usenix
THE ADVANCED
COMPUTING SYSTEMS
ASSOCIATION

Authentication on the Go: Assessing the Effect of Movement on Mobile Device Keystroke Dynamics

Heather Crawford and Ebad Ahmadzadeh, *Florida Institute of Technology*

https://www.usenix.org/conference/soups2017/technical-sessions/presentation/crawford

This paper is included in the Proceedings of the
Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017).

July 12–14, 2017 • Santa Clara, CA, USA

ISBN 978-1-931971-39-3

Authentication on the Go: Assessing the Effect of Movement on Mobile Device Keystroke Dynamics

Heather Crawford
Harris Institute for Assured Information
Florida Institute of Technology
Melbourne, FL
hcrawford@fit.edu

Ebad Ahmadzadeh
School of Computing
Florida Institute of Technology
Melbourne, FL
mahmadzadehe2012@my.fit.edu

ABSTRACT

Transparent authentication based on behavioral biometrics has the potential to improve the usability of mobile authentication due to the lack of a possibly intrusive user interface. Keystroke dynamics, or typing behavior, is a potentially rich source of biometric information for those that type frequently, and thus has been studied widely as an authenticator on touch-based mobile devices. However, the typing-while-moving scenario that characterizes mobile device use may change keystroke-based patterns sufficiently that typing biometrics-based authentication may not be viable. This paper presents a user study on the effects of user movement while typing on the effectiveness of keystroke dynamics as an authenticator. Using the dynamic text-based keystroke timings of 36 study participants, we first show that naïvely measuring patterns without considering position (e.g., sitting, standing or walking while typing) results in generic patterns that are little better than chance. We show that first determining the user's position before classifying their typing behavior, our two-phased approach, inferred the user's position with an AUC of above 90%, and the user's typing pattern was classified with an AUC of over 93%. Our results show that user typing patterns are a viable secondary or continuous post-PIN authentication method, even when movement changes a user's typing pattern.

1. INTRODUCTION

Mobile devices have become full computing platforms. The data and services they provide have made protecting them of paramount importance. Most devices use a secret knowledge-based means to protect them, such as a password, PIN, or small sketch (e.g., Android pattern lock). These are appropriate measures for initially protecting the device, but they do not provide protection if the device owner does not use them, or if an attacker gains access to an unlocked device. Keystroke dynamics, or the way in which a person types, has been suggested as a possible means to improve authentication by allowing it to be both *continuous*, protecting the device even after the initial password has been entered, but also *transparent* in that the user need not be distracted from their main task in order to authenticate regularly [42]. This has the potential to not only provide a higher device security level by continuing to authenticate the user after initial password entry, but also improve usability by removing a potentially disruptive request for repeated authentication.

Many of the existing keystroke dynamics studies have relied on the user typing a fixed word or phrase, such as adding keystroke dynamics to password entry, a practice known as *password hardening* [35], but not on dynamic text that changes from sample to sample. Also, much effort has gone into selecting the "best" classifier or the "best" set of features, with only small changes in the apparent distinctive nature of either.

This paper presents a keystroke dynamics user study designed to determine whether user typing patterns change enough during movement that it can no longer be used as an authenticator. We found on initial analysis that typing patterns over three positions (sitting, standing and walking) were insufficiently distinct to be used as evidence for authentication. This poor result is due to the additional movement that classification algorithms must overcome while typing. We have developed a phased classification approach, seen in Figure 1, that takes advantage of such movement. Our phased approach begins with using gyroscope data gathered at each keypress to determine the user's position (sit, stand, or walk). Next, classification models are created for each of the three positions under study that are then used to classify new data. The work presented here is a feasibility study to determine whether the collected gyroscope data is suitable for determining user position. The main novelty in our work is showing that modeling user typing based on their position improves classification rates over building a single, position-independent model. Our results show an improvement in AUC from 66% to 97% when position is considered before classifying keystroke dynamics data. These results indicate that our phased approach has merit; future work includes simulating the classification model to determine its use in practice.

2. BACKGROUND AND RELATED WORK

2.1 Mobile Device Authentication

In addition to existing password- and PIN-based authentication methods, research has begun to emerge on alternative authentication methods that consider the mobile device's needs more closely; in particular interest in using graphical passwords as an authenticator has been demonstrated [38, 41]. However, these methods still provide an all-or-nothing

Symposium on Usable Privacy and Security (SOUPS) 2017, July 12–14, 2017, Santa Clara, California.

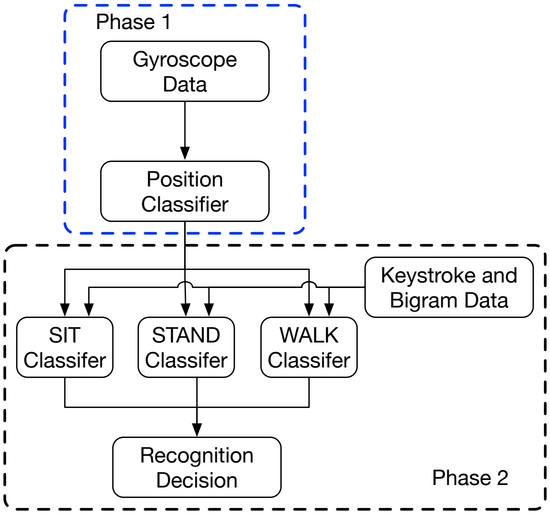


Figure 1: Our two-phased classification model

approach to device protection in that once the user is correctly authenticated, they are granted access to all data, services and apps on the device. In response, researchers have begun to study methods that continue to authenticate the user invisibly in the background while other tasks are completed. This is called *transparent, continuous* authentication. This type of authentication gathers behavioral biometric data such as keystroke dynamics [9, 32], touches [20, 42], etc. to continuously ensure that the device owner is the one currently using the device. Methods by which access to apps, data and services on the device can be restricted based on the identity of the current user have also begun to emerge [19].

2.2 Keystroke Dynamics

Keystroke dynamics is a behavioral biometric that uses patterns in how a person types to distinguish them from other users. It uses metrics such as *key hold time* (the amount of time between pressing and releasing the same key) and *inter-key latency* (the amount of time that passes between releasing one key and pressing the next) to identify these distinctive patterns. Researchers have examined other potential biometrics such as touch [11], facial recognition [15] and device movement [17]. Keystroke dynamics began with studies on desktop and laptop computers [30, 36] and in recent years has moved to mobile devices such as smartphones [13, 32]. Many keystroke dynamics studies attempt to replicate a password hardening situation in which the data gathered during the study is based on a known password that each study participant types a set number of times [28]. This practice can increase the strength of traditional passwords, but still provides an all-or-nothing approach to authentication. More recently, research has focused on providing transparent authentication that protects throughout device use rather than just at the beginning. It is this research that maps most directly to ours; thus we will focus on the current work in this area.

