Multi-touch Authentication Using Hand Geometry and Behavioral Information

Yunpeng Song†
MOE KLINNS Lab
Xi'an Jiaotong University
Xi'an, China
ypsong@sei.xjtu.edu.cn

Zhongmin Cai†*
MOE KLINNS Lab
Xi'an Jiaotong University
Xi'an, China
zmcai@sei.xjtu.edu.cn

Zhi-Li Zhang
University of Minnesota
Twin Cities, USA
zhzhang@cs.umn.edu

***Abstract*—In this paper we present a simple and reliable authentication method for mobile devices equipped with multi-touch screens such as smart phones, tablets and laptops. Users are authenticated by performing specially designed multi-touch gestures with one swipe on the touchscreen. During this process, both hand geometry and behavioral characteristics are recorded in the multi-touch traces and used for authentication. By combining both geometry information and behavioral characteristics, we overcome the problem of behavioral variability plaguing many behavior based authentication techniques – which often leads to less accurate authentication or poor user experience – while also ensuring the discernibility of different users with possibly similar handshapes. We evaluate the design of the proposed authentication method thoroughly using a large multi-touch dataset collected from 161 subjects with an elaborately designed procedure to capture behavior variability. The results demonstrate that the fusion of behavioral information with hand geometry features produces effective resistance to behavioral variability over time while at the same time retains discernibility. Our approach achieves EER of 5.84% with only 5 training samples and the performance is further improved to EER of 1.88% with enough training. Security analyses are also conducted to demonstrate that the proposed method is resilient against common smartphone authentication threats such as smudge attack, shoulder surfing attack and statistical attack. Finally, user acceptance of the method is illustrated via a usability study.**

Keywords—Multi-touch Gesture; Mobile Authentication; Hand Geometry; Behavioral Variability; Usable Security

I. INTRODUCTION

The convenience of smartphones coupled with its increasing functions (e.g., provided by various APPs) has enabled users to collect and store various kinds of data, much of which are highly personal or sensitive such as photos, emails, phone call logs, chat messages, location traces or even confidential business documents, and access them at anytime and anywhere. As they are carried nearly everywhere we go, smartphones are also prone to be lost or stolen, or subject to unwanted access. Securing mobile devices such as smart phones against unauthorized access is therefore critical in protecting user's personal data and privacy. The most common authentication approach for smart phones is to use a PIN or pattern lock when reactivating the screen. However, such an approach is vulnerable to shoulder surfing [1] and smudge attacks [2].

Biometric authentication is a common approach that has been adopted for addressing this issue. Under such an approach, a person is authenticated using either her physiological information (i.e., physiological biometrics) which is stable and relatively accurate, or behavioral characteristics (behavioral biometrics) which may vary over time. The most popular physiological biometrics used for smartphone authentication is fingerprint. This requires special hardware (fingerprint sensor) installed on smartphone (as well as software and license fee), thus incurs additional cost and is not universally available on all smartphones. As fingerprint is often used for other purposes (in particular, by law enforcement), many users are reluctant to use fingerprint authentication; compromising fingerprint data stored in smartphones could also have severe implications. In addition, many current deployments of fingerprint still incorporate the password mechanism (e.g., TouchID) and researchers have shown that smartphone fingerprint sensors can be fooled with a clay finger [3].

Behavioral biometrics leverage people's operational habits and preferences as identity information, such as gait recognition [4-6], keystroke dynamics [7-11] and gesture based authentication [12-14]. A behavioral biometrics-based authentication typically employs touch screen and/or sensors such as accelerometer and gyroscope that are part of most of today's smartphones to measure a user's behavior characteristics (e.g., gesture, keystroke or gait) for user verification or identification. Hence no additional hardware is needed. Unfortunately, it has been shown [12,14,15] that all existing behavioral biometrics-based authentication methods suffer a crucial problem: users' behavioral variability is uncontrollable, which causes an evident and inevitable performance deterioration over time. This severely undermines the accuracy and user experience of behavioral biometrics-based authentication in real applications.

A. Overview of Approach

In this paper we present a simple and reliable authentication method which combines physiological information of hand geometry with behavioral characteristics. While it is specifically designed for smartphones, it can be

†These authors contributed equally to this work.
*Corresponding Author

also applied to other mobile devices equipped with multi-touch screens such as tablets and touch-screen laptops. The proposed method, referred as *multi-touch authentication with TFST gestures* (here TFST stands for "touch with fingers straight and together"), incorporates physiological information of *hand geometry* with behavioral characteristics by *a specific set of multi-touch gestures* performed on the touch screen. Both hand geometry information and behavioral characteristics (such as the touch pattern, area, time, pressure, etc.) are recorded and used for user authentication. The TFST gestures are designed to be natural and easy-to-perform by common smartphone users. By asking a user to perform multi-touch operations with fingers straight and together, she must stretch her fingers and put them together to perform a multi-touch gesture, with her hand posture conforming to a fixed hand geometry. As a result, this produces a more stable behavioral pattern, thereby significantly reducing behavioral variability plaguing existing behavioral biometrics based authentication methods. The TFST gestures also require much less touch area than existing multi-touch operations proposed in the literature. Hence multi-touch authentication can be deployed on smart phones and mobile devices with smaller screens. Another advantage of TFST gestures is that they only require 0.75s in average to perform. As Harbach et al. showed that mobile authentication solutions lasting longer than 2 seconds are unlikely to be used, given the frequency of daily phone unlocks [16].

To investigate the performance of the proposed gestures in multi-touch authentication, we recruit 161 subjects and collect their multi-touch behaviors over 2 months to establish a large multi-touch authentication dataset. The collection procedure is designed to guarantee that the behavioral variances are recorded in the dataset. Using this dataset, we conduct an extensive evaluation of our approach by analyzing its performance with respect to different gestures, feature sets, classifiers and sizes of training sets. Utilizing the long-term behavioral data collected in our dataset, we also perform a thorough examination of the behavioral variability of TFST multi-touch gestures and its impact on authentication performance. Furthermore, we carry out a security analysis of our multi-touch authentication method under four common types of attacks. Lastly, we investigate user acceptance of the method with a usability study.

B. Contributions

The major contributions of this paper are summarized below:

- We propose a simple, fast, reliable and secure authentication method based on a set of TFST gestures for smartphones and mobile devices equipped with a multi-touch screen. By asking a user to perform multi-touch operations with fingers straight and together, we established a close correspondence between the multi-touch traces and hand geometry. This makes features in multi-touch gestures more stable and the behavioral variability can be largely reduced.
- We design a set of TFST authentication gestures in accordance with their ease of performing on screens of different sizes, and examine their accuracies for authentication. Extensive experiments show that, on 5" or larger screens, the "4-finger L swipe" multi-touch gesture is able to produce an EER (Equal Error Rate) of 5.84% with only 5 training samples and a better EER of 1.88% with enough training; while on smaller 4" screens, a simple "3-finger vertical swipe" gesture produces an EER of 4.10% with sufficient number of training samples. This provides more options for mobile users in search for a tradeoff between security and usability.
- We create a large multi-touch dataset from 161 subjects with an elaborately designed procedure to guarantee that behavior variability over time is captured.
- We also perform a thorough examination of behavioral variability of TFST gestures and their impact on multi-touch authentication by utilizing the long-term data collected. The results show that the fusion of behavioral characteristics with hand geometry information leads to effective resistance to behavioral variability over time.
- We carry out security analysis of our proposed multi-touch authentication under four common types of attacks and perform a usability study to understand user acceptance of the proposed method.

The remainder of paper is organized as follows. In Section II we present the design goals and the threat model. Section III describes the gesture design, feature definition and data collection process. In Section IV and V, we analyze the features and introduce the classifiers for authentication. In Section VI we describe the evaluation framework and discuss the experimental results in Section VII. In Sections VIII and IX we present the security analysis and usability study, respectively. In Section X we provide a brief overview of related work, and discuss the strengths and limitation of the proposed method and future work in Section XI. The paper is concluded in Section XII.

II. Design Goals And Threat Model

In this paper we aim to develop a local, usable and (*sufficiently*) secure authentication method to protect everyday usage of smartphones specifically. The proposed method can also be used to protect other mobile devices equipped with a multi-touch screen such as tablets and touch-screen laptops. Studies [17-19] have shown that smartphones have very different usage patterns from conventional computer systems. They are used frequently (on the average, more than 50 times a day) and often for a short duration. Hence usability is a key requirement for secure user authentication on mobile devices such as smart phones. A secure but inconvenient authentication mechanism will be quickly disabled by most of the users.

A. Design Goals

Taking both security and usability into account, we design a novel multi-touch authentication method that meets the following design goals:

1. Simple, usable and universally applicable: The authentication process should be easy-to-use, fast and

convenient. It should not incur too much cognitive loads on the user. The method should be deployable on most mobile target devices without requiring installation of new hardware components.

2. Reliable: The method should be capable of verifying the legitimacy of a user with high accuracy. Its performance should not deteriorate with the elapse of time.

3. *(Sufficiently)* Secure: The method should be able to protect a user's smartphones for daily usage and secure it against unwanted authentication attempts by a random stranger, an acquaintance or a friend, e.g., when the smartphone is left unattended, lost or stolen. In the following, we will elaborate on the assumptions about the adversary and threat model.

B. Assumptions about the Adversary and Threat Model

As smart phones are frequently used by users, in "insecure" environments, e.g., on a crowded bus where a stranger can observe a user unlocks her phone, or left unattended in the office. They are also prone to get lost or stolen. To secure smartphones for daily usage, we assume that the adversary could be someone who has no personal knowledge of the user but somehow has access to her smartphone, or someone who may interact with the user in certain settings and has the opportunity to observe her phone unlocks or have access to her phone immediately afterwards. In other words, we do *not* assume that the adversary has the power to produce an apparatus (e.g., an artificial hand, or the user's twin sister) with the exact same hand geometry while also being able to observe and replicate the behavior characteristics (hand gestures with the same touch trace and pressure). In particular, our method is designed to secure against the following common types of attacks:

1. Zero-effort Attack: The attacker tries to pass the authentication by chance without any knowledge of the inputs of the legitimate user during authentication.

