MandiPass: Secure and Usable User Authentication via Earphone IMU

Jianwei Liu[1], Wenfan Song[1], Leming Shen[1], Jinsong Han[1], Xian Xu[1], and Kui Ren[1,2,3]

[1]School of Cyber Science and Technology, Zhejiang University, China
[2]Key Laboratory of Blockchain and Cyberspace Governance of Zhejiang Province, China
[3]Alibaba-Zhejiang University Joint Research Institute of Frontier Technologies, China
liujianwei@stu.xjtu.edu.cn,{wenfansong,lemingshen,hanjinsong,xianxu,kuiren}@zju.edu.cn

***Abstract*—Biometric plays an important role in user authentication. However, the most widely used biometrics, such as facial feature and fingerprint, are easy to capture or record, and thus vulnerable to spoofing attacks. On the contrary, intracorporal biometrics, such as electrocardiography and electroencephalography, are hard to collect, and hence more secure for authentication. Unfortunately, adopting them is not user-friendly due to their complicated collection methods and inconvenient constraints on users. In this paper, we propose a novel biometric-based authentication system, namely *MandiPass*. *MandiPass* leverages inertial measurement units (IMU), which have been widely deployed in portable devices, to collect intracorporal biometric from the vibration of user's mandible. The authentication merely requires user to voice a short 'EMM' for generating the vibration. In this way, *MandiPass* enables a secure and user-friendly biometric-based authentication. We theoretically validate the feasibility of *MandiPass* and develop a two-branch deep neural network for effective biometric extraction. We also utilize a Gaussian matrix to defend against replay attacks. Extensive experiment results with 34 volunteers show that *MandiPass* can achieve an equal error rate of 1.28%, even under various harsh environments.**

***Index Terms*—Inertial Measurement Unit, Biometrics, User Authentication, Deep Learning**

I. Introduction

User authentication plays an essential role in the security-relevant scenarios, such as access control and commercial transaction. With the prevalence of mobile computing, user authentication usually functions as the first defense for the device and system, e.g., unlocking a mobile phone. Prior works have widely adopted PIN-based [1] and pattern lock-based [2] mechanisms, which follow the principle of 'something a person has or knows' [3]. In this case, if someone has the credential, i.e., the 'something', he would be authenticated as the genuine user, no matter who he really is. Therefore, these approaches are vulnerable to many attacks, including the stealing, guessing, and shoulder-surfing attacks [4].

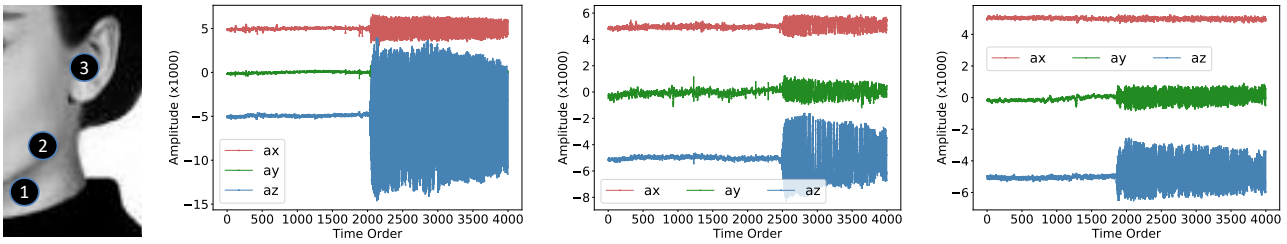
On the other hand, biometric-based authentication is known as 'something a person is or does' [3]. It shows advantages in terms of high security, convenience, non-transferability, and low possibility to be faked or stolen. However, existing pervasively adopted biometrics, including fingerprint, facial feature, and voice-print, are still prone to duplication attacks, because they are easily collected from body surfaces or remote positions. For example, fingerprint can be easily forged and is vulnerable to spoofing attacks [3]. FaceID adopts depth sensor like dot projector and infrared depth camera to improve its security, but it still could be spoofed [5], [6]. Voices can be captured within a relatively large range, thus the voice-based authentication is also vulnerable to replay attacks [7].

Recently, researchers exploit some 'unobtrusive' biometrics for authentication, such as brain waves, cardiac activities, and ear canal features. These biometrics are more secure because they are usually collected from tissues and organs inside human bodies. Capturing, recording, or cloning them is extremely difficult. However, the collection of these biometrics is usually not user-friendly. For instance, users have to pose specific gestures for collecting the cardiac activities, e.g., measurements via electrocardiography (ECG) [5] and photoplethysmography (PPG) [8]. Meanwhile, extra sensing devices lead to inconvenience to users and hence impede adopting these intracorporal biometrics in authentication. For example, stable collection on the electroencephalograph (EEG) requires users to wear cumbersome sensing devices on their heads [9]. Collecting ear canal feature requires deploying extra hardware [5]. Even worse, some of intracorporal biometrics are not stable, e.g., ECG and PPG are susceptible to human motion and emotion changes [8]. Therefore, utilizing the intracorporal biometrics for authentications urgently requests stable, accurate, and easy-to-operate methods for the feature collection and extraction.

Recent years have witnessed the pervasive implementation of inertial measurement units (IMU) in portable devices. Among them, earphone has become one of the most ubiquitous individual computing devices [10]. For instance, WT2 plus earbud [11] has integrated neural network to realize real-time language translation. With these observations, we aim to explore a new biometric inside human body, which can be stably captured by the earphone's IMU, to achieve secure and accurate user authentication. Such an authentication system can also serve as the trusted portable device to securely connect with other devices, such as the smartphones, smart appliances, and autonomous vehicles. In particular, it is well suitable for the hands-free scenarios, e.g., driving and sports.

However, to achieve this goal is challenging. First, it is difficult to find a brand-new biometric inside the human

Jinsong Han is the corresponding author.



(a) Three attaching locations. (b) Location 1: throat. The standard deviation of az is 3805. (c) Location 2: mandible. The standard deviation of az is 1050. (d) Location 3: ear. The standard deviation of az is 761.

Fig. 1. The standard deviation values of three locations.

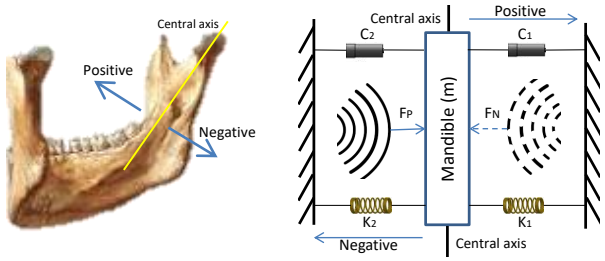
body to meet the user authentication requirements. Second, the sampling rate of common IMU is extremely low (not more than 500Hz [12]) and the raw IMU data contains too much noise, constraining the distinguishability of collected biometric. Last but not least, resisting replay attacks remains an open issue for biometric-based authentications.

In this paper, we propose a novel biometric-based authentication system, namely *MandiPass*. *MandiPass* is based on a new intracorporal biometric, termed as *MandiblePrint*, which is extracted from the vibrations of human's mandibles. *MandiPass* collects *MandiblePrint* via the IMU embedded in the earphone [13]–[15]. During the authentication, a user that wears the earphone only needs to voice 'EMM' for a very short time. The vibration generated by the throat will propagate through the mandible component, reach the ear, and finally be captured by the IMU in the earphone. To validate the feasibility of *MandiblePrint*, we build a one degree-of-freedom theoretical model and conduct a vibration propagation experiment. Moreover, to deal with the challenge of low sampling rate and inferior quality of IMU data, we perform a series of denoising means on raw IMU data and leverage a two-branch deep neural network to extract high-distinguishability *MandiblePrint*. Finally, we utilize a Gaussian matrix to transform *MandiblePrint* into a cancelable template to defend against replay attacks. Once the cancelable template is stolen, user can change the Gaussian matrix to generate a new cancelable template, leading the replay attack to fail due to the dissimilarity between the stolen and new cancelable templates.

