A Novel Identification/Verification Model Using Smartphone's Sensors and User Behavior

Dandachi Ghina, Bachar El Hassan, Anas El Husseini

A Novel Identification/Verification Model Using Smartphone's Sensors and User Behavior

Ghina Dandachi
Centre Azm, EDST
Lebanese University
Tripoli, Lebanon
G_dandachi@hotmail.com

Bachar El Hassan
Department of Electricity and Electronics
Lebanese University
Tripoli, Lebanon
Bachar_elhassan@ul.edu.lb

Anas El Husseini
Centre Azm, EDST
Lebanese University
Tripoli, Lebanon
SAMOVAR, Telecom SudParis
Evry, France
Linux.anas@gmail.com

***Abstract*—Smartphones are increasingly entering people's life; every person in the house carry one or two smartphones (Android, iPhone, Tab…). They use explicit authentication, which is inefficient; once the smartphone is stolen, a thief can steal personal information stored on the phone and can access all services that might have the password stored. In addition, elderly and physically impaired users need to have their medical profile secured and easily accessed without password limitation. For this reason, smartphone sensors are good candidates for providing an implicit authentication.**

This work introduces a new perspective of context-based user authentication: users can be authenticated implicitly using data captured by sensors of the smartphone and user behavior; these data are used in the creation of a unique profile for each user. The proposed model is supposed to be as secure as traditional authentication methods.

Keywords—implicit authentication; smartphone; sensors; context; physical impairement; elderly; user behavior

I. INTRODUCTION

The smartphone world is expanding at a rapid pace introducing a wide demand of online applications, services, and Internet access. Thus, smartphone security has several threats, a thief can be authenticated explicitly, and access personal data if he knows the password or the pin code which is not hard to know. Implicit authentication does not need a clear input from users; it is possible to authenticate the user from the context around him, using smartphone's sensors and behavioral information collected about the user. Generally, authentication and identification processes are increasing with the increase of services, causing the user, in general, to use short, easy, and repetitive passwords. These facts in addition to the possibility of stealing or losing the mobile, leads to security threats on the personal data [1]. More particularly, elderly and physically impaired users find difficulties by entering pin codes or screen patterns, or might be unable to do explicit authentication. Searching a novel solution for this problem, researchers need to find the better user-friendly authentication process that protects the data and offers a better security for the user without involving him at every issue.

However, solutions were found such as behavioral-based authentication and context-aware authentication that process implicit authentication without involving users in the authentication procedure unless in case of security threats. These solutions study only one area of the user context: software or hardware. This study benefits from all possible data in order to improve the implicit authentication of the user: it combines sensor-based authentication and some features of the behavioral based authentication.

In this paper, we propose a novel authentication/identification technique based on the values measured by the mobile sensors and complementing them with user behavioral information collected gradually from user's inputs and acts. In other words, we leverage on Hardware and Logical information found in the smartphone. This combination of information provides a higher authentication accuracy.

This technique benefits from the sensors implemented in the smartphone and the information collected in order to achieve a more secure, more power-aware and user-friendly authentication/identification process. Our challenge is to build a model that implements this type of implicit authentication, create the client's profile, ensure his privacy towards service providers in order to ensure the data safety and offer the user an authentication method with the same level of security that could have using explicit authentication methods.

The main contribution of our work includes (i) sensors study and choosing of an independent and effective set of sensors (ii) Behavioral aspects choosing for User identification (iii) Profile creation and assignment and (iv) System model implementation that provide implicit user's authentication/identification.

This paper is organized as follows: Section 2 presents a literature review. The proposed model is described in section 3. The methods and results are presented in section 4 and 5 respectively. Section 6 contains the conclusion and the future work.

II. Literature Review

The main purpose in this study is to propose an alternative to users of smartphone, especially dependent people, different from explicit authentication by using context information, leading to an implicit authentication. In a comparison with the state-of-art in the context-based authentication field, we represent a set of related works already done.

A. Behavioral User Authentication/Identification

A user behavioral model is provided in [1], the corresponding research is based on the idea that the person is a creature of habit, therefore each event has a correlation between two fundamental attributes: space and time. The proposed architecture uses resources found in the mobile devices: User calls, user schedule, GPS, device battery level, user applications, and sensors. This model is clearly implemented by same authors in [2].

