This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3049217, IEEE Transactions on Mobile Computing

Scanning the Voice of Your Fingerprint with Everyday Surfaces

Aditya Singh Rathore, Chenhan Xu, Weijin Zhu, Afee Daiyan, Kun Wang, *Senior Member, IEEE*, Feng Lin, *Senior Member, IEEE*, Kui Ren, *Fellow, IEEE* and Wenyao Xu, *Senior Member, IEEE*

Abstract—Due to the premise of uniqueness and acceptance, fingerprint has been the most adopted biometric technologies in high-impact applications (e.g., smartphone security, monetary transactions and international-border verification). Although there are an array of commercial fingerprint scanners across different sensing modalities including optical, capacitive, thermal and ultrasonic, existing fingerprint technologies are vulnerable to spoofing attacks via fake-finger [1]. In this paper, we investigate a new dimension of fingerprint sensing based on the friction-excited sonic wave (in simpler words, "voice of fingerprint") from a user swiping his fingertip on everyday surfaces. Specifically, we develop *SonicPrint* to leverage the intrinsic fingerprint ridge information in sonic wave for user identification. First, the complex ambient noise is isolated from the sonic wave using background isolation and adaptive segmentation models. Afterward, a series of multi-level friction descriptors that highlight the target fingerprint information is extracted. These descriptors are fed to a specially designed ensemble classifier for user identification. *SonicPrint* is practical as it leverages in-built microphones in smart devices, requiring no hardware modifications. As the first exploratory study, our experimental results with 31 participants over three different swipe actions on 12 different types of materials show up to a 98% identification accuracy.

Index Terms—Adoptable biometrics, fake-finger spoofing, surface friction, fingerprint-induced sonic effect, user identification.

1 Introduction

This paper asks the following question: can we enable everyday surfaces in the daily environment as fingerprint scanners, while ensuring security against spoofing attacks? Such a capability can transform the biometric domain by removing the dependency on special fingerprint hardware and reshape our interaction with surrounding objects. For instance, common surface materials (e.g., leather, plastic, fabric) could enable user identification while having resilience against fingerprint phantoms, i.e., fake-fingers [2], [3]. The smart-devices (e.g., smartwatch, voice assistant, curved smartphone) that have unique designs can now provide fingerprint sensing without any hardware modifications. Even more, we could enable biometrics without borders by allowing the users to transfer fingerprint-related attributes over communication platforms.

Till date, various types of fingerprint scanners have been proposed utilizing optical, capacitive and thermal sensors [4]. However, these scanners share a fatal weakness: vulnerability to fake-finger spoofing [1]. Even the upcoming in-display ultrasound sensors, targeted towards enhancing usability, are susceptible to 3D finger models [5]. As a countermeasure, researchers have suggested a secondary dimension of security (e.g., blood flow [6], precipitation [7]), yet they have poor generalization across spoof materials besides introducing additional hardware overhead. Other biometrics, including voice [8], [9] and faceID [10] can also be compromised using replay and impersonation attacks and thus fail to achieve high user acceptance.

Fig. 1. SonicPrint: A new dimension of fingerprint sensing by using sonic wave from surface-swipes for secure user identification.

Taking a step back, the requirement for our target biometric application is four-fold: (1) *Cost-effective:* the new scanner-less method should utilize low-cost off-the-shelf sensors that are widely used in smart-devices; (2) *Accessible:* the biometric trait should be available from surfaces with diverse flexibility, texture and composition; (3) *Easy-to-use:* we hope to enhance the user acceptance by providing free-form sensing; (4) *Secure:* compared to the traditional fingerprint methods, our proposed approach should be resilient against spoofing attacks using fake-fingers.

It is a known fact that when two objects slide against each other, kinetic energy is released in the form of sonic and heat. The harmonics of this friction-excited sonic are dependent on the surface characteristics of objects and their internal composition. Our key contribution is the observa-

- *A. Rathore, C. Xu, W. Zhu, A. Daiyan, W. Xu are with the Department of Computer Science and Engineering, University at Buffalo, the State University of New York, Buffalo, NY, 14261.*
- *K. Wang is with the Department of Electrical and Computer Engineering, University of California, Los Angeles, CA, 90095.*
- *F. Lin and K. Ren are with the Department of Computer Science and Technology, Zhejiang University, China, 310027.*

tion that the sonic wave from a user swiping his fingertip on a surface can serve as biometric traits. Since every person has a unique fingerprint, we hypothesize that the sonic waves resulting from two users swiping their fingertips on a common surface should be different. Although the statistical properties of sonic may change depending on user's swiping speed, pressure or surface roughness, the inherent uniqueness is dependent on the surface texture (i.e., fingerprint ridge patterns) and the finger's constitution. If this hypothesis holds, the fingerprint-induced sonic effect (FiSe) can be acquired from the microphone in smart devices. The goal of this work is to explore the knowledge and validation of a new fingerprint sensing modality and open discussions for emerging mobile security research.

Our goal is to transform everyday surfaces into fingerprint scanners. To achieve this, three challenges need to be addressed: (1) The FiSe is typically of low power and submerged in dynamic background noises. How to acquire the target FiSe without any information loss? (2) To enable high accessibility and acceptance, it is important to provide freedom to the users while swiping the surface. In the case where user's swiping speed and pressure is not controlled, how to select appropriate features which closely resemble the fingerprint? (3) For real-world applications, it is critical that the FiSe cannot be compromised. How to evaluate the vulnerability of our system which relies on characteristics of both fingerprint and audio domain?

In this work, we propose a systematic framework that leverages the FiSe of a user swiping on smartphone and other surfaces as a new biometric. We first validate the uniqueness of sonic patterns by comparing the resulting spectrum of fingerprints with different textures. Then, we leverage the underlying microphone in a smartphone to acquire the FiSe and investigate a sequence of spectral and wavelet denoising approaches for background isolation. An adaptive segmentation method is designed to remove the tap noise and other entities which can be easily misinterpreted as the target signal. Afterward, we propose a novel taxonomy that highlights the semantic relationship between fingerprint and audio domain, and identifies multi-level features that fundamentally share the same concept as fingerprint. Based on these insights, we design and implement our system, *SonicPrint*, to facilitate secure sensing of FiSe for user identification. Finally, a comprehensive evaluation is performed with 31 participants on 12 surfaces across six sessions over two months to validate the effectiveness and inclusiveness of *SonicPrint* under real-world scenarios.

Summary: Our contribution in this work is four-fold:

- We explore a novel fingerprint-based biometric approach for user identification. We find that when a user swipes his fingertip on a surface, the sonic wave contains intrinsic fingerprint information.
- We design and implement *SonicPrint*, an end-to-end biometric system to facilitate secure, accessible and user-friendly fingerprint sensing on everyday surfaces in practice.
- We validate the effectiveness and inclusiveness of *SonicPrint* through extensive experiments with results showing up to 98% accuracy. We conduct comprehensive studies to show the resilience of *SonicPrint* against fake-finger, replay, side-channel and ultrasonic attacks.
- We perform two case studies to demonstrate the promising applications of *SonicPrint* for group authentication and object identification.

2 Background and Preliminaries

In this section, we provide a background on friction-excited sonic waves and the rationale behind its uniqueness in terms of human-to-material interaction. We also perform a feasibility study to prove this concept.

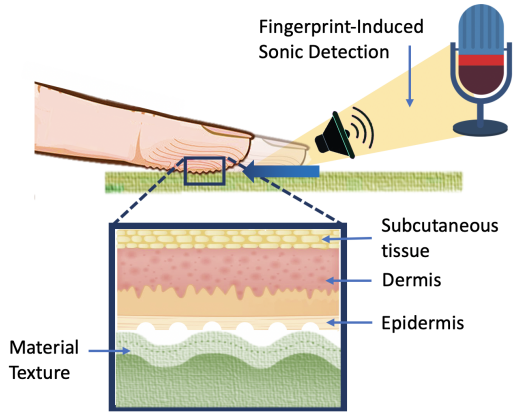


Fig. 2. FiSe arises from the friction between fingerprint and surface and can be sensed by a conventional microphone.

2.1 Fingerprint-Induced Sonic Effect

Friction develops from two surfaces sliding against one another irrespective of the intensity of their relative motion. This friction leads to distinct waves and oscillations within the interacting mediums resulting in the emission of sonic waves to the ambient environment [11]. In daily life, there are several instances of friction-excited sonic waves from an interaction between sneakers on the floor or chalk on the blackboard. In this paper, the context of sonic wave differs from the roughness noise, which is generally random (e.g., rubbing of two sandpapers). Under strong contact conditions, the sliding surfaces become a coupled system and generate an intricate and often nonlinear response. Previous studies have shown that physical parameters, including speed and pressure, only affect the magnitude of power spectral density to a certain extent, but not the overall distribution [12]. The roughness of the sliding surfaces impacts the sound pressure level (SPL) as:

$$\Delta SPL = 20\log_{10}(\frac{R_2}{R_1})^m, \tag{1}$$

where R_2 and R_1 correspond to the roughness of friction pair and m is an empirical factor varying based on the surface texture. The SPL of sonic waves can be similar between different friction pairs and thus impacts its sensing rather than uniqueness. A person with rough fingertip would produce a more audible sonic wave when rubbing a surface, in contrast to a soft skin fingertip. More importantly, for different friction pairs (e.g., finger against metal vs. finger against plastic), the uniqueness of sonic waves arise from the interface properties (i.e., texture) and the constitution of

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3049217, IEEE Transactions on Mobile Computing

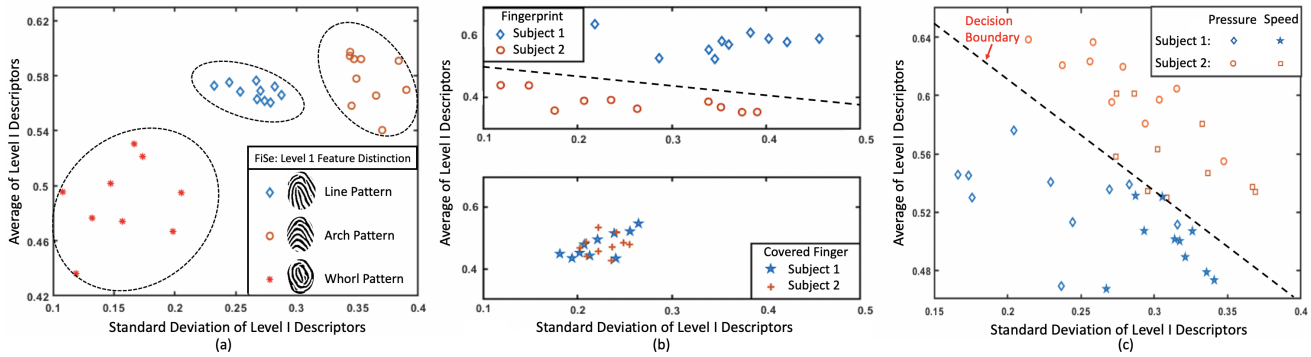


Fig. 3. A proof-of-concept (three subjects) for FiSe-based identification under the impact of (a) different fingerprint patterns; (b) fingerprint and covered finger interaction with surface; (c) human dynamics (i.e., swiping speed and pressure).

objects (e.g., weight distribution). The surface deformation during contact is highly minute [13] and its intensity is inconsequential to surface roughness.

Hypothesis: When a user swipes his fingertip on any surface (refer to Figure 2), the resulting friction-excited sonic wave depends on the intrinsic fingerprint patterns, underlying structure of finger and opposing material. Since every user has a unique fingerprint, the FiSe from two users swiping on the same surface should be different. Moreover, the low SPL of FiSe provides a strong resilience against spoofing attacks.

2.2 A Feasibility Study

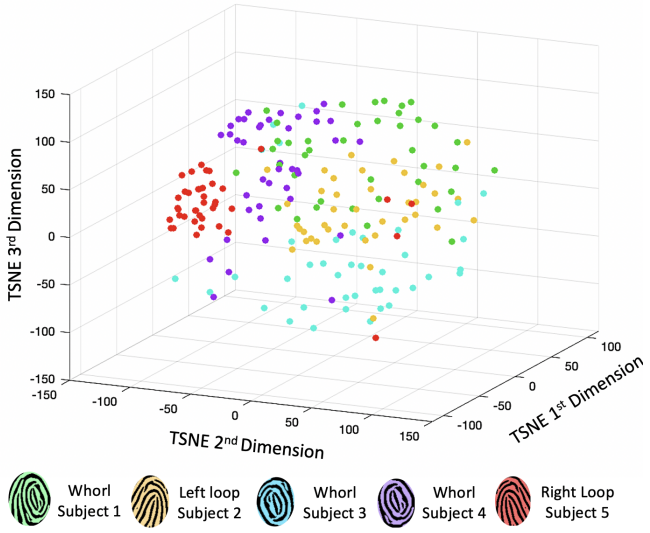


Fig. 4. A multidimensional representation of Level I friction descriptors from five unique fingerprints.

