Towards Baselines for Shoulder Surfing on Mobile Authentication

Adam J. Aviv
United States Naval Academy
aviv@usna.edu

John T. Davin
United States Naval Academy
john.t.davin@gmail.com

Flynn Wolf
University of Maryland, Baltimore County
flynn.wolf@umbc.edu

Ravi Kuber
University of Maryland, Baltimore County
rkuber@umbc.edu

ABSTRACT

Given the nature of mobile devices and unlock procedures, unlock authentication is a prime target for credential leaking via shoulder surfing, a form of an observation attack. While the research community has investigated solutions to minimize or prevent the threat of shoulder surfing, our understanding of how the attack performs on current systems is less well studied. In this paper, we describe a large online experiment ($n = 1173$) that works towards establishing a baseline of shoulder surfing vulnerability for current unlock authentication systems. Using controlled video recordings of a victim entering in a set of 4- and 6-length PINs and Android unlock patterns on different phones from different angles, we asked participants to act as attackers, trying to determine the authentication input based on the observation. We find that 6-digit PINs are the most elusive attacking surface where a single observation leads to just 10.8% successful attacks (26.5% with multiple observations). As a comparison, 6-length Android patterns, with one observation, were found to have an attack rate of 64.2% (79.9% with multiple observations). Removing feedback lines for patterns improves security to 35.3% (52.1% with multiple observations). This evidence, as well as other results related to hand position, phone size, and observation angle, suggests the best and worst case scenarios related to shoulder surfing vulnerability which can both help inform users to improve their security choices, as well as establish baselines for researchers.

CCS CONCEPTS

•**Security and privacy → Graphical / visual passwords; Social aspects of security and privacy;**

KEYWORDS

Shoulder surfing; mobile security; password security; usable security; graphical passwords; PIN passwords; mobile authentication.

1 INTRODUCTION

Personal and sensitive data is often stored on or accessed via mobile devices, making these technologies an attractive target for attackers [17]. In the physical domain, the first line of defense against a proximate attacker seeking to gain access to the device is the unlock authenticator, the method used to authenticate the device owner to the device, e.g., by entering a 4-digit PIN.

One type of attack faced when authenticating via a mobile device is shoulder surfing, a form of an observation attack by which an attacker attempts to observe the authenticator of a victim while the authenticator is being entered on the device [43]. One of the most cited dangers for smartphone unlocking mechanisms are shoulder surfing attacks [28].

While many users utilize biometric authentication as a supplement to the dominant PIN and graphical (stroke-based) pattern password entry mechanisms, this does not provide universal protection from shoulder surfing. Biometrics are a promising advancement in mobile authentication, but they can be considered a reauthenticator or a secondary-authentication device as a user *is still required to have a PIN or pattern* that they enter rather frequently due to environmental impacts (e.g., wet hands). There are also known to be high false negatives rates associated with biometrics [7]. Further, users with biometrics often choose weaker PINs as compared to those without [10], suggesting that the classical unlock authentication remain an important attack vector going forward.

There is much related work that both proposes and studies shoulder surfing resistant authentication mechanisms [11–14, 17, 19, 20, 25, 28], but *research related to understanding the susceptibility to shoulder surfing of currently used unlock authentication, namely PINs and Android graphical pattern unlock, is limited in nature.* Further, as researchers propose methods and authentication schemes that offer protections from shoulder surfing attacks, we do not have clear baselines of comparison for improvement (or lack thereof) to current schemes.

In this paper, we report the results of a comprehensive study of shoulder surfing based on video recordings of a victim authenticating. Our participants, upon viewing the videos, were asked to recreate the authentication sequences, simulating basic shoulder

surfing attacks. While prior work has considered visual observations of Android graphical passwords, such as smudge attacks [6] and animated tracing [39], prior research only considered a single dimension. We attempt to account for multiple conditions.

- Authentication Type: we compared 4- and 6-digit PINs, and 4- and 6-length Android graphical patterns, with visible line feedback and without.
- Observation Angle: we considered 5 different observation angles based on videos recorded simultaneously during authentication.
- Repeated Viewing: we consider situations where the participant has a single view of authentication or multiple views.
- Phone Size: we consider two different touchscreen sizes that are common in today's market.
- Hand Position: we considered two different hand positions to interact with the device, single handed thumb input, and two handed index-finger input.

We constructed a comprehensive web-based survey and recruited participants locally from our institution ($n = 91$) and online via Amazon Mechanical Turk ($n = 1173$) for a mixed-factorial subject study. Participants, acting as attackers, were presented with a set of randomized conditions and asked to view a video of an authentication. They then attempted to recreate the authentication.

Analyzing the results, we find that in all settings, Android's graphical pattern unlock is the most vulnerable, especially when feedback lines are visible; a single observation successfully attacked the pattern 64.2% of the time with 79.9% for multiple observations of a 6-length pattern. Shorter patterns were even more vulnerable. Removing feedback lines during the pattern entry improved the security, finding 35.3% successful attacks with a single view and 52.1% success with multiple views for 6-length patterns (confirming prior work [39]). PINs, however, proved much more elusive to attack than anticipated. A single observation was sufficient to attack just 10.8% of the 6-digit PINs, degrading to 26.5% after two observations.

These results support what we as a community have believed to be true anecdotally, and further demonstrates that current authentication methods provide stronger security against shoulder surfing than one might expect. Further, these results suggest that baselines of shoulder surfing success can be applied to this space, to better support mobile authentication users. Future work should allow for improvements over the current worst settings and best settings for shoulder surfing.

2 BACKGROUND AND RELATED WORK

Mobile Authentication Unlock Choices. In order to secure access to a mobile device, users are able to use three main mechanisms to unlock the screen.

- PIN based authentication, sometimes referred to as a passcode on iPhones: where a user is asked to recall a PIN of a least four digits. Newer iPhones, however, require a 6-digit passcode [18]. (A sample PIN layout, as used in our experiments, is shown in Figure 1.)
- Pattern based authentication: where a user is asked to recall a gesture that interconnects a set of 3x3 contact
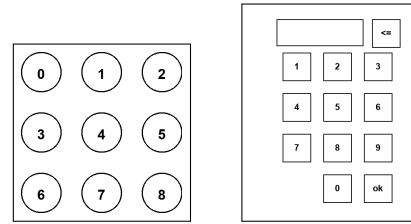

Figure 1: Pattern contact points (left), with label indexing beginning at 0, ending at 8, and a PIN layout (right) with digits 0-9, an OK button, backspace button and display screen

points. On the Android OS, four or more points should be selected. The user must maintain contact through the authentication, may not reuse contact points, nor jump over points previously un-contacted in connecting two points. Figure 1 shows the grid layout for patterns, as well as our labeling scheme of starting with index 0 through 8. For example, the L-shaped pattern would be 03678.

Additionally, pattern based authentication occurs in two flavors. The traditional setting is that visual feedback line is displayed as the user traverses the contact points (so called, with-lines). The second version requires the user to do the same input, but the feedback or tracing lines are not displayed (so called, without-lines). In prior work, it has been suggested that the without-lines version of the Android pattern is more secure from observation attacks, like shoulder surfing [39].

- Password based authentication, sometimes called an alphanumeric passcode: where a user is asked to recall a standard text-based password (entered using a soft-keyboard) to unlock the device.

The usability and security of PINs [9, 40], patterns [4, 36] and passwords [22, 29] have been well documented by researchers. Beyond these methods, picture-based [46] and biometric based mechanisms are also used to unlock mobile devices. The latter is becoming more commonplace. Fingerprint readers (e.g. TouchID on iPhone v5 or later) and face identification (through apps such as FaceCrypt or FastAccessAnywhere) can be used to verify the identity of the user. Biometrics are often utilized as a secondary authentication method, and a user with biometric authentication enabled must also have a PIN set. While biometrics offer promise to promoting quick authentication, threats related to spoofing and the vulnerabilities associated residual information left on sensors by victims can pose challenges to users [35]. In this study, we focus on two most widely used authentication mechanisms [17, 21], PINs and graphical Android patterns, with- and without feedback lines.

