# A TRIDENT SCHOLAR PROJECT REPORT

NO. 456

---

**Baseline Measurements of Shoulder Surfing Analysis and Comparability for Smartphone Unlock Authentication**

by

Midshipman 1/C John Thomas Davin, USN

---

UNITED STATES NAVAL ACADEMY
ANNAPOLIS, MARYLAND



USNA-1531-2

| **REPORT DOCUMENTATION PAGE** | *Form Approved OMB No. 0704-0188* |
|---|---|



| **1. REPORT DATE** *(DD-MM-YYYY)* | **2. REPORT TYPE** | **3. DATES COVERED** *(From - To)* |
|---|---|---|
| 05-22-17 | | |

| | |
|---|---|
| 4. TITLE AND SUBTITLE<br>Baseline Measurements of Shoulder Surfing Analysis and Comparability for Smartphone Unlock Authentication | **5a. CONTRACT NUMBER** |
| | **5b. GRANT NUMBER** |
| | **5c. PROGRAM ELEMENT NUMBER** |
| **6. AUTHOR(S)**<br>Davin, John Thomas | **5d. PROJECT NUMBER** |
| | **5e. TASK NUMBER** |
| | **5f. WORK UNIT NUMBER** |
| **7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)** | **8. PERFORMING ORGANIZATION REPORT NUMBER** |
| **9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**<br>U.S. Naval Academy<br>Annapolis, MD 21402 | **10. SPONSOR/MONITOR'S ACRONYM(S)** |
| | **11. SPONSOR/MONITOR'S REPORT NUMBER(S)**<br>Trident Scholar Report no. 456 (2017) |

**12. DISTRIBUTION / AVAILABILITY STATEMENT**

This document has been approved for public release; its distribution is UNLIMITED.

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**

In this research, we explore a novel approach to measuring the susceptibility of smartphone unlock authentication to shoulder surfing attacks. We have created a series of video recordings where researchers enter authentication sequences into mobile devices (e.g. PINs, graphical patterns with lines, and graphical patterns without lines) in a controlled setting. These videos are designed to simulate shoulder surfing settings under varied attack conditions. Camera angles have been selected to mimic the locations where observational attacks may take place. Participants have taken the survey and played the role of attackers, viewing video-recorded footage of PIN and graphical pattern authentication input with various camera angles, hand positions, phone sizes, and authentication length and strength. Based on the collected data, there are significant differences in success rates between the different authentication types. For PINs with a single view, the average success rate is 23.04%. The pattern with lines authentication has more than triple the success rate with a single view at 72.44%. The goal of this research is to identify more effective guidance for mobile device users to avoid observational attacks. We also aim to advance the methodologies used to measure the shoulder surfing attack surfaces where baselines of comparisons to preexisting systems (e.g. PINs and patterns) are not standardized. Utilizing the methodology and recordings, other researchers may build upon this approach to analyze future systems and replicate our results.

**15. SUBJECT TERMS**

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

| **16. SECURITY CLASSIFICATION OF:** | | | **17. LIMITATION OF ABSTRACT** | **18. NUMBER OF PAGES** | **19a. NAME OF RESPONSIBLE PERSON** |
|---|---|---|---|---|---|
| **a. REPORT** | **b. ABSTRACT** | **c. THIS PAGE** | | 28 | **19b. TELEPHONE NUMBER** *(include area code)* |

**Standard Form 298 (Rev. 8-98)**
**Prescribed by ANSI Std. Z39.18**

U.S.N.A. --- Trident Scholar project report; no. 456 (2017)

# BASELINE MEASUREMENTS OF SHOULDER SURFING ANALYSIS AND COMPARABILITY FOR SMARTPHONE UNLOCK AUTHENTICATION

by

Midshipman 1/C John Thomas Davin
United States Naval Academy
Annapolis, Maryland

________________________________________

Certification of Adviser Approval

Assistant Professor Adam J. Aviv
Computer Science Department

________________________________________

__________________________

Acceptance for the Trident Scholar Committee

Professor Maria J. Schroeder
Associate Director of Midshipman Research

________________________________________

__________________________

USNA-1531-2

## Abstract

In this research, we explore a novel approach to measuring the susceptibility of smartphone unlock authentication to shoulder surfing attacks. We have created a series of video recordings where researchers enter authentication sequences into mobile devices (e.g. PINs, graphical patterns with lines, and graphical patterns without lines) in a controlled setting. These videos are designed to simulate shoulder surfing settings under varied attack conditions. Camera angles have been selected to mimic the locations where observational attacks may take place. Participants have taken the survey and played the role of attackers, viewing video-recorded footage of PIN and graphical pattern authentication input with various camera angles, hand positions, phone sizes, and authentication length and strength. In this study, we recruited 94 midshipmen participants as well as 1164 more respondents via Amazon Mechanical Turk, an online service to recruit survey participants. Based on the collected data, for example, measurements of the success rate of an attack and the recording methodology developed, we provide insight into the factors of mobile unlock authentication which best and least resist shoulder surfing attacks, as well as examine scenarios where weaknesses may occur. There are significant differences in success rates between the different authentication types. For PINs with a single view, the average success rate is 23.04%. The pattern with lines authentication has more than triple the success rate with a single view at 72.44%. The goal of this research is to identify more effective guidance for mobile device users to avoid observational attacks. We also aim to advance the methodologies used to measure the shoulder surfing attack surfaces where baselines of comparisons to preexisting systems (e.g. PINs and patterns) are not standardized. Utilizing the methodology and recordings, other researchers may build upon this approach to analyze future systems and replicate our results.


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

## ACKNOWLEDGEMENTS

My Trident project has been the most challenging yet rewarding work that I have completed here at the Naval Academy.  These challenges helped me to develop both as a student and as a future Marine Corps Officer. I am thankful for the opportunities afforded to me, but I could not have completed this project without the help of many important individuals.

My Trident advisor, Dr. Adam Aviv, has provided me tireless guidance and mentorship for nearly two years. Starting the Trident application in the fall of 2/C year, he has advised me throughout the entire process and has never failed to provide sound direction for my project.

Dr. Ravi Kuber and Flynn Wolf (PhD. candidate) played an integral role in making this project a reality. I would like to thank them for their many hours spent in support of this research.

I would also like to thank the entire Faculty and Staff of the Computer Science Department at the Naval Academy, for without the fundamentals I learned 3/C and 2/C year, this project would not have been possible. I would especially like to thank LCDR Jeff Kenney for his support and patience with the database portion of the project.

Finally, I would like to thank ENS Forrest Cooke, a Trident Scholar from the Class of 2016, for his mentorship throughout my Trident research and my parents, Richard and Rosemary Davin, for spending many hours helping me review applications and reports.

I am very grateful for all the time these individuals have invested in my development.

## Table of contents

# I. INTRODUCTION

## A. Motivation and Overview

Personal and sensitive data is often stored on mobile devices, making these technologies an attractive target for attackers. This has resulted in a heightened focus on the vulnerabilities of mobile unlock authentication, and the susceptibility of these authentication methods to shoulder surfing attacks, or when an attacker directly observes a user authenticating entry in order to acquire a password or other sensitive information from the mobile device [1, 2]. One of the most cited dangers for smartphone unlocking mechanisms are shoulder surfing attacks [3].

Many users utilize biometric authentication as a supplement to the dominant PIN and graphical (stroke-based) pattern password entry mechanisms. However, one study showed that nearly 70% of respondents reported that they utilize either a PIN or Android graphical pattern as their mobile authentication mechanism [4]. While biometrics and other forms of authentication are present, these mechanisms still require a pattern or PIN to utilize the device. Thus, PIN and pattern mechanisms still exist even in the realm of biometrics and, given the opportunity, an attacker will attack the PIN or pattern, not the biometric authentication. Biometrics are also unlikely to ever stand alone as an authentication mechanism due to the concerns of reliability, privacy, security, and ease of use of other technologies [2].

There is much related work that both proposes and studies shoulder surfing resistant authentication mechanisms [3, 4, 5, 6, 7]. We believe our research will further this prior work in providing avenues for new methodologies to test resilience to shoulder surfing attacks compared to conventional PINs or patterns, as a baseline measure. Similar work has utilized cameras to recreate the pattern authentication based on oily residues, or smudges, left on the screen after the user successfully authenticated [8]. Closer to the work we are performing, von Zezschwitz et al. measured the susceptibility of Android's graphical passwords to observation attacks by utilizing simulated observations focusing only on a single dimension, the visibility of the line [9]. These authentications were simulated and single dimensional, while we include multiple dimensions that compare different authentication types and multiple camera angles. Other areas examined include device size, hand position, length of authentication sequence, among others.

Obtaining a user's PIN or pattern may not be very difficult and may not limit an attacker solely to the data stored on the mobile device [7]. In one study, half of the users admitted to choosing PINs based off PINs that they used elsewhere (e.g. bank PINs or physical locks) [4], meaning that a third party may be able to enter multiple systems without the user's knowledge. With regard to difficulty and password strength, graphical passwords suggest trends with respect to easily guessed and non-complex passwords [10, 11]. These studies confirm the need for multidimensional research in the realm of PIN and pattern vulnerability analysis as users suffering from shoulder surfing attacks are exposing themselves to greater risk than the content of their mobile device.

This method of shoulder surfing vulnerability analysis provides a small set of baselines for researchers to utilize. Whether to confirm prior work in the realm of shoulder surfing analysis or to test and compare new mobile authentication systems, this multifaceted approach has the potential to create a standard capable of being replicated.

