Virtual Step PIN Pad: Towards Foot-input Authentication Using Geophones

Yan He, Hanyan Zhang, Edwin Yang, and Song Fang
University of Oklahoma, Norman OK 73019, USA
{heyan, hanyanzhang, edwiny, songf}@ou.edu

***Abstract*—The use of personal identification numbers (PINs) for authentication is ubiquitous due to their simplicity and flexibility. In this work, we present *virtual step PIN pad*, a novel and practical PIN entry scheme that allows a user to enter a PIN through foot tapping on the ground. The virtual step PIN pad utilizes geophones to collect structural vibration signals caused by foot tapping. When a user generates the activation signals by performing a predetermined sequence of foot taps within the target area, the virtual step PIN pad will be launched, and takes the foot tapping input by the user. The system then demodulates the corresponding structural vibration signals into a PIN. We have developed a prototype of the virtual step PIN pad and conduct a suite of experiments to evaluate its practicality and security. Experimental results show that the virtual step PIN pad can achieve an average success rate of 96.5% for inputting a human-chosen 4-digit PIN. Meanwhile, the success rate for an adversary at a distance of more than 2.5 meters away from the PIN pad to infer the target PIN decreases to below 3%.**

***Index Terms*—structural vibration, foot tapping, authentication**

I. Introduction

The personal identification numbers (PINs) are sequences of digits utilized to authenticate user identity to computer systems. A great amount of building entrances and payment terminals apply either traditional physical button based or touchscreen based PIN pads so that users can type in PINs with their fingers for getting access to the buildings or finishing the payment transactions. Finger-input based authentication systems, however, do not fully satisfy the needs of all the population. For example,

- Touchscreen based systems usually do not react to a wet (sweaty) finger while almost 3% of the general population in the United States experience hyperhidrosis or excessive sweating of the palms [1].
- It is often difficult or even impossible for a user with hand or finger injuries to temporarily or permanently press a physical button or operate on a touchscreen. There are 45,000 amputations annually due to traumatic injuries to the hands and fingers in the United States [2].

Finger-input based authentication systems obviously impose a practical hurdle for those users.

On the other hand, finger-input based methods may bring unexpected privacy disclosure risks. To input a PIN, we normally unconsciously or have to let our finger skin directly contact with the physical or on-screen PIN pad, and this process may disclose our fingerprint and thus jeopardize all applications using fingerprints. It has been shown that touchscreen PIN entry systems may suffer from smudge attacks [3] and also oily residues left by tapping fingers on a touchscreen enable attackers to reveal fingerprints [4]. The two aforementioned limitations with finger-input based authentication systems motivate us to develop another scheme that works for people including who cannot use fingers for typing PINs and also mitigates privacy disclosure risks.

Besides PINs, there are emerging biometric authentication systems, which determine the identity of a person by comparing a biometric data capture (e.g., fingerprint [5], iris [6], face [7], voice [8]) to authentic biometric information stored in a pre-collected database. However, the deployment of such systems is expensive as it requires both the corresponding dedicated user interface and a database including biometric information of all users. Meanwhile, if the system discloses such sensitive private biometric information, it may cause serious consequences. [9]–[11] propose to utilize eye gaze interaction to enter a PIN, i.e., correlate the position of the user's gaze on screen to a digit input. However, such methods usually not only need to pre-deploy an on-screen keyboard before inputting the PIN, but also require a camera to monitor the user's eye movement. If the captured eye movement by the camera is unclear (e.g., in low light conditions), the accuracy of the entered PIN would decrease significantly.

Other than traditional input ways such as using finger or eye gaze, is there another way for humans to input PINs that is convenient and secure? Intuitively, we consider to utilize foot for PIN entry as people walk with their feet every day. Among all foot gestures, we select foot tapping (i.e., raising and lowering the toes or heel) as the input method since it requires comparatively low effort and is also inattentive, which can be mistaken for natural walking [12]. Then the question becomes how to convert the foot taps on ground into inputted digits. A geophone [13] is a device that translates ground movement (i.e., the velocity of a monitored surface) into voltage, which can be easily read by a microcontroller. Recent work [14], [15] have shown that the geophone sensor can be utilized to measure footstep-induced structural vibrations. In this paper, we then design and implement a foot-input based authentication scheme leveraging geophones. We refer to this technique as *virtual step PIN pad*. Specifically, a user first initializes the virtual step PIN pad on the ground with specific activation signals, and then taps his foot on the PIN pad to enter a PIN. Foot tapping at different locations will

generate different structural vibrations which can be captured by geophones and then mapped into corresponding digits.

Unlike finger-input based methods that may disclose fingerprints, the foot-input based method does not have such privacy breach concerns as people usually wear shoes and shoe-prints are not unique [16]. Also, since there is no actual physical PIN pad and foot taps are usually regarded as habitual behaviors or leg workouts, foot-input based PIN entry has better concealment and is not easy to be destroyed. Nevertheless, multiple challenges need to be addressed in order to make the proposed method to work accurately in practice.

First, the deployed system and the target user should agree on an exact location of the virtual step PIN pad so that the identified PIN for authentication matches the PIN entered by the user with foot tapping. We design a customized PIN pad activation method to determine the PIN pad layout. The system is then ready to demodulate the following PIN input.

Second, the system needs to distinguish the foot taps in different areas of the virtual step PIN pad in order to correctly translate each foot tap into a digit. We extract the structural vibration signals corresponding to the foot taps for entering a PIN, and propose an inter-peak interval based method to correlate each observed vibration signal pattern with a digit.

Third, an important question is about the security of the proposed scheme, mainly from two aspects. On concern is whether an adversary can identify PINs used for authentication by deploying a geophone-based sensing system around. In general, as the foot tapping induced waves travel outward, the energy that they contain becomes dissipated. Therefore, the vibration becomes weaker the further it is from the source. As a result, an adversary may be unable to capture signals with enough energy to infer the tapping locations if the distance between herself and the tapping location exceeds a certain threshold. Another concern is whether the authentication scheme still works when an advanced attacker obtains the legitimate user's PIN (e.g., via secretly videotaping the foot tapping process). We also propose to utilize the legitimate user's vibration characteristics to further refuse access launched by unauthorized users even when they obtain the correct PINs. We perform real-world experiments to explore the performance of the virtual step PIN pad against such attacks.

We point out the virtual step PIN pad does not aim to replace existing PIN input methods. Same with finger-input based schemes, the proposed scheme cannot work for all the population as well. For example, a person who has difficulties (e.g., paraplegia) in performing foot tapping may not be able to use the virtual step PIN pad. Instead, the proposed scheme is positioned as an alternative input way. It can be complementary to existing finger-input based techniques, and make the PIN authentication system serve for a greater population.

