Challenge-Response Behavioral Mobile Authentication: A Comparative Study of Graphical Patterns and Cognitive Games

Manar Mohamed*
mohamem@miamioh.edu
Miami University

Prakash Shrestha
prakashs@uab.edu
University of Alabama at Birmingham

Nitesh Saxena
saxena@uab.edu
University of Alabama at Birmingham

ABSTRACT

The most researched behavioral biometrics for mobile device authentication involves the use of touch gestures as the user enters a graphical pattern password (like the one used on Android) or otherwise interacts with the device. However, due to the inherent *static* nature of these schemes, they are vulnerable to *impersonation attacks*. In this paper, we investigate *challenge-response* mechanisms to address this security vulnerability underlying the traditional static biometric schemes. We study the performance, security, and usability of two schemes of such challenge-response interactive biometric authentication geared for mobile devices and contrast them to static graphical pattern based biometrics. The first scheme is based on random graphical patterns. The second scheme, recently introduced for PC class of devices (not mobile), is based on a *simple cognitive game* involving semantic interactive random challenges. Our results show that the accuracy of user identification with these approaches is similar to static pattern based biometric scheme. Finally, we argue that utilizing interactivity and randomization significantly enhance the security against impersonation attacks. As an independent result, our work demonstrates that the use of motion sensors available on mobile device serves to improve the identification accuracy of schemes that only use touch-based gestures (static and interactive).

CCS CONCEPTS

• **Security and privacy → Authentication**.

KEYWORDS

Mobile Authentication, Behavioral Authentication, Graphical Patterns, Cognitive Games

ACM Reference Format:
Manar Mohamed*, Prakash Shrestha, and Nitesh Saxena. 2019. Challenge-Response Behavioral Mobile Authentication: A Comparative Study of Graphical Patterns and Cognitive Games . In *2019 Annual Computer Security Applications Conference (ACSAC '19), December 9–13, 2019, San Juan, PR, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3359789.3359838

1 INTRODUCTION

Traditional user authentication suffers from various well-documented usability and security issues. These issues are more severe on mobile devices due to their small screen size. In particular, passwords and graphical patterns (e.g., Android login patterns) are prone to dictionary guessing automated attacks, shoulder surfing attacks and smudge attacks [6, 31, 34, 38]. Other physical biometrics, such as fingerprint and face recognition are also susceptible to spoofing and impersonation attacks [20].

Behavioral biometrics have been studied extensively over the last decade aiming to solve the problems associated with traditional user authentication methods. Such biometric schemes can be used as a stand alone way for authenticating the user [14, 18, 19, 23] or to be added to another authentication scheme [15, 33] so as to provide a second layer of security. However, more research is still needed to enhance the security and usability of behavioral biometrics. This is because all of the proposed schemes suffer from high rejection rate of legitimate users (and hence low usability), high acceptance rate of other users (and hence low security against *zero-effort attacks*) and susceptible to impersonation attacks. Many impersonation attacks have been explored that compromise the security provided by behavioral biometrics. These include training humans to mimic victims [37], and building robots [32] and malware programs [28] to mimic a victim via external observation or internally through a compromised device, respectively. These attacks have been shown to bypass the security provided by the behavioral biometric schemes with up to 100% accuracy.

The above security vulnerabilities of existing behavioral biometric schemes stem from their inherent *static* nature. In this paper, aiming to enhance the security of these schemes against impersonation attacks, we investigate the use of interactive biometrics, taking the form of *challenge-response* authentication. The motivation of such challenge-response interactive behavioral biometrics is to prevent attacks that try to record the user interaction with the authentication construct and replay it later to impersonate the user as in the case of static biometric schemes.

Specifically, we study two schemes of such challenge-response behavioral biometric authentication on mobile devices. These schemes can be utilized to authenticate the user, e.g., to the device, to an app or to a remote web-site. The first challenge-response biometrics we study, called *CR-Pattern* (Figure 1b), is based on graphical patterns (akin to Android pattern based login). However, rather than asking the user to enter her static pattern, we display a random challenge (pattern) and ask the user to re-enter it. This pattern is not used as a password (unlike the case of Android) but rather to extract biometric features. The second interactive biometrics we

ACSAC '19, December 9–13, 2019, San Juan, PR, USA
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-7628-0/19/12…$15.00
https://doi.org/10.1145/3359789.3359838

[*] Work done at UAB

study, called *Gametrics* (Figure 1c), is based on simple semantic matching challenges. We present such challenges as simple drag and drop games, where the user has to drag randomly moving objects to their corresponding target objects, where the moving objects and the target objects are semantically related. Such type of challenges allows us to identify the user based on her unique cognitive abilities, e.g., the time taken to complete the challenge, in addition to other features captured by the touch behavior. Biometrics based on semantic challenges has been previously explored in [27] and shown to be effective on the *PC class of devices*. In this work, we design and evaluate its effectiveness in the context of *mobile devices*, not studied before.

As a baseline for our study, we further study a method called *S-Pattern* biometrics (proposed in [15] – Figure 1a), in which we try to authenticate the users based on the way they enter the *static* graphical pattern on mobile devices. This scheme is representative of static behavioral biometrics, but is vulnerable to multiple attacks in the literature that mainly record the user interaction with authentication construct or learn the user biometrics from her template and then try to reproduce the user biometrics automatically, e.g., malware or robot, or by expert human attacker [28, 32, 37]. In our study, we compare *CR-Pattern* and *Gametrics* with *S-Pattern* and study the feasibility of the former two approaches in enhancing the security of *S-Pattern*.

Our Contributions: In this paper, we study challenge-response behavioral mobile authentication. The main contributions of this paper are summarized below:

(1) *Design and Implementation of Three Behavioral Biometrics Schemes:* As part of our study, we design and implement three Android applications that we utilized to record the user interactions corresponding to the three methods, *S-Pattern*, *CR-Pattern* and *Gametrics* captured by touch, motion and position sensors (e.g., accelerometer and gyroscope). Then, we extract several features from each challenge solving instance and apply machine learning techniques to identify the users.

(2) *Evaluation of the Three Schemes under Benign Settings and Zero-Effort Attacks:* We collected data from multiple users in a lab setting and we show that we can identify the legitimate users and the zero-effort attackers with high accuracy using all schemes (F-measure of up to 89% for *S-Pattern*, 86% for *CR-Pattern* and 83% for *Gametrics*). As an independent result, we further show that utilizing the motion and the position sensors improve the classification accuracy of the three studied biometric schemes (F-measure of up to 100% for the three schemes). Prior schemes [15, 27] have not studied the use of such sensors. We also assess the usability of the three schemes and find that all three offer an acceptable level of user experience.

(3) *Evaluation of the Three Schemes under Active Attacks:* We show that challenge-response schemes are more resilient to active impersonation attacks compared to existing static biometric schemes. Moreover, we argue that the multiple round of interactions and the semantic challenge embedded in *Gametrics* enhance the security of *Gametrics* compared to *CR-Pattern*.