Shoulder surfing vulnerability. Numerous types of attacks exist where mobile authentication sequences can be obtained and used by third parties (e.g., simple guessing, smudge attacks, malware attacks [17]). Mobile device users are particularly susceptible to observational attacks, as these devices are used in a range of public and unfamiliar environments where threats may be present. Inputs can be observed and recreated. Furthermore, accessibility features such as magnification of the typed character or displaying the

last typed character as cleartext in the password entry field may compromise security [32].

Attacks may be performed through direct observation (potentially enhanced through binoculars or low-power telescopes), or through the use of recording devices (e.g. video cameras for later playback) which can be used to covertly obtain or infer credentials [23, 43]. Even if the user attempts to shield the screen from onlookers, security may be compromised through eavesdropping; listening to secure information which can later be used for purposes of recreating entry to a mobile device. Research reveals that human adversaries, even without recording devices, can be more effective at eavesdropping than expected, in particular by employing cognitive strategies and by training themselves [26].

Mechanisms to minimize occurrences of shoulder surfing. Solutions to reduce shoulder surfing include methods of obscuring entry (e.g., through the use of screen filters, such as Amzer Privacy Shield, described by [45]), limiting the ability of third parties to view authentication stimuli input from a specific angle. Drawmetric solutions also exist where input is made on the back and/or on the front of the mobile touchscreen device, obscuring the onlooker's view (e.g. the XSide system [13]). Other drawmetric approaches utilize behavioral biometrics, which can provide an additional authentication factor, to verify the user [37].

Decoy or randomization scenarios have also been proposed [38, 45], where, even after an observation, it challenges observers in recreating the authentication because he/she cannot differentiate between true and random input. Touch sensitivity can also be effective. A prescribed level of pressure during input is difficult for an attacker to recreate [27]. Similarly, unobservable, tactile feedback can also be used to thwart a shoulder surfer [1, 15, 24], where the device informs the user which of a set of passwords (or nonces) to expect.

Kim et al. [23] suggest that current approaches to reducing shoulder-surfing typically also reduce the usability of the system; often requiring users to use security tokens, interact with systems that do not provide direct feedback or require additional steps to prevent an observer from easily disambiguating the input to determine the password [30, 43]. Bianchi and Oakley [8] suggest that authentication becomes a difficult, challenging task as some systems targeting security against malicious attackers typically place high demands on users. Wiese and Roth [44] highlight the difficulties in ascertaining the efficacy of shoulder-surfing-resistant technologies due to the lack of comparative studies. The researchers have highlighted that as set-ups and assumptions made vary by author, it can be difficult to determine the security and usability of solutions.

Additionally, most of these studies do not compare directly to the current state of the art in mobile authentication, namely, how well do PINs or patterns (or other current methods) perform under attack. We attempt to fill in that gap here by providing some baselines for what level of security to expect from current authentication choices.

Evaluations of shoulder surfing using video recordings. According to [44], in order to determine the resistance of an interface to shoulder-surfing, the three main methods used by researchers include: (a) participants are cast into the roles of adversaries and users, where adversaries observe authentication sessions of users; (b) an expert adversary observes the authentication sessions of all participants; or, (c) participants are cast into the role of adversaries and observe authentication sessions of an expert user. While each method has its own advantages and disadvantages, considerations should be made regarding learning, motivation and aptitude, to develop a more reliable perception of risk.

Most related to this work is when researchers present participants with sets of video recordings depicting actors attempting to authenticate entry. Recordings generally aim to simulate an over the shoulder view. Setting up the videos in this manner ensured that the attackers would not be affected by inconsistency caused by the target [13, 34]. In prior work, the choice of number of observations appears to be arbitrary in nature [44]. Schaub et al. [33] aimed to determine how participants fare when attempting to recreate authentication sequences, comparing those watching video footage vs live attempts (i.e., physically viewing over a user's shoulder). Findings revealed that the success rate of video observations are lower for almost all schemes than the respective live results, with a few exceptions deviating by only 1–2 observations. Wiese and Roth [44] recommend preferring live observations to study human shoulder surfing unless good reasons speak in favor of using video.

As our study attempted to perform a large-scale, controlled study to systematically compare the two authentication methods, we opted to use video recordings of a single "expert user" being attacked by our participants. This allowed us to perform finely tuned randomizations and make comparisons between conditions. As such, as suggested by prior work [33, 44], one can consider these results as lower-bounds on the security. Live observations from the same angle would likely increase the vulnerability to shoulder surfing.

Baselines and guidance. While researchers have extensively explored ways to address shoulder-surfing attacks, recommendations have been proposed on ways to design and conduct these types of study. For example, Wiese and Roth [44] recommend rather than arbitrarily selecting a number of observations, that the number of observations made by adversaries should match their assumptions about the scenario and the environment where the scheme will be deployed. Observation strategies should also be taken into account, to gain a more detailed view of feasible strategies. In terms of set-up, Sahami Shirazi et al. [31] propose recording video footage from four different angles: front, rear, left and right, in order to compensate the loss of 3D information in 2D videos. While limited detail was provided about the relative positioning of each camera, this type of technique would be useful to better simulate shoulder surfing scenarios. Schaub et al. [33] have highlighted different ways that users hold and interact with the device. Occlusion by the user's hand and fingers may reduce visibility for shoulder surfers and enhance observation resistance.

As we will describe in the next section, we attempt to account for many of these factors and suggestions. Namely, we apply video recordings from multiple angles, allow for repeated observations and repeated entries, and we also consider different form factors and hand positions for our mobile devices.

3 METHODOLOGY

We designed a mixed-factorial design with both between- and within-subject factors in order to reduce the duration of the study to an acceptable length. Between subjects, we randomized participants into 12 groups based on the authentication type (3-treatments), hand position (2-treatments), and phone type (2-treatments). Within each group, participants were shown a series of videos for a set of 10 authentications. After each video, each participant attempted to recreate the authentication observed. As part of a within group analysis, the observation angles, the number of observations, and the number of attempts to recreate the authentication were randomized.

Based on this design, we intended to address the following set of hypotheses:

- **H1**: The type of unlock authentication, PIN, Pattern with-lines, Patterns without-lines, affects the shoulder surfing vulnerability.
- **H2**: Repeated viewings of user input increase the likelihood of a shoulder surfing vulnerability.
- **H3**: Multiple attempts to recreate the input increase the likelihood of a shoulder surfing vulnerability.
- **H4**: The angle of observations affects shoulder surfing vulnerability.
- **H5**: The properties of the unlock authentication, such as length and visual features, affect shoulder surfing vulnerability.
- **H6**: The phone size affects shoulder surfing vulnerability.
- **H7**: The hand position used to hold and interact with a device affects shoulder surfing vulnerability.

In the remainder of this section, we outline the settings of our experiment and the design choices made. We first discuss the settings of our video recordings that dictate the participant groups, following which we discuss the password/PINs used in the experiments, how they were selected, and the properties they exhibit. Finally, we discuss the survey mechanisms, training, and other procedures.

3.1 Video Recording Settings

Phone settings. We used two phones in our experiments: Nexus 5 and the OnePlus One. The Nexus 5 is a mid-range size phone, with a 5" display. The OnePlusOne has a larger form factor of 6" (compared to 5.4" of the Nexus 5) with a screen size of 5.5". Both phones have the same resolution of 1080x1200 pixels. These two phones are similar to a wide variety of displays and form factors available on the market today, for both Android and iPhone. In charts and tables, we refer to the phones by their coloring, *red* for the Nexus 5x and *black* for the OnePlus One.

