NTNU – Trondheim
Norwegian University of
Science and Technology

"Tell Me Who You Are and I Will Tell You Your Unlock Pattern"

Marte Dybevik Løge

Master of Science in Computer Science
Submission date: July 2015
Supervisor: Lillian Røstad, IDI

Norwegian University of Science and Technology
Department of Computer and Information Science

Abstract

Graphical passwords, like the Android Pattern Lock, are a popular security mechanism for mobile devices. The mechanism was proposed as an alternative to text-based passwords, since psychology studies have recognized that the human brain have a superior memory for remembering and recalling visual information.

This thesis aims to explore the hypothesis that human characteristics influence users' choice of graphical passwords. A collection of 3393 user-created patterns were analysed in order to examine the correlation between people's choice of pattern and their characteristics, like hand size, age, gender and handedness.

This thesis first gives a detailed summary of related research on graphical passwords. Then it shows how an online survey was used for collecting user-selected passwords and information about the respondents. Lastly, the thesis explains how the data was analysed in terms of length and visual complexity in order to gain further insight in users' choice of passwords.

Although the data could not provide significant evidence to accept the hypothesis, the results show that password strength significantly varies between gender, age and IT experience. Additionally, analysis of all the collected patterns shows a significant bias towards the selection of pattern starting position.

Sammendrag

Grafiske passord, som Android Pattern Lock, er en populær sikkerhetsmekanisme for mobile enheter. Mekanismen var foreslått som et alternativ til tekstbaserte passord, siden studier innen psykologi har vist at menneskehjernen er overlegen når det gjelder å huske og å gjenkjenne visuelle inntrykk.

Denne masteroppgaven har som mål å utforske hypotesen som påstår at menneskelige karakteristikker påvirker brukeres valg av passord. En samling av 3393 brukeropprettede passord ble analysert for å undersøke om det finnes en korrelasjon mellom menneskers valg av passord og deres karakteristikker, som håndstørrelse, alder, kjønn og håndpreferanse.

Masteroppgaven gir først en detaljert gjennomgang av relatert forskning om grafiske passord. Deretter viser den hvordan en spørreundersøkelse på internett ble brukt for å samle inn brukervalgte passord sammen med informasjon om innsenderne. Til slutt beskriver oppgaven hvordan lengde og visuell kompleksitet på mønstrene ble analysert for å oppnå en dypere forståelse av brukernes valg av grafiske passord.

Selv om dataene ikke kunne gi signifikante bevis for å akseptere hypotesen, viser resultatene at passordstyrken varierer betydelig mellom kjønn, alder og IT-erfaring. I tillegg viser analyse av alle innsamlede mønstre at det er en skjevfordeling i hvilke noder som blir brukt som startnoder.

Acknowledgements

I would like to thank Per Thorsheim for being my co-supervisor. I do not know anyone so enthusiastic and committed to security as he is. Thank you for all support and advice you have given me. Per have given me opportunities that I would never have experienced if not getting to know him. I am going to miss all late night discussions about passwords and listening to all your crazy ideas.

I Would like to thank my supervisor Lillian Røstad for her guidance and support throughout the project. I appreciate her wanting to supervise this project, without her, this thesis would never have become a reality.

I would like to thank my friends and family for all support. I would like to give a special thank to Agnethe Søraa, Thea Mathisen, Marthe Frogner Bekkevold, and Rolf Erik Lekang for all help and support. I am lucky having such good friends! A special thank to my friend Aleksander Gisvold for telling me about Per Thorsheim's crazy idea; the idea behind all the work in this thesis.

Finally, saving the most important person until the end, I would like to thank Håkon Åmdal for always supporting me when needing it the most.

Contents

List of Figures

List of Tables

Abbrevations

IT = Information Technology

ALP = Android Lock Pattern

RQ = Research Question

NTNU = The Norwegain University of Technology and Science

HTTPS = Hypertext Transfer Protocol Secure

SSL = Secure Socket Layer

OS = Operating Sysyem

PIN = Personal Identification Number

BDAS = Background Draw a Secret

DAS = Draw a Secret

CHC = Convex Hull Click

1 | Introduction

Mobile devices play a significant role in our everyday life. In the last decade, mobile phones have improved in terms of capability, interaction and context of use. Mobile phones are no longer only a tool for simple communication, but also a tool for paying bills, reading email and keeping up with social media. Due to the large amount of sensitive data stored on the devices, there is an increased need for security, which makes mobile authentication an important topic for research.

Screen locks are used as a protection mechanism to prevent sensitive information leakage from mobile devices. Historically, screen locking mechanisms were developed to avoid accidental use, for instance if the device was carried in a pocket. Today, the goal is information protection, and the locks have evolved into mechanisms like PIN codes, fingerprints, pattern locks and passwords. However, the problem with these mechanisms are threefold: First, people find it hard to remember a long and secure password. Second, long, text-based passwords can be troublesome to input on a small touch screen. Third, the use of more complex locking mechanisms means more time spent unlocking the device, which is especially disadvantageous if the device is used frequently.

Passwords are often user-selected secrets which are connected to that user as a person. When creating an alphanumeric password, people tend to use associations to something they know, are, or recognize; passwords are more than just arbitrary words and numbers. Examples of this are people who use their date of birth as their PIN, or their favorite sports team as their password. This kind of predictability makes alphanumeric passwords less secure and illustrates one of the main shortcomings of using these for authentication.

Because of the shortcomings with alphanumeric passwords [25], there is an increased interest in graphical passwords. Graphical passwords were proposed as an alternative to PINs and alphanumeric input because humans in general remember graphical elements better than letters and numbers [12]. This method is a promising alternative to alphanumeric passwords, as it offers better usability and helps the user creating complex passwords that are easy to remember. The smartphone is a well-suited platform for graphical passwords because the touch screen allows for intuitive manipulation of graphical elements. This is easier than typing letters and numbers.

One of the commonly used graphical password mechanisms is the Android Pattern Lock which was introduced on the Android platform by Google in 2008. The Android Pattern Lock enables the user to connect dots in a 3×3 grid, forming a pattern. Compared to PIN codes, which have 10.000 unique combinations, the Android authentication mechanism allows for 389.112 possible combinations. However, this number is only the theoretical password space. In 2013, a research group conducted the first large-scale user study on the Android Pattern Lock where 2900 user-selected patterns were collected and analyzed [39]. They found bias in the pattern-making process and claimed that the password space in practice is less than the theoretical.

However, not many researchers have studied the correlation between human characteristics and choice of patterns. As previously mentioned, PINs are biased towards birth dates and passwords have a higher probability of being the person's favorite sports team. Can any similar behavior be found in graphical passwords? Can your choice of pattern be connected to you as a person?

1.1 Hypothesis and Research Questions

Based on preliminary work [26], the following null(H_0) and alternative(H_1) hypotheses were chosen:

> H_0: `Human properties have no influence on a user's choice of graphical passwords`
>
> H_1: `A user's choice of graphical passwords is influenced by the human properties of the user`

In order to test and further investigate these hypotheses, the following research questions were chosen:

RQ1: Is there a correlation between age and choice of graphical passwords?

RQ2: Is there a correlation between gender and choice of graphical passwords?

RQ3: Is there a correlation between handedness and choice of graphical passwords?

RQ4: Is there a correlation between experience with IT and security and choice of graphical passwords?

RQ5: Is there a correlation between reading orientation and choice of graphical passwords?

RQ6: Is there a correlation between handsize and choice of graphical passwords?

RQ7: Are there any similarities in the choice of graphical passwords in the entire population?

RQ8: Is the choice of graphical passwords determined by its context of use?

The selected human properties stated in RQ1 through RQ6 are closely related to the overall goal; answering the hypotheses. A detailed review of the human characteristics included in the six first research questions is provided in Section 3.3. RQ7 and RQ8 are a result of the selected research design selected in the preliminary work for this dissertation [26]. It may be useful to have an overall understanding of the population to be able to see whether an observation relates to a distinct subgroup or the entire population in general. The data collection will introduce three contexts of using a pattern, making it interesting to see whether the context of use impacts the choice in patterns as described in RQ8.

1.2 Methodology

This section provides an overview of the methods used for conducting this research. The first part contains the literature study, providing an overview of published research and related theory relevant to this research. The second and third part contains the research design and analysis of collected data.

1.2.1 Literature Study

The literature study was performed to place the work in this dissertation in a context of research that has already been published. This study used sources that are considered to be of high quality, as well as being through sufficient reviews and quality controls by an external review board. This study considered *ACM* [1], *IEEE* [23], and *Springer* [33] as highly rated journals in information systems and computing. In addition to journal articles, sources like books and conference papers have been used. Due to lack of quality control of content on web pages, the use of content from web pages was avoided when possible.

Conducting a literature study is challenging work due to the massive amount of literature available. Keywords listed in Table 1.1 were put together with logical operators like `OR` and `AND` to build a query for narrowing the search for literature. Whenever literature satisfying a high level of quality was found, the reference list was further utilized as a source for exploring new and relevant research.

Android	Pattern lock	Graphical password
Passwords	Usability	Security
Authentication	Mobile authentication	Mobile security
Human factors	Psychology	Visual Memory

Table 1.1: Keywords used to narrow the search for literature

When finding literature matching the keywords, a specific order for reading the literature was used for reading the literature to be able to determine its relevance and quality. Firstly, it was preferable to look at the abstract first as it often includes important information about the research objectives, the methods used, and the results. Secondly, if the abstract was promising, the result, discussion, and methodology were studied. Lastly, if the research was very interesting, the whole publication was studied from the start to the end. Table 1.2 is a list of quality criterias that was created as a checklist to be used while reading published research.

#	**Quality Criteria**
QC 1	The research is published in a known digital library, journal or conference.
QC 2	There is a clear statement of the aim of the research.
QC 3	The study is cited by other researchers.
QC 4	There is a clear description of the method used in the study.
QC 5	If the research includes an experiment, user study, or other research strategies, there should be a reasonable sample size used.

Table 1.2: Quality criteria for literature review

1.2.2 Data Collection

In order to answer the hypotheses, an online survey was used for collecting patterns and information about the respondents. The need for large amounts of data makes an online survey a suitable method because it simplifies the data collection process, as well as making it possible to reach people living in different geographical locations. The survey was designed to use a self-selection sampling technique [28], meaning that anyone who wants to participate may answer. The self-selection sampling technique also supports the requirements of anonymity where no overview of respondents exists.

Chapter 3 will present a detailed description of the survey design, including the questions asked, and the structure of the survey.

1.2.3 Data Analysis

Before starting the analysis of the collected data, the data set were preprocessed and validated to obtain results of high quality that can be used to answer the hypotheses. The data were also preprocessed before they were used to avoid including outlines or data including noise in the data. The process of validating and preprocessing the data is further described in Section 5.4.

After obtaining a validated and preprocessed data set, the data are presented as results in Chapter 5. The results examine the patterns created by the different user types as listed in the list of research questions. All the patterns produced by the various user types are analyzed in terms of creation time, length, and visual complexity. For pattern length and visual complexity, a t-test are performed for tesing for significant differences in the created patterns. In addition to creation time, length, and visual complexity, other observations may be presented in the result chapter if interesting results are found.

The validity of the results are evaluated by performing a two-tailed t-test with a significance level of 0.5. The t-test can be used to see if there is a significant difference in the patterns created by the different user types.

1.3 Thesis Structure

Chapter 2: Related Work on Graphical Passwords An introduction to the related work and theory published on graphical passwords and mobile authentication.

Chapter 3: Data Collection Presents the research design in detail and describes the process of collecting data in detail.

Chapter 4: A Detailed Descripton of the Survey Presents the survey application, including the resquirements, desciption of how the survey works and how it looks.

Chapter 5: Results Presents the results observed by analyzing the collected data. The results are presented according to the stated research questions.

Chapter 6: Discussion A discussion of the results found according to the stated research questions. The discussion will further be a basis for the conclusion.

Chapter 7: Conclusion and Future Work Presents the conclusion; acceptance or rejection of the hypotheses. This section will also provide a conclusion of the listed research questions. The section for future work provides suggestions for further work based on the results form this research.

2 Background Theory and Related Work

This chapter looks into the background theory needed for futher reading this thesis, as well as providing an insigh to related work. The chapter starts with Section 2.1, providing an overview of the shortcomings with text-based passwords; the origin of graphical passwords. Section 2.2 looks at graphical passwords from a historical point of view. When did it all start, and where are we now? The section describes and visualises graphical passwords schemes proposed over the past years until today. Section 2.3 are looking at research evaluating different graphical password schemes from usability and security point of view. Section 2.4 looks at graphical passwords focusing on the human aspects of security, as well as introducing some material from psychology. As mobile devices is a crucial part of our daily lives, Section 2.5 are looking into graphical passwords and mobile devices. At last, Section 2.6 looks particularly into one of the most commonly used graphical passwords schemes on mobile devices; the Android Pattern Lock.

2.1 Shortcomings of Text-Based Authentication

User authentication is a central part of security systems. Despite the extensive number of options for authentication, text-based passwords remain the most common authentication scheme. Text-based authentication is the authentication scheme widely adopted because it is easy and inexpensive to implement, and users are familiar with the scheme. Text-based authentication also avoids the privacy issues raised by the use of biometric authentication, as well as preventing the need for a physical security device used in token-based authentication schemes. However, text-based authentication suffers from both security and usability disadvantages. As users need to remember an increasing number of passwords, users adopt bad password habits. The term *habit* is often a bad thing when talking about security. A habit is often hard to change and is often predictable because it tends to occur in similar situations repeatedly.

Password reuse is one of the known password habits among users, caused by human limitations for being able to remember text-based passwords.Another habit introduced as a cause of dealing with the problem of remembering passwords is to create short and meaningful passwords that are easier to remember. However, a consequence of creating short passwords is a vulnerability for brute-force attacks, introducing a security risk. Furthermore, having an expanding number of accounts requires users to manage a set of different passwords across multiple devices. The problem is not just to remember all the required passwords, but also to remember which passwords belong to which account or device. The increased number of accounts and devices is an another cause for users to reuse passwords across multiple accounts and devices.

One of the first large-scale studies on web password habits was conducted in 2007 by Microsoft Research [17]. They analyzed text-based passwords used by 544.960 Internet users over a period of 3 months. In order to collect the passwords, Microsoft used a Windows Live Toolbar observing activities like login frequency. They were also able to observe how many unique passwords each user had and how the passwords were used across separate URLs. Microsoft observed that a typical user has an average of 7 distinct passwords. Out of the seven unique passwords, five of them were re-used on different web pages. An estimate of the average number of accounts per user was estimated to 25 accounts per user.

Password schemes have what is called a *theoretical password space* that is the number of possible combinations of passwords that a user can make. Research has reported that when creating passwords, users to not utilize the entire password space, but uses only a subset of the conceivable password combinations. The password space in use can be seen as the *practical password space*, making the practical password space less than the theoretical password space (Figure 2.1). The selected passwords show that the security of a password scheme relates to its practical password space rather than its theoretical password space.

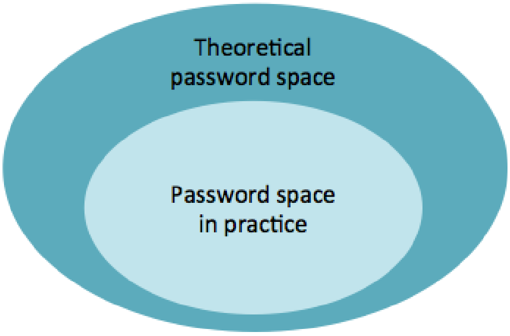


Figure 2.1: Theoretical password space vs. Password space in practice

In a case study of 14.000 Unix passwords, a research group found that 25% of the passwords were a group of words forming a dictionary of 3×10^6 words [25]. This dictionary shows that an attacker can have a relatively high success rate for an attack, despite the fact that there are roughly 2×10^{14} 8-character passwords consisting of digits, and upper and lower case letters. As a result of people choosing weak passwords that are easier to remember, a significant number of user-chosen passwords falls into a small dictionary, e.g. the password space in practice [35]. A well-designed dictionary is considered to be a tiny subset of the full password space, e.g. the theoretical password space, which further may be prioritized according to the likelihood for a password to be chosen. It is, therefore, commonly stated that the security of a password scheme is related to the size of its password space in practice, rather than its theoretical password space. The high success rate of dictionary attacks against text-based passwords is considered to be a significant cause of the recall capabilities of humans and how they choose their passwords.

As a result of the shortcomings of text-based authentication, graphical authentication is getting increased attention as an alternative to text-based authentication. Graphical passwords are attempting to help the users to be able to create secure passwords that are also easy to remember. Instead of consisting of text and numbers, graphical passwords make use of images and visual objects in the authentication process. When comparing the use of text against the use of visual objects, the human brain is more capable of remembering images than text [12]. As a consequence of humans being more capable of recognizing images, users will be more capable of creating more complicated passwords that are harder to guess.

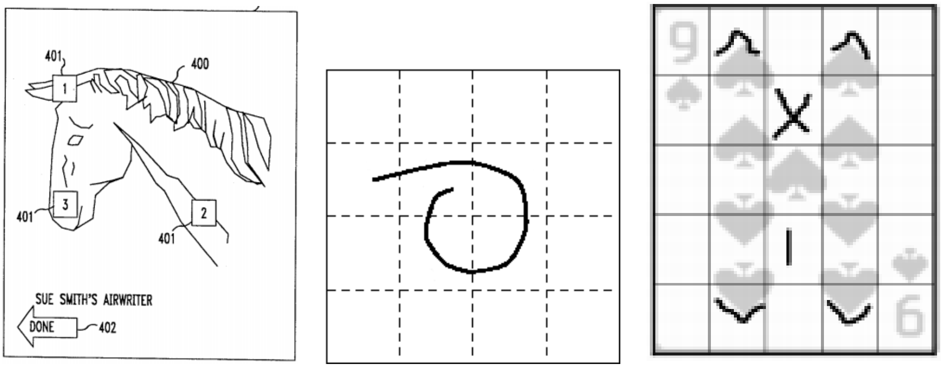
The next section will look further into the history of graphical passwords; when did it all start and what does the situation look like today?

2.2 The History of Graphical Passwords

This section will look at related work on graphical passwords from a historical point of view. The graphical password schemes reviewed will provide an elaboration of how the scheme works as well as a graphical illustration of its design

Since it all started around 1996, there have been many suggestions for graphical password schemes. When proposing a new password scheme, there are several aspects of the scheme that needs to be considered. A password scheme needs to be secure in terms of entropy, and it needs to be hard to guess, as well as being intuitive to use. The history of graphical passwords is important to know because each scheme is attempts to improve various aspects of an earlier proposed scheme. A detailed understanding of today's situation can be understood by studying graphical passwords from a historical point of view. The review starts by looking at where the first graphical passwords originated from, ending up looking at today's situation some of the newly proposed schemes.

Greg Blonder initially described the idea of graphical passwords in a patent published in 1996 [6]. The graphical password scheme proposed was requiring the user to touch on a predefined set of points on an image to pass the authentication process. The patent was just a proposal and did not further explore the power of graphical passwords, nor did it analyze the security aspects of the patent. Figure 2.2(a) is an illustration from Blonder's patent of the first graphical password scheme.