Results Summary: The paper shows that the challenge-response schemes offer similar level of usability to traditional pattern unlock measured by the user experience survey and provide higher level of security against various types of impersonation attacks. On the negative side, the challenge-response schemes require longer time for authentication. This suggests that selection of the biometric scheme can be dependent on the application. For example, for the applications in which login speed is vital and security demands are not that high (e.g., in phone locking application since this already requires the attacker to have physical possession of the phone), *S-Pattern* is better-suited. On the other hand, for the applications with larger time budget and higher security demands, like web authentication, banking apps, *CR-Pattern* or *Gametrics* can be used. The multiple levels of interaction in *Gametrics* provide extra level of security compared to *CR-Pattern* and therefore could be chosen for high-security scenarios such as financial applications. The results of our study show that utilizing the motion and position sensors on mobile devices enhances the classification accuracy both in reducing the rate of rejecting legitimate user and reducing the rate of accepting zero-effort attacker, shoulder-surfing attacker and other forms of impersonation attackers. In this light, we recommend the sensor recordings be always included as classification features.

Paper Outline: The rest of this paper is organized as follows. In Section 2, we lay out the evaluation criteria and the threat model. In Section 3, we describe the design and implementation of the three behavioral biometric schemes. In Section 4, we elaborate on our data collection methodology and procedures. Then in Section 5, we describe our feature extraction methods and our classification models. In Section 6, we provide the classification results in benign setting and against zero-effort attackers. Then in Section 7, we present the usability of the three studied-schemes in terms of completion time and user experience. In Section 8, we evaluate the three biometric schemes against active attacks. In Section 9, we discuss further aspects of our work and provide future research directions. In Section 10, we provide an overview of prior behavioral biometric systems. Finally, in Section 11, we conclude our work.

2 EVALUATION CRITERIA AND THREAT MODEL

The goal of any behavioral authentication scheme is to authenticate the user efficiently with high accuracy while preventing different kinds of impersonation attacks as much as possible. To this end, in our study, we set out to analyze the three behavioral authentication schemes (i.e., *S-Pattern*, *CR-Pattern*, and *Gametrics*) and compare them in terms of authentication accuracy and susceptibility to the impersonation attacks. In particular, we evaluate the three authentication schemes with respect to the criteria described below.

(1) **Usability**
   (a) Accuracy of user identification. The authentication system should identify the legitimate user with high accuracy and with minimal false alarm.
   (b) User experience and perception. The authentication system should have minimal user-effort in its authentication process.
   (c) Time taken to identify the user. Time taken by the authentication system to identify the user should be reasonably low.

(2) **Security** against the following types of deliberate impersonation attacks.

Table 1: Sensors utilized for our study.

Sensor Name	Sensor Type	Description
Accelerometer	Motion	The acceleration force including gravity
Gyroscope	Motion	The rate of rotation
Linear Acceleration	Motion	The acceleration force excluding gravity
Rotation Vector	Motion	The orientation of a device
Gravity	Motion	The gravity force on the device
Game Rotation	Position	Uncalibrated rotation vector
Magnetic Field	Position	The ambient magnetic field
Orientation	Position	The device orientation

(a) *Smudge Attacks*: An external attacker should not be able to learn the authentication token based on the screen smudges and use this knowledge to mimic/impersonate the user [6].

(b) *Shoulder-Surfing Attacks*: An external attacker who monitors the user while she is authenticating herself to the system should not be able to mimic and impersonate the user at a later point of time.

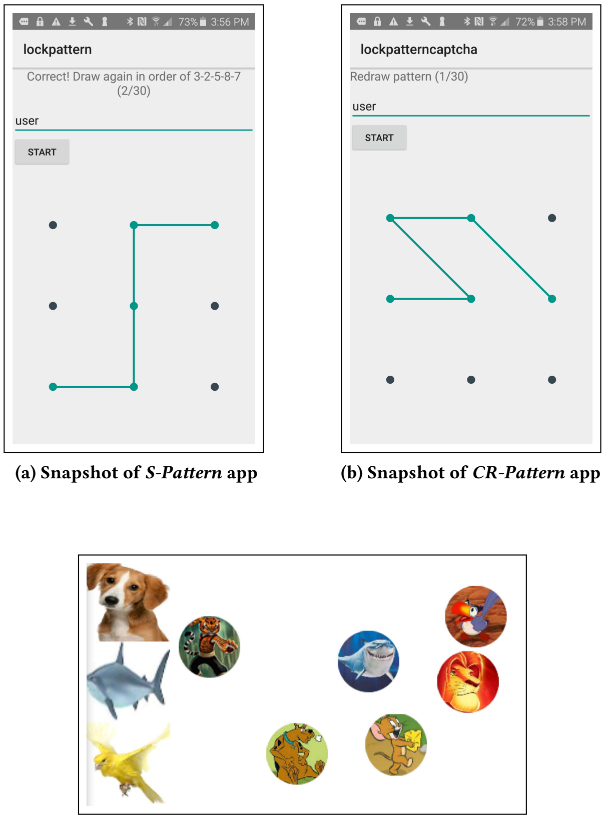
(c) *Automated Attacks*: The attacker who steals a user's authentication template (e.g., by hacking into the device or the server that stores this template) should not be able to authenticate itself to the system in an automated manner [32].

(d) *Internal Attacks*: A malware residing on the authentication device itself may have the ability to record the user's valid authentication token or template and replay it later to authenticate itself on behalf of victim user. Further, it may learn the authentication token/template by recording multiple valid authentication tokens and create the template by itself. Such attackers should not be able to fool the authentication scheme. Other forms of behavioral biometrics schemes have been shown to be vulnerable to such internal attacks [28].

3 METHODS DESIGN & IMPLEMENTATION

In order to evaluate the effectiveness of three behavioral biometric schemes considered in this study, we designed and implemented three Android apps.

- ***S-Pattern* App:** This is a simple Android app that mimics the traditional graphical pattern-lock in Android devices. The app contains instruction text, text box, a start button and a grid of nine dots (a snapshot of the app is shown in Figure 1a). The instruction text displays the instruction for the user during the study. The user provides her name in the text box and hit the start button to start the app. The user is then instructed to input a fixed pattern (3-2-5-8-7)[1]. Once the user has supplied the pattern, the app validates its correctness. On correct pattern entry, the counter gets increased by one, the pattern changes its color to green and then disappears at the end. On incorrect pattern entry, the pattern changes its color to red and informs the user that she has provided the incorrect pattern. During the pattern entry process, the app records the interaction of the user with the device that is captured by touch screen sensor as well as motion and position sensors. The sensors utilized in our study are listed in Table 1. The data collection session ends when the user correctly inputs the pattern thirty times.



lockpattern

Correct! Draw again in order of 3-2-5-8-7 (2/30)

user

(a) Snapshot of *S-Pattern* app

lockpatterncaptcha

Redraw pattern (1/30)

user

(b) Snapshot of *CR-Pattern* app

(c) Snapshot of *Gametrics* app. The target objects on the left are static, the objects on the right move randomly. The user has to match the cartoon animals with the real animal images.