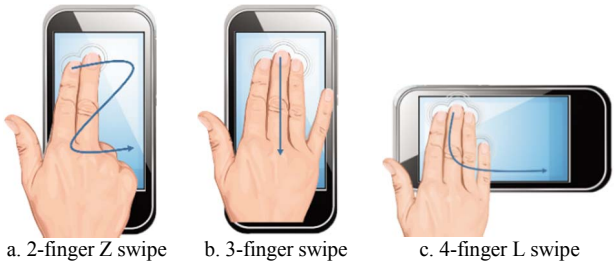
2. Smudge Attack: The attacker utilizes the oily traces left on the screen as a hint to guess the secret to pass the authentication.

3. Shoulder Surfing Attack: The attacker watches the authentication process of a legitimate user and acquires useful hints of the hand gesture to pass the authentication.

4. Statistical Attack: The attacker employs knowledge obtained from the statistics of a group of users as hints to generate authentication attempts.

III. Gesture Design And Data Collection

Multi-touch enabled touchscreens have become a standard configuration for most tablets and phones. Multi-touch enabled laptops and monitors will also become more and more popular. Besides the standard functionality of input, multi-touch behavior is believed to be a rich source of biometric data which implicitly contains information of hand shape [14]. For example, when a user performs five finger pinches or four finger swipes, the traces of touch contains information of lengths of the fingers of the user. However, it is difficult for a user to keep his hand shape unchanged when performing normal touch gestures. He may bend his fingers in one execution while stretch his fingers in another. This leads to changes in hand posture and variations in multi-touch operations, which in turn affects the performance of authentication using multi-touch behaviors. So in order to achieve better performance of multi-touch authentication, we put some restrictions on hand postures which will not affect the user experience while leading to a closer correspondence between the multi-touch traces and hand geometry.



a. 2-finger Z swipe b. 3-finger swipe c. 4-finger L swipe

Fig. 1. Examples of TFST gestures

A. Touch Gesture Design

We introduce a specific set of multi-touch gestures with some restrictions on hand postures, which makes multi-touch traces more closely related to physiological features in hand geometry.

Definition 1. TFST Gestures: A set of multi-touch gestures performed with fingers straight and together. TFST is the abbreviation for "Touching with Fingers Straight and Together".

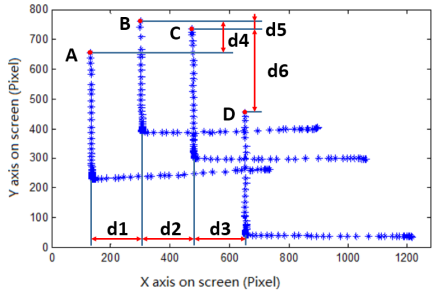
TFST gestures are a specific set of multi-touch gestures performed by adjacent fingers of one hand. As shown in Figure 1, TFST gestures can be performed with two, three or four adjacent fingers of one hand. A TFST gesture may be a simple swipe or a relatively complex pattern like "Z". Actually, in TFST gestures, patterns of touch are not restricted; the only restriction is that users are required to keep fingers straight and together while performing multi-touch operations.

There are two significant advantages of TFST gestures when used for multi-touch authentication. Firstly, when users perform TFST gestures, they must stretch their fingers and put them together. This makes the hand posture conform to a fixed hand geometry, which leads to a more stable behavioral pattern.

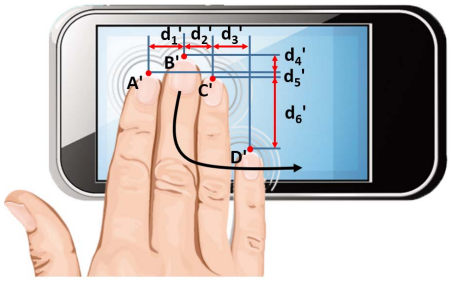
Secondly, TFST gestures require fingers to be together, which requires much less touch area than traditional multi-touch operations. So multi-touch authentication using TFST gestures can be deployed on smaller screen devices such as smartphones, while previous work of multi-touch authentication [14] can only be deployed on large-screen tablets. For example, two or three finger TFST can be performed on 4 inch touch screen (Figure 1a and 1b) and 4-finger TFST gesture can be performed on horizontal 5.3 inch screen (Figure 1c).

B. Features in TFST Gestures

1) Multi-touch Traces: When fingers are sliding on a touch screen, the screen will sample the positions of fingers under a certain frequency and report the information in the form of touch events. For example, Samsung Note 1 (N7000)



a. Physiological features of 4-finger TFST gesture



b. Real features of hand geometry

Fig. 2. Features related to TFST gestures

performs touch sampling at a rate of 60Hz and reports roughly 60 touch events per second. The series of touch events record the moving traces of touching fingers in a multi-touch gesture, which are referred to as **a multi-touch trace** for the gesture. We refer to the touching trajectory of one finger in a multi-touch trace as **a stroke**. A touch event contains information of the XY-coordinates, touch pressure, tool major and touch major of each finger and the timestamp. Tool major and touch major are related to finger size and touching area according to official Android development documents.

2) Physiological Features in TFST Gesture: As shown in Figure 2a, we define 12 distances of the touch trace for a 4-finger TFST gesture: AB, AC, AD, BC, BD, CD, d1, d2, d3, d4, d5 and d6, as physiological features of TFST gestures. A, B, C and D are any set of touch points of the 4 fingers at the same instant in the trace. With respect to edges of the touchscreen, d1, d2 and d3 are horizontal distances between strokes of the multi-touch gesture; while d4, d5 and d6 are vertical distances between strokes of the multi-touch gesture. We assign a feature number from 1-12 to AB, AC, AD, BC, BD , CD, d1, d2 ,d3, d4, d5 and d6 respectively.

Let's assume a user performs the TFST gesture with the directions of his fingers parallel to one edge of the touchscreen, which is the most natural way to perform TFST gestures. Then the 12 physiological features are good measurements of corresponding features of hand geometry. As shown in Figure 2, the distances of d1, d2 and d3 in Figure 2a are good estimations of finger distances of d1', d2' and d3' in Figure 2b. The distances of d4, d5 and d6 are estimations of finger length differences of d4', d5' and d6'. The 6 distances of AB, AC, AD, BC, BD and CD estimate the 6 fingertip distances of A'B', A'C', A'D', B'C', B'D' and C'D'.

Ideally, in the multi-touch trace of a 4-finger TFST gesture, any set of touch points of the 4 fingers at the same instant gives a set of values for the 12 physiological features. In our study, we use the average values for the calculation of the 12 physiological features for a multi-touch trace.

While we can have 12 measurements of hand geometry in the multi-touch trace of 4-finger TFST gesture, in the multi-touch trace of a 2-finger TFST gesture as shown in Figure 1a, we will only have 3 measurements of hand geometry: d1, d4 and AB. In the multi-touch trace of a 3-finger TFST gesture as shown in Figure 1b, we will have 7 measurements: d1, d2, d4, d5, AB, AC and BC.

As the user is required to perform multi-touch operations with fingers straight and together, the 12 measures of his hand geometry are basically unchanged each time he performs a TFST gesture. So the physiological features of TFST gestures are relatively stable and will provide consistent information for identity verification.

3) Behavioral Features in TFST Gestures: Except physiological features, there are many behavioral characteristics in TFST gestures, such as velocity and pressure of touching, and shape of the traces. The behavioral features are defined as the following:

- Length: Length is an important aspect for strokes in a multi-touch trace. We defined 3 length-related features for a stroke: distance (length of a stroke), displacement (displacement between the starting and ending points of a stroke) and ratio of displacement to distance.
- Time: Time is another important aspect of a stroke. We defined 1 time-related feature for a stroke: duration (the duration of touch corresponding to a stroke).
- Velocity: Velocity reflects how fast a user swipes on the screen. We defined 1 velocity-related feature for a stroke: velocity (the mean velocity of the sliding procedure).
- Tool: The tool sequence consists of tool area size for each touch point in a stroke and relates to the size of the touching finger. We defined 2 tool-related features for a stroke: tool mean (the mean of the tool sequence) and tool deviation (the standard deviation of the tool sequence).
- Touch: Touch describes the touch area size of each touch point on the screen. Similar with tool, we defined 2 touch-related features for a stroke: touch mean (the mean of the touch sequence) and touch deviation (the standard deviation of the touch sequence).
- Pressure: The pressure shows how hard a user touches the screen. It is relevant with Tool and Touch. We defined 2 pressure-related features for a stroke: pressure mean (the mean of the pressure sequence) and pressure deviation (the standard deviation of the pressure sequence).
- Angle: We defined the angle between the horizontal line and the connecting line of two adjacent touch points in a stroke. Angle reflects the shape of a stroke and we defined 2 angle-related features for a stroke: angle mean (the mean of the angle sequence) and angle deviation (the standard deviation of the angle sequence).

In total, we have 13 behavioral features for one stroke. Thus, there are 52 behavioral features for a 4-finger TFST gesture, 39 for a 3-finger TFST gesture, and 26 for a 2-finger TFST gesture.

C. Data Collection

To investigate the performance of TFST gestures in multi-touch authentication, we applied and received an approval from the institutional review board of Xian Jiaotong University. We recruited 161 subjects and asked them to perform various TFST gestures. We collected their corresponding multi-touch traces to establish a multi-touch authentication dataset. The data collection lasted for more than 2 months. Each subject was asked to come every week in this period to support collecting behavioral variances.

1) Data Collection Environment: We developed an Android application on a smartphone to collect TFST gesture data. Only one smartphone was used to eliminate the confounding factors introduced by changing software and hardware environments. The smartphone was Samsung Galaxy N7000 (Note 1,1280×800 resolution) with a 5.3-inch screen, 1.4 GHz dual-core processor, and 1GB of RAM. The application installed on the smartphone was a touch data collector for users to perform TFST gestures. Traces of touching fingers were displayed on screen as visual feedback. Each subject was asked to perform the TFST gestures with his right hand.

Subjects were requested to perform horizontal, vertical and L-swipe TFST gestures and datasets were established for 2, 3 and 4 fingers respectively. We did not employ more complex gestures because the behavioral differences between users when performing complex gestures may be more specific to this particular gesture itself, while the differences exhibited in performing simple gestures are more likely to be gesture-nonspecific, which are more objective to demonstrate the performance of TFST gestures.

