Enabling Secure Touch-to-Access Device Pairing based on Human Body's Electrical Response

Yao Wang
Xidian University
Xi'an, China
wangyao@xidian.edu.cn

Tao Gu
Macquarie University
Sydney, Australia
tao.gu@mq.edu.au

Yu Zhang
Macquarie University
Sydney, Australia
y.zhang@mq.edu.au

Minjie Lyu
Xidian University
Xi'an, China
mjlv@xidian.edu.cn

Tom H. Luan
Xidian University
Xi'an, China
tom.luan@xidian.edu.cn

Hui Li
Xidian University
Xi'an, China
lihui@mail.xidian.edu.cn

ABSTRACT

Recent efforts in reducing user involvement during device pairing have successfully introduced *touch-to-access*. To detect whether two devices are being held by the same person, existing *touch-to-access* solutions extract features from a shared information source to generate pairing keys. They focus on validating the device's authenticity by only requiring the user's simple touching of the device, however, ignore the device holder's legitimacy and pairing intent. Moreover, the pairing keys may be vulnerable to eavesdropping attacks since they are exchanged over an open wireless link (e.g., WiFi or Bluetooth). In this paper, we develop a secure device pairing mechanism that essentially uses the human body to generate and transmit user-specific pairing keys, ensuring the user's legitimacy and pairing intent, as well as improving key transmission reliability. Our work is based on the observation that the human body produces a unique response to the electrical signal flowing through it, and different bodies induce distinct responses to the signal. The built-in microphone on devices captures ambient sound as an entropy source and converts it into an electrical signal, which is subsequently processed and transmitted by the human body for device pairing. We build a prototype using off-the-shelf microphones and conduct extensive experiments with 31 participants to evaluate its security performance and usability. The results show that our system achieves a pairing success rate of 97.74% and an equal error rate of 2.28%.

CCS CONCEPTS

• **Security and privacy → Biometrics**; • **Human-centered computing → Ubiquitous computing**.

KEYWORDS

Device pairing, ambient sound, body electrical response

ACM MobiCom '22, October 17-21, 2022, Sydney, Australia

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

ACM Reference Format:
Yao Wang, Tao Gu, Yu Zhang, Minjie Lyu, Tom H. Luan, and Hui Li. 2022. Enabling Secure Touch-to-Access Device Pairing based on Human Body's Electrical Response. In *The 28th Annual International Conference on Mobile Computing and Networking (ACM MobiCom '22), October 17-21, 2022, Sydney, Australia.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

The growing popularity of wearable devices has spawned many applications in health, sport, and fitness, making life more convenient for people. To provide a better user experience, application data are often required to exchange between wearables or synchronize with proximal personal devices in real-time. As the information transmitted in these applications is often sensitive and private, placing a secure pairing procedure between devices is of great importance.

Traditional device pairing typically requires either explicit key input or complicated peer-to-peer protocols for key exchange. The most common method for key input is password entry, i.e., active participation in entering the correct password manually on small screens is required. Such key input approaches are inconvenient, prone to human error, and generally incompetent for wearables due to poor user interface. Key exchange-based protocols (e.g., pre-shared key [49] and Diffie-Hellman [26]) require a public key infrastructure that is computation-intensive, hence they are technically challenging to operate on resource-constrained wearables.

Recent efforts have been made based on the principle of *touch-to-access* to mitigate the limitations imposed by traditional methods. The basic idea is that it is permissible to pair two devices if they simultaneously have direct physical contact with the same human body. During each pairing session, both devices acquire consistent measurements from a dynamic information source, allowing them to agree on the same secret key on-the-fly. To do this, some studies extract time-varying biosignals from the human body such as ECG [34, 50], EMG [21, 46], and heartbeat interval [22], and quantify the measurements into binary sequences to form a symmetric key agreement protocol. However, these solutions may give rise to a practical issue as biosignals are generally captured by dedicated sensors, and it is impractical to expect all wearable devices have a common sensing capability.

To improve usability, many *touch-to-access* schemes rely on readily available sensors on devices. For example, Yan *et al.* [45] and

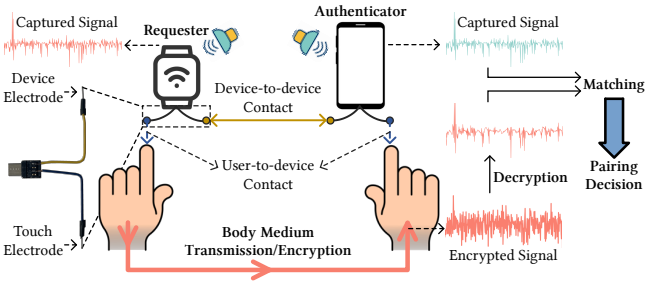


Figure 1: A bilateral *touch-to-access* scheme using the human body for pairing key generation and transmission.

Jin *et al.* [16] employ analog-to-digital converter (ADC) and RF transceiver, respectively, to sense the electric potentials induced by ambient electromagnetic radiations on human body. These systems work specifically based on the fact that human body is electrically conductive, and the electric potentials measured on the same human body are similar whereas those from different bodies are distinct. While these studies provide usable and effortless solutions for device pairing, they present several critical security concerns.

1) We experimentally discover from existing systems using the human body as a conductor [16, 25, 33, 39, 45] that two users are likely to have similar measurements if they achieve equipotential by physically contacting with each other (e.g., handshake) as discussed in Section 4.2. This may create a new vulnerability in which the user's on-body device may be paired with a malicious device carried by a premeditated attacker through physical contact in a densely-populated context, such as subway, airport, and stadium. Additionally, any mistouch from the user may induce an unintended connection between devices. We conclude that the crux for these issues is rooted in their **unilateral touching modality**, i.e., only requiring a simple user contact with the device without further pairing intent confirmation.

2) Most of the *touch-to-access* approaches make an assumption that device pairing is performed by an authorized user, and this essentially implies that anyone who can touch the device will gain access without the need for identification. This **non-user-specific** nature makes these systems vulnerable in real-life scenarios where the device is lost/stolen or temporarily left somewhere by the owner.

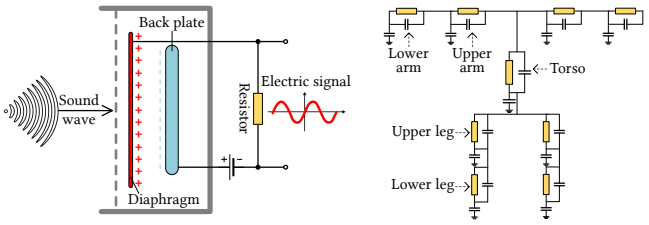
3) To enable both devices to reach an agreement on the same secret key, the established keys are required to be exchanged over an open wireless link, e.g., WiFi or Bluetooth. However, key distribution between devices over a **public** wireless medium is susceptible to eavesdropping and jamming attacks [37].

The security concerns of existing approaches motivate us to develop a more secure *touch-to-access* scheme. We move a step further from human body as a conductor and exploit the human body as a **private** medium to generate and transmit **user-specific** pairing keys. The basic idea of our design is illustrated in Fig. 1. To address the issues caused by the user's unilateral touching of the device, we propose a **bilateral touching** solution, i.e., at the time that the user is in contact with the device, device-to-device contact is also required. The connection between devices essentially serves as a confirmation of the user's pairing intent, thus eliminating unintended pairings triggered by physical contact and mistouch. For key generation and transmission, we utilize the built-in microphone (i.e., commonly available on wearable devices) to capture pervasive ambient sound as an entropy source and convert it into an electrical signal. Then the human body spontaneously encrypts the captured signal from the requester as a pairing key and simultaneously transmits it to the authenticator. Finally, the pairing key is decrypted by a user-related decryption function and compared with the authenticator-captured signal to verify if both devices have direct physical contact with the same legitimate user's body. Our system is based on the key observation that the human body interferes with the electrical signal flowing through it, and different human bodies induce distinct interferences to the signal. The underpinning hypothesis is that the human body can be viewed as a composite conductor composed of resistors and capacitors, which exhibits different conductivity properties due to its diverse physiological structures [23, 25, 39].

To achieve better usability in our system design, we use ambient sound as the entropy source due to its wide accessibility in the real world. Even in a quiet context, users can still perform pairing by proactively making sounds, such as playing audio or speaking. In contrast, gait or gesture-based approaches are hardly applicable to individuals with hand or foot disabilities and those that hinge on electromagnetic radiations [16, 45] are susceptible to denial of service (DoS) when radiations are shielded by strong attackers [6]. Moreover, the touching electrodes required for the system are easy to construct. For one example, metal materials on smart devices (e.g., the metal case, frame, knob, and button) might be designed to be touching electrodes with some engineering efforts. For another, we can also devise a small accessory to serve as the touching electrodes for devices having a charging port as illustrated in Fig. 1, where we show a USB Type-C electrode interface as an example.

In designing an efficient pairing scheme, we have several technical challenges to be addressed. 1) How to synchronize the encryption signal with the authenticator-captured signal for ensuring their consistency? To address this issue, we look for the touch-active segment by examining the variance of the body-transmitted signal in a sliding window and then use this active segment to clip the authenticator-captured signal. 2) How to decrypt the body-encrypted signals for the final comparison? We answer this question by designing a finite impulse response (FIR)-based decryption model to estimate the user's body channel characteristics. Prior to use, users perform a calibration to determine their specific decryption function and store it on the authenticator side. 3) How to effectively distinguish legitimate pairing requests from attacks for maintaining the system's robustness? For this purpose, we train a machine learning-based classifier to make the pairing decision instead of manually defining a threshold that decides whether the similarity result of the signals is a match or non-match. It is critical to declare that the decision classifier is trained offline on a single training set and only once; users need not perform any initial training for classifier construction. In summary, the main contributions of this paper are as follows.