Figure 1: Snapshots of the three apps

- ***CR-Pattern* App:** This app is similar to our *S-Pattern* app, except that the app displays a random pattern each time. A sample of the random pattern displayed to the user is as shown in Figure 1b. Each of the generated patterns has a length of 5 (the same length as that of the pattern used in *S-Pattern* app). The app generates a random pattern and the user is asked to input the displayed pattern. If the user failed in repeating the displayed pattern, the app re-displays the pattern, and the user is instructed to retry entering the pattern. Once the user enters the pattern correctly, the app generates a new random pattern. At each data collection session, the app generates 30 random patterns. As in *S-Pattern* app, the app records all the user interaction with the device measured by touch screen, motion and position sensors.
- ***Gametrics* App:** This Android app first displays an instruction text that provides the details on how to solve the interactive game challenge. The app then shows a text box where the user enters her name. Next, the user presses the start button to proceed with the experiment. The app then displays a semantic interactive challenge, represented as a simple drag and drop game. Each challenge consists of 3 target objects and 6 moving objects. In order to solve a challenge, the user needs to understand the contents of the target and the moving objects/images, the semantic

[1] This pattern was selected as it is one of most common patterns used by the users http://mytrickytricks.blogspot.com/2013/07/commonlockpattern.html

relationship between them, and then requires to drag a subset of the moving objects (answer objects) to their corresponding target objects. After each drag and drop, the game code checks the correctness of the user action. If the object was dropped to its corresponding target, the object disappears informing the user that she has performed a correct drag and drop, otherwise the object is moved to a random location. The game ends, when the user successfully drags all the answer objects to their corresponding targets. At the start of the challenge, the moving objects are placed at random locations on the screen of the phone and then each of them starts moving on a random direction picked from North, East, South, West, North-West, West-East, South-East, and South-West. The object keeps on moving on that direction till it collides with another object or the screen border where upon it picks a new random direction. During the data collection, once the challenge is solved successfully, a new challenge is presented to the user. We implemented 6 different semantic interactive challenges and presented 30 challenges to the user in random order by repeating each of the challenges five times. A snapshot of one of the semantic interactive challenge is shown in Figure 1c. The *Gametrics* app records the user interaction with the device while solving the challenges as in the previous two apps. It also records the moving object locations. The design of this app is in line with the one proposed in [27].

Although, for each of the schemes, we asked the users to input the pattern, or solve the challenge 30 times for our analysis purpose, in the real-world implementation, the user is asked for only once.

4 DATA COLLECTION METHODOLOGY

In our study, twenty participants were recruited by word-of-mouth. The majority of the participants were students at our University. 75% of the participants were males and 25% were females. The participants were composed of educated individuals. The majority of the participants aged between 25 and 34 years and came from Computer Science background. Table 2 summarized the demographics of the participants. The similar number of participants and demographics are well-established in lab-based studies in behavioral biometrics research [12, 14, 33], which serves to demonstrate the viability of the schemes.

Each of the participants were asked to perform each of the tasks ninety times, spanned over three days/sessions. During each session, the participants were asked to perform each task thirty times. We set the time gap between two consecutive sessions to be a minimum of 24 hours. During the study, we did not restrict the participants to a specific phone holding setting. The participants had the choice to use either a single hand or two hands. Further, they could perform the study while sitting or standing. The order of the tasks presented to different participants was derived using 3 × 3 Latin square to minimize the learning effect. The Latin square ensures that each user performed the three tasks in different orders. To avoid any kind of inconsistency, we used only one smartphone (Samsung Galaxy S6) throughout the data collection process.

We conducted the experiment following our University's IRB guidelines. The study and the experiment was approved by the IRB at our institution. The participants were clearly informed about the experiment, such as the data being collected, the purpose of the experiment, and that they can refuse to participate in the middle of the experiment or even request to delete their collected data during or after the experiment has been conducted.

Table 2: Demographics of participants ($N = 20$)

Category	% of participants
Gender	
Male	75
Female	25
Age	
25-35	95
>35	5
Field	
Computer Science	65
Non-CS	35
Education	
Bachelors	20
Masters	50
PhD	30

The participants were subjected to a consent agreement and a demographics form before the study. At the end of the third session, participants' experience in interacting with the three schemes was recorded using a survey form. The survey contains the 10 System Usability Scale (SUS) standard questions, each with 5 possible answers (5-point Likert scale, where 1 represents strong disagreement and 5 represents strong agreement) [8]. SUS is a standard questionnaire that is used to evaluate the usability of software, hardware, cell phones, and websites, and it has been deployed in many prior security usability studies.

5 FEATURE EXTRACTION AND CLASSIFICATION MODELS

In order to build each of the biometrics considered in this study, we utilized the machine learning approach. In this section, we present the features we extracted from the users' logs collected during our data collection campaign. Then, we discuss the classification models and the classifier employed in our study.

5.1 Feature Extraction

***S-Pattern*:** From each of the logs from *S-Pattern* app, we extracted a total of 55 features. These features can be characterized in the following three categories.

- **Touch sensor features:** Start touch size, end touch size, and the average touch size (3 features).
- **Swipe features:** Swipe time (total time taken by the user to enter the pattern), speed, acceleration (i.e., change in speed/time) and distance (4 features).
- **Motion and position based features:** From each of the sensors utilized in our study (i.e., accelerometer, rotation vector, linear acceleration, orientation, gyroscope, gravity and game rotation vector), we extracted following 6 statistical features – mean, standard deviation, minimum, maximum, number of local minima, and number of local maxima (48 features = 8 sensors × 6 statistical features).

Table 3: Performance for 10-fold cross validation of different classifiers for the three schemes.

		FPR	FNR	Precision	Recall	F-Measure
S-Pattern	RF	0.01 (0.01)	0.04 (0.02)	0.99 (0.01)	0.97 (0.02)	0.98 (0.01)
	MP	0.04 (0.02)	0.08 (0.04)	0.96 (0.02)	0.92 (0.03)	0.94 (0.02)
	J48	0.06 (0.03)	0.06 (0.04)	0.94 (0.03)	0.94 (0.04)	0.94 (0.03)
	SVM	0.03 (0.02)	0.10 (0.05)	0.97 (0.02)	0.91 (0.04)	0.94 (0.03)
	NB	0.05 (0.06)	0.11 (0.06)	0.95 (0.06)	0.89 (0.05)	0.92 (0.05)
	L	0.08 (0.03)	0.12 (0.05)	0.92 (0.03)	0.88 (0.05)	0.90 (0.04)
	RT	0.06 (0.03)	0.09 (0.04)	0.94 (0.03)	0.91 (0.04)	0.93 (0.03)
CR-Pattern	RF	0.01 (0.01)	0.05 (0.02)	0.99 (0.01)	0.96 (0.02)	0.97 (0.01)
	MP	0.02 (0.02)	0.09 (0.05)	0.97 (0.02)	0.92 (0.04)	0.94 (0.03)
	J48	0.05 (0.02)	0.06 (0.03)	0.95 (0.02)	0.94 (0.03)	0.95 (0.02)
	SVM	0.03 (0.03)	0.10 (0.06)	0.97 (0.03)	0.90 (0.05)	0.94 (0.04)
	NB	0.06 (0.07)	0.14 (0.11)	0.94 (0.07)	0.88 (0.08)	0.90 (0.05)
	L	0.09 (0.05)	0.15 (0.06)	0.91 (0.05)	0.86 (0.05)	0.88 (0.05)
	RT	0.07 (0.04)	0.08 (0.03)	0.93 (0.04)	0.92 (0.03)	0.92 (0.03)
Gametrics	RF	0.02 (0.02)	0.05 (0.03)	0.98 (0.02)	0.95 (0.03)	0.97 (0.02)
	MP	0.05 (0.03)	0.10 (0.04)	0.95 (0.03)	0.90 (0.03)	0.93 (0.03)
	J48	0.06 (0.03)	0.07 (0.03)	0.94 (0.03)	0.93 (0.03)	0.94 (0.02)
	SVM	0.04 (0.03)	0.12 (0.05)	0.96 (0.03)	0.89 (0.04)	0.92 (0.03)
	NB	0.08 (0.03)	0.16 (0.10)	0.92 (0.03)	0.86 (0.08)	0.89 (0.04)
	L	0.07 (0.04)	0.13 (0.06)	0.93 (0.04)	0.88 (0.05)	0.90 (0.04)
	RT	0.11 (0.03)	0.10 (0.05)	0.89 (0.03)	0.90 (0.05)	0.90 (0.03)