2) Subjects and Dataset: We recruited 161 subjects. 131 of them were sophomores aged from 18 to 20. 18 were master and PhD students aged from 23 to 30. And 12 were faculty members or staffs on campus aged from 30 to 55. The sophomores participated in the experiment as a requirement for a course. They have been informed that their grades and course credits had no relations with the data collection process. The data collection process would provide them the data to be analyzed using the knowledge taught in the course. The grading was solely based on the programs they wrote to analyze the data. The rest of subjects were recruited voluntarily. Totally, the subjects consisted of 26 females and 135 males. All were frequent smartphone users with at least 1 year experience using a touch screen. These subjects may be considered as a convenience sample [20]. However, we decided to focus this study on experienced touchscreen users since touchscreen operations are easy to learn and it is not difficult to become an experienced user.

A 7-session data collection lasted for more than 2 months. Each time the subject finished a session, at least one week passed before the next session. The time interval between sessions was set to be more than one week to guarantee that behavior variability with respect to time was captured in data collection procedure. In each session, subjects were asked to perform each TFST gesture over 20 times. The first session allowed subjects to become familiar with the gestures and the data were not analyzed. There were 6 experimental sessions, namely session 1 to session 6.

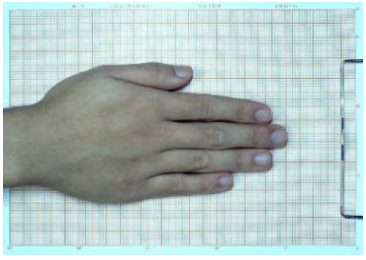


Fig. 3. Example of a subject's hand image

Most subjects completed the 6 sessions. In total, we collected more than 15000 samples for each TFST gesture from 161 subjects after deleting the erroneous samples. It requires 0.3 ~ 1.8 seconds to perform our TFST gestures, with an average of 0.75 seconds for all subjects.

In the data collection process, we also collected hand image data from 144 out of the 161 subjects to provide a supporting dataset for feature analysis of the multi-touch data.

IV. Feature Analysis

In this section we will analyze the basis for multi-touch authentication using TFST gestures and utilize Fisher Score as a feature analysis method to evaluate behavioral and physiological features together.

A. Discernibility of Physiological Features in TFST Gestures

The discernibility of physiological features in TFST gestures are rooted from the discernibility of hand geometry. In this section, we will analyze the discernibility of real features of hand geometry at first, and then demonstrate its connection to the physiological features in TFST gestures experimentally.

1) Real Features of Hand Geometry: Ross shows that hand geometry of different people shows discernibility and can be used as a good source of identity verification [21]. The 12 physiological features of TFST gestures are estimates of the 12 real features of hand geometry as depicted in Figure 2. In the following experiment, we will show the 12 real features of hand geometry have the discernibility to authenticate users.

In the data collection process, we collected hand image data on the coordinate paper from 144 subjects using distance fixed camera as shown in Figure 3. Each subject contributes 21 to 24 hand image samples. We performed rotations to make the directions of hands parallel to the horizontal axis on the paper and manually labelled the fingertip positions on the image. We calculated values of the 12 real features of hand geometry for each hand image as a sample of real hand geometry for one subject. Thus each subject has 21 to 24 samples and in total, we have 3240 samples for all 144 subjects.

By assuming a Gaussian distribution, we built a simple

TABLE I. CORRELATION COEFFICIENTS (CC) BETWEEN FEATURES OF HAND GEOMETRY AND FEATURES OF TFST GESTURE, FEATURE # ARE THE SAME AS THOSE GIVEN IN SECTION III-B

Feature ID	**1**	**2**	**3**	**4**	**5**	**6**
CC	0.75	0.92	0.82	0.73	0.88	0.89
Feature ID	**7**	**8**	**9**	**10**	**11**	**12**
CC	0.70	0.80	0.76	0.70	0.53	0.87

two-sided statistical model from legitimate samples of hand geometry for each user and used two 0.5% percentiles as the legitimate boundaries, which is roughly equivalent to 1% false rejection rate (rejection rate for legitimate samples). By using all samples from the other 143 subjects as the illegitimate samples, we achieved an average false acceptance rate (acceptance rate for illegitimate samples) of 1.08% for each of the 144 subjects.

This result shows that the 12 real features of users' hand geometry exhibit good discernibility to authenticate users.

2) Correlation Analysis between Features of Hand Geometry and TFST Gestures: The 12 physiological features of TFST gestures are estimates of the 12 real features of hand geometry. However, the correspondence between two sets of features are sometimes affected by external factors. For example, small variances in bending and closeness of fingers, and angle of touching may lead to changes in the relative positions of touching points in TFST operations, which in turn leads to changes of the 12 physiological features calculated in TFST gestures. This induces errors in hand geometry estimation using multi-touch traces of TFST gestures.

To explore how much impact the small behavioral variances have on values of physiological features in TFST gestures, we analyze the correlation between physiological features in TFST and corresponding real features of hand geometry.

We calculated average values of 12 features of hand geometry for each of 144 subjects from their hand image data. Then we used recorded multi-touch traces of 4-finger TFST gestures from the 144 subjects to calculate average values of 12 physiological features of TFST gestures for each subject.

The Pearson Correlation Coefficients [22] between two sets of 12 features for the 144 subjects are shown in Table I. We have 144 independent subjects; thus we have 142 degrees of freedom (dof). With this dof, if the coefficient is over than 0.16, the two features are regarded as significantly correlated at the significant level 0.05.

The Pearson Correlation Coefficients of the 12 features are all larger than 0.16, which indicates there are very strong correlations between the two sets of features. This may suggest that although there exists small behavioral variances, the connection between features in TFST gestures and real features of hand geometry is clear and strong. This provides a basis for using TFST gestures to authenticate users.

B. Feature Selection

While physiological features provide a basis for authentication using TFST gesture, behavioral features may complement the decrease in discernibility of physiological features due to measurement errors resulting from behavioral variances. Via the fusion of physiological and behavioral features, we may achieve better performance than only using each individual feature set. In this section, we employ feature selection techniques to search for the best combination of physiological and behavioral features for authentication.

Fisher Score [23, 24] is an effective technique to find discriminant features. The main idea is attempting to find a subset of features which maximizes the between class scatter and minimizes the within class scatter in the data space spanned by the selected features. A simple form of Fisher Score is given by,

$$Fisher(k) = \frac{\tilde{S}_b^k}{\tilde{S}_t^k}$$

where $\tilde{S}_b^k$ and $\tilde{S}_t^k$ are the k-th diagonal element of $\tilde{S}_b$ and $\tilde{S}_t$ respectively, and can be computed from the data of a single feature. $\tilde{S}_b$ is the "between class scatter matrix" and $\tilde{S}_t$ is the "within classes scatter matrix". The definition of the scatter matrices are:

$$\tilde{S}_b = \sum_{k=1}^{c} P_k(\tilde{\mu}_k - \tilde{\mu})(\tilde{\mu}_k - \tilde{\mu})^T,$$

$$\tilde{S}_t = \sum_{k=1}^{c} P_k \sum_{x_i^k \in c_k} \frac{1}{n_k}(x_i^k - \tilde{\mu}_k)(x_i^k - \tilde{\mu}_k)^T.$$

where $\tilde{\mu}_k$ and n_k are the mean vector and size of the k-th class respectively in the reduced data space, $\tilde{\mu}$ is the overall mean vector, c_k is the i-th class, and P_k refers to the priori probability of class i.

Figure 4 presents the Fisher Score of all features for a 4-finger gesture, including the physiological features and the behavioral ones. In general, the physiological features have larger Fisher Scores than the behavioral ones, which implies the discriminability and stability of physiological features are better.

We use a Fisher Score of 0.5 as a threshold to select better features, and list the selected features in Table II. All of the physiological features are selected and some of the behavioral features are selected. Finally, we get 36 selected features; 12 are physiological features and 24 are behavioral features.

V. ONE-CLASS CLASSIFIERS

User authentication is a two-class (legitimate user versus impostors) classification problem from the perspective of pattern-classification. We only have training data from legitimate users so we build a model based only on the legitimate user's data samples, and use that model to detect impostors. This type of problem is known as one-class classification or anomaly detection.

A. K-Nearest Neighbor

K-Nearest Neighbor classifier models a user's profile based on the assumption that new samples from the user will be similar to the samples in the training data. In the training phase, the classifier computes the Manhattan distance matrix between every pair of training samples, and determines a classification threshold based on the distance matrix. In the testing phase, the classifier calculates Manhattan distance

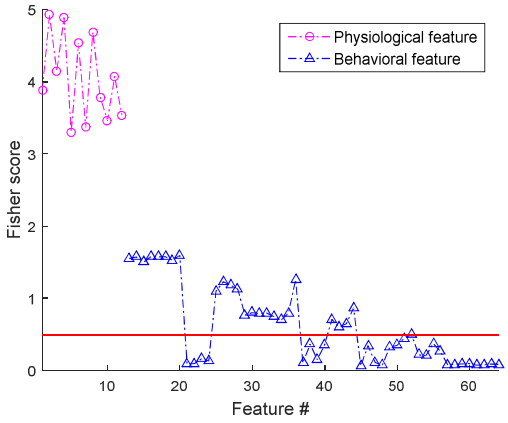


Fig. 4. The Fisher Score of the physiological and behavioral features. Features # from 1-12 are physiological; 13-64 are behavioral.

TABLE II. SELECTED FEATURES USING FISHER SCORE

Category	Feature Name	Selected #
Physiological	Point Distance	6
	Finger Width	3
	Length Difference	3
Behavioral	Length	8
	Time	4
	Velocity	4
	Tool	4
	Touch	4

from the new sample to each of the samples in the training data. The average distance calculated between the new sample to the nearest k samples in the training data is used as the classification score. If the classification score is below the determined classification threshold, we regard the sample as a legitimate one.