II. Foot Tapping Identification

Geophones are easy to install and can be installed anywhere on the ground. As they capture the ambient structural vibration signals, they thus can be utilized to monitor a person's heart rate and respiratory rate [17], perceive human fall [18], and detect the physical events within a house [19].

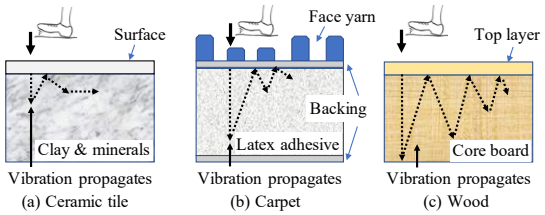


Fig. 1. A foot tapping on different materials.

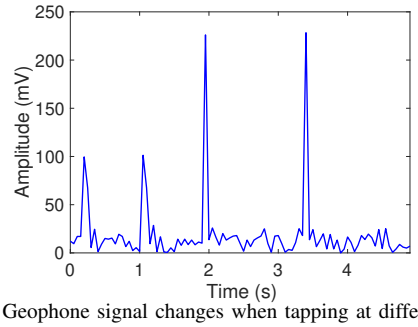


Fig. 2. Geophone signal changes when tapping at different locations.

Ground Structure Impact: Different floor materials with varying vibration absorption ability (i.e., damping capability [20]) may bring different impact for geophone observations induced by foot tapping. Figure 1 gives an example to illustrate how the foot tapping induced vibration may propagate within different floor materials. Empirically, we observe that the damping capability decreases in accordance with ceramic tile, carpet and wood. Specifically,

- Ceramic tile has high stiffness and is rigidly fastened to the subfloor. A floor with ceramic tiles shows high damping capability and would quickly absorb the vibration.
- A carpet is a textile floor covering consisting of four layers from top to bottom, i.e., face yarn, primary backing, latex adhesive and secondary backing. A foot tap on a carpet usually brings minute deformation of the face yarn and the adhesive binder. Meanwhile, the vibration wave would be absorbed after experiencing a few reflections.
- A wood floor normally includes a top layer and a core board. The reflected vibration wave would gradually decay over time.

We detailedly explore the impact of floor material on the performance of the virtual step PIN pad in Section IV-D.

Signal Sensing: We apply a commercial off-the-shelf geophone, SM-24 [13], with a sensitivity of 28.8V/m/s and a natural frequency of 10Hz. The geophone voltage output is too weak to be captured by the built-in Analog to Digital converter (ADC) of Arduino motherboard. Thus we use a 16-bit ADC to amplify the geophone output and observe amplified signals.

Figure 2 shows a stream of geophone records when tapping at two locations that are 30cm apart, each for two times. We observe that when there is no tapping, the air-wave or appliance noise induced vibration amplitude is quite small (less than 4mV), while a foot tap generates a distinguishable vibration wave with larger amplitudes. Also, it can be clearly seen that foot taps at the same location generate highly sim-

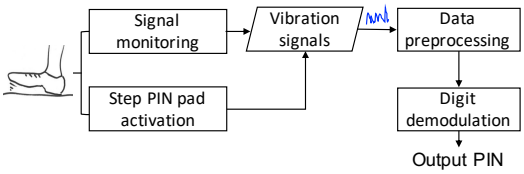


Fig. 3. Virtual step PIN pad schematic.

ilar vibration waves, while varying tapping locations lead to different geophone signals. This observation can be exploited for tapping location inference. The vibration waves propagate outward in all directions from the tapping location through the floor material. They may undergo reflection or refraction at structural boundaries and even dispersion. We refer to such a phenomenon as *multipath effect*. Thus a foot tap may induce multiple vibration signals and a geophone captures the superposition of them. In Section III-E, we take advantage of this multipath effect to develop a technique to distinguish foot taps in different areas on the virtual step PIN pad.

III. Virtual Step PIN Pad Design

A. Adversary Model

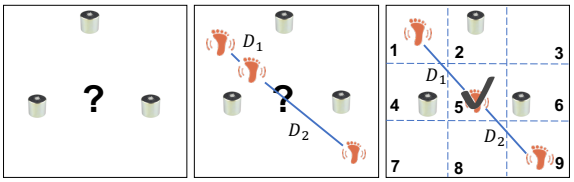
Finger-input based authentication schemes often suffer from shoulder surfing attacks, which acquire the pressed keys by peeping over the victim's shoulder. The PIN in the proposed scheme may also be compromised if an attacker can observe the foot tapping process. However, as aforementioned, the virtual step PIN pad has good concealment due to inconspicuous foot tapping gesture and no tangible PIN pad. We assume if the PIN is solely utilized for authentication, the user can protect the foot tapping (e.g., setting up barriers to cover the foot movement) during the PIN entry process from being directly observed or videotaped by others. Also, when such a barrier is not available, we utilize pre-registered vibration profiles to further characterize users. Correspondingly, the two attacks below are considered to evaluate any PIN authentication schemes (e.g., [21], [22]), including ours.

Side-channel Attack: The adversary aims to sniff PINs by deploying hidden geophones (e.g., behind some regular object) on the floor near the target to capture the structural vibration signals generated during the PIN entry process.

Knowledgeable Observer Attack: The adversary is able to stealthily observe the legitimate user's foot movement through shoulder surfing or secret video recording and thus infer the target PIN. The adversary then tries to imitate the behavior of the legitimate user to bypass the authentication.

B. System Overview

The basic idea underlying the proposed system is to achieve PIN recognition by analyzing unique location-dependent features from the received structural vibration signals caused by foot tapping. The geophones located on the ground continuously monitor the vibration signals. When a user inputs the activation signals with foot tapping on the target area, the PIN pad will be activated and the system prepares to take the following vibration signals and translate them into PIN input.



(a) Standby mode (b) Failed attempt (c) Successful activation

Fig. 4. PIN pad activation. (a) No foot tap is detected; (b) Since $|D_1{-}D_2| > \delta$, the activation fails; (c) With $D_1 > D_{min}$, $D_2 > D_{min}$, and $|D_1 - D_2| \leq \delta$, the PIN pad is launched.

The system then pre-processes the collected vibration signals and identifies the specific features for each foot tap. Finally, each identified feature will be correlated to a respective location on the PIN pad, i.e., a corresponding digit. The output would be the entered PIN for authentication. Figure 3 shows the flowchart of the virtual step PIN pad system.