We invited 34 participants to perform comprehensive experiments. The results show that *MandiPass* is highly accurate in user verification. The results also demonstrate the effectiveness and robustness of *MandiPass* in real-world scenarios, including using different sides of ears, eating food, and performing different activities.

In summary, our contributions are as follows.

- We propose a secure and user-friendly biometric-based authentication system, *MandiPass*. It leverages a brand-new intracorporal biometric, *MandiblePrint*, which is extracted from the vibration of the mandible.
- We build a theoretical model to prove the feasibility of *MandiblePrint*. We also design a novel two-branch deep neural network for extracting high-distinguishability *MandiblePrint*.



(a) (b)

Fig. 2. The vibration model of the mandible component.

- We implement a prototype of *MandiPass* and conduct experiments with 34 volunteers. The experimental results show that *MandiPass* is robust and secure, with a low equal error rate (EER) of 1.28%.

II. Feasibility Study

In this section, we first validate that the vibration produced by throat can pass through mandible before reaching ear, which enables *MandiPass* to capture vibration signals containing mandible characteristics at earphone. Then a theoretical model is built to study the feasibility of extracting person-distinguishable biometrics from vibration signals.

A. Vibration Propagation Path

MandiPass employs IMUs to capture the desired biometric. A typical IMU contains two components, an accelerometer and a gyroscope. Each component has three axes (x, y, and z) of vibration information, which are time-series real numbers. The x-, y-, and z-axis of the accelerometer are represented by *ax*, *ay*, and *az*, respectively. Likewise, *gx*, *gy*, and *gz* respectively represent the x-, y-, and z-axis of the gyroscope. To validate that the vibration indeed propagates from throat to ear and can be eventually captured by an IMU, we conduct the following experiment. We first attach IMUs to three different locations on a volunteer's head, i.e., to a volunteer's throat, mandible, and ear (shown in Fig. 1(a)). Next, we ask the volunteer to keep silent for a while and then voice an 'EMM' sound to collect the vibration signal. As shown in Fig. 1(b), the standard deviation of *az* is high at the throat location, indicating that the vibration is drastic at the throat. When the vibration propagates along the mandible, the standard deviation of *az* becomes lower, as shown in Fig. 1(c). At the location of ear, as shown in Fig. 1(d), we observe the lowest value of

the standard deviation. The experimental results demonstrate that the vibration generated by the throat can propagate along the path 'throat-mandible-ear', although with a strength decay. During the propagation, the vibration first passes from the throat to the mandible, and then from the mandible to the ear. Moreover, since vibration fades slower in medium with larger density [16], and the density of bone is much larger than that of air and other tissues in human body, the collected vibration signals are mainly composed of vibration components propagating through the mandible. Therefore, the collected vibration signals contain the biometrical feature of the mandible, which is unique to a specific user.

B. Theoretical Model

We model the mandible's vibration based on its physiological structure. When the mandible starts to vibrate, the vibration period can be divided into two phases according to the moving direction of the mandible: positive-direction vibration and negative-direction vibration. These two phases appear alternately. We illustrate our one degree-of-freedom vibration model in Fig. 2. To simplify the model, we neglect the procedure that the mandible moves from the outer vibration boundaries to the central axis (shown in Fig.2(b)).

The m is the mass of the mandible. The c_1 and c_2 are the damping factors of the two dampers. The k_1 and k_2 are the two coefficients of elasticity of the two springs. The vibration resistance, i.e., the dampers and springs, is introduced by the tissues (*e.g.*, muscle and fat) surrounding the mandible. Apparently, the tissues on both sides of the mandible are not symmetrical, we thus have $c_1 \neq c_2$ and $k_1 \neq k_2$.

In the positive-direction phase, the two springs and damper c_1 hinder the positive-direction motion of the mass. Meanwhile, making the mandible vibrate is equivalent to applying a force on the mandible component. Suppose that the positive-direction force caused by the throat vibration is $F_P(t)$. According to the Newton's second law, we have:

$$F_P(t) = mx^{''}(t) + c_1 x^{'}(t) + (k_1 + k_2)x(t), \tag{1}$$

where $x(t)$ is the positive-direction displacement of the mass. After performing Fourier transform and term transposition, we have:

$$X_P(w) = \frac{1 - e^{-iw\Delta t}}{-\frac{imw^3}{F_P(0)} - \frac{c_1 w^2}{F_P(0)} + \frac{i(k_1+k_2)w}{F_P(0)}}, \tag{2}$$

where w, $X_P(w)$, i, and $F_P(0)$ are the frequency component, the spectrum of the vibration signal, the imaginary component, and the constant positive-direction force induced by the positive-direction vibration of the throat, respectively.

If we denote the vibration propagation attenuation coefficient, the propagation distance from throat to ear, and the received positive-direction spectrum at ear as α, d, and $Y_P(w)$ respectively, we obtain the following formula according to [17]:

$$Y_P(w) = X_P(w)e^{-\alpha d}. \tag{3}$$

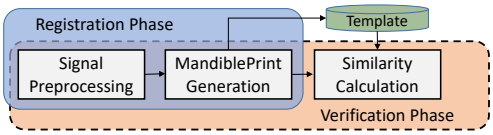


Fig. 3. The architecture of *MandiPass*.

Through replacing the term $X_P(w)$ in Eq. 3 with the right side of Eq. 2, we have:

$$Y_P(w) = \frac{e^{-\alpha d} - e^{-iw\Delta t - \alpha d}}{-\frac{imw^3}{F_P(0)} - \frac{c_1 w^2}{F_P(0)} + \frac{i(k_1+k_2)w}{F_P(0)}}. \tag{4}$$

Likewise, the received negative-direction spectrum can be formulated by:

$$Y_N(w) = \frac{e^{-\alpha d} - e^{-iw\Delta t - \alpha d}}{-\frac{imw^3}{F_N(0)} - \frac{c_2 w^2}{F_N(0)} + \frac{i(k_1+k_2)w}{F_N(0)}}. \tag{5}$$

Thus, the $Y(w)$ of a complete period, which equals to $Y_P(w) \cup Y_N(w)$, can be formulated as:

$$\begin{aligned} Y(w) = & \frac{e^{-\alpha d} - e^{-iw\Delta t_1 - \alpha d}}{-\frac{imw^3}{F_P(0)} - \frac{c_1 w^2}{F_P(0)} + \frac{i(k_1+k_2)w}{F_P(0)}} \\ & \cup \frac{e^{-\alpha d} - e^{-iw\Delta t_2 - \alpha d}}{-\frac{imw^3}{F_N(0)} - \frac{c_2 w^2}{F_N(0)} + \frac{i(k_1+k_2)w}{F_N(0)}}, \end{aligned} \tag{6}$$

in which $\Delta t_1 + \Delta t_2$ equals to the time interval of a vibration period. The m, c_1, $c2$, k_1, and k_2 vary among different persons [18]. Although $F_P(0)$, $F_N(0)$, Δt_1, and Δt_1 are identity-irrelevant noise components, they are relatively stable for a specific person, because human's speaking habit and vocal frequency remain stable after puberty [19], especially when a person only produces a single-tone voice 'EMM'. Hence, the received vibration signals, which record the characteristics of mandible, contain sufficient biometrics (*i.e.*, m, c_1, $c2$, k_1, and k_2) and are potential to be utilized to identify individuals. In this paper, we extract these biometrics, which are termed as *MandiblePrint*, both from positive-direction and negative-direction vibration signals to achieve accurate authentication.

III. System Design

In this section, we first introduce the overview of *MandiPass*, and then detail each module in *MandiPass*.

A. System Overview

As illustrated in Fig. 3, the architecture of *MandiPass* consists of two phases, i.e., registration phase and verification phase. The registration phase contains two modules: *signal preprocessing* module and *MandiblePrint generation* module. The verification phase is not only composed of the two modules contained in the registration phase, but also the *similarity calculation* module.