B. *Support Vector Machine*

We also implemented a one-class Support Vector Machine (SVM) classifier. One-class SVMs have been successfully applied to a number of classification problems, such as mouse dynamics, signature verification and keystroke authentication [25-27]. SVM generalizes the ideas of finding an optimal hyper-plane in a high-dimensional space to perform a classification. In the training phase, SVM builds models based on the training samples of the legitimate user. In the testing phase, the testing samples are projected onto the same high-dimensional space, and the distances between the samples and the hyper-plane are computed as the classification scores. If the classification score is over the threshold, we regard the sample as a legitimate one.

VI. EVALUATION METHODOLOGY

In this section, we discuss the evaluation methodology for our proposed multi-touch authentication. The evaluation is performed on the dataset described in Section III. We present the training and testing procedure for one-class classifiers and show the criterion to evaluate the performance of our approach.

A. *Training and Testing Procedure*

TABLE III. CONTINGENCY TABLE FOR MCNEMAR'S TEST

n_{00}: # of samples misclassified by C_A and C_B	n_{01}: # of samples misclassified by C_A but not by C_B
n_{10}: # of samples misclassified by C_B but not by C_A	n_{11}: # of samples misclassified by neither C_A and C_B

As we have 161 subjects in our evaluation dataset, we designated one of our subjects as the legitimate user, and the rest (160 subjects) as impostors. We trained the classifier and tested its performance to recognize the legitimate user and impostors as follows:

- We randomly selected a portion of samples from the legitimate users, and with these samples, we trained the classifier to build a profile of the legitimate user.
- We tested the performance of the classifier to authenticate the legitimate user with the remaining samples from the legitimate user.
- We tested the performance of the classifier to detect impostors with all samples from the 160 impostors.

This process was then repeated, designating each of the subjects as the legitimate user in turn. In the training phase, 10-fold cross-validation [28] was employed to search for reasonable parameters of the classifiers. Since we used a random sampling method to divide the data into training and testing sets, and we wanted to account for the effect of this randomness, we repeated the procedure 50 times for each legitimate user, each time with independently selected samples from the entire dataset.

B. *Evaluation of Classifier Performance*

We employ the false-acceptance rate (FAR) and false-rejection rate (FRR) as our main evaluation criteria. The FAR is defined as the ratio between the number of falsely accepted illegitimate samples and the number of all illegitimate testing samples; the FRR is defined as the ratio between the number of falsely rejected legitimate samples and the number of all legitimate testing samples. Via varying the threshold on classification score, we calculate the corresponding FRR and FAR pairs, and obtained a performance curve known as the receiver operating characteristic (ROC) curve. We also calculate the equal-error rate (EER) from the ROC curve where FAR equals FRR.

C. *McNemar's Test*

To compare the performance of our two different classifiers, we employed McNemar's test [29], a frequently used test for binary matched-pair data. First, we divide our available data set S into a training set R and a testing set T. We train both of our two classification algorithms on the training set R and obtain two classifiers C_A and C_B. Then we test these classifiers on the testing set T and record the classification results in a contingency table (Table III).

Under the null hypothesis, the two algorithms should have the same error rate, which means $n_{01} = n_{10}$. So the following statistic is distributed as χ^2 with 1 degree of freedom; it incorporates a continuity correction term to account for the fact that the statistic is discrete while the χ^2 distribution is

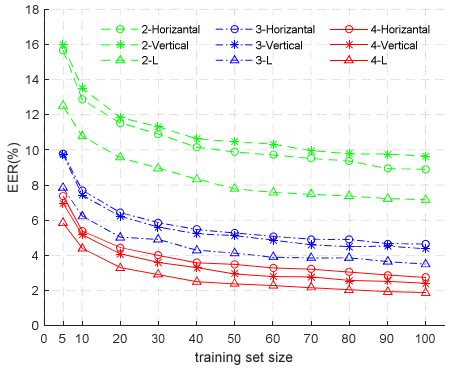


Fig. 5. EER curves for 9 types of gestures at varying training set sizes

continuous:

$$\frac{(|n_{01}-n_{10}|-1)^2}{n_{01}+n_{10}} \sim \chi^2(1),$$

If the null hypothesis is correct, then the possibility that this quantity is greater than 3.84 is less than 0.05. So we may reject the null hypothesis in favor of the hypothesis that the two algorithms have different performance when trained on the particular training set *R* and regard the classifier with low error rate as the better one.

VII. Reliability Evaluation

This section presents an evaluation of the reliability of TFST gesture authentication. We systematically investigate the accuracy of verifying the legitimacy of a user with respect to different TFST gestures, feature sets, classifiers and sizes of training sets. We also examined the impact of behavioral variability utilizing the long-term behavioral data in our dataset to see whether the performance will deteriorate with the elapse of time.

A. Comparison of TFST Gestures

In this experiment, we compare authentication performance when subjects are requested to perform horizontal, vertical and L swipes using 2-finger, 3-finger and 4-finger TFST gestures respectively.

As described in Section III, we have 52, 39 and 26 behavioral features, and 12, 7 and 3 physiological features, for a 4-finger, 3-finger and 2-finger TFST gesture respectively. For L swipe TFST gestures, we separate the trace into two sub-gestures: one vertical TFST swipe and one horizontal TFST swipe. So L swipe TFST gestures have twice as many features as other TFST gestures. The classifying features are selected from combined physiological and behavioral features of TFST gestures with the Fisher Score over 0.5. The size of the training set ranges from 5 to 100. The classifier is One-Class KNN with k set to be 3. Figure 5 shows the average EERs for the nine gestures.

Figure 5 shows that TFST gestures with more fingers achieve better results. However, the simplest 2-finger gesture can achieve an EER of 7.17% with enough training samples. Trading off security and convenience, 3-finger swipes can be used as a promising authentication method on small screen smartphones with EERs less than 5% assuming enough training samples.

Among all the gestures, the 4-finger L swipe achieves the best performance. We speculate that the 4-finger swipe contains more biometric information of hand geometry, and swipes in both the horizontal and vertical directions make the estimates of features of hand geometry more accurate. These lead to the better performance. We will use the 4-finger L swipe as the evaluation TFST gesture in the following experiments.

Figure 5 also exhibits how the authentication performance changes with size of training data, which is an important perspective of a behavioral authentication technique and will be discussed in Section VII-C.

B. Effect of Feature Subsets and Classifiers

In Section III-B, we defined physiological and behavioral features to characterize TFST gestures. With feature selection, we obtain 4 different feature subsets: (1) physiological subset; (2) behavioral subset; (3) the whole set which contains all the features in (1) and (2); and (4) selected subset where features in (3) are selected by a Fisher Score over 0.5.

In this experiment, we investigated the effect of feature subsets and classifiers on the performance of authentication of the TFST gesture of 4-finger L swipe. We employed one-class SVM and K-Nearest Neighbor classifiers with the inputs set to 4 different feature subsets.

To select parameter k for KNN, multiple tests with k ranging from 1 to 20 were performed. The best parameter k = 3 is selected. For SVM, we employed the radial basis function (RBF) kernel after comparative studies of linear, polynomial, RBF, and sigmoid kernels according to the average classification accuracy. The SVM parameter υ and kernel parameter γ (using LibSVM [30]) were set to 0.05 and 0.015 respectively.

The size of the training set was set to 30. Subjects with at least 30 samples were designated as legitimate users. For each legitimate user, the other 160 subjects were set to be imposters. Figure 6 shows the ROC curves for four types of feature subsets using different classifiers.

As shown in Figure 6, the authentication accuracies with the physiological subset (1), whole set (3) and selected subset (4) are all much higher than with the behavioral subset (2), which suggests that our approach to incorporate hand geometry information into multi-touch authentication brings significant improvements over a purely behavioral based approach.

Another observation is that the performance of the selected subset is superior to the other feature subsets. It indicates that the Fisher Score for feature selection is an effective way to fuse physiological and behavioral features. The fused feature set performs better than each of individual feature set. This provides evidence for our assumption: behavioral features may complement the decrease in discernibility of physiological features due to measurement errors resulting from behavioral variances.

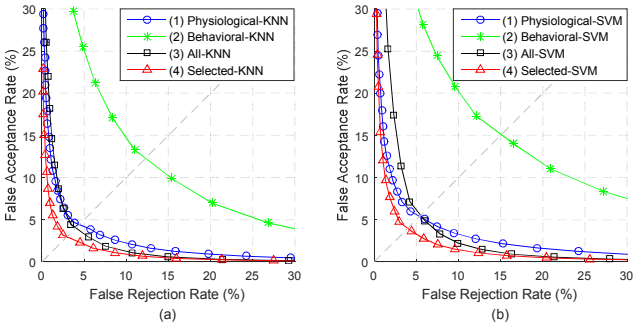


Fig. 6. ROC curves for 4 types of feature subsets using 2 types of classifiers: (a) K-Nearest Neighbor, (b) SVM

TABLE IV. COMPARISON OF TWO CLASSIFIERS USING MCNEMAR'S TEST

Better classifier	# of cases	Proportion (%)
KNN	126	87.5
SVM	6	4.17
Equivalent	12	8.33

We employed McNemar's test to evaluate the performance of the classifiers. The significance level α is set to 0.05. Table IV shows the result. In 12 cases, there is no significant difference between the performance of 3-Nearest-Neighbor and SVM; in 126 cases, 3-Nearest-Neighbor outperforms SVM; and SVM has a better performance only in 6 cases. This result shows that the 3-Nearest Neighbor classifier has better performance than SVM in general. This may be due to the situation that our training set is small and the SVM classifier is trained inadequately, so the few support vectors cannot describe the complex profile of positive samples perfectly. The 3-Nearest Neighbor classifier, using Manhattan distance to measure the distance between samples, nonlinearly builds a more reasonable boundary to distinguish positive and negative samples. Given the better performance of 3-Nearest Neighbor classifier, we will use it as the main classifier in the later experiments.

C. *Effect of Training Set Size*

In the previous experiment, 30 samples from each subjects were used as the training dataset for authentication. Next we investigated using different training sample sizes.

We employed 3-Nearest Neighbor as the one-class classifier in this evaluation, with the input features set to be the physiological subset, the behavioral subset, and the selected subset respectively. We changed the size of the training set from 5 to 100 in a step of 5 initially and 10 afterwards to investigate the impact of training data size on 4-finger L swipe authentication. Figure 7 plots the average EERs against different sizes of training dataset.

