Use the Force: Evaluating Force-Sensitive Authentication for Mobile Devices

Katharina Krombholz
Ruhr-University Bochum, Germany and
SBA Research, Austria
kkrombholz@sba-research.org

Thomas Hupperich
Ruhr-University Bochum, Germany
thomas.hupperich@ruhr-uni-bochum.de

Thorsten Holz
Ruhr-University Bochum, Germany
thorsten.holz@rub.de

ABSTRACT

Modern, off-the-shelf smartphones provide a rich set of possible touchscreen interactions, but knowledge-based authentication schemes still rely on simple digit or character input. Previous studies examined the shortcomings of such schemes based on unlock patterns, PINs, and passcodes.

In this paper, we propose to integrate pressure-sensitive touchscreen interactions into knowledge-based authentication schemes. By adding a (practically) invisible, pressure-sensitive component, users can select stronger PINs that are harder to observe for a shoulder surfer. We conducted a within-subjects design lab study (n = 50) to compare our approach termed *force-PINs* with standard four-digit and six-digit PINs regarding their usability performance and a comprehensive security evaluation. In addition, we conducted a field study that demonstrated lower authentication overhead. Finally, we found that force-PINs let users select higher entropy PINs that are more resilient to shoulder surfing attacks with minimal impact on the usability performance.

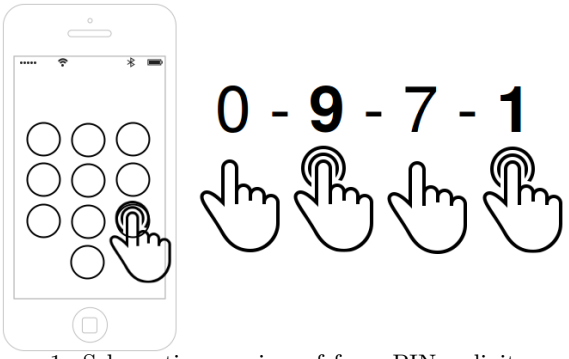


Figure 1: Schematic overview of force-PINs: digits can either be entered with shallow or deep pressure on a pressure-sensitive touchscreen, enhancing the space of four-digit PINs to $20^4 = 160,000$ by an invisible component. The user receives vibration feedback as soon as deep pressure is recognized.

1. INTRODUCTION

With the introduction of pressure-sensitive touchscreens (e. g., Apple recently introduced *3D Touch*[1]), many new kinds of user interaction for smartphones become possible that could also be used to enhance existing authentication schemes. The scientific community has already examined the shortcomings of unlock patterns, PINs and passcodes [2, 16, 19, 25] and presented alternative authentication schemes.

However, none of the proposed systems has shown to be capable of replacing passcodes and unlock patterns as means of authentication. On the one hand, many approaches, e.g., [15, 17] rely on customized hardware that is not available off the shelf and thus makes large-scale deployment infeasible. On the other hand, many alternative approaches, e.g., [13, 23] are time-consuming and therefore increase the authentication overhead. As shown by Harbach et al. [19] in a field study on smartphone unlocking behavior, (un-)locking smartphones produces a significant task overhead. This highlights the need for novel authentication methods that perform equally fast as or even faster than currently deployed systems in terms of authentication speed.

Recently, biometric approaches such as fingerprint sensors and face recognition have found their way into the mobile ecosystem. As with previous authentication methods, however, they have shown to be easy to break by attackers and difficult to use for certain groups of users. For example, Apple's fingerprint sensor as found in some recent iPhone models was soon hacked after being introduced [11] and excludes users with weak fingerprints (e. g., due to manual labor). Furthermore, classic biometric methods and implicit authentication based on user behavior still require users to use a PIN for fallback authentication in case the primary authentication methods fail. Bonneau et al. [8] presented a benchmark to evaluate authentication schemes. Their evaluation shows that many schemes only offer minor improvements over passwords (if any) and that many systems offer a number of benefits in theory but show severe limitations in practice. These observations highlight that it is still worth focusing on improving knowledge-based authentication on smartphones as no other authentication method has proven to be as secure and usable as passwords.

[1] `https://developer.apple.com/ios/3d-touch/`

Symposium on Usable Privacy and Security (SOUPS) 2016, July 22–24, 2015, Denver, Colorado, USA.

In this paper, we propose that device manufacturers integrate pressure-sensitive touchscreen interactions available on mobile and wearable devices into knowledge-based authentication schemes. Our goal is to improve PIN security by enhancing the password space without compromising usability factors such as authentication time, error rate and memorability. This approach enhances traditional four-digit or six-digit PINs with tactile features using pressure-sensitive touchscreens as found in modern consumer hardware. We refer to these enhanced PINs as *force-PINs* and Figure 1 provides an overview of the proposed scheme.

In theory, force-PINs offer the benefit of a larger PIN space by design. Hence they are more difficult for an attacker to guess and are more resilient to shoulder-surfing attacks due to the invisible pressure component. To estimate the task overhead introduced by this security feature, we present a comparative evaluation of force-PINs and standard four-digit and six-digit PINs as currently deployed in modern smartphones. We conducted a lab study with $n = 50$ participants to compare four-digit force-PINs against four- and six-digit standard PINs and performed a small shoulder-surfing experiment.

We found that entering force-PINs is more time-consuming than entering digit-only PINs. However, we also found that the difference in authentication time between six-digit and force-PINs was not statistically significant. The number of both critical and standard errors were rather low for force-PINs even though the participants from our lab study were using force-PINs for the first time. According to our survey results, the participants liked the invisible pressure component as an additional security feature.

In a small shoulder-surfing experiment, we found that the force component is more difficult for an attacker to observe: none of the force-PINs entered while being observed by an attacker was guessed correctly. However, the attackers were able to guess some of the digit sequences correctly. We also analyzed the user-chosen force patterns alongside with the entered digits and found that users create higher entropy PINs. In an additional field study, we collected evidence on learning effects and showed that authentication time decreases with training.

In summary, our contributions in this paper are:

- We propose an enhancement to digit-only PINs with an invisible force component via pressure-sensitive touchscreens.
- We implemented a prototype of the proposed scheme called *force-PINs*.
- We performed an evaluation of force-PINs, including a lab study with 50 participants, a security evaluation, and a field study with 10 participants.

The remainder of this paper is structured as follows: In Section 2, we discuss related work and in Section 3, we introduce the attacker model, the concept of force-PINs, and describe the objectives of this work. Section 4 presents the design and results of our lab study. In Section 5, we provide a security evalution and in Section 6.3, we present the results of a field study to show learning effects of force-PINs deployed in a real-world environment. Sections 7 and 9 discuss our work and its limitations and we conclude this paper in Section 10.

2. RELATED WORK

Given the importance and the practical impact, it is not surprising that there has been a significant amount of work on authentication schemes. In the following, we briefly review work closely related to our approach. We also refer to the work by Bonneau et al. [8], who presented a benchmark for evaluating authentication schemes.

Malek et al. [24] proposed a haptic-based graphical password scheme. They complement graphical passwords with personal entropies based on pressure and argue that the password space is increased. However, they did not conduct a user study to evaluate usability factors and do not provide empirical evidence that supports the theoretical calculations of a larger password space. Furthermore, they did not evaluate their approach against a shoulder-surfing threat model.

Bianchi et al. [3–6] proposed several authentication approaches based on tactile feedback with an emphasis on accessibility and multi-modal feedback. In comparison to our approach, they rely on a tactile wheel to interact with the system, a component which is not available in off-the-shelf devices.

