SilentSense: Silent User Identification Via Touch and Movement Behavioral Biometrics

Cheng Bo[†], Lan Zhang[*], Xiang-Yang Li[†]
[†]Illinois Institute of Technology, USA
[*]Tsinghua University, China
Email: cbo@hawk.iit.edu, zhanglan03@gmail.com, xli@cs.iit.edu

ABSTRACT

In this work, we present **SilentSense**, a framework to authenticate users silently and transparently by exploiting the user touch behavior and leveraging the integrated sensors to capture the micro-movement of the device caused by user's screen-touch actions. By tracking the fine-detailed touch actions of the user, we build a "touch-based biometrics" model of the owner by extracting some principle features, and then verify whether the current user is the owner or guest/attacker. When users are mobile, the micro-movement of mobile devices caused by touch is suppressed by that due to the large scale user-movement which will render the touch-based biometrics ineffective. To address this, we integrate a movement-based biometrics for each user with previous touch-based biometrics. We conduct extensive evaluations of our approaches on the Android smartphone, we show that the user identification accuracy is over 99%.

Keywords

Identification, Touch, Behavioral Biometrics, Security.

1. INTRODUCTION

The blooming digital service for mobile devices has attracted more privacy concern, especially when people are sharing their personalized device to guest users. Since device owners are not willing to take distrust action to reduce permission deliberately before sharing [4], it would be god for devices to silently know exactly who is using it, so as to provide necessary privacy protection and access control.

The most popular mechanism for authentication is using enhanced password patterns [2] with an additional security layer, and establishing guest profiles for access control. Such methods are overelaborated, inconvenient and time consuming, very few users are willing to employ such security mechanism in their devices. Another approach is facial recognition [1], which allows owners to choose the apps to be protected, and uses face as a key to open them. However the accuracy of facial recognition is a great challenge, and have risk of being imitated.

The latest solution exploits the capacitive touch communication as a mechanism to distinguish different users [6], which has potential risk of being imitated. TapPrints [5] indicates that taps on the touch screen could be observed through sensitive motion sensors. Touchalytics [3] only exploits scrolling as biometric for continuous authentication while [7] only considers tap behaviors on certain digit patterns.

In this work, we investigate the feasibility of utilizing the behavior biometrics that can be extracted from smartphone sensors silently for user identification. By exploiting the combination of several features from user's behavior while interacting with the device, we propose a non-intrusive user identification mechanism to substantiate whether the current user is the true owner or a guest or even an attacker who broke the passcode. As long as the current user is identified, necessary access control is triggered accordingly. The tapping behavior could be observed while the smartphone is in relatively static condition. However, the perturbation generated by the tapping may be suppressed by larger-scale user movement if the user is walking. Then we proposed novel techniques to assist extracting motion behavior biometrics for user identification.

We propose a novel model to estimate the probability of the current user being the owner of the device, and we also introduce other side signals to assist the identification, *e.g.*, the prediction of the next app being used according to the owner's usage pattern. In user identification, delay, accuracy and energy consumption are three main issues that we have to consider We design a strategy to optimize the accuracy while guarantee the identification delay and energy consumption.

2. SYSTEM MODEL

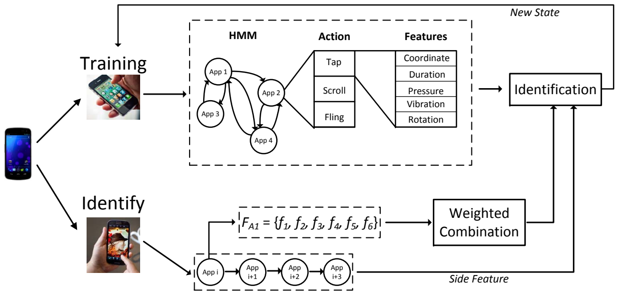


Figure 1: Framework Overview.

The framework model consists of two separate phases: *Training* and *Identification*, as shown in Figure 1. The training phase is conducted to build a biometrics model when the owner is interacting with the device, and after the owner behavior model is established, the system switches to the identification phase. Initially, the framework will train the owner's behavior model by retrieving information of taps or other operation mode, and the reaction from the device while such interaction occurs. For one operation on the device, the framework could capture multiple information, including:(1) the coordinate on the screen of both touch down and release; (2) the duration of one interaction; (3) the sensory data from both accelerometer and gyroscope, (4) the pressure for the finger touching on the screen, and (5) the motion condition of the user.

Since operating certain app may consist of multiple gestures, including tap, fling, and scroll, even different gesture on the same app may generate different reactions of the device. Therefore, we combine the app with certain operating gesture and the information that the framework could capture as the feature patten for modeling the owner, denoted as $O_i = \{A_i, G_i, f_{i1}, f_{i2}, f_{i3}, f_{i4}, f_{i5}, f_{i6}\}$. Here A_i is the App being used, G_i is the gesture, and $f_{i,j}$ ($j \geq 1$) are various features about the gesture.