The result shows that the size of a training set will have a significant effect on the performance of authentication. For all three input feature subsets, the average EER decreases with more training data. A large training set often gives the classifier more information and more training samples characterize the legitimate user more accurately and lead to lower EERs.

Among the three feature subsets, we observed that

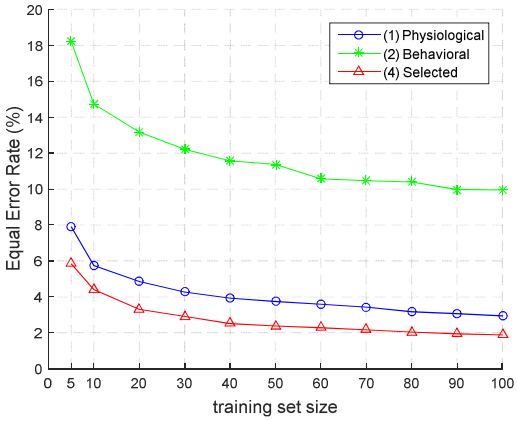


Fig. 7. EERs for 3 types of feature subsets at varying training set sizes

physiological and selected feature subsets exhibited a steeper learning curve. This showed the feature spaces of physiological and selected combined features were more compact so that less training data yielded better learning effects. The selected combined feature set showed the best learning performance with final EER dropping to 1/3 of the initial value (from 5.84% to 1.88%) with enough learning.

The result also showed for the selected feature subset, using over 15000 samples from 160 subjects as illegitimate testing samples, TFST L swipe authentication is able to achieve an EER of 5.84% with only 5 training samples, compared with an EER of 18.21% for the behavioral feature set and an EER of 7.91% for the physiological feature set. This demonstrated the effectiveness of fusion of physiological and behavioral features for multi-touch authentications.

In the experiment of Section VII-A, we also investigated the effect of different sizes of training data on authentication performance of different TFST gestures. From Figure 7, we can see that simple 3-finger TFST L swipe authentication can also achieve an EER of 9.32% with only 5 training samples and an EER of 4.10% with enough training. This suggested that 3-finger TFST gestures may provide an easy and relative secure authentication method on small screen devices such as popular smartphones.

D. *Behavioral Variability*

Behavioral variability is an important issue for behavioral biometric techniques. In this experiment, we focus on long-term behavioral variability and its impact on multi-touch authentication.

As introduced in Section III-C, we separated the data collection process into 6 sessions over about 2 months, with sessions separated by more than a week. We also required the subjects to participate in every session of data collection. By doing this, we recorded long-term variations of touching behaviors for each subject in the collected dataset. In this experiment, the training and testing datasets are organized according to sessions to investigate the impact of behavioral variability with respect to time.

The setting of training and testing data for each EER calculation of session N is shown in Table V. For each subject, there are about 20 samples in one session. We randomly select

TABLE V. TRAINING AND TESTING DATA FOR SESSION N

Experimental Datasets	Description
Training set	5 samples from subject in Session 1
Legitimate Testing set	All samples from subject in Session N (2~6)
Illegitimate Testing set	All imposter samples from all sessions

5 samples in the first session to train the one-class classifier. Then we use his data in subsequent sessions as legitimate testing data and samples from all other subjects in all sessions as illegitimate testing data of imposters.

We employed 3-Nearest Neighbor as the one-class classifier in this evaluation, with the input features set to be the physiological subset, the behavioral subset and the selected subset for the 4-finger TFST L swipe gesture respectively.

The experiment is repeated for each subject as the legitimate user. For each subject, we repeat the experiment 10 times to account for the randomness. Figure 8 show the average EERs when legitimate data in different sessions are used as the legitimate testing data.

Since we only train the one-class classifier once, using data from the first session, the performance of the trained one-class classifier on data from later sessions demonstrate the applicability of the identity model to the behavioral data collected after model training. Our dataset allows us to explore this applicability on behaviors recorded in subsequent two months. This period likely captured substantial amounts of behavioral variability in the collected dataset.

As shown in Figure 8, the EERs for the physiological and selected feature subsets do not vary very much and are relatively constant over different sessions. For example for physiological features, we only observe slight increases of EERs from 6.38% to 7.51% for session periods from 2 to 6. While for behavioral features, the increase of EERs are from 13.39% to 22.14% for the same periods. This means physiological features are better than behavioral features in terms of resisting to behavioral variability over time. We also noticed, for the selected combined features, the EER performance is again the best, with a slight increase from 4.46% to 5.58%. This showed that the fusion of behavioral information with hand geometry features not only reduce the EER levels, but also leads to resistance to behavioral variability over time.

VIII. SECURITY ANALYSIS

In this section, we examine the security of the proposed method according to the threat model presented in Section II. We perform an experimental study of the security of TFST gesture authentication against the so called zero-effort attack, smudge attack, shoulder surfing attack and statistical attack.

A. *Zero-effort Attack*

Zero-effort attack may be the most common type of attack against an authentication system where the attacker guesses

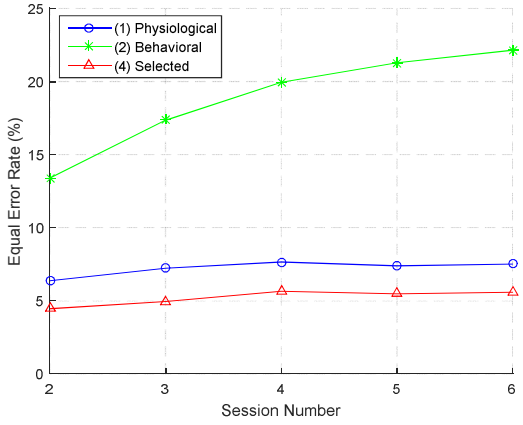


Fig. 8. Long-term EER curves for different feature sets

the secret or tries the authentication procedure without much knowledge of how a legitimate user enters the system. For TFST gesture authentication, we assume a zero-effort attacker only knows which gesture to perform and has no other information.

In section VII, the investigation performed on the various accuracies of TFST gestures is actually an investigation of the resilience to zero-effort attack. For the 161 subjects in our evaluation dataset, we designated one of our subjects as the legitimate user, and the other 160 subjects as impostors, or zero-effort attackers. The results demonstrated the resilience of TFST gesture authentication to zero-effort attacks in general.

Since hand geometry information is very important in TFST gesture authentication and there are chances that an attacker has a similar handshape as the victim being attacked, we further investigate zero-effort attacks considering the factor of handshape similarity. To do this, we selected the user pairs with similar handshapes by calculating a similarity metric, *Sim*, based on the recorded hand images in our dataset.

$$Sim_{ij} = 1 - \frac{\| v_i - v_j \|_1}{\| v_i + v_j \|_1}$$

For user i and user j, v_i and v_j are their feature vectors consisting of the 12 real features of hand geometry depicted in Figure 2. And $\|v_i\|_1$ means the 1 norm of vector v_i. For all 144 subjects with recorded hand images, there are 10296 user pairs.

We use the 4-finger TFST L swipe gesture to evaluate the resilience to zero-effort attack. The authentication model being attacked is a 3-NN classifier trained by 30 legitimate samples with selected features. The experiment was performed 20 times to account for the randomness.

We calculate the *Sim* values for all user pairs in our dataset and simulate users in each pair attacking each other by using one user's data to attack the other's authentication model. We calculate the average FAR at FRR=3% for user pairs with different *Sim* values. The results are shown in Table VI. When hand similarity information was not considered (1st row in Table VI), the average FAR was 4.41% for the selected feature set. For the most similar 59 pairs of users with *Sim* values higher than 0.98 (5th row in Table VI), the average FAR is

TABLE VI. ZERO-EFFORT ATTACK AND HAND SIMILARITY (FARS ARE CALCULATED AT FRR=3%)

Similarity	# of pairs	Avg. FAR (Selected)	Avg. FAR (Physiological)
0	10296	4.41	5.60
0.95	2793	4.64	8.30
0.96	1309	4.82	9.01
0.97	404	4.92	9.12
0.98	59	5.04	9.33

5.04%. The difference between the two FARs showed that even with a similar hand geometry, the likelihood of breaking into an account via zero-effort attack is still not very high for authentication using TFST gestures.

We speculated that the fusion of physiological and behavioral features contributed to the resistance of zero-effort attack with similar handshapes. As shown in Table VI, for the pure physiological feature set, the FARs increased faster with hand similarity.

B. *Smudge and Shoulder Surfing Attack*

Smudge attack [2] and shoulder surfing attack [1] are two types of common attacks in which an attacker manages to obtain some knowledge of the authentication process of the legitimate user. In a smudge attack, attackers utilize the oily traces left on the screen as hints to pass the authentication. In a shoulder surfing attack, attackers watch the authentication process and mimic the behavior to pass the authentication. In this subsection, we demonstrate that TFST gesture authentication is resilient to both types attacks and their combinations.

Experimental Setup: We recruited another 20 students on campus as attackers to attack the 144 subjects in our dataset with recorded hand images. Each attacker attacked 10 victims. 5 victims have the most similar handshape as the attacker, (*Sim* values with the attacker are the highest). The other 5 victims have handshapes that are not similar (*Sim* values with the attacker are among the lowest).

The attackers were asked to try their best to mimic the victims' 4-finger TFST gestures to see if they are able to be accepted by the authentication model. Each attacker is provided with two randomly selected multi-touch traces from each of the 10 victims to mimic. The attackers are allowed to practice as many times as they want and finally each attacker will generate 10 mimicry multi-touch traces for each genuine multi-touch trace provided.

The authentication models being attacked are the 3-NN classifier trained by 30 legitimate samples and 100 legitimate samples respectively. The EERs are calculated according to the decisions made by the corresponding authentication model on all mimicry traces and legitimate traces not used in model training. The experiment was repeated 20 times to account for the randomness.

Smudge Attack: To evaluate the resilience to smudge attack, we drew the genuine multi-touch traces on the screen for the attacker to mimic. This corresponds to the worst situation where the oily residuals are complete and clear and the attacker obtains the complete multi-touch trace.