CR-Pattern: Similar to *S-Pattern*, we extracted the same 55 features from touch, swipe and, motion/position categories. The one exception was that instead of using the exact distance traveled in the distance-based features, we used the difference between the distance traveled and the minimum distance required to enter the pattern.

Gametrics: From each of the logs for *Gametrics*, we extracted a total of 78 features that capture the cognitive abilities of the user while she is solving the challenges as well as the features extracted from the touch, motion and position sensors.

In the previous two methods, the user has to perform a single, relatively long swipe. However, in *Gametrics*, the user has to perform a minimum of three relatively short swipes (drags and drops). From the touch sensor data, we extracted 12 features – average, standard deviation, minimum and maximum of start touch size, end touch size and average touch size. Moreover, for the swipe features, rather than extracting a single feature from each of the speed, the acceleration, and the distance, we extracted the statistical features corresponding to each of the average, standard deviation, minimum and maximum (12 features). As in the previous two tasks, we extracted the same statistical features from the motion and position sensors.

As described in Section 3, in order to solve a semantic interactive challenge, the user has to match the answer objects to their corresponding targets. In order to do that, the user has to understand the content of the images representing the targets and the moving objects, find the relationship between the moving objects and the target objects, and then select a subset of the moving objects (the answer objects), and finally drag/drop them to their corresponding targets. By monitoring the users as they solved the challenges, we found different users take different approaches to solve the challenges. For example, some users start by trying to comprehend the whole challenge and then start the object matching task while some try to find the answer objects corresponding to the target in certain order (i.e., always try to search for the answer object that corresponds to the top most target, and then the second and so on). Others try to pick the object closest to the finger and then check if it matches with any of the targets.

These different mechanisms of solving the semantic challenges are related to the cognitive characteristics of individuals. We capture these characteristics based on the following 6 features (these features are similar the ones used in [27]).

(1) The time between the user pressing the start button and the first recorded touch event. This timing measure captures the time the user takes to comprehend the challenge and start solving it.
(2) The average, standard deviation, minimum and maximum of the times between each of the drops and the start of the next drag. These features capture the time the user takes to find the next answer object.
(3) The total time taken by the user to complete the challenge.

5.2 Classification Metrics & Classifier

Classification Metrics: In our classification task, the positive class corresponds to the legitimate user interaction with the authentication construct and the negative class corresponds to the impersonator (other user or the *zero-effort attacker*). Therefore, true positives (TP) represent the number of times the legitimate user is granted access, true negatives (TN) represents the number of times the impersonator is rejected, false positives (FP) represent the number of times the impersonator is granted access and false negatives (FN) represent the number of times the legitimate user is rejected.

As performance measures for our classification models, we used False Positive Rate (FPR), False Negative Rate (FNR), precision, recall and F-measure (F1 score), as shown in Equations 1-5. Precision and FPR measure the security of the proposed system, i.e., the accuracy of the system in rejecting impersonators. Recall and FNR capture the usability of the proposed system as low recall leads to high rejection rate of legitimate users. F-measure considers both the usability and the security of the system. To make the system both usable and secure, ideally, we would like to have a recall, precision, and F-measure to be as close as 1.

$$FPR = \frac{FP}{TN + FP} \tag{1}$$

$$FNR = \frac{FN}{TP + FN} \tag{2}$$

$$precision = \frac{TP}{TP + FP} \tag{3}$$

$$recall = \frac{TP}{TP + FN} \tag{4}$$

$$F\text{-}measure = 2 * \frac{precision * recall}{precision + recall} \tag{5}$$

Classifier: With the data samples collected in our study, we tested different machine learning algorithms – J48, Random Forest (RF), Random Tree (RT), Multilayer Perceptron (MP), Support Vector Machines (SVM), Logistics (L), and Naive Bayes (NB). We applied 10-fold cross-validation approach to test all these machine learning algorithms. Table 3 summarizes the classification results for the three studied schemes. We achieved the best results with Random Forest classifiers (F-Measure = 98% for *S-Pattern*, and 97% for *CR-Pattern* and *Gametrics*). Therefore, in our analysis, we utilized the Random Forest classifier.

Random Forest is an ensemble learning approach that constructs many classification trees during the learning phase where each tree is generated using a separate bootstrap sample of the data. In the testing/classifying phase, the new data is run down all the trees and the output is the mode of the votings from each individual tree. Random Forest is robust against noise, efficient, can estimate the importance of the features and have shown to give promising results in similar tasks [24, 27].

To avoid the overfitting and improve the classification performance and results, we performed exhaustive search to find the subset of features that results in the best F-measure for each of the classification tasks.

6 CLASSIFICATION RESULTS

In this section, we present the classification results of our study.

6.1 Intra-Session Analysis

As mentioned in Section 4, we collected data from 20 volunteers. In the first day of our data collection experiment, each of the volunteers completed 30 challenges of each of the three studied schemes. We divided the collected data into 60 sets based on the users' identities (ids) and the scheme. In order to build a classifier to authenticate a user, we defined two classes. The first class contains the features data from a given user and a given scheme, and the other class contains randomly selected features data from other 19 users of the same scheme. Then, we divided the data into two sets, one for training and the other for testing. The first 18 instances of each user and 18 instances of the randomly selected set are used to train the classifier, while the remaining 12 are used for testing.

The results of intra-session analysis are shown in the first part of Table 4. Without utilizing the sensors features, we find that the three scheme provide similar classification results. The F-measure came out to be 0.89, 0.86 and 0.83 for *S-Pattern*, *CR-Pattern* and *Gametrics*, respectively. Comparing the F-measures of the three tested schemes using Friedman test, we did not find statistical significance (F(20, 2) = 6.7, p = 0.13)[2].

The second row of the first part in Table 4 show that including the sensors features improves the classification accuracy (F-measure = 1 for all three schemes). We employed the Wilcoxon Signed-Ranked Test (WSRT) with Bonferroni correction to analyze the statistical significance of the F-measure of each of the schemes with and without the sensors features. We found statistical significant difference for all the three schemes with and without sensors features ($p < 0.01$).