Typing patterns while moving have been studied by Clawson *et al.* in an effort to determine whether moving while typing affects accuracy and errors [16]. Their study had 36 participants type set phrases using a hard keyboard on a mobile device (Blackberry Curve 8320) while walking a set path. Their results showed that expert typists made fewer mistakes while walking, but at the cost of a lower typing speed [16]. Clawson *et al.*'s work is supportive of ours since more accurate typing may lead to improved uniqueness in typing patterns. However, Clawson *et al.* were not studying the use of keystroke dynamics as an authentication method, so further comparison of our results to this work are not indicated.

There has been much discussion on the amount and type of text used as input to transparent keystroke dynamics authentication tools. Many of the studies in this area, specifically those to do with password hardening, focus on text that must be repeated, while continuous, transparent authentication methods are likely to be based on any text the user may type. There is also the need for ecological validity – if a user can be expected to type any words and phrases, then basing a user study on specific words or phrases cannot be used to justify results in a more open environment.

2.2.1 *Fixed Text*

Fixed text methods (also called *static text*) assume that the user will type the same word or phrase at both enrollment and at the time of authentication. The text typed is generally short, as typing long texts at the time of authentication is tedious and error-prone on a mobile device [2, 12]. In general, using fixed text allows for more stability as the comparison between enrolled sample and gathered sample share the same keys and are thus similar. In some cases, experiments of this type produce results that either depend on special conditions (such as the attacker knowing the user's password) or have unacceptable accuracy levels [23, 7]. Much research has been done on fixed text methods [14, 28]; a summary of work in this area may be found in [18].

2.2.2 *Dynamic Text*

Also called *free text*, this paradigm assumes that the user may type whatever they wish, and that this input is any length. In reality, dynamic text and free text have several differences; free text is completely without constraints, where dynamic text may have aspects of both fixed and free text. Specifically, dynamic text may be prompted in some way, or may depend on a small number of specific words or phrases [23]. Several studies have examined dynamic text keystroke dynamics [4, 34], including Ahmed *et al.* [1] who report fairly good results, with False Accept Rates well below 1%. Free text keystroke dynamics has also been studied by Gunetti & Picardi [23], although their reported results are not as low as those of Ahmed *et al.* A summary of work on free text keystroke dynamics is available in [3]. The implication is that transparent authentication based on keystroke dynamics is best suited to *true* free text, which removes any restrictions about what or how much is typed. In this way, any characters a user may type can potentially be used as information upon which to base authentication decisions.

Adding realism to mobile device keystroke dynamics experiments has been studied from several points of view. One is that users may change their hand positioning while typing, which may affect their overall typing pattern. Azenkot & Zhai [6] studied user typing patterns when typing with one thumb, both thumbs and one index finger and found that there were pattern differences between these three hand positions. They used these results to suggest changes in keyboard design and layout that can improve typing accuracy. Similarly, Buschek *et al.* [10] studied the same three hand positions, but from the point of view of authentication rather than keyboard improvements. Their results showed that hand position had a strong effect on the ability to authenticate a user [10]. Both of these papers were based on password-hardening techniques, and thus were using fixed text techniques with defined feature vectors. Shen *et al.* studied the use of motion sensor data while typing a mobile device passcode as a potential authenticator [39]. They reported a False Reject Rate (FRR) of 6.85% and a False Accept Rate (FAR) of 5.01% in a user study with 48 participants [39]. Their work is similar to ours since they report results in the seated, standing and walking positions, although they only consider the sensor data when unlocking the mobile device with a passcode.

2.3 Gyroscope Data

Modern mobile devices come equipped with built-in sensors that can measure motion, orientation, environmental conditions such as temperature and humidity, etc. Data from sensors such as accelerometers and gyroscopes has been used for activity recognition [8, 25], to address typing inaccuracies [22], to create keyloggers [33], and to determine on-device input errors [37]. Accordingly, authentication research has begun to consider whether accelerometer and gyroscope data may be used as a unique identifier. Giuffrida *et al.* created what they call "sensor-enhanced" keystroke dynamics in their UNAGI system [21]. They experimented with the use of accelerometer and gyroscope data while 20 participants typed a set of fixed passwords and found that they were able to achieve Equal Error Rate (EER) values of less than 1% [21]. Their use of a fixed password as the stimulus indicates that theirs was a password hardening experiment rather than a dynamic text experiment.

2.4 Contribution

The major contribution of this paper is the determination that while typing patterns do change with user movement while typing, our phased classification model allows keystroke dynamics to be used as a viable secondary authentication method under realistic movement and text-acquisition conditions despite typing changes. To our knowledge, this is the first work to create different keystroke models for different user positions; a step that improves the accuracy of user classification. We also provide evidence that gyroscope data gathered at the time of each keypress is suitably distinctive to distinguish between sitting, standing and walking positions. This is significant because gyroscope data is often sampled essentially continuously, which generates a lot of data, uses significant battery power, and requires significant processing in order to be useful. Overall, our results show that keystroke dynamics can be used as secondary or continuous authentication method.

3. RESEARCH QUESTIONS

Our research questions are as follows:

1. Does gyroscope data captured at the time keypresses were made provide enough information to tell whether the typist is seated, standing or walking?
2. Does creating multiple position-specific models for a typist provide better classification results compared to using a single model trained on all positions?
3. Does a dynamic text-based system based on the above assumptions provide enough data for a sufficiently distinctive user model?

3.1 Hypotheses

Based on the above research questions, we present the following hypotheses for our work:

Hypothesis 1: A mobile device user's typing pattern is distinctive enough to use as a secondary or continuous authentication method as determined by achieving an AUC of at least 90%.

Hypothesis 2: Gyroscope data gathered as a key is pressed is distinctive enough to determine whether the typist is seated, standing or walking while typing, as determined by achieving an AUC of at least 90%.

Hypothesis 3: Determining a user's position and classifying based on data from that position only decreases False Accept and False Reject Rates (FAR and FRR, respectively) when compared to classification without determining user position.

We chose 90% as the AUC for Hypothesis 1 in order to justify using our method as a secondary or continuous authentication method, e.g., one that takes place as a supplement to or after primary authentication such as via a password or PIN. This means that near-perfect accuracy is not required, and the balance between FAR and FRR is not as vital as for a primary authentication method. While we may have chosen a lower AUC, we wish to produce a system that may be viable for primary authentication in the future. Therefore, 90% AUC is a value balanced between these two design choices. We chose AUC of 90% for Hypothesis 2 because high accuracy is not required since position determination is not an authentication decision (although it is related to one via our two-phased approach) and thus does not have the security ramifications that characterize authentication decisions.

