect the global characteristics of signal envelopes, which are verified to effectively represent biometric characteristics. HandPass first extracts mean, variance, maximum difference, and standard deviation of HVRs (time) and 25 bands (frequency). Moreover, skewness, kurtosis, form factor, and crest factor [50] are leveraged to measure the value distribution characteristics of HVR envelopes. Furthermore, Mel-frequency cepstrum coefficients (MFCCs) are widely used in speech recognition to extract the linear and nonlinear characteristics of signals, thereby capturing subtle signal changes. Thus, HandPass segments each HVR into frames with 100 samples to extract the 128-order MFCCs; calculating the first and second order difference of the mel-cepstrum coefficients to further capture



(a) (b)

Fig. 13. New HVR generation using three original ones with two weight groups, i.e., {0.32, 0.41, 0.27} in (a) and {0.16, 0.29, 0.55} in (b). (a) 1st new generation HVR. (b) 2nd new generation HVR.

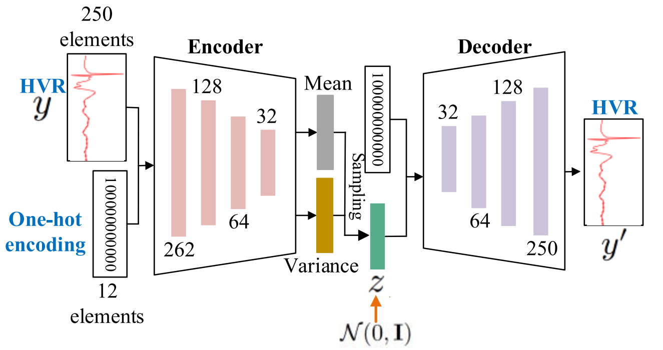


Fig. 14. Structure of our CVAE-based model for nonlinear HVR augmentation.

signal pattern traits. Our system extracts an original vector owning large than 600 elements, including customized and statistical features from each HVR. To speed-up the authentication model training and remove redundant information, we employ the Laplacian Score [51] to measure the importance of features and save the top-86 features in the score ranking, which have scores larger than the threshold 0.5.

VII. Behavior Interference Suppression

Offering sufficient HVR registration samples for building authentication models is a basic as well as nontrivial task, due to the following two reasons. On the one hand, a large-size data set helps the model fully absorb the intrinsic structures of HVRs and thereby enhancing authentication accuracy. On the other hand, data collection is accompanied by burdensome environment configuration, sample labeling, and consuming huge manpower. To break this dilemma, we study HVR variation patterns induced by distinct tapping behaviors, thereby generating newly unseen registration samples.

A. Signal Processing-Based Data Augmentation

As one user taps on distinct positions (in the same subarea), the main difference existing in their HVRs is envelope amplitude. Thus, HandPass can partially adjust amplitudes of original HVRs to simulate data collected at untapped positions. Referring to a DTW-based Barycentric Averaging method [52], we generate new samples by jointly assimilating the characteristics of multiple original HVRs. To be specific, we obtain new samples by the weighted average operation on selected original HVRs, while continuously optimizing the weight by $\arg\min_{w_n^-} \sum_{n^-=1}^{N^-} w_n^- \mathrm{DTW}(y^{\mathrm{new}}, y_{n^-})$. N^- is the total number of selected HVR templates and y^{new} is the newly generated one. By this method, new HVRs nicely follow the manifold envelope shape traits of original HVRs while creating enough diversity in amplitudes. We take three raw sample templates as an example and average them with two groups of weights (i.e., $\{0.32, 0.41, 0.27\}$ and $\{0.16, 0.29, 0.55\}$). As shown in Fig. 13(a) and (b), the newly generated HVRs still retain the original shape but arise changes in four local areas marked as the yellow shadow. Moreover, the tapping force mainly affects the duration of each authentication-triggered vibration event. For HVRs with distinct tapping forces, their envelopes are corresponding warped (i.e., either expansion or constriction). Considering that the common time of tapping events is ranging from 0.25 to 0.4 s, HandPass thus can zoom in/out the envelope of original HVRs in this range to generate new ones representing various forces.

B. Learning-Based Data Augmentation

Except for the aforementioned linear data augmentation, HandPass further explores the learning-based method acting as a complement for completing nonlinear HVR transformation. Currently, benefiting from performance efficiency and stability, the probabilistic generative model named variational autoencoder (VAE) gains wide concern in the deep learning community and is always applied to data augmentation, feature compression, and classification tasks. In HandPass, we construct a customized version relying on it to complete the goal of unseen HVR generation. Fig. 14 depicts the structure and workflow of our model that consists of two parts, i.e., encoder and decoder. The encoder $q_\phi(z|y)$ aims to infer the latent feature representations z of original HVR y and the decoder $p_\theta(y'|z)$ reconstructs output y' derived from z. Due to the prior probability distribution $p_\theta(z)$ of the latent variable z following a Gaussian distribution $\mathcal{N}(0, \mathbf{I})$, thus $q_\varphi(z|y)$ actually learns the distribution parameters presented by a mean and standard deviation. Subsequently, the VAE construction is achieved by maximizing the lower bound of the marginal likelihood [53] that can be written as

