EchoLock: Towards Low-effort Mobile User Identification Leveraging Structure-borne Echos

Yilin Yang
WINLAB, Rutgers University
yilin.yang@rutgers.edu

Yan Wang
Temple University
y.wang@temple.edu

Yingying Chen
WINLAB, Rutgers University
yingche@scarletmail.rutgers.edu

Chen Wang
Louisiana State University
chenwang@csc.lsu.edu

ABSTRACT

Many existing identification approaches require active user input, specialized sensing hardware, or personally identifiable information such as fingerprints or face scans. In this paper, we propose *EchoLock*, a low-effort identification scheme that validates the user by sensing hand geometry via commodity microphones and speakers. *EchoLock* can serve as a complementary verification method for high-end devices or as a stand-alone user identification scheme for lower-end devices without using privacy-sensitive features. In addition to security applications, our system can also personalize user interactions with smart devices, such as automatically adapting settings or preferences when different people are holding smart remotes. To this end, we study the impact of hands on structure-borne sound propagation in mobile devices and develop a user identification scheme that can measure, quantify, and exploit distinct sound reflections in order to differentiate distinct identities. Particularly, we propose a non-intrusive hand sensing technique to derive unique acoustic features in both time and frequency domain, which can effectively capture the physiological and behavioral traits of a user's hand (e.g., hand contours, finger sizes, holding strengths, and holding styles). Furthermore, learning-based algorithms are developed to robustly identify the user under various environments and conditions. We conduct extensive experiments with 20 participants, gathering 80,000 hand geometry samples using different hardware setups across 160 key use case scenarios. Our results show that *EchoLock* is capable of identifying users with over 94% accuracy, without requiring any active user input.

CCS CONCEPTS

• **Security and privacy → Authentication**.

KEYWORDS

user identification; biometrics; internet of things; acoustic sensing

Chen Wang's contribution was as a graduate student at Rutgers University.

ASIA CCS '20, October 5–9, 2020, Taipei, Taiwan
© 2020 Association for Computing Machinery.
ACM ISBN 978-1-4503-6750-9/20/10…$15.00
https://doi.org/10.1145/3320269.3384741

ACM Reference Format:
Yilin Yang, Yan Wang, Yingying Chen, and Chen Wang. 2020. EchoLock: Towards Low-effort Mobile User Identification Leveraging Structure-borne Echos. In *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security (ASIA CCS '20), October 5–9, 2020, Taipei, Taiwan.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3320269.3384741

1 INTRODUCTION

User identification is a fundamental and pervasive aspect of modern mobile device usage, both as a means of maintaining security and personalized services. Verifying oneself is necessary to gain access to smartphones, bank accounts, and customized news feeds; information and resources which must be available on demand. As such, repeated acts of authentication can grow tedious and consume unnecessarily long portions of daily routines involving mobile devices. Studies on cellphone addiction suggest that user identification procedures encompass up to 9% of daily usage time [19], with related inquiries showing strong interest in more convenient practices [39]. Techniques such as facial recognition or fingerprinting do not require considerable effort from the user, but demand dedicated hardware components that may not be available on all devices. This is of particular importance for markets in developing countries, where devices such as the Huawei IDEOS must forgo multiple utilities in order to maintain affordable price points (e.g. under $80) [15, 17]. Secure and effective identification necessitates a lightweight protocol to facilitate tailored services at low cost.

To this end, we propose *EchoLock*, a low-effort user identification scheme for commercial-off-the-shelf (COTS) mobile devices. By allowing an acoustic signal to propagate through the mobile device, it is possible to measure properties of human hand geometry, a biometric indicator known to be accurate for user identification [7], yet rarely employed in mobile applications due to obstacles in obtaining accurate measurements with limited hardware. We show that pressure applied by a person's hand on the device creates unique and observable impacts on structure-borne sound propagation. By using a designated inaudible signal, *EchoLock* can capture such impacts and extract the user's unique hand biometrics for user identification. Our approach is low-effort as no conscious action is required by the user; holding the device itself is the user identification action as shown in Figure 1.

Because structure-borne sound propagation depends on material, dimensions, and external forces , one hand does not produce the same acoustic pattern when holding different devices. Similarly, one device does not produce the same pattern when held by different hands, making our credential a secure key that represents

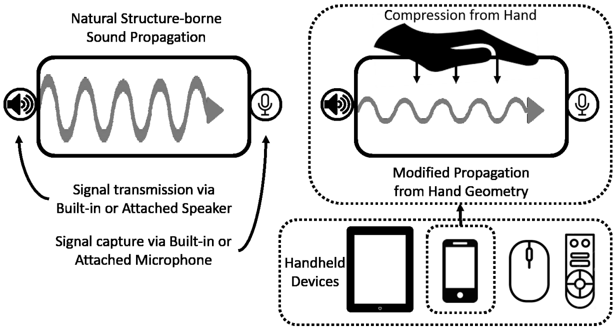

Natural Structure-borne Sound Propagation
Signal transmission via Built-in or Attached Speaker
Signal capture via Built-in or Attached Microphone
Compression from Hand
Modified Propagation from Hand Geometry
Handheld Devices

Figure 1: Capture of hand biometric information embedded in structure-borne sound using commodity microphones and speakers.

a specific hand-device pair. By using only readily available speakers and microphones, our system is non-intrusive and low cost. We envision that the availability of these hardware components will only increase with the rising prevalence of integrated Internet of Things (IoT) devices built with virtual assistants and voice controllers, projected to reach an install base of over 75 billion by 2025 [1].

Existing solutions in the market are typically considered effective, but do have some limitations regarding ease of use. Actions such as password entry, voice utterances, or finger presses demand, however briefly, the user's attention and active participation in the process. In contrast, *EchoLock* is a passive procedure. Our technique serves as a viable standalone identification system, or as a complementary system for multi-factor authentication due to its naturally low involvement. Password security, for example, can be enhanced by simultaneously sampling hand biometrics during typing or swiping actions, compensating for common vulnerabilities (e.g. PIN codes spied on through shoulder-surfing attacks cannot be used if the attacker's hand is not recognized by the device).

As a standalone technique, *EchoLock* is applicable to a wide variety of services. Resources such as financial accounts or health apps on smartphones can trigger a single-instance identification check to verify the user's hand before divulging sensitive information. Private message notifications can be displayed or hidden onscreen depending on which person is currently holding the device. The integration of IoT with home security systems, such as Amazon Blink [8], can enable even unconventional objects to be compatible with *EchoLock*. For example, door handles and safety railings can be used to passively sense structure-borne sound propagation when held and open or lock entrances accordingly. Beyond security, appliances such as smart remotes can also employ our system to enhance the user experience. The Amazon Fire TV stick [4] is equipped with microphones and Alexa support, making our system easy to deploy for personalized user settings and TV channels at no additional cost. The speed of sound propagation is rapid, even when traveling through physical mediums, making our system latency competitive with existing technologies. Continuous authentication can also be implemented via periodic measurements.

Building *EchoLock* for such applications does present many challenges, the most prominent being the development of a non-intrusive approach that leverages a single pair of low-fidelity speaker and microphone to capture unique characteristics of a user's hand biometrics, which usually has only minute differences between people. In addition, the acoustic signal propagating from the device's speaker to its microphone usually experiences the multipath effect, resulting in airborne and structure-borne signals that requires careful separation. Moreover, the ambient noises and acoustic signals reflected off the environment create interference that needs to be accounted for. Finally, many factors could impact the robustness of the proposed approach, such as device shape or material.

Table 1: Qualitative comparison of existing user ID methods.

Identification Technique	Evaluation Category	Personally Identifiable	Physiological Credentials	Behavioral Credentials	Dedicated Hardware
Image [13]	Knowledge	No	No	Yes	No
Face [16]	Visual	Yes	Yes	No	No
Fingerprint [26]	Visual	Yes	Yes	No	Yes
Iris [10]	Visual	Yes	Yes	No	Yes
Gait [43]	Visual	No	Yes	Yes	No
Voice [20]	Acoustic	Yes	Yes	Yes	No
Our Work	**Acoustic**	**No**	**Yes**	**Yes**	**No**

To address these challenges, *EchoLock* utilizes an ultrasonic signal to sense a user's mannerisms when holding a device. A high-frequency, short duration transmission is selected to reduce audible disturbances to the user and provide prompt validation. We distinguish structure-borne and near-surface airborne signals based on differing travel speeds in air and solid materials [2]. The system applies a band-pass filter to remove ambient acoustic noises that do not share the same spectrum as the designated ultrasonic signal. We derive fine-grained acoustic features in the time and frequency domains, as well as acoustic features, to capture the unique hand biometrics. We further develop learning-based user identification algorithms to robustly identify the user when considering various impact factors. Our main contributions in this work are as follows:

- We study the impact of hand biometrics (i.e., hand geometry, holding strengths and holding styles) on structure-borne sound propagation through mobile devices and design an acoustic sensing-based technique to measure these effects using limited hardware in mobile devices. We show that users' unique physiological and behavioral hand biometrics can be captured by using a designated acoustic signal.
- We develop a low-effort user identification system for mobile devices that validates hand biometric information based on acoustic sensing. The proposed system does not require any input from the user and is non-intrusive by utilizing inaudible frequencies.
- We identify unique acoustic features, including time-domain, frequency-domain and acoustic features, to capture the user's hand biometrics. We also develop robust learning-based methods to distinguish users based on their unique hand biometrics.
- We implemented an early prototype of *EchoLock* on various mobile devices and evaluated performance under multiple conditions. With over 80,000 hand geometry samples gathered over 160 trials of key use case scenarios, we show identification accuracy upwards of 94%.