In the registration phase, user needs to provide a segment of vibration signal to generate cancelable template. Specifically, a user first voices 'EMM' for a short time to collect raw signals. Then the identity-irrelevant components in the raw signals are removed by the *signal preprocessing* module. *MandiPass* obtains a 'clear' signal array from this module. Afterwards, the

MandiblePrint generation module extracts a *MandiblePrint* vector from the signal array. The obtained *MandiblePrint* vector is then multiplied by a Gaussian matrix and becomes a cancelable one. Finally, the cancelable *MandiblePrint* vector is deemed as a *MandiblePrint* template and stored in the secure enclave [20] of the earphone.

In the verification phase, the user initiates a verification request by voicing 'EMM' for a short time. Then the collected raw signals are successively processed by the *signal preprocessing* module and the *MandiblePrint generation* module. After that, the obtained *MandiblePrint* vector and the *MandiblePrint* template stored in the secure enclave are utilized to calculate a similarity in the *similarity calculation* module. If the similarity is larger than a threshold we set in advance, the verification request will be accepted. Otherwise, the verification request will be regarded to be from an illegitimate user and rejected.

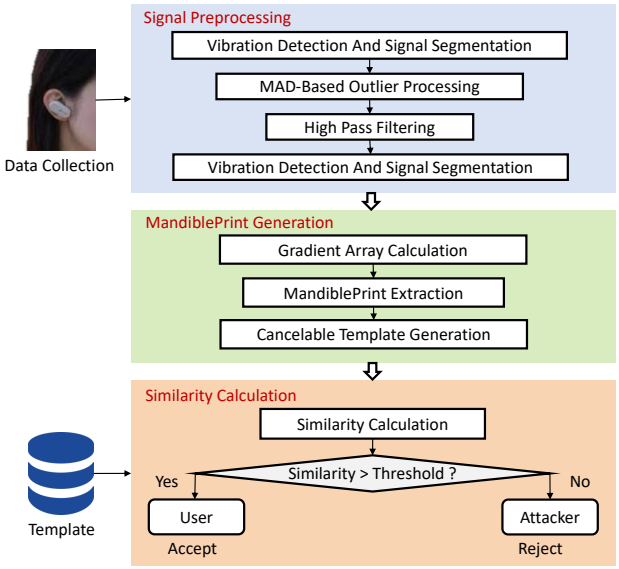


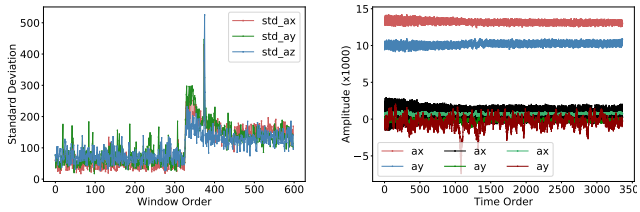
Fig. 4. The workflow of *MandiPass*.

B. Role of Module

The inner operations of the *signal preprocessing*, *MandiblePrint generation*, and *similarity calculation* modules (as shown in Fig. 4) are elaborated in this part.

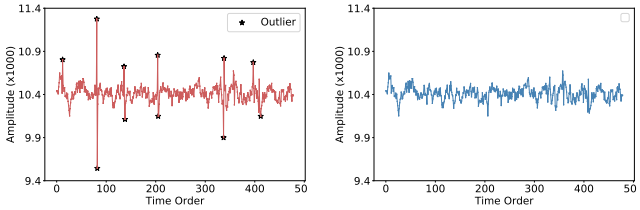
Signal preprocessing: This module is used to remove identity-irrelevant components from raw signals. To this end, *MandiPass* needs to perform four operations. First, *MandiPass* detects the start timestamp of the vibration event. Then, the outliers caused by hardware imperfection and body motion are localized by *MandiPass*. These outliers will be replaced by the mean values of their adjacent normal values. After that, *MandiPass* leverages a high pass filter to remove the noise caused by human movements. Finally, the signal is normalized and the signal values in each axis are concatenated together to form a two-dimensional signal array. The details of each operation are introduced in Section IV.

MandiblePrint generation: This module primarily contains three operations and *MandiPass* obtains a cancelable



(a) The standard deviation becomes large when vibration starts. (b) The beginning values of different axes are different.

Fig. 5. The signal standard deviations and start values.



(a) All outliers are detected. (b) The outlier-replaced signal.

Fig. 6. All outliers are replaced with means.

MandiblePrint vector after the three operations. First, *MandiPass* calculates gradients for each axis of signals in the signal array. A gradient array that contains both positive-direction and negative-direction vibration features is obtained through this operation. Afterwards, the gradient array is fed into a metric extractor (a deep neural network) and the biometric extractor outputs a vector, i.e., *MandiblePrint*. Finally, the *MandiblePrint* vector is multiplied by a Gaussian matrix to get a cancelable one. The design of our biometric extractor and the generation method of the cancelable *MandiblePrint* vector are elaborated in Section V.

Similarity calculation: *MandiPass* calculates the cosine distance [21], i.e., the similarity, between the cancelable *MandiblePrint* vector obtained from a verification request and the stored cancelable *MandiblePrint* template in this module. If the similarity is larger than the threshold, it means that the verification request is initiated by the authentic user. The verification request is thus accepted. Otherwise, *MandiPass* rejects the verification request because it is likely to be initiated by an illegitimate user.

IV. Signal Preprocessing

Vibration detection and signal segmentation: To obtain the signal segment that records mandible vibration, we need to find the start timestamp of the vibration in the raw signal. Since the mandible vibration would make the signal values (in each axis, each timestamp corresponds to a signal value) change drastically, which means that the standard deviation of a certain number of continuous signal values would become large, we determine the start timestamp according to the standard deviation. Specifically, we first divide captured accelerometer signal values into windows and then calculate the standard deviation of each window. Each window has ten continuous signal values and the slide stride is also ten signal values. As shown in Fig. 5(a), if the standard deviation of a window is larger than 250 and the standard deviations of the

subsequent windows are not lower than 100, the vibration is regarded to start at this window. In particular, we consider the timestamp of the first signal value of this window as the start timestamp of the vibration event. Next, we select n continuous signal values behind the start timestamp for each axis to get six signal segments.

MAD-based outlier processing: Due to the hardware imperfection of IMU and motion noise (*e.g.*, walk), the collected raw signals may have some values that are extremely large or small, which should be regarded as outliers. To deal with these outliers, we first detect them by a MAD [22] algorithm, and then replace them with means of normal values. To be specific, We first utilize the MAD outlier detection method to detect all outliers in each signal segment alternatively. As shown in Fig. 6(a), all outliers are found (marked by stars) in a segment, which demonstrates that the MAD algorithm is effective. Afterwards, in order to eliminate the impact of outliers, we perform a two-step mean-based outlier replacing on each signal segment, in which we replace each outlier with the mean of its two previous normal values and two subsequent normal values. The replacing result, shown in Fig. 6(b), proves that our two-step mean-based outlier replacing method is effective.

High pass filtering: Since human activities may generate low-frequency components (LFC), which are irrelevant to the *MandiblePrint*, we need to filter these LFC out. According to the research in [17], the frequency components mostly are less than 10Hz during the body movements. Given that normal people's fundamental frequency of vocal vibration varies from 100Hz to 200Hz [23], a high pass filter is needed to preserve the high-frequency components. Therefore, we utilize a high pass four-order Butterworth filter with a cutoff frequency of 20Hz to remove the LPC from each signal segment alternately.