To make smartphone authentication resilient to shoulder surfers, De Luca et al. [15, 17] presented an authentication mechanism that allows users to enter passwords at the front and the back of their device. While their approach offers benefits with respect to shoulder-surfing resilience, a major limitation of this approach is that there is no such device available at this time that provides users a touch-sensitive back.

Harbach et al. [19] performed a real-world study on smartphone unlocking and found that users spend a significant amount of phone usage time on unlocking their device with PINs and unlock patterns. On average, their study participants unlocked their phones about 47 times throughout the day. This finding shows that mobile device unlocking introduces a severe task overhead and highlights that authentication time is an important factor regarding the usability of the method. It also implies that any time-consuming method is potentially disadvantageous for usability and will therefore have difficulties in getting accepted by users. De Luca et al. [14] found that increased authentication time was a reason for Android users to stop using *Face Unlock* (called *Trusted Face* in later Android versions). Their study also revealed that usability factors are the primary reason keeping users from adopting biometric authentication on mobile devices and that privacy and trust issues only play a secondary role.

A new trending topic in authentication research is implicit authentication. E.g., Buschek et al. [10] studied the feasibility of mobile keystroke biometrics and found that they can be used for user authentication with relatively low error rates. As shown by Khan et al. [22], current methods for implicit authentication are not capable of replacing knowledge-based authentication because their real-world accuracy is significantly lower than in lab settings. Furthermore, they require a certain number of interactions to classify a user correctly. Therefore, these systems are often perceived as disruptive in cases where authentication fails and fallback authentication methods come into play.

3. CONCEPT AND OBJECTIVES

Our approach is based on PIN-based authentication and pressure-sensitive touchscreens as found in modern smartphones (e. g., *3D Touch* available in the iPhone 6s). In the following, we first describe the attacker model and then discuss the design and implementation of *force-PIN*.

3.1 Attacker Model

Throughout the rest of this paper, we assume that the attacker is able to perform a shoulder-surfing attack: she is in close vicinity to the user while authentication takes place and can observe the typing behavior (e.g., in a crowded public or semi-public environment). The key element of a successful shoulder-surfing attack is the ability to clearly observe all sensitive information being entered on the touchscreen.

We also assume that an attacker can gain possession of the user's device. In case the device gets lost or stolen, the design of force-PIN makes a PIN harder to guess due to the theoretically larger PIN space and the pressure component.

3.2 Force-PIN Design

Force-PINs are designed to be more resistant to observation due to the unobtrusive pressure component that helps to obfuscate PIN components and thereby complements regular PIN entry: a user enters a digit either via a shallow or deep pressure on a pressure-sensitive touchscreen. The user receives tactile feedback when entering a digit with deep force. The tactile component and vibration feedback may implicitly help users to memorize force-PINs [9].

An example force-PIN could be 0-**9**-7-**1** where bold and underlined numbers should be pressed more deeply than others on a pressure-sensitive touchscreen (see also Figure 1). The design is not only simple, it is also cheap and easy to deploy as it relies on off-the-shelf hardware. We expect that users who are already using pressure-sensitive touchscreens will find force-PINs as easy to learn as digit-only PINs as they are based on interactions they are already familiar with.

3.3 Implementation

For our study, we implemented a prototype app for iPhones with touch-sensitive screens. The app lets users set a force-PIN and presents a lock screen that looks just like a common lock screen from off-the-shelf iPhones. A force-PIN consists of four digits and a force pattern with two different pressure levels, namely shallow and deep press.

The design decision was based on a small pre-study with 9 participants where we evaluated subjective perceptions on different types of pressure encodings. We evaluated both relative and absolute differences in pressure with different thresholds, respectively. As two-stage pressure with a constant threshold for shallow and deep press performed best; we implemented the prototype app accordingly. We also tested different thresholds and to our surprise it was often not easy to distinguish which threshold was higher and which one was lower. Therefore, we then set the threshold for deep pressure to 50% or more of the maximum possible pressure supported by the hardware.

For our user study, we also implemented apps for four-digit-only and six-digit-only PINs for a comparative lab study and a slightly modified force-PIN app for our field study. The app for the field study had a different main screen and allowed users to submit additional comments to gather in-situ data. Furthermore, the app issued a daily notification to remind the participants of the study task. Each app stored the entered PINs and measured authentication time and failed attempts. The apps with force-PINs also stored the selected four-digit force pattern and arrays of force gradients that were measured for every touch interaction with a pressure-sensitive digit button.

4. LAB STUDY

In the course of a usability lab study, we evaluated force-PINs against digit-only four-digit and six-digit PINs. We chose to evaluate four-digit standard and force-PINs against six-digit standard PINs as they were introduced as the new default in iOS 9. We did not evaluate six-digit force-PINs as we wanted to minimize the additional task overhead. In this section, we describe the methodology and results of this lab study.

4.1 Design and Procedure

Our study is based on a within-subjects design, i.e., every participant is exposed to all conditions. This allows us to perform a comparative evaluation of all subjects exposed to our conditions. We assigned every participant a unique ID and a random order of conditions to reduce learning effects. The three conditions were as follows:

- (C1) four-digit PINs
- (C2) six-digit PINs
- (C3) four-digit force-PINs with shallow and deep pressure

We recruited participants around the university campus over bulletin boards and personal communication mentioning that the study was about their preference of different types of PINs. All of our participants were either employed or currently enrolled as students at the university. We recruited 50 participants for our lab study. They were compensated with a voucher for the university's cafeteria. Table 1 shows the demographics of our participants. All participants were frequent smartphone users and had used digit-only PINs before. To reduce the risk of biased interpretation, we presented the three PIN entry methods equally and did not provide any hints on which method was potentially more secure or not. The participants were not told that the study placed an emphasis on evaluating force-PINs.

The lab sessions proceeded as follows: First, the participants were briefed about the purpose of the study. A subsequent training session allowed them to get familiar with the different types of PINs. This was necessary to minimize the bias introduced by the comparison between a well-known and well-trained authentication method and a newly introduced scheme that users have not yet been exposed to.

Then the participants chose a PIN of the first assigned PIN type and afterwards authenticated with the respective PIN until they had completed three successful authentication sessions. After completing this task, the participant proceeded to the next condition, selected a new PIN and authenticated three times. We instructed the participants to select PINs that they thought were as secure as possible and asked them to remember the PINs just like their own ones in real life. We refrained from assigning PINs as it is a common scenario in the smartphone ecosystem that users

can choose their own PINs. For the same reason, we did not explicitly disallow PIN-reuse.

The metrics we used for our usability evaluation were authentication speed and error rate as defined by De Luca et al. [15]. They defined *authentication speed* as the time between the first touch and the last touch of the authentication session and only counted successful authentication attempts. Regarding the *error rate*, we differentiate between basic and critical errors (as also proposed by De Luca et al. [15]) where basic errors refer to errors within an overall successful authentication session (failed attempts) and critical errors refer to completely failed authentication sessions. Hence, successful authentication sessions may contain failed attempts that influence authentication speed.

In addition to the data collected through our smartphone apps, we gathered quantitative and qualitative data via a questionnaire consisting of 15 closed and open-ended questions to study the perceived security and usability of the three different types of passcodes. The reason why we chose to use open-ended questions was that we wanted to collect meaningful participant statements using their own knowledge, perceptions and interpretations. The questions can be found in Appendix A. After completing the experiments, all participants filled out the questionnaire on a laptop provided by the experimenters.