2 RELATED WORK

Routine identification methods typically assess possession of text or numerical keys such as passwords [35]. In such cases, the user must either commit to memory a complex sequence or settle for a trivial key at the expense of security. Graph-based [38] and image-based [14, 34] methods propose swipe patterns and picture recognition as more intuitive alternatives. While effective, these methods verify knowledge rather than the user and require active input.

In contrast, biometric-based approaches use physical traits of users as credentials, which could enable passive user authentication and reduce user effort. Several popular examples include face ID [5], capacitive fingerprint scanning [12], and iris scanning [9]. Physiological credential are typically unique to a person and do not change abruptly over time, making them ideal for identification systems. However, these approaches require dedicated hardware components to make accurate measurements, which limits the pool of devices on which they can be deployed. Furthermore, theft of these credentials are highly problematic since they involve personally identifiable information [40].

Human behaviors have also been employed as credentials. For example, prior works have shown it is possible to identify individuals based on hand gestures [21, 22], voice commands [5, 41], as well as finger inputs made on touchscreens [29, 31, 32], solid surfaces [23], and wearable devices [3]. These are less personally identifiable, and thus pose a smaller risk to user privacy, but can be challenging to associate with a given identity due to natural inconsistencies users exhibit when asked to reproduce these characteristics

Proposals have been made to measure credentials in a passive, low-effort manner, which we consider closely related to *EchoLock*. Ren *et al.* use accelerometer readings in mobile devices to derive unique gait patterns and passively verify the user as they walk [28]. Zheng *et al.* extract behavioral patterns from touchscreen taps (e.g., rhythm, strength, angle of applied force) using built-in accelerometers, gyroscopes, and piezoelectric sensors to provide non-intrusive user authentication [42]. Zhou *et al.* develop an attack to passively detect lockscreen swipe patterns based on acoustic reflections produced by the user's fingers during input [45].

We summarize the findings of prior work in Table 1. Unlike existing approaches, *EchoLock* leverages novel hand biometrics including hand-related physiological (i.e., hand geometry) and behavioral (i.e., holding strengths and styles) traits to provide convenient and secure user identification. By performing fine-grained acoustic sensing to capture the unique hand biometrics, the user can be identified passively when holding their personal device. The natural availability of speakers and microphones makes acoustic sensing a widely used technique in many mobile computing applications, such as indoor localization [36] and human-computer interaction [33, 37]). To our best knowledge, we are the first work to utilize acoustic sensing to capture hand biometric information for low-effort user identification. Our proposal does not depend on personally identifiable information, active user inputs, or specialized hardware.

3 ADVERSARY MODEL

Malicious users may attempt to attack our system in order to gain access to personal information or deny legitimate users from accessing services. In this section, we introduce attack strategies that may be deployed against *EchoLock*.

Impersonation Attack. The attacker attempts to mimic the holding posture of the legitimate user to gain access to the device. For impersonation attacks, the attacker may either be informed or uninformed. In the uninformed case, the attacker possesses no knowledge on how to circumvent the identification process and naively attempts to mimic the legitimate user's holding behavior. In the informed case, however, the attacker is explicitly aware of the legitimate user's authentication credentials in some form. This may be through passive observations of the user's hands or interactions such as handshaking. Physically faking the legitimate user's profile, however, requires applying forces to the device such that they create structural deformations similar to how the user's own hands would.

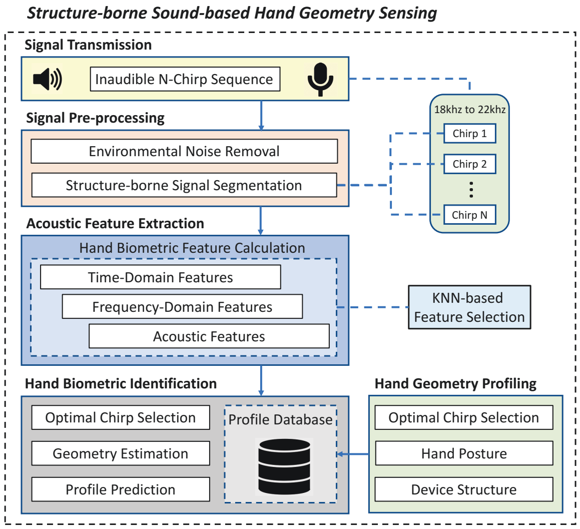


Figure 2: System overview of EchoLock.

Eavesdropping and Replay Attack The attacker attempts to steal acoustic credentials of the legitimate user by eavesdropping instances of identification attempts. This may be done by positioning a microphone near the user as *EchoLock* is deployed. After obtaining an audio sample of a signal used to authenticate the user, the attacker gains possession of the targeted device and replays the audio sample via an external speaker. During ultrasonic sensing, the mobile device will transmit and record our acoustic signal. In order to succeed, the attacker must first suppress or bypass the signal transmission stage to avoid overlap with their attacking signal, which is a non-trivial challenge.

Jamming Attack. The attacker in this scenario is focused on deliberate sabotage of genuine authentication attempts. This may be carried out by playing loud noise or ultrasonic frequencies near the user to disrupt the geometry estimation procedure. The attacker does not necessarily need to know the user's credentials to jam the system. We assume in our assessment that the attacker will utilize ultrasonic frequencies to decrease the chances of detection by the ordinary user.

4 SYSTEM OVERVIEW

4.1 Design of EchoLock

As people have small differences in their hand biometrics, it is critical to design *EchoLock* in such a way that it can perform fine-grained acoustic sensing to capture the small differences among different users' hand biometrics by using low-cost COTS mobile devices. The basic idea of EchoLock is to leverage a speaker and microphone to transmit, receive, and analyze structure-borne sound waves as illustrated in Figure 2. Many mobile devices, such as smartphones, touch pads, and remote controls, are equipped with such components and have many applications regarding security and personalization. *EchoLock* first transmits an inaudible acoustic sequence, consisting of n inaudible chirp signals ranging from 18kHz to 22kHz. Then it immediately records the reflections from the user's hand via onboard microphones. This procedure can be initiated by a pre-determined trigger, such as raise [5] or squeeze [25] detection found in COTS mobile devices.

The recorded response undergoes our *Signal Pre-processing* phase, where we apply a band-pass filter to remove ambient noise and conduct the *Structure-borne Signal Segmentation* to extract the reflection of the transmitted n-chirp signal via the structure-borne propagation paths. After the Signal Pre-processing, we analyze each chirp signal in *Acoustic Feature Extraction* to determine meaningful features capable of differentiating user hand biometrics in the time and frequency-domain, including statistical properties such as average or median, spectral points of the FFT, and MFCC coefficients. To ensure the effectiveness of the candidate features, we perform the *KNN-based Feature Selection* to identify the features that are sensitive to forces exerted by the user's hand. We note that such features may not necessarily be consistent for the same hand when holding different physical structures (e.g., a different smartphone) due to altered sound propagation properties.

Next, our system performs *Hand Geometry Profiling* and *Hand Biometric Identification* to determine the user's identity based on the extracted features. Extracted features representative of the interaction between hand posture and device structure are compiled into a $m \times n$ matrix data structure, where m corresponds to number of features, and saved to a *Profile Database*. This database is then referenced for a profile match when identifying the user. Through our experiments with 20 participants, we empirically find that chirps with different frequency ranges may contain different degrees of useful information. To combat this, we adopt a *Optimal Chirp Selection* method to quantify the likelihood a given chirp signal successfully captured detailed biometric information. We provide a select number of chirp signals as inputs for our *Geometry Estimation* phase to generate a multi-dimensional characterization array using the chirp feature matrices. Finally, we employ a machine-learning based approach to match the extracted hand biometric features with users' profiles in *Profile Prediction*, where our system deduces the most probable hand geometry match by examining the numerical distance discrepancies and concludes with a predicted profile label. This output can control a desired functionality, such as unlocking a device or switching user accounts.