1) We propose a practical and secure device pairing scheme by using human body's electrical response to encrypt omnipresent ambient sounds for key generation and design a decryption model by estimating the body channel characteristics. In contrast to previous studies that do not require user identity confirmation, the user-specific nature of our scheme enhances the security guarantees.



(a) Microphone working principle (b) Human body lumped model

Figure 2: Diagram of the basic principles.

2) We experimentally discover a common security flaw in existing device pairing schemes that use the human body as a conductor, and consequently propose a bilateral touching strategy to mitigate such risk. This solution essentially provides user pairing intent confirmation and can be easily applied to other *touch-to-access* schemes.

3) We develop a prototype and conduct extensive experiments with 31 participants to evaluate the security and usability of our scheme. Results show that our system achieves a pairing success rate of 97.74% and an equal error rate of 2.28%, and is resilient to various attacks.

2 BACKGROUND AND THREAT MODEL

2.1 Microphone

Microphones are transducers that convert sounds into an electrical signal via electromagnetic induction. The most common microphones for commercial use, from wearables to home assistants, are electret condenser microphones (ECMs) and micro-electro-mechanical system (MEMS) microphones. Both types of microphones work on similar principles, they basically have a capacitor that consists of two conductive plates near each other as illustrated in Fig. 2(a). One of the plates is made of a flexible material and serves as a mechanical diaphragm. The diaphragm vibrates in the presence of sound waves, changing the distance between the plates, which varies the capacitance. Since the total amount of electric charge the capacitor holds is constant, the capacitance change leads to a change in voltage across the capacitor which in turn results in a variable electric current to flow [28]. In this way, mechanical sound waves are converted into electrical signals for further processing.

2.2 User-Distinct Electrical Response

The human body is a composite material comprised of liquid water with ions (e.g., sodium, chloride, and potassium) dissolved in them. The aqueous electrolytes have the tendency to conduct electricity and this makes the body a conductor of electricity [30]. As suggested by [33, 51], the human body can be modeled by multiple electronic component units, as shown in Fig. 2(b). Each unit is approximated using a parallel connection of a capacitor and a resistor, as well as coupling capacitance to the ground. Due to the complex and diverse physiological structure of the human body (e.g., differences in bone structure, body water content, fat mass, muscle density, etc.), the parameters for these units vary from person to person, resulting in different degrees of effect on the electrical signal. Existing studies have confirmed that the human body's electrical responses have unique biometric information owing to differences in body structure [23, 25, 39]. Thus, the inherent electrical conductivity properties of the human body can be regarded as a unique biometric used for key generation in this work.

Our pairing scheme relies on a key observation that the electric current received from different parts of the body exhibits the same variation pattern for the same person, and it varies significantly across individuals. Compared with the intra-body communication (IBC)-based approaches, for example, [33] uses the body as a transmission medium to exchange generated keys, performing a symmetric key agreement protocol on both devices; we consider the human body as an encryption key generator. The objectives and principles of our work are distinct from those approaches. In addition, we argue that the approaches are not feasible for wearables since they require either specific transceivers or computing power to manage intricate key agreement protocols.

2.3 Threat Model

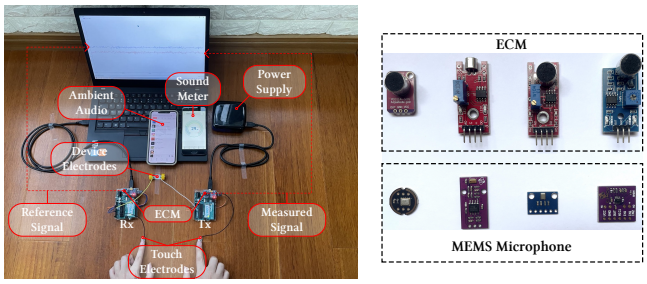
We envision an adversary exploring the existing studies for approaches to breach the security of the system. We assume that the system is secure and that an adversary can neither tamper with the matching mechanism nor steal the user-private decryption function. The adversary can observe how the legitimate user interacts with the device, including touch position and duration, and can even record the ambient sound while the user is pairing the device. Besides, we leverage ambient sound as the entropy source to produce pairing keys, which may consequently bring other security concerns evidenced by [35, 48]. We argue that the threats presented by [48] and [35] primarily target voice assistants on smart devices, while our study focuses on securing the pairing process. The threat vectors are orthogonal to the task of our work and beyond the scope of our study. To ensure the reliability of our system, we mainly consider the following attack scenarios:

Random Attack An adversary randomly touches our system in the hope that the arbitrary touching events can produce similar impacts on the system as the legitimate user does and completes the pairing procedure.

Imitation Attack An adversary observes the legitimate user's touching behavior (i.e., the body position touching the electrode and touching duration) as well as records the ambient sound when the user is performing pairing. The adversary plays the recorded audio and imitates the user's touching behavior, trying to generate the same pairing keys.

Synthesis Attack We assume an adversary is aware that the pairing key is generated by the body channel interference with the sound signal captured by a microphone. The adversary can be more advanced, surreptitiously capturing the ambient sound when the user is pairing and then corrupting the captured signal with noise (e.g., white Gaussian noise). The intention is to synthesize similar distorted signals as the legitimate user's body channel does.

In contrast to wireless-based pairing schemes which expend considerable efforts to ward off man-in-the-middle (MITM) attacks [10, 16, 27, 33, 44–46], our approach is intrinsically immune to it. This is because our solution requires an additional device-to-device contact to function as a confirmation of the user's pairing intent, any strange device access to the user's legitimate device can be easily discerned by the user; whereas wireless channel is vulnerable



(a) System prototype (b) Tested microphones

Figure 3: Measurement setup.

to eavesdropping and modifying for the reason that the adversary may easily get full control of it (e.g., Dolev-Yao model [32]).

The above three attacks seem to impose high requirements for attackers, i.e., requiring attackers to stay in close proximity to the user. They are however likely to occur in real-world contexts. For instance, 1) if attackers are friends and acquaintances, they may embrace or shake hands with the user; 2) if the user is in a crowded environment, such as a subway, concert, or stadium, it is feasible to be in close contact with the attacker; 3) attackers may even pose as someone asking for directions, thus getting close to the user and dispelling the user's wariness. Due to the high return and obvious benefit upon a successful attack, attackers may take whatever is possible to perform attacks no matter how complicated the attack vectors may be. Therefore, the attack scenarios we listed above are viable in practice and serve to validate the pairing scheme presented in this work.

3 SYSTEM OVERVIEW

The basic idea underlying our approach is to analyze the body-transmitted electrical signals to determine whether a device pairing request is initiated by a legitimate user. Two devices are allowed to be paired with each other if and only if they are held by a legitimate user at the same time. The following two features highlight what sets our work apart.

Bilateral Touching Strategy Existing *touch-to-access* schemes typically only require the user's physical touch with the device to perform pairing [16, 33, 44, 45]. This unilateral-touch requirement is experimentally proved to be susceptible to user's mistouch and physical contact with the attacker such as handshake (see Section 4.2). To mitigate the potential vulnerabilities, we develop a bilateral touching strategy, i.e., users maintain contact with the touch electrodes and concurrently connect the device electrodes to establish a closed-loop for signal transmission, which provides a dual confirmation to preserve the security of the pairing process.

User-Specific Key Generation and Confirmation Rather than concentrating on developing a fuzzy commitment framework for key establishment as most of the previous studies do [13, 16, 34, 45], we propose to remove such intricate and resource-consuming algorithms using the human body's unique biometric for key generation. To confirm if the key originates from a legitimate user, we mathematically estimate the user-specific characteristics of the body channel to decrypt the pairing keys for matching. The non-user-specific nature of the previous studies implies that anyone can access the device by simply touching it, exposing risks in cases where the device is lost or the user leaves it.

Table 1: Measurement settings.

Parameter	Default Value
Device Baud rate	9600*bps*
Mic power supply	5*Volt* DC
Cable	Dupont jumper wires
Transmission distance	Finger-to-finger
Skin condition	Dry
Environment	Quiet Lab office, 20°C and 30% RH[1]

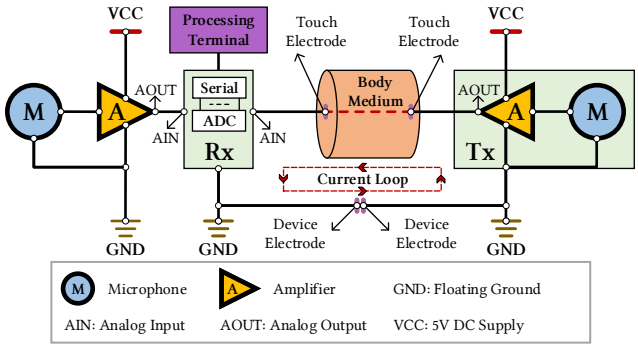


Figure 4: Equivalent circuit of the system.

4 FEASIBILITY ANALYSIS

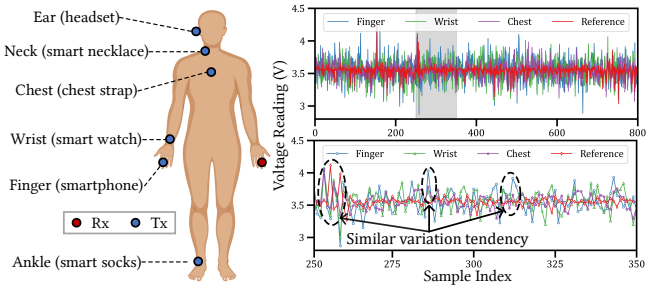
In this section, we investigate the feasibility of exploiting ambient sound as an information source and human body as a medium to generate and transmit secret keys for device pairing. We are primarily concerned with determining 1) whether the same secret keys are generated from different positions on the same human body and 2) whether the keys are discrepant from different bodies.