The participants had to provide their previously assigned experiment ID on the first page of the questionnaire to link the data sets. Except for age, gender and whether the participant had an IT background, no personal data was collected in order to preserve the participants' anonymity. We also collected data on smartphone usage and asked the participants which authentication method they were using at that time on their own smartphones.

The qualitative responses were coded using an iterative coding approach. Two researchers independently went trough the participant responses and produced an initial set of codes. Then, the researchers discussed reoccurring codes, topics and themes, and agreed on a final set of codes. Based on this set, one researcher coded the answer segments for further analysis. As most answers where short and to the point, we did not perform a reliability test of the final coding.

4.2 Results

Given our sample consisting of 50 participants, the quantitative results of our study are based on $3 * 3 * 50 = 450$ authentication sessions (three conditions, every pin type was entered three times by 50 participants). Our study has a repeated-measures design, i.e., every participant was exposed to every condition. Therefore, we analyzed our data with repeated measures ANOVAs. We removed 2 authentication sessions that lasted longer than 30 seconds from the dataset as those occurred when participants where distracted from the study task.

4.2.1 Authentication Overhead

Authentication Speed.

As proposed by De Luca et al. [15], we measured authentication speed from the first to the last touch of a successful authentication session. Hence, an authentication session can also contain a maximum of two failed attempts. After the third failed attempt, the user was locked out of the app.

Table 1: Participant characteristics from the lab study. n=50

Demographic	Number	Percent
Gender		
Male	31	62%
Female	19	38%
Decline to answer	0	0%
Age		
Min.	19	
Max.	56	
Median	25	
IT Background		
Yes	4	8%
No	46	92%
Smartphone		
Android	32	64%
iPhone	14	28%
Windows Phone	2	4%
Other	2	4%
Used Authentication Method		
4-digit PIN	26	52%
6-digit PIN	2	4%
Password (digits/characters)	3	6%
Unlock Pattern	14	28%
Fingerprint Sensor	7	14%
Face Recognition	0	0%
Android Smartlock	1	2%
None	5	10%

The participants had to start the sessions by clicking on a button.

We only considered successful authentication sessions to measure authentication speed. As every user entered every PIN type three times, we calculated the average authentication speed for every user and every authentication method and used this value for further analysis. Overall, 56 force-PINs were selected by our participants. Five of them decided to change their PIN during the experiments, one participant renewed the PIN twice. The participants did not mention any reasons for these decisions. The authentication time was measured based on the most recently selected PIN. Table 2 shows the mean authentication time in seconds and error rate. Figure 2 shows the collected authentication speed measures for all participants and PIN types.

To reveal significant effects regarding authentication speed, we performed a one-way repeated-measures omnibus ANOVA across the 3 PIN types. The results show significant differences in authentication time ($F_{2,147} = 10.19, p < 0.001$). A pairwise t-test with with $t_{0.95,98} = 1.9845$ revealed significant main effects comparing the authentication speed of four-digit with six-digit PINs ($p < 0042$). In addition, authentication speed of four-digit PINs was significantly faster than of force-PINs ($p < 0.001$). The difference in authentication speed between six-digit and force-PINs was not statistically significant ($p = 0.12$).

Errors.

An important factor when estimating the overhead of an authentication method is the number of errors. Similar to

Table 2: Mean authentication time in seconds and error rate with different levels of the independent variables.

Authentication Speed	**Mean**	**SD**
4-digit	2.34	1.21
6-digit	3.33	1.56
Force	3.66	1.96
Error Rate	**Basic**	**Critical**
4-digit	21	0
6-digit	22	0
Force	36	4

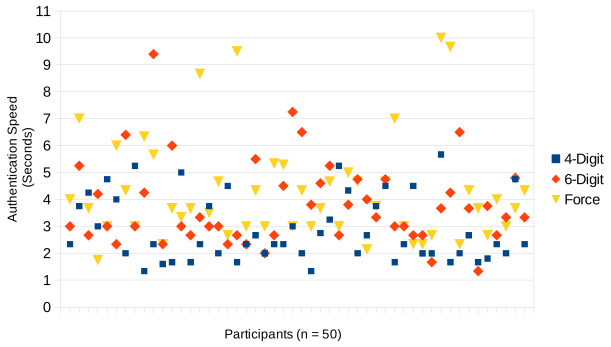


Figure 2: Mean authentication time per participant.

De Luca et al. [15], we distinguished between basic and critical errors. For our authentication scenario, we defined a *basic* error as an erroneous attempt to enter a PIN code. An authentication session can be successful overall, but may take a user two or three times to enter the PIN correctly. We considered an error as *critical* if the entire authentication failed, i.e., a user was locked out after three erroneous attempts as commonly deployed in off-the-shelf smartphone operating systems.

Out of 450 total authentication sessions, four authentication sessions failed (0.9%). All failed sessions involved force-PINs. 36 (8.0%) failed attempts (basic errors) were registered with force-PINs. 22 (4.8%) failed attempts were registered with six-digit PINs and 21 (4.6%) with four-digit PINs.

4.2.2 Perceived Usability and Security

As explained above, participants were asked to fill out a short questionnaire after completing the PIN selection and authentication tasks. In addition to the measurements collected via our iPhone apps, we were interested in participants' perceptions of the three suggested PIN types regarding usability and security. We presented users with closed-ended questions asking which PIN type they thought was the easiest/hardest to remember, fastest/slowest and most/least error-prone to enter and generally most/least secure. The results of these questions are shown in Figure 3.

91% of our participants reported that they thought four-digit PINs were the least secure of the three tested PIN types. 95% also thought that four-digit PINs were the fastest PIN type to enter and 80% thought that they were the easiest to remember. 62% thought that force-PINs were the most secure of the three methods but 55% also thought that this was the most time-consuming PIN type to enter. In

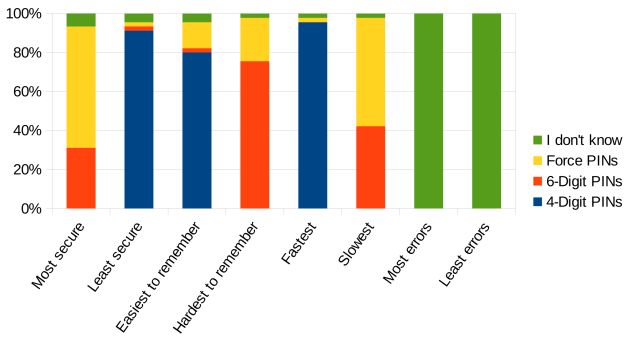


Figure 3: Self-reported usability and security estimation in percent.

comparison, only 31% thought that six-digit PINs were the most secure but 75% also thought that they were the hardest to remember.

To our surprise, all participants chose the "I don't know" option regarding most and least errors when entering any of the suggested PIN types.

On the last page of the online survey, we asked participants three open-ended questions related to their perception of force-PINs. This was the only part of our study where force-PINs received particular attention. These questions were asked at the very end of our lab sessions to minimize the risk of biased interpretation.

After coding the data segments collected through these questions, we found that 38 of the 50 participants thought that a major benefit of force-PINs was the resistance against observation due the haptic and invisible component. 10 participants also stated that they think force patterns are easier to remember than additional digits, as would be the case with longer PINs. Eighteen participants reported that they still think that it requires additional effort to enter digits with different levels of force as they are still not used to this new interaction method with touchscreens.

4.2.3 Informal Participant Statements

In this section, we present informal participant statements and also quote some of the qualitative statements gathered via the open-ended questions from our post-experiment survey. These direct quotes are presented as they were given by the participants prior to coding.