$$\begin{aligned}\log p_\theta(y) &\geq \mathbb{E}_{q_\phi(z|y)}\left[-\log q_\phi(z \mid y) + \log p_\theta\left(y', z\right)\right]\\ &= -D_{KL}\left(q_\phi(z \mid y) \| p_\theta(z)\right) + \mathbb{E}_{q_\phi(z|y)}\left[\log p_\theta\left(y' \mid z\right)\right]\end{aligned} \tag{11}$$

where D_{KL} denotes the Kullback–Leibler (KL) divergence [54], which guides minimizing the distribution difference between $q_\phi(z|y)$ and $p_\theta(z)$. In this case, after utilizing the reparameterization trick and *Stochastic Gradient Variational Bayes* [53] to maximize (11), the VAE model can generate new HVRs as giving a random z sampled from a priori normal distribution.

However, HVRs across subareas present distinct vibration characteristics and thus our model should map one specific HVR distribution for every subarea. To meet this purpose, we first conduct one-hot encoding for all subarea labels and then concatenate them with corresponding HVRs, as presented in Fig. 14. Thus, VAE is translated into CVAE which introduces tapping position information during the training process. In this case, (11) can be written as follows:

$$\log p_\theta(y \mid k) \geq -D_{KL}\big(q_\phi(z \mid y, k) \| p_\theta(z \mid k)\big) + \mathbb{E}_{q_\phi(z|y,k)}\big[\log p_\theta(y' \mid z, k')\big]. \quad (12)$$

To maximize the likelihood $\log p_\theta(y \mid k)$, the training process of our CVAE-based model should minimize the loss function derived from the above lower bound

$$D_{KL}\big(q_\phi(z \mid y, k) \| \mathcal{N}(0, I)\big) + \text{BCE}\,\text{Loss}\big(y, y'\big) \quad (13)$$

where $\text{BCELoss}(y, y') = y \log y' + (1 - y) \log(1 - y')$. The former part plays the role of driving the z' distribution to approach a normal distribution, while the latter measures a reconstruction loss between input and the generated HVRs. When obtaining a trained CVAE model, we sample multiple z from a normal distribution and feed them into the decoder to generate unseen HVRs. The activation function of the last full connection layer is Sigmoid and the batch size is set as 64. The training termination condition is the epoch of 1000 or the averaging HVR reconstruction is 0.01.

VIII. Authentication Model

Authentication models judge user identity by comparing the feature similarity between the registration and authentication HVRs. If the similarity is larger than the predefined metric, the authentication is successful. Common authentication models can be summarized as the following two categories: 1) one-class mode, which only relies on the positive HVRs of the legitimate user for model construction; and 2) two-class mode, which can utilize both positive and negative samples for model training. As the exploratory study using HVRs for continuous authentication, we employ the two modes to fully measure the system performance.

One-Class Mode: After capturing the new HVR, we match the registered HVR owning the same tapping area and similar force. Then, we calculate the cosine similarity of their feature vectors. The similarity value ranges from 0 to 1, where 1 means that the two HVRs are the same. In HandPass, the threshold is the average similarity of any two feature vectors under similar forces and identical tapping areas among the registration data.

Two-Class Mode: We first study the machine learning model to evaluate HandPass's performance. HVRs from the legitimate user with the same tapping subarea and similar forces in the registration data are labeled as 1, and other HVRs from distinct users are denoted as 0. We choose the following models, i.e., support vector machine (SVM) [55] and random forest (RF) [56], which have presented a superior performance in identity classification tasks. We set the key parameters $\{\text{kernel}, \text{tol}\}$ of SVM as $\{\text{rbf}, 0.01\}$ by standard grid-search approach, while the RF's parameters $\{n-\text{estimators}, \text{max}-\text{features}\}$ setting as $\{65, 35\}$.

We then leverage the common multilayer perceptron to construct the deep learning-based model, with five layers owning the size of $\{172, 64, 32, 16, 2\}$. We label the HVR following the same way of our machine learning models. Moreover, we add the activation function Sigmoid to each layer for enabling the feature's nonlinear transition. During the training phase, the amount of data in each batch is 32 and the learning ratio is set as 0.01. The Cross-Entropy loss is leveraged to measure the differences between true and predicted labels and thereby update model parameters.

IX. Implementation

In this section, we introduce the HandPass's implementation, including experiment setting, data collection, and the utilized metrics for evaluating authentication performance.

Experiment Setup: The prototype and data collection of HandPass is implemented on OPPO Reno 6, Samsung Galaxy S8 and HUAWEI Meta30 Pro. Data analysis is conducted on a desktop equipped with an Intel i5-8400 CPU and 16G RAM running Windows 10 with JetBrains PyCharm 2021 software. HandPass employs the built-in accelerometer of smartphones to capture vibration responses when users tap on screens.

