# Mobile User Identification through Authentication using Keystroke Dynamics and Accelerometer Biometrics

Kyle R. Corpus*
kylecorpus@outlook.com

Ralph Joseph DL. Gonzales*
rjdgonzales@gmail.com

Alvin Scott Morada*
scott.morada@gmail.com

Larry A. Vea
Mapua Institute of Technology*
333 Sen. Gil Puyat Avenue
Makati City, Philippines
lavea@mapua.edu.ph

## ABSTRACT

Biometrics is everything that can be measured in a human being. It has two types; behavioral and physiological. This paper discusses the use of keystroke dynamics, a form of behavioral biometrics that deals with the measure of how a person types, and the utilization of accelerometer biometrics as a form of behavioral biometric that measures how a person holds his mobile device. We collected biometric data from 30 volunteer participants by asking them to enter their 8-16-character password specimens 8 times using a customized tool in a mobile phone. The first 6 collection from each participant was set aside for the training set while the other 2 is for the test set. The data were then processed and extracted keystroke dynamic and accelerometer biometrics using a customized tool written in Java. Several well-known classifiers were trained using keystroke dynamic features alone, accelerometer biometrics alone, and the combination of both. Results show that Neural Network classifier using the combined features gave the most acceptable model. The model performance was further improved by removing some low ranking features defined by the Chi Square attribute evaluator and by removing some features that are highly correlated to other features.

## CCS Concepts

• **Security and Privacy → Security Services → Authentication → Biometrics.**

## Keywords

Biometrics; keystroke dynamics; digraph; trigraph; accelerometer.

## 1. INTRODUCTION

In recent years, mobile applications related to online financial transactions have been developed to provide convenience to its users. There are recent innovations in mobile commerce that have enabled users to do transactions using their mobile device. These applications include purchasing goods, banking, and process point-of-sale payments [1]. However, users are concerned with the security of their data. Due to the sensitivity of information, some users are afraid that their personal information will be stolen or hacked, so instead, they prefer not to use these applications and go the traditional way, thus defeating the purpose of mobility [2].

In this study, we aimed to: (1) develop a model for user authentication in mobile applications through keystroke dynamics and accelerometer biometrics; and (2) implement the model in a real world scenario through a prototype.

This study also tries to check if integrating the accelerometer biometrics will improve authentication, since we believe that the mobility or motion of a mobile device depends on how the user holds and type on it.

Our study is limited to the use of capacitive display enabled mobile devices, which is a more suitable type of mobile device display for interpreting user's behavior when typing textual passwords. Our study is limited to mobile devices using Android OS, for testing purposes only.

## 2. METHODOLOGY

### 2.1 Preparation

In preparation to data gathering, we first developed a data gathering tool using Android Studio 1.3.2. This tool records the keypress timestamps, accelerometer biometrics and the textual password. We also customized a QWERTY soft keyboard and a customized Login screen.

### 2.2 Data Gathering

We gathered data from 30 volunteer participants. We asked each participant to create an 8–16 alpha-numeric with at least one (1) special-character password. The data gathered were saved in two (2) text files. The first file contained the following data: keydown timestamp, followed by the value of the pressed key, then the value of the keyup timestamp which are separated by commas. The second file recorded the accelerometer biometrics for x, y and z axes. We collected eight (8) password specimens from each participant.

### 2.3 Feature Extraction

A number of keystroke dynamic features and accelerometer biometrics were used in this study. For the keystroke dynamic features, we used some statistical measure such as average, variance, and standard deviation for digraph (2G) keystroke (measures on two consecutive keys), trigraph (3G) keystroke (measures on three consecutive keys), hold time, and typing's completion time. For the acceleration, we used the average,

*MobileSoft'16,* May 16-17 2016, Austin, TX, USA.
ACM 978-1-4503-4178-3/16/05.
http://dx.doi.org/10.1145/2897073.2897111

variance, standard deviation, minimum, and maximum of the x, y and z axes accelerometer biometrics.

## 2.4 Model Development and Validation

The datasets derived in Section 2.3 were divided into two (2) subsets: the first set is called the training set which is composed of six (6) instances of every participant's password specimen while the second set is called the test set which comprised the remaining two (2) instances.

With the aid of RapidMiner data mining tool, models were developed by training some well-known classifiers used in previous studies [e.g. 3]. These include Decision Tree, J48, Naive Bayes and Neural Networks, and using 10-fold cross-validation for model validation.