4.2 Challenges and Requirements

Using a single built-in speaker and microphone available on a mobile device to sense complex hand geometry is an unexplored area. Because acoustic signals travel rapdily compared to the small dimensions (i.e. sensing area) of mobile devices (e.g. 15 cm between a smartphone speaker and microphone), the existing built-in microphone can only receive limited acoustic samples (< 20 samples [33, 37]) to describe a complete propagation. Additionally, the acoustic signals arriving at the microphone are the combination of structure-borne propagation and airborne propagation, requiring delicate separation. The environmental reflections of the acoustic sensing signals and the ambient noises corrupt the received sound and make the acoustic analysis of the user's holding hand even harder. Besides addressing these challenges, we also need to consider both security and usability when designing the system. In particular, the passive user input to our system should be hard to observe and imitate to meet security requirements.

5 STRUCTURE-BORNE SIGNAL DESIGN

5.1 Sound Propagation on Mobile Devices

Structure-borne sound is most often recognized as vibration and can be perceived both by ear and touch. From Hooke's Law [30], the speed of sound through a medium can be represented as a function, formulated as $c = \sqrt{\frac{K}{p}}$ where K is the bulk modulus of elasticity, or Young's modulus, and p is the medium density. The structure path is more direct compared to in the air due to the greater density and compression resistance of the mobile device, allowing sound to travel and be received more quickly. This trait is of interest as propagation through a physical medium provides natural resilience to reflections from distant obstacles as there is minimal deviation from the sound path.

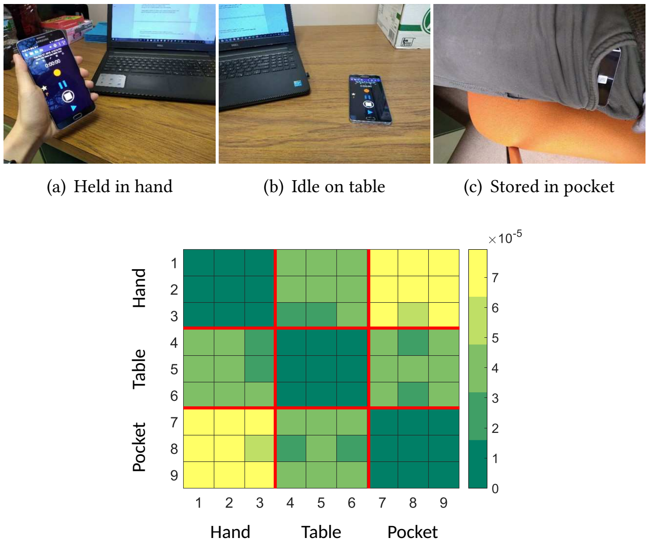
However, structure-borne propagation is much more sensitive to physical disturbances. Interactions such as touching the medium can significantly alter the acoustic patterns as the contact and force exerted upon the medium changes how it reverberates. While this normally poses a challenge for acute acoustic sensing, *EchoLock* exploits this for the purposes of recognizing individual people. The force of a user's grip on the mobile device is integral to the system, essentially extending the medium to encompass both the device and user's hand. Bulk modulus of elasticity can be expressed as:

$$K = \frac{-(p_1 - p_0)}{(V_1 - V_0)/V_0} = \frac{\Delta P}{\Delta V/V_0}, \tag{1}$$

for a given differential change in pressure ΔP and volume ΔV relative to an initial volume V_0. The introduction of a stable additive density V_1 and fluctuating pressure increase p_1 by the user changes K to a dynamic set of elasticity constants. This produces a range of acoustic patterns representative of how the device is held by a specific individual at the time of measurement, uniquely shaped by hand contour, posture, grip pressure, and behavior. Note that this model is an incomplete explanation as it does not account for a distributed application of pressure from different focal points of a user's hand. However, we find this explanation sufficient for the purposes of conveying the intuition behind *EchoLock*'s premise.

Structure-borne Propagation Feasibility. A preliminary experiment was performed to gauge the ability for a mobile device to ascertain environmental conditions using acoustic sensing. Three different use cases were selected with the intent of demonstrating that conditions with distinct forces exerted upon the mobile device could be easily recognized. Figure 3 shows the particular experimental setup for each the three scenarios; having the mobile device resting in the user's hand, on a table, or within the user's pocket.

An inaudible chirp signal sweeping from 18 kHz to 22 kHz was emitted from the bottom device speakers to induce vibration, which is then recorded by a single microphone near the top of the device. This captured recording is then examined for features that may identify environmental origins. The results of these experiments can be



(a) Held in hand (b) Idle on table (c) Stored in pocket

(d) Euclidean distance of amplitudes recorded from different scenarios. Lower distances imply similarity.

Figure 3: Preliminary experiments for environment differentiation.

seen in Figure 3(d). Three samples from each use case were obtained and the average amplitude for each recording was extracted. The absolute difference was then computed for each combination of samples to measure statistical distinction between each use case and show a clear relation between samples originating from different use cases. Samples native to the same use case displayed naturally low amplitude difference, which increased when compared to samples from foreign cases. Note across the diagonal that the difference is zero as these are comparisons between a sample with itself. This suggests that, even with minimal sensors and signal processing, structural-borne sound propagation is capable of communicating critical information about the immediate surroundings.

5.2 N-chirp Sequence for Acoustic Sensing

Measuring biometric properties of a given person accurately using only sound propagation requires our signal to satisfy several design criteria:

- The signal must be designed in such a way to easily distinguish between the structure-borne and airborne sound propagation.
- The transmitted signal should be recognizable such that it can be easily identified and segmented amid interference from ambient noise and other acoustic disruptions.
- The signal should fall within a safe frequency range inaudible to average human hearing. This is primarily for the purpose of usability as a noticeably audible signal may pose a nuisance to some users. Generally speaking, 16kHz is the upper bound of easily detectable sound for ordinary adults [6].
- The signal must be able to be deployed on a COTS device, limiting the viable transmission frequency range. Android devices, for example, are reported to have a maximum sampling rate of around 44kHz, limiting a practical signal to 22kHz at most [18]. Many devices, however, exhibit considerable attenuation problems when transmitting frequencies exceeding 20kHz due to hardware imperfections in onboard speakers [33, 36, 37, 44].

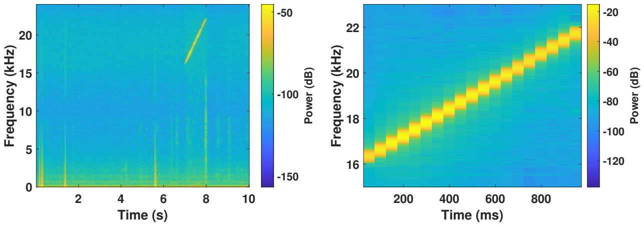
With these considerations, we design a *n-chirp sequence* where n describes a number of repeating chirp signals utilizing the inaudible frequency range from 18kHz to 22kHz. Though this frequency sweep may be perceptible to sensitive groups, such as pets or young children, the duration is brief (i.e. milliseconds) to minimize disturbances. Each chirp consists of 1200 samples, equating to a 25ms duration at a 48kHz sampling frequency. During ultrasonic sensing, the recorded structure-borne propagation of the chirp signal will be embedded with information on the user's hand geometry. While a shorter chirp minimizes exposure to environmental reflections, it also limits the signal-to-noise ratio (SNR). From our experiments, we observed that most COTS speakers struggle to consistently sweep a wide, high frequency band in such a short time period. To balance these considerations, we transmit a series of consecutive chirps, where each chirp is of a singular frequency such that the first chirp is 18kHz while the n-th chirp is of 22kHz. This frequency increments in steps of 1kHz at every $n/5$-th chirp.

In addition, we separate these chirps with 25ms of empty buffers to stagger the arrival of environmental reflections from structure-borne sound. A pilot signal of 22kHz is prefixed to the sequence for the purposes of simplifying signal segmentation in later procedures, but is not used directly to sense information on the user. By utilizing multiple chirps, we can also gather multiple user samples in a single ultrasonic sensing instance. This leads to a natural trade-off dilemma between higher classification accuracy and shorter time delays. We show in further detail the performance accuracy for increasingly large n values in Section 8.5.

5.3 Structure-borne Signal Segmentation

While structure-borne and airborne sound are both capable of carrying information indicative of the surrounding environment, airborne sound is less reliable as a metric due to distortions from the multipath effect. The minute delay of airborne sound may also introduce noise to subsequent structure-borne sound transmissions due to asynchronous arrival time, necessitating separation of the two in order to maintain robust feature extraction and user identification.