The goal of using these two phones is to understand how larger form factors, which provide more viewable space, may affect the attackers ability to shoulder surf (**H6**). There are also side effects for a larger display that we did not anticipate. For example, in Figure 2, with the larger OnePlus One phone, we experienced more glare on the screen as it was a bit more unwieldy. Being larger in the hand, the OnePlus phone moved more during PIN/Pattern entry, particularly one-handed, which caused more opportunities for glare.

Nexus 5 (*red phone*)
5.43"x2.72" form factor, 4.95" display, 1080x1920

| *Thumb* | *Index* |

OnePlus One (*black phone*)
6.02"x2.99" form factor, 5.5" display, 1080x1920
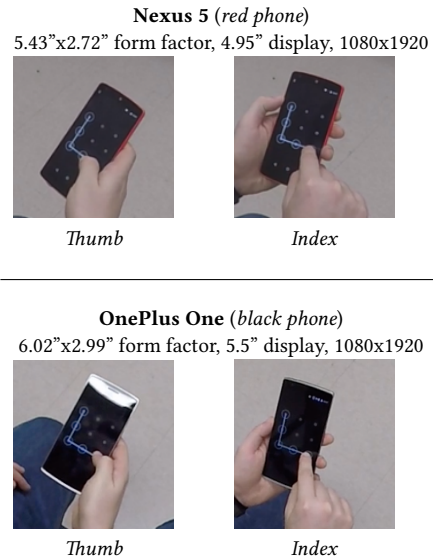

| *Thumb* | *Index* |

Figure 2: Phone Types and Hand Positions: *top* is the Nexus 5x phone and *bottom* is the OnePlus One phone. The Nexus 5x is roughly the size of a iPhone 6s and the OnePlus one is roughly the size of a iPhone 6s+. On the *left* is single handed entry, using the thumb only, and on the *right* is two handed entry using the index finger.

Hand positions. We investigated two different phone-grips (or hand positions) for authentication entry. Figure 2 shows the grips. The images on the top-left and bottom-left show a single handed grip being used, where the thumb is used to enter the authentication. The images at the top-right and bottom-right, the grip is a two handed grip, where the user holds the phone in their left hand and enters the authentication sequence using the index-finger of their right hand. These are both common grip settings for mobile devices [16]. We focus exclusively on right handed entry modes to reduce the complexity of our experiment. In charts and tables, we describe these two hand positions as *thumb* for the single handed grip with thumb entry, and *index* for the two handed grip with index finger entry.

We applied these two conditions because we hypothesized (**H7**) that visual obstructions may impact the vulnerability to shoulder surfing. For example, using an index finger provides the least obstructed view, compared to using the thumb, but it also may increase point-of-view obfuscation where it may appear that contact is being made with the phone, when it is only an illusion due to the the angle of observation.

Angles of recording. We used a camera array to simultaneously record each authentication (e.g., one phone type, one hand position type, one authentication input) from multiple angles. The camera array is shown in Figure 3. The target user, who is seated for the study, is subject of observations from five angles in the camera array to simulate different vantage points. Outlined in Table 1, the angles are, from each side *left* and *right* with a *far* and *near* angle. We also had a *top* angle with a vantage immediate overhead of the target user. In charts and tables, we shorthand these angles as: *nl*

Figure 3: GoPro Camera Array: lower cameras are *near*, higher cameras are *far*, and the middle camera is *top*

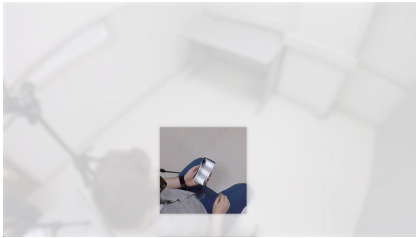| Angle Name	Visual	Description
Near Left (nl)		Over target's left shoulder at a height of 5'
Far Left (fl)		Over target's left shoulder at a height of 6'
Top (t)		Over target's head at a height of 6'
Near Right (nr)		Over target's rigth shoulder at a height of 5'
Far Right (fr)		Over target's right shoulder at a height of 6'

Table 1: Camera Angles

Figure 4: Full screen video with "focus zone" to highlight where to view the authentication and to remove distractions from other factors in the recording space.

for near left, *fl* for far left, *t* for top, *nr* for near right, and *fr* for far right.

We hypothesized that there may exist settings of observations that both hinder and enhance the attackers ability to shoulder surf (**H4**). For example, observations from one side over the other (e.g., left v. right) may provide more or less obstructed views, aiding or hindering shoulder surfing.

Editing Videos. During video recordings, we attempted to make each authentication occur over a consistent length of time with

PINs	properties	Patterns	properties
1328	up/non-adj	0145	up
1955	neutral/non-adj/repeats	1346	left
5962	right	3157	neutral
6702	down/kmove/cross	4572	right/cross
7272	left/kmoves/repeats	6745	down
153525	up/repeat	014763	left/cross
159428	neutral/cross/non-adj	136785	down
366792	right/repeat/kmove/cross	642580	neutral/cross
441791	left/kmove/repeat	743521	up/non-adj
458090	down/repeat	841257	right/kmove/cross

Table 2: PINs and patterns used in experiments. See Appendix B.1 and B.2 for visuals of the authentication.

a consistent hand motion. We further attempted to remove any distractions from the observation area so that participants can focus directly on the task of shoulder surfing. Each video recording, is about 3-5 seconds in length, but this creates a tracking challenge for the participant who needs to quickly determine where to look in a video (occurring from different angles each time) to do the observation. To alleviate this burden, we edited the videos by placing a "focus zone" in the video. See Figure 4 for a visual of this editing. Except for the authentication area, the remainder of the screen is set with a transparent gray so that the participant can quickly determine where to focus their visual attention for the observation task.

3.2 Authentication Settings

As previously mentioned, we aim to analyze two different authentication settings, PINs and Android graphical pattern unlock. Within the Android pattern settings, we also consider settings where the tracing lines are either displayed or not displayed. Recent work has suggested that tracing lines should *not* be displayed for improved security [39]. For each authentication setting, we have chosen a set of 10 representative PINs and patterns that have spatial shifting properties and visual, complexity properties, such as crosses.

In the remainder of this section, we outline how that selection was performed and justify the properties used during selection. Additionally, we describe the application used for performing input and how it was designed to fairly compare the two authentication types.

Pattern Selections. The patterns used in our experiment are shown in Table 2 (graphical representations are presented in the Appendix). These patterns were culled from a set of self-reported patterns collected through an online study [4], and provided to us for analysis and use. From these patterns, we identified five 4-length patterns and five 6-length patterns that exhibited a broad set of representative features.

To determine which features to consider, we hypothesized that there may be locations in the grid space that increase or decrease the effectiveness of the attack (**H5**), as well as complexity features of patterns [2, 3]. We were guided by related work [5, 41] in choosing the features, for both spatial aspects and complexity properties.

- *up* shifted: The contact points of the pattern are in the upper part of the grid space, such as pattern 0145 and 743521.

- *down* shifted: The contact points of the pattern are in the lower part of the grid space, such as pattern 6745 and 136785.
- *left* shifted: The contact points of the pattern are in the left portion of the grid space, such as pattern 1346 and 014673.
- *right* shifted: The contact points of the pattern are in the right portion of the grid space, such as pattern 4572 and 841257.
- *neutral*: The contact points are evenly distributed in the grid space.
- *non-adjacency* (or non-adj): Two, non-adjacent contact points are used, such as pattern 743521, where contact point 3 and 5 are non-adjacent and are connected because contact point 4 was contacted prior.
- *knight move* (or kmove): Two contact points are connected over two and down one (or in any symmetry), like a knight moves in chess, like in pattern 4572.
- *cross*: The sequence of contacts crosses over itself, such as pattern 014673 and 841257 have a perfect 'X', but more obtuse crosses also exist, such as in pattern 4572 due to a knight move.