4.1 Measurement Setup

To simulate the key generation and transmission, we conduct a proof-of-concept study using commercial off-the-shelf Arduino Uno boards [3] and LM386 ECMs [41] as shown in Fig. 3(a). Note that we test different types of microphones including MEMS microphones and ECMs (see Fig. 3(b)), and all of them are found to generate current signals that can travel over the body channel. Since MEMS microphones are more sensitive than ECMs and they work on the same principle as mentioned in Section 2.1, we use ECM in the measurements as a kind of *stress test* for our approach. Table 1 lists the default settings we adopt unless stated otherwise.

Fig. 4 shows the equivalent circuit connection of the system. In the following feasibility studies, we conjoin the two device electrodes to facilitate the measurement. We employ two sets of equipment, the Tx records ambient sounds and produces the corresponding current signals; the Rx captures the current signals sent over the human body and collects ambient sound signals as a reference. For each equipment, the LM386 ECM is wired to the 5*V* DC supply

[1]For the purpose of comparing the measurements, we play the same audio (≈30*dB*) using an external source in a quiet lab office with a temperature of 20°C and relative humidity (RH) of 30%. The same sound source ensures that the microphone produces the same current signal for our feasibility studies. Note that it is not necessary for practical use.



(a) Touch electrode positions (b) Measurement at finger, wrist and chest

Figure 5: Illustration of electrode placements and measurements on the same human body.

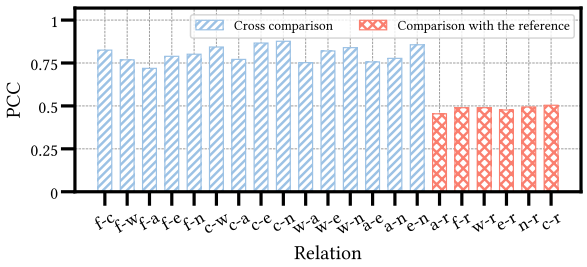


Figure 6: Signal similarity comparison.

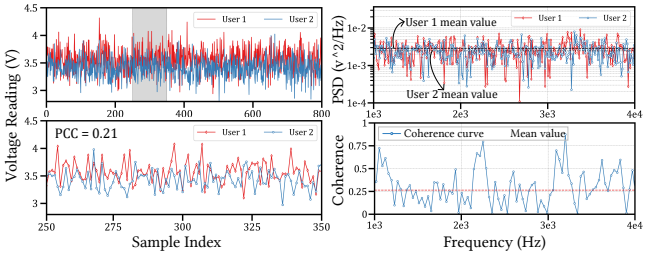
and the floating ground of the Arduino Uno board. The Tx touch electrode is linked to the ECM via a power amplifier to enhance the captured analog signal voltage to the expected $5V$ at the pins on Arduino. The Rx touch electrode is directly connected to the 10-bit analog-to-digital converter (ADC) on the Arduino board to record and process the received signal. To study the measured and reference signals in tandem, we use cross-correlation to align them [31] for further processing.

4.2 Preliminary Results

Case 1: Electrical response on the same human body. In this study, we investigate the impact of electrode position attached to the same human body on the received electrical signal. The study will be also useful in testing if signals can transmit throughout the human body, and it may offer a new opportunity for a variety of wearable applications. Specifically, we recruit a 34 years old male with a height of $170cm$ and weighting $65kg$ to conduct the experiment. The Rx electrode is attached to his right hand, and the Tx electrode is attached to different positions as illustrated in Fig. 5(a). Fig. 5(b) shows the example signals that are collected separately from three positions, where the bottom panel shows a zoomed-in view of the grey data segment. From the figure, we have two observations:

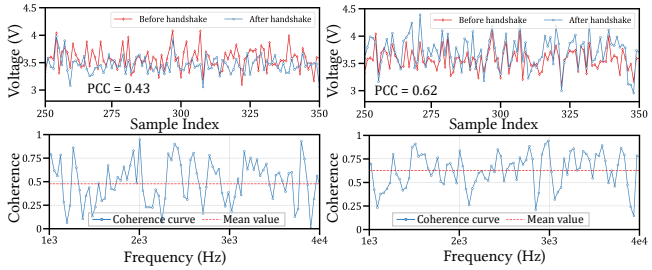
1) Without regard to amplitude, the measured signals exhibit comparable variation tendencies when the Rx/Tx electrodes are on the same human body, as shown by the dashed circle examples. In addition, we cross-compare the six measured signals using Pearson correlation coefficient (PCC) [5]. The results shown in Fig. 6[2] evidence that signals from the same body share a high similarity in variation shapes, with their values all above 0.7.

[2]The abbreviations on the x-axis, i.e., f, c, w, a, e, n, and r, denote finger, chest, wrist, ankle, ear, neck, and reference, respectively.



(a) Comparison in time domain (b) Comparison in frequency domain

Figure 7: Signal measured from two users.



(a) Dry hand contact (b) Sweat moist hand contact

Figure 8: Signal measured when handshaking in different skin conditions.

2) The measured signals show a pronounced distortion compared to the reference. The corresponding PCCs of below 0.5 in Fig. 6 suggest that the body significantly interferes with the signals, and the resulted distortion increases with transmission distance. As described in Section 2.2, the human body can be viewed as an intricate and nonlinear conductor, which modulates the original signal to be somehow encoded. These phenomena inspire us to exploit the uniqueness of the human body to encrypt the ubiquitous entropy source (i.e., ambient sound) to generate secret keys for device pairing.

Case 2: Electrical response on different bodies. In this experiment, we collect data separately from two participants (a 34 years old male with a height of $170cm$ and weighting $65kg$, and a 36 years old female with a height of $165cm$ and weighting $55kg$) by finger-to-finger connection. Fig. 7(a) displays the signals measured from the two subjects, with grey area magnified in the bottom plot. The variation trends of the electrical responses between the two subjects are intuitively less correlated, evidenced by a low PCC of 0.21. In Fig. 7(b), we further compare the two signals over a frequency band of 1-4kHz. The upper panel represents the power spectral density (PSD) distributions [24] of the two signals, it is observed that Subject 1's frequency components are not compatible with that of Subject 2. The bottom panel estimates the coherence of the two power spectra by the Hanning window and Welch's method [11], the mean coherence of 0.26 indicates that they are weakly correlated.

This experiment reflects that it is unlikely to generate identical secret keys from different bodies due to the complexity and diversity of the human body. The above study essentially validates the feasibility of the proposed scheme in this paper: the human body works as a user-exclusive secret key generator; it is also a transmission medium to transfer the encrypted keys for device pairing.

Case 3: Electrical response on different bodies that are physically contacted. We envision an adversary attempting to generate the same pairing key as the legitimate user by means of physical touch (e.g., handshake). To simulate this scenario, we ask the two participants to shake hands and place the Rx and Tx electrodes on the other hand of Subject 1 (legitimate user) and Subject 2 (adversary), respectively. Fig. 8(a) shows the measurements when the adversary's hands are dry, where the red line is Subject 1's original signal and the blue line depicts the signal after hand contact. In comparison to the findings obtained from different bodies (see Fig. 7), the result (a PCC of 0.43 and a coherence of 0.44) indicates that the similarity is enhanced by physical contact but does not unveil strong correlation. To go a step further, we validate the similarity when the adversary is in a moist skin condition (e.g., after gym). As shown in Fig. 8(b), the result (a PCC of 0.62 and a coherence of 0.63) confirms that the two signals are much more correlated. This is because sweat sodium chloride (NaCl) increases the amount of impurities on the skin surface, thus improving the body's electrical conductivity. As we can imagine, if the adversary applies saline solution to the transmission path (e.g., hands and arms), the signal similarity will be much higher.

The above study implies that most of the existing *touch-to-access* pairing schemes [16, 33, 45], which take advantage of the human body as a conductor, may face potential security threats. By increasing the body conductivity (e.g., doping with salt water) and physically touching the victim, an adversary can possibly obtain measurements that are quite similar to the victim's. The fundamental vulnerability of these schemes is that they rely only on the user's unilateral touch without additional intent confirmation, which is also likely to result in connecting to an unwanted device due to user's mistouch. Such security and usability limitations drive us to design the bilateral-touch strategy in this work.

5 DEVICE PAIRING PROTOCOL

Our preliminary studies in Section 4.2 demonstrate the feasibility of using the human body as a key derivation function to encrypt the time-varying and transient ambient sound for device pairing. In this section, we describe the design principle of our device pairing protocol, explain how we exploit the unique characteristics of the human body to develop a simple and effective solution, and discuss why existing encryption schemes fall short.

5.1 Design Motivation

The most intuitive approach for device pairing is to exchange secret source over a secure connection such as transport layer security (TLS), and make an authentication decision based on signal similarity [10, 45]. However, the deficiency with this approach lies in its vulnerability to MITM attacks. Since the secure connection is not authenticated, an attacker may eavesdrop on the channel's information exchange. Once picking up the pairing signal from the requester, the attacker relays it to the authenticator, leading to successful authentication.

To respond to such attacks, symmetric key commitment schemes are employed for pairing agreements [16, 44, 46]. Typically, these schemes predefine several signal templates and use matching methods such as dynamic time warping (DTW) [40] to quantify the

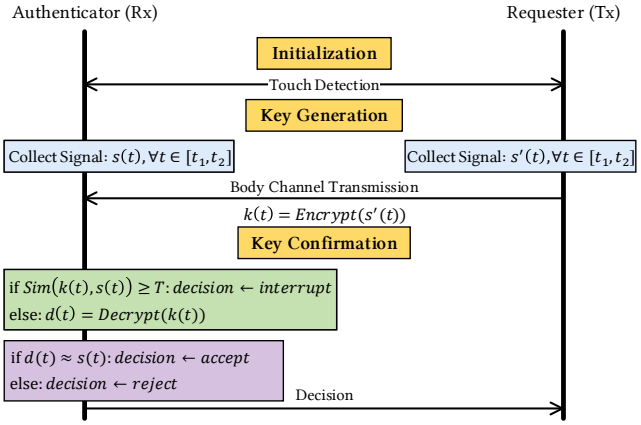


Figure 9: Pairing protocol.

secret source into a binary sequence that serves as a key. After key exchange, two devices conclude a fuzzy agreement on the pairing key using an error correction coding (ECC) [14] algorithm, e.g., Reed-Solomon (RS) [12]. However, these schemes have two problems. For one thing, since key exchange occurs over a public wireless channel (e.g., WiFi and Bluetooth), an attacker may launch a phishing attack on the channel to collect sufficient sensitive information, and then uses an offline dictionary attack to gain access to authenticators [4]. For another, the operations in these schemes are complex and constrained by processing power and computational resources, which may not be practical for wearable devices (e.g., wristband and headset).

