MAUTH: Continuous User Authentication Based on Subtle Intrinsic Muscular Tremors

Yi Jiang, *Member, IEEE*, Hongzi Zhu, *Member, IEEE*, Shan Chang, *Member, IEEE*, and Bo Li, *Fellow, IEEE*

***Abstract*—Continuous authentication is viewed to be increasingly important for mobile devices, which store a wide range of private data and sensitive information of users. Traditional continuous authentication methods need user inputs (e.g. typing, sliding). In this work, we present MAUTH, a zero-effect continuous authentication scheme for mobile devices. With the built-in motion sensors on commercial off-the-shelf (COTS) devices, MAUTH can continuously extract, classify and verify the unique tremor features of users on how their body intrinsically shakes during the normal use of such devices. As a result, it is extremely difficult if not impossible to reproduce the same set of tremors as individuals differ in their muscle development. We implement MAUTH as a software on Android-based smartphones, which demonstrates that MAUTH is light-weight and unobtrusive to its users. We conduct extensive real-world experiments and trace-driven simulations in controlled and uncontrolled environments on 21 volunteers. The results show that MAUTH is difficult to counterfeit and achieves a low average false positive rate (FPR) of 6.73% under real-world spoofing attacks. Moreover, MAUTH is comfortable to use and can achieve a low average false negative rate (FNR) of 2.2% during uncontrolled and continuous usage of devices, leveraging isolation-forest-based classifiers trained with only 40 training samples.**

***Index Terms*—Biometrics, continuous authentication, mobile devices, muscular tremors.**

I. Introduction

With the ever increasing capabilities of modern devices, such as smartphones, tablets and smart watches, a rich set of complex applications like photography, online banking, emails, messengers, fitness tracking, and online social interactions, are made possible to run on such devices. As a result, it is of vital importance to provide secure protection for the private information (e.g., personal photos, bank accounts, emails and contact list) stored on mobile devices. Conventional one-time authentication approach, such as to input a PIN code or to recognize the fingerprints or the face of a legitimate user via dedicated sensors, is intrinsically insecure in two aspects: 1) password-based schemes are vulnerable to shoulder-surfing attacks [1] while biometrics-based schemes suffer from spoofing attacks; 2) they cannot provide continuous protection during the entire session of usage. Therefore, in order to eliminate the potential security risk, continuous user authentication on mobile devices has been considered as a must.

In the literature, there have been a rich set of continuous authentication schemes targeting on mobile devices. One main category of continuous authentication schemes are *behavioral biometrics* based, requiring intensive user-device interactions, such as keystrokes [2], [3], touching and sliding operations [4], [5] [6], [7], and gaze patterns [8], [9]. The performance of such schemes are therefore application-dependent, and cannot be guaranteed when there are no sufficient and timely inputs available. Another category of continuous authentication schemes utilize *physiological biometrics*, such as electric pulse response [10], cardiac motion [11], body surface vibrations caused by voice [12], and electrical activity caused by muscle contraction [13]. Though these schemes generally do not need users to constantly interact with devices, they are built based on special sensors, such as electrodes, electromyogram (EMG) sensors, high-performance accelerometers, and electrocardiogram (ECG) sensors, which are not readily deployed on most mobile devices. As a result, there does not exist an effective solution, to the best of our knowledge, that can provide continuous protection for common commercial off-the-shelf (COTS) mobile devices, especially when there is no operation on the devices.

Our Approach. In this article, we propose *MAUTH*, a zero-effect continuous user authentication scheme that can be used in most COTS mobile devices. MAUTH is based on a key observation that body tremors, caused by muscles when generating active forces, are intrinsic and inevitable for human beings. Such tremors can be perceived by most mobile devices with a low-end build-in inertial sensor. We find that body tremors contain a rich set of stable and distinctive frequency components, which can be leveraged as a new physiological biometric for user authentication on mobile devices. Inspired by the observation, MAUTH silently extracts the unique tremor features of users on how their body shake when they are using their devices for two purposes: 1) to train efficient and effective classifiers

Manuscript received 5 April 2021; revised 20 January 2023; accepted 6 February 2023. Date of publication 9 February 2023; date of current version 8 January 2024. This work was supported in part by the National Key R&D Program of China under Grant 2018YFC1900700, in part by the National Natural Science Foundation of China under Grant 61972081, in part by the Natural Science Foundation of Shanghai under Grant 22ZR1400200, and in part by Ant Group Research Fund under Grant 2021110892158. *(Corresponding author: Hongzi Zhu.)*

This work involved human subjects or animals in its research. The author(s) confirm(s) that all human/animal subject research procedures and protocols are exempt from review board approval.

Yi Jiang and Hongzi Zhu are with the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, P.R. China (e-mail: yj389@cornell.edu; hongzi@sjtu.edu.cn).

Shan Chang is with the School of Computer Science and Technology, Donghua University, Shanghai 201620, P.R. China (e-mail: changshan@dhu.edu.cn).

Bo Li is with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong, P.R. China (e-mail: bli@cse.ust.hk).

Digital Object Identifier 10.1109/TMC.2023.3243687

in the training phase; 2) to constantly verify the legitimacy of users with well-trained classifiers without being noticed in the testing phase. In essence, MAUTH is a *two-factor* authentication scheme, integrating both behavioral biometric characteristics, i.e., arbitrary holding postures of users, and physiological biometric characteristics, i.e., the associated body Tremors.

Challenges and Contributions. There are three main challenges in the design of MAUTH. The first challenge is how to obtain reliable and distinctive tremor features from arbitrary human body movements and noisy raw accelerometer readings. To address this challenge, we separate user movements into three categories. Obvious limb or body movements and small hand movements are discarded with only subtle vibrations (referred to as *tremor-movements*) remained for use. To obtain reliable tremor features, we conduct *cross fast Fourier transform* (CFFT) to remove random noises in the frequency domain.

The second challenge is how to passively characterize legitimate users as they are not required to cooperate with MAUTH. We adopt a *passive mode* to silently collect user tremor-movements during the normal usage of a device. A classifier is automatically trained when a newly identified tremor-movement of a legitimate user cannot be recognized by all existing classifiers.

The third challenge is how to constantly authenticate users in a timely and computational-cost-efficient way on mobile devices with limited resources. To tackle this challenge, MAUTH incorporates the cost-efficient tree ensemble models as classifiers to minimize the computational cost with supreme performance. Furthermore, phone attitude information is measured and used as an efficient classifier index to significantly reduce the number of tree ensemble models required in one authentication process. In this way, MAUTH achieves about a $32\times$ speedup gain compared with using all models for authentication.

Compared with the state-of-art continuous user authentication schemes, the novelty of MAUTH is four-fold: 1) by utilizing the distinctive musculoskeletal structure of individuals, MAUTH is extremely difficult if not impossible for an imposter to forge the complex body tremors of a legitimate user; 2) MAUTH performs unobtrusive authentication, which does not disturb how legitimate users use their devices; 3) MAUTH, running as a daemon in the background, can constantly verify the legality of a user in the period of a few seconds throughout a long session of various applications; 4) MAUTH only needs a low-end accelerometer that is widely available in most COTS devices, making wide deployment easy. Nevertheless, the limitation of MAUTH is also clear that it can provide protection only when users hold their devices in stable postures. To provide full-time protection, MAUTH can incorporate existing behavior-biometric-based continuous authentication schemes. We implement MAUTH on four Google Nexus 4 phones running Android, and evaluate the performance of MAUTH via both real-world experiments and trace-driven simulations. The results show that MAUTH is difficult to counterfeit and achieves a low average false positive rate (FPR) of 6.73% under real-world spoofing attacks. Moreover, MAUTH is comfortable to use and can achieve a low average false negative rate (FNR) of 2.2% during uncontrolled and continuous usage of devices, leveraging isolation-forest-based classifiers trained with only 40 training samples.

In summary, our major contributions made in this work consists of: 1) human muscular tremors can be leveraged as a new physiological biometric for continuous authentication on mobile devices; 2) a novel unobtrusive and continuous user authentication algorithm and a prototype implementation; 3) a systematic evaluation that shows the high accuracy and strong security of MAUTH.