On the other hand, the continuous monitoring and identification will consume a large amount of energy. Our identification process works according to a comprehensive model to balance the identification accuracy, delay and the energy consumption.

2.1 Interacting with Apps

The operation of touchscreen based mobile device mainly consists of four modes: Tap, Scroll, Fling, and Multi-touch. Obviously, most of the apps currently support more than one operating mode. Suppose $A = \{Action_1, Action_2, Action_3\}$ indicate the three operation mode for App *A*. And for each action A_i, we could extract a set of features containing the coordinate of the touch, the duration, fluctuation on both acceleration and gyroscope, and the current motion condition. The set of feature could be presented as $F_{Ai} = \{f_1, f_2, f_3, f_4, f_5.f_6\}$.

2.2 Identification Process

The purpose of the framework is to identify the current user of the device, and prevent sensitive information leakage if the user is not the legal owner. Generally, the three important issues that users concern about are *delay*, *accuracy*, and *energy consumption*.

We employ SVM to judge the identity of the current user according to each interacting behavior observation. Since it is difficult to validate the correctness of the results because of lacking of ground truth, we denote ε_i as the creditability of the result, which is available for the SVM. Obviously, the accuracy of the identification process depends on the amount of observations, thus we denote $\theta_i(X_1, X_2, \cdots, X_i)$ as the accuracy of the identification based on the sequence of accumulated observation until X_i. With the number of observation increases, the framework will be more confident to provide the correct result, and the overall confidence is generated according to historical creditability:

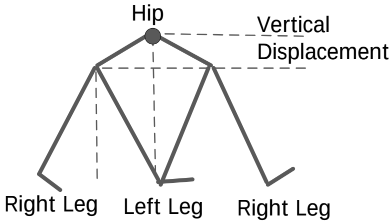


Figure 2: The walk model of users.

$$\theta_i(X_1, X_2, \cdots, X_i) = 1 - (\prod_{j=1}^{n}(1 - \varepsilon_j(X_j)))$$

On the other hand, framework in mobile device cannot neglect the energy consumption, coming from feature extraction and running the identification. We assume the function of $U(E_t, T_t)$ as the identification accuracy that can be achieved under the energy budget E_t and remaining time T_t. Thus, under limited budget on both energy and delay, we have

$$U(E_t, T_t) = \max_{cur \in \{b,c\}} (1 - (1 - U_{cur}(E_{cur}, T_{cur})) \times (1 - U(E_t - E_{cur}, T_t - T_{cur})))$$

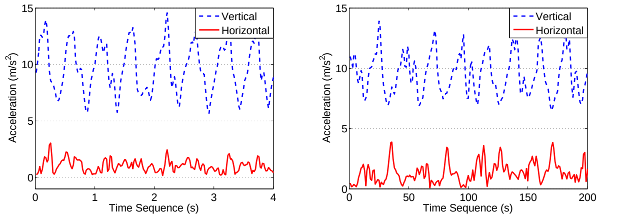
Here cur stands for the current decision, which is either waiting for the next touch behavior (cur is b) or other energy consuming assistant methods such as facial recognition (cur is c). The framework will make dynamic decision so that the expected identification accuracy could be maximized. We can also set the threshold on accuracy while minimizing the energy cost.

2.3 Motion Analysis

The amplitude of the sensory data extracted from the motion will be much larger than that of small perturbation caused by touch action, and the latter may be swamped by the former so that it fails to be extracted as a feature. In our work, we analyze the motion features when the user uses the mobile phone while walking and holding it in front of the chest, and uses the walking features as part of the behavioral biometrics for identification.

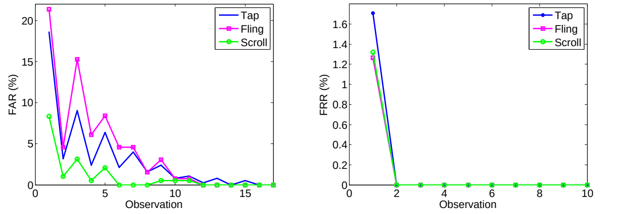
To accurately capture the walking features of different users, we extract raw acceleration vector from the accelerometer in the phone coordinate system, and convert to the earth coordinate system, *i.e.* north, east, gravity. As shown in Figure 2, when a user is walking the vertical displacement of his/her hip is directly correlated to his/her stride length, which is an important feature of different walker. So we obtain the vertical displacement of each step by double integration of the filtered vertical acceleration.

As shown in Figure 3, the step frequency and horizontal acceleration pattern also vary with different users. To sum up, we extract four features for the filtered vertical and horizontal acceleration: (1) Vertical displacement of each step; (2) Current step frequency, calculated by the duration of each step; (3) Mean horizontal acceleration of each step; (4) Standard deviation of each step.



(a) The acceleration in the earth coordinate system of the owner. (b) The acceleration in the earth coordinate system of a guest.