(a) Proposal for a graphical password scheme[6]

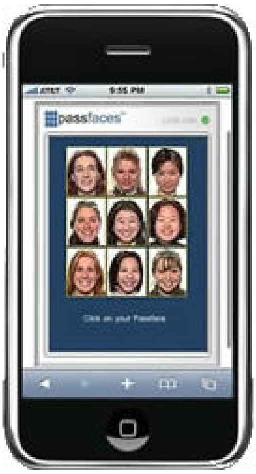
(b) DAS [24]

(c) BDAS [15]

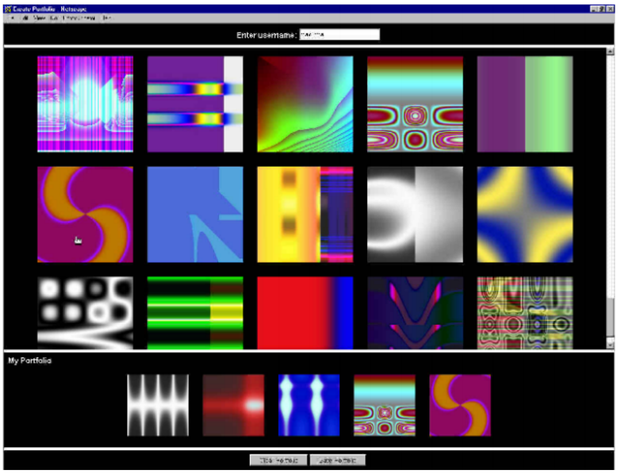
In 1999, Jermyn et al. [24] suggested a new graphical password scheme called *Draw-a-Secret* (abbreviated DAS). DAS was the first recall-based graphical password scheme proposed. The motivation for the DAS was that graphical input devices enabling the user to decouple the position of inputs from the temporal order in which they occur. The decoupling can be used to generate passwords that increased the size of the password space in practice. In order to make a more memorable password, the research group argued that the DAS was more secure than text-based passwords because the users were able to remember longer and more complex passwords. After the DAS scheme was published, Dunphy and Yan [15] added an extra background image to the

DAS and named it "Background DAS" (BDAS). The thought of adding a background to the DAS was to encourage their users to make more complex passwords. Dunphy and Yan believed that the additional image would support the users to remember longer and more complex passwords, and therefore stating that the BDAS was more secure than DAS. Figure 2.2(b) and Figure 2.2(c) are showing images of the DAS and the BDAS respectively.

In 2000, Dhamija and Perrig [13] created a new password scheme called *Deja Vu*. Deja Vu creates its visual content from the hash visualization technique [29], a technique that replaces meaningless strings with structured images. The images end up looking like random art as a result of the hash visualization technique that turns the bits of a meaningless string into an image. Dhamija and Perrig wanted to make a graphical password scheme that solved some of the shortcomings with recall-based authentication like PINs and text-based passwords. A recall-based authentication scheme is a scheme with the characteristics of *something you know*. Dhamija and Perrig designed Deja Vu for purely relying on recognition rather than recall, as well as being hard to write down and share with other people. The images used in Deja Vu makes it hard to share a password because the images are hard to recreate, as well as being easy to remember. Instead of writing art on a grid, the users were asked to select a sequence of images from a random set of images that are generated by the hash visualization technique. The property of being hard to recreate, as well as being easy remember, assists the users in avoiding the habit of writing down the passwords. Figure 2.2(e) are showing the Deja Vu scheme using the hash visualization technique for creating images from random strings looking like random art.



(d) Passfaces [30]



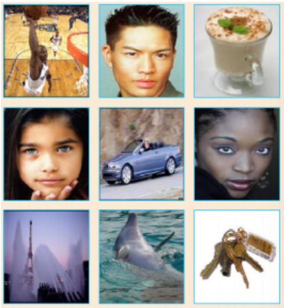
(e) DejaVu [13]

Passfaces is a graphical password scheme developed by Real User Corporation founded in 2000 [30]. The Passfaces scheme asks the user first to select four images that are a visualization of human faces. The four faces selected represent the password, and the user is authenticated by identifying them. The selected faces are shown together with eight other faces not initially included in the set of the preselected faces. The scheme exploits the advantage that people are good at recognizing other people, so when users select the human faces they can use the characteristics of the faces in the process of remembering their password. Passfaces are quite similar to the previously described Deja Vu scheme. The most significant difference between the schemes is that they make use of different association elements in the images, faces, and random art. Figure 2.2(d) illustrates the PassFaces scheme used on a smartphone. Passfaces is one of the few graphical password schemes with a commercial success.

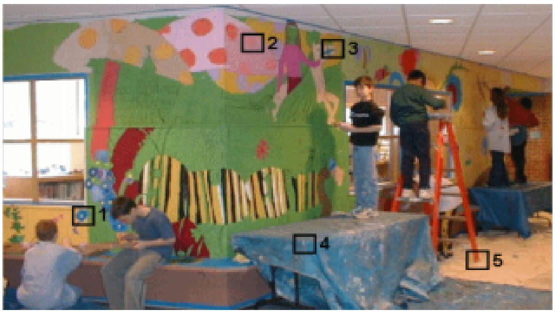
Passdoodle was a new scheme first purposed by Goldberg et al. in 2002 and later studied by Varenhorst in 2004 [19, 42]. Passdoodle is similar to DAS, but allows users to create a freehand drawing as a password, and uses a more complex matching process without the visible grid. To add variability to the doodles, additional characteristics like color, drawing speed, and number of pen strokes, have been suggested. Figure 2.2(h) is an example of a freehand drawn doodle.

In 2004, Davis et al. did a comparison of a light version of PassFace and a new graphical password scheme called *Story* [11]. The Story scheme is making the users choosing images to make up a story instead of just recalling a set of faces. The Story scheme was created to help users remember their passwords by making a memorable story of images. For users to pass the authentication process, the story had to be recalled in the correct order. For supporting memorability, users were instructed to construct a story mentally to connect the everyday images in the set. Figure 2.2(f) is an image of Story scheme demonstrating the images of objects used to create a story.

In 2005 Wiedenbeck et al. proposed a graphical password scheme called *PassPoints* [47]. PassPoints is an extension of Blonder's [6] scheme by eliminating the constraints and allowing arbitrary images to be used. They evaluated their password scheme by testing the scheme for human users. The results showed that PassPoint was a promising scheme with respect to memorability because of the low error rate and low clicking rate. The aim of this study was to gain an understanding of how different images affected user performance in authentication with a graphical password scheme. The preliminary result showed suggested that images may support memorability in graphical password schemes. Figure 2.2(g) is an image of the PassPoints scheme.

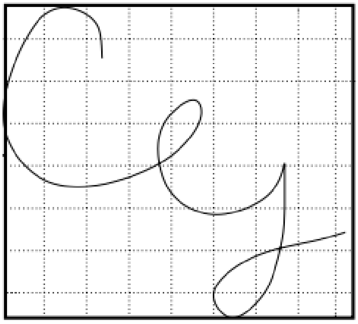


(f) Story [11]

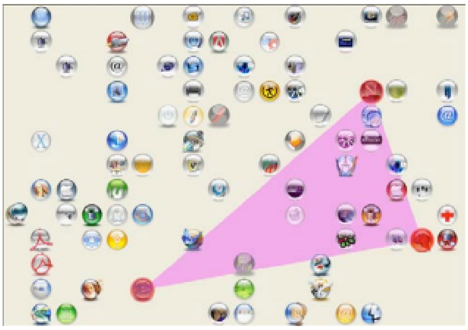


(g) PassPoints [47]

In 2006, a research group wanted to address the problem with graphical passwords and the shoulder surfing problem. They called their password scheme *Convex Hull Click* (abbreviated CHC) [49]. CHC allows the user to use the scheme in secure and insecure locations because users do not directly click on the images in the password. This design makes it hard for attackers to perform a shoulder surfing attack. CHC has a display of small icons. In the authentication process, the user must recognize some minimum number of their chosen images, or *pass-icons*, out of a vast number of randomly placed icons. If the user responds correctly every time in the correct order, the user will pass the authentication. Figure 2.2(i) is a picture of the CHC scheme with three selected icons.



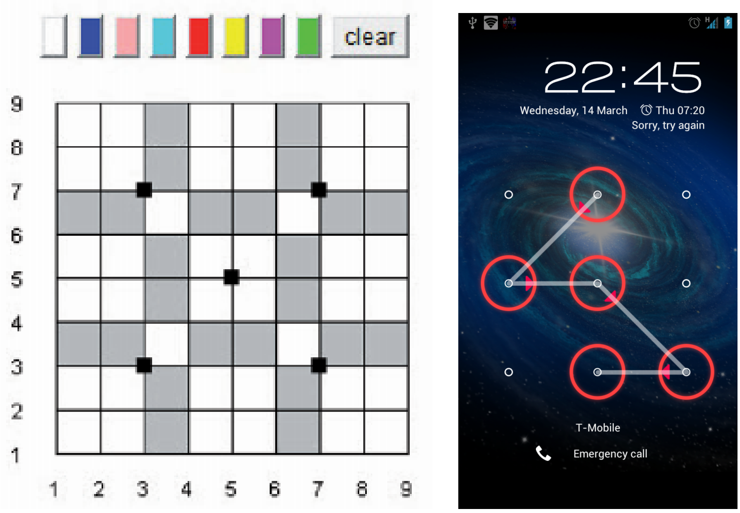
(h) Hand-written Passdoodle [42]



(i) CHC [49]

In 2007, Tao and Adams [35] designed the graphical password scheme*Pass-Go*. The motivation for creating the Pass-Go scheme was to avoid the problem of failing to be able to accurately recreate the drawing as observed in DAS. Pass-Go reduced the problem by using grid intersection points instead of grid cells as used in DAS. The users movements are captured into grid-lines and intersections, eliminating the possibility to reproduce a password where the difference is too high because of the need for precision in DAS. Figure 2.2(j) is an image of the Pass-Go grid used.

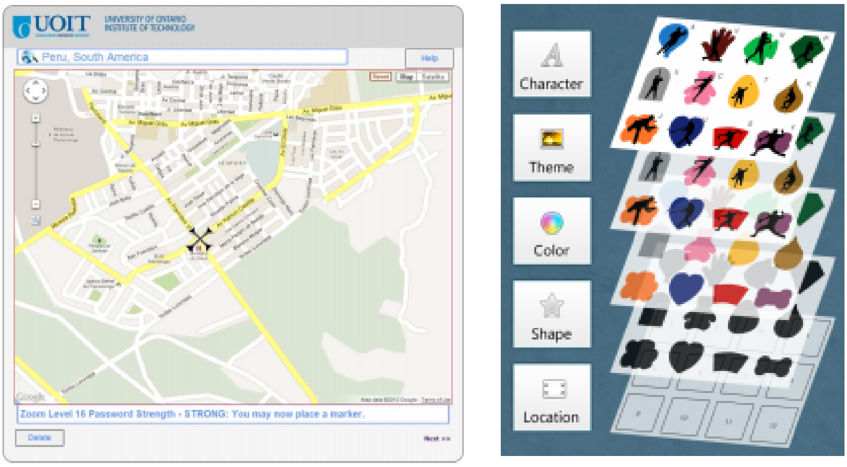
Out of the schemes mentioned until now, most are neither widely known nor widely used. The first known graphical password scheme that has gained increased attention is the Android Unlock pattern. The Android Unlock pattern is a mini version of the "Pass-Go" deployed on Google Android smartphones. Rather than entering a four-digit PIN or a text-based password, the user enters a touch-drawn password on a 3 × 3 grid connecting dots forming a password. Figure 2.2(k) is a visualization of the Android Unlock Pattern in use on a smartphone.



(j) Pass-go [35]

(k) Android Unlock Pattern

Looking at the recently published graphical password we find schemes like GeoPass [38] and PicassoPass [41]. GeoPass uses a digital map for the authentication phase where the user chooses a particular location as their password. PicassoPass is a graphical password scheme presenting a password using a dynamic layered combination of graphical elements. The users can make a story that assists the user in the recognition of the graphical elements. Figure 2.2(l) and Figure 2.2(m) is the two new proposals for graphical authentication schemes, the GeoPass and the PicassoPass, respectively.


Zoom Level 16 Password Strength - STRONG: You may now place a marker.

(l) GeoPass [38]

(m) PicassoPass [41]

Figure 2.2: Graphical password schemes

2.3 Evaluation of Graphical Password Schemes

Authentication with text-based passwords are a traditional approach, but as a result of limitations of recalling text-based passwords, users choose weaker passwords. Graphical passwords came as an alternative solution for overcoming the limitations of text-based passwords because the graphical memory of humans is particularly well-suited to remember graphical information [12]. The problem with many graphical password schemes is that they often promise improved password memorability and thus usability, and at the same time improve the security [5]. The trade-off between usability and security is illustrated in Figure 2.3. The observed trade-off between usability and security are two aspects important to understand while reviewing the literature of graphical passwords.

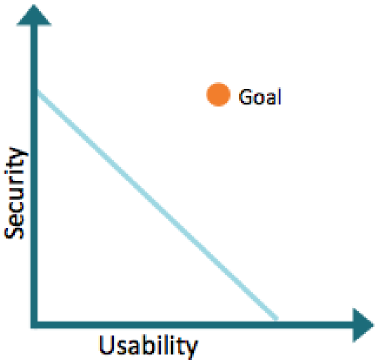


Figure 2.3: Tradeoff between usability and security

2.3.1 Usability and Memorability

An interesting question is what types of graphical password users find memorable. One of the factors supporting users to remember their selected password relates to the usability of the password scheme. What grants a graphical password high usability and what is the effect of having high usability? This section will look at different graphical password schemes focusing on the usability of the scheme.

Deja Vu was one of the graphical password schemes created in order to be easy to remember, but at the same time being hard to reuse and share as a result of the random art used. When *Deja Vu* was first proposed, the creators conducted a user study showing that 90% of all participants succeeded the authentication phase using Deja Vu. On the other side, the creators also revealed that only 70% managed to pass the authentication process by using text-based passwords and PIN codes [13]. The difference in success rate is an example of users tending to have a higher success rate remembering graphical password over text-based passwords and PIN codes.

One of the first graphical password schemes, *DAS* [24], offered a theoretical space comparable with text-based passwords. Based on cognitive studies of visual information, Oorschot and Thorpe [36] investigated the practical password space of the graphical password scheme *DAS* [24]. They found that the password space in practice in *DAS* represented an average length less than or equal to the length of 8 on a 5×5 grid.

Other researchers have revealed that users tend to draw symmetric images with few pen strokes as well as placing their drawings in the center of the grid. The researchers behind *Background Draw A Secret*[15] tried to avoid users placing their drawings in a predictable way and by adding a background image to avoid the predictable behavior. The attached background image resulted in a reduced amount of symmetry within the selected passwords and helped the users make longer passwords that were similarly memorable as for *DAS*.

Davis [11] did a comparison of the memorability between the graphical password scheme *Face* (a light version of the *PassFaces*) and *Story*. The result reported that users had more difficulties remembering the Story password, resulting in a success rate of 85%. The low success rate was observed because the users had to remember the correct sequence of the images, rather than remembering the images in an arbitrary sequence.

When considering usability, we can evaluate the graphical layout of a graphical password scheme to see if the visual elements impact the user's choice of passwords. Ullenbeck et al. [39] considered the Android Unlock patterns and investigated whether a change in the graphical layout would impact the security and usability. The original Android Pattern Lock uses a sequence of dots connected in a 3×3 grid of nodes. Instead of only analyzing the original scheme, they rearranged the points into four separate positions and analyzed patterns created by users for all four rearrangements. The results proved that the number of unique patterns created was doubled by rearranging the points, hence reduced the bias found in the original position of the nodes. However, they did not only remove some of the bias from the original grid, but also introduced new ones. One of the rearrangements was a random approach. Unfortunately, this random arrangement of nodes looked like the mathematical *delta*, an association element that was recognized by several of the participants. The random arrangement scored the worst entropy seeing as many of the users selected the same pattern. People are good at recognizing patterns and using association elements. It would not be surprising if users found similar results in other rearrangements of the grid if it had a shape similar to other symbols or association elements.

Wiedenbeck et al. [46, 47, 48] conducted three lab-based user studies on the graphical password scheme *PassPoint*. The results determined that the participants needed an average time of 63 seconds to create their password, and an additional average time of 171 seconds in training time to remember the created password. The login time took between 9 and 19 seconds on average. The time spent highlights the importance of research on usability and memorability when considering graphical password schemes. The factors that grant a password scheme high usability can be determined by looking at the average creation time, time to remember the password, as well as time used in the login phase.

It is still a problem that published research on graphical passwords focusing on usability are conducted with a pen and paper approach, raising a question about the results' validity. One problem may be that many graphical passwords are not implemented, but rather theoretical and visual suggestions. There is still a need for further research on graphical passwords in their actual intended environment of use.

2.3.2 Security

In knowledge-based authentication, e.g. *something you know*, we classify attacks into two general categories: guessing and capturing attacks. In a guessing attack, the adversary must search through the entire password space; this is often referred to as a brute force attack. A brute force attack validates all possible combinations, making such an attack highly time-consuming. If an attacker has some knowledge of the user or the user's password habits, the attacker would be able to predict the user's password by avoiding searching through the whole password space. Such a reduction in the password space is a reduction in the overall security of the scheme. When managing to reduce the search space, this type of attack is often referred to as a dictionary attack. When talking about capturing attacks, the attackers can directly obtain the passwords by observing the authentication process. One of the known capturing attacks on graphical passwords is shoulder surfing where an attacker is able to observe a user's password as a cause of a visual presence.

When selecting a password, many users select a password that connects to them as a person or to something they know. By using this strategy in the process of creating a password, there will be a lower probability of forgetting the password because the password is something you already know. When using such an approach, it is more likely that the person remembers the password, while at the same time helping an attacker to be able to guess the selected password quickly. There is a tradeoff between what is possible to remember and what is secure enough to use; attackers utilize our predictable behavior. A password created using this predictable behavior is called a biased password. A bias is a prejudice in favor of or against one thing, person, or group compared with another, usually in a way that influences a person's choice of action.

Jermyn et al. [24] evaluated the security of the graphical password scheme *DAS*. One of the statements is that the users do not utilize a uniform distribution of the possible passwords by using Klein's study [25] as an argument. The fact that users do not pick passwords uniformly is no itself a sufficient statement to make a guessing attack successful. They try to cover the possibility of an attacker making a successful attack by making their scheme more complex. The results revealed that the generated passwords were significantly harder to crack in practice than textual passwords. The problem with the conducted tests was that they used computer generated passwords that do not achieve the same validity as user-selected passwords. Neither did they analyze the security of *DAS* by including human factors that earlier have been reported to introduce bias in the password selection process.