II. Design Goals and Models

A. Design Goals

In the design of MAUTH, we consider the following desirable properties:

- *Unobtrusiveness:* The authentication mechanism should be user-friendly and unobtrusive in that it should not interfere with the normal usage, nor should it require extra action for authentication.
- *Strong security:* The authentication mechanism should be able to identify legitimate users and detect both deliberate or unintentional attackers with precision.
- *Immediate verification:* The verification needs to be completed within a very short period of time and can be constantly conducted during the whole session of usage.
- *Cost and energy efficiency:* The authentication mechanism should only relies on the most available built-in sensors to gain large-scale deployment. Moreover, given the nature of mobile devices, it has to operate with low computational cost and power consumption.

B. System and Threat Models

MAUTH has minimum requirements on both mobile devices and their users. We consider the following three entities in the system:

- *Mobile devices:* We require such a target device to have a 3D accelerometer, which can constantly measure the motion and attitude of the device. MAUTH has very limited requirements on the computation and storage capabilities of the device and needs no other special hardware. MAUTH functions as long as the device is associated with the user in some way. For example, a smartphone or a tablet held in one hand or both hands, a virtual reality (VR) headset worn on the head, or a smart watch worn on a wrist.
- *Legitimate users:* MAUTH has no special requirements on how a legitimate user uses his/her device. Comparing with those authentication schemes that rely on operations conducted by the user, such as screen scrolling, keystrokes or walking, MAUTH can deal with the situation where no operations are performed.
- *Imposters:* We consider deliberate or/and unintentional attackers attempting to access private information or conducting unauthorized operations on a mobile device. Note that we only consider the situation when the device is stably held. For cases where an imposter keeps typing or moving while accessing an unlocked phone, other existing behavior-biometric-based continuous authentication schemes should be incorporated. We assume that imposters

cannot have physical access to the device during the training stage of MAUTH. Afterwards, imposters have the following three capabilities. First, they can have physical access to the device when it is locked or unlocked. Second, imposters can launch shoulder surfing attacks by spying or even recording the owner when he/she is using the device. Third, imposters have necessary equipment and technologies to mount biometrics hacking attacks.

III. Preliminaries

A. Rationale of Muscular Tremors

A muscle contains contractile muscle fibers that are embedded within a network of elastic connective tissues [14], [15]. When stretching a muscle by extending a joint, connective tissues are elongated and generate a springlike resistance, which is referred to as *passive tension*. In addition to passive tension, muscle fibers are uniquely designed to contract in response to a stimulus from the nervous system, generating *active force*. In particular, to generate active force, muscle is activated by impulses that are generated within the nervous system, specifically by alpha motoneurons. Each alpha motoneuron has an axon that connects with multiple muscle fibers, forming a motor unit. Muscle fibers connected with small motoneurons, called slow motor units, have twitch responses, that are relatively long in duration (i.e., slow response to a stimulus) and small in amplitude (i.e., small generated force). In contrast, muscle fibers connected with large motoneurons, called fast motor units, have twitch responses, that are relatively short in duration (i.e., fast response to a stimulus) and high in amplitude (i.e., large generated force). Moreover, there is an entire spectrum of intermediate motor units that shows physiologic features somewhere between slow and fast motor units.

With this arrangement, the nervous system can produce a muscle force by first *recruiting* motoneurons and then by driving them to higher rates of sequential stimuli, known as *rate coding*. The highest rate of stimulation that the nervous system of human beings can generate is about 50 Hz. According to the particular demands of a task, the nervous system recruits motor units in line with the *Henneman Size Principle*, where smaller motoneurons will be recruited before larger motoneurons. Through the whole spectrum of motor units, the nervous system is able to activate muscle fibers that sustain stable postures over a long period of time, and when needed, produce high and short-duration bursts of force for more impulsive movements. During the active force generation of a muscle, as a spectrum of different types of motion units are recruited and modulated at distinct rates of stimulation, muscle fibers contacts at different rates, leading to subtle muscular tremors in a wide range of frequencies.

B. Feasibility

Considering the presence of inertial sensors readily available on most mobile devices, we investigate muscular tremor as a new biometric for user authentication on those devices.

We first conduct an extensive data collection campaign. Specifically, we collect inertial sensory data with four Google Nexus 4 smartphones, running Android 4.2 (Jelly Bean). On each phone, raw readings along each axis of its 3D accelerometer

(a) (b) (c)

Fig. 1. Three example postures when using a phone.

can be recorded. The sampling frequency is 100 Hz. We recruit 21 volunteers, six females and fifteen males, aged from 20 to 45, including five undergraduate students, ten graduate students, three faculty members, and three office staff. In general, each volunteer helps collect their tremor data for three times a day, i.e., in the morning, after lunch, and in the evening. For each time, each volunteer is asked to use an experiment phone in six postures, i.e., SLR (*Sitting* with the phone held *Low* in the *Right* hand), SHR (*Sitting* with the phone held *High* in the *Right* hand), SLL (*Sitting* with the phone held *Low* in the *Left* hand), TLR (*sTanding* with the phone held *Low* in the *Right* hand), THR (*sTanding* with the phone held *High* in the *Right* hand), and TLL (*sTanding* with the phone held *Low* in the *Left* hand). Fig. 1 illustrates three example right-hand postures of one male volunteer. In each posture, each volunteer is asked to hold the phone still for three minutes, and to run five commonly used apps of different operational profiles, i.e., messenger, news, Quora for sharing knowledge, Facebook for social networking, and TikTok for viewing short videos, with each app running for three minutes. The data collection campaign lasted for one week from May 22 to May 28 in the year of 2018, resulting to a data set, denoted as **trace** $\mathcal{A}$, of 15,876 pieces of three-minute tremor records.[1]

After data processing (see Section V and Section VI), we plot the normalized frequency response of tremor signals, collected from different volunteers in the same and distinct postures. It can be seen from Fig. 2(a) that the tremor frequency responses of the same volunteer in the same posture are quite similar, though the tremor records are randomly selected over time. In contrast, it is clear to see from 2(b) and 2(c) that, when either the volunteer or the posture is changed, the corresponding tremor frequency response varies significantly. Therefore, it is possible to leverage human muscular tremor in stable postures to authenticate mobile device users.

Furthermore, to learn how people usually operate a device in *uncontrolled* settings, we install our experiment app on volunteers' own phones and collect the acceleration readings if their phones are unlocked and continuously used. Data were collected for two weeks from July 24 to August 6 in the year of 2020. We denote the collected trace as **trace** $\mathcal{B}$.

[1]The IRB of SJTU considers this research is exempt from HRP approval on the base of the low-risk data collection process.

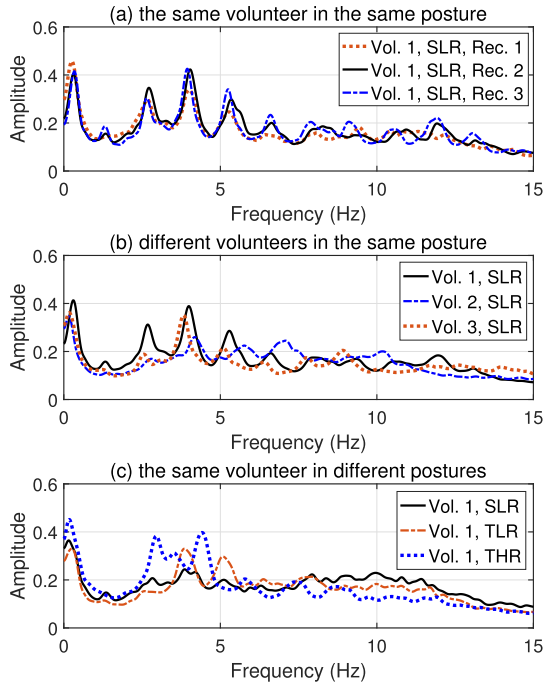

Fig. 2. An illustration of tremor frequency responses of three volunteers in the same and different postures.

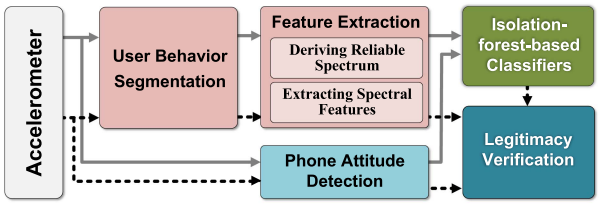

Fig. 3. System architecture of MAUTH, consisting of a training phase, denoted by solid arrowed lines, and a testing phase, denoted by dashed arrowed lines.