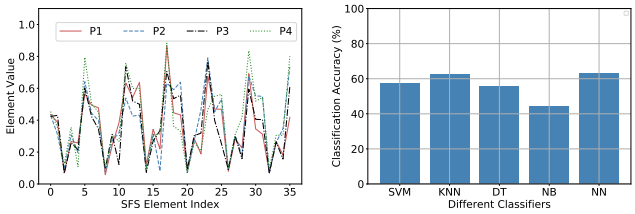
Normalization and multi-axis concatenation: It is noteworthy that the start values of different axes are different, i.e., the elements of some axes oscillate around large values while that of other axes oscillate around small values, as shown in Fig. 5(b). If we directly use un-normalized signals to extract *MandiblePrint*, the contribution of these axes, the values of which are small, would be concealed. Thus, we normalize the signal values through min-max normalization. For each signal segment, the normalized value x_n of each original value x_o can be calculated by:

$$x_n = \frac{x_o - x_{min}}{x_{max} - x_{min}}, \tag{7}$$

where x_{max} and x_{min} are the maximum and minimum values in this signal segment. Moreover, to make full use of captured signals of six axes and provide dimension-consistent input for our biometric extractor, we concatenate six signal segments and obtain a signal array with a dimension of $(6, n)$. Empirically, we set n as 60.

V. MandiblePrint Extraction

In this section, we aim to extract person-distinguishable *MandiblePrint* from the signal array. However, our preliminary experiments show that calculating statistical features is infeasible to extract *MandiblePrint*. We thereby design a novel deep learning model to extract high-quality *MandiblePrint*.



(a) The SFSes of different users are similar.

(b) The accuracy obtained by using statistical features and different classifiers.

Fig. 7. SFS can only achieve low classification accuracy.

A. Infeasibility of Statistical Features

To extract *MandiblePrint*, traditional and intuitive solutions are to calculate some statistical features for each axis. Thus, we conduct a preliminary experiment to explore whether the statistical features of different persons are distinguishable. Specifically, we first invite four volunteers and collect 500 signal arrays for each volunteer. In each signal array, we calculate six common statistical features (*i.e.*, mean, median, variance, standard deviation, upper quartile, and low quartile) for each axis. In this way, we obtain $6 \times 6 = 36$ statistical features for each signal array. Each set of 36 statistical features is called a statistical feature sample (SFS). We then randomly select a SFS for each volunteer and plot the selected four SFSes in Fig. 7(a), where one can find that it is hard to figure out the difference between different SFSes. Further, we label the four volunteers' SFSes by four integers from zero to three. By using 80% SFSes as the training set and the rest 20% ones as the testing set, we utilize four classic classifiers to perform classification: support vector machine (SVM), k-nearest neighbours (KNN), decision tree (DT), naive Bayes classifier (NB), and neural network (NN). The result in Fig. 7 indicates that even the highest classification accuracy is lower than 65%. Therefore, it is infeasible to use statistical features as the *MandiblePrint*.

B. Biometric Extraction

Since convolutional neural networks (CNN) have shown excellent ability of feature extraction [24], we thus attempt to design a CNN-based learning model to mine 'deep-hidden' *MandiblePrint* from signal arrays. Moreover, considering that different biometrics exist in positive-direction and negative-direction vibration signals (according to Eq. 6), we separately perform convolution on these two directions of signals.

In specific, we first separate the positive- and negative-direction vibration signals by calculating gradients for each axis. The i_{th} gradient of the j_{th} axis can be calculated by:

$$g_i^j = \frac{v_{i+1}^j - v_i^j}{|t_{i+1}^j - t_i^j|}, \quad i \in [1, n-1], \quad j \in [1, 6], \tag{8}$$

where v_i^j is the i_{th} signal value of the j_{th} axis, and $|t_{i+1}^j - t_i^j|$ is the normalized time interval between v_{i+1}^j and v_i^j. After

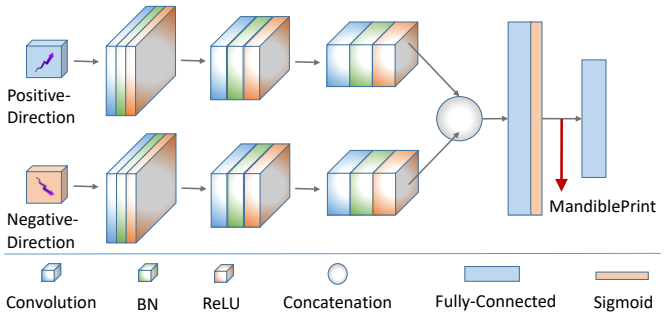


Fig. 8. The architecture of our biometric extractor.

calculating all gradients, we separate them according to their signs, i.e., the gradients that are larger than or equal to zero belong to the positive direction, and the rest gradients belong to the negative direction. In this manner, we obtain approximately $n/2$ gradients for each direction per axis. To provide dimension-consistent inputs for our CNN, we perform linear interpolation to make each direction has $n/2$ gradients. We finally obtain a gradient array with a dimension of $(2, 6, n/2)$, where '2' means the two directions.

Next, we design a two-branch CNN to extract *MandiblePrint* from the gradient array. We notice that the data structure of each axis is time-series values, thus it is reasonable to perform convolution on continuous gradients in each axis to extract temporal features. Meanwhile, since different axes contain different degree-of-freedom features, we also perform convolution among different axes to extract spatial features. Finally, the architecture of our biometric extractor is illustrated in Fig. 8. There are two convolutional branches responsible for extracting temporal-spatial features from the positive- and negative-direction gradients, respectively. Each convolution branch contains three convolutional layers and each of which is followed by a batch normalization (BN) function [25] and a rectified linear unit (ReLU) [26]. The size of each convolutional kernel is 3×3 and the stride size is 1×2. The BN is used to prevent data distribution from offset and the ReLU is used to decrease the inter-neuronal dependence. The BN and ReLU are simultaneously leveraged to improve the effectiveness and robustness of the biometric extractor. After the convolutional operation, we flatten the outputs of the two branches and concatenate them to obtain a feature vector. The feature vector then passes through a fully connected layer and a Sigmoid function [27], and becomes *MandiblePrint*. The output of the Sigmoid function, i.e., *MandiblePrint*, is a biometric vector with a dimension of $(1, 512)$. At last, a fully connected layer is used to project the biometric vector into different classes (*i.e.*, different person IDs), which enables us to train the biometric extractor through loss calculation and back propagation [28].

C. Training Process

To make the biometric extractor learn to effectively extract *MandiblePrint*, we need to train it in a proper manner. However, it is noteworthy that users do not need to provide any vibration signal for the training process, because the biometric extractor is trained by the verification service provider (VSP) (*e.g.*, earphone manufacturer). To be specific, the VSP can hire a large number of people to collect signal arrays. Then these signal arrays are labeled and input to the biometric extractor in a unit of batch. The cross entropy [29] and Adam optimizer [30] can be utilized to calculate loss and update the parameters in the biometric extractor. Once the biometric extractor is well trained, it can be directly deployed on the earphone because it has had the ability of *MandiblePrint* extraction.

VI. Security Enhancement

It is critical to analyze the security of an authentication system. In this section, we first consider four main and potential attacks, and then discuss the defense methods against them.

A. Attack model

Zero-effort attack: In this attack, we assume that the attacker has no awareness of *MandiPass*'s principle. The attacker steals the victim's earphone and attempts to use it to conduct authentication.

Vibration-aware attack: In this attack, our assumption is that the attacker knows the principle of *MandiPass*. The attacker attempts to produce a vibration signal to deceive *MandiPass*.

Impersonation attack: In this attack, we assume that the attacker first observes the verification process of the victim. Then the attacker mimics the voicing manner of the victim to launch the impersonation attack.

Replay attack: Since the vibration propagates inside the human body, it is difficult for the attacker to eavesdrop vibration signals. We assume that the replay attacker steals the *MandiblePrint* template stored in the secure enclave and exhibits it to *MandiPass* to launch the replay attack.