Why do users select the passwords they do, and what strategy do they use in the creation process? Davis et al.[11] evaluated the security of the graphical password scheme *Passfaces*. They found that there was a bias introduced by people's demographics and background. The users tended to choose faces that they liked (their subjective meaning of beauty and attractiveness) and faces they could compare to themselves. The results revealed that with sufficient knowledge of the gender and race of the user, it would be possible to perform a dictionary attack to guess user-selected passwords in the *PassFace* scheme. If the user were male, 10% of the passwords could easily be guessed on the first or second attempt. Similarly, if the race of the user was known to be Asian and his/her gender was also known, then 10% of the passwords could be guessed within the first six attempts. The result indicated that graphical passwords selected by users were heavily biased. The researchers concluded that Passfaces was insecure due to the observed bias in the selection of passwords.

Dirik et al. [14] conducted an experiment by modeling users choice in the graphical password scheme *PassPoints.* The aim of the study was to test whether it would be possible to build a dictionary attack based on a user's choice of clicking points. In this study, they predefined two different images with a different level of salient points. The researchers reported that they could recover 61% of a user's selection of clicking points by searching through a smaller password space based on an analysis of collected click-points. The *PassPoints* scheme provides user-chosen images, but the aim of this study was to investigate whether it was possible to predict user's passwords using a dictionary attack on images after collected data. They observed a slight difference in two out of three images picked by the researchers. The images including few salient points were being stated as less insecure. Since the *PassPoints* scheme enabled user-selected images, the security would rely on the image and clicking points selected by the user, and not the actual scheme itself. The results can not solely state that using *PassPoints* is insecure, but rather highlight the importance of considering the human factors in security as it can influence the overall security of the scheme. The same year, another research group published results on security of *Passpoints* by using two separate research methods [37]. They conducted a user study, as well as a theoretical study image-processing tool, to test whether an attack on the *PassPoints* scheme was possible. They provided empirical evidence that attractive points, e.g. hot-spots, do exist in images. The results from the most efficient attack were generated by harvesting passwords from users to attack other targets. The probability of the guessing attack showed that 36% of the passwords selected by users ended up being guessed within 2^{31} guesses and 12% could be guessed within 2^{16} guesses. The results from the simulated attack using image-processing were slightly less efficient, but they still managed to prove that an attack on graphical passwords is possible.

One of the first large-scale studies on the Android Pattern Lock [39] stated that the entropy of a pattern is lower than its theoretical entropy. The research group compared the security offered by the Android Pattern Lock to be less than the security of a randomly assigned three-digit PIN for guessing 20% of all passwords. In the same research, a Markov model based on collected passwords was built. The patterns created was categorized as offensive and defensive patterns as a result of their research design. They set up a game asking all respondents to create a defensive pattern protecting a possible award, as well as creating offensive patterns used for guessing other participants' defensive patterns. The results showed that it was possible to guess a user's choice of patterns. Within ten guesses, they could guess approximately 4% of the defensive patterns and approximately 7% of the offensive patterns. When increasing the number of guesses to 30 attempts, they managed to guess approximately 9% of the defensive patterns and approximately 19% of the offensive patterns. If we look further into the Android Unlock Pattern, the scheme has roughly 400.000 possible valid combinations of patterns. From a theoretical point of view, such theoretical password spaces are comparable to the security of a 5-digit randomly assigned PIN. The researchers' evaluation of user-chosen patterns explains that they only have an estimated entropy slightly lower than a 3-digit randomly assigned PIN. Another interesting discovery by the researchers is that around 10% of all users use less than 190 patterns, while less than 300 patterns capture around 50% of the whole test popluation. This result is an indication of the theoretical password space not being a representative number when quantifying the security of a password scheme. We should rather look at the password space in practice, e.g. passwords that are used and memorized by users.

Psychology studies have recognized humans' superior memory for recognizing and recalling visual information. This observation supports the statement that users can remember more complicated graphical password from a larger password space than an alphanumeric password. Based on this assumption, the attacker needs to build a bigger and more complex dictionary and spend more time achieving the same success rate as for textual passwords. A clever attacker would narrow down the password space and prioritize guesses to pictures that people are likely to choose. The images that are selected are liable to be the images that users are likely to recall. To understand how an attacker might take advantage of human password choices, psychological studies on humans' visual memory are crucial to comprehend.

2.4 Psychology and Human Factors

In many years, the field of psychology have been important in order to understand how humans interpret and remember information. Psychology studies have recognized that the human brain have a superior memory for remembering and recalling visual information rather than recognizing and recalling verbal or textual information [12]. To be able to go beyond the technical part of security, this section includes related work on passwords focusing on psychology and the human aspects. Combining research from the two different disciplines computer security and psychology, can give a deeper understanding of passwords at a human level.

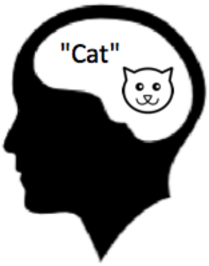


Figure 2.4: Dual-Coding Theory

One known theory from the world of psychology is the *dual-coding theory* [5]. The theory suggests that verbal and non-verbal memory are processed and represented differently in humans mind. Text are verbal information represented by symbols, in contrast to non-verbal information like images that mentally represent perceived concepts assigned to a perceived meaning of what is being directly observed. Both verbal and non-verbal information can be used when recalling information. For example, a person have received stimulus of the concept *cat*, the image as well as the word *cat* (Figure 2.4). When a person is asked to recall the concept of *cat*, a person can retrieve the image or the word individually, or both simultaneously. If a person remembers the word *cat*, the image of the cat is not lost will be possible to retrieve at a later point in time. The ability to code a stimulus in two separate ways can increase human's ability to remember, in contrast to only code the stimulus one way.

When it comes to humans and visual interpretation, studies support the idea that people recall symmetric images better than asymmetric images [4, 18]. A particular interesting observation is that mirror symmetry carries a special status I the human memory [43]. An understanding of psychological studies on visual memory can help to build successful attacks against graphical passwords. If an attacker successfully manages to use the symmetric properties of graphical password schemes, the security of the scheme might be significantly reduced.

Besides choosing symmetric password, humans tend to be influenced by graphical elements in a password scheme. A study on the *PassFace* scheme [11] revealed that there was a high bias in the password selection according to a user's gender and race. When analyzing the choice of faces according to the participant's gender, most of the male and female participants chose female faces. In addition to the bias towards preferring female faces, 60-70% of the participants preferred a model over a typical female/male person. They also looked at the race of the faces, where the results showed that almost all of the participant preferred their own race. This research raises the question if it is possible to analyze user's choice in passwords based on the demographics of a user.

A difference in graphical and text-based password schemes is that graphical passwords can use images with colors that may influence a user's choice in graphical passwords. In a user-study [37] on the image-based scheme *PassPoints*, it was observed that different images were easier guessed compared to other pictures. When analyzing different images and visualizations, gestalt psychology [44] is an important field to understand user's interpretation of visual objects. The picture from the user-study being easily guessed was the picture of cars in various positions and different colors. A possible explanation could be that humans seek to find a pattern in an image that are easily remembered. Structured images can be analyzed by using the principles of grouping, similarities of color, and similarity of size in the picture, e.g. the gestalt principles, helping humans to be able to remember the image.

Password habits may be different across different subpopulation as a cause of background and culture. In 2012, Joseph Bonneau released an analysis of 70 million passwords from *Yahoo!* [7]. The passwords were analyzed in terms of guessing rate by performing a dictionary attack. The collected data contained 328 subpopulations. The results showed that there were no better populations compared to others in the collected data, but there was observed a variation in the different populations. Demographically, gender had a small effect on the guessing rate while age increased across different age groups. The analysis also revealed that language had a significant impact on the password strength where Indonesian-speaking users were among the weakest subpopulations as a contrast to German and Korean-speaking users that provided stronger passwords.

2.5 Graphical Passwords and Mobile Devices

Users are not only dependent on remembering passwords across multiple web pages and systems but do also need to remember passwords for our small mobile devices. The use of handheld devices, such as smartphones and tablets, has seen tremendous growth in the recent years. The smartphone in general has revived increased attention because of its increased capacity and its variety of use. On the first version of the smartphones, users could access their email, participate in social networks, as well as the basic features of a phone like calling and text messaging. During the past years, the gap between a desktop and a smartphone have become smaller and smaller. People today use their mobile phone for work purposes, mobile banking, and online shopping. This progression of the smartphone sets higher standards of the security on smartphones. A smartphone is a handy tool in daily life but do also contain a lot of sensitive of your private life. Mobile devices can easily be lost or stolen due to it small size, making an increased need for protecting the sensitive data from unauthorized access.

The smartphone have emerged as an excellent platform for graphical passwords because its intuitive interaction with the touch screen in contrast to text-based passwords on mobile devices. Graphical passwords on mobile devices seem like a natural fit, as they often require direct manipulation of visual elements. For avoiding unwanted access on smartphones, different locking mechanisms are provided. The history of locking mechanisms was often a solution solely to prevent accidental use while current mobile phones require protection to secure the potentially vast amount of private data that we keep on our smartphones. The situation of our active use of mobile phones, as well as its well-suited platform for using a graphical password, makes authentication on mobile devices an interesting field of study.

When looking at mobile security it's necessary to be familiar with the magnitude of mobile phone usage. As of 2014, over 90% of American adults owned a mobile phone, whereas 58% of American adults owned a smartphone [32]. Another 34% of the users used their phone regularly instead of using other devices such as a desktop or laptop computer for searching on the Internet. The numbers are collected from a population only living in the USA, but still provides insightful information about the usage of mobile devices today.

As stated earlier, as a cause of users storing sensitive information on their phones, it is important to understand the relationship between the use of security features and users risk perceptions. One aspect essential to understand is the reason people choose to use, or not use, screen locks on their smartphone. Engelman et al. [16] published a research paper in cooperation with Google on people's screen locking behavior and attitude towards security on their smartphones. They observed a strong correlation between the use of security features and risk perceptions. They reported that 33% of the smartphone users were thinking about the locking mechanisms as too much of a hassle. At the same time, 26% of the same population didn't think that someone would care about the information stored on their smartphone. Another research group studying the same topic revealed that 46.8% of the participants agreed or fully agreed that unlocking their phone can be annoying. At the same time, 95.5% of the respondents somewhat agreed or fully agreed that they liked the idea that their phone was protected [20]. Locking a smartphone are crucial, even if users do not prefer it because they think that it is annoying. A study reported that 29% did not use any for of locking mechanisms [32] while another research stated that among 35% of mobile users do not lock their phone [40]. The number may vary, but it still highlights that the users want to be secure while at the same time do not wish to use security mechanisms. The results reported might be an indication of a trade-off between the time need to type a password and users risk perception. What is more important, time used or level of protection?

In terms of security, it is interesting to look at the use of mobile devices and locking habits. Services that are in active use are known to be protected by a weaker password as a cause of the overhead in time spend typing the password. In 2014, a group of researchers published a field study of users (un)locking behavior [20]. The problem observed was that that user had to unlock their screen frequently. In the field study, they found a significant overhead in the time used for unlocking their phone. The participants used on average 2.9%, and up to 9% in the worst case, of their time interacting with the smartphone unlocking the screen.

It is stated that many users use their smartphones to perform tasks that involve utilization and storage of sensitive data. Smartphones today do not require their users to use any locking mechanism on their smartphone. As a cause of users tending to choose the easiest way out may result in the choice of not having any locking mechanism at all. By not using any locking mechanisms, the security risks of looking sensitive information are ignored. In a study, over 40% of the users only used the basic *Slide-to-unlock* mechanism on their smartphone, as well as over 16% did not use any locking mechanisms at all [20]. This result highlights an a bad habit among mobile users that may have consequences. What happens if your mobile is stolen? A loss of a mobile phone is not just the cost of replacing the physical device, but also a loss of sensitive data. If the wrong person finds the device, sensitive data on the device may be lost and used for unintended purposes. In 2012, Pew Internet estimated that nearly a third of mobile users have had their mobile device stolen or lost [31]. It is interesting to compare people's locking behavior towards phones that are stolen or lost. The same report also stated that 12% of cell owners say that another person have accessed their phone, making the owners feel that their privacy being exposed to the public.

Besides losing a physical device, what consequences relates to the loss of a smartphone? One point of attack is to get access to people's email account. If you can grant access to someone's email, you probably can get access to a lot more as a cause of password reset sent to the user's email. A study reported that all of their interviewed participants had their email account automatically logged in, as well as 31% of them did not use any locking mechanism at all [16]. The same research group investigated how much information you could gain from getting access to a person email account. The results revealed that both users with or without locking mechanisms found sensitive information in their email account like SSN, Bank Account Number, Email Password and Home Address. Mobile devices might contain more sensitive information than users are aware of.

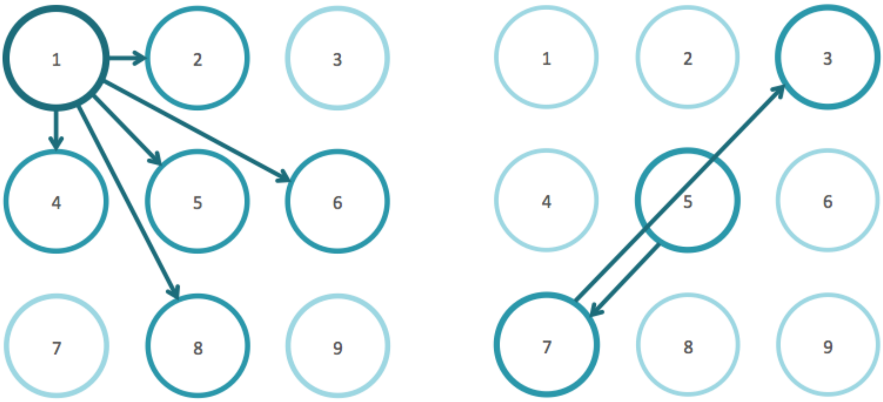
2.6 The Android Unlock Pattern

One of the most popular smartphones in the market is the Android smartphones. Automatic screen lock is one of the most commonly used protection for unauthorized access on smartphones. Android provides several password mechanisms like PIN code, alphanumeric password, pattern lock, face lock, and slide-to-unlock. Among these screen lock options, the slide-to-unlock mechanism only avoids accidental interaction with the screen. Alphanumeric password and PINs are the must commonly used authentication mechanisms used on smartphones, as well as in other systems requiring authentication. An alphanumeric password are created from all writable characters, while PIN codes only use digits. The newly released face-unlock uses image processing to analyze your face to grant access to your phone. The Android operating system are being known for the graphical password called *Pattern Lock* released by *Google* in 2008. This graphical password scheme is at this time available on all Android devices, as well provided on other mobile operating systems apart from Apples' mobile operating system, iOS.

The *Android Pattern Lock* is one of the commonly used screen locks mechanisms on Android devices. For unlocking a device using pattern lock, the user is asked to draw a user-defined sequence of connected dots on a 3×3 grid. Such path is called an lock pattern and is presented in Figure 2.2(k). When creating a pattern, Google has designed several rules for creating a pattern:

1. A pattern needs to be defined by at least 4 dots.
2. A dot can only be selected once meaning that the maximum number of connected dots are 9 (as defined by the dots in the 3×3 grid).
3. The pattern will always connect all dots along a path, expect when a dot already has been selected.
4. A pattern can go through previously connected dots to connect dots along the same path.
5. The dots can be connected horizontally, vertically and by the diagonal.

The first and second rule only states the minimum and maximum number of connected dots in the pattern. The third rule denotes that if a path is drawn from node 1 to node 3, then the nodes in the path will be $1 \rightarrow 2 \rightarrow 3$ as a cause of rule number three. Rule number four states that you can go through a node that is already in the path, but the node will only be selected once. Such path are called an overlap. Pattern having an overlap are illustrated in Figure 2.5(b) by having chosen the path $5 \rightarrow 3 \rightarrow 7$, where node 5 is not selected twice when going from node 3 to node 7. Figure 2.5(a) illustrates rule number five displaying all nodes that reachable and the valid directions; vertically, horizontally, and diagonally.



(a) Reachable nodes from node 1

(b) Create a path over a selected node

Figure 2.5: Illustration of the Android Pattern creation rules

In a historical view, the Android Unlock pattern is seen as a new authentication mechanism as opposite to alphanumeric passwords and PIN codes. In a security perspective, the pattern has a total of 389,112 valid patterns using a 3 × 3 grid. Comparing the Android Lock Pattern with PIN codes, having 10.000 possible codes, the Android Lock Pattern seems to be more secure. Compared to an alphanumeric password, the number of combinations depends on the number of characters included. When looking at published research, users are capable of remembering passwords with an average length of 7 to 8 characters. As introduced, the Android unlock pattern is a more suitable form of authentication for mobile devices due to its interactive and graphical form that suits small touch screens. When using an alphanumeric password on a smartphone, a virtual keyboard are used typing the password, making it less suitable for mobile devices because of the size. Smartphones are being used in various situations during a day, making it desirable to use an authentication mechanism that is quick to type and easy to remember, avoid spending the time unlocking the smartphone. It is no secret that an alphanumeric password with its extensive password space would be more secure if users created long passwords. However, this is not the reality because mobile devices do are not suitable for typing long passwords on the virtual keyboard. To further explore the security and usability of the Android unlock pattern, we will take a look at published research to get an overview.

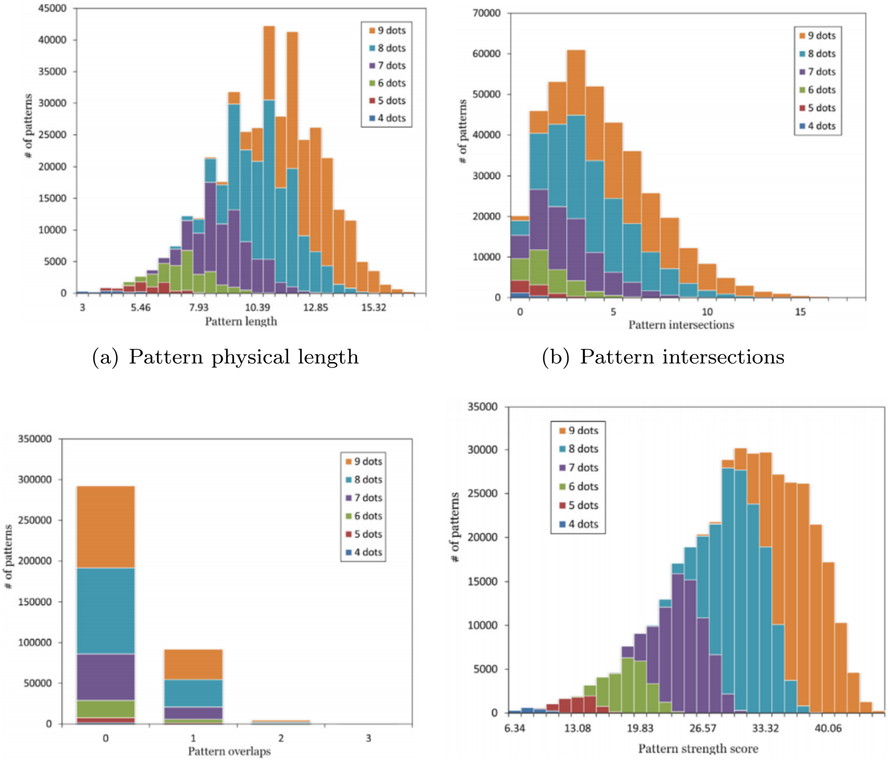
# **Length**	# **Valid combinations**
4	1624
5	7152
6	26,016
7	72,912
8	140,704
9	140,704
Total	389,112

Table 2.1: Number of pattern combinations

As stated, the Android Pattern Lock has 389,112 valid patterns, according to the listed rules. But is this number as secure as it sounds? When looking at the security of a password scheme it can evaluated according to total valid combinations, e.g. the password space or the passwords that statistically are likely of being chosen. The passwords that statistically are likely of being chosen refers to the password space in practice. Table 2.1 summarizes the total number of valid combinations based on the pattern length [34]. An another way of look at the security of the a pattern is to measure the pattern strength. A research paper investigated the use of password meter for measuring the strength of a pattern. Their hypothesis was that the utilization of a password meter was providing more secure user-selected patterns. They stated that there often tended to choose a short and easily guessed pattern due to memorability. A password meter is often shown as a colored bar that is often used as an indication of the strength of a password. The research group used a mathematical equation (2.1) for calculating the strength of Android pattern locks.

$$PS_P = S_P \times log_2(L_P + I_P + O_P) \tag{2.1}$$

PS_p is the strength score of pattern P. S_P, L_P, I_P, and O_P are the number of connected nodes, the physical length, the number of intersections, and the number of overlaps of P, respectively. Using Equation 2.1 on all valid patterns gives a score from 6.340 to 46.807. The formula was being utilized by Sun et al. [34] in their research on Android patterns and password meters. They look at pattern strength in a different way other than just calculating the complexity of a pattern by its length. Calculating password complexity by length seems as a naive approach, making this way more realistic using all the rules of the Android pattern in the equation. When looking at the different characteristics and strength of a pattern used in Equation 2.1, Sun et al. maked distribution graphs of the different characteristics and the strength (Figure 2.6).