IV. Overview of MAUTH

To operate a mobile device, users have to hold the device in a stable way that lets them view its screen while providing input. Meanwhile, for a given holding posture, particular groups of muscles are needed to generate competing forces and inevitable tremors. Based on this key observation, we propose and design MAUTH, which provides *two-factor continuous authentication* for mobile devices, by leveraging the holding postures of users and the corresponding muscular tremors. As illustrated in Fig. 3, the system architecture of MAUTH consists of two phases, i.e., a *training phase* and a *testing phase*, integrating the following five components:

User Behavior Segmentation (UBS). People may carry mobile devices when they are standing, walking, riding a bicycle, or doing just about anything. UBS segments different types of movements of users, and maintains those tremor-movements related to stable holding postures for use in both phases.

Feature Extraction (FE). Forces generated by muscles are controlled by those recruited motoneurons and the stimulus rate acting on them. Therefore, it is natural to inspect muscular tremors in the frequency domain. To this end, FE first conducts

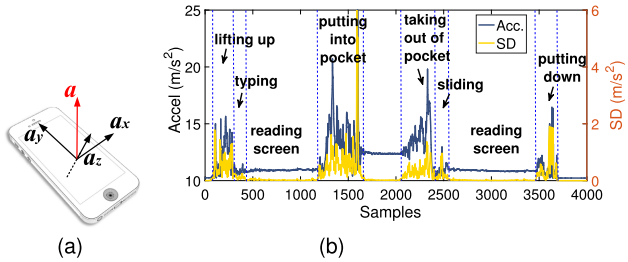

Fig. 4. (a) 3D acceleration readings in the phone coordinate system can be obtained via an onboard accelerometer; (b) human movements are segmented into macro-movements, micro-movements and tremor-movements, based on the windowed standard deviation of the phone acceleration magnitude.

CFFT to obtain reliable spectrum of tremors and then selects major frequency components as features for user authentication.

Classifiers Training and Updating (CTU). For a given holding posture, random frequency components might appear in the spectrum of muscular tremors. As a result, an isolation-forest-based one-class classifiers in the form of an ensemble of weak regression trees are trained and updated along time to distinguish legitimate users and malicious imposters.

User Legitimacy Verification (ULV). The function of ULV is to continuously verify the legitimacy of a user who tries to unlock or use the phone. The decision can be made based on the classification result of one testing posture or the result of multiple voting on individual classification results of multiple postures.

Phone Attitude Detection (PAD). The main function of PAD is to estimate the attitude of the device by calculating the average acceleration components in the device coordinate system. As users are allowed to hold a device in many and arbitrary postures, phone attitude information is used as a classifier index to reduce the number of required classifiers during the training phase and to speed up the verification process during the testing phase.

V. User Behavior Segmentation

MAUTH separates various movements based on the readings of the onboard accelerometer.

Specifically, as depicted in Fig. 4(a), the acceleration of a phone, denoted as $\boldsymbol{a}$, can be decomposed with respect to the device coordinate system. Components of $\boldsymbol{a}$ along x-, y- and z-axis, denoted as a^x, a^y and a^z, respectively, can be measured with an onboard 3D accelerometer. In principle, $\boldsymbol{a}$ is the net result of all forces acting on the device, including the gravity and forces generated by a user. Given that the gravity is nearly constant, therefore, $\boldsymbol{a}$ can reflect the activities of the user. We calculate the measured magnitude of $\boldsymbol{a}$ as the Euclidean norm $\|\boldsymbol{a}\| = \sqrt{(a^x)^2 + (a^y)^2 + (a^z)^2}$. For instance, Fig. 4(b) plots $\|\boldsymbol{a}\|$ in dark color when one of our volunteers normally operates an experiment phone, with (a^x, a^y, a^z) sampled at 100 Hz. In this example, the volunteer picks up the phone from a desk, inputs PIN and reads the screen; then, he puts the phone in his pocket, takes it out and reads the screen again; finally, he puts down the phone on the desk. It can be seen that $\|\boldsymbol{a}\|$ varies

significantly during a series of movements while the variation is relatively mild when the volunteer is in a stable posture.

To distinguish different user movements, instead of using power of acceleration, we calculate the corrected sample standard deviation (SD) of $\|\boldsymbol{a}\|$ using a sliding window as follows,

$$\sigma_t = \sqrt{\frac{1}{W}\sum_{i=t}^{t+W}(\|\boldsymbol{a}_i\| - \overline{\|\boldsymbol{a}\|})^2}, \quad (1)$$

where W is the size of the sliding window; ($\|\boldsymbol{a}_t\|$, $\|\boldsymbol{a}_{t+1}\|$, ..., $\|\boldsymbol{a}_{t+W}\|$) are the measured magnitude of $\boldsymbol{a}$ in the window starting from t; $\overline{\|\boldsymbol{a}\|}$ is the mean value within this window. We empirically take a sliding window of 10 samples (i.e., 0.1 s at the sampling rate of 100 Hz). Fig. 4(b) also plots the corresponding SD of $\|\boldsymbol{a}\|$ in light color. According to the value of σ_t, MAUTH classifies user movements into the following three categories:

1) Macro-movements. Such movements involve obvious limb or body movements, such as lifting up or putting down the mobile device, walking, or riding a bicycle. Macro-movements can lead to huge SD of $\|\boldsymbol{a}\|$. Moreover, the posture of a user may change before and after a macro-movement. For the example in Fig. 4(b), the volunteer may take different postures during the first and the second screen-reading periods. Precisely, we consider that a macro-movement starts when σ_t exceeds a threshold, denoted as ξ_{macro}.

2) Micro-movements. Small movements, such as typing, tapping or sliding on the screen of a mobile device, which only involve palm and finger movements. In fact, though micro-movements would hardly change the current posture of a user, they draw extra energy into the muscular tremor observed in this posture, and, therefore, should also be separated. Similarly, we consider that a micro-movement starts when σ_t exceeds another threshold, denoted as ξ_{micro}, but is less than ξ_{macro}.

3) Tremor-movements. Such movements are the target movements that MAUTH tries to cope with. To operate a device, users have to hold a device in a stable posture that lets them view its screen, while providing input. In this case, the major power of $\|\boldsymbol{a}\|$ stems from muscular tremors. In particular, we consider that a tremor-movement starts when σ_t exceeds a threshold, denoted as ξ_{stable}, but is less than ξ_{micro}. If σ_t is less than ξ_{stable}, the device is considered to be still, such as being placed on a table. It should be noted that, due to hardware noise, σ_t will not be absolute zero when a device is still.

Given a device, it is easy to select appropriate (ξ_{macro}, ξ_{micro}, ξ_{stable}) during an initial training phase of MAUTH (see Subsection VII-A). For instance, in our implementation with Google Nexus 4 smartphones, empirical values $\xi_{macro} = 0.30$, $\xi_{micro} = 0.04$, $\xi_{stable} = 0.01$ are used.

We breakdown the phone usage status for each volunteer in **trace** $\mathcal{B}$ and plot the average results over all volunteers in Fig. 5. It is surprising to see that smartphones are used over nine hours during a day, and tremor-movements account for about 21.58% over all time or about 56.33% when phones are unlocked for operation. Therefore, it is essential to provide security protection when such a device is in stable postures, which however is difficult and unsolved.

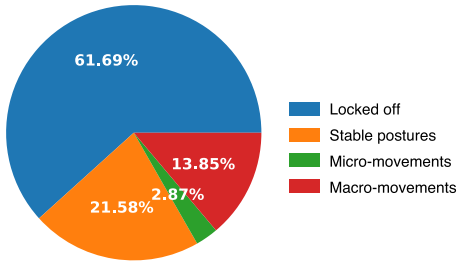

Fig. 5. Device usage status breakdown during one day.

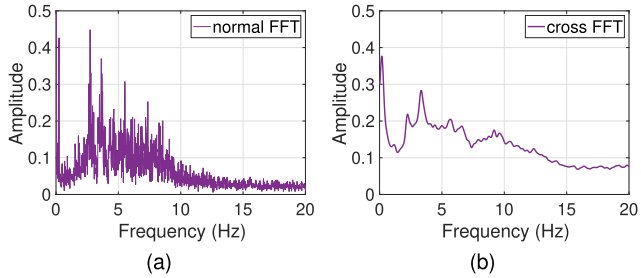

Fig. 6. Frequency spectrums of an example tremor signal of 512 samples, (a) using FFT and (b) using CFFT.