Overall, we were surprised by how easy it was to recruit participants irrespective of the promised reward. We had the impression that all of them found the topic of PIN security important. Based on their comments, we had the impression that most of them seemed to be aware of the richness of private data stored on their smartphones. Most participants also asked for further help in protecting their devices after participating in our study. After their participation, they were given the opportunity to have their questions answered by the experimenters. Even though a few authentication sessions with force-PINs failed, all participants understood the concept of force-PINs and were able to use them. To our surprise, the participants found the concept natural and intuitive even though most of them were using pressure-sensitive touchscreens for the first time.

- *"I like the additional dimension. It is invisible and therefore makes my PIN more secure." (P5)*

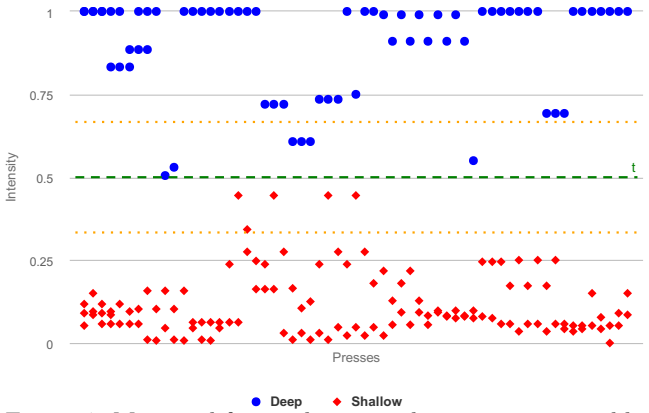


Figure 4: Measured force relative to the maximum possible force. The green line at y = 0.5 represents the threshold for distinguishing between deep and shallow presses. The grey lines at 0.25 and 0.75 indicate two potential thresholds for a three-step force scale (e.g., *shallow-medium-deep*.)

- *"If someone observes me entering my PIN, which is not that secure and probably easy to guess, at least the force component is harder to guess. (P28)"*
- *"I think it might take a while to fully get used to it, as this concept is new to me. (P23)"*
- *"Why not use a six-digit force-PIN? (P12)"*

4.2.4 Force Pressure

As stated in Section 3, we based our design for a two-step scale on our pre-testing with people who had never used *3D Touch* before. Due to the low experience with pressure-sensitive screens, they could not easily distinguish different thresholds to separate deep and shallow press. The app also provided vibration feedback as soon as the user entered a digit with force. Through our lab study, we collected the exact values of the force registered by the device and then used it to evaluate how close or far the registered force was from the threshold and the upper and lower boundaries. Figure 4 shows the force intensities of all logged force-PIN digits during the lab study in percent of the maximum possible force.

5. SECURITY EVALUATION

Based on the data collected during the lab study, we performed an additional security evaluation to evaluate shoulder-surfing resistance and PIN entropy.

5.1 Shoulder Surfing

To evaluate our approach to the attacker model, we performed a small shoulder-surfing experiment in the lab. Similar to the study design of De Luca et al. [15] and von Zeschwitz et al. [26], the attacker tried to shoulder surf the force-PIN entry from the victim. For our evaluation, we considered direct observation, i.e., the attacker was physically standing behind the victim and tried to guess the entered force-PIN and then performed an additional evaluation based on separately recorded video material. Our evaluation is based on the 50 force-PINs which were collected in the course of our lab study and then used for our evaluation of authentication speed and error-rate.

The direct observation attack was performed during the lab study. One experimenter acted as a shoulder surfer and was in close proximity to the victim. Our participants were aware of their entered PINs being tracked via the device used during the experiments but they were not told that one of the experimenters acted as a shoulder surfer. The shoulder-surfing experimenter was perceived as trustworthy. Therefore, the participants did not apply additional measures to prevent their PINs from being observed. We chose this experimental setting as we believe that situations where victims are not aware of being observed are the most dangerous. We furthermore believe that any authentication method should be resilient to direct observation regardless of a specific situation and the user's awareness. In addition, an experimenter entered the collected PINs with their corresponding force patterns while being filmed. Each PIN was entered only once. Another two volunteers, who were university students (one male, one female), then tried to guess the force-PINs based on the recorded material. Each of them tried to guess 25 PINs. They were allowed to re-watch the video sequence up to 5 times if they wanted to.

This first look at shoulder-surfing resistance suggests that force-PINs are capable of making digit PINs more resilient against shoulder-surfing attacks. Out of the 50 entered force-PINs, the shoulder surfer was not able to guess a single one completely. However, 21 out of 50 PINs were partially guessed (i.e., the attacker correctly guessed the digits but not the force pattern). Similar to the direct observation attacks, the attackers in the camera-based attacks were not able to completely guess the force-PINs from the recorded material, but managed to guess 39 of the shown digit sequences correctly. We did not evaluate whether individual digits (with or without force) were guessed correctly.

5.2 Entropy

In theory, the PIN space of four-digit force-PINs is larger than for standard four-digit and smaller than six-digit PINs. In our lab study, we used user-assigned PINs. We gave participants a password policy, namely to choose a PIN that, in their opinion, is as secure yet as memorable as possible and where at least one digit within the four digit pattern is entered with a deep press.

Obviously, the number of possible combinations is $10^4 = 10,000$ for four digit passwords and $10^6 = 1,000,000$ for six digit passwords. Force-PINs augment the four-digit password space to $20^4 = 160,000$ possible PIN codes including four-digit PINs with all digits entered with shallow pressure. As we defined a policy for the lab study which forced participants to choose at least one digit with deep pressure, the password space decreases to $150,000$.

As done by Cherapau et al. [12], we calculate the zero-order entropy, which is a theoretical measure of the entire search space of all possible secrets of a given length and the size of a given alphabet assuming that each character is selected randomly. Zero-order entropy is measured in bits and calculated as $L * \log_2 N$, where L is the length of the secret and N the size of the character set. Hence, for force-PINs, the length is 4 and the character set 20. Thus, the zero-order entropy for force-PINs is 17.28 bits, while four-digit PINs have a zero-order entropy of 13.28 [12] and six-digit PINs 19.93 bits. These theoretical measures are upper bounds for real-world entropy.

In theory, the augmented PIN space is a major improvement compared to standard four-digit PINs. In practice however, users often do not fully exploit this benefit but se-

lect PIN codes and passwords from a much smaller subset that are often easy to predict [21]. Therefore, the search space is smaller and the PIN is therefore easier for an attacker to guess. We therefore evaluate the distribution of force patterns and digit-pressure combinations.

Table 3 shows the occurrences of force patterns selected by our participants. Our results suggest that more than half of our participants selected a force pattern where only a single digit is entered with deep press. In our sample, the most popular positions in the digit sequence were the first and second one with a probability of 14.0%. Even though this trend indicates that our participants did not fully make use of the theoretically larger PIN space and therefore create lower entropy PINs in practice, this is already an improvement over standard four-digit PINs. Our dataset of 56 PINs is relatively small and therefore not sufficient to determine the practical entropy of force-PINs. To provide a rough indicator, we calculate the entropy of the binary force component based on the force-PINs chosen by our study participants. Furthermore, to estimate the entropy gain over digit-only PINs, we compare our results to those from a related study on iPhone passcodes with a larger sample size. In theory, if force patterns were evenly distributed, the theoretical entropy gain would be 4 bits. We calculate the practical entropy gain as $-\sum_{i=1}^{n} p_i * log_2(p_i)$ where p_i is the probability of a certain pattern occurring. Based on our observed probabilities from 56 user-chosen force patterns (as presented in Table 3), the practical entropy gain is 3.41 bits. Bonneau et al. [7] calculated the entropy of four-digit PINs from iPhone users as 11.42 based on a dataset of 204,508 PINs. Comparing our findings with Bonneau et al. [7], an additional binary force component provides an entropy gain of approximately 23% to digit-only PINs of length 4.