Existing studies leverage the difference between the propagation speed of sound waves to distinguish between the structure-borne and airborne signals [37]. Knowing that propagation is faster through physical mediums, we can expect structure-borne sound to always arrive at the microphone earlier compared to airborne sound. Therefore, we apply cross-correlation to derive similarities between the recorded acoustic signal with the original transmission and identify the structure-borne signal based on the time of arrival. In particular, we compute correlation through the function $\sum_{m=0}^{\infty} x^*(m)y(m+d)$ where $x^*(m)$ is the complex conjugate of our transmission and $y(m+d)$ is the recorded propagation sequence time shifted by some unknown delay d. Both $x^*(m)$ and $y(m+d)$ are normalized to compensate for amplitude differences in the differing sound propagation. By locating the index with the highest correlation to our transmission, we can determine the start point of our signal. This recorded signal contains both structure-borne and airborne sound propagation, represented by several amplitude peaks. Environmental reflections arrive at a much slower speed,



(a) Full Recording

(b) Extracted Signal

Figure 4: Signal segmentation on a microphone recording containing the desired acoustic signal.

thus we can safely eliminate interference by keeping our transmission short (i.e. milliseconds long), calculating the signal endpoint based on sequence length, and segmenting audio at this point.

Due to the short distance between the speaker and microphone, structure-borne sound usually overlaps with airborne sound. Therefore, the obtained signals still contain partial airborne signals, which must be accounted for. As such, we apply a third-order median filter to mitigate undesirable outliers introduced by airborne sound in our recorded signal. The output of the median filter is considered to only contain the structure-borne signals. Figure 4 shows an example acoustic signal before and after the proposed structure-borne signal segmentation. A short chirp signal sweeping within the inaudible frequency band is played during a 10 second recording and successfully extracted using our outlined procedure. An inspection of the frequency domain confirms that we have preserved our signal without any persisting interference from airborne sound or environmental reflections.

5.4 Hand Geometry-Induced Chirp Selection

Although *EchoLock* utilizes a sequence of chirp signals to validate the user, we find that not all n chirps necessarily provide equally detailed information on hand geometry. The intuition behind our proposed system is that the form and grip of the user's hand would shape the transmitted chirp signal in such a unique way that it can be used for identification purposes. However, we observe during development that some chirps signals were not transformed in a meaningful way, bearing more resemblance to waveforms originating from scenarios depicted in Figure 3(b) than Figure 3(a).

The absence of new information in these signals may be attributed to a variety of factors, particularly if the user were to temporarily disrupt physical contact with the device due to fidgeting. This contributes to misidentification rates as separate users with similarly uninformative chirp signals consequentially have similar training inputs in our machine learning framework. As such, we develop an optimal chirp selection method to quantify the impact of the user's hand on our n-chirp sequence. The optimal chirps identified by the method are used to build a particular user's identity profile and identify the user during geometry estimation.

In particular, we denote a chirp signal of k points as a vector $\mathbf{u_i} = (u_1, u_2, ..., u_k)$, where i is the i-th chirp of the sequence such that $1 \leq i \leq n$. We also denote the baseline chirp signal, originating from the scenario where the phone is put on the desk (Figure 3(b)), as $\mathbf{v_i} = (v_1, v_2, ..., v_k)$. Then we compute the absolute difference between these vectors such that $\mathbf{d_i} = (|u_1 - v_1|, |u_2 - v_2|, ..., |u_k - v_k|)$ with the intuition that a signal that is properly modified by hand geometry would produce a noticeably distinct signal and therefore a larger $\mathbf{d_i}$ vector. Next, we order the chirps based on the average of $\mathbf{d_i}$ and select the first $n/2$ chirps as the optimal chirps to describe the hand biometrics of the user.

These chirps are used for geometry estimation, where we develop a final characterization of the user's hand biometrics. Similar to the previous selection process, each frequency step (i.e. 18kHz, 19kHz, etc.) is ordered based on the proportion of optimal chirp signals the frequency produced, determined by the i denotation. A multidimensional array is constructed using all feature matrices of the first frequency, followed by s number of features matrices from the second frequency. This subset s is chosen to be $n/10$ through experimentation, with feature matrix selection based off chirp signals with the largest d_i score. The optimal chirp selection method is performed in the profiling stage, where the index of the selected optimal chirp is stored with the recorded acoustic sound as a user's profile. During hand biometric identification, these chirps are referenced for geometry estimation. Note that different users may have different combinations of optimal chirp signals, which increase the diversity of users' profiles and help improve the identification accuracy.

Moreover, we use multiple consecutive n-chirp sequences in testing to capture the user's hand geometry at different times, aiming to improve the robustness of the user identification by performing a majority vote on the results from the acoustic response resulted from the multiple consecutive n-chirp sequences. We specify an odd number of optimal chirps in order to prevent ties when voting on the final decision in the identification process.

6 HAND BIOMETRIC USER IDENTIFICATION

6.1 Acoustic Feature Extraction

After obtaining the structure-borne echos from the received designated signal, the system extracts from it unique features to analyze the interferences caused by the user's hand and derive a hand biometric profile, which integrates both physiological and behavioral traits. A series of candidate features are identified for their potential responsiveness to different user hand biometrics. These features include statistical properties in the time domain, the spectral points in the frequency domain, and acoustic properties such as Mel-Frequency Cepstral Coefficient (MFCC) and Chromagram features.

Time-domain Statistic Features. In the time domain, we choose to analyze signals by its statistics including *mean*, *standard deviation*, *maximum*, *minimum*, *range*, *kurtosis* and *skewness*. We also estimate the signal's distribution by calculating *second quantile*, *third quantile*, *fourth quantile* and *signal dispersion*. Additionally, we examine *peak change* by deriving the index position of the data point that deviates most significantly from the statistical average. Figure 5 illustrates how these features may be applied to differentiate hand geometry. We extract our time-domain features for two distinct users while they are using the same model mobile device. By plotting the kurtosis, standard deviation, and range features, we can see an apparent clustering effect, indicating these statistics show viability as distinguishing factors.

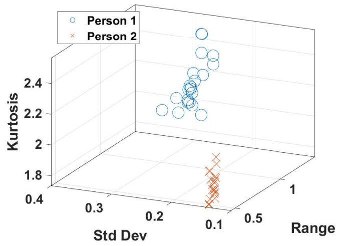


Figure 5: Illustration of the time-domain statistical features to differentiate two people's hands.

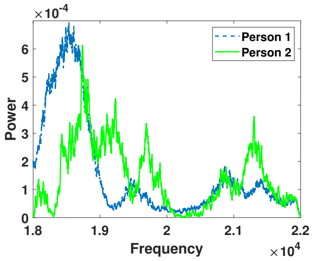


Figure 6: Spectral analysis of the received acoustic signal for two users when holding a mobile device.

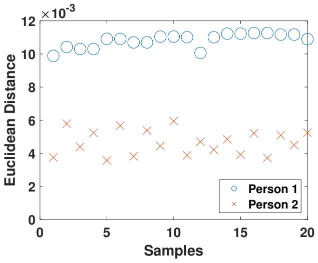


Figure 7: Euclidean distance of the standard deviation feature for two users relative to a separate instance for Person 2.

Frequency-domain Features. In the frequency domain, we apply Fast Fourier Transformation (FFT) to the received acoustic signal and derive 256 spectral points to capture the unique characteristics of the user's holding hand in the frequency domain. This is because the holding hand can be considered as a filter, which results in suppressing some frequencies while not affecting others. Figure 6 shows an example spectral analysis of the received sound for two separate users. We observe unique responses produced by these individuals, which consequentially produce unique FFT points when extracting our frequency-based features. While these two users exhibit mostly identical responses within the 20-21kHz range, our optimal chirp selection process ensures these similarities are not heavily considered when attempting to differentiate between different people.

Acoustic Features. We also derive the acoustic features from the received sound using the MFCC [24] and Chromagram [27]. MFCC features are normally applied in speech processing studies to describe the short-term power spectrum of the speech sound and are good for reflecting both the linear and non-linear properties of the sound's dynamics. Chromagram, often referred to as "pitch class profiles", is traditionally utilized to analyze the harmonic and melodic characteristics of music and categorize the music pitches into twelve categories. We have observed the sensitivity of the MFCC and Chromagram to be sensitive enough to respond to physical biometrics as well. In this work, we derive 13 MFCC features and 12 chroma-based features to describe the different hand holding-related interferences to the sound. Our MFCC features include 13 filter bank coefficients processed using Discrete Cosine Transform whereas our chroma-based features describe a correlation between the recorded signal and one of the 12 tonal pitches along a even-tempered scale.