Data Collection: We recruit 94 volunteers denoted as $V_1 \sim V_{94}$ aged from 23 to 47 for HVR collection. The HVR is collected under three common application scenarios, which are sitting in the lab (static), taking a bus (half-mobile), and walking (mobile). The default number of subareas on a set is 12. Under each scenario, volunteers tap on random positions in each subarea 20 times with habitual forces for registering personal identity. Overall, we collect more than 60 000 vibration responses from all volunteers to construct the data set for our evaluation. In HandPass, we leverage 70% data training our CVAE-based data augmentation module, the threshold and learning-based authentication models. Finally, we enlarge the original data size by six times relying on our augmentation module by default.

Metrics: We leverage false acceptance rate (FAR), false rejection rate (FRR), F-Score and Accuracy as the metrics to comprehensively evaluate HandPass's performance. FAR measures the likelihood that an authentication system incorrectly accepts an access trial from an illegitimate user. FRR represents the likelihood that incorrectly rejecting the access attempt from a legitimate user. Accuracy expresses the overall probability of legitimate users being accepted and illegitimate users being rejected by an authentication system. An outstanding authentication mechanism should have low values of FAR and FRR, while owning high values of Accuracy.

X. Evaluation

In this section, we first evaluate the overall performance and then explore the impact of distinct parameter settings on authentication.

A. Overall Performance

We select one volunteer as the legitimate user (e.g., V_1), and the rest (e.g., $V_2 \sim V_{94}$) as illegitimate users. Relying on the cross-validation approach, each volunteer is selected as the legitimate user in turn, and other volunteers serve as illegitimate ones. Subsequently, the 70% of registered vibration responses from the legitimate user is employed to calculate the similarity threshold for the one-class model and train the

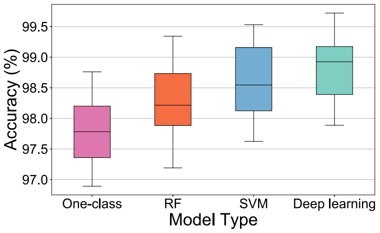


Fig. 15. Overall authentication accuracy of four authentication models.

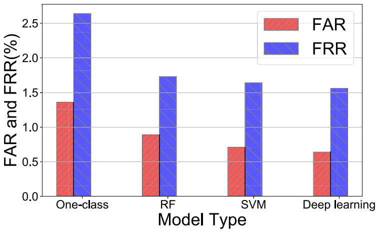


Fig. 16. FARs and FRRs of all authentication models.

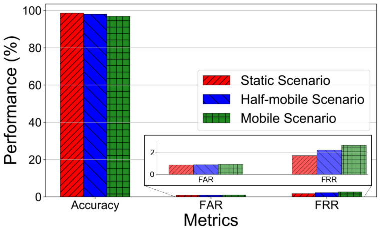


Fig. 17. Authentication performance in three scenarios.

learning-based models, respectively; the other 30% is utilized to test the four models. Figs. 15 and 16 present the overall accuracy distribution, FAR and FRR of all volunteers, while the three metrics' averaging values of all models are 98.36%, 0.84%, and 1.82%. Besides, we observe only the threshold-based model owns a weaker averaging accuracy below 98%. By analyzing the running mechanisms behind these models, we summarize the reason for this difference as follows: the one-class classifier works depend on empirical thresholds that cannot comprehensively represent the feature difference boundary between HVRs, thus lacking resilience and offering a lower accuracy. Nevertheless, the experiment result indicates that all models constructed by our HVR features can accurately recognize the identity of legitimate users and reject the access of illegitimate users.

B. Performance in Different Usage Scenarios

When using a smartphone, a user may be in different scenarios, such as sitting in the lab (static state), taking a bus (half-mobile state), and walking (mobile state). The amount of vibration noise introduced by these scenarios is distinct, which interferes with the original HVRs. In this section, we leverage vibration data from all volunteers in the static scenario to construct our models and evaluate HandPass's performance using data collected from three scenarios, respectively. As shown in Fig. 17, the obtained results in the static scenario are better than those in the other two ones. Nevertheless, the FRR range of all scenarios is from 1.79% to 2.84%, thus legitimate users can successfully pass the authentication with high accuracy even in nonstatic scenarios.

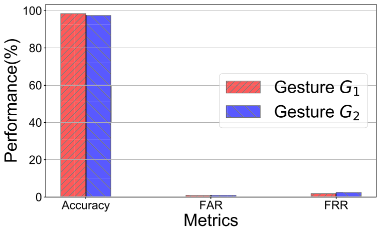


Fig. 18. Authentication accuracy of HandPass when using common gestures.

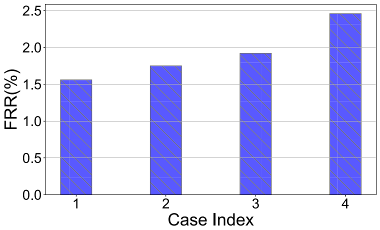


Fig. 19. Performance with/without data augmentation modules.