PIN selections. In order to select PINs, we followed related work in analyzing digit sequences in password datasets [9]. Using the RockYou dataset[1], we extracted 4- and 6-length digit sequences that exhibited similar properties to that of the pattern dataset. The idea being that these digit sequences are likely to be reused as PINs if they appear in passwords.

Matching the PINs to the exact features in the patterns is not perfect, as not all digit sequences found in patterns exist within the RockYou dataset, and further, we wish to include all 10 digits (patterns only use 9 contact points). PINs also have a feature that patterns cannot have, repeated digits, so we wish to include PINs with this property, either a single digit or multiple repeated digits. The final set of PINs selected are available in Table 2, and a visual is provided in Appendix B.2.

Authentication Applications. Another important factor to consider is the applications used for entering the authentication. Critically, the size of each application should be the same and have similar visual properties, so as not to advantage one over the other for shoulder surfing. To this end, we designed two Javascript applications using HTML5 that ran in the Android Chrome browser, setup as a home screen link to simulate a standalone application. Each application mimicked the input used on the device, following the same rules. During the survey, the same applications would be used as embedded Javascript in their browser for the participants to recreate the authentication observed.

For patterns with-line feedback, after the target user completed the application, there would be a brief, 200 ms pause before the screen would go to blank/black screen. This is to simulate the unlock process on the phone. A similar action occurs for patterns without-line feedback, however, no tracing lines or circled contact points would be seen.

For PINs, the input text area would show the number that was pressed, but would fade to an asterisk after one second or after the next number was pressed, similar to how unlock authentication works on smartphones. Only after pressing "ok" would the screen go blank, simulating an unlock.

3.3 Survey Protocol

We designed a protocol around the video recordings by which participants would be assigned a randomized group, receiving training relevant to that group, and then attempt to shoulder surf 10 authentications based on observing videos under different settings. The survey was designed as a web application using a combination of PHP, Javascript, and a MySQL backend. The survey was posted on Amazon Mechanical Turk and participants were also recruited locally at our institution to ensure consistency.

The survey protocol proceeded as follows:

(1) Informed Consent
(2) Demographic and Background Information
(3) Training
(4) Observations and Recreation
(5) Attention Check and Submission

In the remainder of this section we outline each of these survey segments in detail, as well as the randomization and recruitment process.

Informed Consent and Preliminary Instructions. This survey was approved by our institutional oversight board (IRB), and so we require participants to provide informed consent. For online participants, this was done digitally, and for in-person participants, it was done in a traditional manner, following a script.

The informed consent also provided participants with an overview of the experiment, its goals, and initial instructions. For example, it informed participants that they were participating in a research project about shoulder surfing, as well as directions about the procedures:

> The survey will request that you maximize the browser window on your screen. You are not permitted to record the survey or any of its content. The use of pen and paper to write anything down is also strictly prohibited. The survey will request that you watch several videos of a user authenticate into a mobile device. You are to watch the video and attempt to recreate the PIN or pattern you viewed being entered.

Demographic and Background Information. Following acknowledgment of the informed consent, we ask a series of demographic questions. Including:

- Gender (Male, Female, Prefer not to answer)
- Age (drop down box, 18-100)
- Eye Sight (Normal, Corrected with glasses/contacts, Deficient and not corrected)
- Ability with modern cell phones (None, Below Average, Average, Above Average, Professional)

Additionally, we recorded the screen size of the browser, in pixels, to test if participants were following directions as well to get a sense of the different viewing scenarios.

---

[1]Originating from a debunk music sharing web site, the RockYou dataset was leaked in 2009 and contains over 32 million passwords commonly used by researchers [42].

Figure 5: Frame of tutorial video

Conditions	Views	Attempts
A	One	One
B	One	Two
C	Two	One
D	Two (different angles)	One
E	Two (different angles)	Two

Table 3: Five different conditions for each authentication

		In-person			*Online (MTurk)*			
		Male	Female	Total	Male	Female	Non-Spec.	Total
Age	18-24	68	23	91	103	78		181
	25-34				304	221	6	531
	35-44				142	120	1	263
	45-54				47	87		134
	54-64				21	27		48
	65+				7	8	1	16
Sight	Deficient				7	3		10
	Corrected	12	9	21	225	252	3	480
	Normal	56	14	70	392	286	5	684
Skill	Below				17	11	9	37
	Below Average	22	10	32	134	204	6	344
	Above Average	38	12	50	344	277	1	622
	Professional	8	1	9	129	49	1	179
Resolution	< 1300				117	119	4	240
	1300-1500				189	250	1	440
	1500-1800				122	85	2	209
	> 1800	68	23		196	87	1	284
		68	23	**91**	624	541	8	**1173**

Table 4: Demographic Information. The *resolution* refers to the width of the screen resolution, in pixels.

Between Treatment Randomization and Training. At this point in the survey, we randomize the treatments as the remainder portion is dependent on that randomization. We initially randomize into 12 between subject treatment groups:

- Authentication Type: PIN, patterns with feedback lines, or patterns without feedback lines
- Hand Position: index or thumb
- Phone Type: either the red Nexus 5 or the black OnePlus One

Based on the selection, we prepared three training videos that explained the procedure further specifically for each authentication type, and then 12 sample test videos that participants can use to practice shoulder surfing. The test video used the same conditions as the selected treatment, but with a sample PIN (1234) or Pattern (0123). The training video shows the participant how the observation and recall would proceed (a screen-shot of the video is in Figure 5), and once the video completes, test runs are performed using the sample PIN or pattern. Participants are allowed to repeat this training video and test runs as many times as needed before continuing to the main portion of the survey.

Within Treatment Randomization for Observation and Recall. At this point, a participant has been assigned an authentication type, phone type, and hand position. There is now a large set of videos from multiple angles for each of the authentications, but it is not feasible (nor desirable) to display every video to each participant. Instead, we proceed with a within-group randomization to display a subset of those videos under different settings that will support testing hypothesis **H2**, **H3**, and **H4**.

The first stage of randomization is to randomize the order of the authentication that will be displayed. That is, each participant will observe all 10 of the authentications in their selected authentication type, either 10 PINs or 10 patterns, but the order of those must be randomized to handle training effects where the participants become better at the task as time goes on. Once the order is randomized, for each authentication, we then randomize and counterbalance a set of conditions regarding how many views and attempts a participant gets to make, as outline in Table 3.

The "views" refer to how many times a participant gets to view an observation video. For conditions A-C, a random angle is selected, and the participant either gets a single view of that authentication (A,B), or two views from the same angle (C). For conditions D and E, participants get a random first angle selection, and then are assigned a second angle on the opposite side (e.g., first angle is a left side, second angle is a ride side). If the top angle was selected, then a random second angle is used.

The second part of each condition is the number of attempts. After viewing the video, the participant can make either one attempt to recreate the authentication or two attempts.

Prior to each video observation, we informed the participant if they were going to view one or two videos and if they would have one or two attempts.

Submission and Attention Tests. Following the survey, we ask participants to report if they used additional aids, such as pen and paper, in helping them complete the procedure. This acts as both an attention test and a guide for including or excluding results. It also allows us to exclude participants who failed to follow directions. We did not have anyone report that they "cheated."