The three sensors that has been used most by our feature selection algorithm are Orientation, Linear Acceleration and Rotation Vector sensors.

6.2 Inter-Session Analysis

The purpose of the inter-session analysis is to analyze the effectiveness of the studied schemes over multiple sessions/days. For the data instances we collected in day 2 and day 3, we trained the classifier with the data instance collected in the previous day(s) and tested with the data of that day.

The results are shown in the second and the third parts of the Table 4. The results came inline with the results obtained in intra-session analysis. The three schemes have similar classification accuracies. Also, the results show that utilizing the sensors features improves the accuracy for all the three studies schemes. We find that the performance of the classifier degrades slightly compared to the intra-session analysis, which is as expected.

Comparing the F-measures of the three schemes without sensors features using Friedman test, we did not find any statistical significant difference. Also, we did not find statistical significant difference between the F-measures of the three schemes when the sensor features were included.

Furthermore, for each of the tested schemes, both in day 2 and in day 3, we compared the F-measure of the classifier that utilize the sensors features with its correspondent without using the sensors features using Wilcoxon Signed-Ranked Test (WSRT) with Bonferroni correction. For all of the tested pairs, we found statistical difference (p < 0.01 for all the tested pairs).

Summary of Results The results obtained in this section show that utilizing the sensors features improves the accuracy in identifying the users and rejecting the zero-effort attacker for all the

[2] All statistical results reported in this paper are at the 95% confidence level

Table 4: Performance of the classifier for three different schemes. The first part shows the performance of the classifier in intra-session. Part two and three show the performance for inter-session

			FPR	FNR	Precision	Recall	F-Measure
Intra-Session	**Excluding Sensors**	***S-Pattern***	0.15	0.08	0.87	0.92	0.89
		CR-Pattern	0.20	0.10	0.83	0.90	0.86
		Gametrics	0.27	0.10	0.78	0.90	0.83
	Including Sensors	***S-Pattern***	0.01	0.00	0.99	1.00	1.00
		CR-Pattern	0.00	0.00	1.00	1.00	1.00
		Gametrics	0.00	0.00	1.00	1.00	1.00
Inter-Session Day 2	**Excluding Sensors**	***S-Pattern***	0.20	0.23	0.81	0.77	0.78
		CR-Pattern	0.27	0.19	0.75	0.81	0.77
		Gametrics	0.31	0.15	0.73	0.85	0.79
	Including Sensors	***S-Pattern***	0.06	0.09	0.95	0.91	0.93
		CR-Pattern	0.11	0.06	0.90	0.94	0.92
		Gametrics	0.12	0.06	0.89	0.94	0.91
Inter-Session and Day 3	**Execluding Sensors**	***S-Pattern***	0.22	0.16	0.80	0.84	0.82
		CR-Pattern	0.27	0.19	0.75	0.81	0.77
		Gametrics	0.24	0.13	0.79	0.87	0.83
	Including Sensors	***S-Pattern***	0.07	0.07	0.93	0.93	0.93
		CR-Pattern	0.11	0.06	0.90	0.94	0.92
		Gametrics	0.07	0.07	0.93	0.93	0.93

three biometric schemes. The results also show that three schemes have similar classification accuracy. The precision and recall are up to 1 when we include the sensor data in the analysis in the intra-sessions study and above 0.89 for all schemes in the inter-sessions study. Similar results have been reported for other (static) behavioral biometric schemes in the literature [3].

7 USER EXPERIENCE ANALYSIS

In this section, we present the time taken by the participants to solve the challenges of each of the studied tasks. Further, we analyze the user's experience towards each of the studied tasks using the standard usability rating questionnaire, i.e. SUS rating.

Table 5: The average (standard deviation) time taken by the participants to solve a challenge of each of the three schemes.

	Time Mean (Std.)
S-Pattern	0.99 (± 0.32)
CR-Pattern	6.53 (±2.00)
Gametrics	8.22 (±4.61)

Solving Time: The time that users took to perform each of the tasks is summarized in Table 5. The users took on average around 1 second to complete a challenge of *S-Pattern*. The time to solve the task increased to 6.5 and 8.2 seconds on average for *CR-Pattern* and *Gametrics*, respectively. Note that the time for *CR-Pattern* is longer than its correspondent in *S-Pattern* because in *CR-Pattern*, we also considered the time needed to display the random pattern to the user along with the time taken by the user to complete the task. Further, since the pattern in *CR-Pattern* is different each time, the users took a longer time to input the pattern. Comparing the average time taken by the participants to solve the challenges using Friedman Test, we found statistical significant difference (F(1800, 2) = 2745.79, $p < 0.001$). Further, comparing the solving time with Wilcoxon Signed Ranks Test, we found statistical significant difference between all the three pairs ($p < 0.001$).

Table 6: Mean (standard deviation) SUS Score of the three studied biometrics

	SUS Score Mean (Std.)
S-Pattern	82.63 (±12.29)
CR-Pattern	80.50 (±12.05)
Gametrics	77.88 (±12.20)

SUS Score: We next evaluate the data collected during the post-study phase from the participants. The SUS scores of the three studied schemes are summarized in Table 6. The mean SUS score came out to be the highest for *S-Pattern*, and slightly lower for *CR-Pattern* and *Gametrics*.

Although, the mean SUS scores of *CR-Pattern* and *Gametrics* are slightly lower compared to that of *S-Pattern*, Friedman Test did not find any statistical differences on the mean of SUS scores among these three behavioral biometrics. Given that the system with SUS score greater than 68 is considered above the average [30], our results from SUS show that the three schemes are equally usable.

8 SECURITY ANALYSIS

Previously, in Section 6, we demonstrated that the proposed authentication schemes are robust against zero-effort attacks reflected in the high precision. In this section, we analyze and compare the security of the proposed schemes against active impersonation attacks.

8.1 Smudge Attacks

The first attack that we aimed to prevent in our threat model is the smudge attack. The studied three schemes provide the security against such type of attacks by utilizing the features based on motion-position and touch sensors This is because even if the attacker is able to trace the pattern, he will not get enough information about the behavioral gesture, specifically how to hold the phone and swipe, while entering the pattern or solving the semantic challenges. Further, smudge attack relies on the reconstruction of a secret (i.e., the visual pattern in our case). Since *CR-Pattern* and *Gametrics* do not contain any secrets, these schemes, by design, are able to prevent the smudge attack.

8.2 Shoulder-Surfing Attacks

The second attack that we aimed to prevent in our threat model is the shoulder-surfing (or impersonation attack). We analyze the security of the three schemes against deliberate impersonation attacks. During our data collection, one of the researchers played the role of an attacker (representing a relatively well-trained attacker). He monitored the participants while they were performing the tasks through a video recording. For the impersonation attack analysis, the attacker picked two of the participants at random from the pool, and tried to mimic those chosen participants by solving the challenges in a similar way as the participants did for each of the scheme. The impostor made 30 attempts to impersonate each of the chosen users. Both of the chosen participants were right handed, and preferred to perform the tasks while sitting on a chair similar to the attacker.