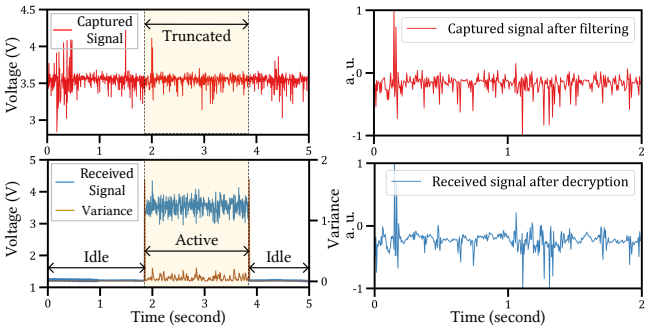
Consequently, our practice to this end is to leverage the intrinsic properties of the human body to spontaneously create and transmit pairing keys, thus reducing computational costs and enhancing communication link security.

5.2 Protocol Overview

We outline our pairing protocol in Fig. 9, which defines two major procedures as follows.

Key Generation The pairing process is initiated upon touch is detected (i.e., circuit connected). The two sides then collect their respective electrical signal caused by ambient sounds $s(t)$ and $s'(t)$ synchronously for a touch duration of τ. Afterward, $s'(t)$ is transmitted to the Rx through the body channel, and concurrently, it is also encrypted by the human body to generate $k(t)$. This procedure avoids cumbersome algorithm-based encryption on Tx, which is more practical for resource-constrained wearable devices.

Key Confirmation Our pairing scheme uses the human body to encrypt transmitted signals. As we demonstrate experimentally in Section 4.2, the human body modulates $s'(t)$ to be encoded, resulting in $k(t) \not\approx s(t)$. We consider a scenario that an adversary may bypass this scheme by connecting the electrodes directly with a wire. To obviate this security hazard, the authenticator first verifies if the signal $k(t)$ is fully encrypted by the body channel by similarity comparison using PCC. If the value is lower than a predefined threshold T, the authenticator determines that $k(t)$ is freshly generated by the body channel. Subsequently, it is decrypted by the user-private decryption function to $d(t)$ and the authenticator matches $d(t)$ with $s(t)$. If $d(t) \approx s(t)$, the authenticator accepts



(a) Temporal alignment (b) Filtering and decryption

Figure 10: Signal pre-processing.

the access request; otherwise it rejects the session. In contrast to password-authenticated key exchange (PAKE) solutions that the keys are generally reusable[8], our pairing keys are dynamic and disposable, hence preventing typical replay attacks.

5.3 Pre-processing

Encryption Detection The challenge in this step is to define an appropriate similarity threshold T that validates whether $k(t)$ is encrypted by the body channel. A high value of T leads to system vulnerabilities, as an inadequately encrypted signal or even an unencrypted signal may complete the pairing process. A low value of T is likely to result in system interruption when the user is in high skin conductivity (e.g., after gym). We investigate T by collecting data from 31 participants immediately after a 10-minute gym session, and calculate the mean value of similarity (i.e., 0.67) as the threshold. We admit that this may limit the usability (e.g., after extensive exercise), however, security and usability are always antithetical to each other for biometric-based systems [36], just as fingerprint recognition is not working if the hand is wet or oily and facial recognition fails in low light conditions.

Temporal Alignment On the authenticator side, the ADC reads signals from the body channel (i.e., received signals) and the microphone (i.e., captured signals) continuously and simultaneously with a clock frequency of $125kHz$. To facilitate comparison, our approach searches for the touch-active segment by monitoring the variance of the received signal in a sliding window, and uses this segment to truncate the captured signal, as shown in Fig. 10(a). The active segment is identified when the variance of the received signal inside a sliding window exceeds a predefined threshold as described in [27]. In our implementation, we empirically use a window of 10 samples (i.e., 0.05 seconds) and a variance threshold of $T_\sigma = 0.219$ to obtain the best result. The detected onset of touch also serves as a trigger to initiate the pairing process, obviating the requirement for any additional operation to indicate the intent to pair a set of devices.

Spatial Alignment Since the authenticator and the requester may have different microphone power voltages, the received and captured signals will be spatially disjointed and hence not directly comparable. We address this by normalizing the magnitude of both signals to range $[-1, 1]$ using the scaling method $x' = 2\frac{x - x_{min}}{x_{max} - x_{min}} - 1$, where x' is the normalized value for variable x, x_{max} and x_{min} are the maximum and minimum values of x, respectively.

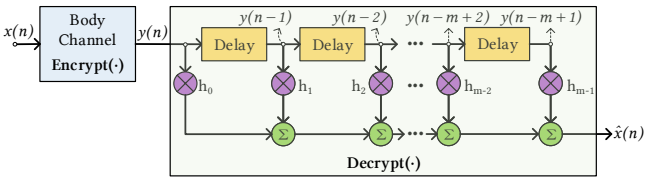


Figure 11: Design of decryption function model.

Captured Signal Filtering During the capturing of the electrical signals produced by ambient sounds, there are many sources of noise. We observe that the most prominent noise is either power-line interference (around $50Hz$) or electrical noise (less than $10Hz$) which is generated by the friction between the device electrodes as well as the touch electrode and the skin. We thus apply a high-pass filter with a cutoff threshold of $55Hz$ to eliminate the effect of the noise. The upper panel of Fig. 10(b) shows the captured signal $s(t)$ aligned and filtered.

5.4 Received Signal Decryption

As we described above, the received signal is encrypted by the body channel, there are various inconsistencies in variation forms between $k(t)$ and $s(t)$ even though they are derived from the same secret source. Consequently, we need to estimate the decryption function in order to fully recover the transmitted signal before matching. Due to the complexity and diversity of the body's physiological structure, the decryption function is distinct from person to person. Before using our pairing scheme, users are required to perform a calibration to determine their private decryption function and store it on the authenticator side.

$Decrypt(\cdot)$ is the inverse of $Encrypt(\cdot)$, which aims to stably restore the body channel-corrupted signals. To achieve this, we design a finite impulse response (FIR)-based decryption model as shown in Fig. 11. Specifically, we use $x(n)$ and $y(n)$ to denote the transmitted and received signals, respectively. The decrypted signal $\hat{x}(n)$ is modeled as:

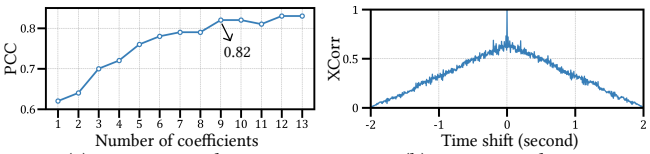
$$\hat{x}(n) = \sum_{k=0}^{m-1} h_k y(n-k) = \boldsymbol{h}^T \boldsymbol{y}, \tag{1}$$

where $\boldsymbol{h}^T = [h_0, h_1, ...h_{m-1}]$ is the impulse response coefficient vector of the decryption model. We use mean square error (MSE) to denote the corresponding estimation error, which is given by:

$$\begin{aligned} E[e^2(n)] &= E[(x(n) - \boldsymbol{h}^T \boldsymbol{y})^2] \\ &= E[x^2(n)] - 2\boldsymbol{h}^T E[\boldsymbol{y}x(n)] + \boldsymbol{h}^T E[\boldsymbol{y}\boldsymbol{y}^T]\boldsymbol{h} \\ &= E[x^2(n)] - 2\boldsymbol{h}^T \boldsymbol{r_{yx}} + \boldsymbol{h}^T \boldsymbol{R_{yy}} \boldsymbol{h}, \end{aligned} \tag{2}$$

where $E[\cdot]$ is the expectation operator, $\boldsymbol{r_{yx}}$ is the cross-correlation vector of the received and the transmitted signals, and $\boldsymbol{R_{yy}}$ is the auto-correlation matrix of the received signal. From Eq. (2), the MSE for our decryption model is a quadratic function of the coefficient vector $\boldsymbol{h}$ and has a single minimum point. Its gradient with respect to the vector $\boldsymbol{h}$ is given by:

$$\frac{\partial E[e^2(n)]}{\partial \boldsymbol{h}} = -2\boldsymbol{r_{yx}} + 2\boldsymbol{h}^T \boldsymbol{R_{yy}}. \tag{3}$$



(a) Parameter selection (b) Cross correlation

Figure 12: Signal comparison.

The minimum MSE is obtained by setting Eq. (3) to zero. Equivalently, we have the optimal solution as follows.

$$\boldsymbol{h}_{opt} = \boldsymbol{R_{yy}}^{-1}\boldsymbol{r_{yx}}. \tag{4}$$

In an expanded form, the decryption model solution can be written as

$$\begin{bmatrix} h_0 \\ h_1 \\ \vdots \\ h_{m-1} \end{bmatrix} = \begin{bmatrix} r_{yy}(0) & \cdots & r_{yy}(m-1) \\ r_{yy}(1) & \cdots & r_{yy}(m-2) \\ \vdots & \ddots & \vdots \\ r_{yy}(m-1) & \cdots & r_{yy}(0) \end{bmatrix}^{-1} \begin{bmatrix} r_{yx}(0) \\ r_{yx}(1) \\ \vdots \\ r_{yx}(m-1) \end{bmatrix}.$$

During user calibration, $y(n)$ is the received signal and $x(n)$ is equivalent to the captured signal after filtering (discussed in Section 5.3), the decryption function parameters for a specific user can be calculated by Eq. (4). To determine the number of the function parameters, our principle is to make the decrypted signal $\hat{x}(n)$ as similar to $x(n)$ as possible. As illustrated in Fig. 12(a), it is observed that the PCC increases with the number of coefficients and it gradually saturates when the number reaches 9 in this example. Afterward, we can obtain the final decrypted signal as illustrated in the lower plot of Fig. 10(b).