3.4 Recruitment

We recruited locally at our institution, and online via Amazon Mechanical Turk. The goal of using both recruitment methods is that for the institutionally recruited participants, we can control the settings, and so we wished to compare these results to those collected online for consistency (see Figure 6). Inconsistent results would suggest that online participants were not taking the survey faithfully. We observed consistent results when comparing similar demographic groups with similar screen resolutions, as described later, suggesting that participants online took the survey in the intended ways. Although, there was some degradation of performance, which may be accounted for by an observation bias or the

	Online (MTurk)					In-person				
	Single View		Multi View			Single View		Multi View		
	4-length	6-length	4-length	6-length	total	4-length	6-length	4-length	6-length	total
PIN	34.92%	10.86%	56.72%	26.53%	32.25%	52.63%	20.75%	76.62%	59.32%	46.02%
NPAT	51.03%	35.28%	71.27%	52.10%	52.28%	72.29%	62.31%	84.31%	95.38%	72.92%
PAT	80.90%	64.20%	88.07%	79.85%	78.27%	94.5%	86.74%	98.61%	83.53%	92.42%
total	55.73%	37.28%	72.15%	52.81%		73.70%	56.70%	86.5%	80.38%	
	46.45%		62.53%		54.54%	65.11%		83.37%		70.81%

Table 5: Single- vs. multi-view for authentication types broken up based on online and in-person participants. NPAT is pattern *without* feedback lines and PAT is with feedback lines. Comparing single vs. multi-view, in all categories, was statistically significant, as well as in-person vs online.

Hawthorne effect; local participants, being observed, were more likely to try and perform the task well to appease their observers.

In total, we recruited 91 participants locally at our institution, and 1173 online participants. The demographic information is available in Table 4. The material used in recruitment mimicked that of the informed consent. The text used in posting the task to Amazon Mechanical Turk is provided in Appendix A.

3.5 Realism and Limitations

We acknowledge that our experimental methodology has a number of limitations. Foremost, we had to reduce the set of authentication tokens to a reasonable size, namely 10, so that we could maintain a reasonable survey length with a reasonable recruitment size. We attempted to mitigate this effect by choosing real authentications, as collected in other datasets, that would be representative of authentication choice broadly. We further did not include text-based passwords, which can form an unlock authentication, as we were unable to develop a protocol to fairly compare to the other authenticators.

We were additionally limited in terms of the observation settings. Our online participants may have used screens that were bigger or smaller than we anticipated. We attempt to manage this limitation by recording the screen size, and, as we will show in this paper, there was an impact on performance with respect to screen size. However, general trend lines remain the same, when we compare the online data to that collected in-person.

3.6 Ethical Considerations

This protocol was reviewed and approved by our institution review board to ensure that participants were treated fairly, such as providing informed consent and an option to opt-out. The survey itself does not elicit ethical challenges as participants are not performing actions that increase the risk to others or themselves in regard to shoulder surfing. It could be argued these participants may be more aware of the risks associated with these attacks after having participated. The identity of the target victim was protected from participants via obfuscation. Finally, the analysis does not include identifiable information about participants.

4 RESULTS

In this section, we describe the results of the survey by addressing each of the hypotheses outlined earlier. We also provide other
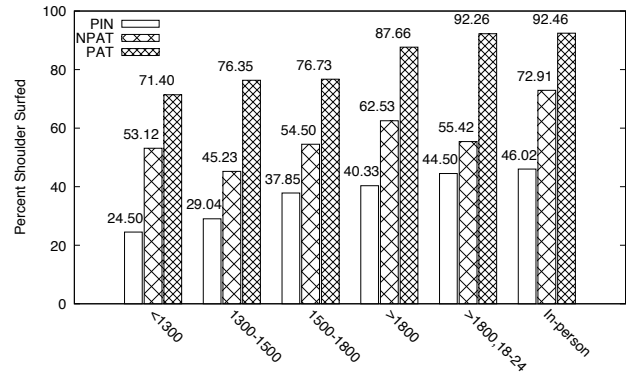

Figure 6: Screen width resolution comparison. In-person subjects used a screen with resolution 990x1840 and are in the age range of 18-24. There is no significant difference (G-Test) between PIN and PAT, there remains a significant difference for NPAT when comparing in-person samples to on line ones with similar resolutions and demographics.

insights as available, particularly related to the realism of the experiment. As we move through the results, it is important to note that in some conditions (C and E) participants had multiple attempts to recreate the observed authentication, which we study in more detail later. Unless otherwise noted, we consider a successful attack if the participants accurately recalled and entered the authentication sequence within either of the attempts.

For statistical testing, our data is categorical and binary, as in a participant either correctly recalled and entered an authentication sequence or did not. As such, in two way comparisons of attack rates, we applied Fisher's Exact Test (or G-test) to test significance, and χ^2 test for comparing for multi-factor analysis. Additionally, we perform a L1-penalty logistic regression analysis to determine the impact (or lack thereof) of all settings of the experiment. A significance level of $p < 0.05$ is used. Finally, unless otherwise stated, each of the tables, when a percentage is displayed, this refers to the rate in which an authentication was successfully recreated in that setting, a so called success or attack rate.

Realism of online results. An important question to consider is if online results are consistent with those collected in-person. As evident in Table 5, there is a significant performance improvement for those in-person participants ($p < 0.005$, using χ^2). Investigating this phenomenon further, we broke down the participants based

	Hand Position			*Phone Type*			*Input Attempts*		
	Index	Thumb	G-test	Red	Black	G-test	One	Two	G-test
PIN	32.74%	32.22%	$p = 0.68$	30.84%	34.04%	$p < 0.05$	35.51%	30.43%	$p < 0.00005$
NPAT	53.82%	50.75%	$p < 0.05$	76.22%	80.23%	$p < 0.05$	56.56%	49.38%	$p < 1 \times 10^{-7}$
PAT	79.93%	76.69%	$p < 0.05$	53.40%	50.97%	$p < 0.0005$	79.90%	77.16%	$p < 1 \times 10^{-9}$
total	55.30%	53.78%	$p < 0.05$	52.92%	56.06%	$p < 0.0005$	57.46%	52.56%	$p < 0.005$

Table 6: Hand position, phone type, input attempts, and observation angle impact for online participants. For hand position and phone type, single and multi-view treatments are considered.

	Observation Angle				
	far-right	far-left	near-right	near-left	top
PIN	20.39%**	18.88%**	21.77%**	32.46%**	22.09%**
NPAT	46.09%	35.62%**	46.57%	41.89%**	45.14%*
PAT	70.31%**	68.23%**	73.47%**	71.50%**	79.08%**
total	45.77%**	41.03%**	47.43%*	49.15%	49.18%

Table 7: Impact of observation angle on shoulder surfing. Single-view treatments and only multi-view treatments of the *same* angle are considered (see Table 5 for single- vs. multi-view). Using χ^2 testing, $*$ indicates $p < 0.05$, $**$ indicates $p < 0.005$.

on the width of the screen resolution used while taking the survey, which is a good approximation for the size of their viewing area. These results are presented in Figure 6, and one can clearly see that as the resolution width increases, so does the performance. As we controlled our in-person computing setup, we know precisely the screen resolution of 990x1840, and further, our in-person participants (being undergraduates) are between the ages of 18-24. When isolating this demographic group, we find no statistical differences between the PIN and PAT results, with remaining difference for patterns without traceback lines (NPAT). This difference is likely the result of an observation bias, by which having the researchers present led the participants to want to perform the task "better." As such, we find that these results suggest that online participants likely took the survey in the intended manner, and variations in screen size (and other factors) probably realistically mimic the realities of shoulder surfing in the wild. The remaining results focus solely on the online dataset.

H1: authentication type. We applied both Fisher's exact test and χ^2 test to the data in Table 5 and found all comparisons between authentication type to be significant. Focusing on the online results, we find that the authentication plays a significant role. PINs proved the most elusive in all settings, with combined performance of 32.25% attack rate. Patterns with traceback lines was the worst performing, 78.27% attack rate across all settings. Removing traceback lines improved results to 58.28%, confirming prior work on this topic [39]. **As such, we *accept* the hypothesis that the authentication type impacts shoulder surfing vulnerability.**

H2: repeated observations. Using the results in Table 5, we find that there are significant differences between the single-view and multi-view settings. Looking at both online and in-person results, participants are about 1.3x-1.4x more likely to correctly attack an authentication if allowed multiple views of the authentication. Later,

as we compare all the features, we find that multiple views, in particular, play an outsized role in the vulnerability of authentication to shoulder surfing. **As such, we *accept* the hypothesis that multiple observations impact shoulder surfing vulnerability.**

H3: multiple input attempts. Recall that we applied a within-group randomization by which some participants on some authentication were provided two attempts to input the authentication. The procedure of the survey informed them of this fact, so participants were aware, prior to viewing the video, that they would have multiple recreation attempts. Table 6 shows these results on the *right* column.