C. Performance in Different Hand-Holding Gestures

Users have several habitual hand-holding gestures when they use the smartphone, and the two most commonly used gestures are shown in Fig. 1. In this section, we evaluate the system performance under the two common hand-holding gestures. We collect vibration response data from volunteers and each volunteer taps 20 times in each subarea under the two gestures. The 70% of vibration responses from the legitimate user is used to train the deep learning-based model, and the rest 30% is used for testing. As shown in Fig. 18, the results obtained for two different gestures are similar. The value of FRR for G_2 is slightly higher than that for G_1, which is due to the unstable operation using only one hand. In general, HandPass can achieve good performance under two frequently used hand-holding gestures.

D. Impact of Data Augmentation Module

Authentication performance is closely dependent on whether there are enough registered HVRs to build the model. Nevertheless, to offer better user experiences, it is basic to allow users to register identity information with only a few HVRs. To solve the above dilemma, HandPass applies a data augmentation module consisting of signal processing and CVAE-based approaches to expand registered HVRs, thereby helping authentication models comprehensively characterize hand structure biometrics with enough training data. In this section, we evaluate the effectiveness of our augmentation modules in four cases: turning on both approaches, only using CVAE or signal processing-based approach, and without using both of them. As depicted in Fig. 19, the averaging FRRs of the former three cases, respectively, decrease by 0.91%, 0.73%, and 0.57%. Therefore, by generating new HVRs under unseen tapping

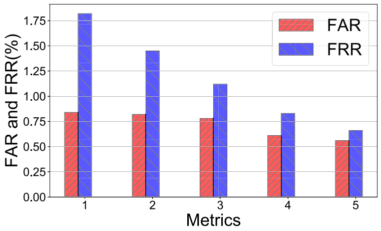


Fig. 20. FRR in different consecutive login times.

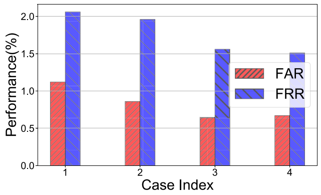


Fig. 21. Authentication accuracy when using distinct feature types.

behaviors, HandPass can further improve authentication performance.

E. Performance Under Multiple Authentication Attempts

Most existing systems allow the user to fail 3–5 consecutive times before being locked. If the legitimate user can pass the authentication with fewer attempts, the system is more user-friendly. In this section, we evaluate the FRR of our system with different consecutive numbers of attempts (1–5) until the legitimate user can successfully pass the authentication. As shown in Fig. 20, FAR and FRR both decrease with an increasing number of attempts. When the user tries to authenticate with four times, the average values of two metrics decrease to less than 1%. This result shows that HandPass can achieve a very high accuracy to authenticate the legitimate user with multiple attempts.

F. Impact of Feature Types

Whether effective features can be extracted to represent HVR is one of the important factors affecting authentication performance. HandPass uses common statistical features and customized features. In this section, we verify their effectiveness in four different configurations, that are, using only statistical features, only customized features, both types of features, and compressed features, respectively. As shown in Fig. 21, compared with only using one type feature, the joint version can improve the FAR by 0.45% and the FRR by 0.52%, respectively. This result shows that these two types of features can effectively complement each other to provide satisfactory performance. In addition, the compressed features can still maintain satisfactory accuracy while reducing computing resources. In general, the extracted features can effectively represent HVR characteristics, and their compressed versions still retain key identity information.

G. Stability of HVRs

The stability of features is a critical factor affecting the practicality of our authentication system. If HVRs change significantly in a short period, it will inevitably lead to an increased rate of rejecting legitimate users. In this section, after building HandPass, we ask ten users to provide one hundred HVRs every half month and then feed them into our models to evaluate the authentication accuracy. As shown in Fig. 22, FARs keep a stable value of around 1.6% in a 3-month time span. Therefore, HVRs and corresponding features captured by HandPass have satisfactory stability and can be used to effectively identify users.

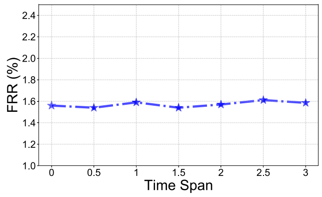


Fig. 22. FRRs in a 3-month usage span.

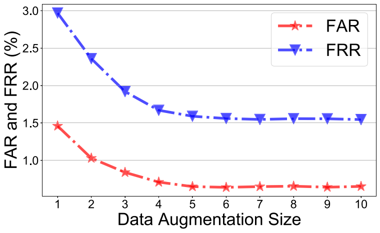


Fig. 23. FAR and FRR in ten distinct data augmentation sizes.

H. Impact of Data Augmentation Size

Registration data size utilized for model construction imposes a critical impact on authentication performance. The authentication accuracy can usually achieve higher by obtaining enough registration data for model training. In this section, we evaluate the HandPass's performance when using a data augmentation module to enlarge the original registered HVR at distinct sizes. We present the authentication accuracy with ten levels relative to the original data size in Fig. 23. The result indicates that when the data size enlarges to six times, HandPass can achieve a satisfactory and stable performance (i.e., 0.64% FAR and 1.56% FRR), with a normal range of user behavior variation. Therefore, in HandPass, we set the default augmentation size of six times relative to the original HVRs.