### B. Hypotheses

H1: The password type (PIN, Pattern with lines, and Pattern without lines) does affect shoulder surfing (SS) susceptibility.
H2: Password shape does affect SS susceptibility.
H3: Single vs multiple views does have a significant difference to SS susceptibility. Multi view is categorized as two views of either the same angle or multi-angle.
H4: The length and strength of the password does affect the SS susceptibility.
H5: The phone size does affect SS susceptibility.
H6: The hand orientation of the user entering the password does affect SS susceptibility.
H7: The angle of observation does affect the SS susceptibility.

## II. Methodology

A within subjects study was designed where participants were exposed to video footage of researchers entering authentication sequences on a mobile device. Participants were asked to view the footage and recreate what they saw to determine the susceptibility of the authentication sequence to observational attacks.

We recorded over 600 videos simulating shoulder surfing in a controlled lab space with the aim to better understand the vulnerabilities of conventional mobile unlocking mechanisms and to study the impact both the user and the environment have upon the attack. These videos, which take into account hand position, phone size, authentication type, and differing camera angles, were compiled into a web-based survey that collected data (e.g. success rates, respondent's biographical information) about the susceptibility to shoulder surfing. We have recruited 1164 responses online via Amazon Mechanical Turk and 94 in person respondents. Amazon Mechanical Turk is a crowdsourcing internet resource that provides an interface between researchers and workers in order to complete human intelligence tasks for compensation. The in person surveys provided a control to compare against the online responses and ensured these respondents provided accurate responses in an uncontrolled setting.

### A. Research Objectives

In examining the videos of simulated shoulder surfing attacks and the data collected from the survey, we hoped to solidify our understanding of the vulnerabilities that PINs and patterns have to observation attacks. Our conclusions not only help identify what type of PINs and patterns are more susceptible to shoulder surfing attacks but also identify environmental factors, user actions, and password features (e.g. length of password) that directly increase or reduce the likelihood of a successful attack. These recordings are also unique in that they provide a public corpus of shoulder surfing attacks for other researchers to study.

## B. Video Recordings

Over 600 videos simulating shoulder surfing attacks were filmed in dedicated lab space at the University of Maryland, Baltimore County (UMBC). These videos account for different authentications, angles, handedness, and cell phone size. The videos have the surroundings grayed to limit distractions and help the respondent focus on the authentication. For reference throughout this report, the tables below have been added to aid in understanding the depth of the variables accounted for in this study. Tables 1-3 below show the variables each respondent was assigned at random at the beginning of the survey. The respondent could only be assigned one variable per table, which are retained throughout the entire survey.

| Name | Phone Model | Dimensions |
|---|---|---|
| Red | Nexus 5x | 5.427” x 2.723” |
| Black | OnePlus One | 6.02” x 2.99” |

**Table 1**: Phones used in experiments, and their short hands - *Red or Black*

| Name | Hand Position |
|---|---|
| Thumb | One Handed |
| Index Finger | Two Handed |

**Table 2**: Phone holding configurations used

| Authentication | Description |
|---|---|
| PIN | Numeric PIN entry |
| Pattern | Android pattern with visible lines |
| No Lines | Android pattern without visible lines |

**Table 3**: Authentication methods being studied

These tables show the independent variables randomly assigned to participants upon start of the survey. These variables were maintained for each respondent throughout the entire survey in order to provide consistency and reduce bias by not confusing the participant with too many variables. The dependent variables that changed with each authentication attempt are shown below in Tables 4-6. Selection of these variables are discussed in later subsections. Every respondent attacked all 10 passwords of either PIN or Pattern while given a random treatment at load time for each attack - shown in Table 5.

| Camera Angle | Description |
|---|---|
| Near Left | Over left shoulder at a height of 5’ |
| Near Right | Over right shoulder at a height of 5’ |
| Far Left | Over left shoulder at a height of 6’ |
| Far Right | Over right shoulder at a height of 6’ |
| Top | Over head at a height of 6’ |

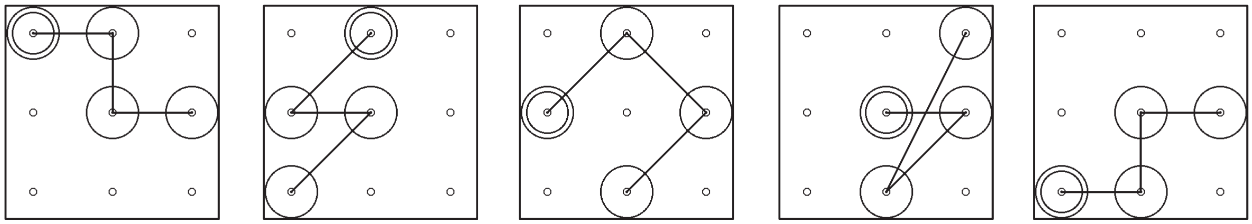
**Table 4**: Camera Locations

| Treatment | Views | Attempts |
|---|---|---|
| A | One | One |
| B | One | Two |
| C | Two | One |
| D | Two (different angles) | One |
| E | Two (different angles) | Two |

**Table 5**: Five different treatments

| 4-Length PINs | Properties | 6-Length PINs | Properties |
|---|---|---|---|
| 1328 | Up Shift | 153525 | Up Shift |
| 1955 | Neutral | 159428 | Neutral Cross |
| 5962 | Right Shift | 366792 | Right Shift |
| 6702 | Down Shift | 441791 | Left Shift |
| 7272 | Knight | 458090 | Down Shift Cross |
| **4-Length Patterns** | **Properties** | **6-Length Patterns** | **Properties** |
| 0145 | Up Shift | 014673 | Neutral |
| 1346 | Left Shift | 136785 | Down Shift |
| 3157 | Neutral | 642580 | Left Cross |
| 4572 | Right Knight/Cross | 743521 | Up Shift Non-Adjacent |
| 6745 | Down Shift | 841257 | Right Shift |

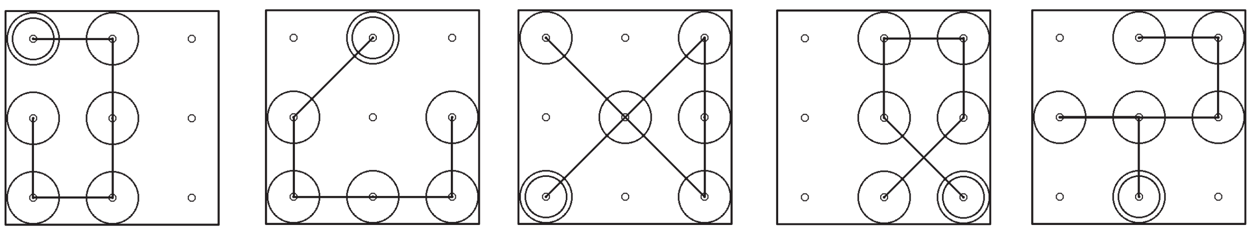
**Table 6**: Ten PIN and Pattern passwords being studied and the properties each one highlights



0145 1346 3157 4572 6745

| | | |
|---|---|---|
| 0 | 1 | 2 |
| 3 | 4 | 5 |
| 6 | 7 | 8 |

Number to Pattern Grid



014673 136785 642580 841257 743521

**Table 6 (continued)**: Pattern Passwords as Displayed on a Mobile Device

### *1) Angles*

The camera setup was an integral part of design. It was important to ensure angles on the left and right accurately mirror each other in distance to the screen and orientation. Figure 1 below shows the setup. As displayed in Table 4, the videos were recorded from five angles: far left, far right, near left, near right, and top. There are five total angles we studied and thus five GoPro cameras were used in the experiment. The two lower GoPros, angles near left and near right, were 5’ high and 2.5’ apart, angled inward at 45 degrees. The actor sat in the center of the setup, holding the phone 3’ high. The two outer second tier of cameras, far right and far left, stood directly 1’ above the lower ones. The fifth camera, directly overhead, was at the same height as the two higher ones centered between them and is referred to as the top angle. Previous work has only touched on three of these angles and allowed the participants to choose the password [12]. Of the numerous research questions we aimed to answer, these angles answer the environmental question whether or not there is an optimal angle for shoulder surfing. Screen shots of these angles are shown in Figure 2.

**Figure 1**: GoPro Camera Array: lower cameras are near, higher cameras are far, and the middle camera is top

Far Left | Far Right | Near Left | Near Right | Top

**Figure 2**: Camera Views

*2) Hand Position*

Referenced in Table 2, hand position was an important factor included in our survey. Based on the initial review of the videos, we hypothesized that videos with a user authenticating with one hand, using their thumb only, would be more difficult to shoulder surf than a user utilizing two hands, one hand to hold the phone and the index finger of the other to authenticate. We came to this projected conclusion based on the partial screen obstruction the one handed user caused to shoulder surfers attacking from a side angle. Figure 3 demonstrates the subtle differences based on hand position.

**Figure 3**: Thumb hand position vs. Index finger hand position

*3) Phone Size*

Similar to the way a user authenticates, the choices users make in regard to phone size may also have an impact on a shoulder surfer's ability to successfully attack an authentication. We sought to answer the question whether larger screens significantly increased the vulnerability of a given authentication mechanism. Referenced in Table 1, the larger phone, the black OnePlus One

(6.02 x 2.99 in), is comparable in size to more popular phones like the iPhone 6s Plus. The smaller phone, the red Nexus 5x (5.427 x 2.723 in), is comparable in size to the iPhone 6s.