Surprisingly, multiple attempts decrease performance, in all cases. We believe this is because participants, knowing they would have multiple attempts, attempted to "game" the process in a way that actually led them to get the pattern wrong in both attempt cases. For example, they would pay attention less well. **We *accept* the hypothesis that multiple input attempts affect shoulder surfing**, but it decreases performance, unexpectedly. From this result, researchers should consider for similar experiments to either not informing participants how many attempts at recreation, or force participants into a regime of single attempts, requiring more attention during that single attempt.

H4: observation angle. Table 7 presents the results comparing performance for the different observation angles. As we wished to isolate the angle, we only consider treatments where a single observation angle was used. We used a χ^2-test to determine significance factors in these scenarios, indicated with *'s in the table. In nearly all cases, within each authentication type, we found that there are significant impacts based on the angle of observation. When performing comparisons in total, we find that the far-right, far-left and near-right angles showed the most significant impact. The far-left angle, in particular, was the most challenging angle, and we believe that this angle provide some obfuscation of when screen touching occurred, making it harder for participant to cleanly determine the location of touch events. **As such, we *accept* the hypothesis that the observation angle affects shoulder surfing.**

H5: properties of authentication. We first consider the length of the authentication, the results of which are displayed in Table 5. The length has a large impact. In most cases, it decreased the rate of shoulder surfing by nearly 50%. While length is far from a perfect approximation for security, it's clear that longer authentication will improve security from observation attack. We further breakdown the vulnerability of the individual authentications in Table 8. Many of the authentications vary from an expected uniform attack rate, as observed by using a χ^2 test within each authentication length. However, there does not appear to be a direct

PIN		NPAT		PAT	
1328	50.36%	0145	81.10%**	0145	92.28%**
1955	47.89%	1346	43.18%*	1346	82.96%**
5962	36.10%**	3157	60.26%**	3157	87.09%**
6702	34.64%**	4572	47.90%	4572	74.12%**
7272	60.53%**	6745	73.70%**	6745	86.51%**
153525	15.40%**	014763	53.44%	014763	84.01%**
159428	17.17%**	136785	41.81%*	136785	73.76%**
366792	19.57%**	642580	46.49%	642580	74.03%**
441791	20.66%**	743521	37.90%**	743521	53.79%
458090	21.83%**	841257	37.78%**	841257	73.93%**

Table 8: Individual authentication attack rate. Significance tested using χ^2 within authentication of the same length, * indicating $p < 0.05$ and ** indication $p < 0.005$, or much less than.

pattern related to the individual spatial properties of the authentication, additional analysis with more authentication types would be needed to draw strong conclusions regarding these features. **As such, we *partially accept* the hypothesis that the properties of authentication impact shoulder surfing,** while features such as authentication length play a large role, the impact of other features is inconclusive.

H6: phone size. Table 6 shows the result of comparing the two phones in the study. Recall that the Red phone refers to the 5" display Nexus 5, and the Black phone refers to the 5.5" display of OnePlus One. Across all conditions, we find that there is a significant difference in shoulder surfing between the two phones. In most cases, the larger Black phone provides less security, except for patterns (PAT), where the smaller Red phone is more secure. After reviewing the videos, we noticed that the larger Black phone experiences more glare during this recording which could account for the difference. Overall, it appears that the larger phones provide less security for shoulder surfing, and **we accept the hypothesis.**

H7: hand position. Recall that we examined two different hand positions. One hand position (or grip) had the victim use a single hand, entering the authentication with his thumb. The second hand position was two handed, holding the phone in the left hand entering the patterns with the right index finger. Table 6 shows the results of comparing these two conditions, *thumb* vs *index*. The results for comparing PINs showed no significant difference; however there was significant, but small, differences between pattern entry for the different hand positions, as well as a small significant difference overall. While there is a difference, the impact factor is challenged, so **we *reject* the hypothesis that hand position impacts shoulder surfing.** These results suggest that researchers can allow for any normal hand position without greatly impacting the results; however, using an index finger provides a more direct view, as opposed to the one-handed thumb blocking portions of the screen) and likely improves results, nominally.

Comparison across features. Finally, we wish to understand how the combination of the features impact the results, asking the question, are there a set of ideal conditions or non-ideal conditions for shoulder surfing that can form a set of baselines? To accomplish this, we performed a logistic regression across all the features using L1-penalties such that features that have small (or

Feature	Coefficient	Feature	Coefficient
PIN	-0.90**	Single-View	-0.09
PAT	1.25**	Multi-View	0.40*
NPAT	0.00	One-Input	0.11*
4-length	0.42*	Two-Input	-0.10*
6-length	-0.33*	far-left	-0.24*
Thumb	-0.03	far-right	-0.06
Index	0.06	near-left	0.12*
Red	-0.03	near-right	0.00
Black	0.09	top	0.14*

Table 9: L1-penalty logistic regression using all features, the average of 100 runs of the regression. 68.7% of the data is explained by the regression. The * indicate top ranked coefficients. The model is significant.

no) effect can have a coefficient of zero. The results of an average of 100 runs of the regression (there were many different minimums) are presented in Table 9.

The regression was set-up using a feature set of binary values, with a one indicating the presence of the feature and zero otherwise. The label on the feature was also binary, a zero indicating that shoulder surfing attack failed and one indicating success. We trained over each trial of the survey, and the resulting model was able to explain 68.7% of the data and was significant.

We can further analyze the coefficients of the features which indicate how much weight they provide to the prediction and also if they increase or decrease the likelihood of shoulder surfing. Negative values imply greater security to shoulder surfing, while positive values indicate more vulnerability to shoulder surfing. As we are using L1 penalty, some coefficients can reduce to zero.

Most surprisingly, the coefficient for NPAT (patterns without tracing lines) is 0. This makes sense if you consider the fact that being a PIN so greatly reducing the likelihood of shoulder surfing, while patterns greatly increase the likelihood. The fact that it is a pattern without lines is not predictive, in comparison to those other two facts. Further, the highest coefficient is that of shoulder surfing a pattern, followed by PINs (in the negative direction). This further supports accepting hypothesis **H1**.

Among the other coefficients, the length factor and having multiple views of the authentication play a large role in shoulder surfing attack rates. The far-left angle proved to be the most challenging for shoulder surfing, as identified earlier while near-left and top were the most beneficial for shoulder surfing.

Based on these results, we can now identify category of the best case scenario for an attacker performing shoulder surfing: attacking a pattern with tracing lines that is of length four when provided multiple with views. Similarly, the worst case scenario is attacking a PIN of length 6 from the far left when provided with just one view. These two scenarios can provide a baseline to compare new systems that offer protections to shoulder surfing, as well as help inform users of stronger authentication choices.

5 CONCLUSION

We presented the results of a large scale, online study of shoulder surfing for the most common unlock authentication, PINs, patterns with tracing lines, and patterns without tracing lines. We find that PINs are the most secure to shoulder surfing attacks, and while both types of pattern input are poor, patterns without lines provides

greater security. The length of the input also has an impact; longer authentication is more secure to shoulder surfing. Additionally, if the attacker has multiple-views of the authentication, the attacker's performance is greatly improved.