4. THREAT MODEL

We assume that an attacker has access to the unlocked mobile device and may have had an opportunity to observe the device owner typing, and thus would know things such as current position, preferred hand position (e.g., index finger, one thumb, both thumbs), device orientation (e.g., landscape or portrait) and a general idea of typing speed. The attacker is assumed to have full knowledge of the biometric authentication system, including all inputs and outputs.

5. STUDY AND DATA COLLECTION

We collected gyroscope data and dynamic text typing data in a user study in a single session. The participant used a custom-built Android app to type phrases provided to them that varied both their position and the device orientation while typing. Specifically, participants were prompted by the experimenter to hold the device in a given orientation (portrait or landscape) and to type while either seated, standing or walking. The participants were told to type as they usually did; specifically, the speed of their typing was not restricted. We also did not provide specific guidance on how to sit, stand or walk. For instance, many participants chose to stand while leaning against a wall, or sit with their arms supported by a table. The only prompts we gave during the experiment were to keep walking if the participant stopped while typing in a walking condition. The study participants filled out a short demographic questionnaire before beginning any typing, and they were allowed to rest between conditions if they wished. Each participant was given the opportunity to practice typing before beginning the first condition; this training data was discarded before analysis. This study was approved by our university's Institutional Review Board (IRB) prior to its start.

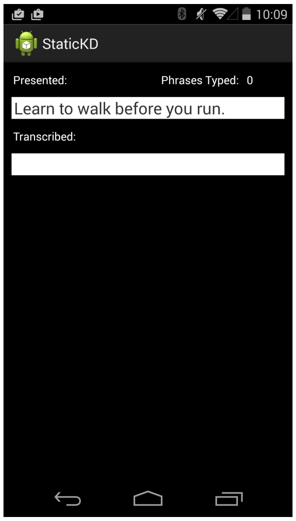
5.1 Participants

We recruited 39 participants (6 female, 33 male) through convenience sampling methods such as personal invitation, emails to mailing lists within our university and word of mouth. The data from three participants was removed from the study due to procedural errors, leaving data from 36 participants (5 female, 31 male). The remainder of this paper, including study results, refers to the analysis of data from the remaining 36 participants. The average age of participants was 28.3 years (SD = 11.3 years). Participants were not required to have any experience with typing on smartphones, although all participants reported that they owned and used a mobile device, most with soft keyboards. 2 participants were left-handed, and 34 were right-handed. Participant experience on their own mobile device varied: 14 participants used an Android-based device, 18 used an iOS device, 2 used another smartphone, and 2 used a feature (non-smart) phone. 2 participants were considered novices (used their device once a week or less), 3 participants as average (used their device more than once a week but not everyday) and 31 participants as experts (used their device every day or several times each day). Most participants were students, faculty or staff at our university; all participants had at least some post-secondary education, ranging from some undergraduate experience to graduate levels. Participants were not compensated for their participation.

5.2 Apparatus

We provided each participant with an LG Nexus 5 smartphone for the duration of the experiment. Each device ran Android version 4.4.4 and contained only the standard Android applications. Text entry was facilitated by the use of two bespoke Android applications. The first (see Figure 2) displayed the phrase to be typed (non-editable), a text box where the user typed the same phrase, and a counter that displayed the number of phrases the participant had typed in the current experimental condition. This app randomly chose a phrase from a modified version of the standard phrase set provided by MacKenzie and Soukoreff [31] (forthwith called the M&S set); duplicate phrases were permitted.

The second Android app used in this study (see Figure 3) was a custom-designed keyboard. It was designed to visually mimic the standard Android keyboard in order to accurately emulate a standard typing environment; the same keyboard design was used by all participants. This app was responsible for gathering the required keystroke and bigram metrics. When the participant pressed a key, the app recorded the key pressed, key hold time, inter-key latency, device orientation, user position and instantaneous gyroscope data (pitch, azimuth and roll). Key hold time is defined as the amount of time that a participant holds down a given key. Inter-key latency is defined as the amount of time between a key up event and the subsequent key down event.



StaticKD

Presented: Phrases Typed: 0

Learn to walk before you run.

Transcribed:

Figure 2: Phrase generation app screenshot

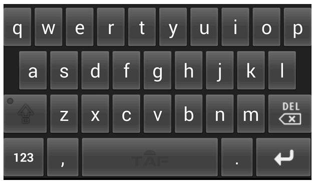


Figure 3: Custom keyboard for metric gathering

Timing of such typing events is a subject of debate in the keystroke dynamics field [27] as incorrect timing accesses can affect the measured typing pattern of a participant, which in turn has an effect on the reported study results. We mitigated such potential sources of error by using a set of four devices of the same model with the same operating system build, all of which had been reset to factory settings before the experiment began. In addition, we used the same Android applications on each device, and removed the previous participant's gathered data and restarted the application between participants. By using these precautions, we have made all possible efforts to minimize the effects of clock discrepancies on the results of this study.

Our keyboard, which runs as a service on the Nexus 5 devices, replaced the default keyboard in the settings of each device. This design means that when the user tapped on a widget that can accept keystrokes, our keyboard was displayed and subsequently used by the study participants. This allowed for gathering keystrokes from all applications that required typing, meaning that this same keyboard could be used in future work on transparent keystroke dynamics-based authentication.

5.3 Study Procedure

Each participant first answered a short demographic questionnaire, then was introduced to the app and bespoke soft keyboard they would use for phrase entry. They were given a choice to practice with the standard Android keyboard if they were unfamiliar with it. Most declined as they felt they had enough experience typing on the standard soft keyboard. The participants were allowed to take short breaks after each experimental condition. The participants were instructed to type in their usual manner, and that speed or accuracy were not being measured. They were told that auto-correction and auto-capitalization were disabled, and that if they made mistakes it was their decision whether or not to correct them. The participants were told to not change the device orientation and to remain in the participant mode (sit, stand, walk) they were placed into by the experimenter. They were not told how fast to walk, nor whether they should (or should not) support the device while typing (i.e., leaning against a wall, or with arms supported on a tabletop while seated). No further specifications were given to participants.

Each participant was placed into each of six experimental conditions (see Section 5.5 for a of the conditions) by the experimenter and asked to return to the experimenter when they had typed at least 22 phrases (there was a counter at the top of the custom phrase app for this purpose). The number of phrases was chosen in order to gather enough data for analysis, and to provide a similar amount of data for each participant. After completing each condition, the participant was asked to return to the experimenter, who would place the device into the next condition and instruct the participant as to their mode and the device orientation (i.e., "Please type the next set of phrases with the device in portrait while you're seated"). Once the participant had completed each of the six conditions, they were thanked for their time and allowed to leave.