I. Impact of Accelerometer Sampling Rate

The sampling rates of the accelerometer determine the captured signal granularity. The higher rate means that more fine-grained HVR data can be obtained. We adjust the sampling rate of the accelerometer from 200 to 400 Hz at a step size of 50 Hz to capture HVRs from twenty randomly selected volunteers. As presented in Fig. 24, the FRR and FAR are 1.56% and 0.64%, respectively, when the sampling rate is 400 Hz. Currently, almost all smartphones own a sampling rate higher than 400*Hz*, thus HandPass can be fully compatible with commodity smartphones.

J. Detection of Potential Attacks

Except for accurately authenticating users, HandPass also requires defending against potential attacks. In this section, we

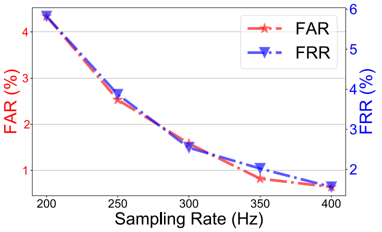


Fig. 24. FAR and FRR in different sampling rate settings.

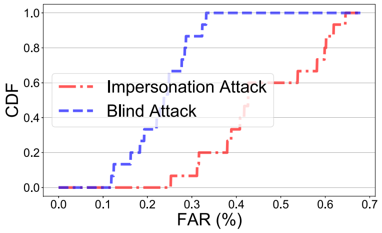


Fig. 25. FARs in two attacks.

utilize two common while typical attacks to evaluate security. We randomly choose fifteen volunteers for the experiments. One of them is selected as the legitimate user and other volunteers serve as attackers. Each attacker tries to pass the authentication system one hundred times in two attack scenarios.

Blind Attack: The attackers do not own any prior knowledge. They try to tap on a potential position, force, and handholding gesture, which may generate similar HVRs to spoof the authentication.

Impersonation Attack: The attacker has observed the registration process of legitimate users, including the tapping behaviors and other details. Then each attacker impersonates the operation of the legitimate user to pass HandPass. As shown in Fig. 25, the average FARs under the above two attacks are both below 0.5%. This result indicates that HandPass can effectively resist attacks to ensure system security.

K. System Latency and Power Cost

As an authentication system, latency is an important parameter worth considering. We set the default authentication frequency as ten times per minute on three smartphones. We measure the averaging time taken for each authentication process in 1 h and the value range is 249 ± 36 ms, demonstrating the capability of achieving real-time authentication. Moreover, HandPass costs an acceptable averaging power of 149.6 ± 5.4 mW on these smartphones. The power consumption includes the process from data collection to identity output. To sum up, HandPass can offer a satisfactory performance on resource cost and computation latency.

XI. Discussion

Although we have evaluated the stability of HVR over a 3-month span in Section X-G. However, with a longer time range, user hand structures may change caused by bone density and human weight variation, thus resulting in HVR variation. To solve this problem, we will study the changing pattern of HVR features over a longer time span in future work and remove its effects by regularly updating model parameters.

In addition to the current continuous authentication mechanism for mobile handholding devices, we can also try to leverage vibration responses triggered by the touchpad to authenticate users for fixed IoT devices such as smart locks. In this application scenario, the distinguishability of vibration signals across users comes from user fingers rather than the whole hand structure. At the same time, we can combine this vibration feature with PINs to achieve two-factor authentication and further improve security.

XII. Conclusion

In this article, we have proposed a continuous authentication mechanism named HandPass for smartphones, which utilized hand structure-dependent hand vibration response features. We have implemented the prototype system on three distinct smartphones and comprehensively evaluated its performance. Extensive experiments demonstrate that HandPass can achieve an overall authentication accuracy of 98.36%. Moreover, HandPass can effectively resist commonly potential threats from blind and impersonation attacks. We believe this is an important step toward the real-life adoption of continuous authentication on smartphones.

References

[1] S. M. Furnell, P. S. Dowland, H. M. Illingworth, and P. L. Reynolds, “Authentication and supervision: A survey of user attitudes,” *Comput. Security*, vol. 19, no. 6, pp. 529–539, 2000.

[2] Z. Sitová et al., “HMOG: New behavioral biometric features for continuous authentication of smartphone users,” *IEEE Trans. Inf. Forensics Security*, vol. 11, pp. 877–892, 2016.

[3] “Data breach.” trendmicro.com. 2019. [Online]. Available: https://www.trendmicro.com/vinfo/us/security_/definition/data-breach

[4] S. M. Furnell, J. P. Morrissey, P. W. Sanders, and C. T. Stockel, “Applications of keystroke analysis for improved login security and continuous user authentication,” in *Proc. IFIP Conf.*, vol. 54, 1996, pp. 283–294.

[5] L. Lu et al., “LipPass: Lip reading-based user authentication on smartphones leveraging acoustic signals,” in *Proc. IEEE INFOCOM*, 2018, pp. 1466–1474.