B. Defense

Zero-effort attack analysis: Since user needs to produce a short-time vibration to perform verification in *MandiPass*, the attacker who is not awareness of this principle cannot provide signal array to *MandiPass*. Thus, the attacker cannot pass the verification, which means that *MandiPass* is capable of defending against zero-effort attacks.

Vibration-aware attack analysis: In *MandiPass*, user is accepted if and only if his provided *MandiblePrint* is similar to the template stored in the secure enclave. The attacker is unable to provide such similar *MandiblePrint*, leading the attack to fail. Hence, *MandiPass* can defend against vibration-aware attacks.

Impersonation attack analysis: Even if the attacker is able to mimic the voicing manner of the victim, his *MandiblePrint* is still dissimilar to the victim's one, resulting in the calculated similarity smaller than the threshold. Therefore, the attack will fail and *MandiPass* is also able to defend against impersonation attacks.

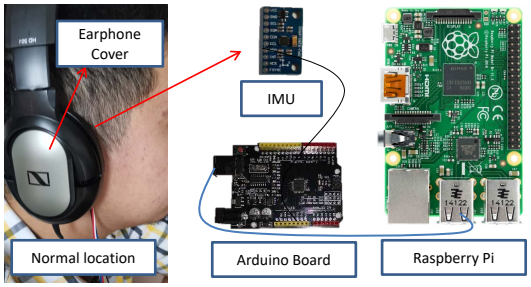


Fig. 9. The experiment setup for MandiPass.

Replay attack defense: To prevent *MandiPass* from replay attacks, we leverage a Gaussian matrix [8] to generate cancelable *MandiblePrint* template. Specifically, the *MandiblePrint* template is transformed by a Gaussian matrix before being stored in the secure enclave in the registration phase. The transformed *MandiblePrint* template is called cancellable *MandiblePrint* template. Let G be a Gaussian matrix and x be a *MandiblePrint* vector. A transformed *MandiblePrint* can be denoted by x' with $x' = x \times G$. In each verification request, the new extracted *MandiblePrint* vector is also transformed as a cancelable one before similarity calculation. In this way, once the cancelable *MandiblePrint* template is stolen, the user can change Gaussian matrix used for transformation, so that the similarity between two *MandiblePrint* vectors transformed by different Gaussian matrices would be smaller than the threshold. The replay attacker, who does not know the changed Gaussian matrix, cannot pass the verification when exhibiting the old template to *MandiPass*. Besides, the attacker cannot calculate the Gaussian matrix by only using the stolen template, which makes the transformation procedure secure. Meanwhile, legitimate authentication would not be impacted since the similarity of two *MandiblePrint* vector transformed by the same Gaussian matrix is still high enough.

VII. Evaluation And Result

We realized *MandiPass* with off-the-shelf devices and conducted extensive experiments to evaluate its performance under real-world environments.

Experiment setup: As shown in the left part of Fig. 9, we built a prototype of *MandiPass* on a Raspberry Pi. This gadget allows us to access the IMU raw data. We used a UNO Arduino board to control the signal collection. While collecting signals, the IMU is attached on the ear by adhesive tapes and covered by a normal earphone cover. We employed two types of IMU, i.e., *MPU-9250* and *MPU-6050*. to conduct experiments. In the default setting, we used *MPU-9250* IMU. The basic frequency of the Raspberry Pi CPU is 160Hz, which is the same as the one in WT2 earbuds and can be achieved by earphone mainboard. The framework used to build the CNN-based biometric extractor is *PyTorch*.

Data collection: We totally invited 34 volunteers (28 males and 6 females) aged from 20 to 45 to participate in our experiments. We collected 23408 signal arrays for overall performance evaluation and each participant provided at least 500 signal arrays. We also collected over 11200 signal arrays in the extensive experiments to evaluate the robustness and security of *MandiPass*.

Metrics: To evaluate the authentication performance quantitatively, we define four metrics: false reject rate (FRR), false accept rate (FAR), EER, and verification success rate (VSR). FRR is the probability that a legitimate user is falsely rejected. It can be represented by the ratio between the number of false rejected signal arrays and the number of all signal arrays. The lower the FRR is, the better performance *MandiPass* has. FRR can be calculated by:

$$\frac{\sum_{i=0}^{V}\sum_{j=0}^{N_i-1}\sum_{k=j+1}^{N_i}\mathbb{1}_{sim(S_i^j,S_i^k)<t}}{\sum_{i=0}^{V}\sum_{j=0}^{N_i-1}\sum_{k=j+1}^{N_i}\mathbb{1}}, \tag{9}$$

where V, t, and N_i are the number of volunteers, the threshold, and the number of signal arrays of the i_{th} volunteer, respectively. The $\mathbb{1}$ equals to one. The $\mathbb{1}_{sim(S_i^j,S_i^k)<t}$ equals to one when the similarity between the *MandiblePrint* vectors extracted from S_i^j and S_i^k is less than t. Otherwise, it equals to zero. The FAR is the probability that an illegitimate user is falsely accepted. It can be represented by the ratio between the number of falsely accepted signal arrays and the number of all signal arrays. The smaller the FAR, the better *MandiPass* is. FAR can be calculated by:

$$\frac{\sum_{i=0}^{V-1}\sum_{j=0}^{N_i}\sum_{k=i+1}^{V}\sum_{l=0}^{N_k}\mathbb{1}_{sim(S_i^j,S_k^l)\geq t}}{\sum_{i=0}^{V-1}\sum_{j=0}^{N_i}\sum_{k=i+1}^{V}\sum_{l=0}^{N_k}\mathbb{1}}. \tag{10}$$

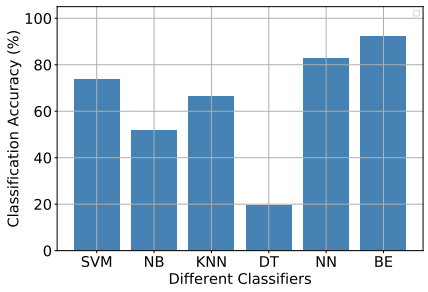
EER is the value of FAR or FRR when FAR equals to FRR. It can be obtained by altering the threshold. The lower the EER is, the better *MandiPass* is. VSR is the probability that a legitimate user is successfully accepted. Higher VSR means better *MandiPass*. It can be calculated by:

$$VSR = 1 - FRR. \tag{11}$$

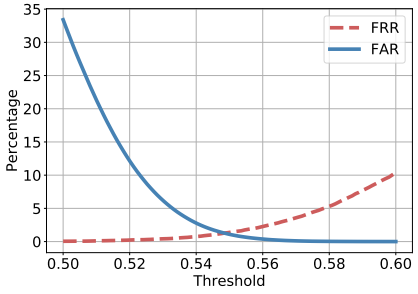
A. Overall Performance

We first evaluated the performance of our biometric extractor by comparing the classification accuracy of different classifiers, i.e., SVM, NB, DT, KNN, NN, and biometric extractor (BE). We randomly selected 80% signal arrays as the training set and the rest 20% ones as the testing set. The classification experiment was performed ten times and we used the mean of ten accuracy as the final classification result. The experimental results are shown in Fig. 10(a). It can be observed that our biometric extractor outperforms other classifiers. It achieves the largest classification accuracy of 90.54%. Therefore, our biometric extractor can effectively extract person-distinguishable mandible biometrics from gradient arrays.

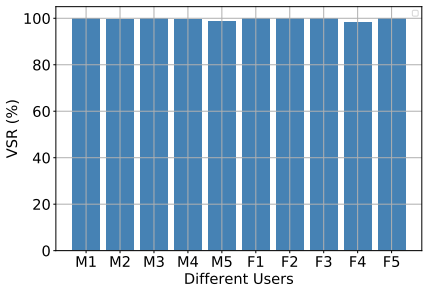
To extract *MandiblePrint*, we treated 33 volunteers' signal arrays as the training set of hired people and extracted the rest volunteer's (plays the role of the user) *MandiblePrint* vectors.



(a) The classification accuracy of different classifiers.