5.4 Phrase Sets

The stimulus item in this experiment – the prompt that encouraged the user to type – was a randomly selected phrase from a modified set of standard phrases (the M&S set) [31]. Much debate has ensued over the choice of phrases used in text entry experiments. The main issues are that having a non-standard phrase set may impact the results of the study in that using a different phrase sets may result in different experimental results [29]. MacKenzie & Soukoreff addressed the issues of experimental validity (both internal and external) [31], and provided a set of 500 phrases that have been used in various studies. Kristensson & Vertanen opine that the phrase set chosen has an effect on study reproducibility in addition to internal and external validity; it is nearly impossible to reproduce an experiment if the actual phrase set is unknown (i.e., taken from random selections from an unspecified source, such as collecting phrases from "the news") [29].

In choosing a phrase set for this experiment, we kept in mind both internal and external validity as well as study reproducibility, while ensuring our phrase set met the requirements of our experiment. Specifically, we required a phrase set that closely matched letter frequencies in English, was large enough to ensure repeated phrases for the same participant were minimized, and contained upper- and lowercase letters, punctuation and numbers. The M&S set met the first two requirements; but not the last one. To remedy this, we edited the M&S set to have upper-case letters at the start of each phrase, changed text numbers to numeric equivalents (i.e., "eight" was changed to "8"), and added punctuation such as ending periods, exclamation points and question marks, as well as commas where grammatically correct. We believe that doing so created a phrase set that was both ecologically valid and made for a repeatable experiment.

Typically, a true free text typing experiment would require the participant to type whatever came to mind. One issue that arises, though, is what to do when the user cannot think of anything to type since this will affect their standard typing pattern. The role of the phrases in our study were to keep the user typing in as natural a way as possible. Otherwise, the content of the phrases did not have an impact on the accuracy, difficulty, or usability of the typing task. By using phrases rather than free text, however, we have change the user's task from one of creation to one of transcription, which may have an impact on their typing pattern. The advantage we gain is that the typing data is captured with a higher degree of freedom and fewer restrictions than with comparable fixed text experiments, which is arguably more similar to real-world typing situations.

5.5 Experiment Design and Analysis

Our laboratory-based study used a within-subjects, repeated measures design, in which the study participants were assigned to one of six experimental groups that differed only in the order in which the participant completed each of the six study conditions (see Table 1). All participants completed all study conditions. Participants were assigned to each study condition using a $6x6$ latin squares design in order to minimize learning effects and fatigue. Each session lasted about one hour.

Study Conditions		
Position	**Device Orientation**	**Description**
Sit	Portrait	**Sitting** in a fixed chair (no casters); Arms optionally **supported** on table;
Sit	Landscape	
Stand	Portrait	Standing, device **unsupported**; Optionally **leaning** against a wall;
Stand	Landscape	
Walk	Portrait	**Walking** around a large space, some obstacles; No set speed; most users walked slowly
Walk	Landscape	

Table 1: Description of study conditions

5.6 Data Gathered

The collected keystroke data was sanitized by removing the %, & and $ characters because the experimenter long pressed these keys to indicate a transition between the six study conditions. We used these keys as indicators of a change in device mode between sit, stand and walk. We chose these three keys for this purpose because they did not appear in any of the 500 phrases used as stimulus items, and thus could safely be removed from the dataset without removing valuable user data.

For bigram data, we collected the two characters that make up each bigram (not used during data analysis), the calculated key hold time for that bigram, the device orientation (portrait or landscape) and the participant's position (sit, stand, or walk). We sanitized the data to once again remove the occurrences of the %, & and $ characters. In the case of bigrams, we removed the entire bigram from the dataset if any of these three characters appeared as either of the two letters saved. For both keystrokes and bigrams, values greater than 3 SD beyond the mean were considered outliers and removed from the dataset prior to classification.

The gyroscope data was sanitized to remove the occurrences of %, & and $ but was unchanged otherwise. Since we gathered the gyroscope azimuth, pitch and roll in an instantaneous (i.e., at the moment of a keypress) rather than continuous manner, it was not necessary to window the data into discrete sections, nor to filter the data to remove high- or low-level frequencies as is common in activity recognition studies. Furthermore, since our gyroscope data is not time-scale data since it is discrete rather than continuous measurements, it was not necessary to transform it to the frequency domain before analysis.

We collected a total of 323,064 keystrokes and 289,520 bigrams from all 36 participants, not including practice phrases. The average number of keystrokes gathered per user was 8,974, and the average number of bigrams was 8,042. Since we gathered gyroscope data on each keystroke, we gathered the same amount of gyroscope data as keystrokes with one exception: we did not record instances of using backspace in the keystroke data, but we retained this information for gyroscope data since we were interested in the device movement on each keypress rather than whether that movement was related to a particular key.

5.7 Feature Vectors

The feature vector for the gyroscope data was simply the x, y, and z coordinates as gathered during each keypress. The makeup of a feature vector for keystrokes and bigrams, however, is much more complex.

In fixed text keystroke dynamics studies, the feature vector used is quite clear – it is the concatenation of the n key hold times corresponding to the keys pressed when typing the password (often with the ending enter keystroke) and the $n-1$ inter-key latencies for the associated bigrams. Since all participants type the same password during a study, the feature vectors are the same for each pattern gathered from each participant; the data is complete without missing values. When used outside an experimental setting, the only comparison is between a person's enrolled keystroke metrics when typing their password, and the subsequent keystroke metrics when typing the same password at a later date.

Keystroke biometrics based on dynamic text are more useful when the goal is to gather keystroke information unobtrusively, such as when continuous, transparent authentication is used to verify the identity of a person after initial login. In this situation, we specifically do not want to interrupt the user in what they are doing in order to retype their password, so we instead gather their keystroke metrics as they type as part of their regular device use. We may gather data from them when typing an email, a paper, or a blog post, all of which will have few phrases that appear in all. We gather this data from a custom extension of Android's standard keyboard so that key hold time and inter-key latency, which depend on the keyboard size and key placement, are collected in the same manner for all study participants.

Since dynamic keystroke biometrics cannot depend on getting a fixed amount of text from each participant, nor guarantee that all participants will type the same values, deciding upon the components of the feature vector is a complex task. Intuitively, selecting the most frequently typed characters and bigrams suggests that the most data possible will be retrieved from each participant. However, in practice the most frequently typed characters may vary from person to person. If the most frequently typed English letters are chosen, there might be gaps in our gathered patterns if the participant did not type that letter. This situation gets far worse when considering the frequency of bigrams. These gaps create a much more sparse dataset upon which to base authentication decisions, so far more data must be gathered to be as predictive and suitable as fixed text keystroke dynamics. To counter this, we used a dynamic feature space where the bigrams and keystrokes involved are those for which we have at least a few instances from the user. For example, early in the data collection process, the classifier may start with a minimum number of bigrams and keystrokes that have been typed thus far (we set this at 4 of each). The feature space then grows as more data is collected. For short text in which all features do not appear, we stochastically estimate the missing values from the user's past typing data.