*4) Password Selection*

There are ten PIN passwords and ten pattern passwords, described in Table 6, that were the subject of the study. These passwords were selected from real world data [10]. The PINs were selected from the RockYou.com data set, a list of leaked passwords from a 2009 security breach that is now utilized in numerous password studies [13, 14]. The patterns were selected from prior research in which users self-reported their pattern password [10]. These passwords each contain distinct properties that are modeled in both the PINs and the patterns. They formed a representative, albeit a small sample, of real world passwords. The properties that these passwords encompass are described in Table 6 to the right of each password. These features have also been identified and studied in prior work [10, 15]. As we are concerned with observational attacks, these properties are visual in nature, e.g. left shifted, right shifted, containing non-adjacency contact points, crosses, and repetitions. A left shift indicates that a majority of the password is on the left side of the screen whereas a right shift indicates the majority of the PIN digits or pattern points are on the right side of the screen. Non-adjacency means that the contacted points are not next to each other and a cross is when the PIN or pattern crosses back over itself. A cross in a pattern is demonstrated below in Figure 4 with the pattern 841257. An example of a PIN that features repetition in our data set is the PIN 7272. Patterns can be represented with numbers when the dots are replaced with numbers. Figure 4 also demonstrates how to draw pattern 841257 given in the form of numbers, starting with the green 8 and ending with the red 7. These properties will help answer the research question whether or not certain PINs or patterns are more vulnerable to shoulder surfing solely on their placement and order on the screen. Figure 5 shows the different authentication screens as seen in the survey.

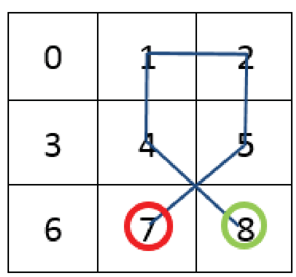



**Figure 4**: Converting Numerical Password to Graphical

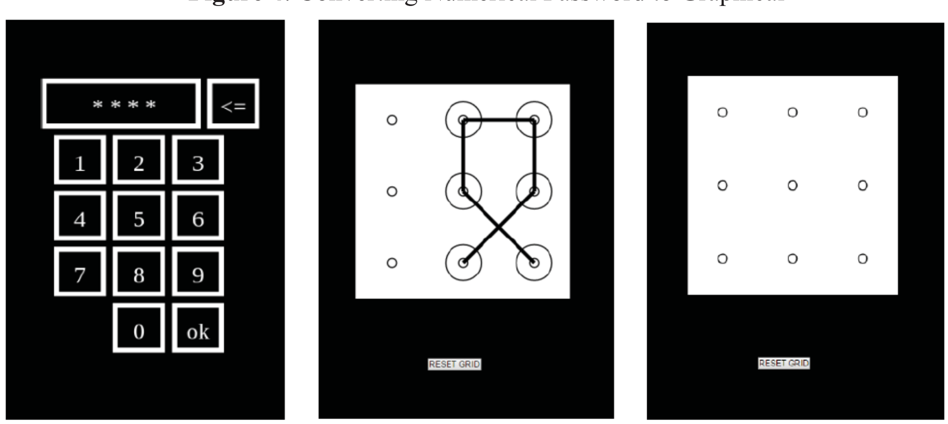



**Figure 5**: PIN, Pattern, and Pattern without lines authentication mechanisms

### C. Realism and Limitations

In accounting for and analyzing all the variables described above, there were some limitations and unintended variables that were not included in the study. Examples include the environmental and situational considerations which have been used to evaluate mobile interfaces [16]. Glare was not addressed because it makes the simulations more realistic given that shoulder surfers cannot control glare on the victim's screen. In addition, the position we had the user sit provided a highly controlled setting. Text-based passwords were not included because of their limited use for mobile authentication in the wild and difficulty selecting similar strength passwords compared to PIN and pattern. Lastly, we acknowledge that the small sample of passwords is a subset of all possible patterns/PINs and the properties they contain.

## III. Website and Database

Integration of the survey with the website and backend database was an integral part of data collection. When a respondent accessed the main page, they were prompted for an authentication code to continue on to the Institution Review Board consent form, the interactive training, and then the survey. For the in person survey administration, the authentication codes were assigned randomly as the people entered the lab room for the survey. For online respondents, authentication codes were generated when the person accepted the terms via Amazon Mechanical Turk.

### A. Institution Review Board (IRB)

Upon entering the authentication code, each respondent read and consented to the USNA IRB statement. This statement made the respondent aware of their rights as an individual taking the survey and ensured that their data and any identifiable information about them is safeguarded and not made public. If a respondent declined to consent, they were opted out of the survey and did not participate.

### B. Survey – Initialization

The initialization of the survey was critical to ensuring that the respondent followed instructions and that the data collection was balanced for a thorough analysis. Once the respondent started the survey, they were randomly assigned three variables that remained constant for the entire survey – Phone (Table 1), Handedness (Table 2), and Authentication Type (Table 3). In order to ensure the respondent followed the instructions that did not allow the survey to be taken on a mobile device or tablet, the survey at this stage tested for those conditions and did not allow them to proceed if they were not abiding. Similarly, we requested that the respondent maximize their screen. To make sure this was taking place, the resolution of each respondent's screen was recorded. If their resolution was too small, they were either making the browser small thus indicating they were not giving their full attention to the survey or their monitor was very small. In either case, we omitted this data from the set. The respondent was also asked to report their sex, age, eye sight, and skill level with modern cell phone technologies. The eye sight question had four options: normal, corrected with glasses/contacts, deficient and not corrected, and not sure. The skill level question provided the options: none, below average, average, above average, and professional. The biographical information not only helped in screening accurate data but also provided insight into the characteristics of attackers which appeared more adept at observational attacks.

### *C. Survey – Training*

After accepting the terms and conditions of the IRB consent page and entering the biographical information, each respondent went through a video step-by-step tutorial. This tutorial video could be replayed as many times as the respondent needed. The left portion of Figure 6 below shows part of the tutorial video. Following the tutorial video, each respondent was given an interactive tutorial in which they shoulder surfed a simple authentication (e.g. 1234 for a PIN) and then were sent to a recreation page in which they were expected to enter the password; the same task they would perform ten times throughout the survey. The right portion of Figure 6 below shows the recreation page for a respondent tasked with PINs. All respondents with the same authentication type received the same training, thus creating a baseline. To ensure a thorough understanding of the task, each respondent could repeat the second phase of training as many times as needed before beginning the survey.

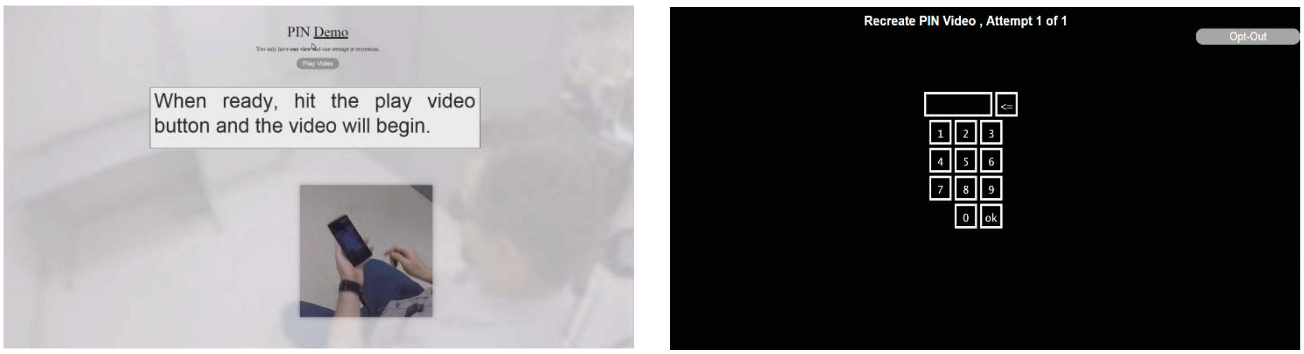



**Figure 6**: Tutorial Video and Recreation Page

### *D. Survey – Randomization*

Effectively randomizing the order and the type of videos was crucial to avoid the introduction of bias or any negative impact on data collection. When the respondent finished training and began the survey of the ten authentications, the website tracked the respondent's assigned phone (red Nexus 5x or black OnePlus One), finger (index or thumb), and authentication type (PIN, pattern with lines, pattern without lines). To ensure balanced, yet still random, data collection, the website utilized a shuffle function on three lists to grab the password, angle, and treatment for each trial. Figure 7 shows a flow chart of how the randomization operated. Since each respondent shoulder surfed ten passwords, the list of treatments (five treatment options - shown previously in Table 5) was doubled upon initialization to match the number of passwords being attacked. Since each participant attacked ten different passwords and there were only five treatments, this explains why each person who took the survey encountered each treatment twice throughout the survey. Even with the best randomization functions, some videos were underrepresented in the data at the end of collection. To address this near the end of data collection, the random function selecting angles (shown as the leftmost box below) was replaced with a function that calculated which angles were underrepresented and filled those in order to better balance the data set.