C. Activation

Unlike traditional finger-input based PIN pads, whose layout is known to the user, the proposed scheme does not have a physical PIN pad. Instead, we design a virtual PIN pad that enables the user to utilize foot tapping for PIN entry, and such a PIN pad can be deployed anywhere on the ground as long as the structural vibration signals caused by foot tapping can be captured by the geophones. First of all, in order to enter a PIN, a user need to activate the virtual step PIN pad.

The activation signal should not only let the system prepare to demodulate the following foot taps into a PIN, but also confirm the PIN pad location for both the user and the system. Correspondingly, the activation signal should not be an activity that a user exhibits in daily walking (e.g., a single foot tap), otherwise the PIN pad may be frequently and falsely activated. Meanwhile, the location of the PIN pad should enable the user to perform foot tapping without spending too many efforts (e.g., turning around). To achieve both goals, we enable the user to first tap once at three spots in a line, and the interval distances between two successive spots are D_1 and D_2. Initially, the system is in standby mode: monitoring received signals and waiting for the virtual step PIN pad activation signals, as shown in Figure 4(a).

We design the criteria of a successful activation as: (1) the time period between two successive taps does not exceed P, which is used to help clean invalid activation signals (we set $P = 2$ seconds); (2) $D_i > D_{min}$ ($i \in \{1, 2\}$) and $|D_1{-}D_2| \leq \delta$, where D_{min} denotes the designed minimum interval distance and δ is a pre-determined threshold that can be adjusted based on the PIN pad size and user preference. D_{min} is usually larger than the foot length as it can help the user to easily distinguish the locations on different rows or columns. As shown in Figure 4(b), the system will be still in standby mode with a failed activation, while a successful activation will launch the virtual step PIN pad, as shown in Figure 4(c).

After the PIN pad is activated, the system enters the *ready* phase: it collects the following structural vibration signals and inputs them into the next module.

D. Data Preprocessing

Like pressing digit buttons on a physical PIN pad with a finger, a user performs a sequence of foot taps with a leg on the virtual step PIN pad for PIN entry. We assume that the user taps once for each digit with a normal speed (1∼2 sec per tap). Data preprocessing phase removes the noise from the structural vibration signals observed by geophones and segments the time series of the structural vibration signals into individual samples, each corresponding to a foot tap.

1) Noise Reduction: It is observed that the variations of geophone observations caused by foot tapping lie at the low end of the spectrum. To preserve valuable signal components and mitigate the noise introduced by environmental vibrations or hardware imperfection, we employ the Fast Fourier Transform (FFT) filter [23]. In particular, we first perform FFT on geophone observations to detect the dominant frequency range corresponding to foot tapping, and then set amplitudes of other parts as zero, and finally use Inverse FFT (IFFT) to the remaining frequency data to recover the denoised signal.

2) Two-threshold Based Segmentation: The segmentation phase separates the structural vibration signals for individual foot tap. The vibration amplitudes during tapping periods usually show a much larger variance than those happening during non-tapping periods. Motivated by this, we then focus on identifying the segments with the variance which is larger than a predetermined *variance threshold*. Meanwhile, some other actions such as dropping an object, may also lead the system to observe a vibration amplitude that exceeds the threshold. To separate foot tapping with other inferences, we further utilize cross-correlation method, which is widely employed to quantify the similarity of two waveforms [24].

Cross-correlation: Let s_1, s_2 denote two signal waveforms, each of which is represented with an n-point discrete time series. The cross-correlation $r_{s_1,s_2}(l)$ can be calculated by a function of the lag $l \in [0, n-1]$ applied to s_2, i.e., $\sum_{i=0}^{T-1} s_1(i) \cdot s_2(i-l)$, where $s_2(i) = 0$ if $i \leq 0$. To accommodate for different amplitudes of the two series, the cross-correlation can be normalized as $r'_{s_1 s_s}(l) = \frac{r_{s_1,s_2}(l)}{\sqrt{r_{s_1,s_1}(0) \cdot r_{s_2,s_2}(0)}}$. To quantify the similarity between two structural vibration signals, we derive the largest absolute value of cross-correlation, i.e., $\max_l(|r'_{s_1 s_s}(l)|)$, which lies in the range of $[0, 1]$, with 1 indicating perfect correlation, and 0 showing uncorrelation.

With multiple foot tapping induced signals during an empirical profile, we calculate the cross-correlation between each pair of the signals, and select the minimum value as the *similarity threshold* to help detect legitimate foot tapping induced signals during the authentication phase. Specifically, we first search over the data for the segments with the variance under the variance threshold. Each of such a segment denotes an inter-vibration interval. Everything between two successive inter-vibration intervals will be regarded as a potential vibration signal induced by a single foot tap. Next, we compute the cross-correlation between each of such signals with a pre-obtained foot tapping induced signal. If the obtained cross-correlation is above the similarity threshold, we believe that this structural vibration signal is caused by foot tapping, otherwise, it would be discarded as an interference signal.

E. Digit Demodulation

Digit demodulation converts foot taps to corresponding digits. We observe when we tap our foot at different locations, the time interval between the first two local maximum amplitudes in the observed wave signal varies. We refer to such a feature as inter-peak interval, and use it to distinguish foot tapping at different locations on the PIN pad. We divide the virtual step PIN pad into 3×3 subareas (i.e., keys). Neighboring subareas are connected with a (horizontal or vertical) *gap area*.

In modern keyboards, the key gap design can impact the typing performance [25]. Similarly, we also design a gap area between adjacent subareas to improve the accuracy of the input via foot tapping. The gap area serves two functions. First, it can help decrease authentication errors. If neighboring subareas are together without a gap in between, when a foot tapping is performed very close to the edge next to other keys, such an input may be easily misidentified as the other keys. This is because for foot tapping at locations approaching the edges connecting neighboring subareas, the induced vibration signals would be quite similar. Second, a gap area with a moderate width (i.e., distance between edges of subareas) enables the user to better distinguish different keys. If there is no key gap or the key gap is too small, it may increase the probability that a user accidentally taps an adjacent key at the same time as the target key. On the other hand, the gap should not be too wide as well, otherwise the user may frequently tap at the gap area, which generates invalid input.