Table 7: Results of Shoulder-Surfing Attacks

	Average FPR	
	Excluding Sensors	*Including Sensors*
S-Pattern	0.35	0.12
CR-Pattern	0.35	0
Gametrics	0.22	0

Table 7 shows the performance of impersonation attack in terms of false positive/acceptance rate. The results show that the attack success rate decreases significantly when including the features from various sensors under consideration. On average, when sensors features were not included, the success rate of the impersonation attack was 0.35 for both *S-Pattern* and *CR-Pattern*, and 0.22 for *Gametrics*. When sensors features were used, the attack success rate decreased significantly to 0.12 for *S-Pattern*, and 0.00 for *CR-Pattern* and *Gametrics*. Although these results are based on the impersonation attack against only two users, similar results will also apply for other users. This suggests that the three scheme offer high level of resilience against shoulder-surfing attacks, especially when the sensors features are used.

8.3 Automated Attacks & Internal Attacks

In the rest of our analysis, as a generalization of a robot and a malware program, we consider the most powerful attack among them because if a scheme is secure against the most powerful attack, it will also be secure against a relatively less powerful attack. The most powerful attack that we consider in our study is an attack that has the ability to record the touch events as well as other sensors values when the user interacts with the authentication construct. Further, we assume that the attacker has the ability to inject the touch events as well as the motion-position sensors events at will. Such attack has been explored and implemented in [28], where the attack takes a form of malicious code, called SMASheD, that is accidentally installed on the device using ADB (Android Debugging Bridge) and is therefore granted several permissions including reading from and writing to the sensors files.

Although SMASheD attack is extremely powerful and its threat model is very strong, such an attack can be assumed to be a generalization of other types of attacks, including a human impersonator (a human that can be trained to identically mimic another user), or a robot such as the one proposed in [32] with no physical constrains that can be programmed to interact with the mobile device in any way it likes.

Recording user interactions with the device is also not straightforward except for the case of the SMASheD attack. However, multiple other approaches have been explored including: recording the user interaction with a malware that looks like a normal authentication construct (i.e., using social engineering tricks), approximately learning the user interaction with the device by manually watching the user [31, 36], recording the user interaction using (surveillance) camera [4], or hacking the server database to learn the stored authentication token.

Next, we analyze and compare the security of the three proposed behavioral biometrics schemes against the above-defined attack.

Security of *S-Pattern*: If the attacker is able to record the user's valid interaction with the authentication construct underlying the *S-Pattern* scheme, it can theoretically fool the authentication system simply by replaying the recorded values (given the ability we explained above as part of the attack). However, for a real-world attacker (e.g., a robot with physical constraints or a human impersonator), such an attack is not straight forward as the attacker needs to mimic the user interaction with the device measured by the touch screen sensor and the motion-position sensors on the device.

Security of *CR-Pattern*: *CR-Pattern* is harder to attack compared to *S-Pattern* as the randomization/interactivity in *CR-Pattern* adds an extra burden on the attacker. Simply recording and replaying the user interactions do not allow the attacker to bypass the security provided by this scheme. To attack the *CR-Pattern* scheme, the attacker needs to understand the pattern it needs to re-enter, then try to mimic the user. However, mimicking, in this case, is harder as replaying previously recorded values will not work because the pattern shown at a given time is typically different from the one previously shown and recorded. The attacker can try to learn the user's way of interaction with the scheme by recording multiple sessions and then trying to input the new pattern mimicking the user utilizing the learned knowledge. This suggests that *CR-Pattern* is much more robust to impersonation attacks compared to *S-Pattern*, which is the inherent benefit of using challenge-response authentication in *CR-Pattern.*

Security of *Gametrics*: In order to solve a *Gametrics* challenge, the attacker needs to understand the content of the images, find the relationships between the target and moving objects, and then drag and drop the answer objects to their corresponding target. To be able to bypass this scheme, the attacker needs to mimic the user interactions with the challenges measured by the touchscreen, motion and position sensors. Moreover, the attacker needs to match the timing of solving the challenge with the legitimate user (i.e., match the cognitive features).

We argue that attacking such biometrics with a human attack is challenging as the attacker needs to match about 70 features used in our classification models. Automated attacks against *Gametrics* are considerably hard. This is because solving the challenge involves solving two hard AI problems: (1) understanding the contents of the images, and (2) finding the semantic relationships. Such type of challenges is considered hard for automated algorithms to solve and, in fact, this scheme has been suggested to be used as CAPTCHAs [26], a method to determine if the user is a human or a bot.

In summary, interactivity and randomization in *Gametrics* allow us to extract more features that can help in identifying the user. Moreover, they make attacking such challenges harder as the requirement of mimicking many features including the cognitive features.

9 DISCUSSION AND FUTURE WORK

In this work, we studied three different schemes for authenticating the user on mobile devices based on behavioral biometrics. The three schemes differ each other in terms of their usability and security. Since *S-Pattern* involves a short authentication time, it has relatively high usability but is vulnerable to various types of impersonation attacks. *CR-Pattern* and *Gametrics* have similar completion time, both of them have better security compared to *S-Pattern.* We argued that the randomization in *CR-Pattern* and *Gametrics* makes the attack harder against them, more harder in the case of *Gametrics* due to cognitive behavioral properties.

Our study shows that utilizing the sensors in three studied schemes improves the accuracy of identifying the user. Moreover, utilizing additional features makes the attack harder as the attacker needs to mimic all the features to be authenticated.

The most common and traditional mechanism for authentication in mobile phones is password. Previous work has shown that the average time taken by the users to enter a password on a mobile device is up to 21 seconds [25]. Unlike the password, the proposed schemes require relatively short time to solve the challenges (about 1 second for *S-Pattern*, and around 6.5 seconds for *CR-Pattern* and 8 seconds for *Gametrics*), and can fit well for many application scenarios. *S-Pattern* seems more viable for the purpose of phone locking (point-of-entry authentication) while the other two methods seem amenable for the purpose of app authentication (such as banking apps or locking photos for users who are concerned about their photo privacy) or web-site authentication.

Gametrics may be incorporated with graphical passwords [7, 35], such as those involving Random Arts images [29], objects (PassObjects) [41] and faces (PassFaces) [5], as well as on recall or cued recall, such as those involving drawings [16, 21] and selection of points on an image (PassPoints) [40], as a second factor authentication to improve the security of the graphical passwords from spoofing attacks. *Gametrics* may also be used as a method for fall-back authentication. Typically, fall-back authentication does not require fast authentication time as it is not used often by the users. However, in order to build an up-to-date classification model for the user to fall-back, the system may need to ask the user to solve challenges periodically. As an additional use case, the three studied schemes can be used to confirm that the user is the one she claims to be in cases, for example, when the continuous authentication schemes cannot recognize the user with high probability. Further investigation is needed to realize such use cases of the studied schemes. In our future work, we will study different variant of interactive user authentication with the aim to reduce the time required during the authentication process. We will also study how the performance of the biometrics changes with the change in the device used.

The proposed biometrics would suffer, like any behavioral biometrics, from degradation in classification accuracy in case of user's behavioral changes, such as emotional changes [17] or sickness. We plan to study the effect of such behavioral changes in our future work. Moreover, we will study the effect of walking, driving, and other motion scenarios on the authentication accuracy. The results of our study are promising, however, more work is needed to reduce the authentication time, test the proposed schemes on larger/varying pool of users, and compare the three schemes in terms of user perception and acceptability.