VI. Spectral Analysis of Muscular Tremors

As posture changes lead to macro-movements, we consider that a user and his/her associate devices should stay in a stable posture between two consecutive macro-movements, and investigate the frequency response of muscular tremors during a stable posture in this section.

A. Deriving Reliable Spectrum of Muscular Tremors

Given a stationary time series of $\|\boldsymbol{a}\|$ corresponding to a tremor-movement, denoted as $x = (\|\boldsymbol{a}_0\|, \|\boldsymbol{a}_1\|, \ldots, \|\boldsymbol{a}_n\|)$, we conduct M-window N-point *cross fast Fourier transform* (CFFT) with a stride of J samples on x. Specifically, a sliding window of M samples starting from t, i.e., ($\|\boldsymbol{a}_t\|$, $\|\boldsymbol{a}_{t+1}\|$, ..., $\|\boldsymbol{a}_{t+M-1}\|$), are first normalized and padded with zeros to bring it to a length of N samples, $N > M$, then multiplied by a N-point Hamming window to reduce spectral leakage, and run through a N-point FFT as follows,

$$X\left(e^{j\cdot k\cdot 2\pi/N}\right)\Big|_t^{t+M} = \sum_{i=0}^{M-1} h_i \cdot \|\boldsymbol{a}_{t+i}\| \cdot e^{-j\cdot i\cdot k\cdot 2\pi/N}, k \in [0, N/2] \quad (2)$$

where h_i is the i-th point of the Hamming window; $t \in [0, n-M+1]$ and is a multiple of J. We take the average of each X and obtain a reliable frequency spectrum of muscular tremors.

The reason of conducting above CFFT is three-fold. First, muscular tremor signals in frequency domain are quite noisy. For example, Fig. 6 plots the spectral magnitude of a tremor signal of 512 samples, derived by directly conducting a 512-point FFT and conducting a 100-window 512-point CFFT with a stride of 50 samples, respectively. It is clear that CFFT can significantly

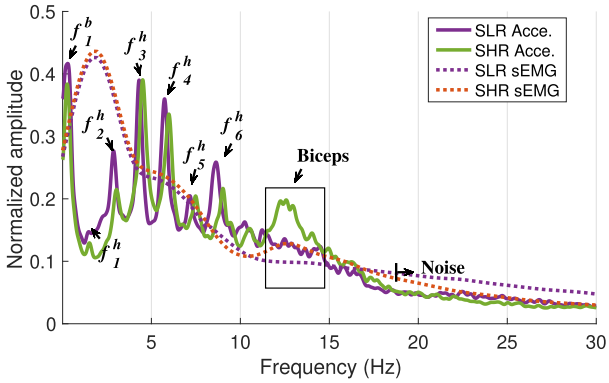

Fig. 7. Major frequency components in spectrum of tremors.

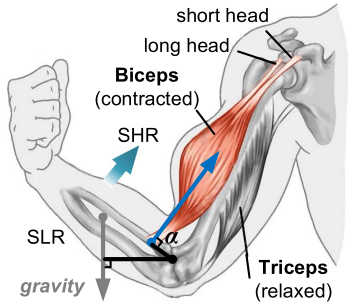

Fig. 8. Competing forces, generated by biceps and gravity, control the angle of the elbow joint.

reduce spectral noise and leads to a more stable spectrum. Second, instead of cutting x into non-overlapping segments, we use a sliding window to obtain more overlapping segments for averaged FFT, which can greatly reduce the required size of x and the response time for online user authentication. Last but not least, a finer frequency spectrum can be obtained by padding tremor samples with zeros before taking each FFT.

B. Extracting Effective Spectral Features

With a derived frequency spectrum of tremors, we first analyze the major components in the spectrum and have the following three observations.

First, breathing and heart beating have significant impact on the vibration signal perceived by a mobile device. For example, we let each volunteer to hold an experiment phone in the SLR and SHR postures, respectively, and measure his/her respiration rate by manually counting and heart rate with a COTS pulse oximeter for three minutes in each posture. Fig. 7 plots the averaged spectrums calculated with CFFT for an example volunteer. In the SLR spectrum, we first find a sharp peak at 0.305 Hz, which nicely corresponds to the measured average respiration rate of 18.3. We denote this fundamental frequency of breath as f_1^b. Next to f_1^b, there is a weak peak at 1.440 Hz, which perfectly matches the measured average heart rate of 86.3. We refer to the peak at 1.34 Hz as the fundamental frequency of heartbeats, denoted as f_1^h. In addition, integer multiples of f_1^h are also found, which are referred to as the second, third, fourth, fifth and sixth harmonic, denoted as f_2^h, f_3^h, f_4^h, f_5^h and f_6^h, respectively. In general, for different people, different heartbeat harmonic patterns in terms of the number of harmonics and their amplitudes can be found, making them valuable for authentication.

Second, muscular tremors have higher frequencies than heartbeat harmonics and vary with different postures. For instance, Fig. 8 illustrates the balance of internal and external torques acting on the elbow joint. The *internal torque* is the product of the force generated by biceps multiplied by the internal moment arm; the *external torque* is the product of gravity and its moment arm. When in the SLR posture, the angle-of-insertion of biceps, denoted as α, is about 90 degrees to the bone and internal moment arm is greatest. In contract, when changed to the SHR posture, the internal moment arm is reduced as α is larger than 90 degrees, which needs biceps to further contract to produce larger force to keep the balance. This change should be reflected in the spectrums of both postures.

To verify this inference, we also measured the power value of the biceps muscle electricity of three of our volunteers when conducting the SHR and SLR postures from September 10 to September 12 in the year of 2022, using a COTS 2-channel surface electromyography (sEMG) module. The sEMG module, consisting of an analog acquisition circuit associated with cables and electrodes, an ADC and a serial communication module, is set with a sampling rate of 1 kHz. We conduct the CFFT operation on all obtained sEMG signals and compare the frequency analysis results of both acceleration and sEMG signals of the same volunteers for comparison. As framed in Fig. 7, it can be seen that there is an obvious peak at around 11-14 Hz in the SHR EMG signal for this volunteer which is aligned with those peaks in the same frequency range in the SHR acceleration signal. In contrast, there is no peak in this range in both SLR signals. We have similar observation for all three volunteers. We consider that the major difference between the SHR and SLR acceleration spectrums is caused by biceps.

Third, after examine all tremor records in **trace** $\mathcal{A}$, we find that frequencies higher than 18 Hz have small amplitude and constitutes a relatively flat region in the spectrum of tremors. It results from the hardware white noise of the accelerometer on the time domain waveform.

In summary, MAUTH chooses the major frequency components, including the fundamental frequency of breath, all harmonics of heartbeats, and those tremor frequency components lower than 18 Hz, as features for user authentication.

VII. Classifier Training

In the training phase, tremor-movements corresponding to one particular holding posture are collected. A set of derived frequency spectrums are used to train a one-class classifier model for this posture based on isolation.

A. Collecting Training Samples

MAUTH adopts both an *active mode* and a *passive mode* to collect training tremor samples. Specifically, in the active mode, a user is first verified through the traditional one-time authentication and then asked to hold a device in a preferred posture (e.g., the SLR posture as shown in Fig. 1) for two minutes, labelled as tremor-movements. The tremor data are divided into segments of

five seconds with each segment recorded as one training sample of the given posture. In addition, the user is also instructed to type and slide on the screen, labelled as micro-movements, and to rest the phone on a table, labelled as being-still, for a few times, respectively. With these labeled behaviors, ξ_{macro} and ξ_{micro} are set to maximal and minimal SD values of micro-movement samples, respectively. ξ_{stable} is set as the maximal SD value of being-still samples. With these learnt thresholds, MAUTH can divide acceleration data into different types of movements.

In the passive mode, MAUTH silently detects tremor-movements of the user when the user is using the device as usual. Once a tremor-movement is identified, the corresponding frequency spectrum is fed into all well-trained classifiers of existing postures for testing. If all classifiers fail in the test, MAUTH locks the device and prompts the user to perform one-time authentication. If the user passes the one-time authentication, this tremor-movement and the successive tremor-movements before a macro-movement are passively recorded for training a new classifier of a new posture.

Initially, MAUTH collects training tremor samples of as many postures as possible in the active mode to boost its availability. Afterwards, MAUTH gradually collects tremor samples of more postures in the passive mode to improve its usability.