(a) Pattern physical length

(b) Pattern intersections

(c) Pattern overlaps

(d) Pattern strength

Figure 2.6: The distribution of the pattern characteristics and strength [34]

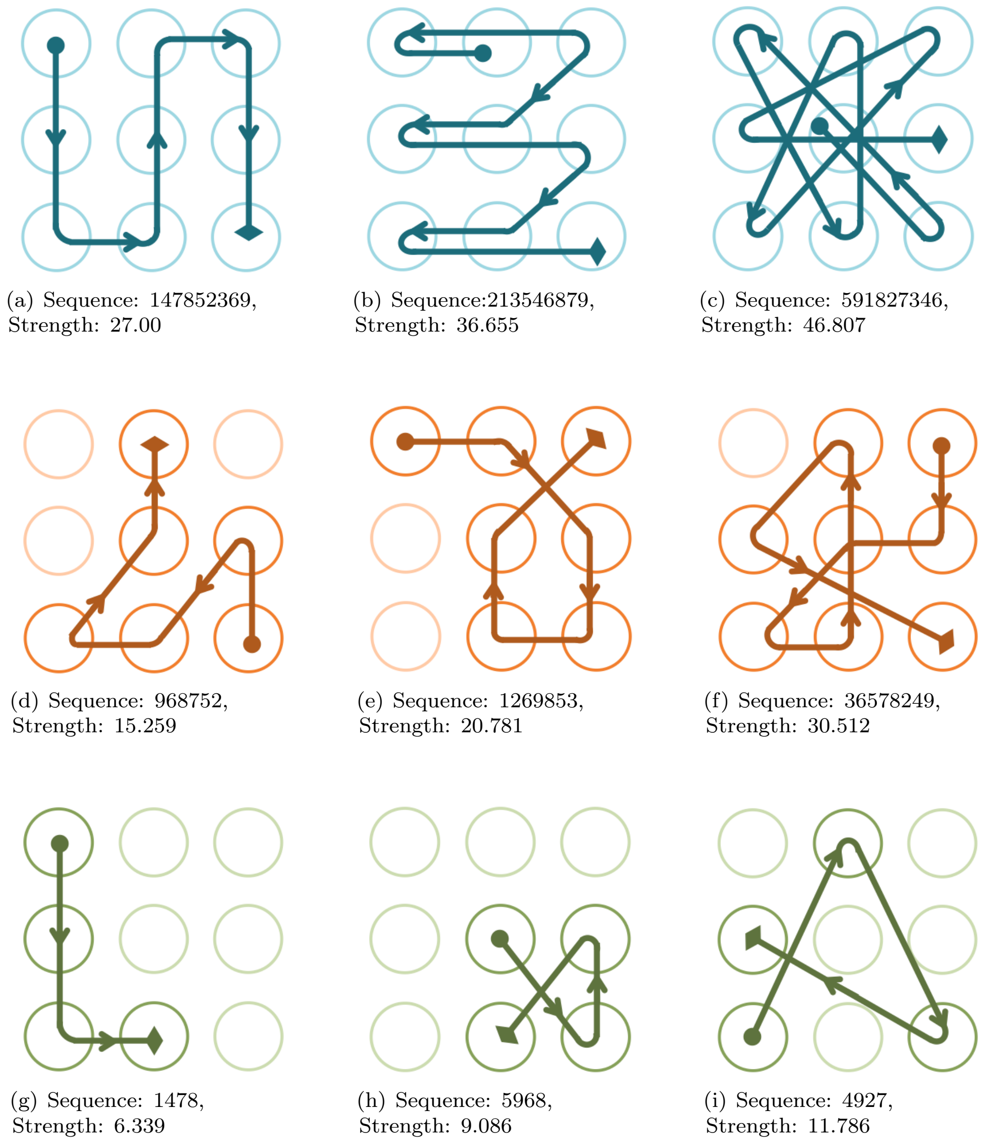
Sun et al. [34] created two different visualizations of a password meter, one looking like a progress bar (Type 1), and one indicating a percent of the strength (Type 2). They recruited 81 participants for a survey testing the strength of user-created patterns. They divided the participants into three separate groups; no password meter (Group A), password meter type 1 (Group B), and password meter type 2 (Group C). The survey randomly assigned the participants to the separate groups. The result revealed that the strength of the created patterns in group B and C had a higher complexity and strength but had a higher error rate when retyping the pattern. As a researcher explained, users are typically more security conscious when they are aware of the need for such behavior [2]. The error rate points to the problems with passwords and security in general. A long and complicated password are harder to guess but are not likely to be selected due to memorability issues. Also, a problem with the Android Pattern is that the pattern is being frequently used as it is provided to grant access to the smartphone in all kind of situations. The results from the survey states that the input convince was the reason that caused the highest number of participants in the user study not selecting a pattern with a high complexity and strength. A pattern containing more dots takes a longer time to type. When looking at patterns containing intersections and overlaps, there is a high chance of accidentally hitting a wrong dot when drawing a pattern. Such an error are causing the user to redraw the pattern, hence spending more time passing the authentication process. The conclusion is that the time used to type the pattern are as crucial as the time used to memorize the pattern when looking at the users choice in patterns.

The pattern strength meter gives a score of the visual complexity of a pattern. The benefit of looking at the visual complexity instead of only looking at the length can be seen in Figure 2.7. When looking at Figure 2.7(a), 2.7(b), and 2.7(c), they have all have the maximum length, but at the same time having a different pattern strength, e.g. a difference in visual complexity. Figure 2.7(g), 2.7(h), and 2.7(i) have all the minimum length of four dots, but they still have a difference in the measured strength. The visual complexity is one of the extra security dimensions to study when working with graphical passwords. The pattern itself is just a sequence of numbers, but the order of the sequence can make a big difference in visual complexity, and can be important when wanting to avoid known attacks like shoulder surfing.

When looking at user-selected passwords, studies tell that many users are using graphical shapes to support memory [45]. Sun et al. [34] analyzed collected Android Lock Patterns and found empirical evidence that some users tended to use patterns which looked like letters or numbers. They found patterns looking like the letters and numbers C, L, N, Z, 2, and 7 that easily can be created on a 3×3 grid. Such a strategy for enhancing memorability by using association elements, e.g something that the user are familiar with, are used in other password schemes. The use of association elements is known to used in PIN codes and alphanumeric passwords where names, objects, dates are used to remember the password instead of a visual representation of a letter or number.

Android Unlock pattern have been shown to have biases when being user-chosen. A research group did one of the first large-scale user study on the security of the Android Unlock Patterns in order to quantify the security of the *Android Pattern Lock* [39]. They analyzed the biases introduced in the pattern making process and added changes to the scheme in order to avoid the known biases in the password scheme. The researchers found that there was a high bias in the pattern selection process, e.g. the upper left corner and three-point long straight lines are likely being selected. If user-chosen patterns were being uniformly chosen, the probability of starting in at any point should be 11%. The results revealed that there was a strong bias towards the starting point in the corners. If the points were uniformly chosen, the probability for all four corners should be 44%, but the results showed that the probability is close to 75% in their pen-and-paper study. In contrast, the center point, the right, the upper, and the lower center points only got a probability of 14% to be selected. Other results from the pen-and-paper study found that the average pattern length was 5.63 with a standard deviation of 1.5. As stated earlier, users tend to take the easiest way out, making users choose short patterns that are easy to remember and type. Looking at the selection of starting node, 43% in the user study, and 38% in the pen-and-paper study selected the upper-left corner as their starting node. This is supporting the researchers claim that users tend to choose less secure patterns “in the wild” than a theoretical evaluation of a password scheme.

What is causing the biases observed in the Android Pattern Lock? The next chapter will introduce a research design for collecting Android Pattern Locks for further analysis focusing on human factors.



(a) Sequence: 147852369, Strength: 27.00

(b) Sequence:213546879, Strength: 36.655

(c) Sequence: 591827346, Strength: 46.807

(d) Sequence: 968752, Strength: 15.259

(e) Sequence: 1269853, Strength: 20.781

(f) Sequence: 36578249, Strength: 30.512

(g) Sequence: 1478, Strength: 6.339

(h) Sequence: 5968, Strength: 9.086

(i) Sequence: 4927, Strength: 11.786

Figure 2.7: Examples of patterns with different length and strength

3 Data Collection

Section 3.1 is the detailed description of the research strategy; the survey. For the selected research strategy, a corresponding sampling technique are described in Section 3.2. For knowing what kind of data to collect, a review of human properties is included in Section refsec:reviewofproperties. When collecting data, it is crucial to know how many samples are needed for being able to get significant results. Section 3.4 is an elaboration of how the sample size were selected. Some of the data properties are might difficult to collect for several reasons. Section 3.5 provides a list of collection strategies that working as predictive countermeasures to be used in the data collection process if needed.

3.1 Research Strategy

This section will provide information about the chosen research strategy. A research strategy is being considered as the selected strategy for being able to answer the main research goals; the hypotheses and the research questions. Briony J. Oates [28] presents six different research strategies: survey, experiment, design and creation, case study, action research, and ethnography. This section will not go further in describing the difference between the six strategies, but rather explain the choice of research strategy. The detailed work behind the choice in research design was carried out in preliminary work beforehand of this thesis [26].

The selected research strategy for this dissertation is a survey. The idea of a survey is that it will obtain some kind of data from a large group of people or events, in a standardized and systematic way [28]. The survey is chosen due to the amount of data needed for the analysis, as well as the lack of time for choosing other approaches, like interviews and other observation techniques. There are different ways of creating a survey, for example, pen-and-paper or using an online survey. For this research, an online survey is selected for several reasons. First, a pen-and-paper survey is too time-consuming to perform and are not scalable for international data collection. It is also desired to keep responses anonymous due to the topic of research. The form of administration is, therefore, self-administered as there is no one present when the respondent answers the survey. More information about the sampling technique is described in section 3.2.

In the preliminary work, wireframes were created as a draft of the survey. These wireframes can be found in Appendix A. Chapter 4 will have a final implemented version giving a more detailed description of the survey.

3.2 Sampling Technique

The *sampling frame* is a list of the whole population of people that could be included in the survey. When looking at the sample of this study, there is not possible to make a limited sampling frame that can be summered in a list. The population of the sample frame is considered as all people with a smartphone because this research looks at people's choice of Graphical Password, in particular the Android Lock Patterns that is a mobile authentication scheme.

When conducting a survey, it needs to be decided how to select people from the sampling frame. There are two different ways of doing sampling: probability sampling and non-probability sampling [28]. Probability sampling means that the sample has been chosen because the researcher believes that there is a high probability that the sample of respondents selected is representative for the overall population being studied. Non-probability sampling means that the researcher does not know whether the sample of people is the overall population.

Because of the broad sampling frame it would be feasible to use non-probability sampling for this research. When using non-probability sampling, we make a decision that it is not practicable to describe a representative sample because there is too much uncertainty about the respondents that will voluntary answer the questionnaire. When choosing a non-probability sampling, we need to select an *sampling technique*. The possible sampling techniques we can choose from is purposive sampling, snowball sampling, self-selection sampling, and convenience sampling [28]. The *purposive sampling technique* requires the researcher deliberately to choose people that are likely to produce valuable data to meet the purpose of the research. This requirement would not be a good technique because I am not able to pick who would participate, as well as the data collection need to be performed worldwide. A Purposive sampling technique would maybe provide a more uniform collection of people, but it is hard to control respondents when the survey is distributed over the Internet. The *snowball sampling technique* utilize the network from one person of the sample frame, and then collects new names from that person. The need for directly communicate with respondents is not possible for me as a researcher. First, I have no contact with the respondents. Second, the respondents, should remain anonymous when answering the questionnaire, as well as there should be possible to track the information back to the respondent. When using the *convenience sampling technique*, the researcher only selects respondents who are convenient for them, because they are easy to reach or willing to help.

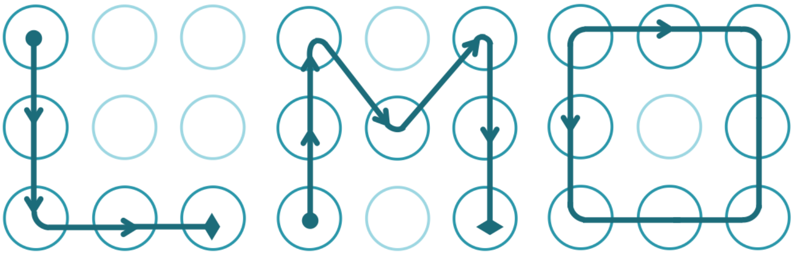
When doing *self-selection sampling*, the researcher advertises their interest in a particular topic and their need for respondents and collect data from anyone who are willing to participate. The self-selection sampling strategy looks like an excellent fit for the research. The survey needs to be distributed over the Internet, and the self-selection strategy will support this choice. People who select themselves for research often do so because they have strong feeling for the subject, or that the research can bring them a personal benefit. A self-selection sampling technique may also reduce the bias that can be introduced when the researcher hand-picks the respondents. With the self-selection sampling strategy allows all interested peoples with a smartphone to participate in the research. The self-selection is a useful technique when directly contact is not achievable. The self-selection sampling remains the technique that are looking as the best fit for this research.

3.3 Review of Human Properties

When using an online survey, the human properties must be carefully selected. *First*, when a survey is distributed on the Internet there is no way back, and we need to be sure that the chosen questions will provide sufficiently and relevant data for answering the hypothesis. *Second*, we need to review all human properties and only select a suited number of them. A too long survey may result in a small number of responses because the time needed to complete the survey if all the properties is included. *Third*, some of the properties may not have a suitable grouping of answers, and may be challenging to include in a survey that needs to be replied to on a mobile device. If such a property is included, it needs to provide irreplaceable and valuable information for the analysis. The survey should try to avoid time-consuming and complicated questions if possible. This section will provide a detailed review of the human properties.

Language preference

By asking the respondent about their language preference it can be used to check whether the alphabet a respondent use impacts their choice in lock pattern. For example, Chinese words are written in a different way then people using the Latin alphabet. In the Android Unlock Patterns, people are able to create patterns looking the same as the letters `'L'`, `'M'`, and `'O'`. The Figure 3.1 there is an illustration the letters that are possible to select a lock pattern. The same possibility could be found in languages using a different alphabet than the Latin. When looking at PIN codes, there are certain numbers that have a different meaning than only its numeric value. In China, people are associating certain words with numbers or things based on the similarities of sound. For example, the number eight is considered as a lucky number because it is pronounced "ba", which sounds like the Chinese word for prosperity [8]. Other numbers like 4 and 775 are pronounced in the same way as "death" and "Kiss me", respectively. Looking at the selection of PIN codes based on people's language, people tend to choose PIN codes that they can associate to a number of special meanings. Examples of this behavior are PIN selection corresponding to date of birth.



(a) The letter L (b) The letter M (c) The letter O

Figure 3.1: Patterns corresponding to letters

Age and Gender

A group of people within a particular group of age may have different risk awareness. A person with an age of 30-40 years and a person younger than 20 years may have different concerns with security. A person of age between 30-40 may use their phone to perform a task with requiring high security, like tasks related for work purposes. A person with an age below 20 may not have the same security awareness because of the different use of mobile devices, as well as experience. Age is also interesting demographic information that can be used to group the respondents into distinct subgroups. When looking at gender, psychological studies have reported that males are more likely to take risks than females [9]. In the literature review, there were no results found related to gender and risk awareness. However, researchers have found bias in the password selection process of *PassFace* when looking at gender.

Handedness

An interesting property of humans is the fact that people write with either left or right hand (and sometimes both). Handedness can influence the way a person are holding a mobile phone, and may impact their choice of lock pattern. In the literature review, it was not found any studies that reported results of people choices in patterns based on handedness. Published research [39] found that over 40% of the participants in their study started their Android pattern by beginning in the upper-left corner, but did not record the hand used when collecting the patterns. A question to be answered is whether a left-handed person using the left hand while interacting with the screen increases the probability for starting in the upper right corner. Figure 3.2 illustrates handedness and likely starting point as a hypothesis, where the percentages attached to the nodes are indicating the probability for starting in the node. The right part of Figure 3.2 are collected from the research by Uellenbeck [39] while the numbers on left part are only hypothetical numbers. Since it is more likely that people are right-handed, are we able to mirror the results in the right figure for left-handed people? My hypothesis as stated earlier is that people that are left-handed are more likely to start in the upper right corner, while people that are right-handed are more likely to start in the upper left corner.

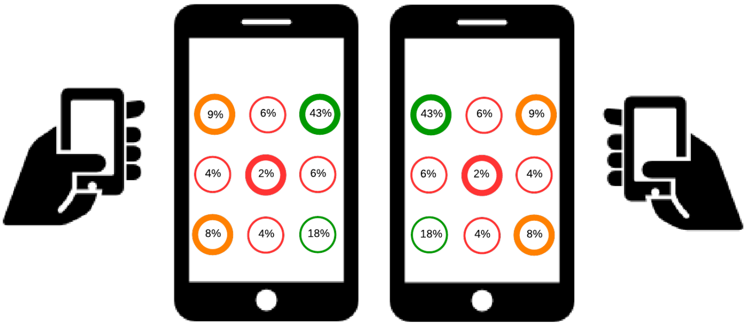


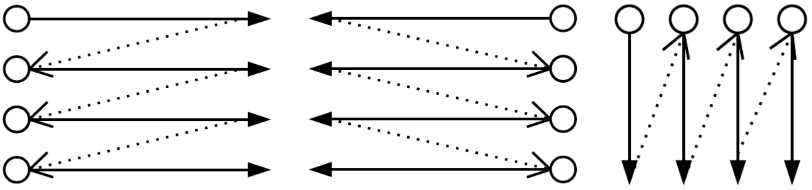
Figure 3.2: Likely chosen initial starting point [39] and handedness

Nationality

The nationality of users is often used in research on graphical passwords. When asking a person about their PIN code, a nationality can be used to find likely numbers to appear because some cultures associates a number with a historical or religious event. In a research on the *PassFace*, the nationality of the participants was proven to be valuable because people tended to choose faces from the same race and nationality. In this research, it is uncertain how much this information will help to prove any connection between nationality and users choice in patterns. Properties like language preference and reading/writing orientation look more promising for this research. However, the nationality is a data property that are useful when getting an overview of the population.