6.2 Hand Biometric Feature Selection

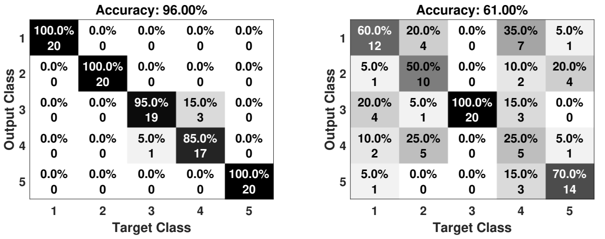
After feature extraction, we obtain 12 time-domain features, 256 frequency-based features, and 25 acoustic features for 293 total features. Some features are more sensitive to the minute differences of handwhile the others may not be very effective at distinguishing between them. Moreover, mobile devices from different vendors may have their speaker and microphone embedded at different positions. These hardware distinctions introduce further uncertainty when measuring the user using our features. We choose a wrapper-based strategy for selecting our features. Though computationally intensive, we believe the optimization of the classifier problem to be more valuable than filter-based methods such as variance threshold or correlation coefficient, which are simpler to implement but less model oriented. In this work, we develop a K-nearest neighbour (KNN) based feature selection method to find the more salient features for *EchoLock*.

In particular, we apply KNN to each type of feature to obtain the clusters for different users. We then calculate the Euclidean distance of each feature point to its cluster centroid and that to centroids of other clusters. The purpose is to calculate the intra-cluster and inter-cluster distances to measure whether a feature is consistent for the same user and simultaneously distinct for different users. Next, we divide the average intra-cluster distance over the average inter-cluster distance and utilize an experimental threshold to select the features. From our list of candidate features described in Section 6.1, we narrow our selection to the best 6 time-domain features, 12 frequency-domain features, and 12 acoustic features. The selected features based on KNN are not only sensitive to the user' hand holding activity but also resilient to the other factors such as acoustic noises.

6.3 Learning-based Holder Identification

We develop learning-based algorithms to learn the unique characteristics of the user's hand holding activity based on the derived acoustic sensing features and determine whether the current device holder is the legitimate user or not. The classifiers are trained during the user profile construction phase that is detailed in Section 6.4. During the user verification phase, *Echolock* first classifies the testing data to one user based on the user profile. For each analyzed chirp signal, our algorithm utilizes the prediction probabilities returned by the classifier as a confidence level and applies a threshold-based method to examine the classification results. If the confidence level of the classification is above the threshold, a user label is predicted. A majority vote of user labels for all processed chirp signals is conducted to determine the final identity. If the confidence is beneath threshold certainty or the algorithm is unable to achieve consensus during voting, our system will determine the user as "unknown" and respond accordingly. This can be used for multiple applications, such as immediately adjusting user settings when a registered user is detected, or locking devices when an unknown user attempts to gain access.

We explore various candidate machine learning-based classifiers including Bagged Decision Trees (BDT), Linear Discriminant Analysis (LDA), K-Nearest Neighbor (KNN), and Support Vector



	1	2	3	4	5
1	100.0% 20	0.0% 0	0.0% 0	0.0% 0	0.0% 0
2	0.0% 0	100.0% 20	0.0% 0	0.0% 0	0.0% 0
3	0.0% 0	0.0% 0	95.0% 19	15.0% 3	0.0% 0
4	0.0% 0	0.0% 0	5.0% 1	85.0% 17	0.0% 0
5	0.0% 0	0.0% 0	0.0% 0	0.0% 0	100.0% 20

(a) Top microphone

	1	2	3	4	5
1	60.0% 12	20.0% 4	0.0% 0	35.0% 7	5.0% 1
2	5.0% 1	50.0% 10	0.0% 0	10.0% 2	20.0% 4
3	20.0% 4	5.0% 1	100.0% 20	15.0% 3	0.0% 0
4	10.0% 2	25.0% 5	0.0% 0	25.0% 5	5.0% 1
5	5.0% 1	0.0% 0	0.0% 0	15.0% 3	70.0% 14

(b) Bottom microphone

Figure 8: Classification accuracy for two available microphones based on signal transmission from a bottom-positioned speaker.

Machines (SVM) and find SVM to be the most effective among all the classifiers. The SVM architecture is thoroughly studied and well suited for lightweight and mobile platforms such as Android, which helps ensure our system is deployable on many devices. SVM relies on a hyperplane to divide the input acoustic sensing feature space into the categories with each representing a user. The hyperplane is determined during the training phase with the acoustic sensing data from the registered users. We use LIBSVM with a cubic kernel to build the SVM classifier [11].

6.4 User Profile Construction

EchoLock requires the user to provide the training data for user profile construction during registration. Our *n*-chirp sequence is transmitted and, based on the optimal chirp selection, a subset of the chirp signals are processed for features to represent the user. These features are shaped not only by hand geometry, but the device structure itself. As structure-borne sound propagation is specific to material, dimensions, and external forces (i.e. a firm grip by the user), we construct a secure credential that is specific to a particular user and device pair. This also prevents an isolated security breach from compromising other devices and services of a given user as structure-borne properties will differ from model to model. An existing data set of anonymized user profiles is included to serve as negative labels when classifying the user during identification attempts. Registering multiple users to the same device also serves to expand this data set. To ensure robustness and low false negative rates, the user is advised to vary holding behavior multiple times rather than remain still to train the data. This can be verified using motion sensors to detect change in device holding posture.

7 IMPLEMENTATION

7.1 Sensing Microphone Selection

Contemporary mobile devices are often built with multiple on-board microphones, particularly smartphones which employ them for noise cancellation during phone calls. These devices also usually contain multiple speakers, which are typically located at the extremities to reserve central space for screens or buttons. We find that for devices with more than a single on-board microphone, reception of acoustic signals from the speakers can vary considerably. For example, speakers and microphones positioned adjacently generally gather less structural information as the propagation path is extremely short. When positioned at opposite ends, however, the signal is able to propagate through nearly the entire device, allowing it to be shaped by the positioning of palm and fingers. Figure 8 shows a small-scale example of two microphones' ability to distinguish 5 people on a device using a bottom-oriented speaker. When evaluated using samples provided from 5 people, we observe identification accuracy of 96% and 61% for the top and bottom microphones, respectively.

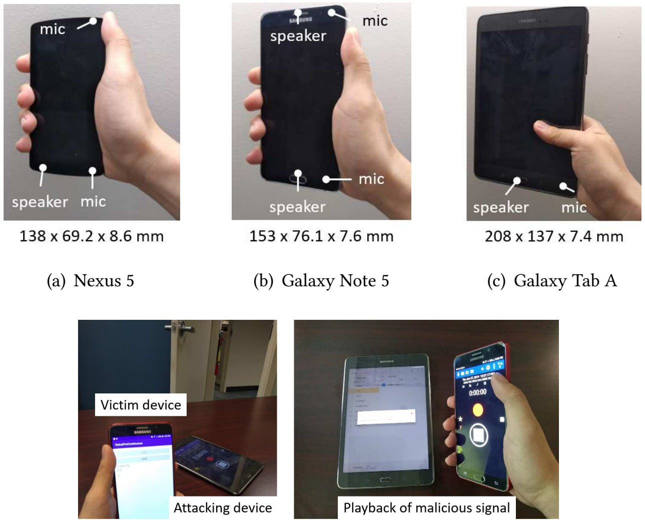
This stipulates that our speaker and microphone must be orientated further apart such that they encompass as much of the device as possible for ideal measurement. As such, we implement our signal transmission and obtain our results using data provided from speaker and microphone pairs with the greatest amount of separation when an option to select exists. This can be achieved through a one-time configuration process by estimating time-of-flight of a sample ultrasonic signal for all speaker-microphone combinations.

7.2 Posture Stabilization Detection

The design of *EchoLock* allows for the identification process to be initiated by many types of actions. For seamless and low-effort usage, this process can be triggered by automatically detecting actions such as the user picking or grabbing the intended device. Motions such as picking the device are common triggers in existing API for iOS and Android, enabling easy integration with our system. However, triggering acoustic sensing immediately after these actions may risk recording unnecessary audio, such as shuffling of items or accidental collisions while the user is moving their hand. To circumvent this, we ensure our system initiates sensing only when movement of the device has stabilized. When movement is not yet stable, the user may still be adjusting their grip on the device, leading us to obtain external sensor readings in order to monitor this behavior. For smartphones and tablets in particular, we can leverage motion sensors such as gyroscopes, accelerometers, and magnetometers to obtain a trace of the motion path. We can then compute variance of this trace across a sliding interval (e.g. every 0.1 seconds) and compare it with a pre-determined threshold to pinpoint when major hand or arm movements have subsided. This ensures minimal disturbances from non-acoustic sources.