Proof-of-concept Setup: For studying the relationship between FiSe resulting from different fingerprints, we organize a preliminary study with 5 subjects between age group of 21∼30 years. Each subject is asked to use their right index finger for performing straight-downward swipes, 40 times each, on the back surface (aluminum) of a commodity smartphone. The subjects are told to swipe naturally without exerting intense pressure or speed, thereby controlling the bias from behavioral or soft characteristics. During the second trial, we cover the subject's fingertip with a scotch tape and repeat the swipe actions. In another experiment, we ask two subjects to repeat 15 swipes with gradually increasing pressure and speed in each trial. For the sake of isolating environmental dependency, this study is performed in a conference room (21°C) with low ambient noise. After processing the fingerprint-induced sonic waves, we aim to extract features that can provide a clue towards the inherent fingerprint.

Level I Friction Descriptors: Level I characteristics of the fingerprint depend on its macro details, i.e., the pattern and ridge flow and can be visually perceived through naked eye [14]. Similarly, in the audio domain, power-based temporal features highlight the changes in signal over time and perceptual features (e.g., pitch, loudness) have semantic meaning to a human listener. Therefore, we select eight features including harmonicity, pitch and spectral features (e.g., centroid, crest, decrease, entropy, flatness) as Level I friction descriptors. For ease of the comparison, Figure 3 illustrates the variations against average and standard deviation of descriptors after normalization. Each FiSe yields a data point on the graph and the points from multiple FiSe by the same fingerprint exhibit a cluster.

Multi-dimensional Analysis: For identifying relevant patterns in the high-dimensional features from FiSe, the Level I friction descriptors need to be strategically converted to a lower dimension space while preserving the distance between the samples. T-distributed Stochastic Neighbor Embedding (TSNE) [15] is a promising technique that can preserve the local structure, implying that the samples which are closer in the high-dimension would tend to be close even after dimensionality reduction. To do this, it converts the similarity between samples to joint probabilities and aims to minimize the Kullback-Leibler divergence between the joint probabilities of high-dimensional data and lower dimensional embedding. We set the initialization for embedding to be computed from Principal Component Analysis (PCA) to retain the global structure while considering the nearest neighbors, i.e., perplexity=30 [16], [17]. For a more detailed representation of Figure 3(a), Figure 4 illustrates the three-dimensional graph of descriptors.

Insights and Summary: The feasibility study reveals that (1) every user has a unique fingerprint pattern (e.g., loop, whorl pattern in Figure 3(a) and 4) which generates a unique FiSe during the swipe action; (2) Figure 3(b) proves that distinc-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3049217, IEEE Transactions on Mobile Computing

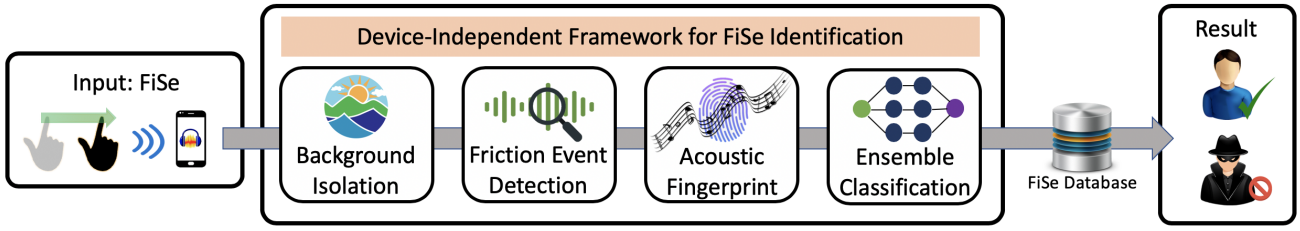


Fig. 5. The overview of *SonicPrint*, a fingerprint-biometric based user identification system.

tiveness of FiSe is dependent on the fingerprint rather than the overall geometry of the fingertip; (3) variation in pressure and speed has a limited effect on the identifiability of FiSe (see Figure 3(c)) However, only Level I descriptors are insufficient to differentiate multiple subjects in presence of contrasting behavioral traits (e.g., swiping speed, fingertip roughness) during the sensing process. To summarize, we prove that FiSe depends on the underlying fingerprint. For improving the accuracy, we continue to recruit appropriate features highlighting the intrinsic fingerprint information (Level II and Level III) from the sonic waves. The application of FiSe is discussed for smartphone security while evaluating different interacting surfaces in Section 9.1.

3 Threat Model

We consider a scenario where an innovate attacker, namely Alice, intends to steal intellectual property (IP) from the victim's smartphone. The smartphone is integrated with a singular defense system, i.e., *SonicPrint*. Unlike a traditional attacker who primarily focus on zero-informed attacks, Alice studies the fundamental operation of *SonicPrint* and even explores the past literature for proven methods to compromise the security of fingerprint and audio channel. Specifically, we consider the following attack scenarios:

• **Fingerprint phantom attack:** Typically, Alice can either exploit the social media of victim or leverage high-resolution cameras to remotely capture the target fingerprint. Afterward, the fingerprint and overall finger geometry can be utilized to create a fingerprint phantom (i.e., fake-finger). This fake-finger is highly identical to the victim's live-finger and can be used to spoof the system. It is worth mentioning that conventional fingerprint scanners can be compromised using this stealthy attack [18].

• **Replay and Side-Channel attack:** Without the victim's knowledge, Alice places a high-sensitive microphone near the smartphone and records the FiSe during an access attempt. This recording is replayed to the target device through direct FiSe matching or vibration injections by leveraging sophisticated hardware. Studies show that this attack can compromise the security of traditional voice authentications within five trials [19].

• **Denial-of-Service attack:** If Alice is unsuccessful even after launching aforementioned attacks, she aims to decrease the trust of victim towards the defense mechanism, ultimately leading the victim to either change or turn off the device security features. To do this, Alice can leverage additional speakers to project white noise towards the target while victim is performing swipe action. Instead of audible white noise, "hidden" ultrasound signals (f = 20KHz) can be utilized. This attack has been recently shown to compromise the security of speech recognition systems (e.g., Siri, Alexa) [20].

To this end, we make a few practical assumptions. Firstly, Alice cannot position the recording microphone in immediate proximity of victim's smartphone (i.e., $< 20cm$) considering the malicious device would be within line-of-sight of the victim, raising his suspicion. Secondly, Alice does not possess the advance manufacturing knowledge or economic capability to leverage organic 3D printers for developing biological replica of victim's finger.

4 SonicPrint System Overview

By analyzing the FiSe caused by fingertip and surface interaction, *SonicPrint* can reveal fingerprint dependent characteristics in the received signal. Figure 5 illustrates four primary modules of *SonicPrint*: (1) Background isolation; (2) Friction event detection; (3) Acoustic fingerprint analysis; (4) Ensemble classification. First, when a user swipes his fingertip on the smartphone surface, the inbuilt microphone is used to capture the FiSe. A series of pre-processing techniques including clutter suppression, target enhancement and ambient denoising are applied to acquire the precise sonic wave. Once its position is verified, a multi-level representation of acoustic fingerprint is obtained from specific features of the target signal. Finally, the representation is input to an ensemble classifier to precisely identify the legitimate user.

5 FiSe Processing Schemes

In this section, we discuss the nature of friction excited sonic waves from a coupled system consisting of fingertip and material. When a user swipes his fingertip on the smartphone surface, a FiSe is generated, which can be captured by the inbuilt microphone and can span the entire frequency band (0-22KHz).

5.1 Pre-processing

The sonic wave is typically submerged in the dynamic ambient noises (e.g., human talking, music) due to its low power. Considering the diverse and known frequency bands in the noise spectrum, it is effective to use high-order cutoff in one-pass filters. However, this also eliminates the intrinsic fingerprint information in the lower frequency bands. To remove the low frequency noise from human speech and music, we employ a high-pass filter with cutoff 2.2KHz to remove the arbitrary clutter and recover the signal with a frequency range from 2.2KHz to 22KHz.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3049217, IEEE Transactions on Mobile Computing

5.2 Sonic Effect Enhancement

The recent development in voice biometrics indicates that excessive background clutter, retained during preprocessing, makes it difficult to localize the phoneme [21]. Although the human voice and background clutter can be separated based on information content, FiSe might be perceived as generic noise due to its low power. The spectral subtraction [22] is a widely used method to enhance the target signal that is degraded by additive noise. However, it also introduces a distortion in the signal, referred as a musical note. A multi-band spectral subtraction technique was proposed as a countermeasure to deal with distortion [23]. Given that noise does not affect the entire frequency band of FiSe uniformly, we need to ideally subtract the appropriate noise spectrum from each frequency bin. This would restrict any excessive subtraction of intrinsic fingerprint information. We acquire the clean and enhanced spectrum of FiSe in the ith frequency band by:

$$|\hat{S}_i(k)|^2 = |Y_i(k)|^2 - \alpha_i \delta_i |\hat{D}_i(k)|^2 \quad b_i < k < e_i, \qquad (2)$$

where Y_i is the power spectrum of noisy FiSe signal, $\hat{D}_i$ is the noise estimate, b_i and e_i are starting and ending frequency bins. α_i is an over-subtraction factor and δ_i is empirically chosen for each frequency band. For calculating δ_i, we leverage a pre-recorded two second audio sample in daily environment with human voices as noise estimate. We update over-subtraction factor α_i as:

$$\alpha_i = c_1 \cdot log_{10}\left(\frac{\sum_{k=b_i}^{e_i} |Y_i(k)|^2}{\sum_{k=b_i}^{e_i} |\hat{D}_i(k)|^2}\right) + c_2, \qquad (3)$$

where c_1, c_2 are empirically chosen values. After nonlinear power spectrum subtraction, the enhanced FiSe is derived from its spectrogram. However, there still exists residual clutter between the intervals of FiSe.

5.3 Denoising-Aware Wavelet Reconstruction

In the past decade, wavelet-based noise removal has gained immense recognition due to two primary advantages: (1) it provides an optimal resolution of time-series signal in both the frequency and time domain; (2) it facilitates a precise multi-scale analysis [24]. Therefore, we employ wavelet denoising to eliminate the residual noise from the FiSe that remains even after sonic effect enhancement. Using maximal overlap discrete wavelet transform (MODWT) [25], the signal is first subjected to decomposition to acquire detail coefficients (α_k) and approximation coefficients (β_k):

$$\alpha_k^{(J)} = \sum_{n \in Z} x_n \overline{g}_{n-2^J k}^{(J)} \quad \beta_k^{(l)} = \sum_{n \in Z} x_n \overline{h}_{n-2^l k}^{(l)}, \qquad (4)$$

where the levels $J \in Z$ and $l \in \{1, 2, 3, ...J\}$. We choose the Daubechies 3 wavelet (dB3) and reduce the FiSe to 6 levels. Afterward, we apply the detail coefficient threshold for each level to discard the ambient clutter. Finally, a level-dependent reconstruction is employed using all the coefficients as:

$$x_n = \sum_{k \in Z} \alpha_k^{(J)} \overline{g}_{n-2^J k}^{(J)} + \sum_{l=1}^{J} \sum_{k \in Z} \beta_k^{(l)} \overline{h}_{n-2^l k}^{(l)}, \qquad (5)$$

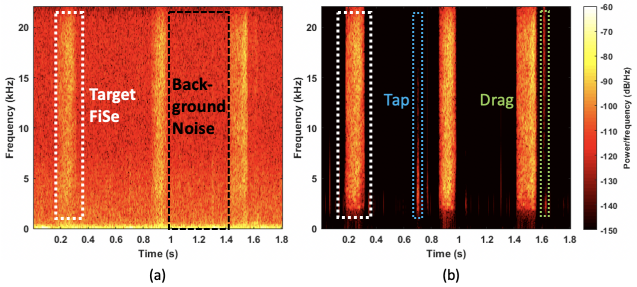


Fig. 6. The spectrogram of (a) original and (b) denoised FiSe from three swipe actions.

where $\overline{g}$ and $\overline{h}$ are rescaled discrete orthogonal functions. The spectrogram of FiSe before and after the processing stage is shown in Figure 6 with the signal-to-noise ratio (SNR) significantly improved from -3 to 23 decibels.