*Shuffle Angle List*
-Grab angle from Table 4
-Remove angle from list

➤

*Shuffle Treatment List*
-Grab treatment from Table 5
-Remove treatment from list

➤

*Shuffle Password List*
-Grab password from Table 6
-Remove password from list

**Figure 7**: Individual Shoulder Surfing Page Randomization Flow Chart

### E. Integration of Website and Database

Linking the website and the database via PHP and MySQL was critical to ensure both the full functionality and security of the website. Throughout each part of the survey, the database was consulted to ensure the respondent took the survey in the correct order and did not manipulate the website in any way. The backend database ensured that each respondent was not on any pages they were not supposed to be and limited the video views per each authentication according to the treatment assigned. It was also critical to keep the database design in order to make survey operations less time consuming and to ensure it could handle many users at once. Taking the load capacity of the server into account, we determined our threshold to be around 25-30 people concurrently taking the survey via test trials on Amazon Mechanical Turk. Figure 8 below is an Entity Relationship (ER) Diagram of the database that demonstrates how this complex survey can be managed on the backend by a relatively lightweight database.

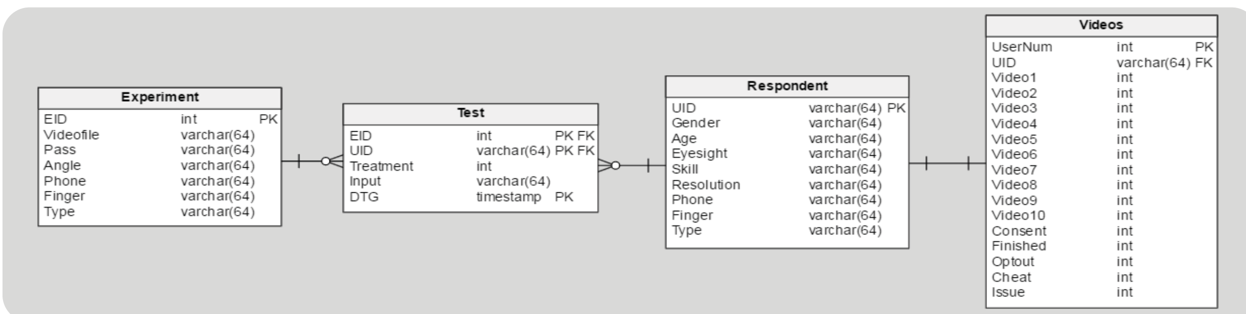



**Figure 8**: Shoulder Surfing Database ER Diagram

## IV. DATA COLLECTION

The survey and website were prototyped and tested in a lab environment with volunteers taking the survey to ensure everything ran properly. The test runs resulted in minor changes to the survey that both increased the accessibility throughout the website and decreased a few areas of confusion by simplifying actions expected of the participant. The data collected during this phase was scrubbed and was not included in the results.

### A. In Person Data Collection

Following the test runs, the survey was run in a controlled lab environment in which midshipmen took the survey. Upon entering the room, the midshipmen were instructed to log in to a computer and await instructions. Once everyone was ready, they were all read a script that mirrors the instructions provided to online participants in the survey.

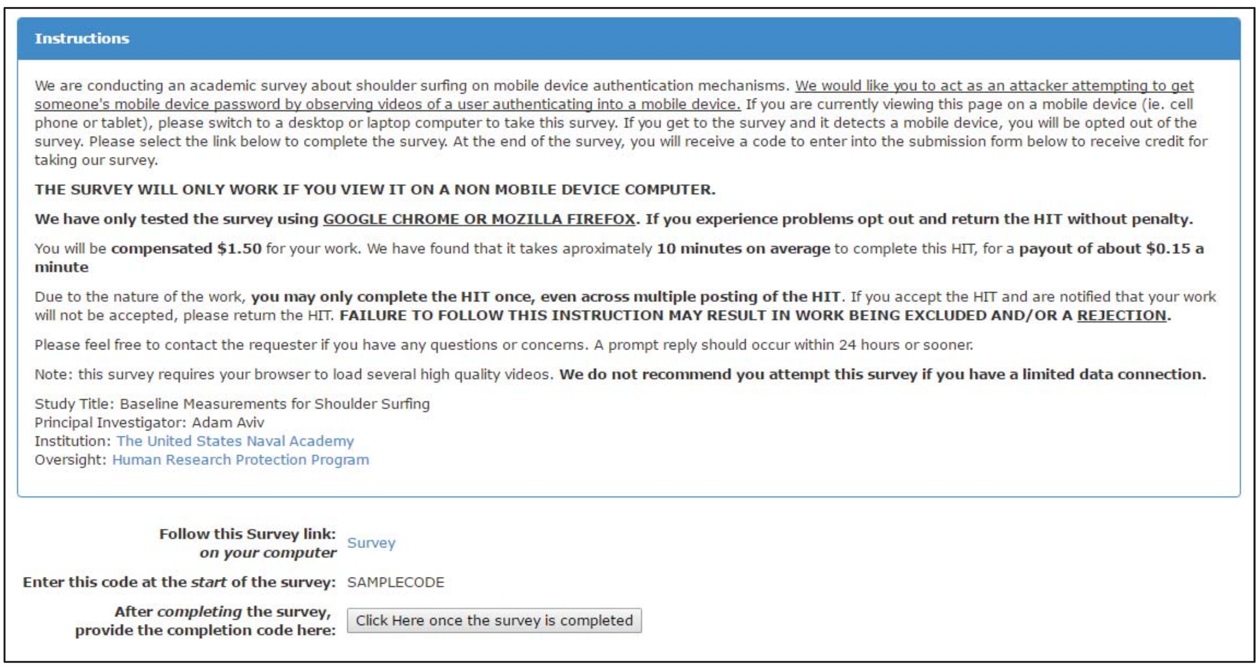
The script read: “Welcome to the Shoulder Surfing Survey. We are conducting an academic study about shoulder surfing on mobile device authentication mechanisms. We would like you to act as an attacker attempting to get someone's mobile device password by observing videos of a user authenticating into a mobile device. Some of you will be working with traditional PINs, some of you will be using Android graphical patterns. While PINs may be familiar, Android patterns may not be for some of you. Android patterns have the following rules: (1) They must contact at least 4 points without lifting; (2) intermediate points cannot be avoided by going outside the grid;

(3) points cannot be contacted twice. We aim to provide insight into the factors of mobile unlock authentication which best and least resist shoulder surfing attacks and examine scenarios where weaknesses may occur. The goal is to identify more effective guidance for mobile device users to avoid observational attacks. Please sign the last page of the consent form. You will read the consent form as a part of the online survey. Please use Google Chrome to access the following site-(URL given here). Do not have any other tabs open. Do not distract other users by talking or making any noise. Thoroughly read instructions. Once you have completed all the tasks, please stay seated and do not leave until the end of the advertised session."

Following the end of each session, the participants were asked to not discuss the survey with anyone in order to eliminate the introduction of bias to future participants.

### *B. Online Data Collection*

Following multiple successful runs of in person data collection, the survey was made available to online participants via Amazon Mechanical Turk. To reiterate, Amazon Mechanical Turk is an online service that connects researchers and online workers to fulfill online human intelligence tasks. Similar to the script, each survey respondent had to read the following prompt before accepting the task and proceeding to the survey. Figure 9 below is a screenshot of the prompt.



**Instructions**

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 aproximately **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.**

Study Title: Baseline Measurements for Shoulder Surfing
Principal Investigator: Adam Aviv
Institution: The United States Naval Academy
Oversight: Human Research Protection Program

**Follow this Survey link: *on your computer*** Survey

**Enter this code at the *start* of the survey:** SAMPLECODE

**After *completing* the survey, provide the completion code here:** Click Here once the survey is completed

**Figure 9**: Amazon Mechanical Turk Prompt Page

*C. Validation of Online Results*

In order to confirm the data collected online as a valid set of data, the in person results were compared against the data collected online via Amazon Mechanical Turk. In these comparisons, we focused on the difference between the two data sets or the delta. Figure 10 shows the comparison of single view treatments comparing the two data sets. The deltas are: PIN - 13.85%, Pattern without lines (NPAT) - 15.78%, and Pattern (PAT) - 24.25%. The average delta for this overall comparison is 17.96%.

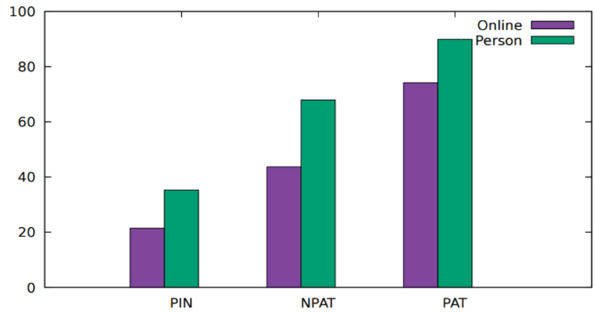



**Figure 10**: All Authentication Method Success Rate - Single View Comparison

Similar comparisons for single view treatments were done against the two hand orientations, Thumb (one hand) and Index Finger (two hands), and against the two phone sizes, Red (smaller phone) and Black (larger phone). The deltas for the hand orientation are: Thumb - 21.24%, Index - 15.49%. The phone size deltas are: Red - 20.91%, Black - 15.38%. These deltas are nearly identical. The visual representation of these rates are shown in Figures 11 and 12.