7.3 Environmental Noise Removal

We capture our transmitted *n*-chirp sequence as an audio recording using on-board microphones, which may also contain interference in the form of ambient noise and high frequency distortions. Generally speaking, naturally occurring sound from daily activity such as movement or light conversation is unlikely to reach the inaudible frequency range used by *EchoLock*, however noise may still be introduced as a result of speaker imperfections, loud public spaces, or malicious actions by attackers. As a countermeasure, we filter our obtained recording prior to detailed signal processing. In particular, we design a band-pass filter with the pass band through 18kHz to 22kHz, which is the expected frequency range of the transmitted chirp signal. We apply a Butterworth low-pass and high-pass filter at the specified frequencies to achieve this. We use a third-order filter in order to minimize passband ripple in our signal amplitude and avoid distorting the biometric information embedded within.



(a) Nexus 5

(b) Galaxy Note 5

(c) Galaxy Tab A

(d) Example attack model setup.

Figure 9: Experimental setup for evaluating *EchoLock*. (a) through (c) show example holding styles and devices specifications, (d) depicts evaluation of the adversary model.

8 PERFORMANCE EVALUATION

We study the performance of *EchoLock* in a variety of common use case scenarios as well as on several mobile devices. Our experiments test the capability of our system to lock or unlock access to mobile devices as an example application, approved by our institute IRB. We present our findings and detailed analysis below.

8.1 Experimental Setup

Devices and Scenarios. A prototype application for *EchoLock* was developed for use on Android. Three smartphones, the *Nexus 5*, *Nexus 6*, *Galaxy Note 5*, and two tablet devices, the *Galaxy Tab A* and *Lenovo Tab 4* , were selected for their varied designs (e.g., speaker and microphone positions) and dimensions, pictured in Figure 9. The smartphone devices include two onboard microphones whereas the tablets are equipped with one. We evaluate our system in typical office and public scenarios. The *office scenarios* consist of quiet, enclosed spaces with minimal disturbances whereas the *public scenarios* are locations with large volumes of people and traffic. We maintain an average noise level of approximately 30dB and 60dB for the two environments, respectively. Sources of noise for the public environment include nearby conversations, walking, and dining. We gauge the ability of our system to accurately identify the user in face of these obstacles. We also investigate the impact of accessories that may transform the properties of the user's hand or device structure, such as gloves or smartphone cases. Additionally, we assess the viability for adversaries to compromise our system using various strategies. From these identified factors, we devised several scenarios to study.

Data Collection. 10 use case experiments were conducted in total, divided into 3 general categories. The first category examines the ability of our prototype to successfully identify the current user holding the device. The second category studies performance differences when used in a public environment. The third category considers usage via indirect physical contact, which includes when the user has equipped a protective case to their device and when the user is wearing a glove while holding their device. To reduce the number of factors at play, we confine our study to office environments unless otherwise noted. Mobile devices are provided on a table for the participants to pick up, hold, and place down. Although our system can support automatic data collection as described in Section 7.2, we initiate collection manually through the press of a button to eliminate the possibility of lost data due to undetected triggers. No specific instruction was provided on how to hold the device to encourage more natural interactions. However, we find that almost all participants favored holding postures similar to those shown in Figure 9. This is both beneficial and challenging as this adds consistency to our data set while also making it less simple to differentiate usage behaviors.

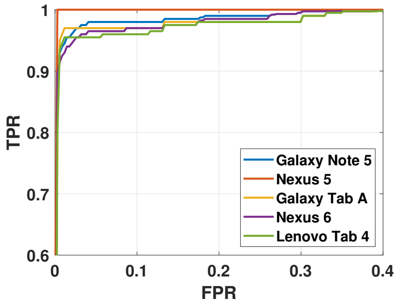
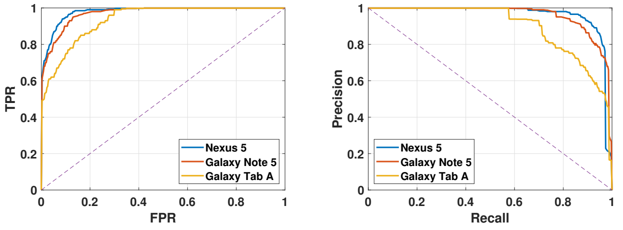


Figure 10: Overall performance for different mobile devices.

We recruit 20 volunteers, 14 males and 6 females ranging from ages 18-35, to participate in our study. We collect 40 n-chirp sequences, where n=10, for each test case based on the procedure in Section 6.4 for a total of 80,000 hand geometry samples. The profiles of all volunteers collectively act as the negative label during classification, with the exception of the target user undergoing identification. While this data is used for user identification purposes, we do not consider it to be personally identifiable information (as mentioned in Table 1). Nonetheless, we are currently maintaining this data privately and do not plan to release it publicly as a precaution.

8.2 Evaluation Metrics

We describe the accuracy of our system by evaluating the relation between *precision* and *recall* as well as usage of standard ROC curves. Precision is defined as the percentage of True Positive (TP) classifications out of all positive classifications recorded, notated as $P = \frac{TP}{TP+FP}$ where FP is the false positive rate. Recall is defined as $R = \frac{TP}{TP+FN}$ or the percentage of true positive classifications out of all target class instances. For our purposes, higher precision describes lower probability for different people to be mistaken for the legitimate user while higher recall describes the lower probability that the legitimate user is misidentified as someone else. The receiver operating characteristic (ROC) curve graphs the TP rate over the FP rate. The ideal system has a simultaneous 100% TP rate and 0% FP rate, i.e. all legitimate users are correctly identified while all attackers are denied access.



(a) ROC curve.

(b) Precision-Recall curve.

Figure 11: Performance under impersonation attacks.

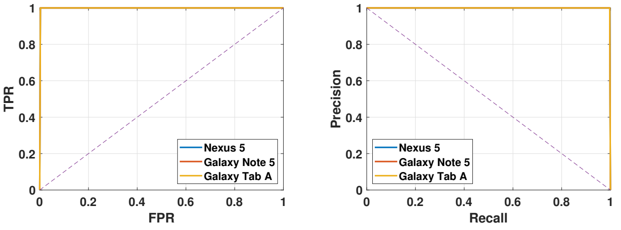
8.3 User Identification Performance

From our signal processing procedures, we obtain several features used to identify the user and present them to a machine learning classifier for identification. To evaluate the performance of our implementation, we consider Bagged Decision Trees (BDT), Linear Discriminant Analysis (LDA), K-Nearest Neighbor (KNN), and Support Vector Machines (SVM) as our candidate classifiers. From our initial comparisons of each classifier's ability to distinguish between 5 different users on a Nexus 5, we observe SVM utilizing a cubic kernel function to demonstrate the strongest performance. As a result, we present our findings for the SVM in our extended evaluations. We choose 10-fold cross-validation during the training process to best utilize our data set and minimize selective bias, allocating 50% for training and the remaining 50% for testing. Figure 10 illustrates the capability of our system to correctly identify the legitimate user, showing average TP rate of 94% for a 5% FP rate across a variety of different mobile devices. We detail the effects of adverse conditions and multiple impact factors in the following subsections.

8.4 Attacks on User Credentials

Impersonation Attacks. We evaluate the possibility of potential attackers to impersonate hand profiles of other users as a means of gaining unauthorized access to devices and information. For our assessment, we consider the worst case scenario; a limited (i.e. 5) training sample size for the victim user and multiple informed attackers. Our system is trained on user samples from 10 of our 20 participants, with one user acting as our victim and the remainder serving as negative labels. The 10 participants not involved in the training process act as attackers in our study. Attackers are allowed 30 seconds to observe the designated victim using our hardware platforms and given 10 attempts to imitate their hold. We conduct this process for 2 of our smartphones and 1 of our tablet devices. Observation of the victim is conducted from behind (i.e. shoulder surfing) and directly across (i.e. sitting in front).

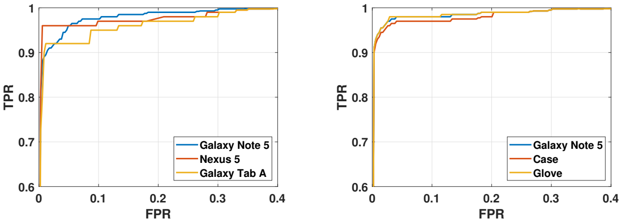
Our results suggest that visual observation alone is insufficient to prepare an attacker for impersonation of another hand profile. We measure the TPR and FPR, plotted as the ROC curve in Figure 11(a), showing FPR as low as 6% for a 90% TP rate for devices such as the Nexus 5. Shoulder surfing in particular was found to be unhelpful for the attackers as the victim hand is mostly obscured by the device. This would suggest that successful impersonation is dependent on physical similarity between the attacker and victim, which the attacker cannot control.