5.4 Friction Event Detection

Considering the FiSe is caused by a user swiping his fingertip on the smartphone surface, there are three challenges in tracing the target's precise location in the measured signal:

- The duration of FiSe would vary among intra-sessions (same user with different swipes) and inter-sessions (different users with different swipes) and typically lies between 0.05 and 0.3 seconds.
- The traditional segmentation approaches in speech recognition rely on threshold-based separation of speech vs. non-speech frames [26] – such methods are inadequate without optimization due to the fluctuations in sound pressure level from roughness or speed during the swipe action (see Section 2.1).
- The ideal signal would only comprise of the sonic wave. However, there may be an initial tap sound (i.e., finger colliding with the surface) or closing drag sound (i.e., finger slipping during lifting) enclosing the FiSe. Since the amplitude of the tap and drag sound are arbitrary, peak detection methods are ineffective.

To this end, we specially design our segmentation process (see Figure 7), to address the above challenges and isolate the starting and ending periods of each FiSe.

i) Adaptive Detection via HMM model: The hidden Markov Model (HMM) has proven to be an effective method for acoustic event detection [27]. It computes the probability of an occurrence of FiSe in every segment of the recorded signal and only consider those with high probability as friction events. Specifically, we first divide the recorded sample in non-overlapping frames, where each frame is 0.01 second period. A discrete fourier transform (DFT) is applied to each frame, after which an unbiased noise variance is calculated based on the optimally smoothed power spectral density estimate and spectral minima from each frequency band [28]. Finally, a widely used log-likelihood ratio test and HMM-based hang-over scheme [29] is used to determine the probability of friction event. To regulate the prior SNR [30] in log-likelihood, we define two additional parameters, i.e., TargetToSilence (TTS) probability and SilenceToTarget (STT) probability. For ensuring the identification of FiSe with even low audibility, we design an adaptive technique

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3049217, IEEE Transactions on Mobile Computing

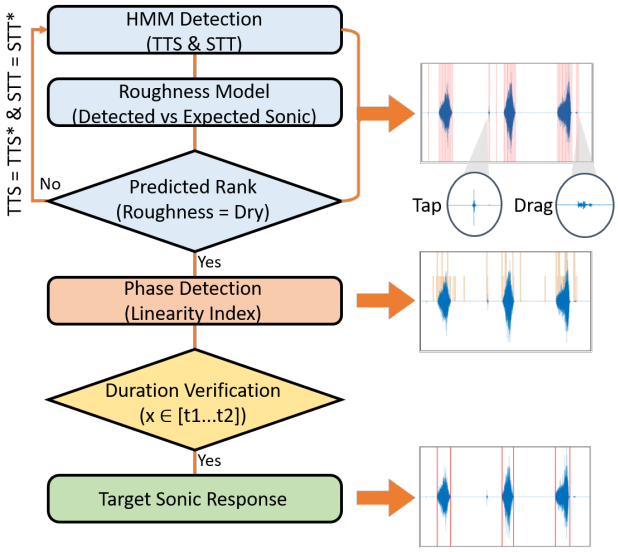


Fig. 7. Roughness-aware sonic detection.

that ranks the roughness of user's fingertip based on the statistical analysis of the signal. In particular, the roughness can be categorized as dry, balanced or soft by comparing the number of detected FiSe vs. the expected FiSe based on overall period. Depending on the predicted roughness, the TTS and STT probabilities are optimized to retrace the optimal friction events. In scenarios where the SPL of FiSe is very low, our adaptive detection can raise the number of identified events by more than 84% (counted manually).

ii) Phase-based Detection: The tap sound and drag sound are of arbitrary characteristics and challenging to remove by conventional statistical methods (e.g., maximum amplitude, mean, standard deviation). Previously, phase-based detection schemes have been proposed to suppress the impact noise [31]. The acoustic signal is first divided into non-overlapping frames of $0.01s$. Considering that there is only one dominant pulse of magnitude a at n_0 in the current frame, the signal $x(n) = 0$ except at $n = n_0$. Afterward, a DFT is applied to individual frames with the kth frequency bin and the phase slope as:

$$X(k) = |X(k)|e^{j\theta(k)} = ae^{-j2\pi kn/N}, \tag{6}$$

$$\Delta\theta(k) = \tan^{-1}\frac{Im(\overline{X}(k)\cdot\overline{X}^*(k-1))}{Re(\overline{X}(k)\cdot\overline{X}^*(k-1))}\quad \overline{X}(k) = \frac{X(k)}{|X(k)|}, \tag{7}$$

where $*$ represents the complex conjugate. Lastly, based on the phase slope and the n_0 position in current frame, a linearity index is defined as:

$$LI_\theta(k) = \Delta\theta(k) - \frac{-2\pi n_0}{N}. \tag{8}$$

The linearity index varies significantly between the FiSe and residual noise. However, its magnitude for tap/drag sound is similar to the FiSe, implying that they are of similar phase. Therefore, we employ the last processing step to select optimal FiSe events.

iii) Duration Verification: The sequence of occurrences with a high magnitude linearity index differs between the tap/drag sound and the FiSe. Based on the insights from HMM model and the linearity index, we conduct a final check by removing the segments whose duration does not lie from 0.05 to 0.3 seconds. Our notable contribution is that the aforementioned event detection is applicable for acquiring FiSe across different smart devices and surfaces (see Section 9.1) by making limited to no assumption with respect to the swiping behavior of users.

6 Taxonomy of Acoustic Fingerprint

The uniqueness of friction-excited sonic wave is dependent on the texture of contact surface, i.e., the fingerprint. The traditional fingerprint recognition relies on three-level vision-based characteristics [14]. As shown in Section 2.2, Level I friction descriptors are not sufficient since they can only relate to Level I optical fingerprint patterns. Therefore, we ask a question: *which features of FiSe can profoundly describe Level II and Level III fingerprint information?* To this end, we propose a novel taxonomy (see Figure 8) that bridges the gap between fingerprint and acoustics to select valid features for FiSe classification.

6.1 Level II Friction Descriptors

In the fingerprint domain, Level II features involve Galton characteristics, also known as minutiae points (e.g., hooks and bifurcations). These features possess a high variance between fingerprints of different users and are actively used in classification models. For the discrimination of audio sources, features such as the mel-frequency cepstral coefficients (MFCC) are essential since they can capture the timbral characteristics. Other cepstral features generally employ the perceptual filter bank and autoregression model to approximate the spectral envelope. Based on this semantic relationship, for the Level II friction descriptors, we select 14 MFCC (with Δ and $\Delta\Delta$), 12 linear prediction cepstral coefficients (LPCC) and 27 perceptual linear predictions (RASTA-PLP [32]). These descriptors can provide insights into the minutiae features of the fingerprint.

6.2 Level III Friction Descriptors

Although being unique, Level II fingerprint features are prone to spoofing since they could be visually perceived through the naked eye or even in low-resolution images. Thus, Level III fingerprint features are proposed based on the dimensional ridge information, including width, pores and edge contour. Similarly, short-time fourier transform and adaptive time-frequency decomposition can reveal various physical attributes of FiSe. These features have inferior meaning to human perception [33] and thus are difficult to spoof. To reveal the intrinsic fingerprint from FiSe, we select 12 linear prediction coefficients (LPC), 12 linear spectral frequencies (LSF), 26 log filter bank, spectral statistics (i.e., flux, kurtosis, skewness and slope) and 16 wavelet cross-level coefficients as Level III friction descriptors. The overall feature vector composed of ***162 friction descriptors*** is fed to our classification model.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3049217, IEEE Transactions on Mobile Computing

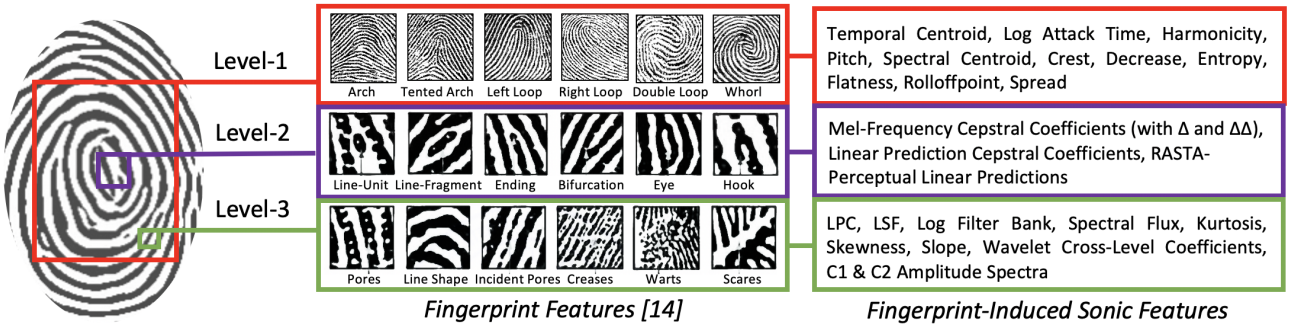


Fig. 8. A taxonomy of multi-level friction descriptors corresponding to intrinsic fingerprint.

6.3 SonicPrint Identification

Ensemble Classifiers: As the first exploratory study using FiSe for biometrics, we employ the following prediction models which have shown superior performance in user identification [34], [35], [36], [37]:

- Logistic Regression (LR): It models the outcome through logistic sigmoid function to deliver a probability measure which is further mapped to a specific class. We set the maximum iterations as 1000 and a cross-entropy loss for multi-class problem.
- Support Vector Machine (SVM): It is a statistical learning method that determines an optimal hyperplane to divide classes by maximizing the margin between closest points. The points lying on the boundary are referred to as support vectors. We choose a linear kernel.
- Random Forest (RF): It fits specific decision tree classifiers on the sub-samples and employs averaging to reduce overfitting. We set the estimators as 200 and use an entropy criterion for prediction.
- Linear Discriminant Analysis (LDA): By utilizing the Bayes' rule and approximating class conditional densities to samples, it creates a linear decision boundary to separate the classes. We select singular value decomposition as the solver.
- Gaussian Mixture Model (GMM): It provides a parametric probability distribution of audio signal and related features and characterizes the weighted sum of Gaussian components as a density function. We assume 5 components in our model.

From our empirical analysis, LDA is most suited for FiSe classification, followed by RF and SVM. Therefore, we assign a weight to each classifier (LR, SVM, RF, LDA, GMM) as 1, 2, 2, 3, 1, respectively. Finally, we perform hard voting on the observations generated from the classifiers to decide the legitimate user.

7 Evaluation Setup

7.1 Experimental Settings

We conduct a pilot study to validate the uniqueness of FiSe caused by the swipe motion on a smartphone. From reviewing the recent development in touch-based biometrics [38], we observe that two swipe actions are the most convenient and acceptable among users, as shown in Figure 9. **1Hand Swipe:** a user holds his phone naturally in right-hand and uses the index finger of the same hand to swipe on the surface. **2Hand Swipe:** left-hand firmly holds the phone while the other is used to perform the swipe. The 2Hand swipe is more robust to artifacts and allows for precise stroke capture. To provide a better understanding of the experimental process, we create a code to describe the performed swipe action. The code comprises of three parts, i.e., **Swipe-Sensing Distance-Surface**. The swipe could vary between 1Hand and 2Hand; sensing distance differs among $1cm$, $7cm$ or $11cm$ from inbuilt microphone; and surface could be aluminum, glass or others. Our experimental setup for the pilot study involves the participants to sit on a chair in a conference room with low ambient noise. The participants are asked to perform *1Hand-7cm-aluminum* swipes in a straight-downward direction on the back of the smartphone. Afterward, they are required to complete *2Hand-1cm-glass* swipes at the front of the smartphone. *To ensure that the obtained insights are applicable in real-world scenarios, physical attributes (i.e., speed, pressure or roughness) of the finger are not controlled during the swipe action, throughout the remainder of this paper.* We employ the Google Pixel 2 smartphone with a 0-22KHz range microphone to record the FiSe caused by the swipe action. It is $14.4cm(5.7inch)$ x $6.8cm(2.7inch)$ x $1.5cm(0.6inch)$ in size and weighs only $161.5g$, which is lightweight for easy use in daily life. It works on a Qualcomm Snapdragon 835 with an Octa-Core processor. The recorded signal is fed to *SonicPrint* for further analysis.