5.5 Matching Decision

After we finish processing the captured and received signals, the last stage is to generate a decision on matching. We measure the similarity of the two signals using the normalized cross-correlation (NCC) [9], as shown in Fig. 12(b). In our work, we do not adopt more advanced compare methods such as dynamic time warping (DTW) which require a heavy computational burden and may discard temporal information. To enhance matching accuracy, we employ a machine learning classifier to make the decision rather than manually defining a threshold that decides whether the cross-correlation result is a match or a non-match. Specifically, our decision classifier is described as follows.

Feature Set We use the NCC values of the two signals as the classifier's feature set. To guarantee that the number of features is consistent among all matching operations, we first identify the maximum of the NCC (i.e., the point with value 1), and then pick 400 data points (i.e., 2 seconds) from the left and right sides of the maximum, respectively. As a result, we have a feature set of 801 values. The determination of data length will be discussed in Section 6.2.

Classifier We apply a classifier–radial basis function-based support vector machine (RBF-SVM) due to its effectiveness in processing data with high-dimensional features. To train our decision classifier, we pre-collect a small amount of data from several participants in a manner described in Section 6.1. Note that the data used to train the classifier is separate from the data used to evaluate system performance. Then, we implement the 5-fold cross-validation and grid search method to optimize the classifier's parameters [38]. Specifically, the pre-collected data is randomly split into 5 subsets. Among the subsets, one single subset is reserved as validation data for the classifier, and the remaining 4 subsets are used for training. The process is then repeated 5 times, with each of the 5 subsets being used strictly once as the validation set. For parameter selection, we prebuild a set of potential parameters, which contains 7 values that are logarithmically spread from 10^{-3} to 10^{3}. After iterating through the parameter set, the classifier's optimal parameters C and γ are 1 and 10^{-3}, respectively.

In our prototype, we use the constructed classifier to determine if the two signals (i.e., the captured signal and the received signal) are similar, instead of manually specifying a threshold that decides whether the similarity result of the signals is a match or not. The similarity decision classifier is trained offline only once with a small dataset, and it is user-agnostic, i.e., users are not required to undertake any initial training for establishing the decision classifier before use. During enrollment, users are only needed to calibrate their individual decryption functions as described in Section 5.4.

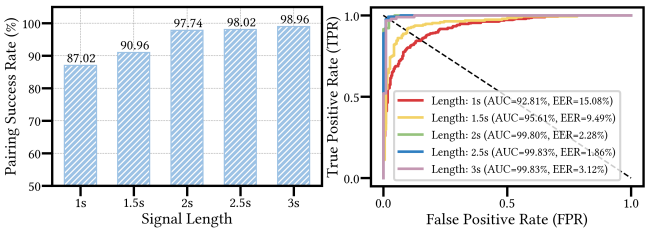
6 SYSTEM EVALUATION

We now move to evaluate system performance based on the prototype we have developed (see Section 4.1). We first describe data collection, we then present the performance results of authenticating legitimate pairing and the resilience to various attacks. We finally evaluate system performance under a range of usage scenarios for potential real-world deployment.

6.1 Data Acquisition

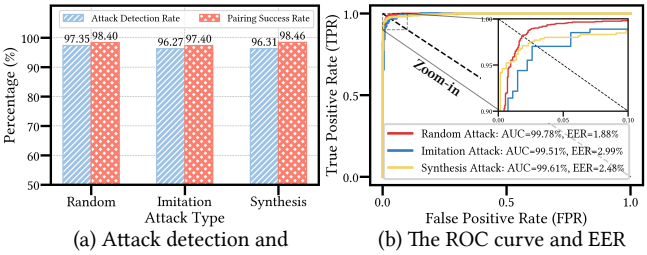
Legitimate Data We recruit 31 participants (11 females and 20 males) between 19-35 years old for our experiments. Their heights and weights fall in the range of 1.6-1.9 meters and 45-80 kilos, respectively. Participants are well informed prior to experiments that the prototype device poses no risk to human health and that their personal data are well protected (i.e., de-identified and stored locally). Unless specified otherwise, we follow the settings listed in Table 1. Our data are collected in multiple rounds over a period of two months. During collection, we randomly play pre-recorded background sounds including ambient office noise, crowds, and road traffic. Participants are simply needed to maintain the circuit connected by touching, without deliberately placing the device stable for operation. For each participant, we collect the captured and received signal pairs for 10 minutes to evaluate legitimate user matching.

Attack Data We evaluate the system's resilience to common attacks described in Section 2.3. The attack data are collected as follows: 1) *Random Attack*. We randomly choose one participant as a legitimate user and the rest as adversaries to launch the attack. Each adversary randomly touches the device several times, and we obtain 2-minute data. In total, we collect 60-minute random attack data. 2) *Imitation Attack*. We randomly invite one participant as a legitimate user and the rest as adversaries. Specifically, we play the same background sound and ask the adversaries to repeat the same touching positions with the same touching duration as the legitimate user does. Each adversary mimics the legitimate user 10 times, we thus obtain 300 signal pairs for evaluation. 3) *Synthesis*



(a) Pairing success rate (b) The ROC curve and EER

Figure 13: Performance with different signal lengths.



(a) Attack detection and pairing success rate (b) The ROC curve and EER

Figure 14: Performance under various attacks.

Attack. An adversary first records the ambient sound when the user is pairing devices, then simulates the pairing key by adding white Gaussian noise with different signal-to-noise ratios (SNRs) to the recorded signal, which is modeled as follows.

$$noise = Gauss(n) \cdot \sqrt{\frac{\sum x^2}{n \cdot 10^{\frac{SNR}{10}}}}, \tag{5}$$

where $noise = Gauss(\cdot)$ follows distribution $\mathcal{N}(0, 1)$, x is the recorded signal, and n is the signal length. Due to the restriction of encryption detection discussed in Section 5.3, we randomly choose 10 different SNRs from range $[5, 30]$. We synthesize 100 pairing keys for each SNR, resulting in a total of 1000 forgery keys to attack the legitimate user.

6.2 Overall System Performance

We begin by describing the following evaluation metrics. 1) *Pairing success rate (PSR)* is the percentage of correctly matched instances performed by a legitimate user during the pairing process. 2) *False Acceptance rate (FAR)* specifies the percentage of unauthorized instances that are incorrectly accepted. 3) *True positive rate (TPR)* denotes the percentage of attack instances that the system correctly rejects. 4) *False positive rate (FPR)* represents the percentage of legitimate instances that the system falsely rejects. 5) *ROC Curve* illustrates the comparison between TPR and FPR under different discrimination thresholds. A bigger area under the ROC curve (AUC) implies that the system performs better. 6) *Equal Error Rate (EER)* is the rate when the discrimination threshold is adjusted to the point where the false positive rate and false negative rate are equal. A lower EER means the system is more accurate.

We first evaluate the pairing success rate and the pairing time of our system, and both are important to real-world deployment. The pairing time mainly depends on the signal length required to complete a match (discussed in Section 5.5). To determine an optimum length of the required signal, we train the classifier with signal lengths of 1s, 1.5s, 2s, 2.5s, and 3s, respectively, and evaluate the system performance accordingly. Fig. 13(a) shows the average pairing success rate with different signal lengths. We observe that the success rate improves with the increase of signal length and it gradually saturates when the length reaches 2s (i.e., 97.74%). This is expected since a longer signal contains plentiful information for accurate pairing, and no more effective information is provided when the length exceeds 2s.

Further, we examine the performance using ROC curves and calculate the corresponding AUC and EER with different signal lengths, as shown in Fig. 13(b). Consistent with the pairing success rate, the AUC increases with signal length and does not improve substantially when the length is raised from 2s to 3s. The corresponding EERs from 1s to 3s are 15.08%, 9.49%, 2.28%, 1.86%, and 3.12%, respectively. Based on the observations, we conclude that the signal length of 2s is optimal for the classifier's training and testing. The results also demonstrate the effectiveness of our system in authenticating the device pairing of legitimate users.

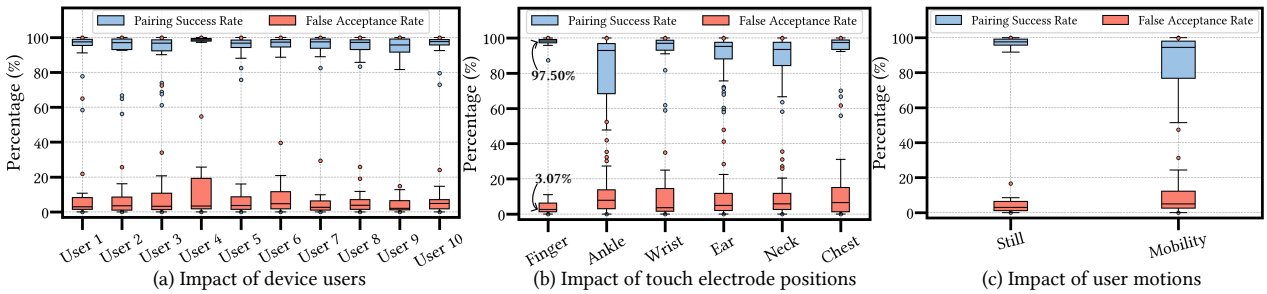
6.3 Resilience Against Attacks

We now evaluate the system performance under the attacks described in Section 2.3. Results are shown in Fig. 14,

Random Attack In this experiment, our system achieves an attack detection rate of 97.35% and a pairing success rate of 98.40%. In terms of the ROC curve, the AUC and EER are 99.78% and 1.88%, respectively. The results indicate that the system is effective in distinguishing legitimate pairing requests from random attacks. This is expected because 1) the adversaries are unaware of system settings (e.g., touching duration), and the system may reject a pairing request due to insufficient signal length; 2) the adversaries' body-transmitted signals are difficult to generate similar impacts by the legitimate user's decryption function, which is consistent with our methodology described in Section 5.4.