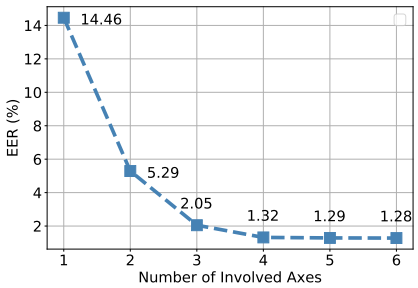


(b) The FAR and FRR curve of *MandiPass*.

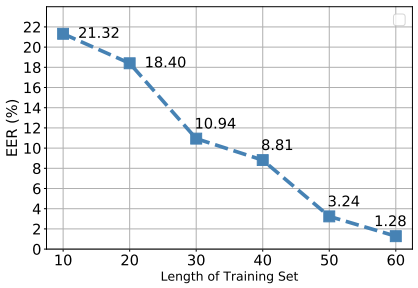


(c) The VSRs of five males and five females.

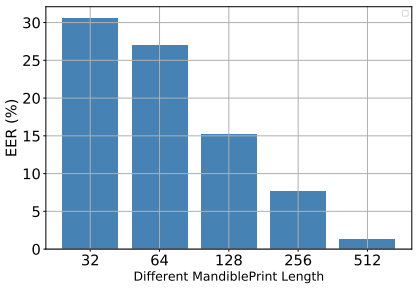
Fig. 10. The overall performance of *MandiPass*.



(a) The effect of number of involved axes.



(b) The effect of training set length.



(c) The effect of *MandiblePrint* length.

Fig. 11. The effect of system setting.

In this way, we extracted *MandiblePrint* vectors of all the volunteers alternatively. We first calculated the mean similarity of a same user and different users. The results indicate that the mean similarity between different *MandiblePrint* vectors of a same user is 0.4884 while that of different users is 0.7032. We then increased the threshold from 0.5 to 0.6 to calculate FAR and FRR. The experimental results are shown in Fig. 10(b). It can be found that when the threshold is 0.5485, the FRR equals to FAR, where we obtain the EER, 1.28%. The low EER demonstrates that *MandiPass* performs significantly well in user verification. In the following experiments, we fixed the threshold to 0.5485.

To explore if the authentication performance is fair to different genders or users, we randomly selected five males and five females and calculated their VSRs. The experimental results, shown in Fig. 10(c), indicate that *MandiPass*'s performance is fair to different genders as well as different users with the same gender.

As aforementioned, we used two types of IMUs for *MandiPass* evaluation, we find that the EERs of *MPU-9250* and *MPU-6050* are 1.28% and 1.29% respectively. There is no apparent EER difference between the two types of IMUs, which shows that *MandiPass* has outstanding device scalability.

B. Effect of System Settings

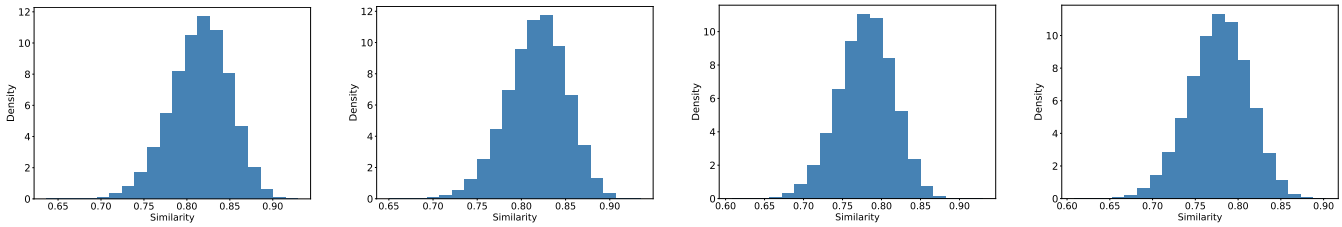
In this part, we evaluated the performance of *MandiPass* under different system settings, including the number of involved axes, the length of the training set, the length of the *MandiblePrint* vector, and the side of the ear (left or right).

The effect of involved axes: In this experiment, we considered the axis order as '*ax*, *ay*, *az*, *gx*, *gy*, *gz*'. The involved axes were selected according to this order. For example, one axis means *ax*, two axes means '*ax*, *ay*', and so on. The experimental results are shown in Fig. 11(a). The results indicate that involving more axes can generate lower EER. Besides, using an accelerometer only can achieve a EER as low as 2.05%.

The effect of training set length: The length of the training set is the time duration of collecting vibration signals for each hired person. We increased the length from 10 seconds to 60 seconds with a stride of 10 seconds. As shown in Fig. 11(b), with the increase of the training set length, the EER keeps decreasing. When the length is 60 seconds, the EER achieves 1.28%. Therefore, collecting one-minute vibration signals for each hired person is sufficient to train the biometric extractor.

The effect of *MandiblePrint* length: It is worth noting that our default *MandiblePrint* length is 512. To explore if the *MandiblePrint* length affects *MandiPass*'s performance, we selected other four commonly used biometric length: 32, 64, 128, and 256. The experimental result shown in Fig. 11(c) indicates that the EER decreases with the increase of *MandiblePrint* length. When the length is 512, the EER is less than 1.5%. Thus, it is reasonable to set the length of *MandiblePrint* as 512.

The effect of ear side: In our default setting, *MandiPass* collects vibration signals from right ears. To validate the feasibility of left ear, we collected a batch of vibration signals from users' left ears. The experimental result shows that the VSR of left ear is as high as 98.02%. Thus, using left ear in



(a) Lollipop. (b) Water. (c) Walk. (d) Run.

Fig. 12. The similarity distributions of lollipop, water, walk, and run. The outer numeric interval is the similarity range and the inner number is the corresponding percentage.

MandiPass is feasible as well.

C. Impacts of Related Factors

We also considered the impacts of four factors from users' daily life. We categorized the factors into two groups, food, and activity.

Food: We took the lollipop and water as the representatives of food since users may use *MandiPass* when they are eating food or drinking. We first conducted an extensive experiment with lollipops, in which we collected testing signal arrays with lollipops in users' mouths. The similarity distribution shown in Fig. 12(a) indicates that lollipop has negligible impact on MandiPass, because all the similarity between the normal signal arrays (without lollipop) and the testing signal arrays (with lollipop) are larger than the threshold. Likewise, we conducted another extensive experiment with water. The similarity distribution shown in Fig. 12(b) proves that water also has negligible impact on *MandiPass* (the VSR is larger than 99%).

Activity: To assess the robustness of *MandiPass* towards human activity, we asked volunteers to walk or run while collecting testing signal arrays. We then calculated the similarity between the normal signal arrays (static) and the testing signal arrays (moving). The similarity distributions shown in Fig. 12(c) and Fig. 12(d) indicate that activity does not affect the performance of *MandiPass*. Thus, *MandiPass* is significantly robust.

D. Effect of Orientation and Tone

Since the orientation of the earphone and the tone of voicing may affect the performance of *MandiPass*, we also evaluated *MandiPass*'s performance with different orientations and tones.

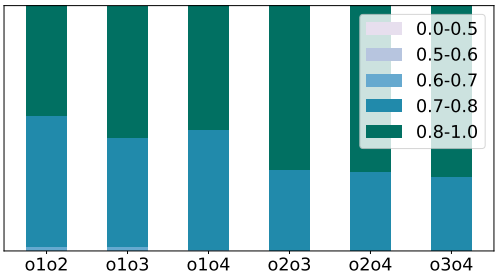


Fig. 13. The effect of tone.

Orientation of IMU: To explore the effect of orientation, we collected four groups of signal arrays and the gap between any two continuous groups is 90 degrees. We then calculated the similarity distributions of signal arrays between any two groups. The results are shown in Fig. 13, which indicate that the similarity of any two signal arrays with different orientations is still higher than the threshold. Therefore, *MandiPass* is robust to the orientation variation.