7.2 FiSe Collection and Partition

As the first exploration of utilizing FiSe for user identification, we recruit 31 subjects (25 males and 6 females) within the age-group of 18-50 years in our study. None

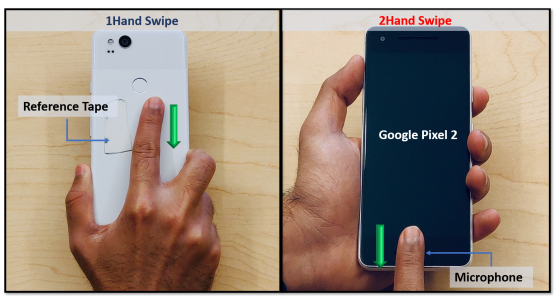


Fig. 9. The evaluation setup with subject performing 1Hand and 2Hand swipe on the smartphone surface with right index finger.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3049217, IEEE Transactions on Mobile Computing

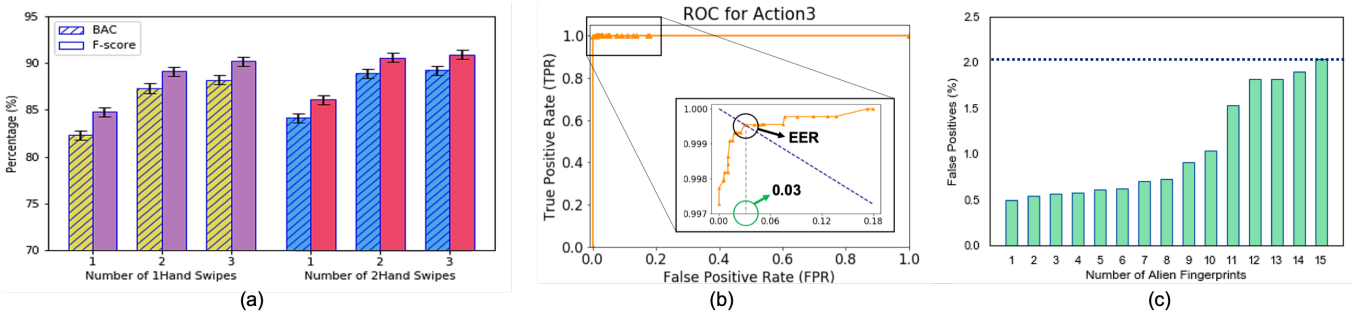


Fig. 10. The performance comparison between (a) *Action1*, *Action2*; (b) *Action3*; (c) *Action3* (unsupervised).

of the subjects have any damage to their fingerprint. For both the experiments involving 1Hand and 2Hand swipes, every subject performs six trials each. In each trial, the subject swipes at the specific position 30 times continuously. A $15min$ break separates every two consecutive trials to ensure non-uniform speed and pressure during swipes. Furthermore, the six trials for each experiment are spread across three weeks. A trial consists of $1min$ recording for each person. In total, every subject performs 180 *1Hand-7cm-aluminum* and 180 *2Hand-1cm-glass* swipe actions. The generated FiSe is recorded by the inbuilt microphone (sampling rate of 44.1KHz) and later fed to *SonicPrint*. After denoising and segmentation, a total of 4099 1Hand swipes ($\sim$130 per participant) and 4405 2Hand swipes ($\sim$140 per participant) are selected for training and testing. A 10-fold stratified cross-validation approach is applied to normalized features during user identification. The reason behind choosing stratified approach relates to the bias in classification models. During prediction, every instance is weighted equally, implying that a few over-represented classes can dominate the evaluation metrics. Thus, a stratified model ensures that each fold in cross-validation is representative of the whole dataset, thereby optimizing the bias and variance [39]. We employ other cross-validation and direct matching algorithms, in Section 9, to evaluate the inclusiveness of *SonicPrint* in real-world scenarios.

Evaluation Metrics: We introduce balanced accuracy (BAC), F-score, equal error rate (EER) and receiver operating characteristics (ROC) curve [40], [41] as metrics in our evaluation model. They are insensitive to class distribution which is critical for identification schemes. We also consider two additional metrics, i.e., Precision and Recall in Section 9 for robustness against unbalanced dataset.

7.3 SonicPrint Usability & Social Acceptance

SonicPrint requires the users to naturally swipe on their smartphone cover to acquire the unique FiSe. To assess the practicality and acceptance of *SonicPrint* in the real-world, we surveyed the 31 participants recruited in our pilot study. Of all the 31 participants, 80% are male and 20% are female. The participants are requested to answer multiple-choice questions belonging to the following two categories:

Smartphone Usability: We first ask the participants about the duration they operate smartphone in daily life. 41% of the participants spend 3 to 5 hours on their smartphone while 32% spend less than 2 hours. Within the spent time, 83% of the respondents primarily commit to communication (call or text) while 61% allocate the time on social media. On a per day basis, 54% of the participants unlocks their smartphone for more than 30 times. It is worth noting that among the unlock attempts by participants, around 60% are either performed multiple times or resorted to the password mechanism due to the insensitivity of fingerprint mechanism.

Security Awareness: We inquire the participants about their preferred biometric platform on the smartphone and their opinion on its security. 87% of the participants opt for fingerprint recognition while others evenly preferred the voice, face and password-based mechanism. From the total 31 participants, 64% think that fingerprint biometrics is not secure, 29% mentioned were unsure and 6% believed in its resilience against spoofing attacks. After informing them about the security risks, the perception of majority of participants shifted considerably towards taking cautionary steps to mitigate the threats. Only one participant still believed that the current state of fingerprint biometrics is reliable.

After completing the experiments, we ask the participants a few questions regarding their experience with our system. 71% of them preferred to perform 2Hand swipes on the front surface of the smartphone, while 29% preferred 1Hand swipes on the back cover. On a scale of 1 to 10, all the participants are requested to rate the comfortability while performing multiple swipe actions. We record an average score of 9.35, validating the ease-of-use of *SonicPrint*. Furthermore, we employ a 4-point Likert scale (ranging from Strongly Disagree to Strong Agree) [42]. This scale determines the participant's willingness to adopt *SonicPrint* in daily life for unlocking a smartphone or accessing protected information. 80% of the participants answered with a score of 4 points, while the rest gave a score of 3 points. These results show high acceptance of *SonicPrint* among subjects, especially when made aware of the threats in traditional fingerprint scanners.

8 Accuracy & Reliability Study

As a potential breakthrough technology, it is critical to evaluate the performance and reliability of *SonicPrint*. Our smartphone-based pilot study comprises user identification using FiSe obtained from two actions: (1) *Action1: 1Hand-7cm-aluminum* swipes; (2) *Action2: 2Hand-1cm-glass* swipes. For each action, we make a comparison of evaluation

metrics by increasing the number of swipes per sample performed by the user.

i) Action1 performance: After performing 10-fold stratified cross-validation on 4099 samples, the observed BAC and F-score are shown in Figure 8(a). The number of inputs, i.e., swipes per sample is increased from one to three and the variation in performance is recorded. The BAC for 1, 2 and 3 inputs is 82.3%, 87.3% and 88.2% while the F-score is 84.8%, 89.1% and 90.2% respectively. From ROC curve, the area-under-curve (AUC) is observed to be 85.3%, 89% and 88.6% as swipes per sample increases.

ii) Action2 performance: We report the BAC and F-score for 10-fold stratified cross-validation on 4405 samples in Figure 8(a). The BAC for 1, 2 and 3 inputs is 84.15%, 88.9% and 89.2% while the F-score is observed as 86.1%, 90.6% and 90.9% respectively. We compute AUC as 85.8%, 88.2% and 88.7% for increasing inputs. For *Action1* and *Action2*, the performance improves by augmenting more swipes per access attempt.

Performance Reliability: To ensure that the observed performance is not dependent on the size of training and testing dataset, we vary the number the splits in K-fold (from 3 to 10) and note the results. For both *Action1* and *Action2*, the BAC and F-score remain stable, within a margin of $\pm$2%, exhibiting the reliability of *SonicPrint* even under less amount of training samples.

Insights: While the previous results demonstrate the uniqueness of FiSe as a biometric trait, they also provide vital clues to improve *SonicPrint*. One reason for the lower performance of 1Hand (*Action1*) to 2Hand (*Action2*) swipes is due to its sensing distance from the microphone. A close proximity of swipe action with microphone ensures high SNR and allows for more precise capture of the FiSe. The 2Hand swipes provide a superior control to the users to ensure that their fingerprint properly interacts with the opposing surface. A rich textural material facilitates strong coupling between the fingerprint and surface to produce a more distinct FiSe. Since the glass material in *Action2* is a smooth surface, the performance can be enhanced by selecting a more suitable material to interact with the fingerprint.

iii) Action3 performance: Based on these insights, we conduct another experiment, *Action3*, to analyze the *SonicPrint* performance under ideal conditions. We place the smartphone in a common protective case made from synthetic leather and ask the 31 subjects to perform *2Hand-1cm-leather* swipes. We collect 4572 FiSe during swipe events and perform 10-fold stratified cross-validation. The BAC and F-score for one swipe per sample is 98.3% and 98.4%, respectively. Figure 8(b) shows the ROC curve where the observed EER and AUC are 0.03 and 97.5%. We examine the performance reliability by changing the splits in K-fold from 3 to 10, with results showing a $\pm$1% variation in scores.

Alien Fingerprint: To examine the vulnerability of *SonicPrint* against alien fingerprints (i.e., samples not trained in advance), we randomly choose 16 subjects and train the model using their *2Hand-1cm-leather* swipes. The remaining 15 subjects are used for testing in Figure 8(c). Our system can successfully reject the alien fingerprints using the threshold value of classification score. The results prove our insights and confirm that the users can be precisely recognized by *SonicPrint*.

Identification vs Recognition: A conventional fingerprint scanner in smart devices grant access to a user by matching his input to a pre-trained template. This task is similar to binary classification in authentication problems [43]. Our previous results show the capability of *SonicPrint* to perform a more challenging task of user identification (in other words, multi-class classification) which is desirable in the IoT environment (e.g., smarthome). Nevertheless, we also evaluated *Action1* and *Action2* performance for user recognition (i.e., each subject is compared against others, in a one-against-one fashion) to observe comparable evaluation metrics (+2%). Furthermore, we vary the number of randomly selected subjects from 2 to 30 and note the BAC score in Figure 11. As the number of subjects increase, the performance decreases. An interesting observation is that after 15 subjects, our model learns to effectively determine features that can accurately differentiate the subject-specific FiSe. A comprehensive evaluation of relative entropy in FiSe can be a lucrative venue for future work.

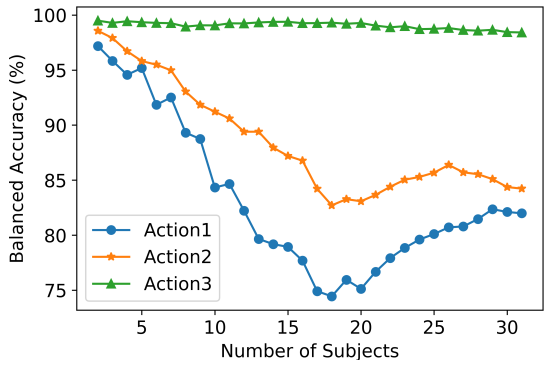


Fig. 11. The trend of balanced accuracy with increasing number of subjects.

9 Inclusiveness Study

To provide further insights about *SonicPrint* capability for user identification, we consider multiple scenarios that might contrast during real-world deployment. In the following, we recruit 5 subjects (vary between the experiments) to perform swipes actions using their right index finger. We evaluate the base performance by considering one swipe per sample through 10-fold cross validation.

9.1 Surface Exploration

We envision that *SonicPrint* can be integrated with not only the smart devices but also common materials or commodities found in the daily environment. To achieve this, it is vital to evaluate a wide variety of interacting surfaces and their impact on the uniqueness and SPL of FiSe. Each of the 5 subjects are asked to perform 150 swipe actions on 10 diverse materials, i.e., paper, foam, coarse leather, polycarbonate, silicon, engraved plastic, smooth plastic, rubber, fiber, smooth leather. The BAC, F-score, precision and recall are illustrated in Figure 12.