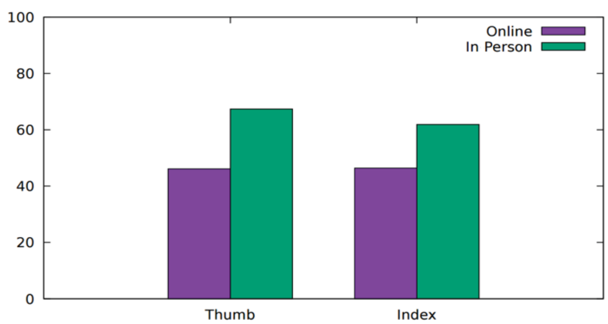



**Figure 11**: Thumb vs. Index Single View Success Rate

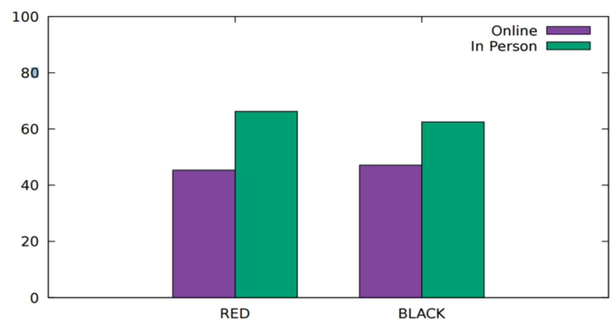



**Figure 12**: Red vs. Black Single View Success Rate

The differences in the success rates between the two groups can be attributed to many factors. One difference is that the effort one puts into a task is typically higher when there is a person supervising. This known bias contributed to the higher success rates among the in person participants. Another reason the success rates are off by a similar factor between the two data sets is the demographic differences between college students and the random composition of people taking the survey online. The average age of the online participant was 34 whereas the average age of the in person sample was 21. Similarly, the distribution of sex was close to 50/50 in the online study but was 70/30 in favor of males in the in person sample. Lastly, the physical computer setup one takes the survey on may have contributed to the difference. The screen resolution for the in person surveys was 990x1840 whereas the average screen resolution for online participants was 820x1508 with a standard deviation of 170 pixels for width and 290 pixels for height. These resolutions indicate that most of the online participants had smaller screens than the in person respondents. The similar trends between these two data sets validates the online data, which is the data utilized in the final results described below.

## V. Results

*A. H1: The password type (PIN, Pattern with lines, and Pattern without lines) does affect shoulder surfing (SS) susceptibility.*

The main comparisons done with the final results are composed of comprehensive comparisons between single view and multiple view treatments. The following graphs represent differences between treatments A and B from Table 5, or single view, vs treatments C, D, and E, or multiple view variants. Figure 13 below shows the broadest results of the research, the success rates comparing the three authentication mechanisms. These results confirm hypothesis one: The password type (PIN, Pattern with lines, and Pattern without lines) does affect shoulder surfing (SS) susceptibility.

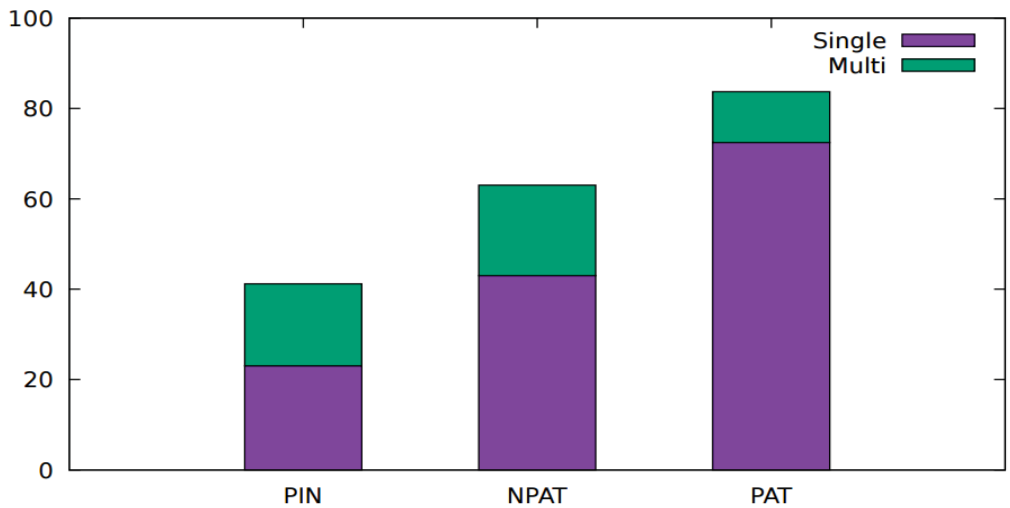



**Figure 13**: Overall Authentication Success Rate Comparison

This graph demonstrates that PAT, or the pattern with lines authentication mechanism, is the most vulnerable mechanism to shoulder surfing with an average success rate of 72% in the single view category. The traditional PIN authentication is the most resilient to shoulder surfing with an average single view success rate of 23%. NPAT, or pattern with lines, is in between these two mechanisms with an average single view success rate of 54%. In all three authentication mechanisms, the multiple view treatments significantly increase the success rate. The success rate increases for multiple views for PIN, NPAT, and PAT are 18%, 20%, and 12%, respectively. Figure 14 below shows a breakdown of the PIN passwords and the success rate per password.

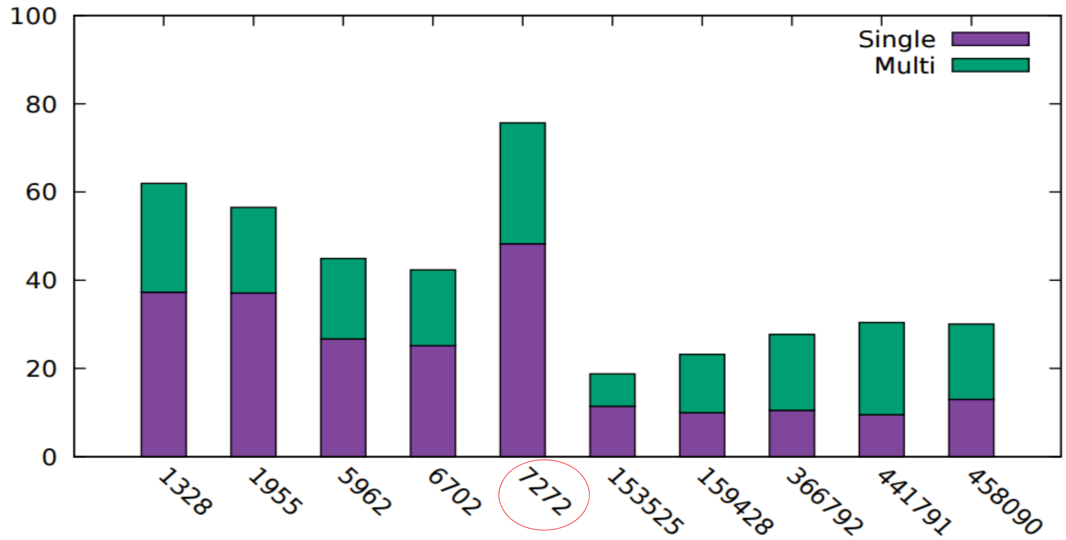



**Figure 14**: PIN Passwords Success Rate Comparison

The most significant cause for differences in success rate for PINs, as demonstrated by the stark contrast in Figure 14, is directly correlated to the length and strength of the PIN. The stronger PINs that are of length six are drastically lower in both the single view and multiple view success rates. Additionally the outlier in this data set, 7272 (circled in Figure 14), demonstrates one of the weaknesses of certain types of password features. 7272 represents PINs with the feature of repetition and the data demonstrates that this feature can nearly double the susceptibility to observation attacks in both single and multiple view treatments.

*B. H2: Password shape does affect SS susceptibility.*

Patterns without lines, the authentication type between PINs and Pattern with lines in terms of susceptibility, shows an interesting breakdown by pattern. Figure 15 breaks down the success rate by password.

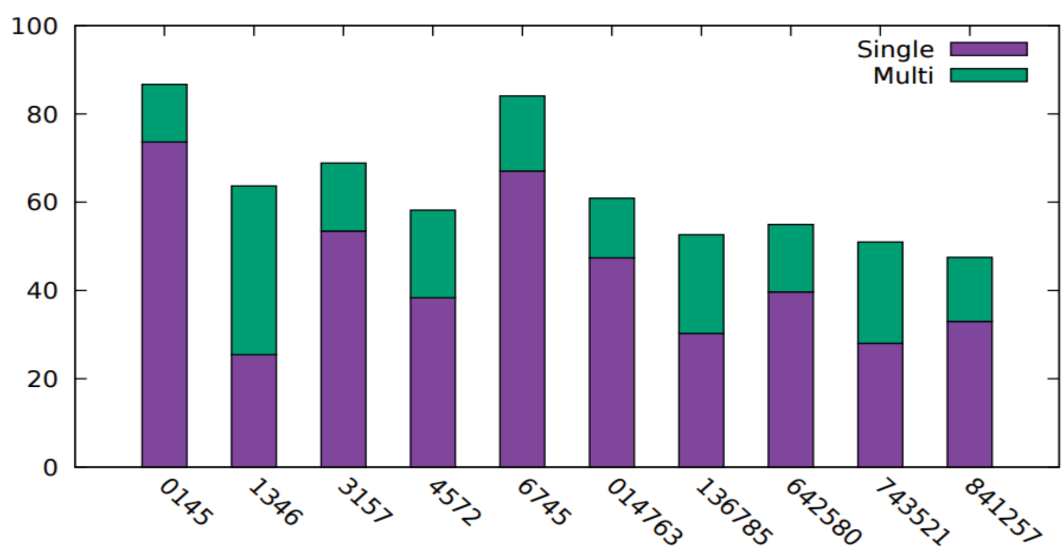



**Figure 15**: NPAT, Pattern Without Lines Pattern Success Rate Comparison

Patterns that did not match the general trends of similar data were looked at more closely to determine why they had significantly lower averages. Pattern 1346, shown in Figure 16 was a pattern that had a much lower single view success rate even when compared to the average single view success rate of length six patterns.