Table 3: Force patterns selected by the lab study participants where S = shallow press, D = deep press. n = 56 user-selected PINs. The table is sorted in descending order. The pattern SSSS was excluded as the PIN selection policy required participants to enter at least one digit with deep press.

Force Pattern	Number	Percent
DSSS	8	14.0%
S**D**SS	8	14.0%
SSS**D**	7	12.2%
SS**D**S	6	10.5%
DSS**D**	6	10.5%
S**DD**S	5	8.7%
DDDD	5	8.7%
SS**DD**	4	7.0%
S**DDD**	2	3.5%
S**D**S**D**	2	3.5%
DDSS	1	1.7%
DS**D**S	1	1.7%
DDS**D**	1	1.7%
DDDS	0	0.0%
DS**DD**	0	0.0%

6. FIELD STUDY

In addition to the lab study, we conducted a field study to show that authentication time for four-digit force-PINs decreases with training. The latter is an important metric when comparing the usability performance of digit-only PINs with force-PINs as we assume that users will initially perform better with digit-only PINs as they are already trained to use them.

Table 4: Digits and their occurrence entered with either shallow or deep press. Deep pressed digits are in bold; sorted in descending order.

Digit (shallow/deep press)	Number
1 (shallow)	27
0 (shallow)	22
5 (shallow)	16
4 (shallow)	15
3 (shallow)	14
2 (shallow)	12
0 (deep)	12
1 (deep)	12
6 (shallow)	11
2 (deep)	10
9 (deep)	10
3 (deep)	9
6 (deep)	9
9 (shallow)	8
4 (deep)	7
7 (shallow)	6
7 (deep)	6
8 (deep)	6
5 (deep)	6
8 (shallow)	5

6.1 Study Design and Procedure

We recruited 10 participants and deployed an iOS app on their personal devices and asked them to enter as many force-PINs as possible (we required a minimum of 300 successful authentication sessions) over a period of two weeks. At the end of this period, we conducted short debriefing interviews with the participants. In contrast to the lab study, the participants were aware that the focus of the study was to evaluate force-PINs.

Due to the low propagation of compatible iPhones in our region, we were able to recruit only 10 participants. In spite of the relatively low number of participants, we still believe that the gathered data provides useful insights and rough indicators on learning effects. Furthermore, deploying force-PINs in a real-world environment helped us to gather in-situ reactions on authentication problems with force-PINs.

We based our study design on findings from Harbach et al. [19], who found that users unlock their phone on average 47.8 times a day (about three unlocks per hour assuming a user is awake for 16 hours per day).

Due to the restrictions in iOS, we were not able to replace the actual PIN scheme on the participants' devices with force-PINs. We also had to reject our plan to issue notifications based on the participants' unlocking behavior as iOS does not offer to activate third-party apps after an unlock event. Therefore, we were not able to collect the respective data from the users' own devices. As everyday routines and smartphone usage habits are highly diverse, we refrained from requiring force PIN entries at fixed timepoints throughout the day and opted for a more realistic and

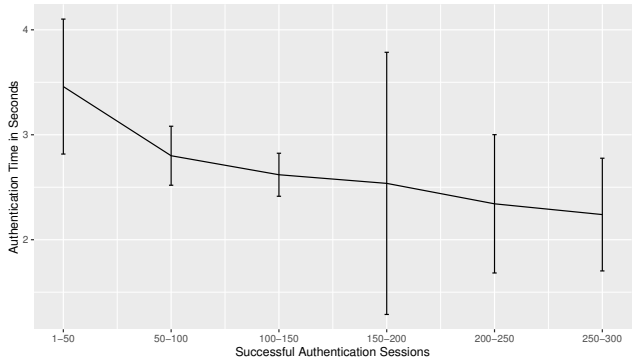


Figure 5: Authentication time development based on the first 300 successful authentication sessions across all participants.

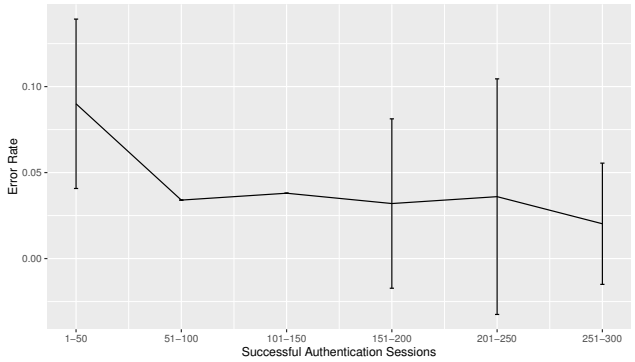


Figure 6: Error rate development (basic errors) based on the first 300 successful authentication sessions across all participants.

less disruptive setting. To evaluate different timing options for notifications, we conducted a small pilot study with different notification patterns. The participants from this pilot study perceived the notifications as disruptive and annoying regardless of whether they were issued at fixed or adaptive time points. Based on the participants' responses, we decided to reduce the number of daily notifications to a single daily reminder at an arbitrary point in time and left it up to the participants when and how often to entered their force-PINs. We are confident that this study design reflects realistic usage habits and reduced the risk of participants dropping out early from the study.

We instructed our participants to enter force-PINs whenever they took out their phone before or after their primary task. We suggested they distribute the PIN entries over the given period of time (i.e., about 20 PINs a day), but also told them that it was their own decision when exactly and how often to enter them. The participants were also instructed to choose as secure and memorable PINs as possible with at least one digit entered with force.

The main screen of our app had a button that redirected the participants to a lock screen to start an authentication session with a force-PIN. It was designed to look exactly like the standard iPhone lock screen. Our app also displayed a counter of successful authentication sessions and provided users with two extra buttons, one to send us an e-mail in case of questions and another one to leave a comment to a situation. We also provided users with an option on the main screen to set a new force-PIN. Upon clicking on this button, a password-forgotten event was logged and the participants were able to set a new force-PIN.

6.2 Results

Overall, our participants successfully completed 3,748 authentication sessions with force-PINs. The results are summarized in Table 5. Among the successful sessions, 254 failed attempts (basic errors) were registered and five participants had entirely failed authentication sessions (critical errors). The number of critical errors (i.e., failed authentication sessions) was low. The entirely failed authentication sessions were registered at the very beginning of the study. The error rates in Table 5 are given in percent of authentication sessions completed by the user. For the quantitative analysis, we removed authentication sessions that lasted longer than 30 seconds from our sample. As observed in our lab study, authentication sessions longer than 30 seconds usually occurred when the participant was interrupted or distracted from the study task.

The mean authentication speed over all authentication sessions was 2.69 seconds (median=2.26, SD=0.59), which is an improvement over the results from the lab study. The shortest authentication session was only 1.02 seconds long. In comparison, Harbach et al. [18] determined the average authentication speed for digit-only PINs as 1.9 seconds.

All participants attended the debriefing session and participated in the debriefing interviews. One participant did not complete the initially requested 300 successful authentication sessions and had only 210 completed authentication sessions. Although this did not meet our desired goal, we included the data and conducted the debriefing interview with the participant as the number of participants was low.