Insights: A high textural surface ensures a more robust coupling with the fingerprint during the swipe action. Furthermore, a smooth surface has lower roughness measure

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3049217, IEEE Transactions on Mobile Computing

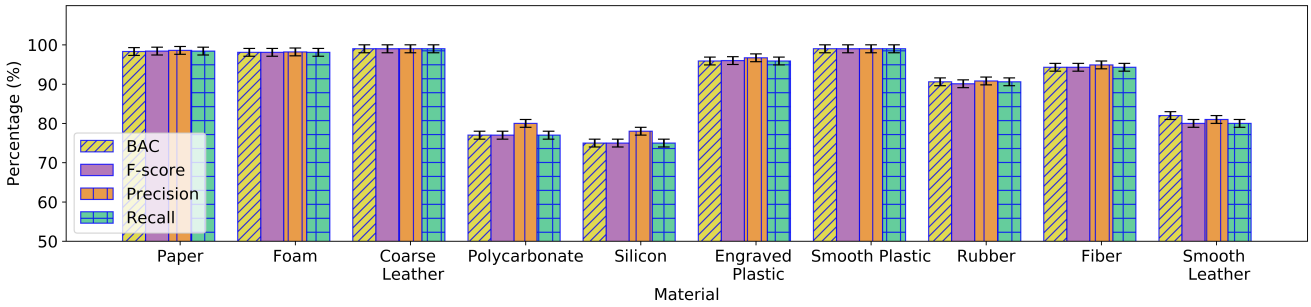


Fig. 12. SonicPrint performance on diverse and accessible surfaces found in smart devices or commodities.

leading to low SPL (see Section 2.1), raising the challenge for preprocessing module to differentiate between FiSe and generic noise. Either a high sensitive microphone or shorter sensing distance can elevate the system performance; FiSe is the first promising biometric trait which can be accessible and adoptable across wide range of materials with satisfactory texture.

9.2 Multi-Fingerprint Sensitivity

To achieve high acceptance among users, it is vital for *SonicPrint* to possess higher degree of freedom than traditional fingerprint biometrics. Furthermore, a user may prefer for *SonicPrint* to be capable in recognizing FiSe generated from multiple fingertips during a single access attempt. We ask the subjects to vary the number of fingers (from one to three) while performing 150 *2Hand-1cm-glass* swipe actions.

Insights: Our initial assumption states that cohesively using multiple fingers, with individual unique fingerprint, would increase the randomness of sonic waves. However, Table 1 demonstrates a unique way to increase the entropy of biometric trait. It is worth mentioning that no modifications were required for *SonicPrint* to facilitate this experiment; whereas, popular biometrics such as face or fingerprint would require either wider sensing region or advanced processing algorithms. Building on this multi-fingerprint approach, we aim to further evaluate its robustness against real-world challenges (e.g., fingertips with different moisture states or motion patterns) in our future work.

9.3 Skin Condition

Previous studies demonstrate that elderly users suffer from statistically lower finger friction coefficient, moisture and elasticity [44] as compared to younger age group. Equation 1 states the relation between the roughness of surface to the SPL of sonic wave, making it crucial to confirm that FiSe from diverse age groups, with different skin conditions,

TABLE 1
SonicPrint adaptability to multi-fingerprint swipe actions.

	One Finger	*Two Finger*	*Three Finger*
BAC	87.5%	93%	97.6%
F-score	86.9%	93%	97.6%
Precision	86.7%	94%	97.7%
Recall	87.5%	93%	97.6%

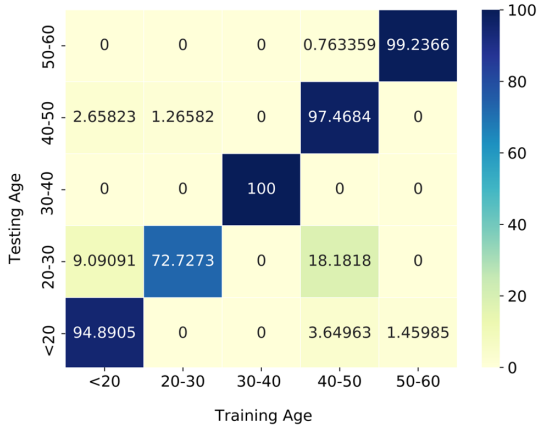


Fig. 13. Evaluation among age groups (years). The subject in each group possess different skin condition in terms of roughness, elasticity.

can be recognized by our system. We randomly choose 5 subjects from different age group (i.e., 18∼60 years) to perform 150 *2Hand-1cm-glass* swipes.

Insights: Considering elderly users have dry fingertips, their FiSe recordings comprised of high SPL making it easier for *SonicPrint* to trace the sonic wave in overall measured signal. This is evident from the stable performance observed among different age groups in Figure 13. The lower performance for age group of 20-30 years is due to the subject using lotion on their fingertip prior to the experiment (thereby leading to highly smooth fingertip). We further discuss the potential improvements in Section 12.

9.4 Swipe Stability

In a real-world setup, it is unlikely that the swipe action performed by user is regulated and monitored as in our pilot study. It would be ideal if FiSe is sufficiently resilient to human artifacts. To this end, the subjects perform 200 *1Hand-7cm-aluminum* and 200 *2Hand-1cm-glass* on the smartphone. During the later 100 swipes in each experiment, the subjects are periodically pushed on their back body at random intervals. The intensity of these artifacts are controlled to prevent huge disruption in the entire body (*e.g.*, pushing with both hands forcefully) but are sufficient to influence the upper body posture of the subject. The results are shown in Figure 14.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3049217, IEEE Transactions on Mobile Computing

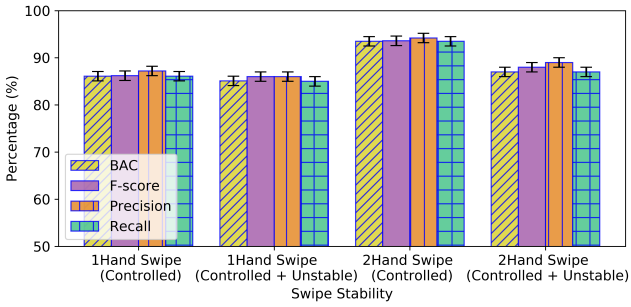


Fig. 14. Evaluation under controlled and unstable swipe actions.

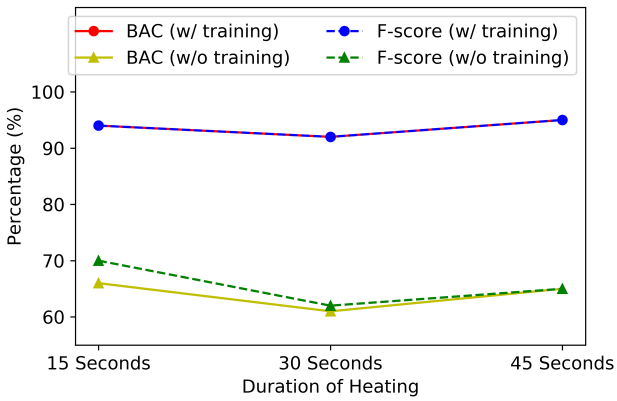


Fig. 15. Evaluation under different duration of heating.

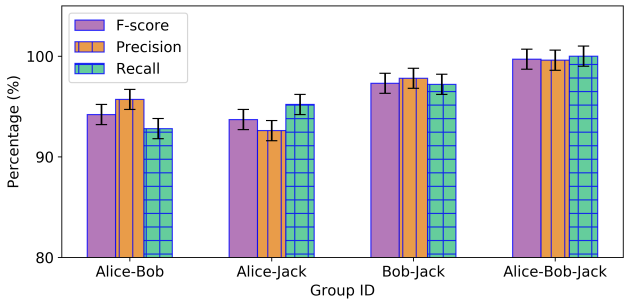


Fig. 16. *SonicPrint* performance to simultaneously identify multiple users from their integrated sonic wave.

Insights: Human artifacts have limited influence on the performance of *SonicPrint*. Intuitively, the *2Hand* swipes should have minimal impact since holding the smartphone with one-hand and using the opposing finger to swipe ensures a more continuous flow. However, the influence of artifacts on *2Hand* is more severe when compared to *1Hand* due to the variations in magnitude of artifacts within experiment. Nevertheless, these findings are valuable as we can envision similar results for users with movement disorders (e.g., parkinson).

9.5 Device Temperature

In traditional biometrics (e.g., fingerprint, face, voice), device temperature is rarely considered as a factor of evaluation. Yet, recent studies demonstrate the adverse influence of temperature on embedded sensors (e.g., stability of cameras [45]). *SonicPrint* relies on conventional microphones in smart devices to sense the FiSe which may be influenced from temperature. For evaluation, the subjects are required to individually conduct 600 *2Hand-1cm-glass* swipes. After every 150 swipes, we increase the temperature of smartphone by using an off-the-shelf hot-air blower for 15, 30 and 45 seconds duration.

Insights: Figure 15 shows that high temperature has an adverse effect on the sensitivity of in-built microphone, leading to decrease in system performance. These results matches with the known fact that MEMS microphone experience a loss of sensitivity and frequency response while suffering from distortion above the operating limit [46], [47]. Nevertheless, the FiSe signals can still be recognized if *SonicPrint* is sufficiently trained to tackle adverse conditions without requiring modifications in the sensing hardware. A more comprehensive study on the effect of temperature on material surface and FiSe is retained for future work.

10 Case Study

10.1 Group Authentication

Over the last decade, biometric technologies has transformed the user security by analyzing diverse physiological and behavioral traits via unique frameworks, e.g., multimodel, unobtrusive and continuous authentication [48]. Yet, one problem remains to be addressed: conventional biometrics provides a one-to-one connection between the measured signal and user's identity. For instance, if users belonging to a group (e.g., family, colleagues) needs be authenticated at a single timestep (e.g., border verification in airports), multiple sensors are required with increased resolution and field-of-view. Moreover, the software algorithms need to individually assess each biometric trait making the computational time complexity similar between identifying the group together vs. each person separately. Considering the promising results shown by *SonicPrint* using multi-fingerprint approach (see Section 9.2), it can lead to a breakthrough if FiSe from different groups of users can be identified without any change in system architecture.

To this end, we recruit 3 subjects (namely, Alice, Bob, Jack) and organize them into four groups (Alice-Bob, Bob-Jack, Alice-Jack and Alice-Bob-Jack). Subjects in the same group are requested to sit next to each other and place their right index finger on a common blank paper. The smartphone measures the FiSe resulting from each group while they concurrently perform 150 *2Hand-2cm* swipes. By using a visual cue (i.e., pointer traversing across the smartphone screen at fixed speed), the swipes of users are controlled to have consistent start and end time. The results of identifying a group in comparison to others are illustrated in Figure 16. The average BAC is 96.3%. *SonicPrint* can not only perform accurate group authentication but is also robust to the number of users in a group.

10.2 Object Identification

The uniqueness of FiSe relates to the fingerprint minutiae, surface texture and the underlying composition of human

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3049217, IEEE Transactions on Mobile Computing

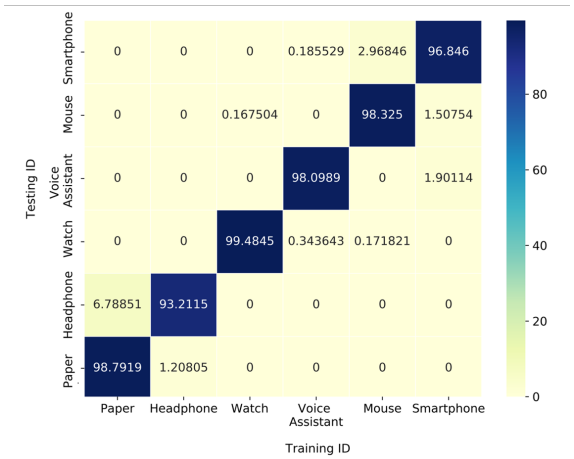


Fig. 17. *SonicPrint* performance to identify interacting object based on swipe actions.

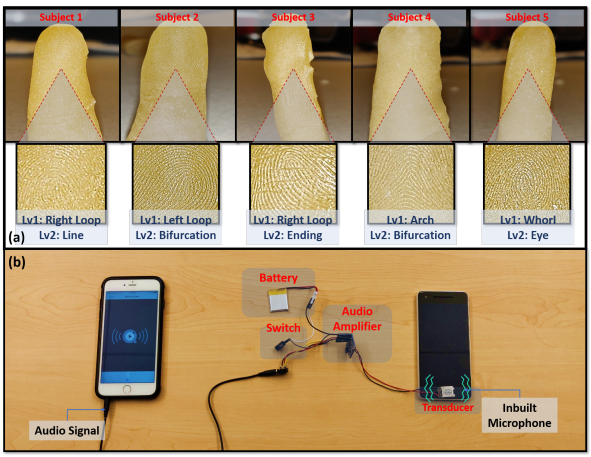


Fig. 18. (a) Gelatin fake fingers with multi-level fingerprint textures; (b) vibration injections via audio transducer.

fingertip. Its dependency on surface texture raises an interesting question whether *SonicPrint* can be applied for object identification. Recently, object tagging without Near Field Communication (NFC) tags have received immense attention for robotics [49] and mobile applications [50].