6. RESULTS AND DISCUSSION

We now present our study results and related them back to the hypotheses defined in Section 3.1.

6.1 Position Independent Results

We begin by reporting the results of the naïve method, in which we do not use the gyroscope data to first determine user position. In this case, we mixed data from the sit, stand and walk positions and classified only based on the key hold times and inter-key latencies for two classification algorithms: Decision Tree and Logistic Regression. We chose Decision Tree because of its use in human activity recognition studies [5, 24] and because it is quick to train and classify data. Logistic regression was chosen for its simplicity and ease in understanding feature significance and removing those found to be insignificant. Furthermore, like Decision Tree, logistic regression has a low computation load for training and classifying data, which is an important feature on the constrained memory, power and processing environment on mobile devices. We considered each participant in turn the device owner (their data was considered the positive class), and the other participants as non-owner (their data was considered the negative class). The owner's model was trained on 2/3 of their supplied key hold times (keystrokes) or inter-key latencies (bigrams) plus an equal amount of data randomly selected from the other study participant's data. We used 10-fold cross-validation and report the averages from the 10 folds in Table 2. We have reported False Accept Rate (FAR), False Reject Rate (FRR), and the Area Under the Curve (AUC) for the Receiver Operating Characteristic (ROC) curve. We chose to report AUC because it provides, in a single value, the ability of our classifier to distinguish between owner typing patterns and those of others. An AUC value equal to 50% represents a method that is no better than chance; an AUC value equal to 100% is indicative of a perfect classifier.

As can be seen in Table 2, the FAR and FRR are very high for both keystrokes and bigrams. For instance, the FAR value of 41.9% for keystroke results using DT indicates that there is a 41.9% probability that an attacker will gain access. This is unacceptably high for any authentication system since it means that nearly half of all attackers will gain access to the mobile device. Similarly, the FRR of 23% for keystrokes using DT represents a nearly one in four likelihood that a legitimate user will be forced to reauthenticate. While reauthentication is less risky in terms of security, it represents an annoyance to users and a reduction in system usability since a legitimate user will have to reauthenticate once out of every four attempts.

The AUC values in Table 2 are not much better. Values in the 60-69% range represent a classifier that is only 10-19% better than chance, which is not acceptable even for secondary authentication. Overall, these results indicate that a person's typing pattern changes sufficiently over the three studied positions (sit, stand, walk) that much of the uniqueness in those typing patterns is lost.

Due to these uninspiring results, we chose not to combine key hold time and inter-key latency features as a way of improving classification rates in favor of a potentially better solution: our dual-phased classification model, which is based on first determining the user's position, then classifying using a model built using only user data from that position.

	Metric	Classifier Metric (%)		
DT		**FAR**	**FRR**	**AUC**
	Keystrokes	41.9	23.0	66.9
	Bigrams	49.3	30.5	60.3
LR		**FAR**	**FRR**	**AUC**
	Keystrokes	39.0	35.6	66.2
	Bigrams	43.3	41.7	60.7

Table 2: FAR, FRR and AUC (%) averaged over all participants for keystroke data (key hold time) and bigram data (inter-key latency) using Decision Tree (DT) and Logistic Regression (LR) classifiers. These results do not consider user position (e.g., sit, stand or walk) and are used as a baseline for comparison purposes.

6.2 Position Dependent Results

The first phase of our two-phased approach is to determine the user's position while they are typing, then classify their typing into owner or not owner based on a model trained only on data from that position. To determine position, we gathered gyroscope data from the mobile device at the moment each key was pressed. Our intuition is that the gyroscopic movement (as measured by the device's pitch, azimuth and roll) will be different when typing while seated, standing or walking. We chose not to measure accelerometer data since it is likely that the accelerometer readings will be different for the walking condition and relatively similar for seated and standing, thus making the latter two positions difficult to distinguish.

6.2.1 *Gyroscope Data*

In order to address Hypothesis 2 regarding the ability of gyroscope data gathered at each keypress to distinguish between the three user positions of sit, stand and walk, we analyzed this data using two classifiers: C4.5 Decision Tree (DT) and Logistic Regression (LR). We used the Weka implementation of these classifiers [26], which were chosen because of their use in activity recognition and keystroke dynamics work, respectively. We used 10-fold cross validation as with the previous classifications.

	Pos.	Classifier Metric (%)		
DT		**FAR**	**FRR**	**AUC**
	Sit	4.5	10.5	97.3
	Stand	10.3	20.2	91.5
	Walk	9.2	23.3	92.2
LR		**FAR**	**FRR**	**AUC**
	Sit	10.8	18.6	90.8
	Stand	15.8	39.8	82.3
	Walk	17.7	31.1	84.5

Table 3: Gyroscope data FAR, FRR and AUC (%) results averaged over all participants for Decision Tree (DT) and Logistic Regression (LR) classifiers.

As can be seen in Figure 3 our results were promising for both DT AND LR, although slightly better for DT. AUC is a valuable measure of classifier accuracy for binary classification problems; Table 3 reports the AUC for the position in question considered the positive class, and the other two positions considered the negative class. For example, the AUC of 97.3% for the Sit position for DT is measured based on using Sit as the positive class and Stand and Walk together as the negative class. In general, values of greater than 90% for DT indicate that the gyroscope data gathered is very good at distinguishing between the three user positions. Note that the AUC values for both classifiers for the Sit position are higher than those values for Stand and Walk. We believe this is because users tended to prop their arms on a table while typing during the study, which may mean that the mobile device moved less (or at least differently) compared to the unsupported arm positions while in the Stand and Walk conditions. These promising results show support for accepting Hypothesis 2.

The FRR values in particular, though, are a bit worrisome as they are high for both classifiers. However, these results are not being used to determine authentication suitability, but only to justify using gyroscope data to determine user position. Thus, there is little security risk associated with misclassifying the user's position; such a misclassification simply means the wrong model may be used for classifying keystroke and bigram data. The selection of the wrong model may result in rejecting the legitimate user, which would require reauthentication and thus could affect usability. We intend to explore the impact of such misclassifications in future work.

6.2.2 *Keystrokes*

Once the user's position has been determined, key hold time and inter-key latency data from the user's typing patterns will be classified as owner or not-owner based on three trained models based on data from the three user positions of Sit, Stand and Walk. This section discusses the results of a feasibility study in which the study participants' keystroke and bigram data was classified using position-based models with the DT and LR classifiers to allow for easy comparison to the naïve results shown in Table 2.