TABLE VII. EERS(%) OF SMUDGE ATTACK ON MODEL WITH 30-SAMPLE TRAINING

Type of Attack	Physiological	Behavioral	Selected
Zero-effort attack	4.06	12.10	3.02
Similar-handshape smudge attack	4.57	11.84	3.08
Dissimilar-handshape smudge attack	2.53	11.61	1.99

TABLE VIII. EERS(%) OF SMUDGE ATTACK ON MODEL WITH 100-SAMPLE TRAINING

Type of Attack	Physiological	Behavioral	Selected
Zero-effort attack	2.94	9.95	1.88
Similar-handshape smudge attack	3.16	9.13	2.00
Dissimilar-handshape smudge attack	1.69	8.66	0.96

The EERs are shown in Table VII and VIII for both smudge attacks with similar handshapes and dissimilar handshapes with regard to authentication models trained with different feature sets and different legitimate samples. The baseline was the zero-effort attack results not considering handshape similarity. From Tables VII and VIII, for smudge attack with similar handshapes, there are EER increases under the physiological models compared with the baseline. The EER decreased with the behavioral models which may mean that the oily residues do not help the attacker to mimic the behavioral features. As a result, the EER under the selected combined models only increases slightly from 3.02% to 3.08% with 30 sample training and from 1.88% to 2.00% with 100 sample training.

For smudge attack with dissimilar handshapes, the EERs are significantly decreased even with the physiological models compared with the baseline EERs. This showed that the leaked information of hand geometry cannot help an attacker with a dissimilar handshape to overcome the fundamental difficulty of hand dissimilarity when attacking our TFST gesture authentication method. We have similar results for the other two attacks with dissimilar handshapes. For space, we only present the results for attack with similar handshapes in the following experiments.

The above results and analysis show that our method is resilient to smudge attacks with both similar and dissimilar handshapes.

Shoulder Surfing Attack: To evaluate the resilience to shoulder surfing attack, we asked the attackers to watch an animation showing the movements of the victim's fingers on the screen of the testing device. The animation accurately replicates the temporal information recorded in a mimicked multi-touch trace. The attackers are allowed watch the animation as many times as they want.

The EERs are shown in Tables IX and X for shoulder surfing attacks with similar handshapes with regard to authentication models trained with different feature sets and different legitimate samples. The baseline is the zero-effort

TABLE IX. EERS(%) OF SHOULDER SURFING AND COMBINED ATTACK ON MODEL WITH 30-SAMPLE TRAINING

Type of Attack	Physiological	Behavioral	Selected
Zero-effort	4.06	12.10	3.02
Shoulder surfing	4.92	12.88	3.31
Combined	5.20	13.34	3.67

TABLE X. EERS(%) OF SHOULDER SURFING AND COMBINED ATTACK ON MODEL WITH 100-SAMPLE TRAINING

Type of Attack	Physiological	Behavioral	Selected
Zero-effort	2.94	9.95	1.88
Shoulder surfing	3.61	10.18	2.06
Combined	4.18	10.44	2.27

attack results not considering handshape similarity.

Tables IX and X also show the results for the combined smudge and shoulder surfing attacks in the row named combined attack. This attack is similar to the shoulder surfing attack except that the multi-touch trace of a victim will be left on the screen when the animation is finished. This corresponds to the worst situation that the attacker knows information of both hand geometry and behavior. The EERs are also for the attacks with similar handshapes. In Tables IX and X, the differences between EERs under attacks and baseline EERs exhibit the resilience of authentication with 4 finger L swipe against shoulder surfing attacks. Under shoulder surfing attack, for the selected combined model trained with 30 samples, the EER increases from the baseline value of 3.02% to 3.31%; while for the model trained with 100 samples, the EER increases from the baseline value of 1.88% to 2.06%. Under the more serious situations of the combined attacks of both smudge and shoulder surfing, the EERs are still not very high: 3.67% and 2.27% respectively.

Recalling that all attacks are examined under more difficult situations of attacking with similar handshapes, the above results demonstrate that our method has the resilience to both smudge attacks and shoulder surfing attacks.

C. *Statistical attack*

Statistical attacks have been shown to be effective against behavior based authentication methods [31-33]. In [31],touch-based authentication systems were attacked successfully using forgeries generated by a simple "Lego" robot driven by input gleaned from general population swiping statistics.

The basic idea of statistical attack is to estimate the probability density functions (pdf) of features from a group of people and then use the most probable feature values to generate the forgery. According to the attack method shown in [31-33], we developed Algorithm 1 to generate synthetic attack samples. The inputs of the algorithm are RealFeatures and NumberOfBins. RealFeatures is a matrix consisting of the feature vectors for multi-touch traces generated by genuine users. Each column of RealFeatures is a feature vector for one multi-touch trace. Each row is a series of values for one feature. In the algorithm, the feature values in each row of RealFeatures is "binned" to approximate the pdf of a feature. NumberOfBins controls the granularity of the approximation.

ALGORITHM 1: Generating forged features for statistical attack

Input: RealFeatures[ ]; //Population feature vectors
Input: NumberOfBins; //Number of bins for each feature
Output: ForgedFeatures[ ]; //Feature vectors used for attack
NumberOfFeatures = NumberOfRows(RealFeatures);
for *i*=1 **to** NumberOfFeatures
do
  BinnedFeatures[*i*] = Binning(RealFeatures[*i*], NumberOfBins);
  //Generate bins according to RealFeatures[*i*] and NumberOfBins
  KeyBin[*i*] = SortBinsByFrequency(BinnedFeatures[*i*]);
  //Sort bins in descending order of frequency
  LowerBound[*i*],UpperBound[*i*] = GetBound(KeyBin[*i*]);
  //get the bound of the first sorted bins
  AttackFeatures[*i*] = uniform(LowerBound[*i*],UpperBound[*i*]);
Return AttackFeatures[ ]

Then the bin with the highest frequency is selected to generate a forged feature value for the output ForgedFeatures.

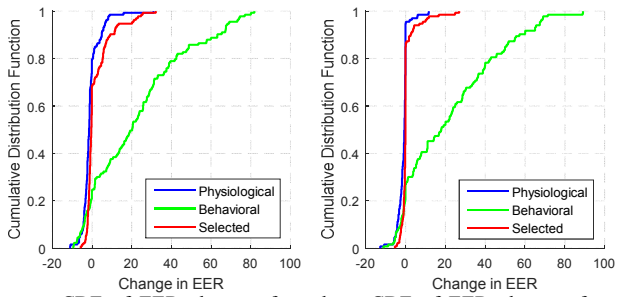
We evaluated the effect of statistical attacks on the authentication model built from 4 finger TFST gestures. We used all samples from the 161 subjects in our dataset as the input matrix of RealFeatures to investigate the worst situation where the attacker has the knowledge about the statistics of the whole population. We also perform a search for the optimal parameter of NumberOfBins in the range from 10 to 100 so that we have the best effect of attack.

Using the above RealFeatures and NumberOfBins as inputs to Algorithm 1, we generated 10000 synthetic feature vectors to attack the authentication models for each of the 161 subjects trained with 30 and 100 legitimate samples respectively. To account for randomness, we repeated the experiments 20 times and used the average EERs to show the effect of attacks.

Figure 9 shows the changes of average EERs between statistical attack and zero-effort attack for each of 161 subjects in cumulative distribution function (CDF) graphs. Using EERs of zero-effort attack as the baseline, EER changes less than or equal to 0 indicate the statistical attack has no effect. For EER changes greater than 0, the statistical attack exhibits a positive effect. The larger the EER changes, the more significant is the statistical attack.

Figure 9 shows that while the accuracies of the behavioral authentication models are undermined by the statistical attack, the performance of physiological and combined selected models are less influenced. Figure 9a shows, for physiological models trained with 30 samples, 80% of the subjects are not affected by statistical attacks. For the selected model, about 70% of the subjects are not affected by the statistical attacks. Figure 9b shows if trained with more legitimate samples, the resilience to statistical attacks will become better. For the selected model trained with 100 samples, we have about 87% of the subjects not affected by the statistical attack.

Tables XI and XII show averaged EERs under statistical attacks over all 161 subjects for models trained with 30 and 100 samples respectively. With the averaged EERs of zero-effort attacks as baselines, the tables show the overall impact



a. CDF of EER changes for model with 30-sample training

b. CDF of EER changes for model with 100-sample training

Fig. 9. Impact of statistical attacks. For each user, we subtracted the EER under the zero-effort attack from that under the statistical attack and then plotted the CDFs of these changes in EER.

TABLE XI. EERS (%) OF STATISTICAL ATTACKS ON MODEL WITH 30-SAMPLE TRAINING

Scenarios	Physiological	Behavioral	Selected
Zero-effort Attack	4.06	12.10	3.02
Statistical Attack	4.35	39.23	4.69

TABLE XII. EERS (%) OF STATISTICAL ATTACKS ON MODEL WITH 100-SAMPLE TRAINING

Scenarios	Physiological	Behavioral	Selected
Zero-effort Attack	2.94	9.95	1.88
Statistical Attack	2.17	32.67	2.43

of statistical attacks for the whole population. The results show that the statistical attack does not have much effect on the authentication model built using only physiological features. In fact, for physiological model trained with 100 samples, we note that the average EER under statistical attacks is 2.17%. It is even lower than the baseline EER of 2.94%. Thus although the authentication model built on pure behavioral features are undermined, with help of the physiological features, the authentication model using both selected physiological and behavioral features also shows some resilience to the statistical attack. Compared with the behavioral model, its EERs only increased from 3.02% for baseline to 4.69% for statistical attack in Table XI and from 1.88% for baseline to 2.43% for statistical attack in Table XII.

We speculate that the strong resilience to statistical attacks with physiological models is due to the stability and distinctiveness of physiological features. This makes the legitimate ranges of physiological features small and feature values of different subjects widely separated. Thus the most probable feature values drawn from a population in statistical attacks actually cannot fit into the ranges of many legitimate users.