Imitation Attack In this experiment, the system identifies imitation attacks and legitimate pairings at a rate of 96.27% and 97.40%, respectively. We observe that the rates are slightly lower than those in random attacks by approximately 1%, this is because random attacks contain data samples that are insufficiently lengthy (i.e., less than 2s), hence the system rejects these samples without matching. Besides, the AUC and EER are 99.51% and 2.99%, respectively, showing that the adversaries are hard to pass the matching even if they have the knowledge of the legitimate user's touching behavior and use the same ambient sound. Since the transmitted signal is encrypted by the adversary's body channel and the human body's physiological structure is distinct from person to person, it is nearly impossible for the adversary to derive the same pairing keys as the legitimate users do.

Synthesis Attack From the result, we observe that the system still retains strong resilience to more sophisticated synthesis attacks, with an attack detection rate of 96.31% and a pairing success rate of 98.46%. The AUC of 99.61% and EER of 2.48% further evidence that, even if the pairing keys are synthesized from the original secret source, our system is still effective in discriminating forgery keys and properly confirming valid users in the meanwhile. This is largely owing to the complexity of the human body channel and the robustness of our decryption function, it is scarcely possible for



(a) Impact of device users (b) Impact of touch electrode positions (c) Impact of user motions

Figure 15: Performance under usage factors.

adversaries to model the encryption process without prior knowledge of the legitimate user's body channel information. It is worth noting that some existing schemes are susceptible to synthesis attacks since the signals they adopt as pairing keys such as cable emanations [45], gestures [10, 27], and gaits [44], present certain patterns and can be readily predicted.

6.4 Performance in Different Usage Scenarios

In this section, we evaluate how the system performs under different usage scenarios. The study provides practical guidance in deploying our system for a better user experience in real-world settings. We randomly invite 10 people from our participants and use the default settings listed in Table 1 unless stated otherwise.

Device User We first evaluate the performance with different device users. Fig. 15(a) shows the results for the ten users at different classifier's discrimination thresholds. We observe that the PSR and FAR display slightly different across the users, while their medians are relatively steady, with an average of 97.33% and 3.18%, respectively. The result confirms that our decision classifier is not user-specific, and users are not required to undertake additional training prior to use as mentioned in Section 5.5. It is noteworthy that users need to calibrate and store their private decryption functions on the authenticator side; whereas some existing *touch-to-access* approaches [16, 33, 45] do not involve such user-related constraints. This may result in an adversary being able to access the device with a simple touch if it is lost/stolen or left somewhere. Apparently, our scheme avoids this deficiency.

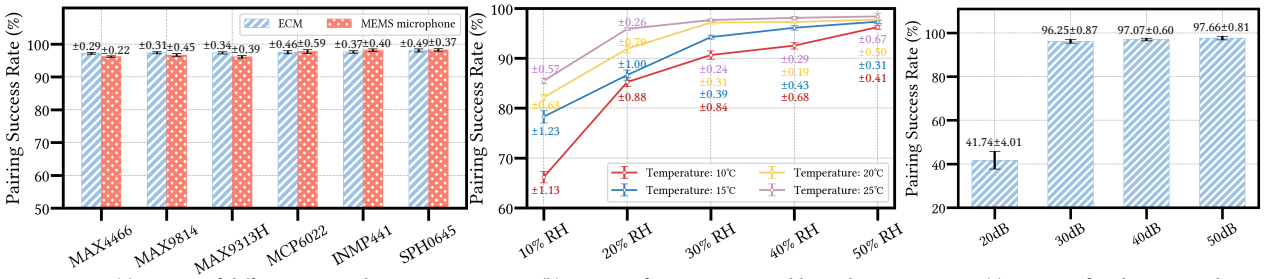
Touch Position To analyze the impact of touch electrode position, we conduct experiments with six different body positions as illustrated in Fig. 5(a). Fig. 15(b) shows the pairing performance in which we use the finger-to-finger manner as a control group. At the positions of wrist, ear, neck and chest, the PSRs and FARs slightly decrease compared to those in the control group, with an average rate of 95.98% and 4.91%, respectively. We observe that the PSR and FAR in the position of ankle (i.e., 92.93% and 8.15%, respectively) exhibit a certain difference. This is because the decryption function and the decision classifier are estimated using finger-to-finger data, and the signal distortion increases with the transmission distance as we demonstrate experimentally in Section 4.2. In comparison to other four positions, the ankle is the farthest from the finger, leading to the difference in performance. To address this problem, we may model several decryption functions and train the classifier with data from multiple touch positions, which we leave for our future work.

User Motion In this experiment, we evaluate how user motion affects the performance. We test two cases, i.e., participants stand still and walk. Fig. 15(c) depicts that as user status changes from still to mobility, the median of PSR reduces from 97.62% to 94.24% and the median of FAR rises from 3.03% to 4.81%. In addition, the bigger box indicates that the results in mobility becomes more discrete and volatile. The main reason for this phenomenon is that the electrodes may not be firmly contacted when the user is moving, resulting in a loss of signal. 1 the impact of motion, the pairing performance is acceptable in practice. It is noted that some approaches generate pairing keys from human body movements, such as gesture [1, 10] and gait [7, 44], which are rendered ineffective when users stand still.

Different Devices In practice, different devices may possess different kinds of microphones. To ensure usability, it is essential to verify whether our system is device-insensitive. In this experiment, we employ an ECM and a MEMS microphone at the authenticator side, respectively, and evaluate the performance using different microphones as shown in Fig. 3(b) on the requester side. Fig. 16(a) shows that the pairing performance is encouraging in all cases. All microphones work on the same principle, i.e., converting a sound signal into an electrical signal, demonstrating our system is device-independent. Besides, the ubiquity of microphones in wearable devices makes our approach more practical than those that require dedicated sensors [34, 43, 46].

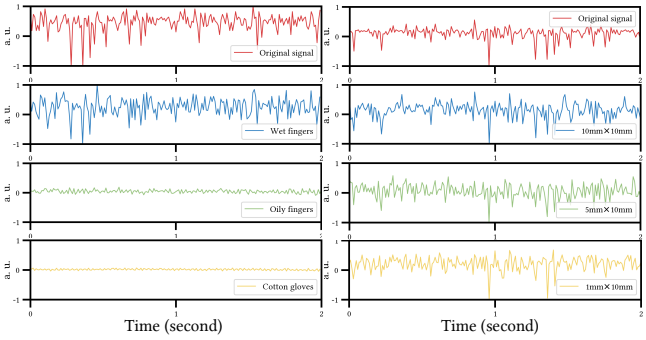
Ambient Environment In this experiment, we investigate how ambient temperature and humidity affect the performance. To do so, we use an air conditioner and a humidifier to control the ambient environment. Specifically, we vary the room temperature from 10°C to 25°C with a step size of 5°C. Under each temperature level, we control the relative humidity (RH) from 10% to 50% with a step size of 10%. Fig. 16(b) depicts the corresponding pairing accuracy. We observe that the increase of either temperature or RH improves the PSR. When the temperature is low, humidity has a great influence on PSR. Under 10°C, for example, the PSR enhances from 66.14% with a standard deviation (STD) of 1.13% to 96.25% with an STD of 0.14% as the RH rises from 10% to 50%. When the RH is 30%, the average PSR at all temperatures reaches over 95% with around 0.45% STD. The phenomenon mainly stems from the fact that the lower the air humidity, the dryer the skin surface, resulting in higher body resistance. To improve usability in dry weather conditions, an immediate solution is to increase the skin surface moisture when pairing devices, e.g., breathing out on the touching fingers.

Sound Level We finally study the impact of sound level on pairing accuracy. In a quiet room, we control the background noise by



(a) Impact of different microphones (b) Impact of temperature and humidity (c) Impact of ambient sound

Figure 16: Performance under practical factors.



(a) Different finger conditions (b) Different electrode sizes

Figure 17: Signal patterns under different conditions.

playing random audio on four different levels: 20dB (e.g., whisper), 30dB (e.g., computer noise), 40dB (e.g., normal conversation), and 50dB (e.g., traffic noise). Fig. 16(c) shows that the system is hardly available at a sound level of 20dB, and volume has little effect on the performance once the sound level is above 30dB. As we leverage ambient sound as a source of entropy to generate secret keys for device pairing, the pairing accuracy is significantly affected when the sound is too faint to be captured effectively by the microphone. This suggests that operating in an environment with a sound level above 30dB is sufficient to achieve good pairing performance. This requirement is not strenuous; even in a quiet environment (i.e., less than 30 dB), users can still perform pairing by proactively making sounds, such as playing audio or speaking, which shows the resilience of our scheme to the ambient environment. In contrast, approaches that rely on electromagnetic radiations [16, 45] are inapplicable if radiations are shielded.

6.5 Usability Study

In this section, we investigate the potential factors that may affect the usability of the system.