Reading and writing orientation

In different cultures, there is a difference in the reading and writing orientation. Cultures of Europe and America usually write and read horizontally from left to right, but there are other cultures that do otherwise. Figure 3.3 illustrates three different ways of reading and writing. Traditionally, Chinese, Japanese, and Korean are writing text vertically in columns from top to bottom, from right to left. Another writing orientation is horizontal from right to left used in Arabic speaking countries. Today, the vertical orientation from top to bottom is often in a horizontal way due to the influence of English and the increased computerized typesetting, but both ways are still in use. There is research that have reported that the writing orientation are affecting the visual attention and memory [10]. They found that the reading orientation affected the way a person would memorize objects. They reported that English and Chinese speakers tended to remember an image that appeared in the top, left-hand side of the screen. The Taiwanese speakers in the study tended to remember images on the opposite side of the screen, the images on the upper-right side of the screen. An interesting aspect of the reading and writing orientation is to see if people from different cultures are choosing different patterns due to their writing orientation.



(a) Left-to-right

(b) Right-to-left

(c) Top-to-bottom, left-to-right

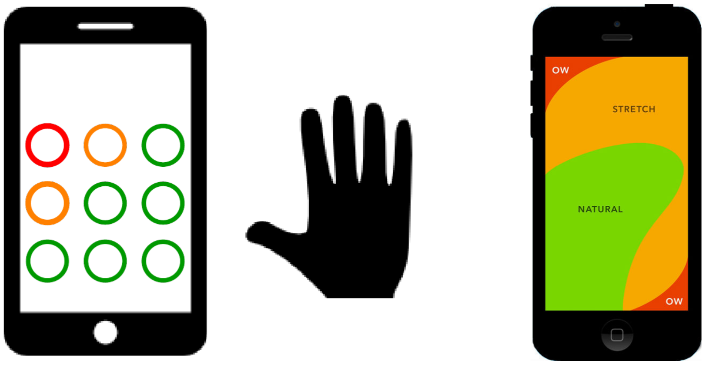
Figure 3.3: Reading and writing orientation

Profession and Occupation

The current profession of a person may say something about persons knowledge and background. When looking at profession, a person with a profession in IT may be more certain of the security aspects than people in other professions. It may cause bias in the data if people with enough knowledge of security overcompensate their choice of lock pattern because they want to prove their knowledge. Occupation is valuable information due to the knowledge level of the respondents. The occupation of the a respondent is simply if a person is a student, employed, not employed or retired.

Finger, handsize, and screensize

When looking at the finger used when creating the pattern, it impacts the reachable areas on the smartphone. When interacting with a smartphone, the most common way is to either use the thumb or the forefinger. When combining this property with the screen size and the size of the hand, we might be able to predict the selection of patterns by eliminating areas that are harder to reach. In a book called *Designing Mobile Interfaces* [22], they used an expression called the *thumb zone*. The Thumb Zone is the most comfortable area for a person to touch when holding a smartphone using one hand. Figure 3.4(b) is showing the thumb zone where the green area is where the thumb can easily access. The orange and the red areas are part of the screen that is harder to reach. The thump zone can be used when looking at users choice of patterns a pattern that are easy to type can be more likely of being created. Smartphones today tends to get bigger and bigger in size. An interesting analysis could be done by looking at user's choice in patterns based on the size of their hands and size of the screen. By looking at a situation where a person with a small hand is interacting with a big screen, it may be hard to reach certain areas of the screen when holding the smartphone in one hand. Maybe a right-handed person with a small hand interacting with a large screen will not be able to reach the upper left corner? The situation are illustrated in Figure 3.4(a).



(a) Reachable points on screen

(b) The thumb zone

3.4 Calculating the Sample Size

It is essential to determine an appropriate sample size to be able to make any conclusions with the data. When a sample size is too big, it will lead to an unnecessary waste of time in this study due to the time frame of the thesis. On the other hand, if the sample size is too small, the results can not be considered to be used for statistical tests, and it might not be possible to come up with a reliable conclusion. When using known formulas for calculating the sample size, you need to know the population size, preferred margin of error, desired confidence interval, and the percentage of the population that is likely to answer. In this study, these parameters are hard to determine because of the non-probability sampling technique that are being selected for this study. For example, the targeted population size is hard to estimate because it includes all people with a smartphone worldwide. Because of the uncertainty connected to the sampling technique, there do not exist a known formula for calculating the sample size. We are not able to calculate the sample size by a known formula, but the sample size still needs to be decided based on my subjective meaning. The sample size is determined by what is achievable with the time frame available, as well as what I as a researcher think of as a good enough sample to ensure sufficient data for getting any results.

The greater the accuracy required by my claim that my sample size represent the whole population adequately; the larger your sample size needs to be. Statisticians have produced tables that correlate population size against the sample size for the required level of confidence and accuracy. Table 3.1 is a recommended sample size for using a survey (using 95% confidence interval and +/- 3% accuracy range) estimated by the targeted population size [28]. As stated in Table 3.1, the sample size does not increase at the same rate as the target population. When looking at all people that owns a mobile phone worldwide, we can argue that the targeted population size is greater than 1,000,000, and we, therefore, need a sample size of at least 1000. It would be desirable to get a sample size bigger than 1000, but due to the time frame of this research, a sample size of 1000 should me achieved.

Target Population Size	**Required sample size**
50	47
5000	760
100,000	888
900,000	895
>1,000,000	1000+

Table 3.1: Target population and sample size [28]

Since the sampling size is quite high, and the survey is being distributed with a self-selection sampling strategy over the Internet, it is still important to provide some level of control of the respondents. Using a self-selection sampling technique implies that there is no control of who will answer the survey, hence likely that some subgroups will have a higher representation that others. In a situation where lacking respondents from a subgroup, it should be recruited more participants from the subgroup. However, when using such a strategy, it might introduce some bias because I would then influence the sample. Section 3.5 will look at some strategies for responses from different subgroups that may be underrepresented when using a self-selection sampling technique.

3.5 Response Rate and Non-responses

The survey is distributed over the internet, making it difficult to maintain control over how many people are willing to respond and how many people from different subgroups that are reachable. To be able to achieve the amount of data needed for an analysis, a strategy are need if a subgroup are being underrepresented in the sample. Below, there are listed a strategy for collecting data from the different subgroups. The strategies are being included as a cause of some subgroups might being less willing to respond, or harder to reach, because they are outside the network of the people involved in this research.

People with age of 60 or higher: People with the age over 60 may not own a smartphone or may be hard to reach for other reasons. It would be beneficial to find networks where containing a high representation of people with an age of 60 or higher. In Norway, there is a network called "Seniorweb." There is also a magazine called “vi over 60” for senior citizens with members with an age over 60. Both networks are possible to contact if there is a lack of respondents from this subgroup.

People with a different field of interest than IT and security: The network of the people involved in this research is being overrepresented by people with a profession in IT and security. It is needed to find other networks to be able to reach out to other people with other jobs. For reaching out to people with other professions, it is a possibility to use the network from my family or directly contact a company in an another field. The university is a good start because of the high representation of different fields of study.

People with a reading orientation from right-to-left: The primary reading orientation is from right-to-left, except some Arabic and Asian nationalities. My network does not include a high sample of people that have another reading and writing direction than my own. To reach out to other nationalities, NTNU is a possibility for getting help to distribute the survey to countries where NTNU have exchange programs.

People living in a different country than Norway: The sample should include respondents from different nationalities to obtain diversity in the sample. It is not easy to reach out to other countries if not having a big international network. NTNU has exchange students at the campus from different parts of the world. The preliminary work for this thesis was presented at a conference named Passwordscon. At the conference, many security interested people met at the conference are potential help for the distribution of the survey to other countries.

People that are left-handed: There is a significant higher percentage of people that are right-handed, meaning that there is a possibility of getting more responses from right-handed people. Selecting a strategy for this is hard because there is not a known official network for left-handed people. There is a possibility of using groups at Facebook like “Left-Hander’s Club.”

4 A Detailed Description of the Survey

The challenge of analysing user-selected passwords is that you need to have any passwords to analyze. The majority of the research on passwords are often conducted using passwords from leakages or passwords distributed from other sources. Looking at the Android Lock Pattern, there exist no distributed sources and all patterns locks being used by people are stored locally on each device. To be able to analyse user-selected patterns, all patterns are being collected from the survey described in this Chapter.

The survey is a custom designed and implemented specifically for this research. The survey will be described in detail throughout the chapter, starting with an elaboration of the requirements in Section 4.1 and a detailed description of how the survey works in Section 4.2. As described in Chapter 3, when starting sending out the survey, there is no way back for changing the survey. Therefore, Section 4.3 contains a detailed description of how the survey was tested before being distributed. The last section of this survey is an elaboration of the ethical aspects of this research.

4.1 Requirements

When addressing the difficulties of collecting patterns, it is important to define requirements for the survey application. This section will walk through a list of requirements specified for the survey. Each requirement are being described in detail below.

R1:	The survey should be able to stand for all commutation to the respondents
R2:	The survey should be considered as trustworthy
R3:	The survey should be implemented on a technical device reflecting the environment of the Android Lock Pattern
R4:	The survey should be easy to understand and easy to complete
R5:	The survey should be visual appealing
R6:	The survey should be provide easy navigation between the questions
R7:	The survey should provide high security

Table 4.1: Survey requirements

R1 - One way communication

The survey will be an online application, meaning that the user interface of the survey application is the only way being able to communicate with the respondents. The respondents will not be able to directly talk to me during the survey, hence important how designing the application for communicating with the respondents.

When sending out the survey, there are no record of who will receive the survey. The respondents might come from different cultures and countries, hence requiring a universal way of communicating with the respondents. The survey are being written in English, but there is a chance that there persons entering the survey not being able to understand English properly. Instead of using too much words, it is preferable to use icons and illustrations. The use of icons instead of text is not just beneficial for English-speaking respondents, rather a design choice making the whole application easier to understand for all respondents.

Besides the language used, the words used for formulating the questions need to be carefully selected considering all respondents having a different background. The use of academic and technical words are required being avoided, hence ensuring everyone being able to understand the purpose of the question. It should also be considered that some respondents do not know what Android, locking mechanism, or pattern lock is. It should be visualized, explained, as well as provide practice for the respondents needing it. None of the respondents should leave the survey feeling stupid or insecure as a cause of a poorly designed application. The participant should leave the survey with a positive experience.

R2 - Trustworthiness

Gaining trust is important for getting people to wanting to participate. When being asked to take part in a study, respondents might ask several questions regarding the reliability of the source. For example, are the data being used for the right reasons and handled according to what are being specified in the survey?

When respondents open the survey, a description of the research and the researcher should be provided. It is reasonable to provide contact information, as well as information about myself so participants able to read about the person asking them to participate. The survey should being placed on a sub-domain on my personal domain for gaining trust from the respondents. My personal domain is *marteloge.no*, containing the sub-domain *survey.marteloge.no* for the survey. The contact information is also important in terms of ownership of data. When linking to my web page, the respondents can see who are in the possession of the collected data. Who are reading the data and how is it handled? It should be clearly stated that the information collected are only available to the research group containing myself, my supervisor, and my co-supervisor. It should be clearly stated that the data gathered only will be used for research purposes.

The visual appearance is an important part of gaining trust from the respondents. This is explained as an own requirement.

R3 - Environment of use

When collecting data, the environment should be considered because it can impact the data in terms of introducing bias. When looking at the Android Unlock Pattern, the scheme is best known for being used on mobile devices like smartphones and tablets. When looking at different characteristics of mobile devices, tablets and smartphones should be defined as two separate environments of use. *First*, the tablet are often used in various settings than a smartphone that a user carry and use every day. *Second*, the physical interaction are often different. A smartphone are smaller than a tablet and can be interacted with by using one hand. When collecting data, it is desired to capture specific characteristics of the environment a pattern were created. As a cause of different characteristics with the interaction on smartphones and tablet, it is desirable only to collect data from one the environments. The Android Lock Pattern are most commonly used on smartphones, making this study only collecting data for this particular environment.

R4 - Complexity and length

Since the only option for answering the survey are by using a smartphone, the time used for complete the survey, as well as the complexity, needs to be configured. The smartphone have a limited ability for interaction where the small touch screen are the only interaction available.

For ensuring that people completes the survey, the questions needs to be short and concise formulated, as well as being easy to answer. As a cause of using smartphones for data collection, the number of questions might need to be reduced. The questions should also be prioritized according to their importance in case some respondents leave the survey before completion. It is better to get some data than nothing at all; each provided answer to a question should directly be stored after submission.

R5 - Visual appearance

The visual appearance of a system is not only being related to a design preference but are also related to other aspects like psychology. To be asked to create patterns can for some respondents be a daunting task as it is related to security.

To avoid respondents having a bad experience by visiting the survey, several design requirements should be used for ensuring a pleasant experience while answering the survey. *Firstly*, use bright colors that are calming. The use of colors is not related something friendly. *Secondly*, the use of icons can be welcoming. The use of icons, especially icons of a childish touch should be used for giving a good experience, hence avoid scaring off potential respondents.

R6 - Navigation

The use of a smartphone as a platform requires navigation to be easy and intuitive to use. When evaluating the usability of a system, the number of clicks used for reaching a goal can be used. The navigation should in some way be automatic when the participant selects their answer to avoid too much of a time used for navigating. However, when sending the respondent to a new state, the transformation should be intuitive without the respondent losing track of the state. As well as being fast, the navigation should clearly visualize the selected answer before changing any content.

The number of questions and the scope of a survey are essential information when deciding to participate or not. To signalize the number of questions, a progess bar of some extent should be included in the survey.

R7 - Security

Since the survey is required to be anonymous for ethical concerns, the communication should be transferred over an encrypted channel using SSL. Using SSL avoid the probability of someone monitoring the traffic. The use of SSL does also provide the requirement for anonymity, as well as a visual appearance of the implemented security mechanisms. The use of SSL provides the secure HTTP (HTTPS) flag in the navigation bar, helping to build upon the trustworthiness of the survey. Security is also a part of being trustworthy.

On the backed of the application, all logging and traces of the users should not be stored. An IP address is a piece of information that can be used to trace the position of the user, hence should not be stored in any terms. The only data stored form the survey should be the answers provided by the participants. Besides turning off any logging as well as using SSL, other security measures should be implemented to obtain a secure application. The application should be reviewed by someone with a high competence in information security before starting the data collection. It is not desired to lose any data, or a situation where someone can steal the data in any terms. If any cookies are used, it should be clearly stated in the introduction. The introduction should also include a description of any security related information concerning a respondent.

4.2 Layout and Structure

The layout and design were first drafted in my specialization project in 2014, and the first wireframes can be found in Appendix A. Since the first drafts were made, a technical implementation and redesign have been carried out. Last section included an elaboration of the requirements of the survey while this section will go a step further and describe the survey and its functionality in use.

The survey can be divided into three unique parts; introduction, pattern creation, and background information. All three parts are being described in detail throughout this section. All figures of the application in this section are pictures of the actual application used for collecting data.

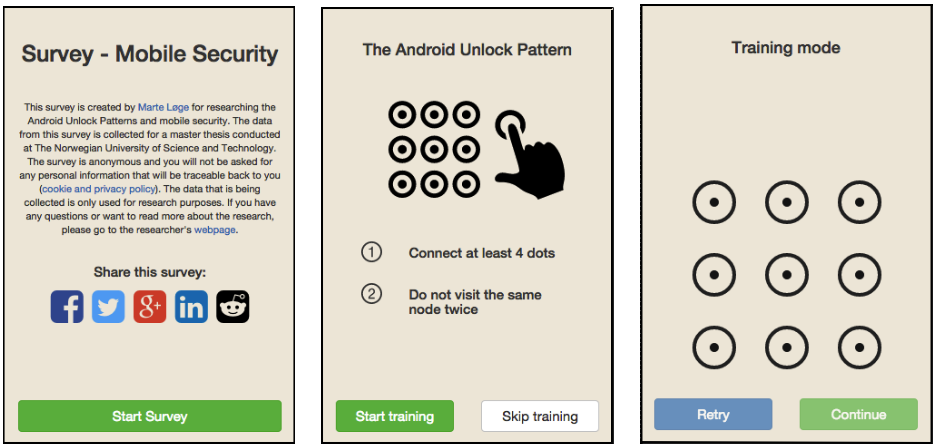
4.2.1 Introduction

When entering the survey on a smartphone, all users will be presented with an introduction to the survey, the research, and the author. The survey will not start collecting any data before a visitor decides to participate by pressing "Start Survey", as illustrated in Figure 4.1(c). When the visitor clicks on the green button, the visitor becomes a participant in the study. Figure 4.1(d) is the first screen in the survey, explaining how the Android Pattern Lock works. It is important to give a brief explanation to users not familiar with the Android Pattern Lock while at the same time avoid giving too much information causing any bias.

Figure 4.1(d) shows the next path in the survey; either enter the training mode or skip the training mode. If a respondent has never used the Android Pattern Lock before, there is added a training mode. Such training mode can avoid uncomfortable situations where the user is getting the feeling of being tested for something they're not managing. The training mode is an opportunity for the participant to play with the pattern without feeling any pressure to perform. The training mode is also an opportunity for collecting extra data in an another context than the three patterns types obtained later in the survey.

Patterns created in the training mode can be as valid as the other pattern types collected later in the survey. The patterns created in training mode might be the first patterns that pop into the respondent's mind, hence avoiding respondents to trying to overcompensate as a cause of being under pressure. As far as this research know, there are not found any research on how people think when asked to create a password or being asked to give away a password. It is believed that asking people to "give away" a password or pattern will introduce the effect of people overcompensating by creating longer passwords than typically created.

Figure 4.1(e) shows the training mode. When creating a pattern, the view will give feedback to the respondents according to the rules specified in Figure 4.1(d). In the training mode, the participant can create as many patterns as they like by pressing the "Retry" button. For leaving the training mode continuing with the survey, the respondents clicks the "continue" button. The respondents entering the training mode are asked the same questions as the participant selected the "Skip training" in Figure 4.1(d).



Survey - Mobile Security

This survey is created by Marte Løge for researching the Android Unlock Patterns and mobile security. The data from this survey is collected for a master thesis conducted at The Norwegian University of Science and Technology. The survey is anonymous and you will not be asked for any personal information that will be traceable back to you (cookie and privacy policy). The data that is being collected is only used for research purposes. If you have any questions or want to read more about the research, please go to the researcher's webpage.