B. Constructing Tree Ensemble Models

An effective statistical one-class classification model is required to classify such tremor spectrums. We choose isolation-based model as classifiers for their strong capability to identify anomalies and learn the complex data dependencies. In particular, we adopt the isolation forest (IF) [16], an anomaly detection algorithm that explicitly isolates abnormal data, as it achieves state-of-the art result on many standard classification benchmarks and has a linear time complexity with a low memory requirement.

Specifically, for a given training sample set of a particular posture with n samples of m frequency features $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}(|\mathcal{D}| = n, \mathbf{x}_i \in \mathbb{R}^m, y_i = 1)$, we recursively partition $\mathcal{D}$ by first randomly selecting a feature and then a random split value within the range of the selected feature to build decision trees. In principle, as anomalies lie further away from normal samples in the feature space, anomalies are more susceptible to isolation under random partitioning. Hence, when a forest of random trees collectively produce shorter path lengths for some particular samples, then they are highly likely to be anomalies. An anomaly score s for x_i is defined based on the depth of this sample in all trees in the isolation forest as follows,

$$s(x_i, n) = 2^{-\frac{E(h(x_i))}{c(n)}}, \tag{3}$$

where $h(x_i)$ is the path length of sample x_i in an isolation tree, $E(h(x_i))$ is the average of $h(x_i)$ from a collection of isolation trees, and $c(n)$ is the average path length of unsuccessful search in an isolation tree with n external nodes. $c(n)$ can be calculated as $2H(n-1) - (2(n-1/n))$, where $H(i)$ is the harmonic number and it can be estimated by $ln(i) + 0.5772156649$.

It can be seen that 1) when $E(h(x_i)) \to c(n)$, $s \to 0.5$, the sample x_i does not really have any distinct anomaly; 2) when $E(h(x_i)) \to 0$, $s \to 1$, the sample is definitely an anomaly; 3) when $E(h(x_i)) \to n-1$, $s \to 0$, the sample is quite safe to be regarded as a normal sample. Given the training sample set $\mathcal{D}$, a threshold of s is automatically set so that 95% training samples are considered normal.

VIII. User Authentication

In MAUTH, user authentication is implicitly conducted in the passive training mode. Specifically, when a user is using a device, the frequency spectrum of a newly detected tremor-movement, referred to as a *testing sample*, is fed into pre-trained classifiers for verification.

We have observed that if the user takes a similar posture when using the device, the phone attitude is also similar. Therefore, MAUTH adopts an efficient *attitude indexing method*, where the phone attitude information is used as the classifier index to reduce the number of classifiers required in the verification. Specifically, given the time series of accelerometer readings corresponding to a tremor-movement, denoted as $(\boldsymbol{a}_0, \boldsymbol{a}_1, \ldots, \boldsymbol{a}_{n-1})$, the phone attitude is represented by a triple $(\overline{a^x}, \overline{a^y}, \overline{a^z})$, where $\overline{a^x} = \frac{1}{n}\sum_{i=0}^{n-1} a_i^x$, $\overline{a^y} = \frac{1}{n}\sum_{i=0}^{n-1} a_i^y$, and $\overline{a^z} = \frac{1}{n}\sum_{i=0}^{n-1} a_i^z$. Such phone attitude information is used in both the training phase and the testing phase. In the training phase, the phone attitude of a training sample is measured and stored when a new classifier is trained. In the testing phase, the phone attitude of a testing sample is first compared with the phone attitude of each classifier. If the attitude difference measured with root mean square (RMS) error is less than a given threshold[2], the frequency spectrum of the testing sample is fed to those corresponding classifiers for verification. Otherwise, the testing sample is considered to be associated with a new phone attitude. In this case, the passive training mode as described in Section 7.1 is activated, i.e., the user is required to conduct the traditional one-time authentication and a new classifier would be trained if the user is legitimate.

IX. Evaluation

We evaluate the performance of MAUTH through both trace-driven simulations and real-world experiments. We consider the following two metrics:

- *Accuracy (ACC):* It is the percentage of correct classification results, defined as $\frac{TP+TN}{TP+TN+FP+FN}$ where TP, TN, FP and FN denote the number of true positive samples, the number of true negative samples, the number of false positive samples and the number of false negative samples, respectively. We balance the number of positive samples and negative samples in testing data so that ACC can well measure the accuracy of the authentication system.
- *EER:* It refers to the error rate when false positive rate (FPR) equals false negative rate (FNR) with FPR and FNR defined as $\frac{FP}{FP+TN}$ and $\frac{FN}{FN+TP}$, respectively. The lower the equal error rate value, the higher the accuracy of the authentication system.

For each setting, we repeat that experiment for ten times and present the average ACC and the average EER.

[2]According to our empirical study on **trace** $\mathcal{A}$, 0.11 is an appropriate RMS threshold over all phone attitudes.

TABLE I
PERFORMANCE OF DIFFERENT CLASSIFICATION ALGORITHMS

	IF	RF	SVM	CNN
ACC (%)	96.13	93.52	94.01	96.10
EER (%)	6.68	33.15	26.17	15.22
Training time (ms)	167.13	161.80	168.73	183.55
Testing time (ms)	0.02	0.03	0.34	0.03

A. Classifier Comparison

In this experiment, we compare the performance of IF-based one-class classifier with that of the following three candidate classification algorithms commonly used on mobile devices through trace-driven simulations:

- *Random forest (RF):* We train a random forest classifier consisting of ten decision trees by repeatedly resampling training data with replacement. A consensus prediction is made by voting the trees. A RF classifier is a specific type of bootstrap aggregating that can achieve good classification accuracy at very low computational cost.
- *Support vector machine (SVM):* We train a one-class SVM classifier with the Radial Basis Function (RBF) kernel function for each holding posture of each volunteer using the libSVM [17]. The most appropriate configuration of the two key parameters, i.e., the penalty parameter c and the gamma parameter g, in the RBF kernel function is identified through a grid search [18].
- *Convolutional neural networks (CNN):* Considering the limited computational capability of mobile devices, we adopt a typical CNN structure consisting of one input layer, two one-dimensional convolution layers, one fully connected layer and one output layer with ReLu as the activation function and the cross entropy as the loss function.

Specifically, we divide the tremor data collected when experiment phones are hold still in **trace** $\mathcal{A}$ into two parts, i.e., training set $T = \{D_1, D_2, \ldots, D_5\}$ and testing set $S = \{D_6, D_7\}$, where D_i is the set of tremor records collected on the ith day since May 22. For each holding posture of each volunteer, we randomly select 40 five-second tremor segments from T as positive samples and randomly generate the same number of negative samples to train classifiers, using IF, RF, SVM and CNN, respectively. In the testing, we randomly select 100 five-second tremor segments from S for each holding posture of each volunteer and treat each volunteer as a legitimate user once and treat the rest volunteers as imposters for the current legitimate user. To be fair, we do authentication 100 times for a legitimate user and 5 times for each imposter, which makes the number of tests from the legitimate user and that from all imposters balanced.

Table I lists the average ACC and the average EER over all postures and over all volunteers for each classification algorithm. It can be seen that IF achieves the best performance with an average ACC of of 96.13% and an average EER of 6.68%. Meanwhile, given the same number of training samples, IF also consumes the minimum CPU time for training one single classifier and verifying one single testing sample on a Nexus 4 with a quad-core 1.5 GHz CPU and 2 GB memory.

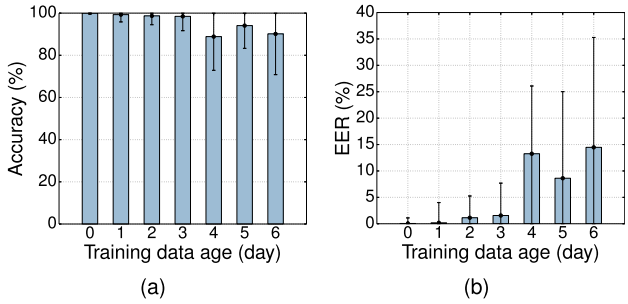

Fig. 9. Impact of history training data.

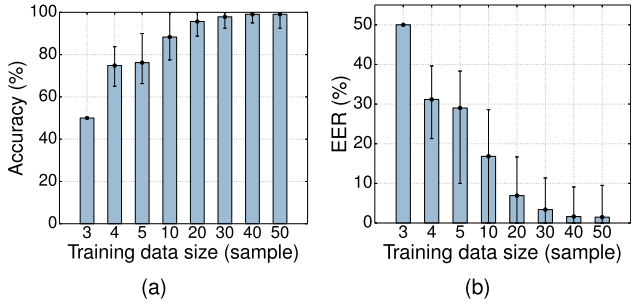

Fig. 10. Impact of training data size.