(a) ROC curve.

(b) Precision-Recall curve.

Figure 12: Eavesdropping and replay attack performance.



(a) Public environments.

(b) Different physical contact.

Figure 13: Performance in successfully identifying the legitimate user under various common circumstances.

Eavesdropping and Replay Attacks. We assessed the viability of eavesdropping information during our studies on standard usage behavior. Figure 9(d) shows an example hardware configuration during these experiments. The victim is seated at a desk and uses our system on a mobile device. A separate mobile device positioned 0.2m from the victim acts as a malicious sensor, listening for the validating signal. Many smartphones and tablets are able to activate their microphones without any obvious indicators onscreen, making this attack strategy highly plausible. We recruit 10 of our participants to act as 9 victims and 1 attacker. The profiles of the remaining 10 participants are used as negative labels during identification. We train our system on data from the 9 victims, though purposely exclude the attacker to simulate an attack by an unknown user. Each victim first authenticates themselves to generate a signal for the attacker to eavesdrop. The attacker then uses this signal to attempt to falsify their identity to our system.

Our experiments show *EchoLock* is able to correctly recognize each victim while also blocking the attacker when attempting to use an eavesdropped *n*-chirp sequence. Filtered signals recovered from these simulated attacks in an office setting showed recognizable chirp patterns, however clarity is lost due to the multipath effect. This is reflected in Figure 12, resulting in 0% false positive readings and perfect precision and recall. Recorded signals must travel through two airborne paths, once from the victim device to the eavesdropper and vice versa, causing significant attenuation and loss of genuine structure-borne properties. This level of attenuation is too severe for an attacker to compensate without intimate knowledge of the victim's authentic signal. We note that an attacker may attempt to deceive a naïve user into installing a malicious app in the mobile device, compromising the security of user identification techniques, including *EchoLock*. Protecting users from deception by attackers is a larger security problem to study, but beyond the scope of this paper.

8.5 Impact Factor Study

Impact of Device Models. We consider the performance discrepancies when operating *EchoLock* on different mobile devices. Our participants are provided devices with our prototype installed and given a short explanation on its functionality. We note that a single demonstration less than 10 seconds long is sufficient for our participants to grasp how to operate *EchoLock*, indicating its ease of use. We graph TP and FP in Figure 10 as a ROC curve for each of our mobile devices and observe that performance can be correlated to the size of the device used, as users more easily acclimate to a consistent posture for smaller devices compared to larger devices. We find an average TP of 94% for a fixed FP of 5%. In particular, we find our top performing device, the Nexus 5, capable of TP consistently exceeding 97%. For particularly large devices such as our tablets, the size difference between the hand and physical medium diminishes the impact of the holding posture. This suggests our measurements are most reliable when conducted on a small, well-defined space. We refer to these results as a benchmark for other experiments.

Impact of Environmental Noises. We study performance in public environmnets using a subset of our mobile devices. Our smartphone devices showed greater resilience under these conditions compared to our tablet device. This may be attributed to the lack of secondary microphones on tablet devices, which limits noise-cancellation capabilities compared to smartphone. Figure 13(a) indicates the introduction of significant noise produces measurable degradation, showing average TP decline ranging from 2% to 6% at a fixed 5% FP. We observe that the higher end of this degradation may be produced by loud vocalizations, such as shouts or laughter. We attribute this to our acoustic features, specifically our usage of MFCC. As MFCCs are most commonly used in speech processing, the presence of loud voices nearby causes our biometric measurements to be overwritten by the more dominant speech characteristics. This indicates that usage of our system may be challenging in certain situations, such as if the user is in the middle of a conversation.

Indirect Physical Contact. We also consider the influence of factors that may transform properties of either the device or user's hand when using *EchoLock*. To simulate these conditions, we equip our Galaxy Note 5 device with a smartphone case to change the physical properties of the medium. We also provide wool gloves roughly 2mm thick for the user to wear during separate sets of experiments. Figure 13(b) shows our results for usage during situations where the user's hand does not directly make contact with the mobile device. Our findings do not show statistically significant deterioration for these conditions tested. On the contrary, some users showed slightly improved accuracy ranging from 1-2%, particularly when wearing gloves. We suggest that the material of the gloves conformed well to the curvature of the hand while simultaneously suppressing variance in grip behavior, such as minor shifting.

N-Chirp Sequence Length. We investigate the effect of chirp sequence length on identification accuracy as a consideration to improve performance. While increasing the size of n used for sensing enhances accuracy, we observe an onset of diminishing returns rapidly after roughly 5 iterations. Stability around 90% can be maintained for n values as low as 3 in optimal test cases (i.e. smaller devices, quiet settings), shown in Figure 14.

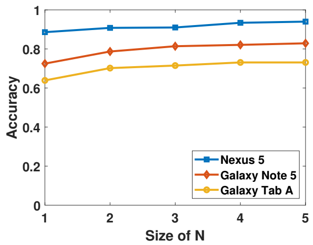


Figure 14: Performance over length of n-chirp sequence.

As mentioned in Section 5.2, a single chirp iteration requires only 25ms, or 50ms when accounting for buffering between iterations, meaning our current performance can be feasibly achieved in execution times competitive with techniques such as fingerprinting. This also indicates that an individual's hand biometrics are recognizable such that our machine learning framework may begin to identify them with moderate success using relatively few samples.

9 DISCUSSION

Jamming Attacks. We also consider the possibility of acoustic disruptions to our performance via jamming strategies. To do so, we study the ability for 10 users to use our system when an attacking device plays a continuous signal within the operational frequency of *EchoLock*. We play an attacking signal continuously sweeping from 18kHz to 22kHz during ten identification attempts per user at a distance of 0.2m. We find that detection of these jamming signals is feasible using a threshold scheme. We note that although detection is possible, negating the interference still poses a challenge. However, jamming attempts from distances greater than 1m were observed to lose potency, functioning more similar to public environment conditions described in Section 8.5. Methods to avoid threshold detection may attempt to limit the extent they exceed normal intensity, however this requires careful synchronization on the scale of milliseconds. Additionally, the jamming signal must match the length of the n-chirp sequence, which the attacker cannot predict. We are further improving resistance to these attacks as part of our future work.

Potential Hardware Constraints. During our selection of candidate devices to experiment on, we became aware of certain hardware configurations unsuitable to implement *EchoLock* on (e.g. speakers on front face, microphone on back). We find that our implementation requires our hardware components to be orientated such that they are as distant from each other as possible to allow for uninterrupted sound propagation, which cannot be accurately measured with a opposite-facing components. Adherence to this may be relaxed by leveraging additional sensor measurements to compensate for inconvenient speaker or microphone placement, though this also adds more hardware requirements to an intentionally minimalist design.

10 CONCLUSION

We have proposed *EchoLock*, a low-effort, and lightweight identification protocol deployable on commodity mobile devices. Our system verifies the user based on how the they hold their devices through a novel technique leveraging acoustic sensing of structure-borne sound to measure biometric characteristics. This technique can enable seamless identification checks or personalize user services when using smartphones, tablets, and similar devices. A prototype of *EchoLock* has been implemented on Android and evaluated in 160 trials of key use case scenarios, obtaining 80,000 hand geometry samples from multiple participants. Our technique is quick to conduct, low effort to use, and demonstrates accuracy over 94%. For future work, we intend to integrate *EchoLock* with existing authentication techniques and assess the possibility of elevating current security rates. We also intend to improve our defense against more sophisticated attack models and develop a more robust implementation to realize secure, low-effort identification.

ACKNOWLEDGMENT

This work was partially supported by the National Science Foundation Grants CNS1566455, CNS1826647, CNS1954959, CCF1909963, CCF2000480, CNS1801630, CNS1820624, and ARO Grant W911NF-18-1-0221.

REFERENCES

[1] 2019. Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025 (in billions). https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/.

[2] 2019. Material Sound Velocities. https://www.olympus-ims.com/en/ndt-tutorials/thickness-gage/appendices-velocities/.

[3] Abbas Acar, Hidayet Aksu, A. Selcuk Uluagac, and Kemal Akkaya. 2018. WACA: Wearable-Assisted Continuous Authentication. In *IEEE Symposium on Security and Privacy Workshops*.

[4] Amazon. 2018. Fire TV Stick. https://developer.amazon.com/docs/fire-tv/device-specifications-fire-tv-stick.html.