To find the most suitable model for mobile identification through authentication using keystroke dynamics and/or accelerometer biometrics, we conducted several experiments:

First, we trained the aforementioned classifiers using three (3) features sets: keystroke dynamic features only; accelerometer biometrics only; and the combination of both. This was done to determine the effect of accelerometer biometrics when combined with some keystroke dynamic features. Results showed that the model generated by the neural network classifier performed much better than the others in terms of accuracy and kappa statistic. It was also observed that the model is best when both keystroke and accelerometer features were combined compared when the two were separated.

Next, to find an optimized model performance, we conducted feature selection in two ways: First, we ranked the features using Chi Square attribute evaluator and iteratively removed features on the lowest rank one at a time until we found the model with the highest performance in terms of accuracy and/or kappa. We found out that when the variances of the 2G, 3G and acceleration features were removed, the model performance increased from 61.11% to 66.11% accuracy rate; second, we also removed some features that are highly correlated to other remaining features. This additional process increased the accuracy to 68.89% and the kappa to 0.675. Also, the remaining significant features needed to identify mobile users using keystroke dynamics and accelerometer biometrics are as follows: (1) the average time between the press of a key and the release of the same key; (2) the time between the press of the first key and the release of the last key; (3) the average and the standard deviation time between the press of a key and the press of the succeeding key; (4) the average and the standard deviation time between the press of a key and the release of the succeeding key; (5) the average and the standard deviation time between the release of a key and the press of the succeeding key; (6) the average and the standard deviation time between the release of a key and the release of the succeeding key; (7) the average and the standard deviation time between the press of a key, the press of the succeeding key, and the press of the third key; (8) the average and the standard deviation time between the press of a key, the press of the succeeding key, and the release of the third key; (9) the average and the standard deviation time between the release of a key, the press of the succeeding key, and the release of the third key; (10) the average and the standard deviation time between the release of a key, the press of the succeeding key, and the press of the third key; (11) the average value, maximum value, minimum value, standard deviation of x-axis acceleration; (12) the maximum value of y-axis acceleration; and, (13) the maximum value of z-axis acceleration.

## 2.5 Model Testing

### 2.5.1 *Testing using a pre-labeled Test Set*

The most acceptable model generated in Section 2.4 was tested using the pre-labeled test set that was defined during data processing. The test achieved an accuracy of 73.33%, a False Acceptance Rate (FAR) of 27.59% and a False Rejection Rate (FRR) of 26.67%, respectively.

### 2.5.2 *Real World Testing trough a Prototype*

In this section, we modified the data gathering tool we developed in Section 2.1 and included some of the processes needed for the prototype. The mobile application included the enrolment of the username and four (4) password specimens. The remaining processes (re-modeling and testing) were implemented in a server using a web service using Play Framework in Eclipse. After the raw data acquisition, the features were computed using the same formula in Section 2.3. The mobile application posts the data in a csv file in the web service. Then re-modeling/authentication is done at the server side using the Rapid Miner libraries.

In this test, we requested six (6) volunteer participants who have no password specimen in the database. Each participant was asked to input their password of choice four (4) times. Then, each user was asked to test his/her own password. We also asked them to enter the actual passwords of their co-participants. Results show that the accuracy of the model/prototype is 60% or a False Rejection Rate (FRR) of 40% while the False Acceptance Rate (FAR) is 7.0%. This low FAR indicated that the prototype is good at blocking illegal access of other's account.

# 3. CONCLUSION

It was observed in this study that integrating accelerometer biometrics with keystroke dynamic features improved the model performance from 49.44% to 61.11%. It is evident therefore that accelerometer biometrics is significant in biometric-based user authentication. In addition, the low FAR (7.0%) of the prototype indicates that the model is good at blocking illegal access of other's account. However, its recognition is just slightly above average (60% -70%), which is not conclusive that the model can accurately identify mobile users. Thus, the outcome suggests that the model/prototype can only be used as a supplement to the traditional static textual password user identification scheme by setting some thresholds.

# 4. FUTURE WORK

With the hope of producing a more accurate behavioral biometric-based authentication model, we will be adding other parameters in our future work. These include gyroscope biometrics, typing pressure, fingertip size, and other parameters that can be read by the available sensors in a mobile device.

# 5. REFERENCES

[1] Ruggiero, P. and Foote, J. 2011. Cyber threats to mobile phones. *United States Computer Emergency Readiness Team.*

[2] Boyles, J.L., Smith, A. and Madden, M. 2012. Privacy and data management on mobile devices. *Pew Internet & American Life Project, 4.*

[3] Maiorana, E., Campisi, P., González-Carballo, N. and Neri, A. 2011. Keystroke dynamics authentication for mobile phones. In *Proceedings of the 2011 ACM Symposium on Applied Computing*, (Mar. 2011), 21-26. DOI= http://doi.acm.org/10.1145/1982185.1982190 .