B. Training Data Age and Size

In this experiment, we first study how human tremors evolve along time. The experiment setting is similar to the above experiment except that we set $T = \{D_1\}$ and $S = \{D_2, D_3, \ldots, D_7\}$. For each holding posture of each volunteer, we build an IF classifier using 40 tremor training samples randomly selected from T. In the testing, we randomly select 100 tremor samples from each S_i for each holding posture of each volunteer with i denoting the age of the training data and the corresponding classifiers. Fig. 9(a) and (b) plot the average ACC and EER as a function of the data age, respectively, calculated by taking the average over all postures and all volunteers. It can be seen that, in general, the performance decreases as the training data ages, especially when the training data are older than three days. In contrast, as shown in Fig. 7, though sEMG signals were measured more than four years later than the acceleration signals from the same volunteer, the frequency peaks in both acceleration and sEMG signals corresponding to the biceps muscle are well aligned, indicating that using tremor frequency peaks as features is stable over time. The reason for this performance drop may lies in the gradual posture changes when collecting tremor data over time, which leads to tremor spectrum changes. This implies that MAUTH can silently collect positive tremor samples for three days before a new classifier is trained for a given posture.

We then study how much tremor data is sufficient to profile a user. From above results, we set $T = \{D_3, D_4, D_5\}$ and $S = \{D_6, D_7\}$ and vary the number of training tremor samples from 3 to 40 with an interval of one sample. Fig. 10(a) and (b) depict the average ACC and the average EER as a function of the number of training samples, respectively. It can be seen that the average EER drops as the training size increases and gradually stabilizes.

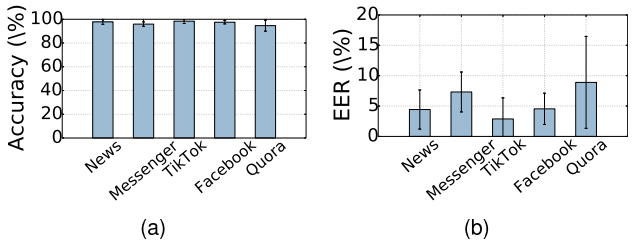

Fig. 11. Different mobile Apps have distinct user interaction patterns, affecting authentication accuracy.

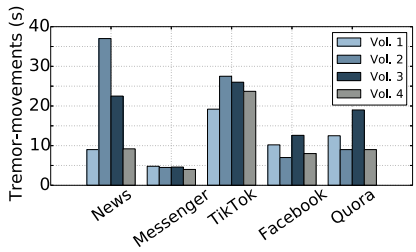

Fig. 12. Tremor-movement duration in different Apps.

When there are more than 40 training samples, the average ACC and the average EER are 99.06% and 1.47%, respectively. As more training samples cannot bring obvious gain in performance but can increase the computation cost for training classifiers, we choose to train tree ensemble models with 40 tremor samples.

C. Impact of Mobile Applications

It should be noted that users are continuously authenticated with MAUTH during tremor-movements. In this experiment, we examine the impact of mobile applications to the authentication accuracy. Different applications have distinct user interaction patterns, which may affect the performance of MAUTH in two aspects: 1) frequent user interactions slightly vary a posture and therefore the extracted frequency features of tremor-movements also change, which harms the authentication accuracy; 2) interactions also interrupt tremor-movements so that it may take longer for MAUTH to collect sufficient tremor data for verification.

The experiment setting is similar to the above experiment except that we use the tremor data in **trace** $\mathcal{A}$ collected when volunteers are using five popular apps, i.e., Messenger, news, Quora for sharing knowledge, Facebook for online social networking, and TikTok for viewing short videos. We set $T = \{D_1, D_2, \cdots., D_5\}$ and $S = \{D_6, D_7\}$ and train classifiers for each app and for each posture of each volunteer.

Fig. 11(a) and (b) depict the average ACC and the average EER as a function of five apps, respectively, calculated by taking the average over all postures and all volunteers. It can be seen that MAUTH achieves supreme authentication accuracy in terms of both ACC and EER for all apps. The average ACC and the average EER over all apps are 96.85% and 5.61%, respectively. However, compared with the accuracy results using tremor data collected when phones are held still, the average ACC drops a little and the average EER also increases slightly. This may be due to posture changes caused by frequent user interactions as mentioned above.

Fig. 12 shows the average duration of a tremor-movement obtained when volunteers are using the five apps on the experiment phones. The results of four volunteers are depicted for illustration. It can be seen that, in general, long tremor-movements are easier to find when users are browsing news or viewing videos while short tremor-movements are mostly seen in Messenger app in that users type messages to friends. In addition, it can also be seen that different volunteers have different content preferences and interaction patterns and therefore have different tremor-movement durations. In any case, MAUTH can provide constant protection for common apps in mobile devices by taking tremor-movement segments of five seconds for verification.

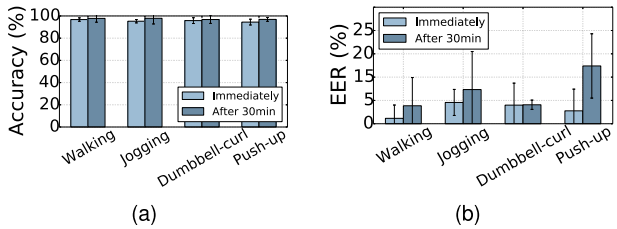

Fig. 13. The impact of two aerobic exercises and two anaerobic exercises is examined.

D. Impact of Physical Exercises

MAUTH basically leverages the biometric characteristics of human muscles for authentication. We investigate whether physical exercises and workouts would affect the performance of MAUTH.

In specific, we randomly select five male volunteers and train classifiers for two postures, i.e., TLR and TLL, for each volunteer using tremor data collected when the volunteers are calm. Then, we collect tremor data of each volunteer in each posture immediately after they have done exercises for five times, deriving a data set denoted as S'. We also collect tremor data half an hour after exercises, deriving another data set denoted as S''. The following four types of exercises are considered: 1) jogging for 2 km at a speed of 8 km/hour; 2) standing 20 lb dumbbell curl for two sets of 12 repetitions; 3) pushing up for two sets of 15 repetitions; 4) walking for 1 km at a speed of 5 km/hour. In the testing, for each holding posture of each volunteer, we randomly select 40 five-second tremor samples from S' and from S'', respectively, and treat each volunteer as a legitimate user once and treat the rest volunteers as imposters for the current legitimate user. To be fair, we do authentication 40 times for a legitimate user and 10 times for each imposter to make the number of positive samples and the number of negative samples even.

Fig. 13 plots the average ACC and average EER of authentication just after and half an hour after doing different exercises. It can be seen that MAUTH generally performs steadily before and after doing exercises. For all the exercises examined, MAUTH achieves better authentication accuracy when testing data in S'' are used. Meanwhile, the EER witnesses a slight increase in S''. We explain that as the breath and heartbeats get faster after exercising, the relation among their harmonics and the main tremor features would be slightly changed and they will

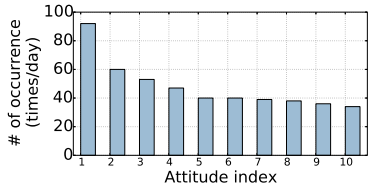

Fig. 14. The average occurrence for the ten most-preferred phone attitudes over all volunteers.

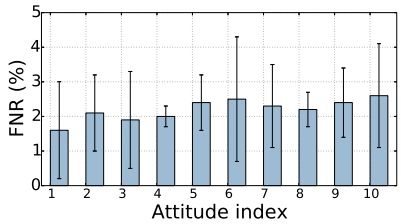

Fig. 15. The average FNR for the ten most-preferred phone attitudes over all volunteers.

get restored once the user calms down (i.e., after 30 minutes). MAUTH achieves good authentication accuracy even for testing data in S'. On the other hand, muscles like biceps and chest muscle get tired just after doing dumbbell curl or push-ups and relieved after half an hour. As MAUTH updates classifiers in three days, it can capture the long-term muscle improvement for fitness enthusiasts.