Share this survey:

The Android Unlock Pattern

① Connect at least 4 dots

② Do not visit the same node twice

Training mode

(c) Start screen

(d) ALP introduction

(e) Training mode

Figure 4.1: Survey screens

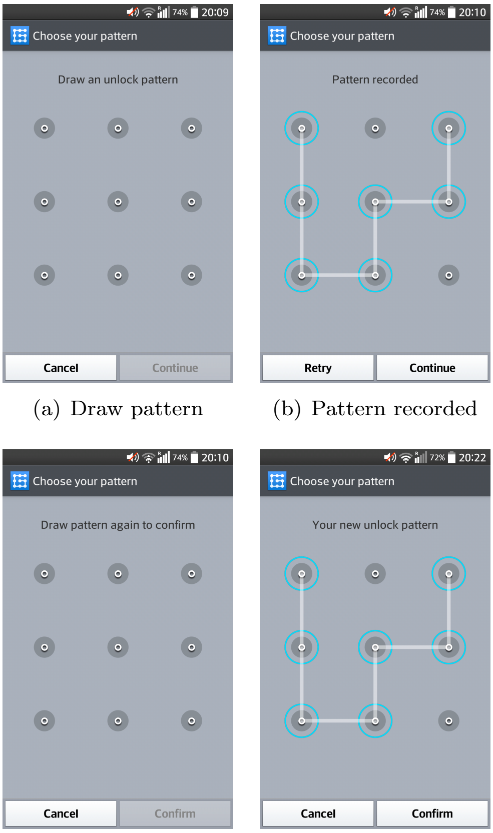
4.2.2 Pattern Creation Process

After introducing the research and the Android Pattern Lock, it is time to start collecting the three main selected patterns types; shopping account, smartphone, and banking account. The reason for dividing the pattern collection process into three separate patterns is to put the pattern creation into a context. The decision to add three patterns also works as a preventive measure for avoiding data being submitted by respondents just trying to finish the survey as fast as possible.

By introducing three separate pattern types instead of only one type, can introduce both positive and negative effects. When asking respondents to create a pattern for other types than its intended environment, some people might be confused. The choice of asking people to create a pattern for a banking account can for someone be an uncomfortable situation, might causing respondents to leave the survey. By asking the participants to create three patterns instead of one pattern takes more time and requires more attention and creativity from the participant. When asking respondents to create three patterns can introduce a situation where some respondents are creating the same pattern for all pattern types.

In the survey, respondents are asked to create a pattern for a shopping account, a smartphone, and a banking account. The three types are being carefully selected of how people would categorize different situations from a security point of view. As mentioned, the three pattern types are introduced to set the pattern creation process into context, avoiding dishonest attempts for answering the survey.

When collecting the patterns, it is desirable to copy the pattern creation process used by Android. The process consists of two steps: create a pattern of at least four connected dots, after that retype the same pattern for completing the process. Figure 4.2 shows the pattern creation process from a real Android smartphone. First, the user selects a pattern of at least four connected dots as illustrated in Figure 4.2(a). Second, after creating a pattern, the user are asked to retype the same pattern as illustrated in Figure 4.2(d). If the redrawn pattern was correctly typed, the pattern creation process are finished. The pattern creation process used in the survey are visualized in Figure 4.3 and 4.4.



(a) Draw pattern (b) Pattern recorded

(c) Redraw pattern (d) Confirm new pattern

Figure 4.2: The Android pattern creation process

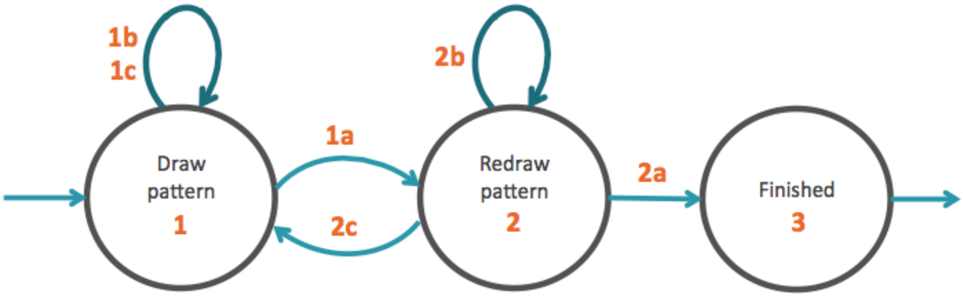
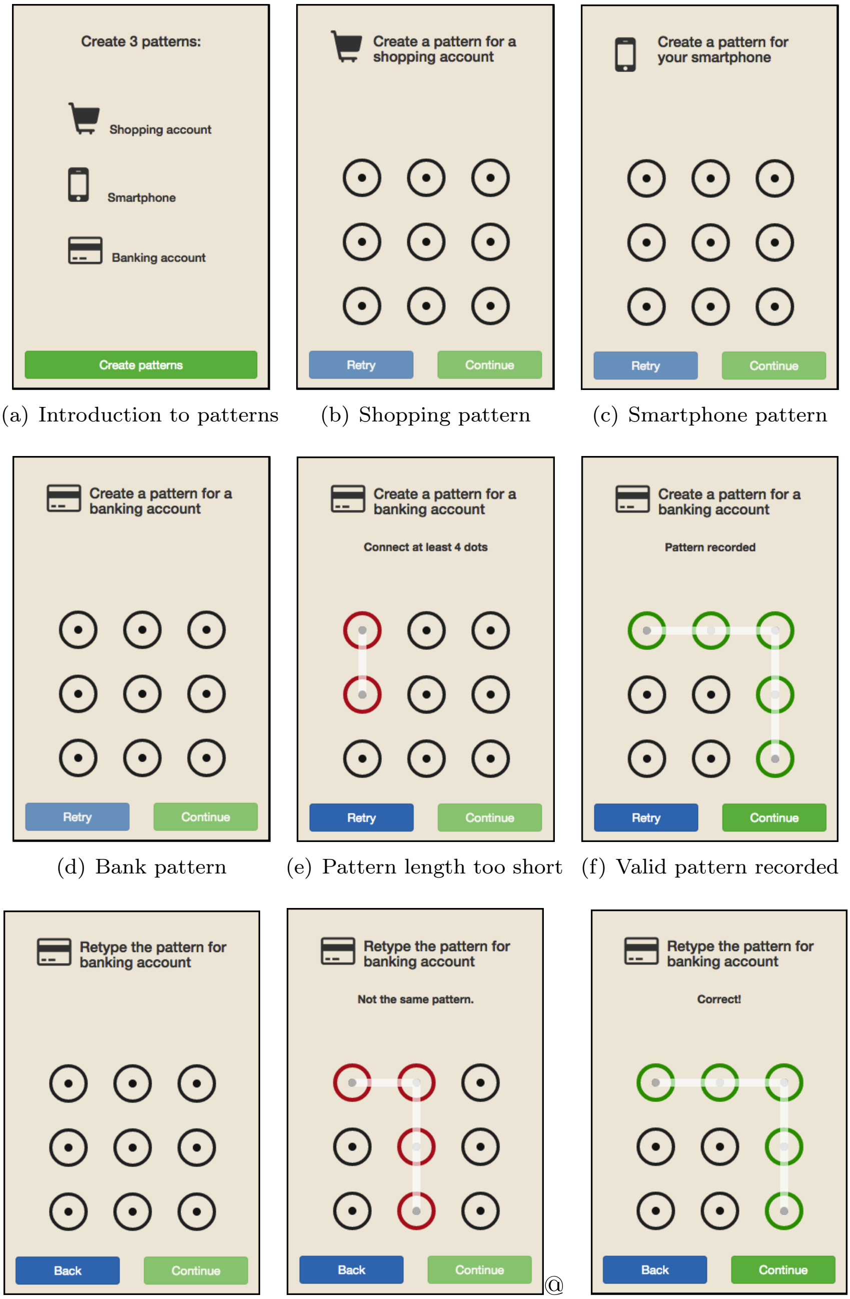


Figure 4.3: Pattern creation workflow

1. The first step is tp draw the pattern. The patter are drawn by connecting the nodes on the grid creating lines between the nodes (Figure 4.4(d)). The user will not be able to proceed before a valid pattern is drawn.
   (a) If the pattern is a valid pattern, the pattern turns green, and the message "pattern recorded" are shown. After, the user are able to press the *"Continue"* button to proceed to step 2 (Figure 4.4(f)).
   (b) IIf the pattern not being valid pattern, the pattern turns red, and the message *"Connect at least 4 dots"* are shown (Figure 4.4(e)). The user can press the button *"Retry"* for redrawing a new pattern.
   (c) If a pattern drawn is a valid pattern, but the user want to create a different pattern, the user can always press the "Retry" button to reset the pattern already created.
2. When the respondent has created a valid pattern in step 1, they are proceeded to step 2 for redrawing the created pattern. The view is very similar to the view in step 1 beside the buttons and the text (Figure 4.4(g)).
   (a) If the user successfully redraws the pattern from step 1, the pattern turns green, and the message *"Correct!"* appears as in Figure 4.4(i). The user can complete the pattern creation process by clicking the green button *"Continue"* to proceed to step 3 and complete the pattern creation process.
   (b) If the user unsuccessfully manages to redraw the pattern cerated in step 1, the pattern turns red, and the message "Not the same pattern" appears (Figure 4.4(h)). The user can try to remember the pattern and try to draw again. A situation like this can appear when the user do not remember the pattern or incorrectly draws the pattern.
   (c) If the user do not remember the pattern created at all, the *"Back"* button can be hit to go back to step 1 where the user are able to recreate a new pattern.
3. When the user successfully redraws the pattern in step 2, the pattern is successfully created.



Create 3 patterns:

Shopping account

Smartphone

Banking account

Create a pattern for a shopping account

Create a pattern for your smartphone

(a) Introduction to patterns (b) Shopping pattern (c) Smartphone pattern

Create a pattern for a banking account

Create a pattern for a banking account

Connect at least 4 dots

Create a pattern for a banking account

Pattern recorded

(d) Bank pattern (e) Pattern length too short (f) Valid pattern recorded

Retype the pattern for banking account

Retype the pattern for banking account

Not the same pattern.

Retype the pattern for banking account

Correct!

(g) Retype pattern (h) Retype wrong (i) Retype correct

Figure 4.4: Survey - Create and retye patterns

4.2.3 Background Information

After going through the pattern creation process for the three pattern types, the respondent will now be asked several questions about their mobile and personal characteristics. Figure 4.5 shows screenshots from the survey application and the questions asked. The views in Figure 4.5 is the final design used used for collecting data. The changes made during the design and implementation are described in Section 4.3.

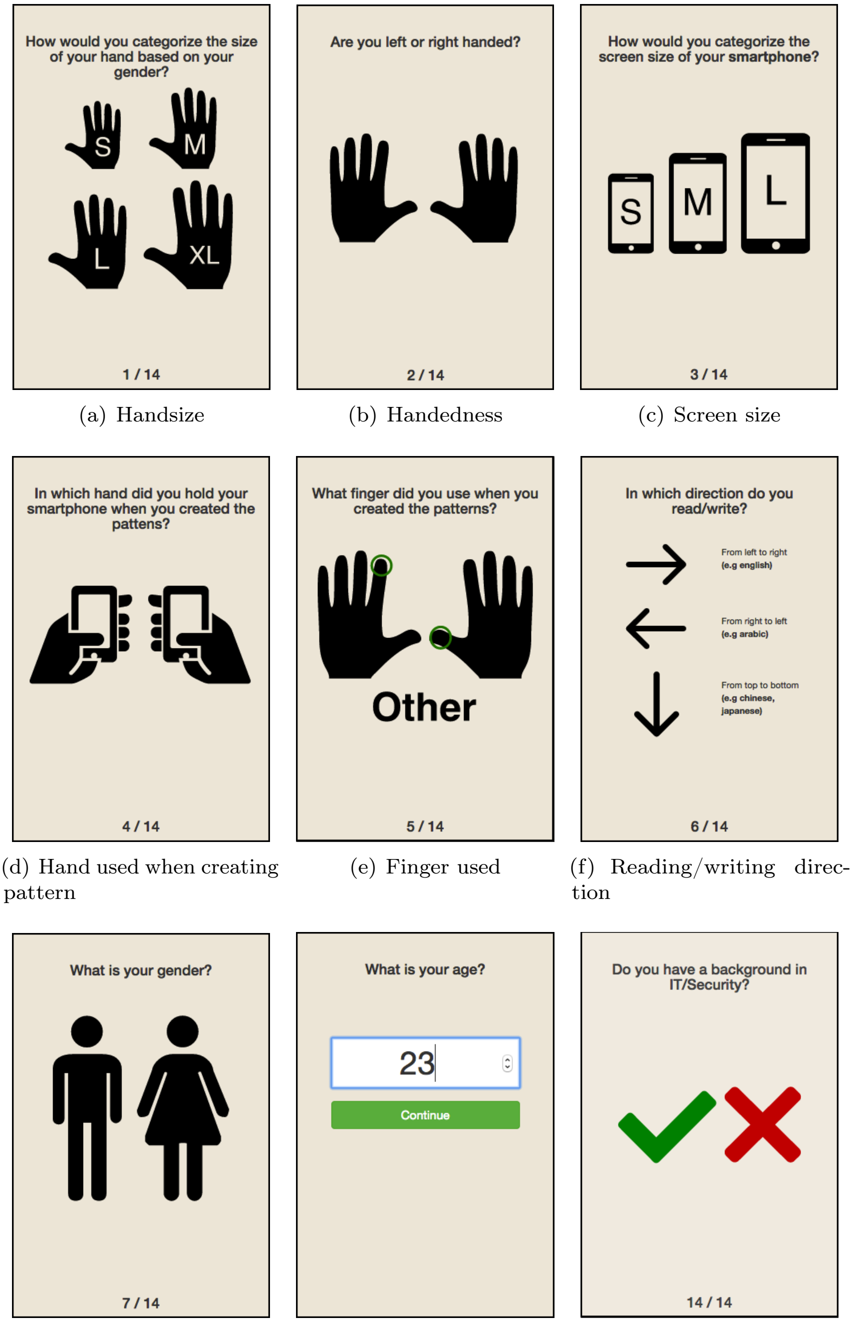
There are some core functionality implemented in these questions that is important to notice. *First*, the traditional list of alternatives are replaced by icons. As discussed in Section 4.1, it is important to customize the interface for easy interaction on a mobile touchscreen. The images used are easy to interact with because of its size, and it requires less text because of the semantic meaning of the icons used. The icons used in the survey has been tested in a own usability test that are described later in Section 4.3.1. *Second*, to keep the user motivated, a progress bar is added to the views. The questions are listed after their appearance in the survey. Not all the questions are visualized in figure 4.5, but all will be described in in detail in this section.

The first questions to be asked is the handsize of the respondent, ranged from small to extra large as visualized in Figure 4.5(a). The question is a subjective categorization, and there is no way of doing any countermeasures for avoiding wrong classification. There is often a difference in hand size of genders, so the question is asking the respondents to categorize the size of their hand according to their gender. By specifically asking for hand size based on gender will give a more precise answer. The male respondents claiming to have a small hand will probably have a bigger hand compared to female respondents claiming to have a the same size.

Figure 4.5(b) and 4.5(c) are the handedness of the respondent and the size of the screen used, respectively. Whether a person should be straight forward, where your dominant are either left- or right-hand. Screen size has the same problem as asking for the hand size, because how a respondent categorizes the screen size depends on the subjective evaluation made by the respondent. A countermeasure for having the possibility to evaluate respondents categorization of the screen sizes is to store the size of the screen in pixels. The task of storing the pixel width and height of the screen are stored automatically when the respondent selects the screen size.

Figure 4.5(e) and 4.5(d) is questions asking for the hand and finger used for creating the patterns. Typically, a respondent will either use their left or right hand for holding the smartphone while interacting with the screen by using either the forefinger or the thumb. The option of using another finger than forefinger and thumb are also applied. Instead of using words like thumb and forefinger, a circle is applied to a hand indicating the finger used.

Figure 4.5(f) are asking about the reading and writing direction preferred by the respondent. The three alternatives includes a small example to avoid any misinterpretation of the icons. Figure 4.5(g), 4.5(h), and 4.5(i) are asking about the respondent's gender, age, and experience with IT and security. The alternatives for gender are visualized by using a male and female icon. The last question asks whether the respondent have any experience with IT and security. The screens missing from Figure 4.5 are the question asking for the screenlock the respondent use, country, and the last screen thanking the respondents for participating.



How would you categorize the size of your hand based on your gender?

(a) Handsize

Are you left or right handed?

(b) Handedness

How would you categorize the screen size of your smartphone?

(c) Screen size

In which hand did you hold your smartphone when you created the pattens?

(d) Hand used when creating pattern

What finger did you use when you created the patterns?

(e) Finger used

In which direction do you read/write?

From left to right (e.g english)

From right to left (e.g arabic)

From top to bottom (e.g chinese, japanese)

(f) Reading/writing direction

What is your gender?

(g) Gender

What is your age?

(h) Age

Do you have a background in IT/Security?

(i) Experience

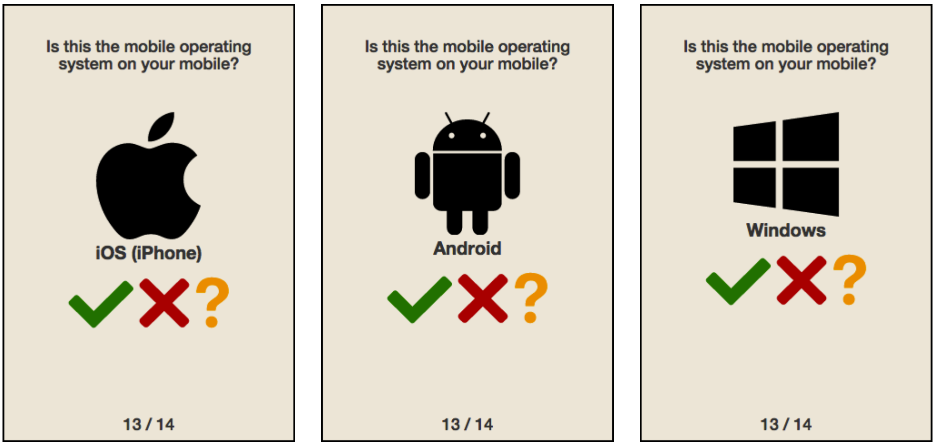
Figure 4.5: Survey - Questions

There is one special question in the survey that is not being mentioned until now; the mobile operating system used. This question is one of the hard questions to ask because it is no intuitive way for asking about the operating system of a mobile without using any technical terms.

Instead of listing all the possible mobile operating systems for mobile devices, the mobile operating system of the mobile is detected and visualized. In other words, the survey checks for the mobile OS and present what is being detected. With this information, a generic question like *"Is this your mobile operating system?"*, hence avoiding a long list asking *"Which one is your mobile OS?"*. The alternatives presented, `Yes`, `No`, and `Don't know`, are designed in way that the answer provided by the respondent are valuable whatever they select. For each mobile OS, each corresponding alternative are stored as `mobileOS_yes`, `mobileOS_no`, and `mobileOS_unknown`.