Overall, the goal of this research is to work towards establishing baselines for how current authentication performs against shoulder surfing, as well as provide insight into settings of current authentication that can protect users from shoulder surfing. Based on our analysis, researchers should consider comparing their performance of new systems to the most secure setting, namely using at least 6-digit PINs with just a single view, as well as to the least secure setting of using a 4-length pattern with visible lines with multiple views. Additionally, these results suggest, for users, that 6-digit (or longer) PINs provide the best security from shoulder surfing.

REFERENCES

[1] Abdullah Ali, Adam J Aviv, and Ravi Kuber. 2016. Developing and evaluating a gestural and tactile mobile interface to support user authentication. *IConference 2016 Proceedings* (2016).

[2] Panagiotis Andriotis, Theo Tryfonas, and George Oikonomou. 2014. Complexity metrics and user strength perceptions of the pattern-lock graphical authentication method. In *Human Aspects of Information Security, Privacy, and Trust*. Springer, 115–126.

[3] 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 (WiSec'13)*. 1–6.

[4] Adam J. Aviv, Devon Budzitowski, and Ravi Kuber. 2015. Is Bigger Better? Comparing User-Generated Passwords on 3x3 vs. 4x4 Grid Sizes for Android's Pattern Unlock. In *Proceedings of the 31st Annual Computer Security Applications Conference (ACSAC 2015)*. ACM, New York, NY, USA, 301–310. https://doi.org/10.1145/2818000.2818014

[5] Adam J. Aviv and Dane Fichter. 2014. Understanding Visual Perceptions of Usability and Security of Android's Graphical Password Pattern. In *Proceedings of the 30th Annual Computer Security Applications Conference (ACSAC '14)*. ACM, New York, NY, USA, 286–295. https://doi.org/10.1145/2664243.2664253

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

[7] Chandrasekhar Bhagavatula, Blase Ur, Kevin Iacovino, Su Mon Kywe, Lorrie Faith Cranor, and Marios Savvides. 2015. Biometric Authentication on iPhone and Android: Usability, Perceptions, and Influences on Adoption. (2015).

[8] Andrea Bianchi and Ian Oakley. 2014. Multiplexed input to protect against casual observers. In *Proceedings of HCI Korea*. Hanbit Media, Inc., 7–11.

[9] Joseph Bonneau, Sören Preibusch, and Ross Anderson. 2012. A Birthday Present Every Eleven Wallets? The Security of Customer-Chosen Banking PINs. 25–40.

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

[11] Alexander De Luca, Martin Denzel, and Heinrich Hussmann. 2009. Look into My Eyes!: Can You Guess My Password?. In *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS '09)*. ACM, New York, NY, USA, Article 7, 12 pages. https://doi.org/10.1145/1572532.1572542

[12] 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 *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 987–996. https://doi.org/10.1145/2207676.2208544

[13] Alexander De Luca, Marian Harbach, Emanuel von Zezschwitz, Max-Emanuel Maurer, Bernhard Ewald Slawik, Heinrich Hussmann, and Matthew Smith. 2014. Now You See Me, Now You Don'T: Protecting Smartphone Authentication from Shoulder Surfers. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2937–2946. https://doi.org/10.1145/2556288.2557097

[14] Alexander De Luca, Katja Hertzschuch, and Heinrich Hussmann. 2010. Color-PIN: Securing PIN Entry Through Indirect Input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 1103–1106. https://doi.org/10.1145/1753326.1753490

[15] Alexander De Luca, Emanuel Von Zezschwitz, and Heinrich Hußmann. 2009. Vibrapass: secure authentication based on shared lies. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 913–916.

[16] Rachel Eardley, Anne Roudaut, Steve Gill, and Stephen J. Thompson. 2017. Understanding Grip Shifts: How Form Factors Impact Hand Movements on Mobile Phones. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4680–4691. https://doi.org/10.1145/3025453.3025835

[17] Serge Egelman, Sakshi Jain, Rebecca S. Portnoff, Kerwell Liao, Sunny Consolvo, and David Wagner. 2014. Are You Ready to Lock?. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS '14)*. ACM, New York, NY, USA, 750–761. https://doi.org/10.1145/2660267.2660273

[18] Cyrus Farivar. Jun 8, 2015. Apple to require 6-digit passcodes on newer iPhones, iPads under iOS 9:Stronger passcode ups the ante: there will be one million possible permutations. (Jun 8, 2015). http://arstechnica.com/apple/2015/06/apple-to-require-6-digit-passcodes-on-newer-iphones-ipads-under-ios-9/.

[19] Alain Forget, Sonia Chiasson, and Robert Biddle. 2010. Shoulder-surfing Resistance with Eye-gaze Entry in Cued-recall Graphical Passwords. In *Proceedings of the 2010 CHI Conference on Human Factors in Computing Systems (CHI '10)*. ACM, New York, NY, USA, 1107–1110. https://doi.org/10.1145/1753326.1753491

[20] H. Gao, Z. Ren, X. Chang, X. Liu, and U. Aickelin. 2010. A New Graphical Password Scheme Resistant to Shoulder-Surfing. In *2010 International Conference on Cyberworlds*. 194–199. https://doi.org/10.1109/CW.2010.34

[21] Marian Harbach, Alexander De Luca, and Serge Egelman. 2016. The Anatomy of Smartphone Unlocking: A Field Study of Android Lock Screens. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 4806–4817. https://doi.org/10.1145/2858036.2858267

[22] Patrick Gage Kelley, Saranga Komanduri, Michelle L. Mazurek, Richard Shay, Timothy Vidas, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, and Julio Lopez. 2012. Guess Again (and Again and Again): Measuring Password Strength by Simulating Password-Cracking Algorithms. 523–537.

[23] Sung-Hwan Kim, Jong-Woo Kim, Seon-Yeong Kim, and Hwan-Gue Cho. 2011. A new shoulder-surfing resistant password for mobile environments. In *Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication*. ACM, 27.

[24] Katharina Krombholz, Thomas Hupperich, and Thorsten Holz. 2017. May the Force Be with You: The Future of Force-Sensitive Authentication. *IEEE Internet Computing* 21, 3 (2017), 64–69.

[25] Manu Kumar, Tal Garfinkel, Dan Boneh, and Terry Winograd. 2007. Reducing Shoulder-surfing by Using Gaze-based Password Entry. In *Proceedings of the 3rd Symposium on Usable Privacy and Security (SOUPS '07)*. ACM, New York, NY, USA, 13–19. https://doi.org/10.1145/1280680.1280683

[26] Taekyoung Kwon, Sooyeon Shin, and Sarang Na. 2014. Covert attentional shoulder surfing: Human adversaries are more powerful than expected. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 44, 6 (2014), 716–727.

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

[28] Shushuang Man, Dawei Hong, and Manton M Matthews. 2003. A Shoulder-Surfing Resistant Graphical Password Scheme-WIW. (2003), 105–111 pages.

[29] 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 (CHI '16)*.

[30] Volker Roth, Kai Richter, and Rene Freidinger. 2004. A PIN-entry method resilient against shoulder surfing. In *Proceedings of the 11th ACM conference on Computer and communications security*. ACM, 236–245.

[31] Alireza Sahami Shirazi, Peyman Moghadam, Hamed Ketabdar, and Albrecht Schmidt. 2012. Assessing the vulnerability of magnetic gestural authentication to video-based shoulder surfing attacks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2045–2048.

[32] 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 (MUM '12)*. ACM, New York, NY, USA, Article 13, 10 pages. https://doi.org/10.1145/2406367.2406384

[33] Florian Schaub, Marcel Walch, Bastian Könings, and Michael Weber. 2013. Exploring the design space of graphical passwords on smartphones. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*. ACM, 11.

[34] Michael Sherman, Gradeigh Clark, Yulong Yang, Shridatt Sugrim, Arttu Modig, Janne Lindqvist, Antti Oulasvirta, and Teemu Roos. 2014. User-generated free-form gestures for authentication: Security and memorability. In *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*.