[6] I. Sluganovic, M. Roeschlin, K. B. Rasmussen, and I. Martinovic, “Using reflexive eye movements for fast challenge-response authentication,” in *Proc. ACM CCS*, 2016, pp. 1056–1067.

[7] N. Farhi, N. Nissim, and Y. Elovici, “Malboard: A novel user keystroke impersonation attack and trusted detection framework based on side-channel analysis,” *Comput. Security*, vol. 85, pp. 240–269, Aug. 2019.

[8] H. Feng, K. Fawaz and K. G. Shin, “Continuous authentication for voice assistants,” in *Proc. ACM MobiCom*, 2017, pp. 343–355.

[9] F. Lin, C. Song, Y. Zhuang, W. Xu, C. Li, and K. Ren, “Cardiac scan: A non-contact and continuous heart-based user authentication system,” in *Proc. ACM MobiCom*, 2017, pp. 315–328.

[10] J. Li, K. Fawaz, and Y. Kim, “Velody: Nonlinear vibration challenge-response for resilient user authentication,” in *Proc. ACM CCS*, 2019, pp. 1201–1213.

[11] S. Elliott, I. Stothers, and P. Nelson, “A multiple error LMS algorithm and its application to the active control of sound and vibration,” *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 10, pp. 1423–1434, Oct. 1987.

[12] G. Ke et al., “LightGBM: A highly efficient gradient boosting decision tree,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–9.

[13] J. Kim, J. Kong, and J. Son, “Conditional variational autoencoder with adversarial learning for end-to-end text-to-speech,” in *Proc. PMLR*, 2021, pp. 5530–5540.

[14] J. Yan, A. Blackwell, R. Anderson, and A. Grant, “Password memorability and security: Empirical results,” *IEEE Security Privacy*, vol. 2, no. 5, pp. 25–31, Sep./Oct. 2004.

[15] G. Ye et al., "Cracking android pattern lock in five attempts," in *Proc. ISOC NDSS*, 2017, pp. 1–15.

[16] D. Zhang, F. Liu, Q. Zhao, G. Lu, and N. Luo, "Selecting a reference high resolution for fingerprint recognition using minutiae and pores," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 3, pp. 863–871, Mar. 2011.

[17] C. Yan, Y. Long, X. Ji, and W. Xu, "The catcher in the field: A fieldprint based spoofing detection for text-independent speaker verification," in *Proc. ACM CCS*, 2019, pp. 1215–1229.

[18] W. Huang, W. Tang, H. Jiang, J. Luo, and Y. Zhang, "Stop deceiving! An effective defense scheme against voice impersonation attacks on smart devices," *IEEE Internet Things J.*, vol. 9, no. 7, pp. 5304–5314, Apr. 2022.

[19] H. Cao et al., "LiveProbe: Exploring continuous voice liveness detection via phonemic energy response patterns," *IEEE Internet Things J.*, early access, Dec. 13, 2022, doi: 10.1109/JIOT.2022.3228819.

[20] P. Perera and V. M. Patel, "Face-based multiple user active authentication on mobile devices," *IEEE Trans. Inf. Forensics Security*, vol. 14, pp. 1240–1250, 2019.

[21] N. Erdogmus and S. Marcel, "Spoofing face recognition with 3D masks," *IEEE Trans. Inf. Forensics Security*, vol. 9, pp. 1084–1097, 2014.

[22] M. Shirvanian and N. Saxena, "Wiretapping via mimicry: Short voice imitation man-in-the-middle attacks on crypto phones," in *Proc. ACM CCS*, 2014, pp. 868–879.

[23] H. Jiang, H. Cao, D. Liu, J. Xiong, and Z. Cao, "SmileAuth: Using dental edge biometrics for user authentication on smartphones," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 3, pp. 1–24, 2020.

[24] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song, "Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication," *IEEE Trans. Inf. Forensics Security*, vol. 8, pp. 136–148, 2013.

[25] Y. Song, Z. Cai, and Z.-L. Zhang, "Multi-touch authentication using hand geometry and Behavioral information," in *Proc. IEEE SP*, 2017, pp. 357–372.

[26] L. Zhang, S. Tan, J. Yang, and Y. Chen, "VoiceLive: A phoneme localization based liveness detection for voice authentication on smartphones," in *Proc. ACM CCS*, 2016, pp. 1080–1091.

[27] N. Zheng, K. Bai, H. Huang, and H. Wang, "You are how you touch: User verification on smartphones via tapping behaviors," in *Proc. IEEE ICNP*, 2014, pp. 221–232.

[28] S. Eberz, G. Lovisotto, K. B. Rasmussen, V. Lenders, and I. Martinovic, "28 blinks later: Tackling practical challenges of eye movement biometrics," in *Proc. ACM CCS*, 2019, pp. 1187–1199.

[29] 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. ACM MobiHoc*, 2019, pp. 201–210.

[30] L. Yang, W. Wang, and Q. Zhang, "VibID: User identification through bio-vibrometry," in *Proc. IEEE ICNP*, 2016, pp. 1–12.