When the OS have been detected, one of the screens in Figure 4.6 will show. It is implemented support for four different mobile operating systems that are covering the majority of the operating systems on smartphones: iOS, Android, Windows, and BlackBerry. Each OS will have the three alternatives yes, no, and question mark.

One pitfall is the formulation of the question. I should avoid any formulations stating that I detects their information. A question like *"It is detected that you have an Android phone. Is this correct?"* can be daunting, hence causing the respondents the feeling of being monitored. The feeling of someone knowing anything about you can be scary for anyone not knowing the technical details. The mobile operating system are easily detected on the client side with one line of code with *JavaScript*.



Is this the mobile operating system on your mobile?

Is this the mobile operating system on your mobile?

Is this the mobile operating system on your mobile?

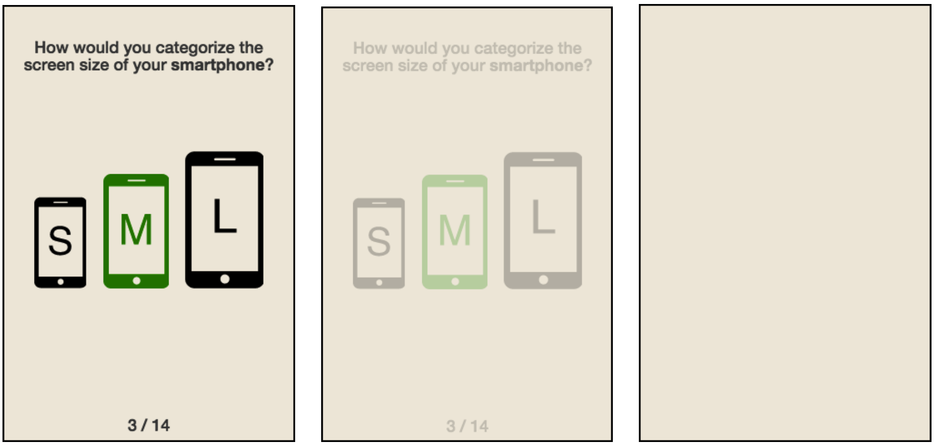
(a) iOS

(b) Android

(c) Windows

Figure 4.6: Survey - Mobile OS

Figure 4.7 shows an example of what happens when the respondent clicks on an icon. When an icon is being press, the icons are highlighted by turning green like in Figure 4.7(a). When an icon is being clicked on, a slow fading effect starts for indicating the state changes to the next question. The fading effect are visualized in Figure 4.7(b) and 4.7(c). When the question is faded out, the next question appears. The fading effect implemented is a cause of the described navigation requirements stated in Section 4.1. By using the automatic navigation while clicking on an image, the highlighting and the fading effect supports the respondent keeping track of the state. If a view is not utilizing icons, the fading effect is still used for indicating a change of state.



How would you categorize the screen size of your smartphone?

3 / 14

How would you categorize the screen size of your smartphone?

3 / 14

(a) Select icon (b) Icon fading out (c) Icon faded out

Figure 4.7: Survey - Icon selecting effect

4.3 Usability Testing of the Survey

The usability of the survey was being tested before starting collecting data for different purposes. *First*, when sending out the survey, there is no much room for changing the layout and content of the survey. The layout and content should be looking the same for all respondents. A change in the formulation or a change in the layout can cause respondents to interpret a question differently. *Second*, the survey is in a non-controlled environment, meaning that knowing who the participants are and where they are from are not known. The questions and layout needs to be created as universal and as intuitive as possible. The questions, the flow, and the graphical elements should be intuitive across different countries and cultures.

The usability testing is divided in two main parts: usability testing in a controlled environment and usability testing in an uncontrolled environment. Beside the specific usability tests performed, it has also been performed testing on the selected icons as well as the wireframes created in the preliminary work /citeforprosjekt. A summary of usability test of the icons will be provided in this section. The usability test performed on the wireframes provided in Appendix A resulted in the first implemented version being tested in this study.

4.3.1 Testing the Icons

As a described in the requirements, respondents should be able to understand and interpret the questions correctly regardless of the language spoken. As a cause of the requirement, icons were chosen to be used instead of using text. If the icons should work as intended, the icons should be properly tested.

When testing the icons, the goal is to get the participants in the test to interpret what being asked for by only looking at the icons. In other words, the respondents had to guess the formulation of the question by looking at the icons. The test was performed by using all the icons selected for the survey. The test were performed on 12 students; 5 female and 7 male students.

All the respondents managed to guess the questions asked for most of the questions to some extent. Each attempt had some variations but were considered to be correctly enough for passing. When the icons are used in the survey, the question is provided. Some of the participants had some problem interpreting the different locking mechanisms in the screen asking for the locking mechanism currently in use, especially the *slide-to-unlock*. Some of the questions used a question using the alternatives with a green check mark, and a red cross mark indicating the answer no. For the check mark and the cross, the respondents were asked to interpret thir purpose. All of the respondents either categorized the icons as "agree/disagree" or "yes/no".

4.3.2 Testing the Survey in a Controlled Environment

To perform a usability test in a controlled environment, meaning that one responsable person have set up an environment being observable, as well as that person being present. The test was conducted in a quiet meeting room by testing 10 students, 5 boys and 5 girls from the Norwegian University of Science and Technology. The participants had different background, but the majority were studying Computer Science.

To ensure the test was being conducted in an environment close to how the survey works, they brought their own device and they were not being told much about the research. All the participants got the same introduction and instructions of how the test worked:

1. The participant were asked to speak aloud during the test and tell what they were thinking and reason about their choices.
2. The participants were told that the test was not testing their ability to finish the test.
3. The participants were explained that they could quit the test if they felt uncomfortable

After giving an introduction to the test, the participant got the URL to the survey, and they were asked to start whenever they felt ready. In the list below there is listed comments from the participants that was stated during the test. The following statements listed below were useful statements from the persons testing the application. The statements are either useful comments provided during the test or observation might causing a need for doing changes in the application:

"How do I get back if I press a wrong answer?"
One of the persons asked how it was possible navigating back to a previous question if a wrong answer were being selected. The participant expected a button working similar to an undo button for being able to undo a wrong answer.

"Too many icons on the screen for representing different screen locks"
The person did find it confusing with all the icons representing the different screen locks. The participant also found it hard to interpret and select the intended screen lock. As this caused problems, the respondent would like this being presented in an another way, hence eliminate the ambiguity and confusing visualization by using too many icons. One suggestion was to add a description to each of the icons or to replace the icons using a list with only text.

"I need to choose a pattern that is hard to guess for the banking account!"
Most of the participants used more time creating a pattern for a banking account than for the other patterns. Some of them also commented that they did spend more time creating this than the other patterns because they felt the need for creating protecting their banking account more than for the other types. It was also noticeable that the participants used more time thinking before creating the pattern for the banking application. Some of the respondents also acknowledged that they used a pattern for smartphone that they had used or were currently using.

"Do I choose the size of my hand based on my gender?"
Four of the participants was uncertain on how to categorize the size of their hand.

Some of the participants commented on whether they selected their hand size based on their gender or categorizing the hand based on unisex sizes.

"Tody, my smartphone would probably be categorized as a medium smartphone"
The question is by definition subjective, but there is no easy for asking this differently. The provided answer might depend on the person's technical experiences with smartphones. Today, most of the smartphones would be categorized as a medium size, while older smartphones would probably fall into the category small size. For avoiding any ambiguities, an new solution should be found ensuring quality in the data collected.

"I expected to type the pattern after creating the pattern"
One of the participants expected to be asked to retype the pattern. The respondent did not have any preferences of when the patterns should being asked to be retyped. Two suggestions was to either copy the pattern creation process of Android or asking the respondents to recreate the patterns at the end of the survey.

"I do read from left to right, but I do also read lines from top to bottom."
Three of the participants either was unsure what the question asked about or did not understand the icons used. The view should be redesigned to avoid any ambiguities in the presentation. One provided from the respondents was to used examples next to the icons.

"The last question about my experience with IT and security took a while to understand."
The question was long and complex. Many of the participants used a long time interpreting the question and some of them asked me during the test if they had understood the question correctly. The formulation should be changed for avoiding such situation.

4.3.3 Testing the Survey in an Uncontrolled Environment

Observation of incoming data is not a typical usability test but is rather a quality check to see whether the collected data is reasonable. When sending out the survey without being present, it is important to validate the incoming data to see if any respondents have any problems related to completing the survey or interpret the questions. As a test, the survey was first distributed to a small group of selected people from Norway (ISO country code 'NO').

When looking through the data, it was observed that one person were registered rare and small country in Africa with the ISO country code starting with 'N'. It is reasonable that this was a typing error or something wrong with the component used in the question. The component used in the question was a third party dropdown list with a flag icon for each country. It were observed that the component used were using a long time loading the flag, hence lagging while the respondents interacted with the component. The component should be changed and optimized to avoid such situation.

The majority of the people asked to participate in an early release were living in Scandinavia, hence having a reading and writing direction from left to right. When looking through the collected data, 2 out of 25 respondents had selected otherwise. Based on statistics, there should only be persons with reading and writing direction from left to right. One assumption is that the icons used are being ambigiuos, hence interpreted wrong.

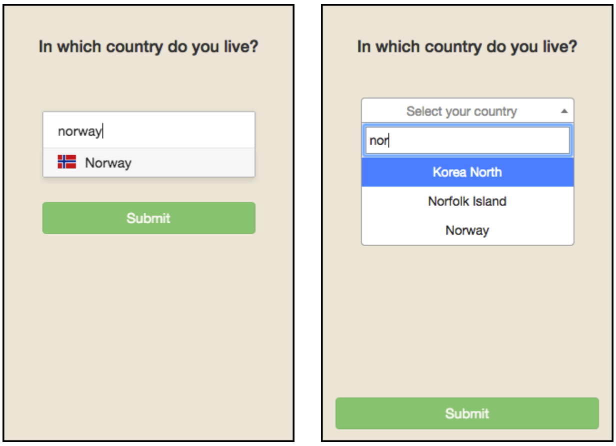
Besides looking at the data, I received feedback by email from some respondents answering the survey for testing purposes. They commented that the dropdown were lagging, as well as being confused when selecting the screen lock they were currently using.

4.3.4 Changes Made to the Survey

Before distributing the survey there were conducted two types of tests providing useful insight of how the users experienced the application. The tests were crucial because of the high need for high-quality data collected from the survey. Any ambiguities or errors causing bias in the data are not wanted. This section will provide information on all changes made to the application as a result of the observations and the feedback from the tests.

1) Country dropdown

In both tests, there were many participants that unintentionally selected the wrong country because the dropdown component lagged. The component used had a flag attached to the country name that used many resources for loading, causing the users experiencing the component to be slow. The component was changed, and the flag icons were removed. Figure 4.8(a) and 4.8(b) are showing the old and the new version of the country component, respectivley.



In which country do you live?

norway

In which country do you live?

nor

(a) Old view

(b) Final view after changes

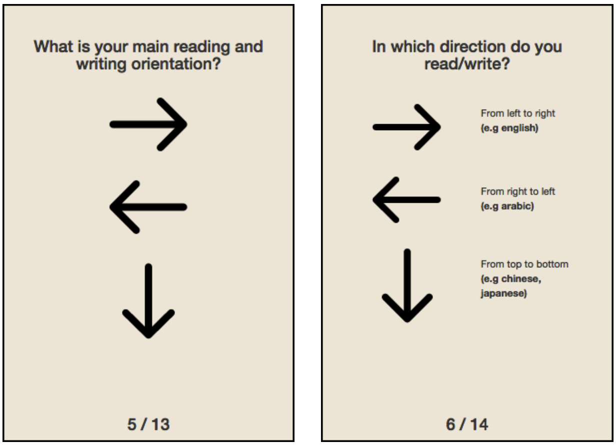
Figure 4.8: Changes in country selection view

2) Experience question

The question was rather complicated and hard to read. The question was changed from "Do you work with or study/studied IT and/or security full time?" to "Do you have a background in IT/Security?".

3) Reading and writing direction

The first version of the survey just listed three different icons of arrows indicating the reading and writing direction, but it seemed that people misinterpreted the icons. This view was changed by adding a description next to the arrow to remove any misinterpretations of the icons. The old and the new presentation of the question are shown in Figure 4.9(a) and 4.9(b), respectively.



What is your main reading and writing orientation?

→

←

↓

5 / 13

In which direction do you read/write?

→ From left to right **(e.g english)**

← From right to left **(e.g arabic)**

↓ From top to bottom **(e.g chinese, japanese)**

6 / 14

(a) Old view (b) Final view after changes

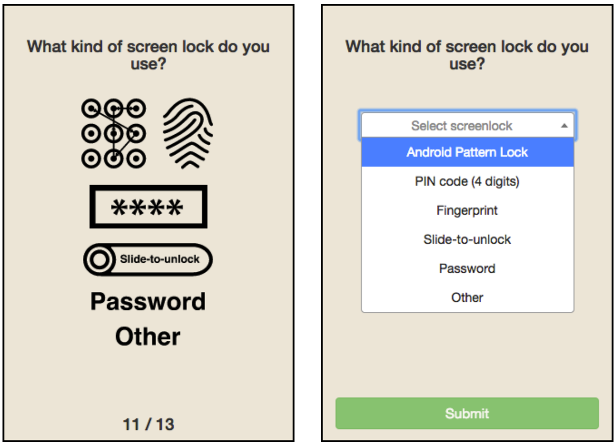
Figure 4.9: Changes in reading/writing direction view

4) Apply time used into creation of patterns

It was observed that some of the participants in the usability test in the controlled environment was spending more time creating the pattern for banking account. During the test, it was not easy to track the time used on different tasks, but this can be implemented in the application for observing the time used for creating a pattern of a spesific type. The observation from the test was that some of the participants took their eyes off the screen thinking harder about the pattern they wanted to create for the banking account. In this test, I was present and could observe such situations, but are not possible when distributing the survey. Such time frame can also be used in the analysis. An example of use could be to eliminate dishonest attempts causing bias in the data.

5) Selection of screenlock

The first version of the view had a list of all the screen locks that are provided on the majority of the smartphones. Based on feedback, the participants found it confusing to select the correct screen lock for several reasons.*First*, some of the participants did not understand all the icons used. *Second*, it was observed that some of the users did not interpret all the alternatives and selected the first matching alternative. *Third*, it seems that some had a screen lock that could be categorized as several of the presented icons, making it hard to pick the correct one. The new version of the view replaced all the icons with a drop-down listing all the alternatives. When the user selects a screen lock, an image or description appears below the dropdown, giving an extra layer of confirmation of the selected screen lock. The alternative "PIN" was also changed from "PIN" to "PIN (4-digits)" for eliminate any ambiguities for those using more than four digits. Figure 4.10(a) and 4.10 are the old view and the new view, respectively.



(a) Old view

(b) Final view after changes

Figure 4.10: Changes in screenlock selection view

6) Changing handsize question

Based on feedback from the test conducted in the controlled environment, the participants did not know what to compare their hand size to. The size selected depended on whether they compared themselves to their gender or not. For being consistent, the question explicitly stated in the question that the user should select their size based on their gender.

7) Adding pattern retype

During the tests, it was received feedback that many of the participants expected to be asked to retype their selected patterns. Two suggestions from the respondents were to copy the process used on Android devices or asking respondents to retype in the end of the survey. It is not desired to ask users to retype their patterns at the end of the survey because such situation can make respondents feeling if not being able to remember the pattern. As stated in the requirements, respondents should not leave the survey with a bad experience. If asking the respondents to retype the pattern at the end of the survey, this should be clearly stated before creating the patterns. By stating that the users should remember the patterns for being able to complete the survey could cause respondents feeling uncomfortable or even cause the respondents to leave the survey. The best solution was to copy the same process that Android are using.

8) Collecting screen pixels of screen used

By observing the participants in the first test, it was observed that the selected choice in patterns was different from person to person. Since this question is a subjective meaning of the person answering, an extra layer of information is applied to the dataset by saving the pixels width and height of the screen. One problem with pixels size of a mobile screen is that they do not match the physical size of the screen. The pixels could still provide information for being able to categorize the size correctly by comparing the OS and pixels to the respondents answer.

9) Shuffle the ordering of the patterns

One concern with asking the users to create the patterns in a predefines order as discussed until now, respondents might be influenced by the order of the patterns.

To be sure that the ordering is not introducing bias, a method named *Latin Square* can be used. A Latin Square is a table filled with n different symbols in such a way that each symbol occurs exactly once in each row and exactly once in each column. Instead of using three different orderings, all possible combinations of the three pattern types will be used. This results in 6 different ways of ordering the patterns.

By using the Latin Square, it ensures that the ordering do not impact the respondent's choice of patterns. Or in other words, the data can be validated afterwards to see if the ordering have affected the collected data. In practice, respondents will create the same patterns, but in a different order.

4.4 Ethical Considerations

My research is performed by collecting demographics and other information about the respondents and the devices used. The data are not by definition considered to be sensitive information, in other words, the information can not be linked back to the respondent. It is still important to evaluate the work being done for avoiding conducting research not considered being ethical. As this research are collecting user-selected patterns, a set of countermeasures have been implemented for ensuring anonymity of the respondents.

Before the survey starts, the respondents are informed about the purpose of the research and how their contributed data will be managed. Respondents are also advised that they have the right to leave the survey whenever wanted. The questionnaire should be entirely anonymous, and the identity and location of the respondent should not be possible to track back to the respondent. As the selected sampling technique supports, a list of respondents is not kept in any form as the respondents participate voluntarily if receiving the survey. Other measures applied are turning off the logging of the IP addresses on the server side. The application are made by the researchers and for this research in particular. Therefore, no third part is included or have access to the data, making it possible to ensure anonymity.

Before collecting data, it is required to get an allowance for collecting data from the *NSD*, in Norwegian called *Personvernombudet for forskning* [27]. The NSD is the privacy ombudsman for approximately 150 research and educational institutions in Norway, where NTNU is one of them. The research was reviewed and accepted by the supervisor of this project and the NSD contact person from NTNU.

5 | Results

This chapter presents the results provided by the collected data from the survey. Section 5.1 is a overview of the population while Section 5.2 and 5.3 looks at the results with respect to entire population and different subgroups, respectively. Section 5.4 is the last section including a description of how the data were preprocessed before being analyzed. The subgroups included in Section 5.3 are gender, age, handedness, experience with IT and security, hand size and screen size. For the subgroups, a simple two-tailed t-test are performed to test a significant difference in two samples looking at pattern length and visual complexity.

The results in Section 5.2 and 5.3 are divided into pattern creation time, pattern length, and visual complexity. The pattern creation time is the time used for creating a pattern, recorded from the start when the grid appeared until the user submitted the pattern. The pattern length are defined as the number of dots used to form a pattern. Each dot has an own sequence number and can only appear once. The minimum length of a pattern is 4 dots while the maximum length is 9 dots. The number of unique patterns of all possible pattern lengths is described in Table 2.1. The pattern length are visualized in two different ways; the average pattern length and the distribution of pattern length. Pattern complexity, e.g. pattern strength, is calculated from a mathematical formula that utilizes the visual aspects of patterns. The formula and parameters used are described in detail in Section 2.6. In short, the formula uses the size (number of nodes), physical size (length) of the pattern, number of intersections and number of overlaps. The minimum strength is 6.340 and the maximum strength is 46.807.