Table 4 shows the FAR, FRR and AUC metrics that result from classifying key hold times over the three user positions. The results for DT for all three metrics are better than those for LR; FAR values for LR in the 18.7% to 20.42% range indicate an unacceptably high one in five chance that an attacker will be mistaken for the legitimate device owner. Furthermore, FRR values of about 23% for LR show a usability problem since nearly one in four authentication attempts by the legitimate owner will fail. Since keystroke dynamics is best used as secondary or continuous authentication method, such a high failure rate is not as great a problem as for primary authentication methods. However, it is still an unacceptably high reauthentication rate. Therefore, we intend to use DT as the classifier of choice in future work.

	Position	Classifier Metric (%)		
DT		**FAR**	**FRR**	**AUC**
	Sit	8.5	8.4	90.3
	Stand	8.3	9.3	89.8
	Walk	7.4	8.3	91.0
LR		**FAR**	**FRR**	**AUC**
	Sit	18.70	23.18	82.76
	Stand	19.63	23.73	82.32
	Walk	20.42	23.55	82.14

Table 4: Keystroke data (key hold time) FAR, FRR and AUC (%) results averaged over all participants for Decision Tree (DT) and Logistic Regression (LR) classifiers.

6.2.3 *Bigrams*

Previous studies have shown that bigrams on mobile devices are not distinctive as authenticators on mobile devices [40]. However, our results refute this result, perhaps due to the use of position as an initial classification. Table 5 shows that bigrams are, in fact, a quite accurate means of authentication. The table shows the results of classifying Sit, Stand and Walk data as separate classification problems; for example, the Sit row for each classifier shows the results of classifying only Sit data into Owner and Not Owner classes; similarly for the Stand and Walk rows.

Table 5 shows that the DT classifier outperforms the LR classifier for FRR results, while remaining only slightly higher than LR for FAR values. The AUC values show that inter-key latency is perhaps even slightly more distinctive than key hold time since the bigram AUC values are slightly higher than those of keystrokes. Given that our intent is to use keystroke dynamics for secondary or continuous authentication, AUC values of 89.82% to 93.61% for DT over the three positions are highly encouraging. As with the keystroke data results, we intend to use the DT classifier in future work since the AUC values are comparable to LR, but the FRR values for DT are considerably lower, indicating less likelihood of reauthentication, thereby supporting improved usability.

6.2.4 *Keystrokes + Bigrams*

Due to the encouraging keystroke and bigram results after position classification, we combined the key hold time and inter-key latency features while still classifying only one position at a time. Table 6 shows the results of this classification; as expected, combining features showed an increase in AUC for both classifiers, although the increase is more notable for the LR classifier. Furthermore, the FAR and FRR values from LR classification are lower for the combined features when compared to those features alone. AUC results of around 97% over all positions for LR move keystroke dynamics into a range we consider suitable for primary authentication, although this must be validated via simulation to determine the impact of battery and processor use, which we leave for future work. Thus, we recommend that keystroke dynamics be used only for secondary or continuous authentication.

	Position	Classifier Metric (%)		
DT		**FAR**	**FRR**	**AUC**
	Sit	6.9	6.0	89.8
	Stand	6.6	6.9	93.6
	Walk	6.9	7.4	92.7
LR		**FAR**	**FRR**	**AUC**
	Sit	5.0	13.0	92.2
	Stand	4.3	12.7	93.1
	Walk	5.3	13.5	91.2

Table 5: Bigram data (inter-key latency) FAR, FRR and AUC (%) results averaged over all participants for Decision Tree (DT) and Logistic Regression (LR) classifiers.

	Position	Classifier Metric (%)		
DT		**FAR**	**FRR**	**AUC**
	Sit	5.6	6.1	93.2
	Stand	6.1	5.3	93.3
	Walk	4.8	5.6	93.9
LR		**FAR**	**FRR**	**AUC**
	Sit	1.7	7.0	97.3
	Stand	1.8	5.5	97.7
	Walk	1.4	6.2	97.7

Table 6: Combination of keystroke (key hold time) and bigram (inter-key latency) data FAR, FRR and AUC (%) results averaged over all participants.

The approximately 90% and up AUC values for DT over keystrokes, bigrams and their combination indicates that using keystroke dynamics as a distinctive information source for authentication is viable, and shows support for accepting Hypothesis 1 of this work.

6.2.5 *Comparison to Position Independent Results*

We now move to comparing the naïve, position independent keystroke and bigram results shown in Table 2 to the relevant data in Tables 4 and 5. The highest AUC for position independent results (Table 2) is 66.9% for key hold time data, and 60.7% for inter-key latency data, while the highest AUC values when position is considered are 91.01% for key hold time and 93.61% for inter-key latency. These increases are considerable, and show that considering position before authentication classification is a plausible approach to using keystroke dynamics as a secondary or continuous authentication method. This result is supported by the overall reduction in FAR and FRR values: from lows of 41.5% (FAR) and 23% (FRR) without considering position, to lows of 4.25% (FAR) and 6% (FRR) when position is considered. These reductions indicate that the two-phased approach is better able to minimize both attacker access and reauthentications compared to not considering position. These results show strong support for Hypothesis 3 regarding improvements in classification results when considering device position.

6.2.6 Implications

Our threat model outlined in Section 4 described possible attacks that can affect the system described in this paper. In particular, we stated that it is possible that the attacker may observe the device owner typing, and thus may be able to gather information that would allow the attacker to imitate the legitimate device owner. Given that the position-independent results showed us that a user's typing patterns are variable across positions, an attacker would have to learn different typing styles across all positions, which we consider unlikely.

The other implication to consider is what might happen if the first phase of the model (determining position) is incorrect. The effect would be that the wrong model would be used for matching the gathered keystroke information, which may result in rejecting a legitimate user. Given that our method is intended to be used for transparent authentication, there are two possibilities: either that multiple rejected authentication attempts are required to completely block access to the user, which enhances usability since additional user action is not required, but also has serious security ramifications since it increases the possible attack window. The second option is to prompt the device owner to enter a password or PIN when transparent methods are rejected, which has usability implications due to requiring additional user effort, but reduces the possible attack window. The preference for one of these options over the other depends on what type of system it is implemented in; a high-security system may require the latter.