IX. USABILITY STUDY

Usability is a very important factor for authentication systems on smart devices such as smartphones. We investigated the usability of TFST gesture authentication by inviting more than 158 undergraduate students on campus to try different TFST gesture authentications on smartphones with different screen sizes. These participants are different from the subjects involved in previous data collections and were not familiar with the TFST gestures. They were asked to rate TFST gesture authentication from the following 4 perspectives and to compare it to the commonly used methods of passcode and pattern lock: 1) Is it easy to memorize? 2) Is it fast to login in? 3) Is it convenient to perform? 4) Is it less error prone? For each question, we use 3 (1-3) levels representing responses of "disagree", "neutral" and "agree".

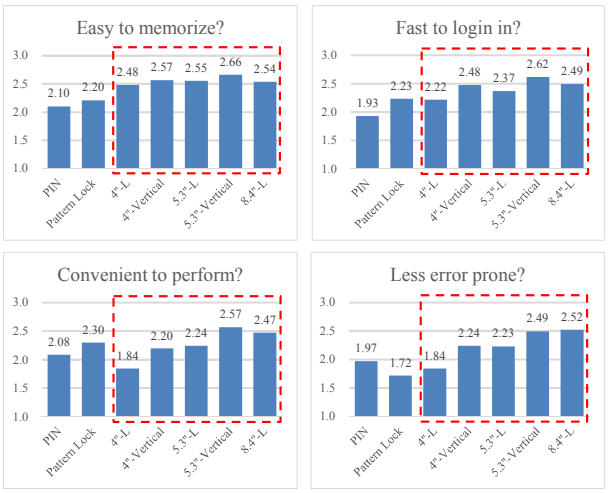


Fig. 10. Average ratings for the four usability questions

The subjects were asked to evaluate TFST gesture authentications on two smartphones and one tablet of 3 different screen sizes (4", 5.3" and 8.4"). The subjects were asked to perform the 3-finger vertical swipe and 3-finger L swipe on 4" smartphone, the 4-finger vertical swipe and 4-finger L swipe on the 5.3" smartphone and the 4-finger L swipe on the 8.4" tablet. Before answering the questionnaire, each subject tried all of the gestures more than 20 times on each device. The average ratings for different authentication methods calculated from the resulted 158 questionnaires are shown in Figure 10 with results of TFST gestures highlighted in the red box.

Figure 10 shows that compared to the standard methods, the evaluated gestures are regarded easier to memorize; 4 out of 5 gestures are regarded faster for login and less error prone. For convenience, there are 4 out of 5 gestures rated better than passcode and 2 out of 5 rated better than pattern lock. The 3 finger L swipe on the 4" screen is the lowest rated TFST gesture, but it is still rated better than pattern lock in 2 out of 4 aspects, and better than passcode in 3 out of 4 aspects.

This result demonstrates that user acceptance of TFST gesture authentication method is high compared with traditional methods of passcode and pattern lock. Our methods are generally regarded as easy, fast and convenient. The gestures do not incur too much cognitive load on the user. Thus, for devices with small screen sizes such as 4", 3 finger TFST gestures are able to provide good usability with relatively good security. For devices with 5.3" or larger screens, 4 finger TFST gestures are able to provide good usability and good security simultaneously.

X. Related Work

Behavioral biometrics, which identify a person by analyzing his own behavioral traits, are becoming a hot topic on smartphones, with the help of various embedded sensors and devices such as accelerometer, gyroscope and touch screen. Authentication approaches using gait recognition, keystroke dynamics, and touch gestures are proposed and investigated on smartphones and other mobile devices.

Gait Recognition: This authentication method relies on sensor data from accelerometers and gyroscopes and analyzes the way how a user walks to authenticate the user. Mantyjarvi [4] proposed an approach of identifying persons by using characteristics in the acceleration signal produced by walking with a portable sensor device in 2005. In 2010, Kwapisz [5] utilized accelerometers in smartphones to identify and authenticate cellphone users based on their normal daily activities. But both Mantyjarvi and Kwapisz assumed the smartphone is fixed on some part of the user's body, which is too strict for normal usages of smartphones. In 2014, Lu [6] came up with a gait verification system for mobile phones without assumption of body placement or device orientation. However, all gait recognition approaches require the observation of a certain period of users' behaviors which is not suitable for common authentication scenarios such as screen unlocking requiring a quick authentication.

Keystroke dynamics: This methods analyzes the manner and the rhythm of typing characters on a keypad or soft keyboard to authenticate users. It was first applied in authenticating users on PC systems with desktop keyboard [7, 8]. Afterwards, keystroke dynamics has been investigated to authenticate users of mobile devices. Clarke et al. [9] in 2003 conducted a feasibility study of using keystroke dynamics to authenticate users on mobile handsets. Then in 2012, Zheng et al. [10] utilized tapping behaviors for user verification on touchscreen smartphone, and extracted four types of features to distinguish the legitimate user and impostors. Giu□rida et al. [11] proposed a new biometric mechanism by sensor-enhanced keystroke dynamics in 2014. They asked subjects to type fixed-texts and characterized users' identity via the combination of traditional keystroke dynamics and accelerator information. Keystroke dynamics rely on behavioral features and the impact of long term behavioral variability has not been extensively investigated in previous work.

Gesture based authentication: With the popularity of smartphones, touchscreens have recently become a leading input device. Gesture based authentication utilizes the detailed information of how users perform certain gestures on touchscreens to verify the identities of users [12-14,34,35]. In 2012, De Luca et al. [12] proposed a pattern lock enhancement method. They designed an unlocking pattern with 4 gestures: horizontal, vertical, vertical with two fingers and diagonal, and collected 30,720 unlocks data from 48 subjects. They reported a best accuracy of 57% among four gestures. They also implemented the gesture based authentication as a hidden security layer for a pattern lock. With 34 subjects' data consisting of XY-coordinates, pressure, size, time and speed, they achieved an average 77% accuracy. De Luca et al. performed a pioneering study on mobile authentications using simple touch gestures. However, the reported accuracies are not satisfactory for real-world implementations and the important issue of behavioral variances was not closely examined. Cai et al. proposed another approach to authenticate users using common touch gestures such as drag, zoom-in and zoom-out [34]. But the authentication process lasts more than 6 seconds, which makes it difficult to be used for the frequent phone unlocks.

Shahzad et al. [13] presented GEAT, a gesture based user authentication system for the secure unlocking of touch screen devices. They designed 39 gestures and chose the 10 most e□ective gestures. They extracted features like finger velocity, accelerator readings, and stroke time to recognize users. With samples from 50 volunteers, the authors achieved an average EER of 4.8% with 1 gesture and 1.7% with 3 gestures. Sherman et al. [35] studied free-form multi-touch gestures with any number of fingers for mobile authentication. The technique is different from the TFST gestures proposed in this paper. Both Shahzad and Sherman used behavioral characteristics for authentication, but they did not perform extensive investigations of behavioral variability either.

Sae-Bae et al. proposed an interesting approach to use five-finger gestures on multi-touch screen for authentication [36,14]. They defined a set of 22 multi-touch gestures and extracted 20-dimensional features from the multi-touch traces. They used DTW to calculate the dissimilarity scores for the features. Their method achieved an EER of 7.88% on average and 2.98% for the best case "user-defined" gestures. Since no restrictions are posed on user's gestures, the proposed features are subject to behavioral variability introduced by variations in separation and bending of fingers during multi-touch operations. This led to performance deterioration with time as they reported EERs close to 20% for inter session authentications [14]. Moreover, the proposed gestures require a large touch screen to perform. This limits the applicability to the majority of smartphones with medium or small sized screens.

For all investigated behavioral biometrics on smartphone platforms, variations in behaviors or behavioral variability constitute a serious challenge to undermine the accuracy and user experience in real applications. To deal with this problem, we developed a new approach to multi-touch authentication by using physiological information of hand geometry and behavioral characteristics simultaneously, so that behavioral variability can be largely reduced.

XI. Discussions and Future Work

A. Authentication Time

Authentication time is an important aspect in the usability of an authentication system on smart devices such as a smartphone. This time is related to action time, verification time and enrollment time for our TFST gesture based authentication.

Action time is the time required for a user to perform a TFST gesture on the touch screen. For the most complex 4-finger TFST L swipe gesture, it takes 0.75 second in average for a subject to complete. Verification time is the time

required for the smart device to verify the legitimacy of a user using the multi-touch trace of his TFST gesture performed. A prototype system we developed on Samsung Note 1 takes about 0.2 second to perform the verification with a system overhead of 20M memory and 1% CPU.

Enrollment time consists of time to provide the training samples and time of model training. As shown in Figure 7, for an EER of 3%, a user should provide 20 training samples of 4-finger TFST L swipe. It takes about 1 minute for an inexperienced user to complete. The model training with KNN on Note 1 takes about 2-3 seconds. To save enrollment time, we may allow new users to provide 5 training samples with an enrollment EER of 5.84% (Figure 7), and update the authentication model in the subsequent authentication stage.

B. Advanced Attacks

In Section VIII, we have shown our method is resilient to the four common types of attacks described in our threat model. For biometric authentication, replay attack is another relevant attack which is effective against fingerprint and face recognition [37]. In replay attack [38], an attacker replays a legitimate user's previously recorded authentication action to the authentication system.

For our method, replay attack can be done at the touchscreen interface outside a smart device, or inside the device by injecting recorded samples directly into the dataflow of the authentication system. The latter approach requires access to the inner operation system, which means local protection should be breached at first. This is out of the scope of our method as a local protection mechanism. For the former approach, its success relies on the replication of both the hand geometry and multi-touch behavior. If not impossible, it will be very difficult to be accomplished by the type of adversaries we assume to defend against in Section II.

C. Future Work

In this paper, we only analyze simple TFST gestures such as vertical, horizontal and L swipes, and investigate their basic capabilities for user authentication. In fact, there are more types of complex TFST gestures worthy of exploration, such as Z swipe and multi-touch signatures. The stability and discernibility of these gestures are good topics for future research.

Another important work to pursue in future is to expand the dataset. Currently, we have established a reasonably large dataset consisting of more than two months of data from 161 subjects. But all subjects are from within campus, it will be helpful to collect data from other population categories such as workers and children, and from different ethnic groups. These data can be used to evaluate whether the results achieved in this paper are generalizable to a more diverse population and provide a more comprehensive basis to show the effectiveness of our approach.