[31] W. Wang, L. Yang, and Q. Zhang, "Touch-and-guard: Secure pairing through hand resonance," in *Proc. ACM Ubicomp*, 2016, pp. 670–681.

[32] J. Liu, C. Wang, Y. Chen, and N. Saxena, "VibWrite: Towards finger-input authentication on ubiquitous surfaces via physical vibration," in *Proc. ACM CCS*, 2017, pp. 73–87.

[33] W. Chen et al., "Taprint: Secure text input for commodity smart wristbands," in *Proc. ACM MobiCom*, 2019, pp. 17:1–17:16.

[34] H. Cao et al., "HandKey: Knocking-triggered robust vibration signature for keyless unlocking," *IEEE Trans. Mobile Comput.*, early access, Oct. 25, 2022, doi: 10.1109/TMC.2022.3216868.

[35] X. Xu et al., "TouchPass: Towards behavior-irrelevant on-touch user authentication on smartphones leveraging vibrations," in *Proc. ACM MobiCom*, 2020, pp. 1–13.

[36] T. Sabhanayagam, V. P. Venkatesan, and K. Senthamaraikannan, "A comprehensive survey on various biometric systems," *Int. J. Appl. Eng. Res.*, vol. 13, no. 5, pp. 2276–2297, 2018.

[37] H. Cao, D. Liu, H. Jiang, R. Wang, Z. Chen, and J. Xiong, "LIPAuth: Hand-dependent light intensity patterns for resilient user authentication," *ACM Trans. Sens. Netw.*, to be published.

[38] R. Hooke, *Lectures de Potentia Restitutiva, or of Spring Explaining the Power of Springing Bodies*, John Martyn, 2016.

[39] P. Bruce, "Newton's interpretation of Newton's second law," *Arch. Hist. Exact Sci.*, vol. 60, no. 2, pp. 157–207, 2006.

[40] D. Wang, H. Lu, and C. Bo, "Visual tracking via weighted local cosine similarity," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1838–1850, Sep. 2015.

[41] D. G. Karczub and M. P. Norton, *Fundamentals of Noise and Vibration Analysis for Engineers*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[42] H. Ma et al., "Retrieval-based gradient boosting decision trees for disease risk assessment," in *Proc. ACM SIGKDD*, 2022, pp. 3468–3476.

[43] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, no. 2, pp. 1–25, 2012.

[44] E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowl. Inf. Syst.*, vol. 7, no. 3, pp. 358–386, 2005.

[45] R. Carbó-Dorca, "A study on the centroid vector of a polyhedron," *J. Math. Chem.*, vol. 54, no. 1, pp. 61–71, 2016.

[46] E. Meijering, "A chronology of interpolation: From ancient astronomy to modern signal and image processing," *Proc. IEEE*, vol. 20, no. 3, pp. 319–342, Mar. 2002.

[47] N. Pezzotti, B. P. F. Lelieveldt, L. V. D. Maaten, T. Höllt, E. Eisemann, and A. Vilanova, "Approximated and user steerable tSNE for progressive visual analytics," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 7, pp. 1739–1752, Jul. 2017.

[48] W. G. Cochran, "Testing a linear relation among variances," *Biometrics*, vol. 7, no. 1, pp. 17–32, 1951.

[49] R. Johnson and T. Zhang, "Accelerating stochastic gradient descent using predictive variance reduction," *News Physiol. Sci.*, vol. 1, no. 3, pp. 315–323, 2013.

[50] Z. Li, Z. Yang, C. Song, C. Li, Z. Peng, and W. Xu, "E-eye: Hidden electronics recognition through mmWave nonlinear effects," in *Proc. ACM SenSys*, 2018, pp. 68–81.

[51] X. He, D. Cai, and N. Partha, "Laplacian score for feature selection," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 18, 2005, pp. 507–514.

[52] G. Forestier, F. Petitjean, H. A. Dau, G. I. Webb, and E. Keogh, "Generating synthetic time series to augment sparse datasets," in *Proc. IEEE ICDM*, 2017, pp. 865–870.

[53] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2013, *arXiv:1312.6114*.

[54] D. Johnson and S. Sinanovic, "Symmetrizing the Kullback–Leibler distance," *IEEE Trans. Inf. Theory*, submitted for publication.

[55] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, Jul./Aug. 1998.

[56] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

Hangcheng Cao (Student Member, IEEE) is currently pursuing the Ph.D. degree with the College of Computer Science and Electronic Engineering, Hunan University, Changsha, China.

From 2021 to 2022, he was a joint Ph.D. student with the School of Computer Science and Engineering, Nanyang Technological University, Singapore. He has published papers in ACM Ubicomp/IMWUT, IEEE ICDCS, IEEE TRANSACTIONS ON MOBILE COMPUTING, *IEEE Communications Magazine*, ACM MobiCom Workshop, and IEEE INTERNET OF THINGS JOURNAL. His research interests lie in the area of IoT security.

Mr. Cao is served as the reviewer for IEEE TRANSACTIONS ON MOBILE COMPUTING, *ACM Transactions on Sensor Networks*, and *Wireless Networks*.