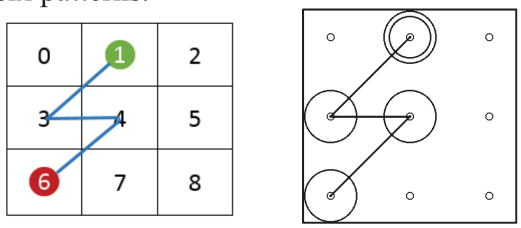



**Figure 16**: NPAT Pattern 1346 – Correctly Entered

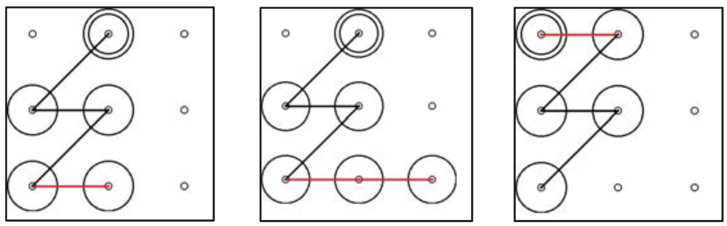


**Figure 17**: NPAT Incorrect Patterns Entered for 1346: 13467, 134678, 01346, Respectively

The reason this pattern is significantly more difficult to shoulder surf is demonstrated in the data of all the incorrect patterns entered. The pattern without lines authentication mechanism results in confusion by the attacker as to when the attacker starts and/or finishes the pattern. The incorrect patterns entered in Figure 17 represent the top three incorrect patterns entered for the 1346 pattern. 13467 represents 43% of the incorrect patterns entered while 134678 and 01346 represent 28% and 10% respectively. It is important to note that although this pattern had the lowest success rate, it is likely that it could still be guessed after a few initial failed attempts because the incorrect entries are so close.

The pattern with lines authentication method, which has been proven to be the most vulnerable, intuitively shows the highest average success rates among single view and multiple view treatments. Figure 18 shows the pattern with lines success rate per password.

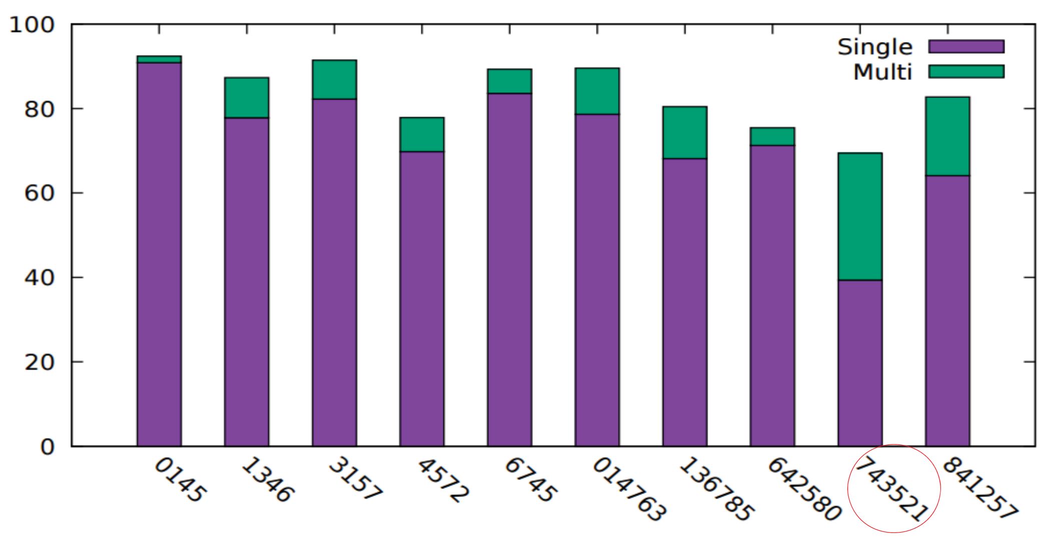



**Figure 18**: PAT, Pattern With Lines Pattern Success Rate Comparison

Similar to the analysis done on the pattern 1346 in the NPAT authentication, the PAT authentication pattern of 743521 (circled in Figure 18) also shows a similar issue. In this particular pattern, the most common mistaken pattern was not an overshoot of the original but rather an undershoot: 74352. Of all the incorrect patterns entered for the pattern 743521, 85% of participants responded with the pattern 74352. The commonality between the two patterns with anomalies, 1346 and 743521, is that they both are only affected significantly in the single view treatment category. Both multiple view treatments are in an acceptable range among each's respective groups of patterns. These patterns accept the second hypothesis that password shape does affect SS susceptibility.

*C. H3: Single vs multiple views does have a significant difference to SS susceptibility.*

In all three authentication mechanisms, the amount of views affected the success rate. Across the board, providing multiple views vs a single view increased the likelihood of an attacker obtaining the password. For the PIN authentication mechanism, the average success rate for a single view was 22.9% whereas the average success rate for multiple views was 41.2%, an 18.3% increase. For the NPAT authentication mechanism, the average success rate for a single view was 43.6% whereas the average success rate for multiple views was 62.8%, a 19.2% increase. For the PAT authentication mechanism, the average success rate for a single view was 80.3% whereas the average success rate for multiple views was 83.6%, a 3.3% increase. The implications of these results indicate that a single view, or possibly an attack from a stranger, is more resilient to a shoulder surfing attack versus someone (e.g. a family member, coworker) that has been exposed to multiple views. These findings confirm the third hypothesis: Single vs multiple views does have a significant difference to SS susceptibility.

*D. H4: The length and strength of the password does affect the SS susceptibility.*

Increasing the length and complexity of the password decreased the observability and vulnerability to shoulder surfing attacks. For the PIN authentication mechanism, the average success rate for single view length four PINs was 34.9% whereas the average success rate for single view length six PINs was 10.9%, a 24% decrease. For the NPAT authentication mechanism, the average success rate for single view length four patterns was 51.6% whereas the average success rate for single view length six patterns was 35.7%, a 15.9% decrease. Lastly, for the PAT authentication mechanism, the average success rate for single view length four patterns was 80.8% whereas the average success rate for single view length six patterns was 64.3%, a 16.5% decrease. These findings confirm the fourth hypothesis: The length and strength of the password does affect the SS susceptibility.

*E. H5: The phone size does affect SS susceptibility.*

The phone size and its role in determining the susceptibility of a user to a shoulder surfing attack was one of the original hypotheses established in this research. Although the graph appears to indicate that phone size did not play a role, it did impact vulnerability. Figure 19 displays the difference in success rates between the two phones. Red was the smaller phone of dimensions 5.427" x 2.723" and black was the larger phone with dimensions 6.02" x 2.99". The Black phone does have a larger success rate and thus can be deemed more vulnerable to shoulder surfing. There are statistically significant differences between the two phone sizes, just on a smaller magnitude. Thus, these findings accept the fifth hypothesis.

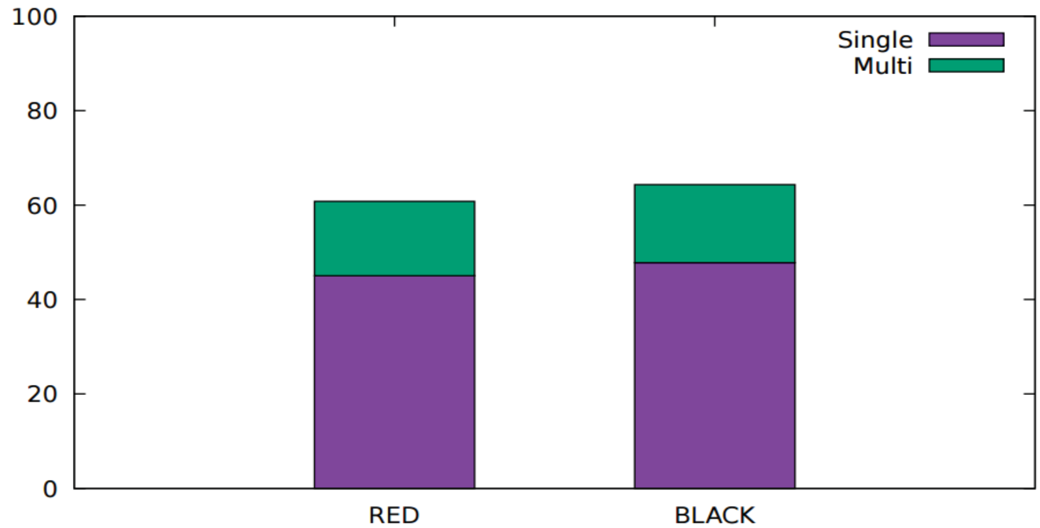



**Figure 19**: Phone Size Success Rate Comparison

*F. H6: The hand orientation of the user entering the password does affect SS susceptibility.*

Similar to the phone size hypothesis, hand orientation or how the user holds the mobile device was anticipated to be a driving factor. The data indicates that there is a slight difference in the two variations and proves the hypothesis that index finger authentication is more vulnerable than just the user using a single hand or their thumb. Index finger hand orientation is more vulnerable due to the fact that the screen is more stable when two hands are being utilized as well as the lack of obstructions to the attacker point of view as compared to the single hand (thumb) orientation. Figure 20 demonstrates the subtle difference.