7. LIMITATIONS

As with other user studies, ours has several limitations that must be considered in light of the results provided. Users often walked very slowly during the walking conditions; their focus was on their perceived goal (to enter the phrases) rather than on actually walking. It is likely that in a real-world situation, the user will be intent on walking rather than typing (i.e., if they are running late). Similarly, we observed users propping their arms on a table while typing during the Sit condition, and leaning against a wall during the Stand condition. It is possible that these postures introduced bias in that the static positional data may be more static, thereby further distinguishing this data from that gathered in the Walking condition. This may have resulted in better FAR, FRR and AUC values than in a real-world environment. The phrases themselves may have caused some bias in typing patterns (and removed some ecological validity) as the participant was transcribing the given phrases rather than creating true free text. Furthermore, our study required participants to use an unfamiliar mobile device with an unfamiliar keyboard, which may have had an effect on the participants' typing speed, as well as possibly changing how the keyboard reacts to touch events. We also disabled the predictive and corrective text actions, which affects ecological validity as these are widely used features on soft keyboards. We also did not consider hand postures during our study; participants were permitted to switch between typing with one thumb, both thumbs or any finger while in any of the six experimental conditions. We tested only a small set of classifiers (DT and LR) with few features. Many more possible classifiers exist, including those that take an anomaly detection approach, in which the classifier is trained only on the owner's data rather than adding in some representative negative samples. An anomaly detection approach is considered by some to be more valid for a single-user mobile device as it is unlikely that there will be a significant sample of other people's typing that can be used to create the negative class [10]. While other studies have achieved improved FAR and FRR values by using fused features in a multimodal biometric [10], we chose to use only inter-key latency and key hold time first to conform to other similar studies and also as a minimum baseline result to which future work can be compared. Finally, we collected data in a single session of only one hour in duration, which does not effectively study possible changes in a person's typing pattern over time.

Our final limitation is on the selection of Sit, Stand and Walk as user positions. We chose these based on our intuition that these are the most likely positions in which a user may type. It is unlikely, for instance, that a user will choose to (or be able to effectively) type while running, and positions such as laying down are very similar to both sitting and walking. We plan on addressing this issue with one of two approaches: either create an full activity recognition system that encompasses more positions, or narrowing the positions into those that are similar, such as ambulatory (e.g., walking, running) versus static (e.g., sitting, standing).

While each of these design decisions results in bias that will have differing effects on the results of this study, we believe that the largest effect will be in the overall classification rates, which in the worst case would be artificially high, which would give an inaccurate representation of the predictive power of gyroscope and keystroke data. We note that our results are similar to other studies in this field [39], and plan on removing some of these limitations (particularly those to do with the custom keyboard and disabling predictive and corrective text functions) in the simulation of our phased approach that we mention as future work.

8. CONCLUSION

In this paper we have presented the results of a user study designed to test the efficacy of keystroke dynamics as a potential continuous, transparent authenticator on mobile devices. We first determined via gyroscope data whether the typist was seated, standing or walking, then trained and tested three different models based on dynamic text from each of those three positions. We found that determining position first before classifying typing data resulted in an AUC increase of 30%. Our two-phased model approach of determining position first, then classifying keystroke information thus has merit and should be further examined via simulations. Both our experimental design and our threat model were chosen to provide as much ecological validity as possible given that the study was lab-based. We hope that taking a step back in assessing how much information is required per keystroke, and mimicking how users type in the wild, will provide an important advance in the field of keystroke dynamics.

Overall, our results support continuing research in keystroke dynamics as a transparent authenticator. We removed the need for a feature vector and the associated pre-processing required by them, while supporting a realistic evaluation scenario. We refuted previous results that showed bigram inter-key latency is not as distinctive as hoped for dynamic text, meaning that this feature may now be considered along with key hold time. We also provided support for the idea that transparent authentication may indeed be viable, which may help remove the need for a potentially intrusive and unusable authentication interface.

9. FUTURE WORK

We have begun creating a simulation of the phased approach pictured in Figure 1; we will use the simulation to test the effect of the phased approach on device battery and processor consumption, the amount of time needed for a classification decision, and the amount of data needed to reach suitable FAR, FRR and AUC values for continuous authentication. The use of a simulation as a first step will allow us to more closely model real-world typing conditions since our results were from a lab study. With the results of the simulation as a guide, we also plan on creating a prototype of this authentication method for Android devices, which we will test via a longitudinal user study. We will also use the simulation to innovate solutions to the sit-stand confusion, as well as to determine whether a catch-all classifier is suitable for situations where the user is neither sitting, standing nor walking while typing.

Acknowledgements

The authors wish to thank Thomas C. Eskridge, Philip Bernhard, and Abdullah Alharbi for valuable discussions that shaped this work.

10. REFERENCES

[1] A. A. E. Ahmed, I. Traore, and A. Almulhem. Digital Fingerprinting Based on Keystroke Dynamics. In *Proceedings of the Second International Symposium on Human Aspects of Information Security & Assurance (HAISA 2008)*, pages 94 – 104, July 2008.

[2] J. M. Allen, L. A. McFarlin, and T. Green. An In-Depth Look into the Text Entry User Experience on the iPhone. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 52(5), pages 508 – 512, 2008.

[3] A. Alsultan and K. Warwick. Keystroke Dynamics Authentication: A Survey of Free-text Methods. *International Journal of Computer Science Issues*, 10(4), 2013.

[4] A. Alsultan and K. Warwick. User-Friendly Free-Text Keystroke Dynamics Authentication for Practical Applications. In *IEEE International Conference on Systems, Man and Cybernetics*, pages 4658 – 4663, 2013.

[5] A. Anjum and M. U. Ilyas. Activity Recognition Using Smartphone Sensors. In *Proceedings of First Workshop on People Centric Sensing and Communications*, pages 914 – 919, 2013.

[6] S. Azenkot and S. Zhai. Touch Behavior with Different Postures on Soft Smartphone Keyboards. In *Proceedings of 14th International Conference on Human Computer Interaction with Mobile Devices and Services*, pages 251 – 260, 2012.

[7] F. Bergadano, D. Gunetti, and C. Picardi. User Authentication Through Keystroke Dynamics. *ACM Transactions on Information and System Security*, 5(4):367–397, November 2002.

[8] T. Brezmes, J.-L. Gorricho, and J. Cotrina. Activity Recognition from Accelerometer Data on a Mobile Phone. In *10th International Work-Conference on Artificial Neural Networks (IWANN)*, volume 5518 of *Lecture Notes in Computer Science*, pages 796 – 799, 2009.

[9] A. Buchoux and N. Clarke. Deployment of Keystroke Analysis on a Smartphone. In *Proceedings of the 6th Australian Information Security Management Conference*, pages 40 – 47, 2008.

[10] D. Buschek, A. D. Luca, and F. Alt. Improving Accuracy, Applicability and Usability of Keystroke Biometrics on Mobile Touchscreen Devices. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems (CHI)*, page to appear., 2015.

[11] T.-Y. Chang, C.-J. Tsai, and J.-H. Lin. A Graphical-Based Password Keystroke Dynamic Authentication System for Touch Screen Handheld Mobile Devices. *Journal of Systems and Software*, 85(5):1157 – 1165, May 2012.

[12] T. Chen, Y. Yesilada, and S. Harper. What Input Errors do you Experience? Typing and Pointing Errors of Mobile Web Users. *International Journal of Human-Computer Studies*, 68(3):121–182, 2010.