Finger Condition In this study, we would like to examine the impact of different finger conditions on system usability. We measure the signal under three cases: wet fingers, oily fingers, and wearing cotton gloves. Fig. 17(a) shows the corresponding signal patterns in time domain, where the top figure plots the original signal from the microphone and the bottom three figures depict the decrypted signals from the three finger conditions. We can observe that in the case of wet fingers, the decrypted signal nearly recovers the pattern of the original signal, making it easier to perform pairing. In situations involving oily fingers and wearing gloves, the electrical responses are extremely weak or hard to be captured. This is because water helps conduct electricity while oil and cotton fiber are poor conductors of electricity. The conductive properties of finger coverings have a significant impact on the sensing performance, so we suggest to avoid covering the electrodes with insulating material when touching them for better pairing performance.

Electrode Size In practice, wearable devices typically tend to be limited in size, which poses challenges in designing electrode size. To ensure that our solution can be effectively deployed on wearables, we further conduct measurements to verify how the electrode size affects the results. Particularly, we employ three different sizes of aluminum sheets as touching electrodes: $10mm \times 10mm$, $5mm \times 10mm$, and $1mm \times 10mm$. The original signal captured by the microphone and the decrypted signals measured by these three kinds of electrodes are shown in Fig. 17(b). It is observed that the three signals are highly recoverable to their original state. Theoretically, a larger contact area is beneficial to improve conductivity; however, in practice the skin commonly has sweat on it, leading to an increase in conductivity. Hence, the signal may be transmitted quite effectively even with electrodes as small as a pinhead ($1mm \times 10mm$). The results confirm the system's resilience in practice and demonstrate the feasibility of designing tiny electrodes for wearable devices with constrained dimensions.

7 RELATED WORK

We divide the existing *touch-to-access* device pairing schemes into the following two main categories.

Biometric-based Biosignals have been studied for on-body device pairing. Rostami *et al.* extract time-varying randomness from ECG signals to form a symmetric-key commitment [34]. Yang *et al.* define several templates to quantify EMG signals into binary sequences that serve as pairing keys [46]. Zhang *et al.* propose a key generation framework that converts the interpulse intervals of both ECG and PPG signals to digital binaries and reconciles them between legitimate devices [50]. A similar idea is adopted in [22] that uses heartbeat interval signals for key construction. These approaches require some dedicated sensors or complicated design of communication transceivers, which restricts their practical use. In addition, they generally need careful placement on the human body and may perform poorly in real-life settings [45]. To improve usability, more commonly available sensors on devices such as accelerometers have been exploited to capture human movements for establishing a shared secret key (e.g., gait [42, 44] and gesture

Table 2: Comparison between different approaches.

Work	Entropy	Sensor	PSR	FNR
Jin *et al.* [16]	RF noise	RF transceiver	96.9%	2.8%
TouchAuth [45]	Cabling radiation	ADC	98.9%	2.0%
VoltKey [18]	Power-line noise	Designed circuit	90.0%	N/A
Perceptio [13]	Event timing	On-device sensors	94.9%	N/A
P2Auth [20]	User operations	Inertial measurement unit	99.6%	<3%
ShakeUnlock [10]	Gesture	Accelerometer	N/A	18.0%
Gait-Key [44]	Gait	Accelerometer	98.3%	N/A
H2B [22]	Heartbeat pulse	Piezo sensor	95.6%	N/A
Our work	Ambient sound	Microphone	97.6%	3.0%

[10, 27]). However, attacks are possible on the keys derived from behavior-based approaches if the behavior is recorded by a hidden camera with motion analysis.

Environment-based To reduce the limitations above, a series of approaches leverage on-board sensors to extract random information from the surrounding environment that can be translated to shared keys. For instance, Yan *et al.* [45] and Jin *et al.* [16] employ the common components in electronic devices, i.e., analog-to-digital converter (ADC) and RF transceiver, to sense the electric potentials (EP) induced by ambient electromagnetic radiations on the human body. The EP measurements are then transformed for key establishment by fuzzy commitment. Besides, ambient noise and power-line noise are also used as a common context by devices to authenticate each other's physical proximity [17, 18]. Han *et al.* verify device co-presence using the fact that devices in the same environment can observe the same events over time [13]. However, these approaches are non-user-specific, which poses risks when attackers are co-present with legitimate devices. Despite not requiring user involvement, Perceptio [13] may be unavailable for devices, such as those on separate floors, that observe distinct environmental information. Considering the user's legitimacy, An *et al.* propose to leverage the human hand as a user-specific infrared light source for information decryption [2]. Nevertheless, special coating materials are required to form the coding pattern, which extremely limits its real-world applications.

In Table 2, we compare several state-of-the-art device pairing approaches with ours. In terms of PSR and FNR, our work is comparable to the literature [16, 44, 45]. More importantly, we use ambient sound as the entropy source for key generation, which is more resilient to the environment. Even in a tranquil setting, users can artificially produce sounds to perform pairing. In contrast, [16] and [45] rely on RF noise and indoor cabling radiation, respectively, and thus are inapplicable in places where the electromagnetic radiation is shielded and outdoors. Gait and gesture-based approaches [10, 44] are not friendly to people with hand or foot disabilities. [20] requires the user to wear a wristband as a kind of token and perform a series of predefined operations to accomplish pairing, which is inconvenient and obtrusive.

8 DISCUSSION

Resource Consumption Wearable devices typically have limited resources compared to mobile devices (e.g., battery life and storage). Therefore, designing a resource-efficient pairing scheme for wearable devices is critical. As for existing approaches, resource consumption generally results from data acquisition, key generation, key transmission, and key confirmation. In our work, we harness the human body to spontaneously generate and transmit the pairing key, thus saving the resources of key generation by encryption algorithms and key transmission by networking protocols. For data acquisition, we use the device's built-in microphone (e.g., ECM and MEMS microphone) to sense the ambient sound. With the adoption of zero-power listening technology in microphones [47], the power consumption of microphones is as low as a few tenths of a milliampere. Besides, the constructed decision classifier only has a size of 5.9MB, and knowledge distillation [29] can be used to further reduce the model size in our future work.

User Safety Our scheme involves an electric current flowing through the human body, which naturally raises concerns about health safety. According to ICNIPR [15], the safe contact current for human body should be smaller than 20mA. An off-the-shelf microphone on smart devices typically requires a small DC power supply, usually from 1 to $5V$. The human body's resistance normally ranges from 1000 to 5000Ω, depending on external conditions [25]. For example, the overall resistance is reduced if hands are wet or have cuts. According to Ohm's law, the worst-case electric current in our scenario is $\frac{5V}{1k\Omega} = 5mA$, which is far below the safe limit.

Sound as the Entropy Traditional modalities rely on what we refer to as static biometrics, such as fingerprint, voiceprint, and face. Since these biometrics are invariable and reusable, the problem lies in that once the bio-information has been divulged, the user cannot recover [19]. Motivated by this vulnerability, we exploit the human body in this work to encode time-varying and transient ambient sound, generating dynamic and disposable pairing keys. In terms of practical use, our experiment suggests keeping a sound level over 30dB for better pairing performance. Even in a quiet context, users can still perform pairing by proactively making some sounds, such as playing audio on the device, speaking, or doing any activity that may produce sounds, including clapping hands and knocking on an object.

9 CONCLUSION

This paper proposes a novel touch-to-access approach to wearable device pairing that uses the human body to generate and transmit secret keys. We leverage the on-device microphone to capture ambient sounds and convert them to an electrical signal as a common source for key generation. Then we propose a user-specific decryption model and present a pairing protocol that enables both devices to agree on the mutual information source. We build a prototype using off-the-shelf microphones and microcontroller boards. Extensive experiments show that our scheme is resilient against various attacks and effective in verifying pairing requests from legitimate users in different usage scenarios.

ACKNOWLEDGMENTS

We would like to thank our shepherd and the anonymous reviewers who helped to improve this paper with their insightful comments and constructive suggestions. We would also like to thank Xiaoluan Zhang and Qianfeng Wang for their supports to this paper. This work is supported by NSFC (Grant No. 62002278).

REFERENCES

[1] Imtiaj Ahmed, Yina Ye, Sourav Bhattacharya, Nadarajah Asokan, Giulio Jacucci, Petteri Nurmi, and Sasu Tarkoma. 2015. Checksum gestures: continuous gestures as an out-of-band channel for secure pairing. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)*. 391–401.

[2] Shun An, Wen Shang, Modi Jiang, Yini Luo, Benwei Fu, Chengyi Song, Peng Tao, and Tao Deng. 2021. Human hand as a powerless and multiplexed infrared light source for information decryption and complex signal generation. *Proceedings of the National Academy of Sciences* 118, 15 (2021).

[3] Arduino 2022. *Arduino Home*. Retrieved Jan. 12, 2022 from https://www.arduino.cc/.

[4] José Becerra, Peter YA Ryan, Petra Šala, and Marjan Škrobot. 2019. An offline dictionary attack against zkpake protocol. In *IFIP International Conference on ICT Systems Security and Privacy Protection*. Springer, 81–90.

[5] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. Pearson correlation coefficient. In *Noise reduction in speech processing*. Springer, 1–4.

[6] Electromagnetic Shielding 2022. *6 Ways to Keep Your Home Safe from Radiation Exposure*. Retrieved Feb. 11, 2022 from https://www.vesttech.com/6-ways-to-keep-your-home-safe-from-radiation-exposure/.

[7] Lamiaa A Elrefaei and Ashwaq M Al-Mohammadi. 2019. Machine vision gait-based biometric cryptosystem using a fuzzy commitment scheme. *Journal of King Saud University-Computer and Information Sciences* 34, 2 (2019), 204–217.

[8] Andreas Erwig, Julia Hesse, Maximilian Orlt, and Siavash Riahi. 2020. Fuzzy asymmetric password-authenticated key exchange. In *International Conference on the Theory and Application of Cryptology and Information Security*. Springer, 761–784.

[9] Huan Feng, Kassem Fawaz, and Kang G Shin. 2017. Continuous authentication for voice assistants. In *Proceedings of the Annual International Conference on Mobile Computing and Networking (MobiCom)*. 343–355.