E. User Experience

We study whether MAUTH may work well in practical settings, where a phone is in continuous usage, by conducting trace-driven simulations using **trace** $\mathcal{B}$. Specifically, for each volunteer, we randomly select a window of three days from his/her trace as training data and the trace of the next day from the window as testing data. Given that we do not know the ground truth of holding postures of volunteers when collecting the trace, we re-train a new classifier for each volunteer when one tremor-movement of a particular phone attitude is identified in **trace** $\mathcal{B}$. We then use all identified tremor-movements of the same phone attitude in the testing data to verify the corresponding volunteer. For each volunteer, we repeat the experiment for ten times and examine the FNRs.

Fig. 14 plots the average number of occurrences per day for the ten most-preferred phone attitudes over all volunteers. It can be seen that in general the occurrence of phone attitudes follows a power law distribution. Fig. 15 plots the average FNRs for the ten most-preferred phone attitudes over all volunteers. The average FNR over all identified attitudes and all volunteers is about 2.2%. The results demonstrates that MAUTH would hardly disturb normal operations of users and has a good user experience. Moreover, with the phone-attitude indexed authentication method, the average number of classifiers involved in an authentication process is significantly reduced from 729 tremor-movements to 23 attitudes, achieving about a 32× speedup gain.

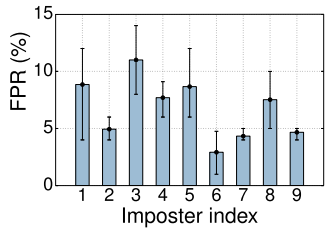

Fig. 16. Spoofing attacks against MAUTH.

F. Real-World Spoofing Attacks

We examine whether MAUTH can defend spoofing attacks via real-world experiments, following the suggestions proposed in [19]. In specific, we randomly select five volunteers, two females and three males, as legitimate users, and nine volunteers, three females and six males, as imposters. For each legitimate user, we first let him/her choose three most comfortable holding postures, and train corresponding tree ensemble models for each posture on one smartphone. Then, we ask each legitimate user to hold the phone in their customized postures for one minute and record the whole process on tape. For imposters, they are allowed to perform live observations on how a legitimate user holds the phone. In addition, they are allowed to watch the taped video as many times as they want as well. We then let imposters rehearse before requiring them to perform twenty authentication attempts.

Fig. 16 shows bar plots of the average FPR of each imposter over all five users. The average FPR over all nine imposters turned out to be 6.73%. The results show that MAUTH is very resilient to spoofing attacks. The human tremors containing a rich set of dynamic features are more complex than most used static biometrics (e.g., face, fingerprint, and iris) and therefore more difficult to forge or reproduce.

G. Power Consumption

In our implementation, UBS adopts a short sliding window of 10 samples to calculate standard deviation over the acceleration signal. FE conducts 100-window 512-point CFFT with a stride of 50 samples over a tremor-movement of 512 samples (i.e., nine 512-point FFTs conducted on acceleration signal of about five seconds) to obtain a reliable frequency spectrum. CTU trains a tree ensemble model using 40 frequency features for one posture. Compared with FE and CTU, the computational cost of UBS can be negligible. The CPU time on the experimental phone with a quad-core 1.5 GHz CPU and 2 GB memory for FE and CTU is 18.06 ms and 110.22 ms, respectively. As CTU is conducted once every three days for a particular posture and the total number of postures are quite limited, the computational cost and power consumption of CTU is low. Therefore, MAUTH does not require mobile devices to have a powerful CPU and has a low power consumption.

H. Limitations

Two main limitations of MAUTH are learnt from above experiments as described in below. First, MAUTH can provide

protection only when users hold/wear their devices in stable postures. However, It is often the case that users may slightly change their habitual postures over time. For instance, from the training data age experiment, it can be seen that the authentication accuracy has a huge drop with more than over 13% EER on average when using the data collected after four days. As shown in Fig. 7, though sEMG signals were measured more than four years later than the acceleration signals from the same volunteer, the frequency peaks in both acceleration and sEMG signals corresponding to the biceps muscle are well aligned, indicating that using tremor frequency peaks as features is stable over time. The reason for this performance drop lies in the gradual posture changes when collecting tremor data over time, which leads to tremor spectrum changes. To deal with this issue, MAUTH tries to train models for stable postures as many as possible and keeps updating models with latest training samples.

Second, in the current stage, MAUTH cannot deal with background vibrations (*e.g.*, walking, jogging, in a car or on a train). The reason is that we need to divide user activities into macro-, micro-, and tremor movements, and only utilize tremor segments for continuous user authentication. One remedy for this is to detect the mobility mode of a user and incorporate existing behavior-biometric-based continuous authentication schemes for full-time protection.

X. Related Work

One-Time Authentication. The conventional user verification scheme is one-time authentication, which is to only verify the user at the start of a session. Common identification mechanisms on mobile devices include using PIN, fingerprint, facial, and iris recognition. Some recent studies also explored other authentication methods. GEAT by Shahzad et al. used the distinguishing features (e.g. finger velocity, device acceleration, stroke time) obtained at the input of password to verify the user [18]. Bichler et al. Mayrhofer et al. and Zhu et al. utilize user's hand shaking features to do user authentication [20], [21], [22]. Also, gait features are explored for authentication [23], [24]. However, these conventional one-time authentication schemes would give adversaries chances to access the system before the user logs out, which leaves much security flaws.

Continuous Authentication. More secure continuous authentication mechanisms are explored to overcome the security flaws of one-time authentication. They mainly fall into two categories.

One is to utilize behavioral biometrics. Keystroke dynamics, namely timing patterns on key pressing and releasing, have been used for continuous authentication by Pinto et al. [2] and Shepherd [3]. Touchpad behaviors have also been explored to continuously verify the user (e.g. Ali et al. [4], Frank et al. [5], Chan et al. [6], Feng et al. [7]). Studies on multi-modal features, which incorporates a set of behavior features (e.g. movement, keystroke, linguistic analysis) have also been carried out by Sitová et al. [25] and Saevanee et al. [26]. Some works utilized eye movement features for continuous authentication (e.g. Eberz et al. [8], Mock et al. [9], Song et al. [27]). However, these methods all require the user to be engaged in the system, either continuously interacting with it via keyboard or touchscreen or watch it.

The other category leverages physiological biometrics. Rasmussen et al. utilized human body's response to an electric pulse for continuous authentication [10]. However, it requires the user to make direct contact with electrodes, which is inpractical and not user-friendly. Feng et al. matched body surface vibration with voice to continuously verify the user [12]. Cardiac Scan by Lin et al. is based on the unique cardiac motion [11]. There are also some work utilizing human muscle information, specifically Electromyogram (EMG, electrical activity caused by human muscle contraction), to realize identification. Venugopalan et al. fused EMG signals with keystroke dynamics for a spoof-resistant authentication system [28]. Belgacem et al. integrated both EMG and electrocardiogram (ECG) to authenticate users [29]. Yang et al. used EMG obtained from user's wrist to generate a secret key and securely authenticate nearby devices [13]. Ataş utilized leap motion devices to capture hand tremor for user authentication [30]. However, these methods mentioned above all require external devices, which is inconvenient, not user-friendly and is limited in real-world application.

XI. Conclusion

In this article, we find that human muscular tremors can be leveraged as a new physiological biometric for continuous authentication on mobile devices. We have proposed an unobtrusive continuous user authentication scheme, called MAUTH, based on human intrinsic muscular tremors. MAUTH relies on a minimum hardware configuration and can be deployed on most COTS mobile devices. By incorporating efficient and effective classifier training and user identity verification algorithms, MAUTH is lightweight and can continuously protect mobile devices at a period of a few seconds, especially when there is no input from the user available. We have implemented MAUTH and conducted intensive trace-driven and real-world experiments. The results demonstrate that MAUTH is accurate and hard to counterfeit in various usage environments.

References

[1] F. Schaub, R. Deyhle, and M. Weber, "Password entry usability and shoulder surfing susceptibility on different smartphone platforms," in *Proc. 11th ACM Int. Conf. Mobile Ubiquitous Multimedia*, 2012, Art. no. 13.
[2] P. Pinto, B. Patrão, and H. Santos, "Free typed text using keystroke dynamics for continuous authentication," in *Proc. IFIP Int. Conf. Commun. Multimedia Secur.*, 2014, pp. 33–45.
[3] S. Shepherd, "Continuous authentication by analysis of keyboard typing characteristics," in *Proc. Eur. Conv. Secur. Detection*, 1995, pp. 111–114.
[4] Z. Ali, J. Payton, and V. Sritapan, "At your fingertips: Considering finger distinctness in continuous touch-based authentication for mobile devices," in *Proc. IEEE Secur. Privacy Workshops*, 2016, pp. 272–275.
[5] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song, "Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication," *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 1, pp. 136–148, Jan. 2013.
[6] A. Chan, T. Halevi, and N. Memon, *Touchpad Input for Continuous Biometric Authentication*, B. De Decker and A. Zúquete Eds., Berlin, Germany: Springer, 2014.
[7] T. Feng, J. Yang, Z. Yan, E. M. Tapia, and W. Shi, "TIPS: Context-aware implicit user identification using touch screen in uncontrolled environments," in *Proc. 15th Workshop Mobile Comput. Syst. Appl.*, 2014, pp. 91–96.