Each subarea is associated with a characteristic inter-peak interval vector $\bar{\mathbf{t}}_i = [\bar{t}_{i1}, \bar{t}_{i2}, \bar{t}_{i3}]$ ($i \in \{1, 2, \cdots, 9\}$), where $\bar{t}_{ij}$ ($j \in \{1, 2, 3\}$) denotes the standard inter-peak interval obtained by the j-th geophone when the source of structural vibration lies in the i-th subarea. Meanwhile, the maximum variation of the observed inter-peak interval compared with its standard value (i.e., $\bar{t}_{ij}$) can be denoted with δ_{ij}. Let the vector $\mathbf{\Delta}_{i_{max}}$ denote $[\delta_{i1}, \delta_{i2}, \delta_{i3}]$.

After the virtual step PIN pad is activated, the digit that each subarea represents is determined. Let D_i represent the inputted digit when foot tapping happens in the i-th subarea. Then when the user enters a digit with foot tapping, each of the three geophones observes a vibration wave. Next, the system performs the following steps to translate it into a digit. Initially, the index i of the subarea is 1.

(a) For the collected signal by the j-th geophone, we search the first two largest peaks and calculate the inter-peak interval t_{ij}. We thus obtain $\mathbf{t}_i = [t_{i1}, t_{i2}, t_{i3}]$.
(b) If $i < 10$, we calculate $\mathbf{\Delta} = |\mathbf{t}_i - \bar{\mathbf{t}}_i|$ and compare it with $\mathbf{\Delta}_{i_{max}}$; otherwise return with invalid input.
(c) If all elements of the vector $\mathbf{\Delta}_{i_{max}} - \mathbf{\Delta}$ are positive, this foot tap will be converted into the digit D_i; otherwise, enable $i = i+1$ and jump to step (b).

The proposed inter-peak interval based method will apply to each foot tap. As a result, an N-digit PIN is entered for authentication with N foot taps. Note that the characteristic

inter-peak interval vector $\bar{\mathbf{t}}_i$ and the corresponding maximum variation vector $\mathbf{\Delta}_{i_{max}}$ can be obtained through an empirical profile. Specifically, we divide each subarea into smaller sub-subareas. We then let the user tap her foot at each sub-subarea, and collect the structural vibration signals through the three geophones. With each geophone's data, we calculate the inter-peak interval for each foot tap and utilize the average value of all inter-peak intervals to denote the characteristic value observed with this geophone. After that, the maximum variation of the inter-peak interval compared with the characteristic value can be calculated. Normally, with more sub-subareas, we can have more fine-grained reference signals, and may generate more accurate results.

Digit '0' Input and Identification: We propose to input the digit '0' via a non-tapping way. Specifically, after the activation phase, the system maintains a timer for each foot tap. The maximum timer interval is set to be 2 seconds, which empirically provides a user with enough time to tap once. If a foot tap is detected before the timer runs out, the system immediately resets the timer for detecting the next input. While if there is no foot tap detected within the maximum timer interval, we regard this silent period as the input of digit '0'. The system then resets the timer for detecting the next one.

F. Dealing With Knowledgeable Observer Attacks

A concern is that whether the virtual step PIN pad still works when the attacker is able to directly observe the legitimate user's foot movement via peeping or videotaping. Apparently, it is no longer enough for distinguishing different users only by PINs. Instead, after the virtual step PIN pad detects that the inputted PIN is correct, the system would then take advantage of the pre-built user-specific feature extracted from the structural vibration signals to determine whether the foot tapping is performed by the registered PIN owner.

Specifically, during the training phase, the system takes the observed structural vibration signals as input when the user enters PINs via foot tapping, then runs the supervised learning algorithm SVM (Support Vector Machine) by LIBSVM library [26] to classify the features. Different user's foot tapping preference varies due to different personal characteristics including personal weight, foot size, leg length and foot tapping habit. Such preference difference will be represented by the distribution of the typed PIN by the regression analysis of SVM training process. The SVM classification results show the personal features unique to each user, and will be used in the testing phase to determine whether a PIN input is initiated by the owner of the PIN or somebody else. It is highly difficult for an adversary to generate a feature that matches with the legitimate user's as the feature integrates both PIN and the user's behavior and physical characteristics.

IV. Experimental Evaluation

A. Evaluation Setup

We build a prototype of the virtual step PIN pad on top of three SM-24 geophones [13] to demonstrate its performance. The system has two controllers, referred to as A and B. As shown in Figure 5, Controller A includes two geophones (G1 and G2), and both connect with one same ADS1115 module, i.e., a 16-bit analog-to-digital converter (ADC) with a programmable-gain amplifier, on an Arduino MEGA 2560 [27], which is connected to a computer for data processing. As it cannot distinguish foot taps coming from both sides of the line between two geophones based on the signals only collected from Controller A, Controller B is utilized to help determine the direction of each foot tap. Figure 6 shows Controller B, consisting of the third geophone (G3), whose output is fed into ADS1115 on an Arduino UNO [28]. In order to synchronize data on both Controllers, we equip each with a 2.4G wireless transceiver module – NRF24L01 [29], considering that wired connection between Controllers may bring inference for the geophones' output.

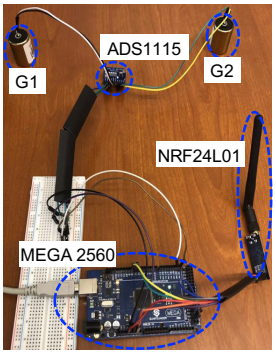


Fig. 5. Controller A.

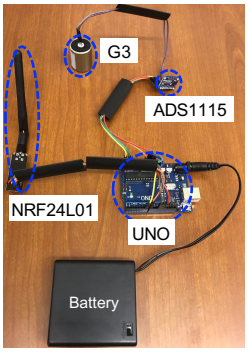


Fig. 6. Controller B.

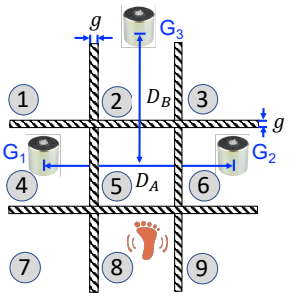


Fig. 7. Deployment of geophones.

The position of three geophones relative to the virtual step PIN pad layout is as shown in Figure 7. The PIN pad follows the typical design of a door lock or payment terminal PIN pad, and consists of a number pad with the digits from 1 to 9. For inputting digit '0', we propose a method without tapping, as described in Section III-E. We enable G1 and G2 at the centers of respective key areas on the second row and set the key pitch (i.e., center-to-center distance between keys) of $D_A = 40$cm. We design a gap area with a width $g = 4$cm empirically, which can help the user effectively perceive two neighboring key areas. The distance between G3 and the line of G1 and G2 is represented with D_B. In order to help distinguish foot tapping at different rows, we then let G3 lie at the mid-perpendicular of the line of G1 and G2, and meanwhile put G3 at exactly above the first row of the PIN pad for the best performance. Thus, we have $D_B = (3/4) \cdot (D_A - 2g) + g$.