[10] Rainhard Dieter Findling, Muhammad Muaaz, Daniel Hintze, and René Mayrhofer. 2016. Shakeunlock: Securely transfer authentication states between mobile devices. *IEEE Transactions on Mobile Computing* 16, 4 (2016), 1163–1175.

[11] Clément Gallet and Claude Julien. 2011. The significance threshold for coherence when using the Welch's periodogram method: effect of overlapping segments. *Biomedical Signal Processing and Control* 6, 4 (2011), 405–409.

[12] Venkatesan Guruswami and Mary Wootters. 2017. Repairing reed-solomon codes. *IEEE transactions on Information Theory* 63, 9 (2017), 5684–5698.

[13] Jun Han, Albert Jin Chung, Manal Kumar Sinha, Madhumitha Harishankar, Shijia Pan, Hae Young Noh, Pei Zhang, and Patrick Tague. 2018. Do you feel what I hear? Enabling autonomous IoT device pairing using different sensor types. In *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 836–852.

[14] W Cary Huffman and Vera Pless. 2010. *Fundamentals of error-correcting codes*. Cambridge university press.

[15] ICNIRP. 2020. Guidelines for limiting exposure to electromagnetic fields (100 kHz to 300 GHz). *Health physics* 118, 00 (2020).

[16] Wenqiang Jin, Ming Li, Srinivasan Murali, and Linke Guo. 2020. Harnessing the Ambient Radio Frequency Noise for Wearable Device Pairing. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS)*. 1135–1148.

[17] Nikolaos Karapanos, Claudio Marforio, Claudio Soriente, and Srdjan Capkun. 2015. Sound-Proof: Usable Two-Factor Authentication Based on Ambient Sound. In *24th USENIX Security Symposium (USENIX Security 15)*. USENIX Association, 483–498.

[18] Kyuin Lee, Neil Klingensmith, Suman Banerjee, and Younghyun Kim. 2019. Voltkey: Continuous secret key generation based on power line noise for zero-involvement pairing and authentication. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–26.

[19] Jingjie Li, Kassem Fawaz, and Younghyun Kim. 2019. Velody: Nonlinear vibration challenge-response for resilient user authentication. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS)*. 1201–1213.

[20] Xiaopeng Li, Fengyao Yan, Fei Zuo, Qiang Zeng, and Lannan Luo. 2019. Touch well before use: Intuitive and secure authentication for iot devices. In *Proceedings of the Annual International Conference on Mobile Computing and Networking (MobiCom)*. 1–17.

[21] Teh-Lu Liao, Hsin-Chieh Chen, Chiau-Yuan Peng, and Yi-You Hou. 2021. Chaos-based secure communications in biomedical information application. *Electronics* 10, 3 (2021), 359.

[22] Qi Lin, Weitao Xu, Jun Liu, Abdelwahed Khamis, Wen Hu, Mahbub Hassan, and Aruna Seneviratne. 2019. H2B: Heartbeat-based secret key generation using piezo vibration sensors. In *Proceedings of the 18th International Conference on Information Processing in Sensor Networks*. 265–276.

[23] Jingna Mao. 2019. Investigating on the Interferences on Human Body Communication System Induced by Other Wearable Devices. In *Proceedings of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 4044–4047.

[24] Rainer Martin. 2001. Noise power spectral density estimation based on optimal smoothing and minimum statistics. *IEEE Transactions on speech and audio processing* 9, 5 (2001), 504–512.

[25] Ivan Martinovic, Kasper B Rasmussen, Marc Roeschlin, and Gene Tsudik. 2017. Pulse-response: Exploring human body impedance for biometric recognition. *ACM Transactions on Privacy and Security (TOPS)* 20, 2 (2017), 1–31.

[26] Ueli M Maurer and Stefan Wolf. 2000. The diffie–hellman protocol. *Designs, Codes and Cryptography* 19, 2 (2000), 147–171.

[27] Rene Mayrhofer and Hans Gellersen. 2009. Shake well before use: Intuitive and secure pairing of mobile devices. *IEEE Transactions on Mobile Computing* 8, 6 (2009), 792–806.

[28] Microphone Basics 2022. *Comparing MEMS and Electret Condenser (ECM) Microphones*. Retrieved Jan. 12, 2022 from https://www.cuidevices.com/blog/comparing-mems-and-electret-condenser-microphones.

[29] Seyed Iman Mirzadeh, Mehrdad Farajtabar, Ang Li, Nir Levine, Akihiro Matsukawa, and Hassan Ghasemzadeh. 2020. Improved knowledge distillation via teacher assistant. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 5191–5198.

[30] Qiongyao Peng, Jingsi Chen, Tao Wang, Xuwen Peng, Jifang Liu, Xiaogang Wang, Jianmei Wang, and Hongbo Zeng. 2020. Recent advances in designing conductive hydrogels for flexible electronics. *InfoMat* 2, 5 (2020), 843–865.

[31] R Quian Quiroga, A Kraskov, T Kreuz, and Peter Grassberger. 2002. Performance of different synchronization measures in real data: a case study on electroencephalographic signals. *Physical Review E* 65, 4 (2002).

[32] Singam Bhargav Ram and Vanga Odelu. 2022. Security Analysis of a Key Exchange Protocol under Dolev-Yao Threat Model Using Tamarin Prover. In *12th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 0667–0672.

[33] Marc Roeschlin, Ivan Martinovic, and Kasper Bonne Rasmussen. 2018. Device Pairing at the Touch of an Electrode. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, Vol. 18. 18–21.

[34] Masoud Rostami, Ari Juels, and Farinaz Koushanfar. 2013. Heart-to-heart (H2H): Authentication for implanted medical devices. In *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security (CCS)*. 1099–1112.

[35] Nirupam Roy, Sheng Shen, Haitham Hassanieh, and Romit Roy Choudhury. 2018. Inaudible Voice Commands: The Long-Range Attack and Defense. In *15th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*. 547–560.

[36] Farrukh Sahar. 2013. Tradeoffs between usability and security. *IACSIT International Journal of Engineering and Technology* 5, 4 (2013).

[37] Yifu Sun, Kang An, Junshan Luo, Yonggang Zhu, Gan Zheng, and Symeon Chatzinotas. 2021. Intelligent Reflecting Surface Enhanced Secure Transmission Against Both Jamming and Eavesdropping Attacks. *IEEE Transactions on Vehicular Technology* 70, 10 (2021), 11017–11022.

[38] Iwan Syarif, Adam Prugel-Bennett, and Gary Wills. 2016. SVM parameter optimization using grid search and genetic algorithm to improve classification performance. *Telkomnika* 14, 4 (2016), 1502.

[39] William J Tomlinson, Stella Banou, Shay Blechinger-Slocum, Christopher Yu, and Kaushik R Chowdhury. 2019. Body-guided galvanic coupling communication for secure biometric data. *IEEE Transactions on Wireless Communications* 18, 8 (2019), 4143–4156.

[40] Ramachandran Varatharajan, Gunasekaran Manogaran, Malarvizhi Kumar Priyan, and Revathi Sundarasekar. 2018. Wearable sensor devices for early detection of Alzheimer disease using dynamic time warping algorithm. *Cluster Computing* 21, 1 (2018), 681–690.

[41] Waveshare 2022. *LM386 Sound Sensor*. Retrieved Jan. 12, 2022 from https://www.waveshare.com/wiki/Sound_Sensor.

[42] Yuezhong Wu, Qi Lin, Hong Jia, Mahbub Hassan, and Wen Hu. 2020. AutoKey: Using autoencoder to speed up gait-based key generation in body area networks. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–23.

[43] Fengyuan Xu, Zhengrui Qin, Chiu C Tan, Baosheng Wang, and Qun Li. 2011. IMDGuard: Securing implantable medical devices with the external wearable guardian. In *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 1862–1870.

[44] Weitao Xu, Chitra Javali, Girish Revadigar, Chengwen Luo, Neil Bergmann, and Wen Hu. 2017. Gait-key: A gait-based shared secret key generation protocol for wearable devices. *ACM Transactions on Sensor Networks (TOSN)* 13, 1 (2017), 1–27.

[45] Zhenyu Yan, Qun Song, Rui Tan, Yang Li, and Adams Wai Kin Kong. 2019. Towards touch-to-access device authentication using induced body electric potentials. In *Proceedings of the Annual International Conference on Mobile Computing and Networking (MobiCom)*. 1–16.

[46] Lin Yang, Wei Wang, and Qian Zhang. 2016. Secret from muscle: Enabling secure pairing with electromyography. In *Proceedings of the ACM Conference on Embedded Network Sensor Systems (SenSys)*. 28–41.

[47] ZeroPower Listening 2019. *ZeroPower Listening*. Retrieved Feb. 12, 2022 from https://www.mouser.com/pdfDocs/ApplicationNoteAN2-IntroducingZeroPowerListeningTMusingVM1010_Rev20.pdf.

[48] Guoming Zhang, Chen Yan, Xiaoyu Ji, Tianchen Zhang, Taimin Zhang, and Wenyuan Xu. 2017. Dolphinattack: Inaudible voice commands. In *Proceedings of*

the ACM SIGSAC Conference on Computer and Communications Security (CCS). 103–117.

[49] Jiansong Zhang, Zeyu Wang, Zhice Yang, and Qian Zhang. 2017. Proximity based IoT device authentication. In *Proceedings of IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 1–9.

[50] Junqing Zhang, Yushi Zheng, Weitao Xu, and Yingying Chen. 2021. H2K: A Heartbeat-based Key Generation Framework for ECG and PPG Signals. *IEEE Transactions on Mobile Computing* (2021).

[51] Renyun Zhang, Magnus Hummelgård, Jonas Örtegren, Martin Olsen, Henrik Andersson, and Håkan Olin. 2019. Interaction of the human body with triboelectric nanogenerators. *Nano Energy* 57 (2019), 279–292.