Figure 3: The acceleration pattern in the earth coordinate system while walking for different users.



(a) FAR (b) FRR

Figure 4: FAR, and FRR by different actions and different number of actions observed.

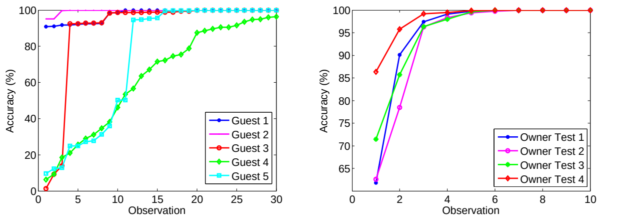
3. PRELIMINARY RESULTS

3.1 Identification in Static Scenario

We extracted features from different users, including interacting duration, pressure on screen, and both the vibration and rotation generated simultaneously when interacting with three different apps. And we notice that the operating pattern difference is large between users, which mainly comes from the habit of holding the phone, the force on the screen when interacting.

We evaluate the performance of operating in three different modes through three different apps, including *Message*, *Album*, and *Twitter*. In Figure 4, we plot the false acceptance ratio (FAR), and false rejection ratio (FRR) of identifying user by different actions with different number of total observed actions. From Figure 4(a), the mean FAR of identification through tap alone is as high as 22% with single Fling action, Tap follows with 18%, and using Scroll information, the FAR is 8%. The FAR is reduced to below 1% after observing about 15 actions. The reason for the high accuracy is that the number of observations is sufficient, and the reaction of the smartphone especially the vibration and rotation when encountering perpendicular touch differentiate. Surprisingly, Figure 4(b) shows that the FRR is almost 0 with only 2 observations for each of the actions.

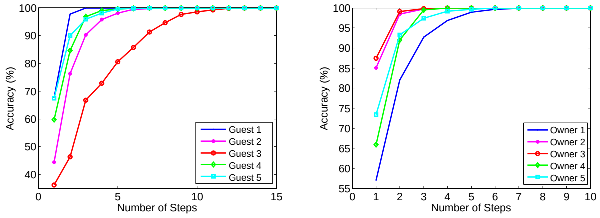
Then, we evaluate the performance of **SilentSense** in a more general scenario. The same 5 users are required to operate with the smartphone without any constraints, which means that the actions are randomly occurred. Since the using habit leads to the difference of behavior pattern, the framework could reach high accuracy after a small amount of events, as shown in Figure 5(a). In our experiments, we discover that if the user is a guest, the framework will have to spend more time to achieve acceptable accuracy, but the identification of owner will be much quicker. Even so, the framework could reach over a 80% accuracy within ten event observations according to Figure 5(a), and the owner will be judged with in 6 observations, as shown in Figure 5(b).



(a) Guest Accuracy (b) Owner Accuracy

Figure 5: Identification based on combined operation.

3.2 Identification in Dynamic Scenario



(a) Guest Accuracy (b) Owner Accuracy

Figure 6: Identification based on walking feature.

In the dynamic scenario, we extract 4 walking features, including vertical displacement, step duration, mean and standard deviation of horizontal acceleration. and establish a SVM model for walking features.

The same users are required to use the smartphone while they are walking freely. We collect their processed vertical and horizontal accelerations in the earth coordinate system. After collecting necessary information, we combine the walking features with touch event features to establish the SVM model. And such touch event features only contains the duration, pressure, and the operation mode. Figure 6 presents the achieved identification accuracy increases with observed steps. As shown in Figure 6(a), after 12 steps, the accuracy to identify a guest can achieve 100%. Similarly, Figure 6(b) shows that after 7 steps, the accuracy to identify the owner can achieve 100%.

4. REFERENCES

[1] Visidon applock. http://www.visidon.fi/en/Home.

[2] De Luca, A., Hang, A., Brudy, F., Lindner, C., and Hussmann, H. Touch me once and i know it's you!: implicit authentication based on touch screen patterns. In *CHI* (2012), ACM, pp. 987–996.

[3] Frank, M., Biedert, R., Ma, E., Martinovic, I., and Song, D. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *TIFS* (2013).

[4] Karlson, A., Brush, A., and Schechter, S. Can i borrow your phone?: understanding concerns when sharing mobile phones. In *ACM CHI* (2009), pp. 1647–1650.

[5] Miluzzo, E., Varshavsky, A., Balakrishnan, S., and Choudhury, R. R. Tapprints: your finger taps have fingerprints. In *ACM MobiSys* (2012), pp. 323–336.

[6] Vu, T., Baid, A., Gao, S., Gruteser, M., Howard, R., Lindqvist, J., Spasojevic, P., and Walling, J. Distinguishing users with capacitive touch communication. In *ACM MobiCom* (2012), pp. 197–208.

[7] Zheng, N., Bai, K., Huang, H., and Wang, H. You are how you touch: User verification on smartphones via tapping behaviors.