10 RELATED WORK

The most studied approaches for behavioral biometrics on smarphones are based on touchstroke footprints and implicit behavioral gestures. Only few researches (e.g., [2, 13]) studied user authentication based on user's cognitive abilities. Al Galib et al. [2] studied the ability of authenticating the users based on their cognitive process captured by visual search, working memory and priming effect on automatic processing. The game they utilized to capture the users' cognitive abilities provides a challenge-response task. In each instance of the challenge-response, the user is given a challenge, which is an object. The user's task is to drag the challenge object onto the matching object inside the search set. After a valid drop, the user then receives a gold coin as a reward and deposits it in a bank. On a correct deposit, the user is challenged with a new object and the game continues as before. From the interaction with the challenges, the authors extracted several features that capture the cognitive abilities of the users, however, they did not look into the mouse dynamics biometrics of the users. Chen et al. [13] proposed a method to solve account hijacking and share problems in an online gaming environment. They propose identifying the user based on her gameplay activities. They show that the idle time distribution is a representative feature of game players. They propose the relative entropy test RET scheme, which is based on the Kullback-Leibler divergence between idle time (i.e., the idle periods between successive moves of a player-controlled character) distributions, for user identification. Their evaluation shows that the RET scheme achieves higher than 90% accuracy with a 20-minute detection time given a 200-minute history size.

Many behavioral biometric based authentication mechanisms on mobile devices have been proposed. *Conti et al.* [14] proposed a system that transparently authenticates the user by analyzing her hand movement gesture while she is making or answering a phone call. It uses accelerometer and orientation sensor to detect the proposed gesture. The system uses the dynamic time warping distance (DTW-D) algorithm to verify if the authorized user is making or answering the phone call. Buriro et al. [9] introduced similar authentication system, *AnswerAuth*, based on the way the user slides the lock button on the screen to unlock the phone and the implicit gesture of bringing the phone towards the ear. It utilizes multiple built-in sensors, such as accelerometer, gyroscope, gravity, magnetometer, and touchscreen, to capture the said gestures, and employs machine learning algorithms to detect the sliding and phone-lifting actions.

Buriro et al. [10] have also proposed another bimodal authentication system, *DialerAuth*, for smartphone based on the user touchstroke pattern and generated phone movements. The proposed system requires the user to tap/enter a random non-secret 10-digit number. It utilizes tap-timing to capture touchstroke pattern while entering the 10-digit number and uses accelerometer and gyroscope sensors to capture the device micro-movements. Similarly, Akhtar et al. [1] introduced multimodal user identification system. Unlike DialerAuth, the proposed system employs user's facial features in addition to phone-movements and touchstrokes patterns to transparently authenticate the user. Similar to DialerAuth, it requires the user to enter a random 8-digit number. It employs device's camera to capture facial features, accelerometer, gravity, and magnetometer sensors to record device movements and touch sensors for touchstroke pattern. Van Nguyen et al. [39] proposed *Draw-A-Pin*, a user authentication system for touch-enabled devices based on PIN and PIN drawing characteristics. In this scheme, user is asked to draw his PIN on the touch screen of the device instead of typing it on a keypad. It employs touch sensors to capture the PIN drawing characteristics.

Further, Buriro et al. [11] have proposed a user authentication scheme on smartphone after the user has unlocked the phone. The proposed scheme is based on the profile of the user's hand micro-movement while the user is using his phone after it has been unlocked. It leverages built-in sensors, such as accelerometer, gyroscope, magnetometer, and orientation, to capture the micro-movement of the user's hands. It compares the captured hand movements profile with the stored template utilizing machine learning algorithms to authenticate the user.

Gascon et al. [19] and Lee et al. [22] presented an approach to continuously authenticate users on smartphones by analyzing their typing motion behavior. Both approaches utilizes built-in motion and position sensors (e.g., accelerometer and gyroscope) to capture behavioral biometrics so as to authenticate the user. Lee et al. also utilizes auxiliary motion information from a wearable wrist-device (e.g., smartwatch) on their authentication system.

11 CONCLUSION

In this paper, we studied two challenge-response methods for behavioral biometric authentication on mobile devices and compared them with an existing static graphical pattern biometric scheme. For each of the studied schemes, we show that utilizing the motion sensors improves the accuracy of detecting the user and security against impersonation attack. Moreover, we argued that utilizing challenge-response schemes improves the security of the authentication, although with an increase in the time taken to authenticate the user (but still less than 10 seconds). Our study shows that game-based biometric scheme has a similar accuracy and completion time as the challenge-response pattern biometric scheme, but it has a higher level of security. Our results suggest that each of the three schemes may be used in different applications depending upon the desired level of security and usability.

ACKNOWLEDGEMENTS

This work is partially supported by Google Faculty Research Award and National Science Foundation (NSF) grants: CNS-1209280 and CNS-1526524.

REFERENCES

[1] Zahid Akhtar, Attaullah Buriro, Bruno Crispo, and Tiago H Falk. 2017. Multimodal smartphone user authentication using touchstroke, phone-movement and face patterns. In *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. IEEE, 1368–1372.

[2] Asadullah Al Galib and Reihaneh Safavi-Naini. 2015. User Authentication Using Human Cognitive Abilities. In *Financial Cryptography and Data Security*. Springer, 254–271.

[3] Abdulaziz Alzubaidi and Jugal Kalita. 2016. Authentication of smartphone users using behavioral biometrics. *IEEE Communications Surveys & Tutorials* 18, 3 (2016), 1998–2026.

[4] Panagiotis Andriotis, Theo Tryfonas, George Oikonomou, and Can Yildiz. 2013. A pilot study on the security of pattern screen-lock methods and soft side channel attacks. In *Proceedings of the sixth ACM conference on Security and privacy in wireless and mobile networks*. ACM, 1–6.

[5] Real User Personal Authentication. 2004. The Science Behind Passfaces. *White Paper, June* (2004).

[6] Adam J Aviv, Katherine Gibson, Evan Mossop, Matt Blaze, and Jonathan M Smith. 2010. Smudge Attacks on Smartphone Touch Screens. *WOOT* 10 (2010), 1–7.

[7] Robert Biddle, Sonia Chiasson, and Paul Van Oorschot. 2009. Graphical passwords: Learning from the first generation. In *Technical Report TR-09-09, School of Computer Science, Carleton University*.

[8] John Brooke. 1996. SUS: a "Quick and Dirty" Usability Scale. In *Usability Evaluation in Industry*, P. W. Jordan, B. Thomas, B. A. Weerdmeester, and A. L. McClelland (Eds.). Taylor and Francis, London.

[9] Attaullah Buriro, Bruno Crispo, and Mauro Conti. 2019. AnswerAuth: A bimodal behavioral biometric-based user authentication scheme for smartphones. *Journal of information security and applications* 44 (2019), 89–103.

[10] Attaullah Buriro, Bruno Crispo, Sandeep Gupta, and Filippo Del Frari. 2018. Dialerauth: A motion-assisted touch-based smartphone user authentication scheme. In *Proceedings of the Eighth ACM Conference on Data and Application Security and Privacy*. ACM, 267–276.