A similarity model for context-aware recommendation system was proposed in [3], in this study, the recommendation system depends of the user's past/current/ future contexts and it is elaborated using the similarity calculations between two contexts and actions. Considering A: action and C: context, we can have user action pattern:

$$< \cdots, (A-1, C-1), (A, C), (A+1, C+1) \ \ldots >$$

According to Clark and Furnelle [4], a continuous and implicit user authentication could be achieved using keystroke analysis. Thus, this study has not a good performance, for two reasons. First, the user's variation of using the mobile could cause problems, and secondly, this method needs a high computational power because of feed forward multi-layered perceptron (FF-MLP) usage.

B. Sensor Based User Authentication/Identification

Researchers proposed SenGuard [5] as a new user identification framework that offers continuous and implicit user identification service for the smartphone. SenGuard is a new passive authentication technique that leverages the sensors available on the smartphone. It uses four sensors: voice, multi-touch, locomotion and location. These sensors are processed together in order to get the user identification features implicitly, explicit authentication is performed only when there is an important evidence that the user has changed.

C. Biometric Based Authentication

Biometric based authentication is an important alternative for explicit authentication. However, according to the limitations in the size, power efficiency, and mainly the cost of a smartphone, the usage of physiological features on a mobile are less attractive and researchers omitted the emergence on these kinds of recognition in a smartphone.

Several studies experimented user authentication/ identification using gait recognition as a possible implicit authentication method. The proposed approach uses acceleration signal and detects the person's way of walking [6]. A motion-recording device was used [7], [8] in order to measure the acceleration according to the three axes, there were multiple algorithm proposed including histogram similarity, and cycle length measurement techniques.

Therefore, these three approaches present several propositions for implicit authentication/verification. We worked on the combination of the best part in these works to deliver a more reliable implicit authentication model.

III. Proposed Model

In order to identify necessary sensors and behavioral aspects that can give a good implicit authentication/identification model, an implementation of data acquisition and analysis process is proposed. Consequently, the work consists in usage of sensor values offered by the context that surrounds the user. Behavioral parameters are used beside sensor values in order to give results that are more accurate as shown in *Figure 1*.

A. Sensors

The first stage of work is to acquire data from smartphone's sensors. The data collection is achieved by implementation of android applications whose role is to write acquired values on the external memory of the smartphone. The programming language chosen is java for android applications, because it offers a native library for working with android.

The sensors in smartphones can be classified into three types: Motion, environmental and position sensors. Since not all smartphones contain environmental sensors, the choice of sensors will be between motion and position types.

According to sensors definition in [9], there is redundancy between different Sensor values; some sensors are hardware, while others are software and get their values by applying mathematical formulas on one or two sensors. Thus, the number of sensors is reduced by eliminating duplicated sensors usage. Consequently, from six sensors-Accelerometer, Linear Acceleration, Gravity, Orientation, Magnetic and Rotation Vector-, there is three independent sensors. These sensors are Accelerometer, Orientation and Rotation Vector sensor.

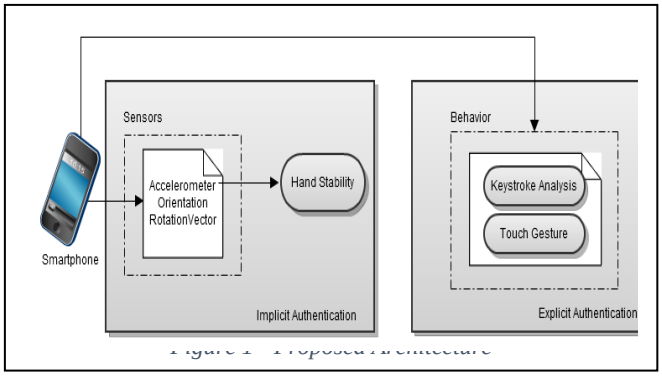
There are also Proximity and sound level sensors, but these sensors are eliminated by applying a classification, presented in Section 4, according to the group of sensors in study, because these sensors did not improve the results of classification and testing. Based on this analysis, the GPS parameters add efficiency for the classification, thus it is chosen to be used in this study.