Tone of voicing: Even if we recommend users to produce 'EMM' voice naturally, users may change their tones unconsciously during authentication, which may further impact the EER of *MandiPass*. Hence, We asked volunteers to raise or lower their tones intentionally when collecting signal arrays in this experiment. Then we calculated the similarity distributions between normal signal arrays (normal tone) and tone-changed ones (high or low tone). The results shown in Fig. 14 indicate that even with a high or low tone, users can still be successfully verified with a high similarity, which proves that *MandiPass* is robust to tone variation as well.

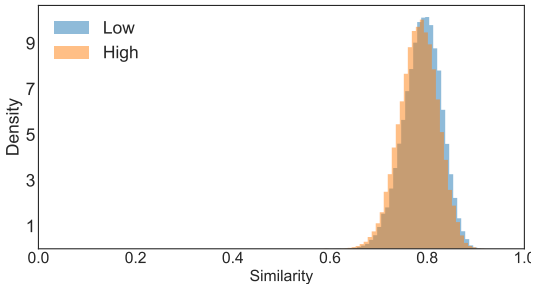


Fig. 14. The effect of tone.

E. Overhead

Time cost: The time cost of *MandiPass* for processing an authentication request mainly comes from three components: vibration signal collection, signal preprocessing, and *MandiblePrint* extraction. First, user needs to voice 'EMM' for a short time to collect vibration signals, which costs 0.2 $(60 \div 350)$ seconds. Second, With the same CPU frequency of WT2 earbud, the signal preprocessing costs less than 0.01 seconds. Finally, with the WT2 earbud's CPU frequency also, biometric extractor outputs an *MandiblePrint* vector within 1 second on average. Therefore, *MandiPass* can process an authentication request with less than 2 seconds and it has outstanding real-time performance.

Storage consumption: The storage consumption of *MandiPass* comes from two components: biometric extractor storage and cancelable *MandiblePrint* template storage. First, the biometric extractor requires approximately 5MB to store

TABLE I
COMPARING *MandiPass* WITH SKULLCONDUCT AND EARECHO.

System	RTC$\leq$ 1s	FRR $\leq$ 2%	RARA	IAN
MandiPass	✓	✓	✓	✓
SKullConduct	✓	×	×	×
EarEcho	×	×	×	×

its parameters. Second, a cancelable *MandiblePrint* template consumes about 1.8KB storage space. Therefore, the total storage consumption is less than 6MB, which is acceptable to an authentication system.

F. Long-Term Observation:

To validate that if *MandiPass* can still authenticate users with a high VSR after a long term, we randomly selected six volunteers to conduct a validation experiment. Specifically, we first collected two batch of signal arrays at time t_1 and t_2, respectively. The time interval between t_1 and t_2 is two weeks. Then we calculated the similarity between the *MandiblePrint* generated by signal arrays collected at t_1 and t_2. The experimental results show that the average VSR of these volunteers is larger than 99.5%. Hence, *MandiblePrint* is stable and *MandiPass* is robust in long term use.

G. Security Assessment

As introduced in Section VI, we need to assess the security of *MandiPass* towards four attack models. In the zero-effort attack experiment, we invited five volunteers (attackers) who do not know the principle of *MandiPass* to initiate authentication requests 20 times per attacker. As a result, the VSR for these attackers is 0%. In terms of the vibration-aware attack, the EER shows that the VSR for attackers is 1.28%. As for the impersonation attack, we first asked five volunteers (attackers) to observe the voicing manners of other five volunteers (victims). Then we collected signal arrays with these attackers. After that, we calculated the similarity between attackers' *MandiblePrint* and victims' *MandiblePrint*. The experimental results show that the VSR for attackers is 1.30%. Finally, to assess the security of *MandiPass* towards replay attacks, we calculate the similarity between cancelable *MandiblePrint* vectors transformed by different Gaussian matrices. The result, a VSR of 0.6%, indicates that nearly all replayed *MandiblePrint* vectors are rejected. Therefore, *MandiPass* can defend against these four types of attacks effectively.

H. Comparing with Existing Works

We compared *MandiPass* with two related works, i.e., SkullConduct [31] and EarEcho [5], in terms of the registration time cost (RTC), EER, replay attack resilience ability (RARA), and immunity against acoustic noise (IAN). SkullConduct is an acoustic signal-based authentication system collecting skull biometrics as authentication credential, which can be deployed on GoogleGlass. EarEcho, a state-of-the-art earphone-based authentication system, collects ear canal biometrics to identify individuals. The comparing results are shown in Table 1. First, *MandiPass* and SkullConduct can finish the registration within one second, but EarEcho does not have such ability. Secondly, the FRR of *MandiPass* is lower than that of SkullConduct and EarEcho. Thirdly, *MandiPass* can defend against replay attacks, while the other two systems cannot. Finally, *MandiPass* is immune to acoustic noise, but the other two systems are susceptible to acoustic noise. Thus, *MandiPass* outperforms SkullConduct and EarEcho.

VIII. RELATED WORK

Authentication on wearable devices: According to the type of authentication credential, existing authentication on wearable devices can be divided into two categories: knowledge-based (something people remember) and biometric-based (something inherent to people) [32]. For knowledge-based authentication, password and pattern are mostly used authentication credential. For example, Vlaenderen *et al.* [33] develop a pattern-based authentication approach deployed on smartwatch, in which camera is leveraged to input user's secrete pattern. Compared with knowledge-based authentications, biometric-based ones are more secure since biometrics are difficult to be stolen or duplicated. For example, Cao *et al.* [8] collect PPG on smartwatch to achieve secure authentication. Gao *et al.* [5] propose to pack a microphone on earphone to extract human's ear canal features. However, existing biometrics collected via wearables either are susceptible to ambient noise or are difficult to be captured. In this paper, we devise *MandiPass* to achieve robust and easy-to-use authentication via earphone IMU.

IMU-based sensing on wearable device: Wearable devices can be fixed at different parts of human body [32], such as head, limb, and back. Various IMU-based sensing technologies are developed at these body parts. For instance, Hwang *et al.* [34] demonstrate that a single head-worn IMU can be used to evaluate interpersonal activities. Based on individual gait events, it can also measure coordination between two gait patterns. To help those who suffering from tetraplegia to use an easy control method for some certain activities, Severin *et al.* [35] place the IMUs on top of headphones to extract features from head movement data. With a slim pedestrian dead reckoning (PDR) sensor bound on the shoe, Gupta *et al.* [36] realize real-time indoor localization in GPS denied environment. Abyarjon *et al.* [37] show that with proper development using sensor fusion algorithms, an IMU-based prototype attached to human upper back can continuously monitor the user's behavior, helping the user form a good posture habit. In this paper, we propose an IMU-based secure authentication technique deploying on earphone.

IX. CONCLUSION

To realize a secure and user-friendly biometric-based authentication, we propose *MandiPass*, which extracts biometrics from the vibration of user's mandible. The feasibility of *MandiPass* is validated via a rigorous theoretical model. We introduce deep learning techniques to improve the efficiency and effectiveness of *MandiPass* in both the biometric extraction and verification. The security of *MandiPass* is further

enhanced via cancellable templates and transformation countermeasures. Extensive experiment results over 34 subjects indicate that *MandiPass* is highly accurate, robust, and secure in various environments.

X. Acknowledgement

This work is supported in part by the National Natural Science Foundation of China (Grants No.: 61872285, 62032021, and 61872081), Research Institute of Cyberspace Governance in Zhejiang University, Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang (Grant No. 2018R01005), Zhejiang Key R&D Plan (Grant No. 2019C03133), major project of the National Social Science Foundation under Grant 20ZDA062, and Alibaba-Zhejiang University Joint Research Institute of Frontier Technologies.

References

[1] S. Rajarajan and P. Priyadarsini, "UTP: a novel PIN number based user authentication scheme," *International Arab Journal of Information Technology*, vol. 16, no. 5, pp. 904–913, 2019.