ACM, 176–189.

[35] Stephen J Tipton, Daniel J White II, Christopher Sershon, and Young B Choi. 2014. iOS security and privacy: Authentication methods, permissions, and potential pitfalls with touch id. *International Journal of Computer and Information Technology* 3, 03 (2014).

[36] 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 (CCS '13)*. ACM, New York, NY, USA, 161–172. https://doi.org/10.1145/2508859.2516700

[37] 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.

[38] Emanuel Von Zezschwitz, Alexander De Luca, Bruno Brunkow, and Heinrich Hussmann. 2015. SwiPIN: Fast and secure pin-entry on smartphones. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1403–1406.

[39] Emanuel von Zezschwitz, Alexander De Luca, Philipp Janssen, and Heinrich Hussmann. 2015. Easy to Draw, but Hard to Trace?: On the Observability of Grid-based (Un)Lock Patterns. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2339–2342. https://doi.org/10.1145/2702123.2702202

[40] Emanuel von Zezschwitz, Paul Dunphy, and Alexander De Luca. 2013. Patterns in the Wild: A Field Study of the Usability of Pattern and PIN-based Authentication on Mobile Devices. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '13)*. 261–270.

[41] Emanuel von Zezschwitz, Malin Eiband, Daniel Buschek, Sascha Oberhuber, Alexander De Luca, Florian Alt, and Heinrich Hussmann. 2016. On Quantifying the Effective Password Space of Grid-based Unlock Gestures. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia (MUM '16)*. ACM, New York, NY, USA, 201–212. https://doi.org/10.1145/3012709.3012729

[42] Matt Weir, Sudhir Aggarwal, Michael Collins, and Henry Stern. 2010. Testing metrics for password creation policies by attacking large sets of revealed passwords. 162–175.

[43] 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 '06)*. ACM, New York, NY, USA, 177–184. https://doi.org/10.1145/1133265.1133303

[44] Oliver Wiese and Volker Roth. 2015. Pitfalls of Shoulder Surfing Studies. In *NDSS Workshop on Usable Security*. 1–6.

[45] Nur Haryani Zakaria, David Griffiths, Sacha Brostoff, and Jeff Yan. 2011. Shoulder surfing defence for recall-based graphical passwords. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*. ACM, 6.

[46] Ziming Zhao, Gail-Joon Ahn, Jeong-Jin Seo, and Hongxin Hu. 2013. On the security of picture gesture authentication. In *22nd USENIX Security Symposium (USENIX Security 13)*. 383–398.

A   SURVEY ADVERTISEMENT

We are conducting an academic survey about shoulder surfing on mobile device authentication mechanisms. We would like you to act as an attacker attempting to get someone's mobile device password by observing videos of a user authenticating into a mobile device. If you are currently viewing this page on a mobile device (ie. cell phone or tablet), please switch to a desktop or laptop computer to take this survey. If you get to the survey and it detects a mobile device, you will be opted out of the survey. Please select the link below to complete the survey. At the end of the survey, you will receive a code to enter into the submission form below to receive credit for taking our survey.

THE SURVEY WILL ONLY WORK IF YOU VIEW IT ON A NON-MOBILE DEVICE COMPUTER.

We have only tested the survey using GOOGLE CHROME OR MOZILLA FIREFOX. If you experience problems, opt out and return the HIT without penalty.

You will be compensated $1.50 for your work. We have found that it takes approximately 10 minutes on average to complete this HIT, for a payout of about $0.15 a minute

Due to the nature of the work, you may only complete the HIT once, even across multiple posting of the HIT. If you accept the HIT and are notified that your work will not be accepted, please return the HIT. FAILURE TO FOLLOW THIS INSTRUCTION MAY RESULT IN WORK BEING EXCLUDED AND/OR A REJECTION.

Please feel free to contact the requester if you have any questions or concerns. A prompt reply should occur within 24 hours or sooner.

Note: this survey requires your browser to load several high quality videos. We do not recommend you attempt this survey if you have a limited data connection.

## B   PATTERNS AND PINS VISUALIZED

B.1   Patterns

Patterns used with properties: a double circle indicates a start point, while a single circle indicates a point included in the pattern. Note that labeling of patterns begins in the upper left with 0, ending in the lower right with 8 (See Figure 1)

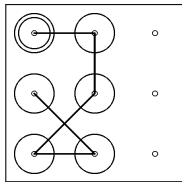0145
up

6745
down

3157
neutral
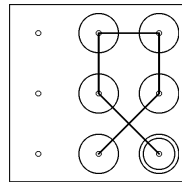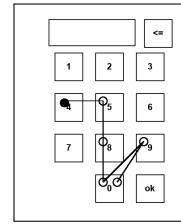
1346
left

4572
right/cross
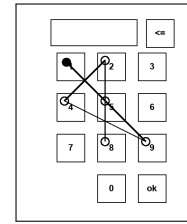
743521
up/non-adj
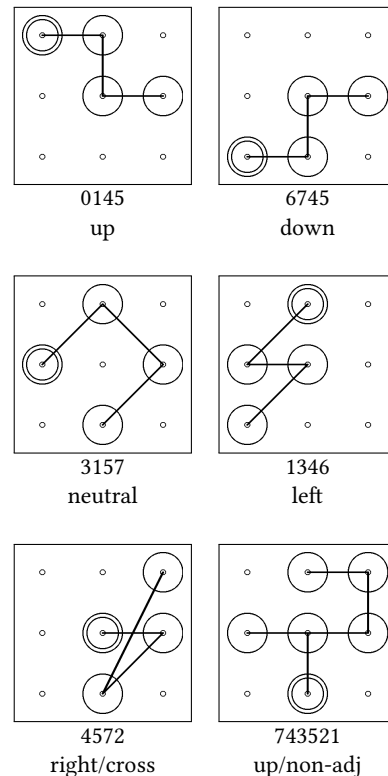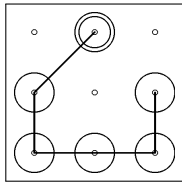
136785
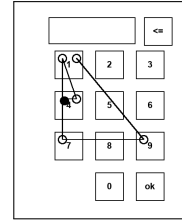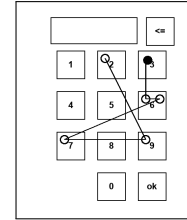down
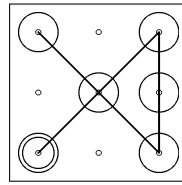
642580
neutral/cross

014673
left

841257
right/kmove/cross
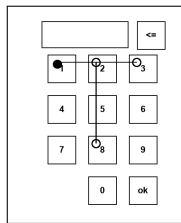
458090
down/repeat

159428
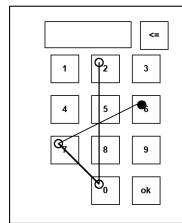neutral/cross/non-adj

441791
left/kmove/repeat

366792
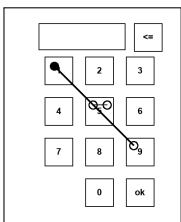right/repeat/kmove/cross

B.2 PINs

PINs used with properties: filled circle is the start point, multiple circles on a number indicate multiple touches.
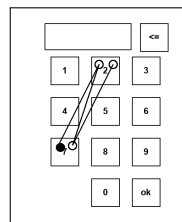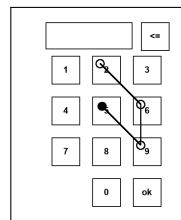
1328
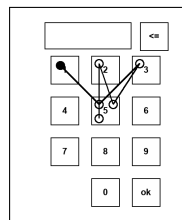up/non-adj

6702
down/kmove/cross

1955
neutral/non-adj/repeats

7272
left/kmoves/repeats

5962
right

152525
up/repat