Intuitively, the accuracy of the proposed scheme may be affected by the setting of PIN length, deployment of geophones,

TABLE I
CONFUSION MATRIX FOR SINGLE DIGIT INPUT.

Digit Inputted \ Digit Detected	1	2	3	4	5	6	7	8	9
1	0.96	0.00	0.00	0.03	0.00	0.00	0.01	0.00	0.00
2	0.01	0.97	0.02	0.00	0.01	0.00	0.00	0.00	0.00
3	0.01	0.01	0.96	0.00	0.01	0.01	0.00	0.00	0.00
4	0.02	0.00	0.00	0.97	0.00	0.00	0.01	0.00	0.00
5	0.00	0.00	0.00	0.00	0.99	0.00	0.00	0.01	0.00
6	0.00	0.00	0.03	0.00	0.00	0.95	0.00	0.00	0.02
7	0.00	0.00	0.00	0.04	0.00	0.00	0.95	0.01	0.00
8	0.00	0.00	0.00	0.00	0.01	0.00	0.02	0.94	0.03
9	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.98

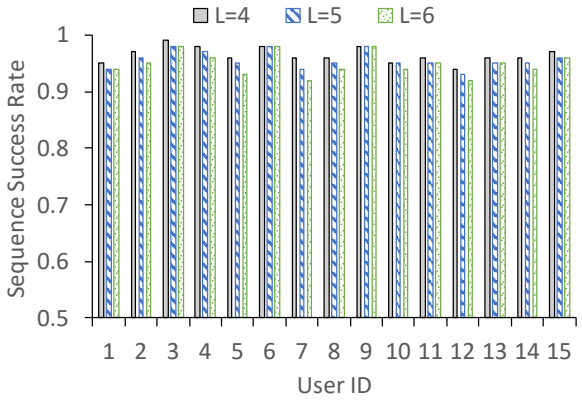


Fig. 8. Sequence success rate vs. sequence length.

and floor material. Next, we explore how these factors can impact the performance of the virtual step PIN pad.

B. Verification Accuracy

We recruited 15 volunteers (6 of them are female) to test the performance of the virtual step PIN pad deployed on a wood floor, as wood flooring is durable and one of the popular floor coverings. In Section IV-D, we will explore the impact of the floor material on the performance of the proposed system. We define the *digit success rate* as the rate of successfully recognizing a single digit. We employ this metric to ascertain the digit verification accuracy of the virtual step PIN pad. To demonstrate the complete digit sequence verification accuracy, we compare the obtained digit sequence by the system with the one inputted by the user to determine whether the digit sequence identification is successful, and calculate the *sequence success rate*, which equals to the ratio between the number of successful digit sequence identification times and the total number of digit sequence entry trials.

1) Digit Success Rate: We let each participant input each digit 100 times via foot tapping. Table I shows the confusion matrix for digits from 1 to 9. It demonstrates that the mean digit success rate of all digits is 96.3%, and the lowest digit success rate is 94%. Meanwhile, we observe that though digit '5' borders the most other digits, its digit success rate is the highest (i.e., 99%). Digit '5' lies in the middle of geophones G1 and G2, which have similar observations when inputting digit '5', enabling the system to determine the candidates: digits '2', '5' and '8'. Meanwhile, the geophone G3 can easily distinguish these three digits as their distances to G3 are different. Thus, the system can further shrink the search space from three digits into one, i.e., digit '5' is correctly verified. Besides, for inputting digit '0' with our customized method, a digit success rate of 100% is always achieved. This accuracy demonstrates the virtual step PIN pad's ability to utilize structural vibration signals to extract inputted digits.

2) Sequence Success Rate: Since no real-life dataset of PINs has ever been publicly available, we utilize the dataset of 4-digit sequences, extracted from 32 million *Rockyou* passwords [30]. Such 4-digit PINs can be utilized to approximate user choices of PINs [31]. We randomly select 100 4-digit PINs from the dataset and let each user input them through the virtual step PIN pad one by one. Meanwhile, the PIN length usually varies from 4 to 6 for achieving the balance of security and usability according to ISO 9564 standard [32]. Thus, to explore the relationship between the sequence success rate and the length of the inputted digit sequence, we also randomly generate 5- and 6-digit PINs for 100 times, respectively, and then input each to the virtual step PIN pad via foot tapping.

Figure 8 shows the sequence success rate when each of the fifteen users inputs digit sequences with different lengths. We observe that the proposed virtual step PIN pad authentication scheme can retain high accuracy (i.e., 93.0% or above) across different users, and meanwhile with the length of digit sequence increasing from 4 to 6, the sequence success rate slightly decreases. Specifically, for inputting a 4-digit PIN, the authentication system can achieve an average sequence success rate of 96.5%, while for inputting a sequence of 5 or 6 digits, the corresponding average sequence success rate becomes 96.0% or 95.2%, respectively.

C. Impact of Inter-Geophone Distance

Different people may have different step length, and thus tap at different distances with one foot away from the opposite foot. The virtual step PIN pad may adjust to such variations. Specifically, when the distance D_A between G1 and G2 (i.e., the key pitch) changes, the PIN pad size changes correspondingly. We then change D_A from 40cm to 30cm and 50cm. Meanwhile, D_B is also changed with D_A. The gap width stays the same. We repeat the above experiments calculating the digit success rate and sequence success rate when D_A varies.

Figure 9 shows the average digit success rate for all digits under different key pitches. We observe that for all pitches, the proposed scheme can always achieve a high average digit success rate (over 90%), and the average digit success rate for each digit has a small fluctuation of 4% or 5%. Besides, for most digits, when $D_A = 40$cm, the average digit success rate is the highest, and the corresponding digit success rate for $D_A = 30$cm or $D_A = 50$cm slightly decreases.

Figure 10 shows the sequence success rates for different users to input 6-digit PINs across three key pitches. We see that the sequence success rates for all pitches are always above 90%. Similarly, when the pitch decreases to 30cm or increases to 50cm from 40cm, the corresponding sequence success rate

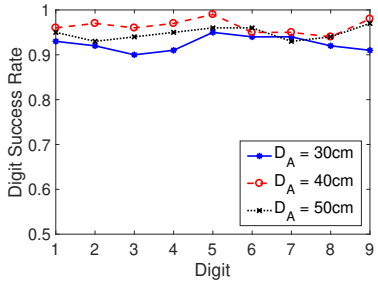


Fig. 9. Average digit success rate vs. key pitch.