[11] Attaullah Buriro, Bruno Crispo, and Yury Zhauniarovich. 2017. Please hold on: Unobtrusive user authentication using smartphone's built-in sensors. In *2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*. IEEE, 1–8.

[12] P Campisi, E Maiorana, M Lo Bosco, and A Neri. 2009. User authentication using keystroke dynamics for cellular phones. *IET Signal Processing* 3, 4 (2009).

[13] Kuan-Ta Chen and Li-Wen Hong. 2007. User identification based on game-play activity patterns. In *Proceedings of the 6th ACM SIGCOMM workshop on Network and system support for games*. ACM, 7–12.

[14] Mauro Conti, Irina Zachia-Zlatea, and Bruno Crispo. 2011. 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.* ACM, 249–259.

[15] Alexander De Luca, Alina Hang, Frederik Brudy, Christian Lindner, and Heinrich Hussmann. 2012. Touch Me Once and I Know It's You!: Implicit Authentication Based on Touch Screen Patterns. In *SIGCHI Conference on Human Factors in Computing Systems (CHI).*

[16] Paul Dunphy and Jeff Yan. 2007. Do background images improve "draw a secret" graphical passwords?. In *CCS '07: Proceedings of the 14th ACM conference on Computer and communications security.* ACM, 36–47.

[17] Clayton Epp, Michael Lippold, and Regan L Mandryk. 2011. Identifying emotional states using keystroke dynamics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 715–724.

[18] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song. 2013. Touchalytics: On the Applicability of Touchscreen Input as a Behavioral Biometric for Continuous Authentication. *IEEE Transactions on Information Forensics and Security* (2013).

[19] Hugo Gascon, Sebastian Uellenbeck, Christopher Wolf, and Konrad Rieck. 2014. Continuous Authentication on Mobile Devices by Analysis of Typing Motion Behavior.. In *Sicherheit.*

[20] Abdenour Hadid, Nicholas Evans, Sébastien Marcel, and Julian Fierrez. 2015. Biometrics systems under spoofing attack: an evaluation methodology and lessons learned. *IEEE Signal Processing Magazine* 32, 5 (2015), 20–30.

[21] Ian Jermyn, Alain Mayer, Fabian Monrose, Michael K. Reiter, and Aviel D. Rubin. 1999. The design and analysis of graphical passwords. In *SSYM'99: Proceedings of the 8th conference on USENIX Security Symposium.*

[22] Wei-Han Lee and Ruby B Lee. 2017. Implicit smartphone user authentication with sensors and contextual machine learning. In *2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN).* IEEE, 297–308.

[23] Lingjun Li, Xinxin Zhao, and Guoliang Xue. 2013. Unobservable Re-authentication for Smartphones.. In *Network and Distributed System Security Symposium (NDSS).*

[24] Roy A Maxion and Kevin S Killourhy. 2010. Keystroke biometrics with number-pad input. In *Dependable Systems and Networks (DSN), 2010 IEEE/IFIP International Conference on.* IEEE, 201–210.

[25] William Melicher, Darya Kurilova, Sean M Segreti, Pranshu Kalvani, Richard Shay, Blase Ur, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Michelle L Mazurek. 2016. Usability and security of text passwords on mobile devices. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.* ACM, 527–539.

[26] Manar Mohamed, Niharika Sachdeva, Michael Georgescu, Song Gao, Nitesh Saxena, Chengcui Zhang, Ponnurangam Kumaraguru, Paul C van Oorschot, and Wei-Bang Chen. 2014. A three-way investigation of a game-CAPTCHA: automated attacks, relay attacks and usability. In *Proceedings of the 9th ACM symposium on Information, computer and communications security.* ACM, 195–206.

[27] Manar Mohamed and Nitesh Saxena. 2016. Gametrics: Strong Behavioral Authentication with Simple Cognitive Games. In *Computer Security Applications Conference (ACSAC).*

[28] Manar Mohamed, Babins Shrestha, and Nitesh Saxena. 2016. SMASheD: Sniffing and Manipulating Android Sensor Data. In *Proceedings of the Sixth ACM Conference on Data and Application Security and Privacy.* ACM, 152–159.

[29] Adrian Perrig and Dawn Song. 1999. Hash Visualization: a New Technique to Improve Real-World Security. In *CrypTEC.*

[30] Jeff Sauro. 2015. Measuring Usability with the System Usability Scale (SUS). February 2, 2011. *URL http://www. measuringusability. com/sus. php* (2015).

[31] Florian Schaub, Ruben Deyhle, and Michael Weber. 2012. Password entry usability and shoulder surfing susceptibility on different smartphone platforms. In *Proceedings of the 11th international conference on mobile and ubiquitous multimedia.* ACM.

[32] Abdul Serwadda and Vir V Phoha. 2013. When kids' toys breach mobile phone security. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security.* ACM, 599–610.

[33] Muhammad Shahzad, Alex X Liu, and Arjmand Samuel. 2013. Secure unlocking of mobile touch screen devices by simple gestures: you can see it but you can not do it. In *Proceedings of the 19th annual international conference on Mobile computing & networking.* ACM, 39–50.

[34] Youngbae Song, Geumhwan Cho, Seongyeol Oh, Hyoungshick Kim, and Jun Ho Huh. 2015. On the effectiveness of pattern lock strength meters: Measuring the strength of real world pattern locks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* ACM, 2343–2352.

[35] Xiaoyuan Suo, Ying Zhu, and G Scott Owen. 2005. Graphical passwords: A survey. In *21st Annual Computer Security Applications Conference (ACSAC'05).* IEEE.

[36] F. Tari, A. Ant Ozok, and S. H. Holden. 2006. A Comparison of Perceived and Real Shoulder-surfing Risks Between Alphanumeric and Graphical Passwords. In *SOUPS: Proceedings of the second symposium on Usable privacy and security.*

[37] Chee Meng Tey, Payas Gupta, and Debin Gao. 2013. I can be You: Questioning the use of Keystroke Dynamics as Biometrics. In *The 20th Annual Network & Distributed System Security Symposium (NDSS 2013).*

[38] Sebastian Uellenbeck, Markus Dürmuth, Christopher Wolf, and Thorsten Holz. 2013. Quantifying the security of graphical passwords: the case of android unlock patterns. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security.* ACM, 161–172.

[39] Toan Van Nguyen, Napa Sae-Bae, and Nasir Memon. 2017. DRAW-A-PIN: Authentication using finger-drawn PIN on touch devices. *computers & security* 66 (2017), 115–128.

[40] Susan Wiedenbeck, Jim Waters, Jean-Camille Birget, Alex Brodskiy, and Nasir D. Memon. 2005. PassPoints: Design and Longitudinal Evaluation of a Graphical Password System. In *International Journal of Human Computer Studies.*

[41] Susan Wiedenbeck, Jim Waters, Leonardo Sobrado, and Jean-Camille Birget. 2006. Design and Evaluation of a Shoulder-surfing Resistant Graphical Password Scheme. In *Proceedings of the working conference on Advanced visual interfaces (AVI).*