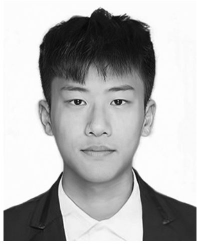
Hongbo Jiang (Senior Member, IEEE) received the Ph.D. degree from Case Western Reserve University, Cleveland, OH, USA, in 2008.

He is currently a Full Professor with the College of Computer Science and Electronic Engineering, Hunan University, Changsha, China. He was a Professor with Huazhong University of Science and Technology, Wuhan, China. His current research focuses on computer networking, especially, wireless networks, data science in Internet of Things, and mobile computing.

Dr. Jiang has been serving on the editorial board of IEEE/ACM TRANSACTIONS ON NETWORKING, IEEE TRANSACTIONS ON MOBILE COMPUTING, *ACM Transactions on Sensor Networks*, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, and IEEE INTERNET OF THINGS JOURNAL. He was also invited to serve on the TPC of IEEE INFOCOM, ACM WWW, ACM/IEEE MobiHoc, IEEE ICDCS, and IEEE ICNP. He is an Elected Member of Academia Europaea, a Fellow of IET and BCS, a Senior Member of ACM, and a Full Member of IFIP TC6 WG6.2.

Kehua Yang received the Ph.D. degree in computer science and engineering from Southeast University, Nanjing, China, in 2005.

He is currently an Associate Professor with the College of Computer Science and Electronic Engineering, Hunan University, Changsha, China. His major research interests include embedded systems, cyber–physical systems, and automotive systems.



Siyu Chen (Student Member, IEEE) received the B.S. degree in communication engineering from Hunan University, Changsha, China, in 2021, where he is currently pursuing the Ph.D. degree with the College of Computer Science and Electronic Engineer, Hunan University.

His research interests lie in the area of WiFi sensing.

Wenqi Wu received the master's degree from the Huazhong University of Science and Technology, Wuhan, China, in 2020.

He is an Algorithm Engineer with Shenzhen Technological Inc, Shenzhen, China. He is engaged in machine learning research, including intent recognition and user behavior prediction, and has extensive experience in industrial implementation.

Jiangchuan Liu (Fellow, IEEE) received the B.Eng. (cum laude) degree from Tsinghua University, Beijing, China, in 1999, and the Ph.D. degree from The Hong Kong University of Science and Technology, Hong Kong, in 2003.

He is a University Professor with the School of Computing Science, Simon Fraser University, Burnaby, BC, Canada. In the past, he worked as an Assistant Professor with The Chinese University of Hong Kong, Hong Kong, and a Research Fellow with Microsoft Research Asia, Beijing. He was an EMC-Endowed Visiting Chair Professor with Tsinghua University from 2013 to 2016. His research interests include multimedia systems and networks, cloud and edge computing, social networking, online gaming, and Internet of Things/RFID/backscatter.

Prof. Liu has served on the editorial boards of IEEE/ACM TRANSACTIONS ON NETWORKING, IEEE TRANSACTIONS ON BIG DATA, IEEE TRANSACTIONS ON MULTIMEDIA, IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, and IEEE INTERNET OF THINGS JOURNAL. He is a Steering Committee Member of IEEE TRANSACTIONS ON MOBILE COMPUTING and the Steering Committee Chair of IEEE/ACM IWQoS from 2015 to 2017. He is a TPC Co-Chair of IEEE INFOCOM'2021. He is a Fellow of The Canadian Academy of Engineering and an NSERC E.W.R. Steacie Memorial Fellow.

Schahram Dustdar (Fellow, IEEE) received the Ph.D. degree in business informatics from the University of Linz, Linz, Austria, in 1992.

He is currently a Full Professor of Computer Science (Informatics) with a focus on Internet technologies heading the Distributed Systems Group, TU Wien, Wein, Austria.

Dr. Dustdar is the recipient of multiple awards, including the TCI Distinguished Service Award in 2021, the IEEE TCSVC Outstanding Leadership Award in 2018, the IEEE TCSC Award for Excellence in Scalable Computing in 2019, the ACM Distinguished Scientist in 2009, the ACM Distinguished Speaker in 2021, and the IBM Faculty Award in 2012. He is a Founding Co-Editor-in-Chief of *ACM Transactions on Internet of Things* (ACM TIoT) as well as the Editor-in-Chief of *Computing* (Springer). He is an Associate Editor of IEEE TRANSACTIONS ON SERVICES COMPUTING, IEEE TRANSACTIONS ON CLOUD COMPUTING, *ACM Computing Surveys*, *ACM Transactions on the Web*, and *ACM Transactions on Internet Technology*, as well as on the editorial board of IEEE INTERNET COMPUTING and IEEE COMPUTER. He is an Elected Member of the Academia Europaea: The Academy of Europe, where he is a Chairman of the Informatics Section, an Asia–Pacific Artificial Intelligence Association President in 2021 and a Fellow in 2021, an EAI Fellow in 2021, and an I2CICC Fellow in 2021. He is a member of the IEEE Computer Society Fellow Evaluating Committee in 2022.