Just like in the lab study, we measured the authentication time of each session as time from the first touch until the user was successfully authenticated (including potentially unsuccessful attempts made during the session). As per the study design, we expected the PIN entries to be unevenly distributed over time across the participants. Our results show that the participants did not make use of the given time and completed the study task in a few days regardless of our daily notifications. Five participants completed their authentication sessions on a single day. They distributed their PIN entries over the morning, late afternoon and the evening of that day. Four participants completed the study task in two or three days and entered their PINs mostly in the morning and late afternoon/evening of these days. One participant spent four days on the study task and distributed the PIN entries over various times of the day. We therefore refrain from a time-based analysis and compare the results based on authentication sessions.

For our analysis of authentication time and error rate, we consider the first 300 successfully completed authentication sessions from all participants. In order to visualize a trend over multiple completed authentication sessions, we grouped the results in bins of 50 sessions across all participants. We selected a bin size of 50 to approximate the average number of phone unlocks per day as determined by Harbach et al. [19]. We believe that this is a good way to simulate a trend over a reasonable period of time. Figure 5 provides a comparison of the average authentication time grouped by 50 successful authentication sessions based on the median

Table 5: Summary of field study results. n=10

Subjects	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10
Completed Authentication Sessions	534	336	453	387	407	335	210	386	343	357
Basic Errors	13	41	69	20	4	26	16	17	21	27
Basic Error Rate	2.4 %	12.2 %	15.2 %	5.2 %	0.9 %	7.7 %	7.6 %	4.4 %	6.1 %	7.6 %
Critical Errors	0	3	0	0	0	1	1	1	1	0
Critical Error Rate	0 %	0.8 %	0 %	0 %	0 %	0.2 %	0.4 %	0.2 %	0.2 %	0 %
Forgot Force-Pin	0	2	1	0	0	0	0	0	0	0
Force-Pins	**5225**	**0229**	1234	**5795**	**5968**	**0000**	1**70**3	**0171**	22**0**4	9**9**99
	-	0**229**	741**2**	-	-	-	-		-	-
	-	**1397**	-	-	-	-	-	-	-	-

authentication time per participant. These results suggest that the authentication time decreases with training. Figure 6 shows that the error rate also decreases with training.

6.3 Debriefing Interviews

During the debriefing sessions, we asked the participants in which situations they used force-PINs and whether they found them feasible in these scenarios. According to the participants, most force-PINs where entered either while they were at home, in their office, or on public transport. Eight participants reported that they found force-PINs a good way to protect their digit PINs from shoulder surfers even though they estimated their susceptibility towards direct observers as relatively low. Three participants said that they would like to use force-PINs to make their existing PINs more secure against close intruders such as family and friends who could easily guess their PIN as it was an important date. According to them, the risk of a close acquaintance spying on their phones was higher than that of shoulder surfing attacks in public spaces.

Nine participants reported that their perceived authentication time decreased with training when they used it several times a day. However, five of them reported that they still think that simple digit PINs are faster for authentication. All participants reported that they did not find force-PINs harder to remember than simple digit PINs.

Participants were also asked if they would prefer to use force-PINs over simple digit PINs. All of them said that they generally liked the idea of an additional invisible component and six participants said that they would maybe use them if deployed on their device. Eight participants reported that they found the training phase in the beginning annoying. Three expressed interest in multiple-step pressure difference.

7. DISCUSSION

Previous research [19] has shown that the task overhead of smartphone authentication is relatively high. Therefore, we argue that the overhead of a technology to replace simple digit PINs should not be higher than the state of the art.

The results from our lab study suggest that the task overhead of force-PINs is initially higher than for digit-only four- and six-digit PINs. Our security analysis and the participants' responses indicate that force-PINs can increase PIN entropy and improve the resilience towards shoulder-surfing attacks. The results from our field study revealed learning effects after a certain number of interactions with the invisible component, and indicate that authentication time and error rate decrease with training and converge towards the metrics for four-digit PINs.

We collected evidence on frequently used force patterns and determined a practical entropy gain of 3.41 bits based on the force-PINs chosen by our study participants. Similar to other user-chosen secrets, the practical entropy does not meet the theoretical measures but still suggests a major improvement when compared to entropy estimations of digit-only PINs.

Apart from the metrics we used to evaluate the performance of the respective PIN types, the self-reported data from our participants suggests that force-PINs were perceived as more secure than six-digit PINs. The open-ended survey questions revealed that this was mainly due to the force component, which our participants perceived as a good countermeasure against observation.

Only two participants forgot and renewed their force-PINs from the field study. The number of critical errors was also low.

Hence, our results suggest that our scheme is able to improve security with a reasonably low impact on task overhead. In comparison to other solutions, our design improves security without requiring the user to memorize longer sequences of digits, which have been shown to be more difficult to remember [20].

To our surprise, none of the 50 participants provided an estimation of which of the PIN schemes was most/least error prone. While our collected data does not explain reasons, we believe that this is because of the manifold sources of errors: As authentication sessions in the wild usually take place in diverse situations, their successful completion is influenced by environmental and situational constraints beyond the design of the authentication method.

According to a study by Harbach et al. [19], users are generally aware of risky situations but this does not influence their general opinion about this threat, which is that this risk is only considered in a low number of everyday situations. However, just because users do not perceive situations as risky does not mean that they are not. Hence, physically shielding the PIN from an observer can only mitigate an attack if the user is aware of the threat and therefore actively taking precautions. Our results suggest that force-PINs can help to protect users from shoulder surfers regardless of their risk awareness, while minimizing the additional effort the user has to invest.

Modern smartphones offer biometric authentication as an alternative. While supporters of these methods often argue that they are harder to replicate and therefore not susceptible to shoulder surfing, it is commonly acknowledged by the scientific community that these methods are non-revocable and can easily be broken [1, 11]. Furthermore, they still rely on passwords for fallback authentication. These examples highlight that it is worth putting effort into making knowledge-based authentication resilient to shoulder surfing.

Our prototype app was implemented for iPhone 6s. Other smartphone models, such as the Huawei Mate S, also have pressure-sensitive screens and are therefore suitable for force-PINs. Furthermore, force patterns like in force-PINs can also be added to character/digit passwords with variable length and Android unlock patterns to make them resilient to shoulder surfing attacks. As future work and as soon as a compatible API and device are available in our region, we plan to evaluate force patterns in combination with unlock patterns and other alternative authentication schemes, respectively.

8. LIMITATIONS

We now discuss limitations of our methodology and the conducted studies.

As we recruited our participants at the university campus, the level of education and technology affinity among our sample were higher than expected from the general population. As the results might differ for other demographics, our results cannot be generalized to the entire population of smartphone users. Since only 28% of the participants in the lab study were iPhone users, we cannot determine whether the measurements based on their input were biased by the lack of practice. However, as this study had a repeated-measures design, we were able to perform a comparative evaluation of all subjects exposed to our conditions. All participants in our field study took part with their own devices and had therefore been exposed to a force-sensitive screen before and were already familiar with the iOS user interface and lock screen, respectively.

It is possible that users would improve even more over a longer period of time and usability metrics converge to those of four-digit standard PINs. Regardless of our suggestion to distribute the authentication sessions over the two weeks, the participants tried to complete the study task as fast as possible and therefore entered all force-PINs within the first three days. Also, the number of successful authentication sessions varies widely across the participants. As the participants did not spread out the PIN entries over the given time, we can neither perform a time-based evaluation nor seriously evaluate memorability. The fact that our participants from the lab study thought that force-PINs are more memorable speaks for the system but does not obviate the need for a future long-term evaluation. Regardless of these limitations, we are confident that our study design reflects real-world usage behavior and due to its flexibility ensured that participants would not drop out early.