B. Behavioral Data

Behavioral data chosen are keystroke analysis, touch gesture analysis and hand stability recognition, these behavioral aspects are important and very different between users for whom the study is achieved. Each user have a different type of impairment, thus a different profile of using the smartphone.

The keystroke analysis is the study of writing velocity, and the error rate. Keystroke analysis is mentioned in several studies as an important feature [4], thus it cannot be used alone because it does not give accurate results.

Touch gesture analysis gives inputs that are user specific; these inputs can be used in implicit authentication process. Touch gesture types are several: flick, spread, pinch, drag and tap. These gestures are differently used between different users.



For example, users suffering from hand problems use tap option instead of spread.

Hand stability can be obtained by performing data analysis on accelerometer and orientation sensors values. Definitely, hand stability of a Parkinson user differs from a user having heart failure problem, and differs from a user that have no hands at all and uses a pen in his mouth or toes. This difference offers an important parameter that boosts the classification accuracy.

This section clearly introduces the proposed model for implicit authentication. The three built-in sensors, the GPS, the hand stability recognition are used for implicit authentication. In the other hand, when implicit authentication fails the system have a backup plan by using keystrokes values and touch gesture analysis. The distinction between users is done using Support Vector Machine (SVM) classification.

IV. METHODS

The validation of this model is performed on a set of three users. The data is recorded on 24 hours, than classified and tested by SVM neural network. The verification uses the data recorded from another day, and pick slot time randomly.

A. Data

Data Collection is applied for three users. The first stage was to collect data from the three chosen sensors (Accelerometer, Orientation and Rotation Vector). The parameter extraction for the three sensors is made on the three axes for each sensor (x, y, and z); these parameters are mean value and standard deviation value. In addition, there is the two values of GPS (latitude and longitude). Consequently, the total number of parameter is 20; these parameters are used in the classification process, testing and validation.

B. Classification

The proposed method of analysis of data is SVM method, SVM is chosen because of many characteristics that are needed in this research. SVM is capable of delivering higher performance in terms of accuracy and tuning of model selection and kernel function, especially non-linear kernel functions to solve classification problems of non-linearly separable data [10]. It scales relatively well to high dimensional data [11]. Furthermore, Radial Basis Function (RBF) is chosen as kernel function because it is the best kernel for multi-class classification [10].

C. Verification

A verification work was achieved on this classification, using data from another set of records. The results obtained by verification could be refined, for this reason, an algorithm of filtering is proposed and tested.

This section presents the methods followed in order to achieve the classification. The results of these methods are shown in the next section.

V. RESULTS

According to the steps listed in section 4, this section comes to prove the model by giving the obtained results during classification, verification and results refinement. The false negative and false positive rates are calculated for the verification phases.

A. Choice of Time Slot

The parameter calculation is made on slot times of 20 minutes, which permits to obtain a profile of the user for all the day of registration. The choice of slot time is chosen among a set of slot times {10, 20 and 40 min}. The classification of data using 20 minutes of slot time shows close results to those using 10 minutes, and better results than those using 40 minutes.

B. Classification

After the classification with 90% of the data, a testing using 10% of the data was made. The results of testing are presented in the *Table 1*. The measured value in the table is the classification error rate for each user.

C. Verification

The SVM classifier gives acceptable results; they can be more accurate by adding more parameters and by filtering the classifier's support vectors. The following *Table 2* gives a glance about results obtained. By classifying user 1 using the three classifiers, the False Negative results for the classification of user 1 by its own classifier is around 20%; using the other two classifiers the false positive rate varies between 0 and 20%.

D. Results refinements

The obtained results in the verification phase are not acceptable, but they can be more accurate by applying a kind of filtering on the support vectors for each one of the three classifiers. The adopted approach, used to achieve this goal, is to eliminate the support vector points that add noise to the classification. For this reason, a MATLAB code is implemented. The code is used as a filter to de-noise the set of support vectors for each classifier.