Building on this, we ask subjects to perform swipe actions on six different types of objects (i.e., paper, Bose headphone, Apple Watch Series 4, Google voice assistant (Echo), Logitech mouse and Google Pixel 2 smartphone). An overall of 3109 sonic waves are processed and analyzed for this experiment. For classification, instead of assigning unique class label to each subject, swipes performed on each object would have same class label irrespective of source fingerprint. Figure 17 demonstrates a high performance with precision, recall, F-score and BAC of 97.9%, 97.1%, 97.6% and 97.4% respectively. We envision that *SonicPrint* ability to sense the nature and type of object that users are touching can have revolutionary impact on accessibility services.

11 Vulnerability Study

In this section, we examine the security of *SonicPrint* against the sophisticated attacks that are known to compromise the security of traditional fingerprint scanners and voice recognition systems.

11.1 Fingerprint Phantom Attack

We assume that Alice has access to the fingerprint and other geometrical characteristics (e.g., width, thickness) of left index finger of a legitimate user. Based on this information, she aims to build a replica of the victim's finger and breach the biometric security. There are two methods to achieve this goal. First, she can utilize an advanced 3D printer to replicate the precise texture patterns of a fingerprint on a finger model. Yet, these printers are economically infeasible and often inaccessible to general public. Furthermore, considering the extensive detail of fingerprint, simulating the minutiae characteristics is computationally expensive, where the complexity increases exponentially with the level of features. Second, Alice can utilize materials commonly found in the household to build a fake-finger. Considering these materials have shown sufficient capability to spoof the fingerprint scanners [51], this is the most plausible attack scenario. Alice utilizes gelatin which can most closely relate the texture of live finger [2] and can even spoof capacitive fingerprint scanners [1]. We recruit 5 subjects with fingers of different sizes and perform the following steps:

- We ensure that the entire finger of each subject is covered by multiple layers (5 to 8) of latex material.
- Between each successive layer, we wait for 10 minutes to lose the moisture; the finger is kept still so that no pressure marks or creases occur on the coating.
- Once the latex coating becomes firm, we gently enclose it with baking powder as we remove the latex from the finger. The baking powder do not harm the target fingerprint since it is placed on the outside while the fingerprint features are on the inside of the coating.
- We prepare a mixture of one part gelatin, glycerin and water and use a conventional microwave to heat the mixture. Finally, we pour the mixture inside the recovered latex coating and leave it to dry for 24 hours. The latex coating is then discarded to obtain the gelatin fake-finger, as illustrated in Figure 18.

We ask each subject to use their live left index finger and perform 100 *2Hand-7cm-aluminum* and 100 *2Hand-1cm-glass* swipes on the smartphone. Afterward, we repeat the process by informing subjects to utilize their fake-fingers to complete swipe actions. We train the *SonicPrint* on recordings from live fingers and test fake-fingers during identification. For the fake-finger recordings, we observe that our preprocessing module discards 300 (out of 500) aluminum and 450 (out of 500) glass FiSe. Out of the remaining, only 32 (6.4%) aluminum and 21 (4.2%) glass FiSe are misclassified as live fingers. These results provides a promising start regarding the sensitivity of our background isolation module to identify the live sonic wave and the resilience of *SonicPrint* against fake-fingers.

11.2 Replay and Side-Channel Attack

We assume that Alice knows the underlying mechanism of *SonicPrint* to sense the sonic waves for user identification.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3049217, IEEE Transactions on Mobile Computing

Through a high-resolution camera, Alice can acquire the victim's fingerprint from a distance of $2m$ [52]; however, no FiSe can be obtained from a similar distance due to its low SPL. Therefore, we envision an unrealistic scenario where she leverages a high-sensitive microphone (i.e., Fifine-K670) and positions it at very close proximity of $20cm$ and $30cm$ facing the target smart device. The microphone captures the FiSe during an access attempt by a legitimate user.

Attack via Microphone: the recording is replayed to the inbuilt microphone of target smartphone by direct FiSe replay. Overall, 4 subjects conduct 500 *2Hand-7cm-aluminum* swipes on Google Pixel 2 and the inbuilt and secondary microphone concurrently records the FiSe. For attack through a direct transfer, merely 4.8% and 3.2% of replayed FiSe match with the original recording, even at a close distance of $20cm$ and $30cm$ respectively. During the sensing phase, all microphones pick up two sounds, "on-axis" from the direction they are designed to pick up and "off-axis" from all other directions which cannot be modelled and follows behavior of microphone. This can have different effects such as change in frequency response, relative volume or character of sound, especially in dynamic environments [53]. The original FiSe is of low sound pressure level and our preprocessing module is designed to be extensive during filtering (wavelet denoising) and selecting friction events (Hidden-Markov Model) from overall signal. During the experiments, we observed majority of replay samples to not pass through the preprocessing module of our system.

Attack via Vibration Channel: we consider a scenario where Alice attempts to forge the swipe action of legitimate user as vibration signals for identification. When the previously recorded audio signal is passed through the coil of transducer, a dynamic electromagnetic field is generated that makes the actuator vibrate the smartphone (see Figure 18). The intensity of these vibrations are controlled through an amplifier to drive its amplitude closer to that of sonic wave. Although these vibrations are propagated from a very close distance (i.e., top of smartphone), all are rejected by *SonicPrint*, making side-channels attacks via hidden transmitters ineffective.

11.3 Hidden Denial-of-Service Attack

Upon realizing the unsuccessful attempts to compromise *SonicPrint* via fake-finger, replay and side-channel attack, Alice aims to manipulate victim's trust in the biometric system by leveraging inaudible noise (i.e., audio signal with frequency of 20KHz). Specifically, while victim is accessing his smart device through swipe action, Alice would project inaudible noise towards the in-built microphone. She envisions that FiSe would suffer from same deterioration in information content as human voice under the influence of inaudible noise [54]. To evaluate this, we ask 5 subjects to perform 150 *2Hand-1cm-glass* swipes each on Google pixel 2. During the sensing process, we project a tone with 20KHz frequency of highest volume supported by the speakers in IPhone 6S smartphone towards the microphone. The recorded FiSe is fed to the background isolation module of *SonicPrint* for further processing. From Figure 19, the noise at 20KHz can be clearly observed in the measured signal. However, the MODWT wavelet denoising algorithm in the

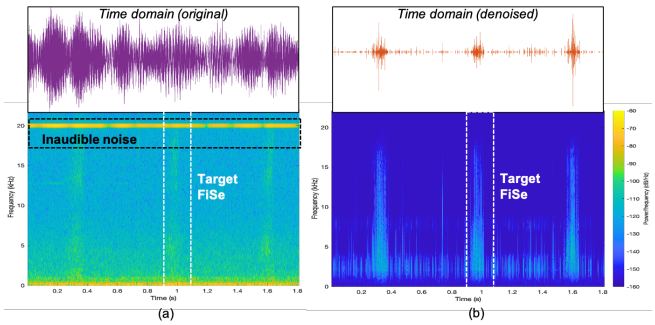


Fig. 19. The time domain representation and spectrogram of (a) original and (b) denoised FiSe from three swipe actions.

background isolation module is robust to inaudible noise and provides a more finer and regularized discretization of the signal [55]. The traces of noise of lower frequency between concurrent FiSe can be further discarded via the friction event detection algorithm (see Section 5.4). While the observed results are favorable, this experiment can be improved by considering higher frequency noise (i.e., above 25/40/60KHz) thus serving as an valuable exploration opportunity for mobile security research.

12 Discussion

Microphone Sensitivity: *SonicPrint* leverages the low-cost microphone of smartphone for FiSe acquisition. Although our system shows a satisfactory performance under ideal conditions, the overall results can be significantly improved by adopting high sensitive microphones. These microphones can precisely detect FiSe from even swipe actions on smooth surfaces in a noisy environment. Users would not be required to perform the swipe as close to the microphone, increasing the level of freedom and user acceptance.

Accuracy and Improvements: *SonicPrint* achieves 84% and 98% identification rates with a single trial on standard and high-texture smartphone surface, respectively. This is comparable to recent low-cost solutions using vibrations [56], [57], gait patterns [58] and passive sensing [59] for authentication. Yet, the most significant contribution of *SonicPrint* is its adoptability across diverse surface materials (refer to Section 9.1) which is not supported by existing solutions. Our proposed approach can also be used as secondary biometrics; improvements in microphone frequency response and deep learning approaches can be considered for our future exploration.

System Considerations: As a starting point, *SonicPrint* is a promising biometric with high adoptability and anti-spoofing capabilities. However, a practical deployment in the real-world requires reflection on following criteria: (1) *Privacy:* The audible nature of FiSe makes it prone to theft via a conventional recording device. For a countermeasure, the user can be asked to perform a specialized gesture (e.g., zig-zag or star pattern) during the training process. These gestures are uncommon in normal user behavior, thereby increasing the difficulty for an attacker to acquire the target FiSe outside the recognition period. (2) *Power consumption:* The power consumption primarily depends on the sensing and processing algorithms used for *SonicPrint*

implementation. Both of the sensing and processing are light and can be managed by a digital signal processor (DSP). For instance, a 5.8mW 48-kHz recording supported by TLV320AIC3212 AudioCodec [60] and a TMS320C553x DSP [61] with 64KB to 320KB memory and 0.15 mW/mHz active power at 1.05 V and 0.15 mW standby power are sufficient for standard audio filtering. In case a CPU is employed, since each authentication takes less than 2 seconds, the power consumption is limited and can be further decreased by using co-processors [62]. The use of statistical classifiers would limit the memory consumption compared to Tensorflow Lite (0.85W on EdgeTPU [63]). Given that our solution is locally-hosted and do not require heavy computational resources, it can be more energy-efficient than voice assistants such as Alexa, Siri which consumes less than 2W [64] on standby despite actively listening. In case DSP has limited memory for continuous listening, a touch trigger [65] can be employed to activate FiSe recording, thereby limiting battery usage in smart devices. (3) *Recognition time:* By employing computationally inexpensive algorithms, *SonicPrint* can identify a user within 2 second period, further facilitating its deployment in smart devices.

User's Perspective: In a real-world application, *SonicPrint*, at its current capability, would require users to swipe up to 60 times during the training phase (1 minute duration) and 1-3 swipes during the login attempt. There are two considerations: (1) The widely-used biometrics (e.g., fingerprint, face) also require users to follow special instructions during the training process, i.e., input biometric trait from multiple orientations and locations which can consume more than a few minutes for non-technical audience. (2) Instead of performing multiple swipes during login attempt, users can perform a single swipe with multiple fingers having higher precision during identification (as shown in Section 9.2).

13 Related Work

Touch-based Biometrics: Touch-based implicit authentication relies on the unconstrained movement patterns of users when they interact with their smartphone. The location of finger taps could be inferred from the motion sensors [66], [67]. Based on this insight, the touch dynamics was explored as a soft biometric trait for user authentication [68], [69]. Different parameters such as the rhythm, strength, angle of applied force [70] or the size and axis length of finger touch area [71] can depict the user's individuality. Despite the enhancements in security [72], [73], [74], it was shown that mimicry attacks have a bypass rate of 86%, even with partial knowledge of the underlying features of touch biometric [75]. Recently, researchers have employed induced body electric potentials (iBEP) or body guided communications as a new biometric [76], [77]. However, it requires the user to continuously wear a token device and can be spoofed through injection attacks. Our method relies on the uniqueness of fingerprint and cannot be spoofed via mimicry or side-channel attacks.

Acoustic Sensing: In 2011, researchers proposed that the acoustic signatures caused by an object impacting with a screen surface could identify its type (i.e., fingernail, knuckle, tip) [78]. Afterward, the domain of acoustics-based touch interaction was enhanced by monitoring continuous sound via structure-borne sound propagation [79] for inferring the finger tapping and movements of the user [80]. When a vibration motor excites a surface, the presence of devices [81] or user-specific gestures [56] can be sensed by the inertial sensors. However, these approaches have limited accessibility due to the requirement of additional vibration transmitters and receivers and more importantly, are vulnerable to the Denial-of-Service (DoS) attacks. The latest advancement in the field of photoacoustics [82], [83] provides multidimensional insight to human palm while researchers have shown to utilize wireless signals for extracting precise audio signals for authentication [84], [85]. Yet, these systems cannot support adoptability in smart devices. A recent study captures the finger sound caused by thumb rubbing the finger for gesture recognition [86], yet requires the user to wear a ring during the sensing process. To the best of our knowledge, we provide the first study on exploring the intrinsic fingerprint information in friction-excited sonic waves for secure user identification.