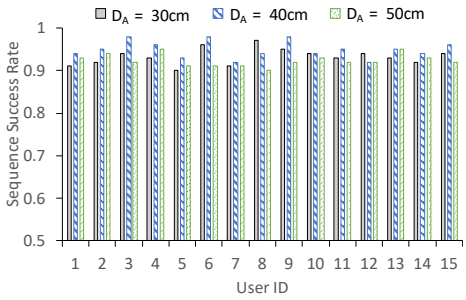


Fig. 10. 6-digit sequence success rate vs. key pitch.

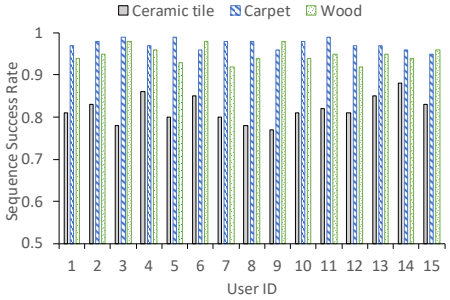


Fig. 11. 6-digit sequence success rate vs. material.

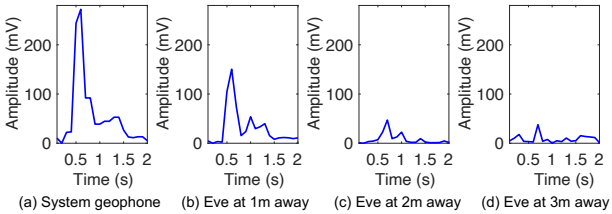


Fig. 12. Foot tapping records at one geophone of the virtual step PIN pad and adversary geophones at different distances away from the tapping location.

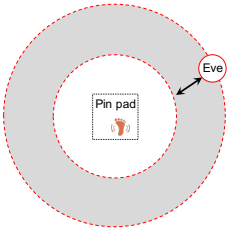


Fig. 13. Adversary (Eve) setup.

drops slightly. The average 6-digit sequence success rates across all users when D_A equals to 30cm, 40cm and 50cm are 93.2%, 94.9%, and 93.0%, respectively. We conclude that an appropriate PIN pad size would help the proposed scheme to achieve the best performance. A smaller or bigger PIN pad, on the contrary, requires the user to tap at a further or closer location for a same digit, and thus may cause the user to accidentally tap in a wrong PIN area, which does not correspond to the digit that the user intends to input.

D. Impact of Floor Material

As mentioned earlier, different floor materials have different structure and vibration absorption ability. To examine the impact of floor material on the verification accuracy of the proposed system, we also perform authentication experiments on another two popular materials, i.e., ceramic tile and carpet. Specifically, with 100 6-digit sequences generated in Section IV-B2, we let each user input each one via foot tapping.

Figure 11 shows the 6-digit sequence success rates for different users across three tested floor materials. We can see that for both carpet and wood, the sequence success rate is consistently high (in a range of 93% to 99%), while for ceramic tile, the users obtain relatively lower sequence success rates, ranging from 78% to 88%. Ceramic tile usually has a large damping capability, which may cause the geophones to capture weak vibrations and thus be unable to correctly distinguish foot tapping at different digit areas. As a result, the corresponding sequence success rate is decreased. Besides, we observe that carpet achieves a slightly higher sequence success rate than wood. This is because the reflected vibration wave has less effect on the geophone measurements for carpet as the carpet absorbs vibrations more quickly than wood.

E. Attack Scenarios

We evaluate the robustness of the virtual step PIN pad under different types of attacks. Specifically, for each round, 1 of 15 participants is alternatively taken as the legitimate user and the rest 14 participants play as attackers.

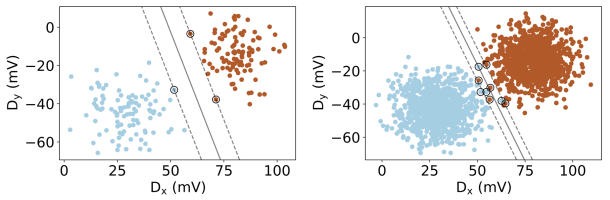
1) Side-channel attack: An adversary (Eve) may deploy a well-camouflaged three-geophone-based system on the ground around the PIN pad and thus infer inputted PINs by analyzing the captured structural vibration signals. We assume that Eve cannot be placed within 1m from the center of the PIN pad, as in this case the exposure risk would be dramatically increased.

We first explore how far a foot tap induced vibration wave can propagate. Correspondingly, we let the user perform foot tapping on the PIN pad as usual with uniform force, and put three adversary geophones at the distances of 1m, 2m, and 3m from the tapping location, respectively. Figure 12 presents an example of records at different geophones for a foot tap. We observe that with a longer distance between the geophone and the tapping location, the geophone obtains a lower signal amplitude. When the adversary geophone is 3 meters away from the geophone, the structural vibration signal almost submerges in noisy background.

Next, we measure the sequence success rates for Eve at varying distances away from the PIN pad to infer the inputted 4- to 6-digit PINs generated in Section IV-B2. As shown in Figure 13, we draw a circle originating at the center of the PIN pad and place Eve at a radius ranging outward from 1m to 3.5m, every 0.5m. For each radius, we enable the user to input each selected PIN. Eve performs PIN inference based on the captured structural vibration signals. For comparison, we also calculate the corresponding sequence success rate of the virtual step PIN pad system. Table II shows the average sequence success rate of the proposed system and Eve at different distances away from the target PIN pad. We can see with the distance or PIN length increasing, the sequence success rate at the adversary drops. Specifically, when the distance reaches 3.5m, the sequence success rate at the adversary for inferring a 4- or 5-digit PIN lowers to 0%, while such a

TABLE II
SEQUENCE SUCCESS RATES AT THE PROPOSED SYSTEM AND EVE.

PIN length	System	1m	1.5m	2m	2.5m	3m	3.5m
4	0.97	0.63	0.09	0.05	0.03	0.02	0
5	0.96	0.52	0.06	0.04	0.02	0.01	0
6	0.94	0.47	0.05	0.02	0.01	0	0



(a) Running 200 foot taps on SVM (b) Running 2000 foot taps on SVM

Fig. 14. Result of running SVM on foot taping induced vibration signals.

distance decreases to 3m for a 6-digit PIN. Meanwhile, we observe that the sequence success rate of the virtual step PIN pad always maintains a high value of 94% or above.