After applying this filtering algorithm, the results are more accurate. As shown in the table 3, the false negative obtained by classifying user 1 by its own classifier is 5% instead of 20% without filtering. The false positive rate obtained by the two other classifiers became 0%.

Table 1 - Classification error rate for SVM

	User 1 v/s others	User 2 v/s others	User 3 v/s others
Acceleration	16	23	24
Orientation	17	10	19
Rotation Vector	12	18	21
3 Sensors + GPS	13.5	7	17.4

Table 2 - Verification false rate without support vectors filtering

User 1 classifier	User 2 classifier	User 3 classifier
20.7%	0%	20%

Table 3 - Verification false rate with support vectors filtering

User 1 classifier	User 2 classifier	User 3 classifier
5%	0%	1%

This section shows the classification and verification results using the SVM neural network. The classification is applied on three users. The classifier for each user is used in the verification phase, thus verification results are improved by applying a filtering algorithm (de-noising).

It should be pointed that this study is not supposed, for the moment, to distinguish every person in the world. The targeted users are the physically impaired and elderly ones, whom each has a particularity that makes his profile somehow different from others.

VI. Conclusion and future Work

As this paper clearly shows, implicit authentication is a very important option that is necessary to employ in smartphones. Physically impaired and elderly users find difficulties by entering explicit passwords and in the other side normal users also find the explicit authentication as a very annoying process. In addition, the explicit authentication does not offers user verification. However, the proposed model offers an implicit authentication implementation. This model shows important results according to the current stage of work, using three motion sensors and GPS values. These results have become more accurate by applying the filtering algorithm on the classifier data.

The future work is adding the hand stability recognition to the study of the implicit authentication, studying the touch gesture and keystroke recognition for the explicit authentication process that is proposed as a back-up authentication model in case of implicit authentication failure. Another aim of the work is the creation of a unique profile for each user from this data. Finally, implicit authentication for smartphones is a very important study that every person can cope and become involved with.

References

[1] J. Lima, C. Rocha, I. Augustin and M. Dantas, "A Context-Aware Recommendation System to Behavioral Based Authentication in Mobile and Pervasive Environments," in IFIP Ninth International Conference on Embedded and Ubiquitous Computing, 2011.

[2] Rocha, Lima, Dantas and Augustin, "A2BeST: An Adaptative Authentication Service Based on Mobile User's Behavior and Spatio-Temporal Context," in Computers and Communications (ISCC), 2011 IEEE Symposium, 2011.

[3] Oku, Nakajima and Miyazaki, "A Recommendation System Considering Users' Past / Current / Future Contexts," CARS-2010, 2010.

[4] Clarke and Furnell, "Authenticating mobile phone users using keystroke analysis," International Journal of Information Security, vol. 6, no. 1, pp. 1-14, 2006.

[5] Shi, Yang, Jiang, Yang and Xiang, "SenGuard: Passive User Identification on Smartphones Using Multiple Sensors," in IEEE 7th International Conference on Wireless And Mobile Computing, Networking And Communication (WiMob), 2011.

[6] Mantyjarvi, Lindholm, Vildjiounaite, m. Makela and Ailisto, "Identifying users of portable devices from gait pattern with accelerometers," in IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.

[7] Gafurov, Helkala and Soendrol, "Biometric gait authentication using accelerometer sensor," JOURNAL OF COMPUTERS, vol. 1, no. 7, pp. 51-59, 2006.

[8] D. Gafurov, E. Snekkenes and P. Bours, "Gait Authentication and Identification Using Wearable Accelerometer Sensor," in IEEE Workshop on Automatic Identification Advanced Technologies, 2007.

[9] "Sensors Overview | Android Developers," [Online]. Available: http://developer.android.com/guide/topics/sensors/sensors_overview.html#. [Accessed 27 03 2013].

[10] D. K. SRIVASTAVA and L. BHAMBHU, "DATA CLASSIFICATION USING SUPPORT VECTOR MACHINE," Journal of Theoretical and Applied Information Technology, 2009.

[11] V. Jakkula, "Tutorial on Support Vector Machine (SVM)," Washington, 2006.