The chi-squared distribution test in Table 7 illustrates the statistical difference. The p-values indicate that there is a meaningful statistical difference. However, as illustrated in Figure 20, this difference is not very large. These findings confirm the sixth hypothesis: The hand orientation of the user entering the password does affect SS susceptibility.

| | P-Value | χ2 | Percentage |
|---|---|---|---|
| Index | 2.68e-05 | 17.62 | 46.7% |
| Thumb | 8.56e-07 | 24.22 | 46.2% |

**Table 7**: Hand Orientation Overall Chi-Squared Test

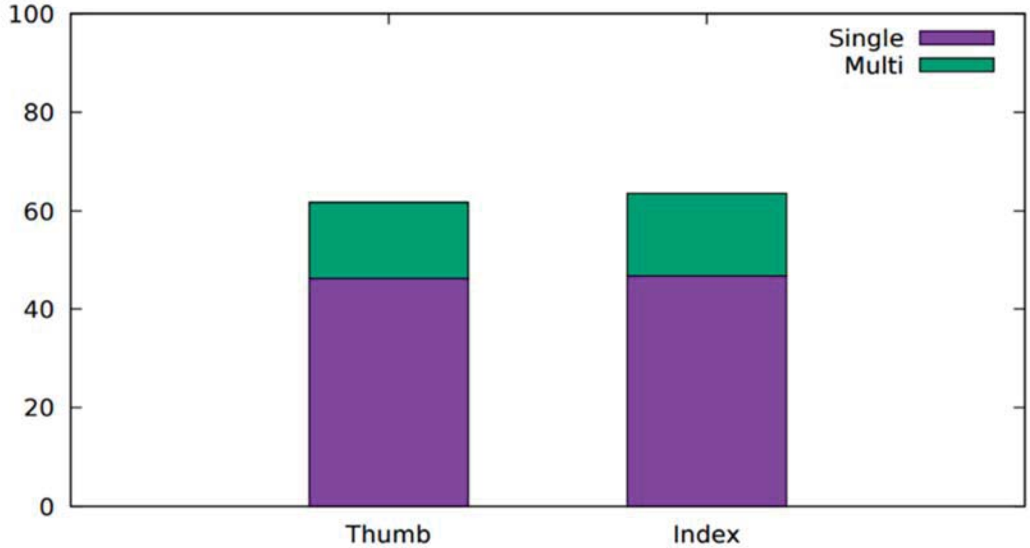



**Figure 20**: Hand Orientation Success Rate Comparison

Due to the similarity of the results, it is appropriate to further breakdown the thumb vs index hypothesis into how each authentication individually performed. Table 8 represents the chi-squared distribution for PINs. The p-values indicate that the results have statistically significant differences. However, the percentages are very close and do not seem to impact the success rate by much. Thus, in the PIN authentication mechanism hand orientation is a factor, but does not appear to be a large determinant of risk.

| | P-Value | χ2 | Percentage |
|---|---|---|---|
| Index | 2.52e-92 | 415.33 | 22.7% |
| Thumb | 6.72e-85 | 381.22 | 23.3% |

**Table 8**: Hand Orientation PINs Chi-Squared Test

The same test was run on the pattern with lines authentication mechanism. Table 9 represents the chi-squared distribution for PAT. Again, the p-values in this table demonstrate that there is a statistical difference between the two hand orientations. The percentages mirror the trend of PINs in that the difference is very minor. Similar to the PIN authentication method, pattern with lines is impacted only slightly by hand orientation.

| | P-Value | χ2 | Percentage |
|---|---|---|---|
| Index | 2.39e-64 | 286.86 | 73.1% |
| Thumb | 2.57e-63 | 282.13 | 71.9% |

**Table 9**: Hand Orientation PAT Chi-Squared Test

This analysis yields different results for the pattern without lines authentication method. Table 10 represents the chi-squared distribution for NPAT. The p-values demonstrate a statistical difference exists. The percentages in this test represent a nearly 5% difference between index and thumb hand orientation. This means that this authentication mechanism is more vulnerable when users utilize the index finger or the two hand method.

| | P-Value | χ2 | Percentage |
|---|---|---|---|
| Index | 3.90e-4 | 12.57 | 45.1% |
| Thumb | 2.22e-11 | 44.76 | 40.9% |

**Table 10**: Hand Orientation NPAT Chi-Squared Test

### G. *H7: The angle of observation does affect the SS susceptibility.*

Table 11 shows the logistical regression run on our data set. This is a binomial logistic regression. The N/A refers to the dependent variables as determined by the algorithm. The shading breaks up the table to demonstrate the dependencies among the data. (e.g. a video has to have either a length four or a length six password, utilize the PIN, PAT, or NPAT authentication mechanism, be from one of the five angles, etc.) In addition to confirming the results of all the previous claims on each hypothesis, this table also highlights that the angle of observation does affect the ability to shoulder surf a password, thus accepting hypothesis seven.

The p-values are important to note in this table. The p-values that are < .05 are statistically significant and indicate that the coefficients or estimates (as referred to in Table 11) are accurate. For a length four password as indicated by the estimate coefficient in the length 4 row, susceptibility is associated with a 95% increase. For the PIN authentication mechanism, the estimate coefficient indicates a 91% decrease in susceptibility. On the contrary, PAT or the pattern with lines authentication mechanism demonstrates a 128% increase in susceptibility. In terms of angle, the far left angle demonstrates a 19% decrease in susceptibility and the top angle demonstrates roughly a 10% increase in susceptibility. Similarly, the far right angle has a negative correlation to susceptibility but is not statistically significant due to the poor p-value. The hand orientation of thumb or one handed authentication demonstrates a 10% decrease in susceptibility. The Red smaller phone is also associated with a 16% decrease in shoulder surfing susceptibility. Lastly, the single view treatment demonstrates an 80% decrease in susceptibility versus the multiple view treatment.

| | Estimate | Std. Error | P-Value |
|---|---|---|---|
| (Intercept) | 0.174415 | 0.055580 | 0.001701 |
| Length 4 | 0.954239 | 0.035766 | < 2e-16 |
| Length 6 | N/A | N/A | N/A |
| PIN | -0.919947 | 0.041751 | < 2e-16 |
| PAT | 1.286946 | 0.044350 | < 2e-16 |
| NPAT | N/A | N/A | N/A |
| Far Left Angle | -0.190882 | 0.055137 | 0.000536 |
| Far Right Angle | -0.001776 | 0.055279 | 0.974375 |
| Top Angle | 0.104981 | 0.055916 | 0.060454 |
| Near Left Angle | 0.055952 | 0.053203 | 0.292953 |
| Near Right Angle | N/A | N/A | N/A |
| Thumb | -0.108644 | 0.035181 | 0.002014 |
| Index | N/A | N/A | N/A |
| Red Phone | -0.166513 | 0.035184 | 2.22e-06 |
| Black Phone | N/A | N/A | N/A |
| Single View | -0.806324 | 0.035715 | < 2e-16 |
| Multiple Views | N/A | N/A | N/A |

**Table 11**: Binomial Logistic Regression

The analysis for the differences between angles is even more compelling when broken down by each authentication mechanism. Table 12 represents the chi-squared distribution for PINs. The p-values on all of the angles indicate that there are statistically significant differences. The two right angles were close to the same while the top angle was just slightly better. It is interesting to note that the percentages indicate that the most successful angle was near left while far left was the least successful. Near left was roughly 10% more susceptible than the average of all the angles for the PIN authentication. The near left angle is the most successful because the user in all the videos was right handed, creating the best line of sight on the close left side.

| | P-Value | Chi^2 | Percentage |
|---|---|---|---|
| Far Right Angle | 6.32e-45 | 197.79 | 20.4% |
| Far Left Angle | 4.11e-50 | 221.56 | 18.9% |
| Near Right Angle | 1.84e-39 | 172.76 | 21.8% |
| Near Left Angle | 4.64e-16 | 65.94 | 32.5% |
| Top Angle | 1.92e-37 | 163.52 | 22.1% |

**Table 12**: PIN Chi-Squared Test

This same analysis was done with the pattern with lines authentication (PAT) mechanism. Table 13 represents the chi-squared distribution for this data. Similar to the PIN chi-squared table, the p-values for this data are all significant as well. However, the percentages were all so similar with a standard deviation of only 4.13% that it leads us to believe that the angles play far less of a role in the success rate for this authentication mechanism. Although, the angle does not have as much of an impact, it is interesting that the lowest percentage angle was the same as PINs: far left.

| | P-Value | Chi^2 | Percentage |
|---|---|---|---|
| Far Right Angle | 1.84e-22 | 95.06 | 70.3% |
| Far Left Angle | 2.13e-18 | 76.56 | 68.2% |
| Near Right Angle | 2.28e-28 | 122.02 | 73.5% |
| Near Left Angle | 8.16e-25 | 105.80 | 71.5% |
| Top Angle | 5.35e-42 | 184.38 | 79.1% |

**Table 13**: PAT Chi-Squared Test

Lastly, the NPAT authentication mechanism was subjected to the same test. Table 14 represents the chi-squared distribution for this entry method. The results of this analysis were less clear than the prior two. The only angles with statistically significant results in this test were far left, near left, and top. Both right angles were right around the mean of the data set. While near left and top held close percentages, the far left angle fell in line with the prior two chi-squared analyses as the worst angle for the attacker or the most resilient for the user. The far left angle has the lowest percentage and this indicates that across all authentication mechanisms, the most unlikely angle for someone to successfully steal a password is at a distance of a foot or greater above the shoulder and to the left side.