2) Knowledgeable observer attack: We utilize the data obtained from geophones G1 and G2 as two features (referred to as D_x and D_y) inputted to SVM. Figure 14 shows the result for analyzing the features for a total of 200 and 2,000 foot taps performed by two different participants when they input the same PINs. We can observe that the features for the users can be separated by a hyperspace. Also, with the amount of foot taps increasing, we can still distinguish different users with the SVM hyperspace. In each round, all users are asked to input a PIN randomly selected by the legitimate user for 10 times. Figure 15 presents the mean sequence success rates for the legitimate user and all attackers to bypass the authentication with different PIN lengths. We can see that the attackers have a significantly low sequence success rate (e.g., less than 0.9% for breaking a 6-digit PIN) while the legitimate user always maintains a sequence success rate as high as over 95.0%, convincingly verifying the effectiveness of the proposed technique against knowledgeable observer attacks.

V. RELATED WORK

In this section, we review two domains of prior works, closely related to the proposed virtual step PIN pad technique.

A. PIN Authentication Schemes

PIN authentication is employed in a wide variety of applications, such as automated teller machine (ATM) and point of sale (POS) transactions, room access and smartphone unlocking. Nowadays, increasing physical PIN pads are replaced with touchscreen ones [33], which are user-friendly and also save space for computer systems. Generally, existing PIN authentication schemes mainly fall into the following categories:

Finger-input based methods: Physical touch or press via fingers is currently a mainstream way of inputting PINs. However, such approaches may not be available or cause inconvenience for the population who have certain disabilities or illnesses that disable them from using fingers. Meanwhile, when the finger skin contacts a touchscreen during the PIN input, the fingerprint may be stolen by an adversary with

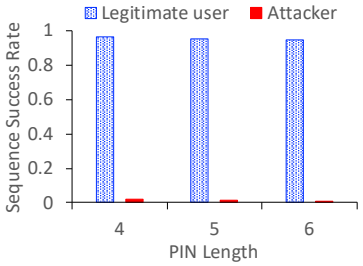


Fig. 15. Mean sequence success rates for the legitimate user and the attacker.

techniques such as smudge attack [3]. [21] develops an authentication system leveraging physical vibration, which enables a user to give finger input on ubiquitous surfaces and can thus defend against the smudge attacks. However, the system has to be reconfigured each time the nearby environment varies.

Eye gaze tracking based methods: It has been long proved that eye gaze tracking can be used for PIN entry [9]–[11], [34], [35]. Those techniques, however, may suffer from so called Midas Touch problem [36], when each fixation on an interface that a user is looking at may become selected even when the user has no such intention. Commercial products, such as Tobii eyes trackers [37], can achieve high accuracy and thus push the pedal on eye gaze tracking based schemes. For example, [11] proposes an eye tracking technology with a dedicated eye tracker hardware that can achieve a high authentication accuracy of around 95% for a 4-digit PIN. Besides, [35] implements an eye gaze based typing system which only takes advantage of common cameras. Nonetheless, the process of live video frames requires strong computing power and also the video recording may violate the user privacy.

Vision-based methods: Vision-based perceptual user interfaces (e.g., [38]–[40]) can be another way to type, especially with the development of computer vision. Facial or head gestures (e.g., month open, brows up) can be utilized to control on-screen mouse printer or virtual keyboard. However, similar with eye gaze based methods which leverage cameras, vision-based methods need to process recorded videos and thus bring privacy concerns, and also if the user's head does not happen in the presence of a camera, the input cannot be detected.

B. Vibration Detection via Geophones

Geophones are sensitive devices and have been widely applied to detect structural vibration signals induced by micro-seismic events (i.e., micro-earthquakes that are typically too small to be felt on the surface [41]), animal behaviors [42], as well as human activity such as walking [15], fall [18], breathing [17], and in-bed motions [43]. Geophones can also detect the displacement of bridge structures by analyzing vibration signals induced by the live load of the bridge [44]. Our work utilizes geophones not only to detect human activity (i.e., foot tapping), but also to further identify the tapping locations, each of which corresponds to a PIN digit input.

VI. CONCLUSION

We propose the virtual step PIN pad technique which maps each foot tapping location into a digit leveraging the observed structural vibration signals induced by foot tapping.

A customized PIN pad activation method is utilized to enable both the proposed system and the user to agree on an identical PIN pad layout and area on the ground. Also, an inter-peak interval based approach is applied to correlate the observed structural vibration signal pattern to the foot tapping location on the PIN pad. Extensive experimental results demonstrate that the proposed technique can achieve an average success rate of 96.5% for inputting a human-chosen 4-digit PIN, whereas the success rate for an adversary to infer the inputted PIN lowers to less than 3% when its distance away from the PIN pad is more than 2.5 meters. Also, by integrating personal vibration profile, the system still rejects the attacker's access with a high probability even when she obtains the correct PIN.

ACKNOWLEDGEMENT

This work is supported in part by the National Science Foundation under Grant No.1948547.

REFERENCES

[1] A. Haider and N. Solish, "Focal hyperhidrosis: diagnosis and management," *Canadian Medical Association Journal (CMAJ)*, vol. 172, no. 1, pp. 69–75, 2005.

[2] D. B. Reid, K. N. Shah, A. E. Eltorai, C. C. Got, and A. H. Daniels, "Epidemiology of finger amputations in the united states from 1997 to 2016," *Journal of Hand Surgery Global Online*, vol. 1, no. 2, pp. 45 – 51, 2019.

[3] A. J. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. M. Smith, "Smudge attacks on smartphone touch screens," in *Proc. of the 4th USENIX Conf. on Offensive Technologies*, WOOT'10, pp. 1–7, 2010.

[4] Y. Zhang, P. Xia, J. Luo, Z. Ling, B. Liu, and X. Fu, "Fingerprint attack against touch-enabled devices," in *ACM Workshop on Security and Privacy in Smartphones and Mobile Devices*, pp. 57–68, 2012.

[5] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar, *Handbook of Fingerprint Recognition*. Springer Publishing Company, Incorporated, 2nd ed., 2009.

[6] J. Daugman, "How iris recognition works," in *The Essential Guide to Image Processing*, pp. 715 – 739, 2009.

[7] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.

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

[9] M. Kumar, T. Garfinkel, D. Boneh, and T. Winograd, "Reducing shoulder-surfing by using gaze-based password entry," in *Proc. of the 3rd Symp. on Usable Privacy and Security (SOUPS)*, pp. 13–19, 2007.

[10] A. De Luca, R. Weiss, and H. Drewes, "Evaluation of eye-gaze interaction methods for security enhanced pin-entry," in *Proc. of the 19th Australasian Conf. on Computer-Human Interaction: Entertaining User Interfaces*, OZCHI '07, pp. 199–202, 2007.