A major limitation of this work is that the participants from both the field study and the shoulder surfing experiment participated voluntarily and did not receive a compensation for their participation. Therefore, the motivation for the shoulder surfers was rather low to actually break the system. Another limitation is that they were new to the concept of force-PINs and therefore perceived the task as particularly challenging. Also, force-PINs do not provide visual feedback and the vibration for digits entered with force is very subtle and therefore not audible on the video material. The participants reported finding it hard to focus on both the digits and the force patterns. The person who entered the PINs in front of the camera was a faculty member who was aware of the hypothesis being tested just like the experimenter who tried to shoulder surf the PINs from the lab study. These limitations imply that further investigation is needed to determine a lower bound for shoulder surfing resistance.

9. ETHICAL CONSIDERATIONS

Our university does not have an ethics board but has a set of guidelines that we followed in our research. A fundamental requirement of these guidelines is to preserve the participants' privacy and to limit the collection of person-related data as far as possible. For both our studies, we did not collect any personally identifiable information, except for age and gender. A major ethical challenge was the collection of PINs. The PINs were chosen by the participants and they were aware that the PINs they selected were being collected. However, we cannot preclude that those were real PINs. Keeping this data confidential and making it impossible to map a physical person with a certain PIN was therefore our primary concern. In similar shoulder surfing studies, participants were re-corded with video cameras to perform attacks based on the recorded material. Although this was our initially planned study setting, we decided not to film the participants directly while they entered their PINs. This decision was made based on the results and feedback from our pilot study, where our participants expressed discomfort about being filmed while entering information as sensitive as a PIN. We therefore chose to let a separate person enter all force-PINs in front of a camera and then used the resulting material for our camera attacks.

10. CONCLUSION

In this work, we proposed integrating pressure-sensitive touchscreen interactions into knowledge-based authentication. These force-PINs enhance digit-only PINs with a force pattern, i. e., an additional pressure-sensitive component that allows users to select higher entropy PINs that are harder for a shoulder surfer to observe.

We were able to collect evidence on the security benefits of force-PINs and their impact on usability. We conducted a lab study with 50 participants and showed that authentication speed of force-PINs is not significantly slower than that of six-digit standard PINs, but still significantly slower than that of 4-digit standard PINs. We also showed that the error rate is rather low in spite of the fact that most participants had not yet been exposed to pressure-sensitive touchscreen interaction. Furthermore, we conducted a small shoulder-surfing study where an attacker tried to observe and guess force-PINs. The attackers were not able to guess a full force-PIN consisting of a digit sequence and a force component. These results suggest that force-PINs can help to mitigate shoulder-surfing attacks in public spaces that are potentially noisy and crowded. In a security evaluation of the collected force-PINs, we showed that the practical entropy is still higher than for standard four-digit PINs although users do not make full use of the larger PIN space. In

an additional field study with 10 participants, we deployed force-PINs in the wild and showed that users improve after being exposed to the technology over a longer period of time.

Our results imply that small enhancements such as an additional pressure component allow users to select higher entropy PINs that are more resilient to shoulder-surfing attacks, while keeping the impact on usability metrics such as authentication speed and error rate low. This is important as users enter their PINs multiple times a day and therefore require methods that do not increase the task overhead.

11. ACKNOWLEDGMENTS

We would like to thank the reviewers for their constructive feedback. We would also like to thank our shepherd Marian Harbach for his suggestions that were very helpful in improving our paper. This research was partially funded by COMET K1, FFG - Austrian Research Promotion Agency.

12. REFERENCES

[1] Android Authority. Android Jelly Bean Face Unlock 'liveness' check easily hacked with photo editing. http://www.androidauthority.com/android-jelly-bean-face-unlock-blink-hacking-105556/, last accessed 2/10/2016.

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

[3] Bianchi, Andrea and Oakley, Ian and Kostakos, Vassilis and Kwon, Dong Soo. The phone lock: audio and haptic shoulder-surfing resistant PIN entry methods for mobile devices. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*, pages 197–200. ACM, 2011.

[4] Bianchi, Andrea and Oakley, Ian and Kwon, Dong Soo. Spinlock: a single-cue haptic and audio PIN input technique for authentication. In *Haptic and Audio Interaction Design*, pages 81–90. Springer, 2011.

[5] Bianchi, Andrea and Oakley, Ian and Kwon, Dong Soo. Counting clicks and beeps: Exploring numerosity based haptic and audio PIN entry. *Interacting with computers*, 24(5):409–422, 2012.

[6] Bianchi, Andrea and Oakley, Ian and Lee, Jong Keun and Kwon, Dong Soo and Kostakos, Vassilis. Haptics for tangible interaction: a vibro-tactile prototype. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*, pages 283–284. ACM, 2011.

[7] J. Bonneau, S. Preibusch, and R. Anderson. A birthday present every eleven wallets? the security of customer-chosen banking pins. In *Financial Cryptography and Data Security*, pages 25–40. Springer, 2012.

[8] Bonneau, Joseph and Herley, Cormac and Van Oorschot, Paul C and Stajano, Frank. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *Security and Privacy (SP), 2012 IEEE Symposium on*, pages 553–567. IEEE, 2012.

[9] A. Bragdon, E. Nelson, Y. Li, and K. Hinckley. Experimental analysis of touch-screen gesture designs in mobile environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 403–412. ACM, 2011.

[10] Buschek, Daniel and De Luca, Alexander and Alt, Florian. Improving Accuracy, Applicability and Usability of Keystroke Biometrics on Mobile Touchscreen Devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1393–1402. ACM, 2015.

[11] Chaos Computer Club. Chaos Computer Club breaks Apple TouchID. http://www.ccc.de/en/updates/2013/ccc-breaks-apple-touchid, last accessed 11/11/2015.

[12] Cherapau, Ivan and Muslukhov, Ildar and Asanka, Nalin and Beznosov, Konstantin. On the impact of touch id on iphone passcodes. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*, pages 257–276, 2015.

[13] S. Chowdhury, R. Poet, and L. Mackenzie. Passhint: Memorable and secure authentication. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems*, CHI '14, pages 2917–2926, New York, NY, USA, 2014. ACM.

[14] De Luca, Alexander and Hang, Alina and von Zezschwitz, Emanuel and Hussmann, Heinrich. I feel like i'm taking selfies all day!: Towards understanding biometric authentication on smartphones. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI*, volume 15, pages 1411–1414, 2015.

[15] De Luca, Alexander and Harbach, Marian and von Zezschwitz, Emanuel and Maurer, Max-Emanuel and Slawik, Bernhard Ewald and Hussmann, Heinrich and Smith, Matthew. Now you see me, now you don't: protecting smartphone authentication from shoulder surfers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2937–2946. ACM, 2014.

[16] De Luca, Alexander and Lindqvist, Janne. Is Secure and Usable Smartphone Authentication Asking Too Much? *Computer*, 48(5):64–68, 2015.

[17] De Luca, Alexander and Von Zezschwitz, Emanuel and Nguyen, Ngo Dieu Huong and Maurer, Max-Emanuel and Rubegni, Elisa and Scipioni, Marcello Paolo and Langheinrich, Marc. Back-of-device authentication on smartphones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2389–2398. ACM, 2013.

[18] M. Harbach, A. De Luca, and S. Egelman. The Anatomy of Smartphone Unlocking. In *Proceedings of the 34th Annual ACM Conference on Human Factors in Computing Systems, CHI*. 2016.