14 Conclusion

Existing fingerprint biometric is vulnerable to spoofing attacks (e.g., fake-fingers) and cannot be adopted in upcoming smart devices due to hardware constraints. In this paper, we introduce a new dimension of fingerprint sensing using the friction-excited sonic wave caused by a fingerprint to surface interaction. We develop *SonicPrint* that utilizes the FiSe from a user swiping his fingertip on everyday smart devices for identification. The system is adoptable, user-friendly and difficult to counterfeit with an identification accuracy up to 98%. We also show the inclusiveness of *SonicPrint* under human artifacts, skin conditions, multi-fingerprint and device temperature. Furthermore, SonicPrint shows immense potential for applications in group authentication and object identification. In the future, we aim to consider users having damaged fingerprints while exploring high-sensitive microphones with ultrasonic range to improve the system accuracy.

Acknowledgement

This work is in part supported by the National Science Foundation under grant No. 1718375 and 2028872.

References

[1] H. Kang, B. Lee, H. Kim, D. Shin, and J. Kim, "A study on performance evaluation of the liveness detection for various fingerprint sensor modules," in *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*. Springer, 2003, pp. 1245–1253.

[2] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino, "Impact of artificial" gummy" fingers on fingerprint systems," in *Optical Security and Counterfeit Deterrence Techniques IV*, vol. 4677. International Society for Optics and Photonics, 2002, pp. 275–289.

[3] S. S. Arora, K. Cao, A. K. Jain, and N. G. Paulter, "3d fingerprint phantoms," in *2014 22nd International Conference on Pattern Recognition*. IEEE, 2014, pp. 684–689.

[4] A. Ross and A. Jain, "Biometric sensor interoperability: A case study in fingerprints," in *International Workshop on Biometric Authentication*. Springer, 2004, pp. 134–145.

[5] D. Winder, "Samsung galaxy s10 fingerprint scanner hacked - here's what you need to know," Apr 2019. [Online]. Available: https://www.forbes.com/sites/daveywinder/2019/04/06/samsung-galaxy-s10-fingerprint-scanner-hacked-heres-what-you-need-to-know/10c88305d423

[6] P. D. Lapsley, J. A. Lee, D. F. Pare Jr, and N. Hoffman, "Anti-fraud biometric scanner that accurately detects blood flow," Apr. 7 1998, uS Patent 5,737,439.

[7] R. Derakhshani, S. A. Schuckers, L. A. Hornak, and L. O'Gorman, "Determination of vitality from a non-invasive biomedical measurement for use in fingerprint scanners," *Pattern recognition*, vol. 36, no. 2, pp. 383–396, 2003.

[8] T. Kinnunen, M. Sahidullah, H. Delgado, M. Todisco, N. Evans, J. Yamagishi, and K. A. Lee, "The asvspoof 2017 challenge: Assessing the limits of replay spoofing attack detection," *INTERSPEECH*, 2017.

[9] L. Zhang, S. Tan, and J. Yang, "Hearing your voice is not enough: An articulatory gesture based liveness detection for voice authentication," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2017, pp. 57–71.

[10] N. Erdogmus and S. Marcel, "Spoofing 2d face recognition systems with 3d masks," in *2013 International Conference of the BIOSIG Special Interest Group (BIOSIG)*. IEEE, 2013, pp. 1–8.

[11] A. Akay, "Acoustics of friction," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1525–1548, 2002.

[12] B. L. Stoimenov, S. Maruyama, K. Adachi, and K. Kato, "The roughness effect on the frequency of frictional sound," *Tribology international*, vol. 40, no. 4, pp. 659–664, 2007.

[13] H. B. Abdelounis, A. Le Bot, J. Perret-Liaudet, and H. Zahouani, "An experimental study on roughness noise of dry rough flat surfaces," *Wear*, vol. 268, no. 1-2, pp. 335–345, 2010.

[14] A. K. Jain, Y. Chen, and M. Demirkus, "Pores and ridges: High-resolution fingerprint matching using level 3 features," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 1, pp. 15–27, 2006.

[15] L. v. d. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.

[16] M. Wattenberg, F. Viégas, and I. Johnson, "How to use t-sne effectively," *Distill*, vol. 1, no. 10, p. e2, 2016.

[17] L. Van Der Maaten, "Accelerating t-sne using tree-based algorithms," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 3221–3245, 2014.

[18] C. Barral and A. Tria, "Fake fingers in fingerprint recognition: Glycerin supersedes gelatin," in *Formal to Practical Security*. Springer, 2009, pp. 57–69.

[19] H. Feng, K. Fawaz, and K. G. Shin, "Continuous authentication for voice assistants," in *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*. ACM, 2017, pp. 343–355.

[20] G. Zhang, C. Yan, X. Ji, T. Zhang, T. Zhang, and W. Xu, "Dolphinattack: Inaudible voice commands," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 103–117.

[21] L. Zhang, S. Tan, J. Yang, and Y. Chen, "Voicelive: A phoneme localization based liveness detection for voice authentication on smartphones," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2016, pp. 1080–1091.

[22] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 27, no. 2, pp. 113–120, 1979.

[23] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise." in *ICASSP*, vol. 4. Citeseer, 2002, pp. 44 164–44 164.

[24] H. Abdelnasser, M. Youssef, and K. A. Harras, "Wigest: A ubiquitous wifi-based gesture recognition system," in *2015 IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2015, pp. 1472–1480.

[25] D. B. Percival and A. T. Walden, *Wavelet methods for time series analysis*. Cambridge university press, 2006, vol. 4.

[26] Q. Li, J. Zheng, A. Tsai, and Q. Zhou, "Robust endpoint detection and energy normalization for real-time speech and speaker recognition," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 3, pp. 146–157, 2002.

[27] Y. Bi, M. Lv, C. Song, W. Xu, N. Guan, and W. Yi, "Autodietary: A wearable acoustic sensor system for food intake recognition in daily life," *IEEE Sensors Journal*, vol. 16, no. 3, pp. 806–816, 2015.

[28] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Transactions on speech and audio processing*, vol. 9, no. 5, pp. 504–512, 2001.

[29] J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," *IEEE signal processing letters*, vol. 6, no. 1, pp. 1–3, 1999.

[30] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 32, no. 6, pp. 1109–1121, 1984.

[31] A. Sugiyama, R. Miyahara, and K. Park, "Impact-noise suppression with phase-based detection," in *21st European Signal Processing Conference (EUSIPCO 2013)*. IEEE, 2013, pp. 1–5.

[32] H. Hermansky and N. Morgan, "Rasta processing of speech," *IEEE transactions on speech and audio processing*, vol. 2, no. 4, pp. 578–589, 1994.

[33] D. Mitrović, M. Zeppelzauer, and C. Breiteneder, "Features for content-based audio retrieval," in *Advances in computers*. Elsevier, 2010, vol. 78, pp. 71–150.

[34] C. Bo, L. Zhang, X.-Y. Li, Q. Huang, and Y. Wang, "Silentsense: silent user identification via touch and movement behavioral biometrics," in *Proceedings of the 19th annual international conference on Mobile computing & networking*. ACM, 2013, pp. 187–190.

[35] J. Angulo and E. Wästlund, "Exploring touch-screen biometrics for user identification on smart phones," in *IFIP PrimeLife International Summer School on Privacy and Identity Management for Life*. Springer, 2011, pp. 130–143.

[36] S. Kwon and S. Narayanan, "Robust speaker identification based on selective use of feature vectors," *Pattern Recognition Letters*, vol. 28, no. 1, pp. 85–89, 2007.

[37] S. Luo, A. Nguyen, C. Song, F. Lin, W. Xu, and Z. Yan, "Oculock: Exploring human visual system for authentication in virtual reality head-mounted display."

[38] Z. Ali, J. Payton, and V. Sritapan, "At your fingertips: Considering finger distinctness in continuous touch-based authentication for mobile devices," in *2016 IEEE Security and Privacy Workshops (SPW)*. IEEE, 2016, pp. 272–275.

[39] R. Kohavi *et al.*, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Ijcai*, vol. 14, no. 2. Montreal, Canada, 1995, pp. 1137–1145.

[40] 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 *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, 2017, pp. 315–328.

[41] F. Lin, K. W. Cho, C. Song, W. Xu, and Z. Jin, "Brain password: A secure and truly cancelable brain biometrics for smart headwear," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, 2018, pp. 296–309.

[42] S.-O. Leung, "A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point likert scales," *Journal of Social Service Research*, vol. 37, no. 4, pp. 412–421, 2011.

[43] N. K. Ratha, R. M. Bolle, V. D. Pandit, and V. Vaish, "Robust fingerprint authentication using local structural similarity," in *Proceedings Fifth IEEE Workshop on Applications of Computer Vision*. IEEE, 2000, pp. 29–34.

[44] L. Skedung, C. El Rawadi, M. Arvidsson, C. Farcet, G. S. Luengo, L. Breton, and M. W. Rutland, "Mechanisms of tactile sensory deterioration amongst the elderly," *Scientific reports*, vol. 8, no. 1, pp. 1–10, 2018.

[45] M. Elias, A. Eltner, F. Liebold, and H.-G. Maas, "Assessing the influence of temperature changes on the geometric stability of smartphone-and raspberry pi cameras," *Sensors*, vol. 20, no. 3, p. 643, 2020.

[46] P. Lall, A. Abrol, and D. Locker, "Effects of sustained exposure to temperature and humidity on the reliability and performance of mems microphone," in *International Electronic Packaging Technical Conference and Exhibition*, vol. 58097. American Society of Mechanical Engineers, 2017, p. V001T01A022.

[47] "Measuring sound with microphones." [Online]. Available: https://www.ni.com/en-us/innovations/white-papers/13/measuring-sound-with-microphones.html

[48] M. Abuhamad, A. Abusnaina, D. Nyang, and D. Mohaisen, "Sensor-based continuous authentication of smartphones' users using behavioral biometrics: A survey," *arXiv preprint arXiv:2001.08578*, 2020.

[49] W. Yuan, S. Dong, and E. H. Adelson, "Gelsight: High-resolution robot tactile sensors for estimating geometry and force," *Sensors*, vol. 17, no. 12, p. 2762, 2017.

[50] K. Ali and A. X. Liu, "Fine-grained vibration based sensing using a smartphone," *arXiv preprint arXiv:2007.03874*, 2020.

[51] T. Chugh, K. Cao, and A. K. Jain, "Fingerprint spoof buster: Use of minutiae-centered patches," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 9, pp. 2190–2202, 2018.

[52] S. Swanson and S. Swanson, "Fingerprints go the distance," Oct 2012. [Online]. Available: https://www.technologyreview.com/s/422400/fin gerprints-go-the-distance/

[53] H. Zhao and H. Malik, "Audio recording location identification using acoustic environment signature," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 11, pp. 1746–1759, 2013.

[54] N. Roy, S. Shen, H. Hassanieh, and R. R. Choudhury, "Inaudible voice commands: The long-range attack and defense," in *15th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 18)*, 2018, pp. 547–560.

[55] F. Ykhlef, M. Arezki, A. Guessoum, and D. Berkani, "A wavelet denoising method to improve detection with ultrasonic signal," in *2004 IEEE International Conference on Industrial Technology, 2004. IEEE ICIT'04.*, vol. 3. IEEE, 2004, pp. 1422–1425.

[56] J. Liu, C. Wang, Y. Chen, and N. Saxena, "Vibwrite: Towards finger-input authentication on ubiquitous surfaces via physical vibration," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2017, pp. 73–87.

[57] J. Li, K. Fawaz, and Y. Kim, "Velody: Nonlinear vibration challenge-response for resilient user authentication," in *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2019, pp. 1201–1213.

[58] Y. Ren, Y. Chen, M. C. Chuah, and J. Yang, "Smartphone based user verification leveraging gait recognition for mobile healthcare systems," in *2013 IEEE International Conference on Sensing, Communications and Networking (SECON)*. IEEE, 2013, pp. 149–157.

[59] W.-H. Lee and R. B. Lee, "Multi-sensor authentication to improve smartphone security," in *2015 International Conference on Information Systems Security and Privacy (ICISSP)*. IEEE, 2015, pp. 1–11.