[11] M. Seetharama, V. Paelke, and C. Röcker, "Safetypin: Secure pin entry through eye tracking," in *Proc. of the 3rd Int. Conf. on Human Aspects of Info. Security, Privacy, and Trust*, pp. 426–435, 2015.

[12] E. Velloso, D. Schmidt, J. Alexander, H. Gellersen, and A. Bulling, "The feet in human–computer interaction: A survey of foot-based interaction," *ACM Comput. Surv.*, vol. 48, pp. 21:1–21:35, Sept. 2015.

[13] "Geophone - SM-24." https://www.sparkfun.com/products/11744, 2019.

[14] S. Pan, N. Wang, Y. Qian, I. Velibeyoglu, H. Y. Noh, and P. Zhang, "Indoor person identification through footstep induced structural vibration," in *HotMobile*, pp. 81–86, 2015.

[15] S. Pan, T. Yu, M. Mirshekari, J. Fagert, A. Bonde, O. J. Mengshoel, H. Y. Noh, and P. Zhang, "Footprintid: Indoor pedestrian identification through ambient structural vibration sensing," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 1, pp. 89:1–89:31, Sept. 2017.

[16] I. Rida, S. Bakshi, X. Chang, and H. Proença, "Forensic shoe-print identification: A brief survey," *ArXiv*, vol. abs/1901.01431, 2019.

[17] Z. Jia, A. Bonde, S. Li, C. Xu, J. Wang, Y. Zhang, R. E. Howard, and P. Zhang, "Monitoring a person's heart rate and respiratory rate on a shared bed using geophones," in *ACM SenSys*, pp. 6:1–6:14, 2017.

[18] Y. Huang, W. Chen, H. Chen, L. Wang, and K. Wu, "G-fall: Device-free and training-free fall detection with geophones," in *IEEE SECON*, pp. 1–9, June 2019.

[19] J. Han, A. J. Chung, M. K. Sinha, M. Harishankar, S. Pan, H. Y. Noh, P. Zhang, and P. Tague, "Do you feel what i hear? enabling autonomous iot device pairing using different sensor types," in *2018 IEEE Symposium on Security and Privacy (SP)*, pp. 836–852, May 2018.

[20] J. Zhang, R. J. Perez, and E. J. Lavernia, "Documentation of damping capacity of metallic, ceramic and metal-matrix composite materials," *Journal of Materials Science*, vol. 28, pp. 2395–2404, May 1993.

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

[22] W. Chen, L. Chen, Y. Huang, X. Zhang, L. Wang, R. Ruby, and K. Wu, "Taprint: Secure text input for commodity smart wristbands," in *ACM MobiCom*, 2019.

[23] X. Liu, Z. Zhou, W. Diao, Z. Li, and K. Zhang, "When good becomes evil: Keystroke inference with smartwatch," in *ACM CCS*, 2015.

[24] C. Yoon, O. O'Reilly, K. Bergen, and G. Beroza, "Earthquake detection through computationally efficient similarity search," *Science Advances*, vol. 1, 12 2015.

[25] H. Madison, A. Pereira, M. Korshøj, L. Taylor, A. Barr, and D. Rempel, "Mind the gap: The effect of keyboard key gap and pitch on typing speed, accuracy, and usability, part 3," *Human factors*, vol. 57, 2015.

[26] C.-C. Chang and C.-J. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, May 2011.

[27] "Arduino mega 2560." https://store.arduino.cc/usa/mega-2560-r3, 2019.

[28] "Arduino uno." https://store.arduino.cc/usa/arduino-uno-rev3, 2019.

[29] "nRF24 series." https://www.nordicsemi.com/Products/Low-power-short-range-wireless/nRF24-series, 2019.

[30] "Passwords." https://wiki.skullsecurity.org/Passwords, 2020.

[31] J. Bonneau, S. Preibusch, and R. Anderson, "A birthday present every eleven wallets? the security of customer-chosen banking pins," in *Financial Cryptography and Data Security*, pp. 25–40, 2012.

[32] "ISO 9564-1:2017 financial services – personal identification number (PIN) management and security." https://www.iso.org/standard/68669.html, 2020.

[33] N. Henze, E. Rukzio, and S. Boll, "Observational and experimental investigation of typing behaviour using virtual keyboards for mobile devices," in *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, CHI '12, p. 2659–2668, 2012.

[34] M. Kumar, A. Paepcke, and T. Winograd, "Eyepoint: Practical pointing and selection using gaze and keyboard," in *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, CHI '07, p. 421–430, 2007.

[35] Z. Li, M. Li, P. Mohapatra, J. Han, and S. Chen, "iType: Using eye gaze to enhance typing privacy," in *IEEE INFOCOM*, pp. 1–9, 2017.

[36] R. J. K. Jacob, "The use of eye movements in human-computer interaction techniques: What you look at is what you get," *ACM Trans. Inf. Syst.*, vol. 9, p. 152–169, Apr. 1991.

[37] "Tobii Pro." https://www.tobiipro.com/product-listing/, 2020.

[38] M. Betke, J. Gips, and P. Fleming, "The camera mouse: visual tracking of body features to provide computer access for people with severe disabilities," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 10, no. 1, pp. 1–10, 2002.

[39] Y. Gizatdinova, O. Špakov, and V. Surakka, "Face typing: Vision-based perceptual interface for hands-free text entry with a scrollable virtual keyboard," in *IEEE Workshop on the Applications of Computer Vision*, pp. 81–87, 2012.

[40] Y. Yan, C. Yu, X. Yi, and Y. Shi, "Headgesture: Hands-free input approach leveraging head movements for hmd devices," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, Dec. 2018.

[41] P. Folger and M. Tiemann, "Human-induced earthquakes from deep-well injection: A brief overview," Congressional Research Service, 2015.

[42] B. Mortimer, W. L. Rees, P. Koelemeijer, and T. Nissen-Meyer, "Classifying elephant behaviour through seismic vibrations," *Current Biology*, vol. 28, pp. R547–R548, 2018.

[43] M. Alaziz, Z. Jia, M. Aldeer, R. Howard, and Y. Zhang, "Motionphone: a wireless geophone-based in-bed body motion detection and classification system," in *1st Int. Conf. on Cybernetics and Intelligent System*, 2019.

[44] M. Micheloni and M. Pieraccini, "Bridge monitoring using geophones: Test and comparison with interferometric radar," in *Proc. of the 13th Int. Conf. on Damage Assessment of Structures*, pp. 25–34, 2019.