[2] P. Andriotis, G. C. Oikonomou, A. Mylonas, and T. Tryfonas, "A study on usability and security features of the android pattern lock screen," *Journal of Information and Computer Security*, vol. 24, no. 1, pp. 53–72, 2016.

[3] J. Liu, X. Zou, J. Han, F. Lin, and K. Ren, "BioDraw: Reliable multi-factor user authentication with one single finger swipe," in *IEEE/ACM International Symposium on Quality of Service, IWQoS*, 2020.

[4] Y. Song, Z. Cai, and Z. Zhang, "Multi-touch authentication using hand geometry and behavioral information," in *IEEE Symposium on Security and Privacy, S&P*, 2017.

[5] Y. Gao, W. Wang, V. V. Phoha, W. Sun, and Z. Jin, "Earecho: Using ear canal echo for wearable authentication," *Journal of Interactive, Mobile, Wearable and Ubiquitous Technologies, IMWUT*, vol. 3, no. 3, pp. 81:1–81:24, 2019.

[6] W. Xu, J. Liu, S. Zhao, Y. Zheng, F. Lin, J. Han, F. Xiao, and K. Ren, "RFace: anti-spoofing facial authentication using cots rfid," in *IEEE International Conference on Computer Communications, INFOCOM*, 2021.

[7] K. N. R. K. R. Alluri and A. K. Vuppala, "IIIT-H spoofing countermeasures for automatic speaker verification spoofing and countermeasures challenge 2019," in *Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2019.

[8] Y. Cao, Q. Zhang, F. Li, S. Yang, and Y. Wang, "PPGPass: Nonintrusive and secure mobile two-factor authentication via wearables," in *IEEE International Conference on Computer Communications, INFOCOM*, 2020.

[9] F. Lin, K. W. Cho, C. Song, W. Xu, and Z. Jin, "Brain password: A secure and truly cancelable brain biometrics for smart headwear," in *ACM International Conference on Mobile Systems, Applications, and Services, MobiSys*, 2018.

[10] "Earphones: The next significant platform after smartphones," http://talks.cam.ac.uk/talk/index/128890, 2019.

[11] "Wt2 plus ai real-time translator earbuds," https://www.timekettle.co/products/wt2-plus, 2020.

[12] Z. Ba, T. Zheng, X. Zhang, Z. Qin, B. Li, X. Liu, and K. Ren, "Learning-based practical smartphone eavesdropping with built-in accelerometer," in *Network and Distributed System Security Symposium, NDSS*, 2020.

[13] APPLE, "The imu in airpods," https://www.apple.com.cn/airpods-pro/specs/, 2020.

[14] Cirmall, "The bma456 accelerometer in tws earphone," https://www.cirmall.com/articles/27711, 2020.

[15] HUAWEI, "The imu in huawei freebuds studio," https://consumer.huawei.com/cn/headphones/freebuds-studio/specs/, 2020.

[17] 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 International Conference on Mobile Computing and Networking, MobiCom*, 2019.

[16] J. F. Hunt, H. Zhang, Z. Guo, and F. Fu, "Cantilever beam static and dynamic response comparison with mid-point bending for thin mdf composite panels," *BioResources*, vol. 8, no. 1, pp. 115–129, 2013.

[18] S. WE, "The gross composition of the body," *Journal of Advances in Biological and Medical Physics*, vol. 4, no. 513, pp. 239–279, 1956.

[19] N. M. Meddy Fouquet, Katarzyna Pisanski and D. Reby, "Seven and up: individual differences in male voice fundamental frequency emerge before puberty and remain stable throughout adulthood," *Journal of Royal Society Open Science*, vol. 3, 2016.

[20] J. POT, "What is apple's "secure enclave", and how does it protect my iphone or mac?" https://www.howtogeek.com/339705/what-is-apples-secure-enclave-and-how-does-it-protect-my-iphone-or-mac/, 2018.

[21] H. Zhang, X. Wang, and Z. He, "Weighted softmax loss for face recognition via cosine distance," in *Biometric Recognition - 13th Chinese Conference, CCBR 2018, Urumqi, China, August 11-12, 2018, Proceedings*, ser. Lecture Notes in Computer Science, vol. 10996. Springer, 2018, pp. 340–348.

[22] Y. Li, Z. Li, K. Wei, W. Xiong, J. Yu, and B. Qi, "Noise estimation for image sensor based on local entropy and median absolute deviation," *Journal of Sensors*, vol. 19, no. 2, p. 339, 2019.

[23] M. Geiger, D. Schlotthauer, and C. Waldschmidt, "Improved throat vibration sensing with a flexible 160-ghz radar through harmonic generation," in *IEEE/MTT-S International Microwave Symposium*, 2018.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[25] C. Garbin, X. Zhu, and O. Marques, "Dropout vs. batch normalization: an empirical study of their impact to deep learning," *Multim. Tools Appl.*, vol. 79, no. 19-20, pp. 12 777–12 815, 2020.

[26] R. Arora, A. Basu, P. Mianjy, and A. Mukherjee, "Understanding deep neural networks with rectified linear units," in *6th International Conference on Learning Representations, ICLR*, 2018.

[27] G. Mourgias-Alexandris, G. Dabos, N. Passalis, A. Tefas, A. Totovic, and N. Pleros, "All-optical recurrent neural network with sigmoid activation function," in *IEEE Optical Fiber Communications Conference and Exhibition, OFC*, 2020.

[28] A. Mukherjee, D. K. Jain, P. Goswami, Q. Xin, L. Yang, and J. J. P. C. Rodrigues, "Back propagation neural network based cluster head identification in MIMO sensor networks for intelligent transportation systems," *IEEE Access*, vol. 8, pp. 28 524–28 532, 2020.

[29] L. Li, M. Doroslovacki, and M. H. Loew, "Approximating the gradient of cross-entropy loss function," *IEEE Access*, vol. 8, pp. 111 626–111 635, 2020.

[30] X. Jiang, B. Hu, S. C. Satapathy, S. Wang, and Y. Zhang, "Fingerspelling identification for chinese sign language via alexnet-based transfer learning and adam optimizer," *Sci. Program.*, vol. 2020, pp. 3 291 426:1–3 291 426:13, 2020.

[31] S. Schneegass, Y. Oualil, and A. Bulling, "Skullconduct: Biometric user identification on eyewear computers using bone conduction through the skull," in *ACM Conference on Human Factors in Computing Systems, CHI*, 2016.

[32] A. Bianchi and I. Oakley, "Wearable authentication: Trends and opportunities," *Information Technology and Management*, vol. 58, no. 5, pp. 255–262, 2016.

[33] W. V. Vlaenderen, J. Brulmans, J. Vermeulen, and J. Schöning, "Watchme: A novel input method combining a smartwatch and bimanual interaction," in *ACM Conference Extended Abstracts on Human Factors in Computing Systems, Seoul, CHI*, 2015.

[34] T. Hwang, A. O. Effenberg, and H. Blume, "A rapport and gait monitoring system using a single head-worn IMU during walk and talk," in *IEEE International Conference on Consumer Electronics, ICCE*, 2019.

[35] I. C. Severin, D. M. Dobrea, and M. Dobrea, "Head gesture recognition using a 6dof inertial IMU," *International Journal of Computer Communication & Control*, vol. 15, no. 3, 2020.

[36] A. Gupta, I. Skog, and P. Händel, "Long-term performance evaluation of a foot-mounted pedestrian navigation device," in *Annual IEEE India Conference, INDICON*, 2015.

[37] F. Abyarjoo, N. O.-Larnnithipong, S. Tangnimitchok, F. R. Ortega, and A. B. Barreto, "Posturemonitor: Real-time IMU wearable technology to foster poise and health," *Design, User Experience, and Usability: Interactive Experience Design*, vol. 9188, pp. 543–552, 2015.