[13] N. Clarke and S. Furnell. Advanced User Authentication for Mobile Devices. *Computers & Security*, 26(2):109 – 119, March 2007.

[14] N. Clarke and S. Furnell. Authenticating Mobile Phone Users Using Keystroke Analysis. *International Journal of Information Security*, 6(1):1 – 14, January 2007.

[15] N. Clarke, S. Karatzouni, and S. Furnell. Transparent Facial Recognition for Mobile Devices. In *Proceedings of the 7th International Information Security Conference*, 2008.

[16] J. Clawson, T. Starner, D. Kohlsdorf, D. P. Quigley, and S. Gilliland. Texting While Walking: An Evaluation of Mini-QWERTY Text Input While On-the-Go. In *Proceedings of 16th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 339 – 348, 2014.

[17] M. Conti, I. Zachia-Zlatea, and B. Crispo. Mind How You Answer Me!: Transparently Authenticating the User of a Smartphone when Answering or Placing a Call. In *Proceedings of the 6th ACM Symposium on Information, Computer, and Communications Security*, pages 249 – 259, 2011.

[18] H. Crawford. Keystroke Dynamics: Characteristics and Opportunities. In *Proceedings of 8th Annual International Conference on Privacy, Security and Trust*, pages 205 – 212, 2010.

[19] H. Crawford, K. Renaud, and T. Storer. A Framework for Continuous, Transparent Mobile Device

Authentication. *Computers & Security*, Vol. 39, Part B:127 – 136, 2013.

[20] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song. Touchalytics: On the Applicability of Touchscreen Input as Behavioral Biometric for Continuous Authentication. In *IEEE Transactions on Information Forensics and Security*, volume 8, pages 136 – 148, 2012.

[21] C. Giuffrida, K. Majdanik, Mauro, and H. Bos. I Sensed It Was You: Authenticating Mobile Users with Sensor-Enhanced Keystroke Dynamics. In *Proceedings of Detection of Intrusions and Malware, and Vulnerability Assessment*, volume 8550 of *Lecture Notes in Computer Science*, pages 92 – 111, 2014.

[22] M. Goel, L. Findlater, and J. Wobbrock. WalkType: Using Accelerometer Data to Accomodate Situational Impairments in Mobile Touch Screen Text Entry. In *Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, pages 2687 – 2696, 2012.

[23] D. Gunetti and C. Picardi. Keystroke Analysis of Free Text. *ACM Transactions on Information and System Security*, 8(3):312 – 347, August 2005.

[24] Q. Guo, B. Liu, and C. W. Chen. A Two-Layer and Multi-Strategy Framework for Human Activity Recognition Using Smartphone. In *IEEE International Conference on Communications (ICC)*, 2016.

[25] P. Gupta and T. Dallas. Feature Selection and Activity Recognition System Using a Single Triaxial Accelerometer. *IEEE Transactions on Biomedical Engineering*, 61(6):1780 – 1786, 2014.

[26] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The WEKA Data Mining Software: An Update. *SIGKDD Explorations*, 11(1), 2009.

[27] K. S. Killourhy and R. A. Maxion. The Effect of Clock Resolution on Keystroke Dynamics. In R. Lippmann, E. Kirda, and A. Trachtenberg, editors, *Proceedings of RAID 2008*, volume 5230 of *Lecture Notes in Computer Science*, pages 331–350. Springer-Verlag Berlin Heidelberg, 2008.

[28] K. S. Killourhy and R. A. Maxion. Comparing Anomaly-Detection Algorithms for Keystroke Dynamics. In *Proceedings of the IEEE/IFIP International Conference on Dependable Systems & Networks*, pages 125 – 134, 2009.

[29] P. O. Kristensson and K. Vertanen. Performance Comparisons of Phrase Sets and Presentation Styles for Text Entry Evaluations. In *Proceedings of 12th ACM International Conference on Intelligent User Interfaces*, pages 29 – 32, 2013.

[30] J. Leggett, G. Williams, M. Usnick, and M. Longnecker. Dynamic Identity Verification via Keystroke Characteristics. *International Journal of Man-Machine Studies*, 35(6):859–870, December 1991.

[31] I. S. MacKenzie and R. W. Soukoreff. Phrase Sets for Evaluating Text Entry Techniques. In *Extended Abstracts on Human Factors in Computing Systems*, pages 754 – 755, 2003.

[32] E. Maiorana, P. Campisi, N. Gonzalez-Carballo, and A. Neri. Keystroke Dynamics Authentication for Mobile Phones. In *Proceedings of the 2011 Symposium on Applied Computing*, pages 21 – 26, 2011.

[33] P. Marquardt, A. Verma, H. Carter, and P. Traynor. (sp)iPhone: Decoding Vibrations from Nearby Keyboards Using Mobile Phone Accelerometers. In *Proceedings of ACM Conference on Computer and Communication Security*, pages 551 – 562, 2011.

[34] A. Messerman, T. Mustafic, S. A. Camtepe, and S. Albayrak. Continuous and Non-intrusive Identity Verification in Real-Time Environments Based on Free-Text Keystroke Dynamics. In *2011 International Joint Conference on Biometrics*, pages 1 – 8, 2011.

[35] F. Monrose, M. Reiter, and S. Wetzel. Password Hardening Based on Keystroke Dynamics. *International Journal of Information Security*, 1(2):69 – 83, February 2002.

[36] F. Monrose and A. D. Rubin. Keystroke Dynamics as a Biometric for Authentication. *Future Generation Computer Systems*, 16:351–359, 2000.

[37] M. F. M. Noor, S. Rogers, and J. Williamson. Detecting Swipe Errors on Touchscreens using Grip Modulation. In *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pages 1909 – 1920, 2016.

[38] Paul Dunphy and Andreas P. Heiner and N. Asokan. A Closer Look at Recognition-Based Graphical Passwords on Mobile Devices. In *Proceedings of the 6th Symposium on Usable Privacy and Security*, pages 26 – 38, 2010.

[39] C. Shen, T. Yu, S. Yuan, Y. Li, and X. Guan. Performance Analysis of Motion-Sensor Behavior for User Authentication on Smartphones. *Sensors*, 16(3), 2016.

[40] T. Sim and R. Janakiraman. Are Digraphs Good for Free-Text Keystroke Dynamics? In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1 – 6, 2007.

[41] X. Suo. A Study of Graphical Password for Mobile Devices. In *Proceedings of 5th International Conference on Mobile Computing, Applications and Services*, pages 202 – 214, 2014.

[42] H. Xu, Y. Zhou, and M. R. Lyu. Towards Continuous and Passive Authentication via Touch Biometrics: An Experimental Study on Smartphones. In *Proceedings of Symposium on Usable Privacy and Security*, 2014.