[8] S. Eberz, K. Rasmussen, V. Lenders, and I. Martinovic, "Preventing lunchtime attacks: Fighting insider threats with eye movement biometrics," in *Proc. Annu. Netw. Distrib. Syst. Secur. Symp.*, 2015, pp. 1–13.
[9] K. Mock, B. Hoanca, J. Weaver, and M. Milton, "Real-time continuous iris recognition for authentication using an eye tracker," in *Proc. ACM Conf. Comput. Commun. Secur.*, 2012, pp. 1007–1009.
[10] K. B. Rasmussen, M. Roeschlin, I. Martinovic, and G. Tsudik, "Authentication using pulse-response biometrics," in *Proc. Netw. Distrib. System Secur. Symp.*, 2014, pp. 1–14.
[11] F. Lin, C. Song, Y. Zhuang, W. Xu, C. Li, and K. Ren, "Cardiac scan: A non-contact and continuous heart-based user authentication system," in *Proc. 23rd Annu. Int. Conf. Mobile Comput. Netw.*, 2017, pp. 315–328.
[12] H. Feng, K. Fawaz, and K. G. Shin, "Continuous authentication for voice assistants," in *Proc. 23rd Annu. Int. Conf. Mobile Comput. Netw.*, 2017, pp. 343–355.
[13] L. Yang, W. Wang, and Q. Zhang, "Secret from muscle: Enabling secure pairing with electromyography," in *Proc. 14th ACM Conf. Embedded Netw. Sensor Syst.*, 2016, pp. 28–41.
[14] D. A. Neumann, *Kinesiology of the Musculoskeletal System: Foundations for Rehabilitation*. Oxford, U.K: Elsevier Health Sciences, 2013.
[15] G. Wolf-Heidegger and P. Köpf-Maier, *Systemic Anatomy, Body Wall, Upper and Lower Limbs*. Berlin, Germany: Karger, 2004.
[16] F. T. Liu, K. M. Ting, and Z. Zhou, "Isolation forest," in *Proc. IEEE 8th Int. Conf. Data Mining*, 2008, pp. 413–422.
[17] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, 2011, Art. no. 27.
[18] M. Shahzad, A. X. Liu, and A. Samuel, "Secure unlocking of mobile touch screen devices by simple gestures: You can see it but you can not do it," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw.*, 2013, pp. 39–50.
[19] O. Wiese and V. Roth, "Pitfalls of shoulder surfing studies," in *Proc. Workshop Usable Secur.*, 2015, pp. 1–6.
[20] D. Bichler, G. Stromberg, M. Huemer, and M. Löw, "Key generation based on acceleration data of shaking processes," in *Proc. Int. Conf. Ubiquitous Comput.*, Springer, 2007, pp. 304–317.
[21] R. Mayrhofer and H. Gellersen, "Shake well before use: Intuitive and secure pairing of mobile devices," *IEEE Trans. Mobile Comput.*, vol. 8, no. 6, pp. 792–806, Jun. 2009.
[22] H. Zhu, J. Hu, S. Chang, and L. Lu, "Shakein: Secure user authentication of smartphones with single-handed shakes," *IEEE Trans. Mobile Comput.*, vol. 16, no. 10, pp. 2901–2912, Oct. 2017.
[23] M. O. Derawi, C. Nickel, P. Bours, and C. Busch, "Unobtrusive user-authentication on mobile phones using biometric gait recognition," in *Proc. IEEE 6th Int. Conf. Intell. Inf. Hiding Multimedia Signal Process.*, 2010, pp. 306–311.
[24] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Cell phone-based biometric identification," in *Proc. IEEE 4th Int. Conf. Biometrics: Theory Appl. Syst.*, 2010, pp. 1–7.
[25] Z. Sitov áet al., "HMOG: New behavioral biometric features for continuous authentication of smartphone users," *IEEE Trans. Inf. Forensics Secur.*, vol. 11, no. 5, pp. 877–892, May 2016.
[26] H. Saevanee, N. Clarke, S. Furnell, and V. Biscione, "Continuous user authentication using multi-modal biometrics," *Comput. Secur.*, vol. 53, pp. 234–246, 2015.
[27] C. Song, A. Wang, K. Ren, and W. Xu, "EyeVeri: A secure and usable approach for smartphone user authentication," in *Proc. IEEE 35th Annu. Int. Conf. Comput. Commun.*, 2016, pp. 1–9.
[28] S. Venugopalan, F. Juefei-Xu, B. Cowley, and M. Savvides, "Electromyograph and keystroke dynamics for spoof-resistant biometric authentication," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2015, pp. 109–118.
[29] N. Belgacem, R. Fournier, A. Nait-Ali, and F. Bereksi-Reguig, "A novel biometric authentication approach using ECG and EMG signals," *J. Med. Eng. Technol.*, vol. 39, no. 4, pp. 226–238, 2015.
[30] M. Ataş, "Hand tremor based biometric recognition using leap motion device," *IEEE Access*, vol. 5, pp. 23 320–23 326, 2017.

Yi Jiang (Member, IEEE) received the BS degree from the Department of Computer Science and Engineering, Shanghai Jiao Tong University, in 2018. She is currently working toward the PhD degree with the Department of Computer Science, Cornell University. Her research interests include mainly focus on processing in memory and resource management in data centers.

Hongzi Zhu (Member, IEEE) received the PhD degree in computer science from Shanghai Jiao Tong University, in 2009. He was a post-doctoral fellow with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology, and the Department of Electrical and Computer Engineering, University of Waterloo, in 2009 and 2010, respectively. He is now a professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University. His research interests include mobile sensing and computing, and Internet of Things. He received the Best Paper Award from IEEE Globecom 2016. He was a leading guest editor for IEEE Network Magazine. He is an associate editor for *IEEE Transactions on Vehicular Technology*. He is a member of the IEEE Computer Society, IEEE Communication Society, and IEEE Vehicular Technology Society. For more information, please visit http://lion.sjtu.edu.cn.

Shan Chang (Member, IEEE) received the PhD degree in computer software and theory from Xian Jiaotong University, in 2013. From 2009 to 2010, she was a visiting scholar with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology. She was also a visiting scholar with the BBCR Research Lab, University of Waterloo, from 2010 to 2011. She is now a professor with the Department of Computer Science and Technology, Donghua University, Shanghai. Her research interests include security and privacy in mobile networks and sensor networks.

Bo Li (Fellow, IEEE) received the BEng (summa cum laude) degree in computer science from Tsinghua University, Beijing, and the PhD degree in electrical and computer engineering from the University of Massachusetts at Amherst. He is a professor with the Department of Computer Science and Engineering, Hong Kong University of Science and Technology. He was the chief technical advisor for ChinaCache Corp. (NASDAQ CCIH), the largest CDN operator in China. He was a Cheung Kong visiting chair professor with Shanghai Jiao Tong University (2010–2013) and an adjunct researcher in Microsoft Research Asia (1999–2007) and in Microsoft Advance Technology Center (2007–2009). His current research interests include: multimedia communications, the Internet content distribution, datacenter networking, cloud computing, and wireless sensor networks. He made pioneering contributions in the field of Internet video broadcast with the system, Coolstreaming, which was credited as the worlds first largescale Peer-to-Peer live video streaming system. The work appeared in IEEE INFOCOM in 2005 and received the IEEE INFOCOM 2015 Test-of-Time Award. He has been an editor or a guest editor for over a dozen of IEEE journals and magazines. He was the Co-TPC Chair for IEEE INFOCOM 2004. He received five Best Paper Awards from IEEE. He received the Young Investigator Award from Natural Science Foundation of China (NFSC) in 2005, and the State Natural Science Award (2nd Class) from China in 2011.