XII. Conclusion

In this paper, we propose a simple, fast, reliable and (*sufficiently*) secure approach to multi-touch authentication using information from both hand geometry and behavioral characteristics. Users are authenticated by performing simple TFST gestures with one swipe on the touchscreen.

TFST gestures require users to stretch their fingers and put them together. This makes the hand posture conform to a fixed hand geometry and leads to a more stable behavioral pattern. Moreover, TFST gestures require much less touch area than traditional multi-touch operations. So multi-touch authentication using TFST gestures can be deployed on a wide range of multi-touch enabled devices from small screen smartphones to large screen tablets or laptops.

To evaluate the reliability of our method, we established a large-scale multi-touch dataset from 161 subjects. Data collection process was elaborately designed to guarantee behavior variability with respect to time was captured. We performed a comprehensive experimental analysis with respect to different TFST gestures, feature sets, classifiers and sizes of training sets. Our approach achieves an EER of 5.84% in verifying the legitimacy of a user with only 5 training samples and the accuracy is further improved to an EER of 1.88% with enough training. Moreover, it is demonstrated that the fusion of behavioral information with hand geometry features leads to effective resistance to behavioral variability over time and our identity model exhibits good applicability to future behavioral data.

Security analyses are also conducted to demonstrate that the proposed method is resilient against common smartphone authentication threats such as smudge attack, shoulder surfing attack and statistical attack. Finally, a usability study shows user acceptance of our method.

Acknowledgments

We would like to thank Prof. Ellen Bass for her valuable comments and careful proofreading. We would also like to thank Prof. Xiaohong Guan, Prof. Qinghua Zheng and Prof. Roy Maxion for their kind support of this work, and the anonymous reviewers for their helpful comments. We also acknowledge the help from Mr. Tao Hua, Ms. Meilan Liu and Hexiang Wang in the data collection process. This work is supported in part by NSFC grants 61175039, 61375040 and 61221063. Zhi-Li Zhang was supported in part by NSF grants CNS-1411636, DTRA grant HDTRA1-14-1-0040 and ARO MURI Award W911NF-12-1-0385.

References

[1] S. Wiedenbeck, J. Waters, L. Sobrado, and J.-C. Birget, "Design and evaluation of a shoulder-surfing resistant graphical password scheme," *In Proceedings of the Working Conference on Advanced Visual Interfaces*, Venezia, Italy, 2006.

[2] A. J. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. M. Smith, "Smudge attacks on smartphone touch screens," *in Proceedings of the 4th USENIX Conference on Offensive Technologies*, Washington, DC, 2010.

[3] iPhone fingerprint sensor hacked with a finger made of clay at MWC 2016, http://www.techworm.net/2016/02/iphone-fingerprint-sensor-hacked-finger-made-clay-mwc-2016.html

[4] J. Mantyjarvi, M. Lindholm, E. Vildjiounaite, S. M. Makela, and H. A. Ailisto, "Identifying users of portable devices from gait pattern with accelerometers," *in Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP). IEEE International Conference on*, 2005, pp. ii/973-ii/976 Vol. 2.

[5] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Cell phone-based biometric identification," *in Biometrics: Theory Applications and*

Systems (BTAS), 2010 Fourth IEEE International Conference on, 2010, pp. 1-7.

[6] H. Lu, J. Huang, T. Saha, and L. Nachman, "Unobtrusive gait verification for mobile phones," *in Proceedings of the 2014 ACM International Symposium on Wearable Computers*, Seattle, Washington, 2014.

[7] F. Monrose and A. Rubin, "Authentication via keystroke dynamics," *in Proceedings of the 4th ACM SIGSAC Conference on Computer and Communications Security (CCS)*, Zurich, Switzerland, 1997.

[8] F. Monrose, K. M. Reiter, and S. Wetzel, "Password hardening based on keystroke dynamics," *International Journal of Information Security*, vol. 1, pp. 69-83.

[9] N. L. Clarke, S. M. Furnell, B. M. Lines, and P. L. Reynolds, "Keystroke dynamics on a mobile handset: a feasibility study", *Information Management & Computer Security*, vol. 11, pp.161 - 166, 2003.

[10] N. Zheng, K. Bai, H. Huang, and H. Wang, "You Are How You Touch: User Verification on Smartphones via Tapping Behaviors," *in Network Protocols (ICNP), 2014 IEEE 22nd International Conference on*, 2014, pp. 221-232.

[11] C. Giuffrida, K. Majdanik, M. Conti, and H. Bos, "I sensed it was you: authenticating mobile users with sensor-enhanced keystroke dynamics," *in International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, 2014, pp. 92-111.

[12] A. D. Luca, A. Hang, F. Brudy, C. Lindner, and H. Hussmann, "Touch me once and i know it's you!: implicit authentication based on touch screen patterns," *in Proceedings of the 2012 ACM CHI Conference on Human Factors in Computing Systems (CHI)*, Austin, Texas, USA, 2012.

[13] 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 Proceedings of the 19th Annual International Conference on Mobile Computing Networking (MobiCom)*, Miami, Florida, USA, 2013.

[14] S.-B. Napa, N. Memon, K. Isbister, and K. Ahmed, "Multitouch Gesture-Based Authentication," *IEEE Transactions on Information Forensics and Security*, vol. 9, pp. 568-582, 2014.

[15] C. Shen, Z. Yong, G. Xiaohong, and R. A. Maxion, "Performance Analysis of Touch-Interaction Behavior for Active Smartphone Authentication," *IEEE Transactions on Information Forensics and Security*, vol. 11, pp. 498-513, 2016.

[16] M. Harbach, A. De Luca, and S. Egelman, "The anatomy of smartphone unlocking: A field study of android lock screens," *in Proceedings of the 2016 ACM CHI Conference on Human Factors in Computing Systems (CHI)*, 2016, pp. 4806-4817.

[17] P. S. Teh, N. Zhang, A. B. J. Teoh, and K. Chen, "A survey on touch dynamics authentication in mobile devices," *Computers & Security*, vol. 59, pp. 210-235, 2016.

[18] M. Harbach, E. von Zezschwitz, A. Fichtner, A. D. Luca, and M. Smith, "It's a hard lock life: A field study of smartphone (un) locking behavior and risk perception," *in Symposium On Usable Privacy and Security (SOUPS)*, 2014, pp. 213–230

[19] 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 Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 136-148, 2013.

[20] K. N. Ross, "Sample design for educational survey research," Evaluation in Education. International Progress, vol. 2, pp. 105-195, 1978/01/01 1978.

[21] A. Ross and A. Jain, "A prototype hand geometry-based verification system," *in Proceedings of 2nd conference on audio and video based biometric person authentication*, 1999, pp. 166-171.

[22] M. Mukaka, "A guide to appropriate use of correlation coefficient in medical research," *Malawi Medical Journal*, vol. 24, pp. 69-71, 2012.

[23] S. Wang, C. L. Liu, and L. Zheng, "Feature Selection by Combining Fisher Criterion and Principal Feature Analysis," *in Machine Learning and Cybernetics, 2007 International Conference on*, 2007, pp. 1149-1154.

[24] Q. Gu, Z. Li, and J. Han, "Linear discriminant dimensionality reduction," *in Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 2011, pp. 549-564.

[25] C. Shen, Z. Cai, X. Guan, Y. Du, and R. A. Maxion, "User Authentication Through Mouse Dynamics," *IEEE Transactions on Information Forensics and Security*, vol. 8, pp. 16-30, 2013.

[26] S. Abe, Support vector machines for pattern classification vol. 2: Springer, 2005.

[27] K. S. Killourhy and R. A. Maxion, "Comparing anomaly-detection algorithms for keystroke dynamics," *in Dependable Systems & Networks (DSN), 2009 IEEE/IFIP International Conference on*, 2009, pp. 125-134.

[28] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *in Proceedings of the 14th international joint conference on Artificial intelligence (IJCAI)*. San Francisco, CA, USA. 1137-1143, 1995.

[29] M. W. Fagerland, S. Lydersen, and P. Laake, "The McNemar test for binary matched-pairs data: mid-p and asymptotic are better than exact conditional," *BMC Medical Research Methodology*, vol. 13, pp. 1-8, 2013.

[30] C.-C. Chang and C.-J. Lin, "LIBSVM: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, p. 27, 2011.

[31] A. Serwadda and V. V. Phoha, "When kids' toys breach mobile phone security," *in Proceedings of the 2013 ACM SIGSAC conference on Computer Communications Security (CCS)*, Berlin, Germany, 2013.

[32] A. Serwadda and V. V. Phoha, "Examining a large keystroke biometrics dataset for statistical-attack openings," *ACM Transactions on Information and System Security*, vol. 16, p. 8, 2013.

[33] V.-D. Stanciu, R. Spolaor, M. Conti, and C. Giuffrida, "On the Effectiveness of Sensor-enhanced Keystroke Dynamics Against Statistical Attacks," *in Proceedings of the Sixth ACM Conference on Data and Application Security and Privacy*, New Orleans, Louisiana, USA, 2016.

[34] Z. Cai, C. Shen, M. Wang, Y. Song, and J. Wang, "Mobile authentication through touch-behavior features," *in Biometric Recognition*, Springer, 2013, pp. 386-393.

[35] M. Sherman, G. Clark, Y. Yang, S. Sugrim, A. Modig, J. Lindqvist, et al., "User-generated free-form gestures for authentication: security and memorability," *in Proceedings of the 12th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys)*, Bretton Woods, New Hampshire, USA, 2014.

[36] N. Sae-Bae, K. Ahmed, K. Isbister, and N. Memon, "Biometric-rich gestures: a novel approach to authentication on multi-touch devices," *in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 977-986

[37] A. Boehm, D. Chen, M. Frank, L. Huang, C. Kuo, T. Lolic, et al., "Safe: Secure authentication with face and eyes," *in Privacy and Security in Mobile Systems (PRISMS), 2013 International Conference on*, 2013, pp. 1-8.

[38] I. Sluganovic, M. Roeschlin, K. B. Rasmussen, and I. Martinovic, "Using Reflexive Eye Movements for Fast Challenge-Response Authentication," *in Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS)*, 2016, pp. 1056-1067.