[60] "Tlv320aic3212 data sheet, texas instruments." [Online]. Available: https://www.ti.com/product/TLV320AIC3212

[61] "Tms320c5532 data sheet, texas instruments." [Online]. Available: https://www.ti.com/product/TMS320C5532

[62] P. Georgiev, N. D. Lane, K. K. Rachuri, and C. Mascolo, "Dsp. ear: Leveraging co-processor support for continuous audio sensing on smartphones," in *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*, 2014, pp. 295–309.

[63] A. Reuther, P. Michaleas, M. Jones, V. Gadepally, S. Samsi, and J. Kepner, "Survey and benchmarking of machine learning accelerators," *arXiv preprint arXiv:1908.11348*, 2019.

[64] N. Horowitz and N. Horowitz, "The energy impacts of smart speakers and video streaming devices," Aug 2019. [Online]. Available: https://www.nrdc.org/resources/energy-impacts-smart-speakers-and-video-streaming-devices

[65] S. Guha and R. Bulusu, "Low power always-on voice trigger architecture," Apr. 23 2015, uS Patent App. 14/060,367.

[66] L. Cai and H. Chen, "Touchlogger: Inferring keystrokes on touch screen from smartphone motion." *HotSec*, vol. 11, no. 2011, p. 9, 2011.

[67] E. Miluzzo, A. Varshavsky, S. Balakrishnan, and R. R. Choudhury, "Tapprints: your finger taps have fingerprints," in *Proceedings of the 10th international conference on Mobile systems, applications, and services*. ACm, 2012, pp. 323–336.

[68] T. Vu, A. Baid, S. Gao, M. Gruteser, R. Howard, J. Lindqvist, P. Spasojevic, and J. Walling, "Distinguishing users with capacitive touch communication," in *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM, 2012, pp. 197–208.

[69] Y. Meng, D. S. Wong, R. Schlegel, *et al.*, "Touch gestures based biometric authentication scheme for touchscreen mobile phones," in *International Conference on Information Security and Cryptology*. Springer, 2012, pp. 331–350.

[70] N. Zheng, K. Bai, H. Huang, and H. Wang, "You are how you touch: User verification on smartphones via tapping behaviors," in *2014 IEEE 22nd International Conference on Network Protocols*. IEEE, 2014, pp. 221–232.

[71] H. Yang, L. Chen, K. Bian, Y. Tian, F. Ye, W. Yan, T. Zhao, and X. Li, "Taplock: Exploit finger tap events for enhancing attack resilience of smartphone passwords," in *2015 IEEE International Conference on Communications (ICC)*. IEEE, 2015, pp. 7139–7144.

[72] M. Shahzad, A. X. Liu, and A. Samuel, "Secure unlocking of mobile touch screen devices by simple gestures: you can see it but you can not do it," in *Proceedings of the 19th annual international conference on Mobile computing & networking*. ACM, 2013, pp. 39–50.

[73] M. Sherman, G. Clark, Y. Yang, S. Sugrim, A. Modig, J. Lindqvist, A. Oulasvirta, and T. Roos, "User-generated free-form gestures for authentication: Security and memorability," in *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*. ACM, 2014, pp. 176–189.

[74] Y. Chen, J. Sun, R. Zhang, and Y. Zhang, "Your song your way: Rhythm-based two-factor authentication for multi-touch mobile devices," in *2015 IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2015, pp. 2686–2694.

[75] H. Khan, U. Hengartner, and D. Vogel, "Targeted mimicry attacks on touch input based implicit authentication schemes," in *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2016, pp. 387–398.

[76] Z. Yan, Q. Song, R. Tan, Y. Li, and A. W. K. Kong, "Towards touch-to-access device authentication using induced body electric potentials," *arXiv preprint arXiv:1902.07057*, 2019.

[77] V. Nguyen, M. Ibrahim, H. Truong, P. Nguyen, M. Gruteser, R. Howard, and T. Vu, "Body-guided communications: A low-power, highly-confined primitive to track and secure every touch," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. ACM, 2018, pp. 353–368.

[78] C. Harrison, J. Schwarz, and S. E. Hudson, "Tapsense: enhancing finger interaction on touch surfaces," in *Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, 2011, pp. 627–636.

[79] Y.-C. Tung and K. G. Shin, "Expansion of human-phone interface by sensing structure-borne sound propagation," in *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2016, pp. 277–289.

[80] K. Sun, T. Zhao, W. Wang, and L. Xie, "Vskin: Sensing touch gestures on surfaces of mobile devices using acoustic signals," in *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. ACM, 2018, pp. 591–605.

[81] M. Goel, B. Lee, M. T. Islam Aumi, S. Patel, G. Borriello, S. Hibino, and B. Begole, "Surfacelink: using inertial and acoustic sensing to enable multi-device interaction on a surface," in *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*. ACM, 2014, pp. 1387–1396.

[82] Z. Li, Y. Wang, A. S. Rathore, C. Song, N. Nyayapathi, T. Vu, J. Xia, and W. Xu, "Pavessel: practical 3d vessel structure sensing through photoacoustic effects with its applications in palm biometrics," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 3, pp. 1–24, 2018.

[83] Y. Zhan, A. S. Rathore, G. Milione, Y. Wang, W. Zheng, W. Xu, and J. Xia, "3d finger vein biometric authentication with photoacoustic tomography," *Applied Optics*, vol. 59, no. 28, pp. 8751–8758, 2020.

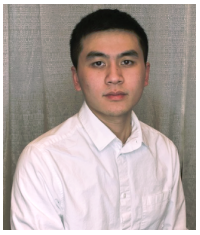
[84] H. Li, C. Xu, A. S. Rathore, Z. Li, H. Zhang, C. Song, K. Wang, L. Su, F. Lin, K. Ren, *et al.*, "Vocalprint: exploring a resilient and secure voice authentication via mmwave biometric interrogation," in *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, 2020, pp. 312–325.

[85] C. Xu, Z. Li, H. Zhang, A. S. Rathore, H. Li, C. Song, K. Wang, and W. Xu, "Waveear: Exploring a mmwave-based noise-resistant speech sensing for voice-user interface," in *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, 2019, pp. 14–26.

[86] C. Zhang, A. Waghmare, P. Kundra, Y. Pu, S. Gilliland, T. Ploetz, T. E. Starner, O. T. Inan, and G. D. Abowd, "Fingersound: Recognizing unistroke thumb gestures using a ring," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, p. 120, 2017.

Aditya Singh Rathore is a 4th year Ph.D. candidate in the Department of Computer Science and Engineering, University at Buffalo where he also received his B.S. degree in Computer Engineering in 2017. During his academic term, he has been awarded multiple honors including UB International Freshman Scholarship, SEAS Senior Scholar Research Scholarship and Presidential Fellowship from University at Buffalo, ACM APSys Student Travel award. Most recently, he received Best Paper Award in ACM MobiSys'20. His research interests include Mobile Security, Biometrics, Internet of Things and Human-Computer Interaction.

Chenhan Xu is a 3rd year Ph.D. student in the Department of Computer Science and Engineering, University at Buffalo. His current research interests include Internet of Things, Mobile Computing, Human-Computer Interaction, and Mobile Health.



Weijin Zhu is a 2nd year Master student in the Department of Computer Science and Engineering, University at Buffalo. His research interests include medical image analysis, big data, and machine learning.

Afee Daiyan is a 5th year undergraduate student studying Computer Science at the University at Buffalo. His current interest includes cyber security, writing technical reports on security breaches and working with iptables to control network traffic flow.

Kun Wang [M'13-SM'17] received two Ph.D. degrees in computer science from Nanjing University of Posts and Telecommunications, Nanjing, China, in 2009, and from the University of Aizu, Aizuwakamatsu, Japan, in 2018. He was a Post-Doctoral Fellow in UCLA, USA from 2013 to 2015, where he is a Senior Research Professor. He was a Research Fellow in the Hong Kong Polytechnic University, Hong Kong, from 2017 to 2018, and a Professor in Nanjing University of Posts and Telecommunications. His current research interests are mainly in the area of Artificial Intelligence and Internet of Things, AI hardware acceleration, and blockchain. He is the recipient of ACM CHI 2020 Horourable Mention Award, ACM FPGA 2020 Best Paper Award Candidate, ACM MobiSys 2020 Best Paper Award, ACM SenSys 2019 Best Paper Award, IEEE GLOBECOM 2016 Best Paper Award, IEEE ISJ Best Paper Award 2019, IEEE TCGCC Best Paper Award 2018, IEEE TCBD Best Paper Award 2019 and CBD 2019 Best Student Paper Award. He is/was the symposium chair/co-chair of IEEE GLOBECOM 2021, IEEE CNCC 2017, IEEE WCSP 2016, etc. He serves/served as an Associate Editor of IEEE Access, an Editor of Journal of Network and Computer Applications, and a Guest Editor of IEEE Network, Future Generation Computer Systems, Journal of Systems Architecture, Peer-to-Peer Networking and Applications, IEICE Transactions on Communications, IEEE Access, Journal of Internet Technology, and Future Internet.

Feng Lin (S'11-M'15-SM'20) received the Ph.D. degree from the Department of Electrical and Computer Engineering, Tennessee Technological University, USA, in 2015. He is currently a Professor with the School of Cyber Science and Technology, College of Computer Science and Technology, Zhejiang University, China. He was an Assistant Professor with the University of Colorado Denver, USA, a Research Scientist with the State University of New York (SUNY) at Buffalo, USA, and an Engineer with Alcatel-Lucent (currently, Nokia). His current research interests include mobile sensing, Internet of Things security, biometrics, AI security, and IoT applications. Dr. Lin was a recipient of the Best Paper Award from ACM MobiSys'20, IEEE Globecom'19, IEEE BHI'17, the Best Demo Award from ACM HotMobile'18, and the First Prize Design Award from the 2016 International 3D printing competition.

Kui Ren received the Ph.D. degree from the Worcester Polytechnic Institute. He is cur- rently a Professor with the Institute of Cyberspace Research, Zhejiang University, and the Director of the UbiSeC Laboratory, State University of New York at Buffalo (UB). He has published 200 papers in peer-reviewed journals and conferences. His cur- rent research interests include cloud and outsourcing security, wireless and wearable systems security, and mobile sensing and crowdsourcing. He is a Distinguished Lecturer of the IEEE, a member of ACM, and a past Board Member of the Internet Privacy Task Force, State of Illinois. He received several best paper awards, including IEEE ICDCS 2017, IWQoS 2017, and ICNP 2011. He received the NSF CAREER Award in 2011, the Sigma Xi/IIT Research Excellence Award in 2012, the UB SEAS Senior Researcher of the Year Award in 2015, the UB Exceptional Scholar Award for Sustained Achievement in 2016, and the IEEE CISTC Technical Recognition Award in 2017. He currently serves on the editorial boards of the IEEE TRANSACTIONS ON DEPENDABLE AND SECURE COMPUTING, the IEEE TRANSACTIONS ON SERVICE COMPUTING, the IEEE TRANSACTIONS ON MOBILE COMPUTING, the IEEE WIRELESS COMMUNICATIONS, the IEEE INTERNET OF THINGS JOURNAL, and the SpringerBriefs on Cyber Security Systems and Networks.

Wenyao Xu received the Ph.D. degree from the University of California at Los Angeles, Los Angeles, USA, and both the Master and Bachelor degree from Zhejiang University, China. Wenyao Xu is an Associate Professor with tenure of the Computer Science and Engineering Department, University at Buffalo (SUNY). His research has focused on exploring novel sensing and computing technologies to build up innovative Internet-of-Things (IoT) systems for high-impact human-technology applications in the fields of Smart Health and Cyber-Security. Results have been published in peer-reviewed top research venues across multiple disciplines, including Computer Science conferences (e.g., ACM MobiCom, SenSys, MobiSys, UbiComp, ASPLOS, ISCA, HPCA, Oakland, NDSS and CCS), Biomedical Engineering journals (e.g., IEEE TBME, TBioCAS, and JBHI), and Medicine journals (e.g., LANCET). To date, his group has published over peer-reviewed 180 papers, won nine best paper awards, two best paper nominations and three international best design awards. His inventions have been filed within U.S. and internationally as patents, and have been licensed to industrial players. His research has been reported in high-impact media outlets, including the Discovery Channel, CNN, NPR and the Wall Street Journal. Currently, Wenyao Xu serves as an Associate Editor of IEEE Transactions on Biomedical Circuits and Systems (TBCAS), the technical program committee of numerous conferences in the field of Smart Health and Internet of Things, and has been a TPC co-chair of IEEE Body Sensor Networks in 2018.