[5] Apple. 2018. Apple iOS. support.apple.com.

[6] Kaoru Ashihara. 2007. Hearing thresholds for pure tones above 16 kHz. *The Journal of the Acoustical Society of America* 122, 3 (2007).

[7] Silvio Barra, Maria De Marsico, Michele Nappi, Fabio Narducci, and Daniel Riccio. 2019. A hand-based biometric system in visible light for mobile environments. *Information Sciences* 479 (2019), 472–485.

[8] Todd Bishop. 2019. Amazon?s Blink unveils new security camera with ?exclusive? chip and two-year battery life. https://www.geekwire.com/2019/amazons-blink-unveils-new-security-camera-proprietary-chip-enables-two-year-battery-life/.

[9] Cam Bunton. 2016. Samsung Galaxy Note 7 iris scanner. https://www.pocket-lint.com/phones/news/samsung/138335-samsung-galaxy-note-7-iris-scanner-what-is-it-and-how-does-it-work.

[10] J. Guerra Casanova, C. Sánchez Ávila, A. de Santos Sierra, G. Bailador del Pozo, and V. Jara Vera. 2010. A Real-Time In-Air Signature Biometric Technique Using a Mobile Device Embedding an Accelerometer. In *Networked Digital Technologies*, Filip Zavoral, Jakub Yaghob, Pit Pichappan, and Eyas El-Qawasmeh (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 497–503.

[11] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)* 2, 3 (2011), 27.

[12] Ivan Cherapau, Ildar Muslukhov, Nalin Asanka, and Konstantin Beznosov. 2015. On the Impact of Touch ID on iPhone Passcodes.. In *Symposium on Usable Privacy and Security (SOUPS)*. 257–276.

[13] Hsin-Yi Chiang and Sonia Chiasson. 2013. Improving user authentication on mobile devices: a touchscreen graphical password. In *Proceedings of the 15th International Conference on Human-computer interaction with mobile devices and services*. MobileHCI.

[14] Sonia Chiasson, Paul C van Oorschot, and Robert Biddle. 2007. Graphical password authentication using cued click points. In *Computer Security–ESORICS 2007*. Springer, 359–374.

[15] Eric Chiu. 2017. Google's CEO Wants 30 dollar Smartphones For Developing Countries. https://www.ibtimes.com/googles-ceo-wants-30-smartphones-developing-countries-2471321/.

[16] Mohammed E. Fathy, Vishal M. Patel, and Rama Chellappa. 2015. Face-based Active Authentication on mobile devices. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*. IEEE.

[17] Jeremy Ford. 2011. 80 dollar Android Phone Sells Like Hotcakes in Kenya, the World Next? https://singularityhub.com/2011/08/16/80-android-phone-sells-like-hotcakes-in-kenya-the-world-next/.

[18] Google. 2019. Android Developer Resources. https://developer.android.com/reference/android/media/AudioRecord.html.

[19] Marian Harbach, Emanuel von Zezchwitz, Andreas Fichtner, Alexander De Luca, and Matthew Smith. 2014. It's a Hard Lock Life: A Field Study of Smartphone (Un)Locking Behavior and Risk Perception. In *Proceedings of the Tenth Symposium on Usable Privacy and Security (SOUP)*. SOUP, 213–224.

[20] R.C. Johnson, Walter J. Scheirer, and Terrance E. Boult. 2013. Secure voice-based authentication for mobile devices: vaulted voice verification. In *Proceedings of Biometric and Surveillance Technology for Human and Activity Identification*. SPIE.

[21] Sven Kratz and Md Tanvir Islam Aumi. 2014. AirAuth: a biometric authentication system using in-air hand gestures. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*. ACM, 499–502.

[22] Jian Liu, Hongbo Liu, Yingying Chen, Yan Wang, and Chen Wang. 2019. Wireless Sensing for Human Activity: A Survey. *IEEE Communications Surveys & Tutorials* (2019).

[23] Jian Liu, Chen Wang, Yingying Chen, and Nitesh Saxena. 2017. 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, 73–87.

[24] Beth Logan et al. 2000. Mel Frequency Cepstral Coefficients for Music Modeling.. In *ISMIR*, Vol. 270. 1–11.

[25] Andrew Martinik. 2018. How to customize Active Edge on the Google Pixel 3. https://www.androidcentral.com/how-customize-active-edge-pixel-3.

[26] Surbhi Mathur, Ankit Vjay, Jidnya Shah, Shreyasi Das, and Adil Malla. 2016. Methodology for partial fingerprint enrollment and authentiation on mobile devices. In *Proceedings of the International Conference on Biometrics*. IEEE.

[27] Meinard Müller, Frank Kurth, and Michael Clausen. 2005. Audio Matching via Chroma-Based Statistical Features.. In *ISMIR*, Vol. 2005. 6th.

[28] Yanzhi Ren, Yingying Chen, Mooi Choo Chuah, and Jie Yang. 2014. User Verification Leveraging Gait Recognition For Smartphone Enabled Mobile Healthcare Systems. *IEEE Transactions on Mobile Computing* (2014).

[29] Yanzhi Ren, Chen Wang, Yingying Chen, Mooi Choo Chuah, and Jie Yang. 2015. Critical segment based real-time e-signature for securing mobile transactions. In *2015 IEEE Conference on Communications and Network Security (CNS)*. IEEE, 7–15.

[30] Jan Rychlewski. 1984. On Hooke's law. *Journal of Applied Mathematics and Mechanics* 48, 3 (1984), 303–314.

[31] Napa Sae-Bae, Kowsar Ahmed, Katherine Isbister, and Nasir Memon. 2012. Biometric-rich Gestures: A Novel Approach to Authentication on Multi-touch Devices. In *Proceedings of ACM SIGCHI*.

[32] Muhammad Shahzad, Alex X Liu, and Arjmand Samuel. 2013. Secure unlocking of mobile touch screen devices by simple gestures: You can see it but you can not do it. In *ACM MobiCom*. 39–50.

[33] Ke Sun, Ting Zhao, Wei Wang, and Lei Xie. 2018. 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*. 591–605.

[34] Xiaoyuan Suo, Ying Zhu, and G Scott Owen. 2005. Graphical passwords: A survey. In *Proceedings of the 21st Annual Computer Security Applications Conference*. IEEE.

[35] TSYS. 2016. 2016 U.S. Consumer Payment Study. https://www.tsys.com/Assets/TSYS/downloads/rs_2016-us-consumer-payment-study.pdf.

[36] Yu-Chih Tung and Kang G. Shin. 2015. EchoTag: Accurate Infrastructure-Free Indoor Location Tagging with Smartphones. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*. 525–536.

[37] Yu-Chih Tung and Kang G. Shin. 2016. 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*. 277–289.

[38] Sebastian Uellenbeck, Markus Dürmuth, Christopher Wolf, and Thorsten Holz. 2013. Quantifying the security of graphical passwords: the case of android unlock patterns. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*. 161–172.

[39] Dirk Van Bruggen, Shu Liu, Mitch Kajzer, Aaron Striegel, Charles R. Crowell, and D'Arcy John. 2013. Modifying Smartphone User Locking Behavior. In *Proceedings of the Ninth Symposium on Usable Privacy and Security (SOUP)*. SOUP, 213–224.

[40] Chen Wang, Yan Wang, Yingying Chen, Hongbo Liu, and Jian Liu. 2020. User authentication on mobile devices: Approaches, threats and trends. *Computer Networks* 170 (2020), 107–118. https://doi.org/10.1016/j.comnet.2020.107118

[41] WeChat. 2017. Voiceprint. https://thenextweb.com/apps/2015/03/25/wechat-on-ios-now-lets-you-log-in-using-just-your-voice/.

[42] Nan Zheng, Kun Bai, Hai Huang, and Haining Wang. 2014. You Are How You Touch: User Verification on Smartphones via Tapping Behaviors.. In *ICNP*, Vol. 14. 221–232.

[43] Yu Zhong and Yunbin Deng. 2014. Sensor orientation invariant mobile gait biometrics. In *Proceedings of the IEEE International Joint Conferene on Biometrics*. IEEE.

[44] Bing Zhou, Jay Lohokare, Ruipeng Gao, and Fan Ye. 2018. EchoPrint: Two-factor Authentication using Acoustics and Vision on Smartphones. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*. 321–336.

[45] Man Zhou, Qian Wang, Jingxiao Yang, Qi Li, Feng Xiao, Zhibo Wang, and Xiaofeng Chen. 2018. PatternListener: Cracking Android Pattern Lock Using Acoustic Signals. In *ACM Conference on Computer and Communications Security*.