| | P-Value | Chi^2 | Percentage |
|---|---|---|---|
| Far Right Angle | 6.69e-2 | 3.35 | 46.1% |
| Far Left Angle | 1.36e-11 | 45.72 | 35.6% |
| Near Right Angle | 1.01e-1 | 2.68 | 46.5% |
| Near Left Angle | 1.87e-4 | 13.95 | 41.9% |
| Top Angle | 2.74e-2 | 4.86 | 45.1% |

**Table 14**: NPAT Chi-Squared Test

### *H. Comparison of Similar Work*

Previous work has been done on shoulder surfing, in particular on the resiliency pattern passwords have to this type of attack. [9] The setup of this prior experiment is a user watches a pattern digitally drawn on a computer screen once and is tasked to recreate what they saw. This pattern is not a recording like the videos featured in our research and does not have anything blocking the screen (e.g. a hand entering the password). Additionally, these simulations do not incorporate any angles or environmental factors associated with shoulder surfing in the wild. The results for their success rates are in Table 15.

| | Length 4 | Length 6 |
|---|---|---|
| Visible (PAT) | 96.9% | 78.9% |
| Invisible (NPAT) | 87.6% | 54.9% |

**Table 15**: The number of successfully observed patterns with respect to their length and line visibility [9].

Table 16 shows the results of our data collection. The in person data represents the midshipmen that took the survey in a controlled lab environment. The online data represents the data collected via Amazon Mechanical Turk.

| | In Person (length 4) | In Person (length 6) | Online (length 4) | Online (length 6) |
|---|---|---|---|---|
| Visible (PAT) | 94.67% | 85.21% | 80.84% | 64.28% |
| Invisible (NPAT) | 71.42% | 60.98% | 51.58% | 35.67% |

**Table 16**: Success Rates: Visible (PAT-with lines) vs Invisible (NPAT-without lines) Pattern Authentication

How our data compares to this previous work studying shoulder surfing provided a unique analysis opportunity as their data did not include any of the environmental factors and focused solely on the features of the patterns and their memorability. Table 17 below shows the deltas, or differences in success rates between our work and that of the similarly conducted study.

| | In Person (length 4) | In Person (length 6) | Online (length 4) | Online (length 6) |
|---|---|---|---|---|
| Visible (PAT) | 2.23% | 6.31% | 16.06% | 14.62% |
| Invisible (NPAT) | **16.18%** | 6.08% | **36.02%** | 19.23% |

**Table 17**: Success Rates Deltas

The average difference, or the average of the deltas shown in Table 17, in success rate between the in person data set and the previous work was 7.7% while the average difference between the online data set was 21.48%. These average differences are important to note because in both cases, the NPAT length four success rates were much higher than the average difference (16.18%, in bold in Table 17, is 8.48% above the mean for in person success rates and 36.02%, also bolded, is 14.54% above the mean for online). This may indicate that the environmental factors added by our research had the most impact on the pattern without lines authentication with length four patterns. While this does not definitely confirm that claim, the rest of the data and its similarity does help validate our data collected.

## VI. Future Work and Conclusions

### *A. Future Work*

These baseline measurements are valuable in assessing future authentication systems. The trial tested methodology proves to support the multivariable functionality needed to test new authentication mechanisms and draw conclusions as to whether they perform better than the traditional methods.

### *B. Contributions*

This work advances the research methodology of shoulder surfing vulnerability analysis. It creates a baseline for shoulder surfing analysis that is easily suitable for recreation by other researchers. Once our work is confirmed by other researchers recreating the experiment, this proven methodology could then be used to test new authentication mechanisms. These tests would compare results in similarly conducted studies against the conventional authentication mechanisms analyzed in this research. Lastly, another contribution of this research is the corpus of over 600 videos of shoulder surfing that will be made available to other researchers.

### *C. Summary of Accomplishments*

All of the goals of this research were accomplished. The results of the individual hypotheses and the logistical regression provide both the baseline measurements of each variable's vulnerability in addition to providing detailed guidance to the end user to mitigate risk of shoulder surfing attacks.

This research has been accepted to the 2017 ACM Chi Conference on Human Factors in Computing Systems in the category of Late-Breaking Work. This work will be featured in a poster presentation at the conference in Denver, CO and the paper will appear in the CHI Extended Abstracts proceedings.

### *D. Conclusions*

As proven through the graphs representing the raw data and the statistical tests paired to each claim, the results of this research provide valuable structure and information not only to fellow researchers but also to everyday smartphone users. Phone size, angle of attack, and hand orientation all impact the risk or resilience an authentication mechanism has to a shoulder surfing attack. Increasing phone size increases the screen surface area and makes the password the user is entering more visible to an attacker. The two angles that were the most vulnerable to shoulder surfing attacks were the near left and top angle with a right handed user. For the PIN authentication especially, the near left angle had a success rate that was 10% higher than the average of all five angles. Although these factors seem to have a small impact relative to changing the authentication mechanism, there is also an effect the length of the password has on susceptibility. This feature, as explored in each authentication mechanism, has the most impact on PINs (as demonstrated in Figure 14) and the least on patterns with lines.

This research not only confirmed prior work by demonstrating similar trends in the two data sets, but also demonstrated the impact added by environmental variables that more accurately represent this type of attack as a whole. It also demonstrated that these environmental variables had the most impact on the pattern without lines authentication mechanism.

The main purpose of this research was to determine the baseline measurements of shoulder surfing vulnerability in the three main authentication mechanisms. Through detailed and careful analysis, we proved the PIN authentication to be the most resilient to shoulder surfing attacks across the board. The six digit PIN, with a 10.9% average success rate in the single view treatment, is suitable as a baseline for other researchers to strive to beat.

## VII. Works Cited


[1] A. Harbach, A. De Luca and S. Egelman, "The Anatomy of Smartphone Unlocking: A Field Study of Android Lock Screens," in *In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI ’16)*, New York, NY, 2016.

[2] S. Wiedenbeck, J. Waters, L. Sobrado and J. Birget, "Design and Evaluation of a Shoulder-surfing Resistant Graphical Password Scheme," in *In Proceedings of the Working Conference on Advanced Visual Interfaces (AVI ’06)*, New York, NY, 2006.

[3] S. Man, D. Hong and M. Matthews, "A Shoulder-Surfing Resistant Graphical Password Scheme - WIW," in *In Proceedings of the International Conference on Security and Management (SAM '03)*, Las Vegas, NV, 2003.

[4] S. Egelman, S. Jain, R. Portnoff, K. Liao, S. Consolvo and D. Wagner, "Are You Ready to Lock?" in *In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS ’14)*, New York, NY, 2014.

[5] A. Forget, S. Chiasson and R. Biddle, "Shoulder-surfing Resistance with Eye-gaze Entry in Cued-recall Graphical Passwords," in *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’10)*, New York, NY, 2010.

[6] M. Kumar, T. Garfinkel, D. Boneh and T. Winograd, "Reducing Shoulder-surfing by Using Gaze-based Password Entry," in *In Proceedings of the 3rd Symposium on Usable Privacy and Security (SOUPS ’07)*, 2007.

[7] A. De Luca, A. Hang, F. Brudy, C. Lindner and H. Hussmann, "Touch Me Once and I Know It’s You!: Implicit Authentication Based on Touch Screen Patterns," in *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’12)*, New York, NY, 2012.

[8] A. Aviv, K. Gibson, E. Mossop, M. Blaze and J. Smith, "Smudge Attacks on Smartphone Touch Screens," in *In Proceedings of the 2010 Workshop on Offensive Technology*, 2010.

[9] E. von Zezschwitz, A. De Luca, P. Janssen and H. Hussmann, "Easy to Draw, but Hard to Trace?: On the Observability of Grid-based (Un)Lock Patterns," in *In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI ’15)*, New York, NY, 2015.

[10] A. Aviv, D. Budzitowski and R. Kuber, "Is bigger better? Comparing user-generated passwords on 3x3 vs. 4x4 grid sizes for android’s pattern unlock," in *In Proceedings of the 31st Annual Computer Security Applications Conference*, 2015.

[11] S. Ullenbeck, M. Dürmuth, C. Wolf and T. Holz, "Quantifying the Security of Graphical Passwords: The Case of Android Unlock Patterns," in *In Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security (CCS '13)*, Berlin, Germany, 2013.

[12] F. Schaub, R. Deyhle and M. Weber, "Password Entry Usability and Shoulder Surfing Susceptibility on Different Smartphone Platforms," in *In Proceedings of the*

*11th International Conference on Mobile and Ubiquitous Multimedia (MUM '12)*, New York, NY, 2012.

[13] J. Bonneau, "The Science of Guessing: Analyzing an Anonymized Corpus of 70 Million Passwords," in *2012 IEEE Symposium on Security and Privacy*, San Francisco, CA, 2012.

[14] D. Malone and K. Maher, "Investigating the distribution of password choices," in *In Proceedings of the 21st international conference on World Wide Web (WWW '12)*, New York, NY, 2012.

[15] A. Aviv and D. Fichter, "Understanding visual perceptions of usability and security of Android's graphical password pattern," in *Proceedings of the 30th Annual Computer Security Applications Conference*, 2014.

[16] L. Barnard, J. Yi, J. Jacko and A. Sears, "An empirical comparison of use-in-motion evaluation scenarios for mobile computing devices," *International Journal of Human-Computer Studies,* vol. 62, no. 4, pp. 487-520, 2005.

[17] S. Brostoff and M. Sasse, "Are Passfaces More Usable Than Passwords? A Field Trial Investigation," in *People and Computers XIV—Usability or Else!*, Springer London, 2000.