In this work, we present SilentSense, a framework to authenticate users silently and transparently by exploiting the user touch behavior biometrics 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 using the smartphone, the unique operating pattern of the user is detected and learnt by collecting the sensor data and touch events silently. 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 90%.

Categories and Subject Descriptors
H.4 [Information Systems Applications]: Miscellaneous

Keywords
SilentSense, Mobile Device, Identification

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 good 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 over-laborated, inconvenient and time consuming. Facial recognition [1] by front camera is another optional strategy to identify user. But it is still annoying to require users to accept frequent shooting. Besides, the accuracy is unstable with changing environment and frequent shooting is power-consuming.

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 [8] only considers tap behaviors on certain digit patterns.

In this work, investigating the feasibility of utilizing the behavioral biometrics extracted from smartphone sensors for user identification, we propose SilentSense, a non-intrusive user identification mechanism to silently substantiate whether the current user is the device owner or a guest or even an attacker. Exploiting the combination of several interacting features from both touching behavior (pressure, area, duration, position) and reaction of devices (acceleration and rotation), SilentSense achieves highly accurate identification with low delay. A great challenge comes from the circumstance when the user is in motion, such as walking. The perturbation generated by the interacting will be suppressed by larger-scale user movement. While most of existing works neglect this circumstance, SilentSense is capable of identifying user in motion by extracting the motion behavior biometrics. As long as the current user is identified, necessary access control is triggered automatically.

Continuous monitoring sensors will provide minimum guest identification delay, but could cause unwanted energy consumption. Facing this challenge, we propose a novel model to estimate the current user leveraging the observation of owner’s sociable habit. An online decision mechanism is designed for the timing to turn on or turn off sensors, which provides a balance between energy cost, delay and accuracy. Our online decision mechanism results in an adaptive observing frequency according to owner’s social habit.

2. SYSTEM MODEL

SilentSense is designed as a pure software-based framework, running in the background of smartphone, which non-intrusively explores the behavior of users interacting with the device without any additional assistant hardware.

2.1 Main Framework

The framework model consists of two basic phases: Training and Identification, as shown in Figure 1. The training phase is conduct-
ed to build a behavior model when the user is interacting with the
device, and the identification phase is implemented to distinguish
the identity of the current user based on the observations of each
individual’s interacting behaviors.

Initially, the framework trains the owner’s behavior model by
retrieving two types of correlated information, the information of
each touch-screen action and the corresponding reaction of the de-
vice. With the owner’s behavior model, the current user will be i-
dentified through SVM classification. As personalized devices, the
initial identification accuracy is usually not high enough because of
lacking of non-owner’s pattern. The model will upgrade the SVM
model by adding newly observed features gradually through self-
learning, which provides more accurate judgement.

2.2 Interacting Model

The operation of touchscreen based mobile device mainly con-
sists of four gestures: Tap, Scroll, Fling, and Multi-touch. Different
gestures usually have different touch features and lead to different
device reactions. Interacting with certain app often involves a cer-
tain set of gestures. For a touch action \( T_i \), we combine the app with
its touch gesture and the features captured by this framework as one
observation, denoted as \( O_i = \{ A_i, G_i, f_{i,1}, \ldots, f_{i,n} \} \). Here \( A_i \)
is the app being used, \( G_i \) represents the gesture (e.g. tap), and \( f_{i,j} \)
\((j \geq 1)\) are features of the observed action.

Two types of features are used in this system: the touch features
and reaction features. The touch features include touch coordinate
on the screen, touch pressure and duration, which can be obtained
from system API. To capture the reaction features, we notice that
diverse gestures and positions for holding the device by individual
users infer different amplitudes of vibration caused by each touch,
which has already been proved by previous works (\cite{5, 7}). Such
tiny reaction of the devices produces an identifiable patterns which
could be observed via accelerometer and gyroscope.

2.3 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 ac-
cording to each interacting behavior observation. Since it is diffi-
cult to validate the correctness of the results because of lacking of
ground truth, we denote \( \epsilon_i \) as the probability of the result, which
is available for the SVM. Obviously, the accuracy of the identifica-
tion process depends on the amount of observations, thus we denote
\( \theta_i(X_1, X_2, \ldots, X_n) \) as the probability of the identity based on
the sequence of accumulated observation until \( X_n \). With the number
of observation increases, the framework will be more credible to
provide the correct result, and the overall probability is generated
according to historical credibility:

\[
\theta_i(X_1, X_2, \ldots, X_n) = 1 - \prod_{j=1}^{n} \left( 1 - \epsilon_j(X_j) \right)
\]

Figure 1: Framework Overview.

Figure 2: The frequency feature of acceleration in the earth
coordinate system while walking.

On the other hand, framework in mobile device cannot neglect
the energy consumption, coming from feature extraction and run-
ning the identification. We assume the function of \( U(E_1, \theta_t) \) as
the utility that could be achieved under the energy budget \( E_t \)
and maintaining the reliable overall probability for the identification.
The framework will make dynamic decision to balance the two fac-
tors so that the expected utility could be maximized. We can also
set the threshold on the overall probability while minimizing the
energy cost.

2.4 Motion Analysis

The amplitude of the sensory data extracted from the motion will
be much larger than that of small perturbation caused by touch ac-
tion, 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, and use the walking features as part of the behavioral bio-
metrics for identification.

To accurately capture the walking features of different users,
three steps are conducted in our method. Firstly, considering a user
could hold the phone in any attitude, we convert the raw accel-
eration vector in phone coordinate system into the earth coordinate
system in real time. There are a lot of walking independent noise in
the acceleration, which will greatly confuse the walking feature de-
tection. We analyze the acceleration while walking in the frequency
domain, Figure 2(a) shows that, the energy mainly locates around
2Hz, which is the user’s walking frequency. The energy in other
frequency comes from noise. To extract the pure walking acce-
eration, secondly, we filter linear acceleration with a band pass filter.
Then we get the vertical acceleration in the gravity orientation and
horizontal acceleration. Figure 2(b) shows the filtered vertical acce-
erlation. A simple step detection algorithm can be performed on
the filtered vertical acceleration in real time. Thirdly, we extract
the walking feature from the processed acceleration data. The ver-
tical displacement of a walker is directly correlated to his/her stride
length and height, hence it is an important feature. Besides, the
step frequency and horizontal acceleration pattern also vary with d-
iferent users. To sum up, we extract four features of walking from
\( E_{A_1} \) and \( E_{A_2} \): 1) Vertical displacement of each step by double
integration of \( E_{A_1} \); 2) Current step frequency, calculated by the
duration of each step; 3) Mean horizontal acceleration for each
step; 4) Standard deviation of \( E_{A_2} \) for each step.
about 20%. But, with about 12 observations of various actions, the FAR and FRR are both reduced to nearly 0, meaning that there is no incorrect identification.

### 3.2 Identification in Dynamic Scenario

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 dynamic 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 5 presents the achieved identification accuracy increases with observed steps. As shown in Figure 5(a), after 12 steps, the accuracy to identify a guest can achieve 100%. Similarly, Figure 5(b) shows that after 7 steps, the accuracy to identify the owner can achieve 100%.

### 5. REFERENCES