[19] Harbach, Marian and von Zezschwitz, Emanuel and Fichtner, Andreas and De Luca, Alexander and Smith, Matthew. It's a hard lock life: A field study of smartphone (un) locking behavior and risk perception. In *Symposium on Usable Privacy and Security (SOUPS)*, 2014.

[20] Huh, Jun Ho and Kim, Hyoungshick and Bobba, Rakesh B and Bashir, Masooda N and Beznosov, Konstantin. On the Memorability of System-generated PINs: Can Chunking Help? In *Eleventh Symposium*

On Usable Privacy and Security (SOUPS 2015), pages 197–209, 2015.

[21] P. G. Kelley, S. Komanduri, M. L. Mazurek, R. Shay, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, and J. Lopez. Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms. In *Security and Privacy (SP), 2012 IEEE Symposium on*, pages 523–537, 2012.

[22] Khan, Hassan and Atwater, Aaron and Hengartner, Urs. A comparative evaluation of implicit authentication schemes. In *Research in Attacks, Intrusions and Defenses*, pages 255–275. Springer, 2014.

[23] R. Kuber and W. Yu. Tactile vs graphical authentication. In *Haptics: Generating and Perceiving Tangible Sensations*, pages 314–319. Springer, 2010.

[24] Malek, Behzad and Orozco, Mauricio and El Saddik, Abdulmotaleb. Novel shoulder-surfing resistant haptic-based graphical password. In *Proc. EuroHaptics*, volume 6, 2006.

[25] Song, Youngbae and Cho, Geumhwan and Oh, Seongyeol and Kim, Hyoungshick and Huh, Jun Ho. 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*, pages 2343–2352. ACM, 2015.

[26] von Zezschwitz, Emanuel and De Luca, Alexander and Brunkow, Bruno and Hussmann, Heinrich. SwiPIN: Fast and secure pin-entry on smartphones. In *Proceedings of the Conference on Human Factors in Computing Systems, CHI*, volume 15, pages 1403–1406, 2015.

APPENDIX

A. LAB STUDY QUESTIONNAIRE

The following questions were answered by the participants of the lab study after they used the three different types of PINs in a randomized order (four-digit/six-digit/force-PIN).

Demographics.

1. What was your ID during the lab experiments?
2. Gender
3. Age
4. Are you studying IT security or are you working in an IT security-related field? *(yes/no)*
5. What kind of smartphone are you currently using? *(single-choice: iPhone, Android, Windows Phone, Other, I don't use a smartphone)*
6. What methods are you currently using to unlock your smartphone? *(multiple-choice: 4-digit PINs, 6-digit PINs, character and digit password, unlock pattern, fingerprint sensor, Android Smartlock, none)*

Estimated security and usability of the three PIN types.

1. Which of the three PIN methods do you think is the most secure? *(single-choice: 4-digit PINs, 6-digit PINs, force-PINs, I don't know)*
2. Which of the three PIN methods do you think is the easiest to remember? *(single-choice: 4-digit PINs, 6-digit PINs, force-PINs, I don't know)*
3. Which of the three PIN methods do you think is the least secure? *(single-choice: 4-digit PINs, 6-digit PINs, force-PINs, I don't know)*
4. Which of the three PIN methods do you think is the most time-consuming? *(single-choice: 4-digit PINs, 6-digit PINs, force-PINs, I don't know)*
5. Which of the three PIN methods do you think is the hardest to remember? *(single-choice: 4-digit PINs, 6-digit PINs, force-PINs, I don't know)*
6. Which of the three PIN methods do you think is the least time-consuming? *(single-choice: 4-digit PINs, 6-digit PINs, force-PINs, I don't know)*

Open-ended questions.

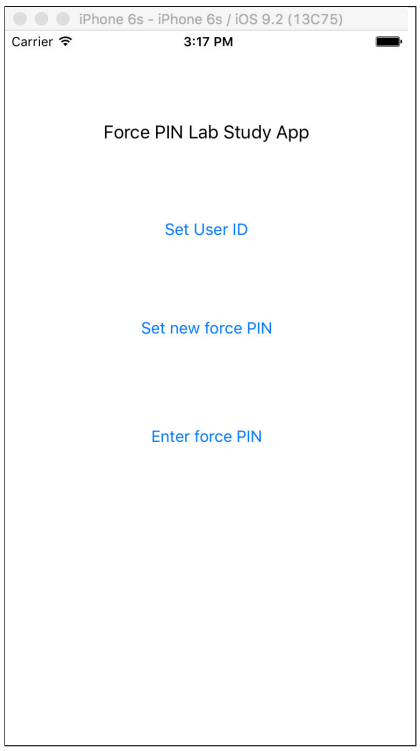
1. What did you like about force-PINs?
2. What did you NOT like about force-PINs?
3. Can you think of a situation where force-PINs would be particularly useful?

B. FIELD STUDY DEBRIEFING INTERVIEWS

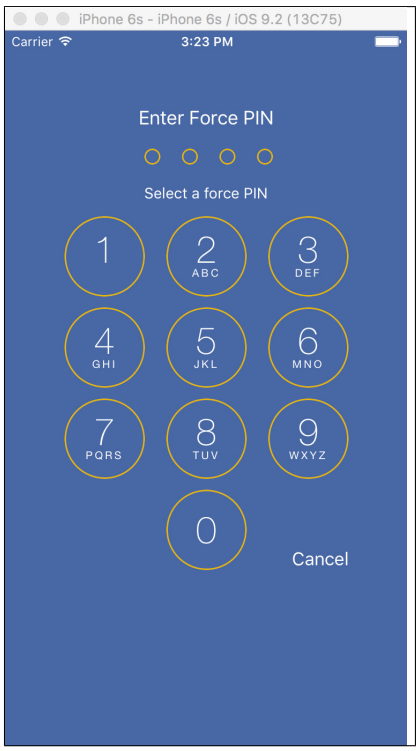
1. Where did you use force-PINs?
2. What did you like about force-PINs?
3. What did you NOT like about force-PINs?
4. Can you think of a situation where force-PINs were particularly useful?
5. Can you think of a situation where force-PINs were annoying?
6. Is there anything else you would like to let us know?

C. STUDY APPS

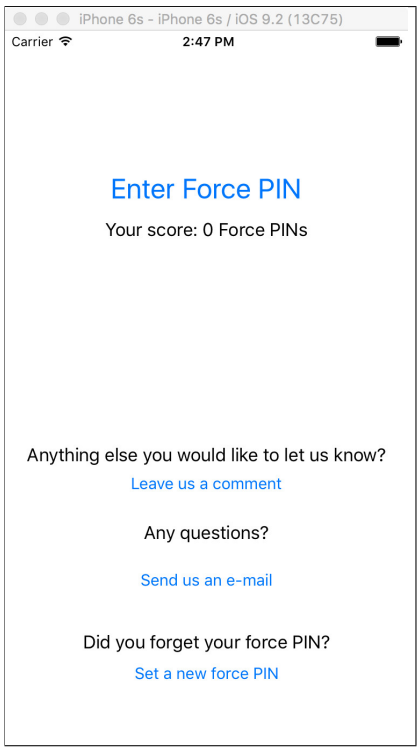
The following screenshots show the user interface of the apps used for the lab and field study. Figure 7a and Figure 7b were used to evaluate force-PINs in the lab study. The apps for the other two conditions had the same layout but evaluated four-digit and six-digit PINs, respectively. Figure 7c shows the main screen of the app used in the field study.



(a) Main screen of the lab study app.



(b) Lock screen of both the lab study and field study app.



(c) Main screen of the field study app.

Figure 7: Screenshots of the study force-PIN apps.