5.1 The Population

This section will be a summary of the population as well as information about the data collected. The presented results are aggregated from the data gathered from the survey described in Chapter 4.

5.1.1 Number of Respondents

Table 5.1 summarizes of the number of respondents and level of completeness. Some respondents just opened the survey without providing any information, while other answered all questions. A total of 802 respondents completed the whole survey. 11 people created patterns and started answering question before leaving the survey and 204 respondents started creating patterns, but did not answer any other question. At last, 81 respondents entered the survey and left without creating any pattern or answering any questions.

Completed the survey	802
Stoped during the survey	11
Started creating patterns	204
Opened the survey	81

Table 5.1: Number of respondents

5.1.2 Demographics and Background Infromation

Figure 5.1 summarizes the populating by looking at gender, handedness, experience with IT and security, and reading/writing orientation.

The majority of the respondents was male, where 529 of the respondents were male, and 278 respondents were female. In total, 66% of the population consist of male respondents while 34% of the population consist of female respondents. By looking at handedness of the participants, 88% of the population were right-handed, and 12% were left-handed.

The percentage of left-handedness in the world is hard to estimate, but it is stated that one in ten people are are left-handed [21]. The frequency of left-handed people seems reasonable compared to the frequency of left-handed respondents in the survey that was 12% of the population.

To be able to distinguish between people with experience with IT and Security it was asked about the participants experience in the survey. The majority had a background in IT and security constituting 59% of the population, while 41% of the respondents did not have a background related to IT and security.

It is hard to reach people outside your own network, especially to reach groups of people with other cultures. In the dataset, it was only 2% that had an another reading and writing direction than left-to-right. Countries operating with a reading and writing direction other than left-to-right are Arabic countries or countries located in Asia.

The reading and writing direction can therefore not be used in any further analysis because the number of participants is too low for obtaining any significant results.

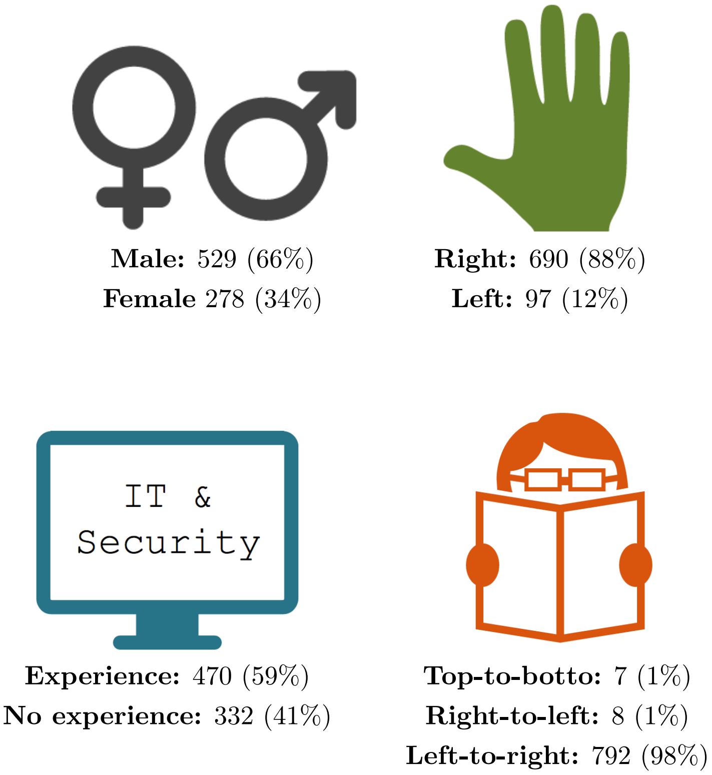


Figure 5.1: Gender, handedness, experience with IT and security, and reading/writing orientation

Figure 5.2 presents a distribution of the respondents' age. The respondents are divided into 8 intervals: under 20, 20-24, 25-29, 31-34, 35-39, 40-49, over 50. The three last intervals have a lower frequency of participants, hence being grouped into smaller intervals. The distribution have a peak at the interval 20-24. Reasons for the skewed distribution can be a cause of the network that received the survey. The majority of the networks contacted were students in their twenties, hence resulting in a skewed distribution. When looking at the right-hand side of the graph, there is a lower frequency of respondents.

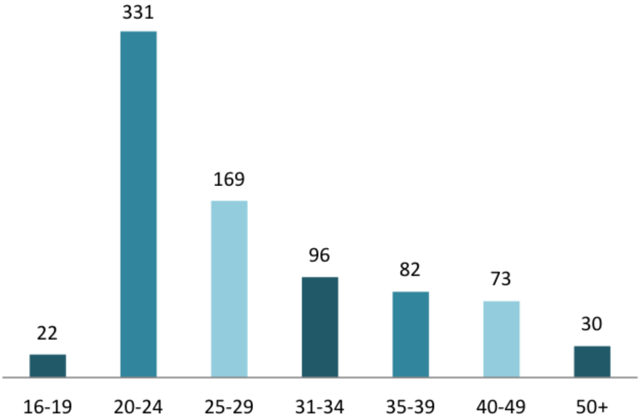


Figure 5.2: Age distribution

Figure 5.3 is the overview of the hand size of the respondents. The registered hand size is considered to be subjective and needs to be carefully used when making any conclusion. The majority of both genders classified their hand size according to their gender as a medium size. Male respondents have a higher frequency of large classifications. A more detailed quality control of the provided dataset of selected hand sizes will be carried out in Chapter 5.3.5.

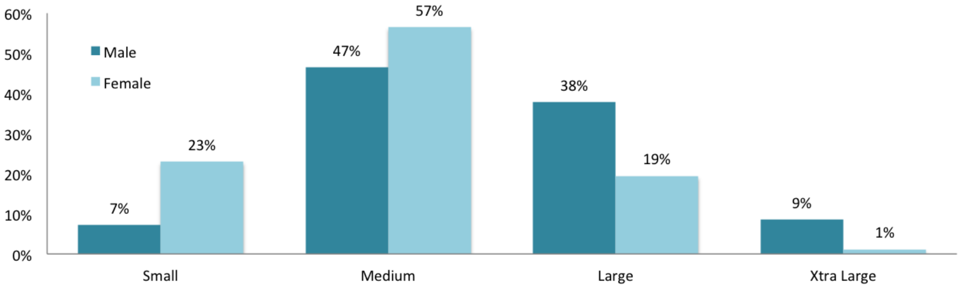


Figure 5.3: Handsize

The survey asked the participants about their country of origin and the population ended up being represented by 39 countries. The majority of the countries represented in the dataset was Norway and United States of America as seen in Table 5.2. The table summarizes all 39 countries, as well as the frequency of respondents from each of the countries. The distribution can not be used for making any conclusion of selection of graphical passwords based on the country of origin, as many of the countries represented in the population do not provide sufficient mounts of data. During the data collection it was a goal to avoid a homogeneous dataset. The majority of the participants is still from Norway, but is was important to get other countries represented as well to obtain as a more heterogeneous data set.

Country	# Respondents
🇳🇴 Norway	517
🇺🇸 United States of America	115
🇩🇪 Germany	33
🇨🇿 Czech Republic	31
🇬🇧 United Kingdom	22
🇷🇺 Russia	13
🇩🇰 Denmark	7
🇸🇪 Sweden	6
🇨🇭 Switzerland	6
🇦🇺 Australia	5
🇳🇱 Netherlands	4
🇨🇱 Chile	4
🇫🇮 Finland	3
🇦🇹 Austria	3
🇺🇦 Ukraine	3
🇨🇳 China	3
Afghanistan, Mexico, North Korea, Pakistan, Vietnam, Luxembourg, Ireland, Tunisia	2
Italy, Greece, Belgium, Indonesia, Malaysia, Bahrain, Botswana, Argentina, Singapore, japan, Canada, South Korea, Hungary, Turkey, Brazil	1

Table 5.2: Respondents country of origin

5.1.3 Screen Lock Habits and Mobile Device Used

Table 5.3 summarizes the respondents habits when selecting screen lock, particularly their use of screen lock mechanisms, and which screen lock they are currently using. The table also lists information about the mobile device used for answering the survey, like screen size and mobile operating system. The table provides as an overview of how the respondents manage the security on their smartphones.

The majority have answered the survey using a mobile running on Android or iOS that are the two most popular mobile operating systems in the market. Together, 98% of the mobile operation system used were either Android or iOS, constituting 58% and 40% of the smartphones, respectiveley.

A total of 65% of the participants had used the Android Pattern Lock before while 35% of the population were not familiar with the Android Pattern Lock. The respondents not familiar with the utilization of the Android Pattern Lock will probably have their first time using the scheme in this survey.

Looking at the screen lock habits in the population, 82% of the participants uses screen lock on their mobile device. Among the listed screen locks in Table 5.3, the majority are using either 4-digit PIN (36%), Android Pattern Lock (31%) and fingerprint (18%). The fingerprint is only available on iOS while Android Pattern Lock are not allowed on iOS. Beside the different screen locks, the mobile devices do also have different screen sizes. The stated screen size by the respondents is a subjective classification of the screen size. Further validation of the screen size and classification of correct physical size are being found in Section 5.3.5.

Screenlock in use		
Android Pattern Lock	202	31%
4-digit PIN	237	36%
Fingerprint	116	18%
Password	44	7%
slide-to-unlock	28	4%
Other	28	4%
Screensize		
Small	108	13.3%
Medium	532	65.4%
Large	173	21.3%

Mobile Operating System		
Android	464	58.0%
iOS	321	40.0%
Windows	16	1.9%
Blackberry	1	0.1%
Use screenlock		
Yes	655	82%
No	149	18%
Used Android Unlock Pattern		
Yes	526	65%
No	278	35%

Table 5.3: Information about password habits and mobile device used

5.1.4 The Created Patterns

This section is looking at the patterns created, and how the patterns were being created due to the hand and finger used. Table 5.4 summarizes the number of patterns created of the different pattern types. There is about the same abount of patterns created for all pattern types. The cause of not having the exact number of each pattern is because some of the respondents did not complete answering the entire survey as listed in Table 5.4.

The number of training patterns are larger than for the other pattern types because the respondents were able to create as many pattern as they liked. A total of 658, constituting 80% of the population, created at least one pattern in training mode. The total number of patterns collected for shopping account, smartphone, banking account, and training was 3393 patterns. The distinct number of pattern created for the different types as summarized in Table 5.4.

On the right side of Table 5.4, a summary are provided of how the respondents created the patterns. By observing normal interaction with a smartphone, the majority of people fall under two main categories in how to interact with a smartphone:

1. Use one hand using the thumb for interaction, whereas the hand are defined by handedness.
2. Use the opposite hand defined by handedness and use the forefinger on the other hand for interacting with the screen.

There are also people interacting with the screen using other fingers than the thumb and forefinger, hence provided an option for selecting otherwise. The majority of the respondents was either using their right hand using their thumb or using their left hand and their forefinger. Handedness may influence these numbers, hence provided more information of the pattern creation process in Section 5.3.3 by lookig futher at handedness, hand used, and finger used.

Patterns created	
Shopping	841
Smartphone	842
Bank	838
Training	872
Total	3393

Hand and finger used			
Left hand	Forefinger	233	28.8%
Left hand	Thumb	60	7.4%
Right hand	Forefinger	72	8.9%
Right hand	Thumb	398	49.3%
Other		45	5.6%

Table 5.4: Information about the collected patterns

5.2 Findings for the Entire Population

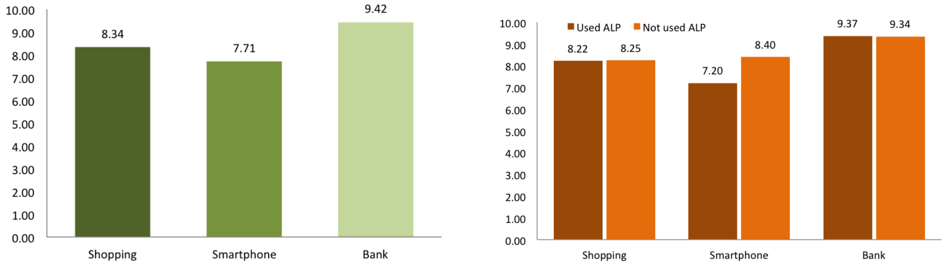
This section will go through the results from the survey based on the entire population. The results will at first look at pattern creation time, pattern length, and visual complexity as described in the introduction. The section will also provide other results found when going through the data for the entire population.

5.2.1 Pattern Creation Time

Figure 5.4(a) gives the pattern creation time in seconds for the three pattern types. By looking at the average creation time for patterns, patterns created for bank accounts have the highest creation time of 9.42 seconds while patterns created for smartphones have an average creation time of 8.24 seconds.

Figure 5.4(b) shows the average pattern creation time in seconds for respondents experienced with the Android Unlock Pattern. The graph reveals that both patterns created for shopping account and banking account are not affected by the participants experience with the Android Unlock pattern.

Patterns created for smartphones have a lower response time for patterns created by respondents experienced with the Android Lock Pattern. Patterns created for smartphones by experienced respondents had an have an average creation time of 7.20 seconds while the patterns created for smartphones by unexperiences respondents had an average creation time of 8.40 seconds. The difference in average creation time results in a difference of 1.2 seconds.



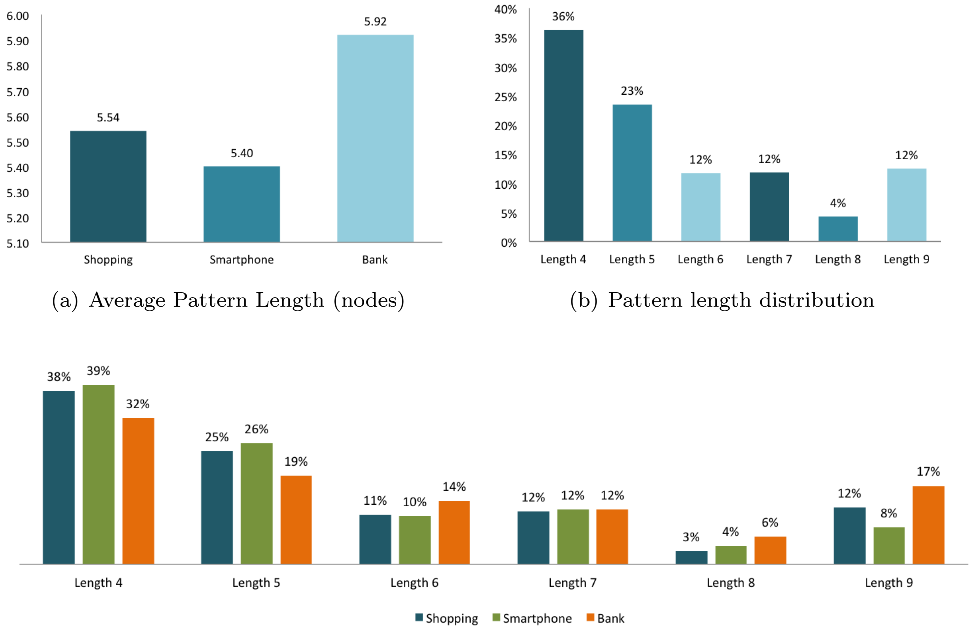
(a) Average pattern creation time (seconds)

(b) Creation time and experience with ALP

Figure 5.4: Pattern creation time for the entire population

5.2.2 Pattern Length

Figure 5.5(a) summarizes the average nodes selected to form a pattern for the each distinch pattern type. The average length is 5.54, 5.40, and 5.92 for patterns created for shopping accounts, smartphones, and bank accounts, respectively. The patterns created for bank accounts have a higher average length than patterns created for shopping and smartphone. The numbers from the graph also show that patterns created from smartphone has the smallest average pattern length. The difference in average pattern length for patterns created for smartphones and banking accounts are on average 0.52 nodes. Patterns created for shopping accounts are slightly longer than patterns created for smartphones but constitute not a significant difference between the two types.



(a) Average Pattern Length (nodes)

(b) Pattern length distribution

(c) Pattern length and type distribution

Figure 5.5: Pattern Length for the entire population

The graph in Figure 5.5(b) is a pattern length distribution indicating what pattern length that are more often selected by the population. Figure 5.5(c) is a more detailed version of Figure 5.5(b), showing the pattern distribution for all pattern types. Both graphs give an indication the respondents selecting patterns having a length longer than 5 nodes less often. Both graphs in Figure 5.5(a) and 5.5(b) contains a low frequency of pattern with length 8, where patterns of length 7 and 9 both have a higher frequency than patterns of length 8.

The pattern created for shopping account and smartphones in Figure 5.5 have the majority of the patterns distributed over the shortest pattern lengths, e.g. patterns of length 4 and 5. Patterns created for banking accounts had the highest average pattern length, but the majority of the patterns created for bank accounts have still a length of 4 or 5.

5.2.3 Pattern Complexity

Table 5.5 is a summary of all the parameters used in calculating the strength of the patterns created by the entire population. The pattern length is not being further described as it was covered in the previous Subsection. The pattern length are named *Size* when talking about the length for calculating the visual complexity.

The physical length are increased when patterns utilize lines between the nodes that not are horizontal or vertical. A longer length are often correlated with the number of intersections and overlaps. Patterns created for bank accounts have the highest strength and highest occurrences of intersections and overlaps, 0.433 and 0.023 respectively. Patterns created for banking account also have a high pattern creation time and a high average pattern length. The average strength of patterns created for banking account are 15.514, also being the highest average strength score compared to the average score obtained by the two other types.

Parameters	**Shopping**	**Smartphone**	**Bank**	**All**
#Patterns	841	842	838	2521
Avg. Size	5.541	5.398	5.920	5.619
Avg. Length	5.050	4.920	5.666	5.212
#Intersections	177	149	363	689
Avg. Intersections	0.210	0.1769	0.433	0.273
#Overlaps	15	12	19	46
Avg. Overlaps	0.0178	0.014	0.023	0.018
Avg. Strength	13.440	12.837	15.514	13.928
Min strength	6.340	6.340	6.340	6.340
Max strength	44.441	43.187	44.441	44.441

Table 5.5: Pattern strength for all patterns types in the entire population

The smartphone is the pattern type with the weakest average strength. The characteristics of patterns created for smartphones is that they have a short pattern length in terms of number of nodes as well as a short physical length. The average occurrences of intersections and overlaps remain the lowest compared to patterns created for bank. The average number of occurrences of intersections and overlaps are 0.1769 and 0.014 respectively. The patterns created for smartphones gets an average strength score of 12.339, being the lowest score compared to the average score obtained by the two other types. Patterns created for shopping account are similar to patterns created for smartphones. All the parameters for patterns created for shopping accounts have slightly higher parameter values than for smartphone, hence a slightly higher average strength of 13.440.

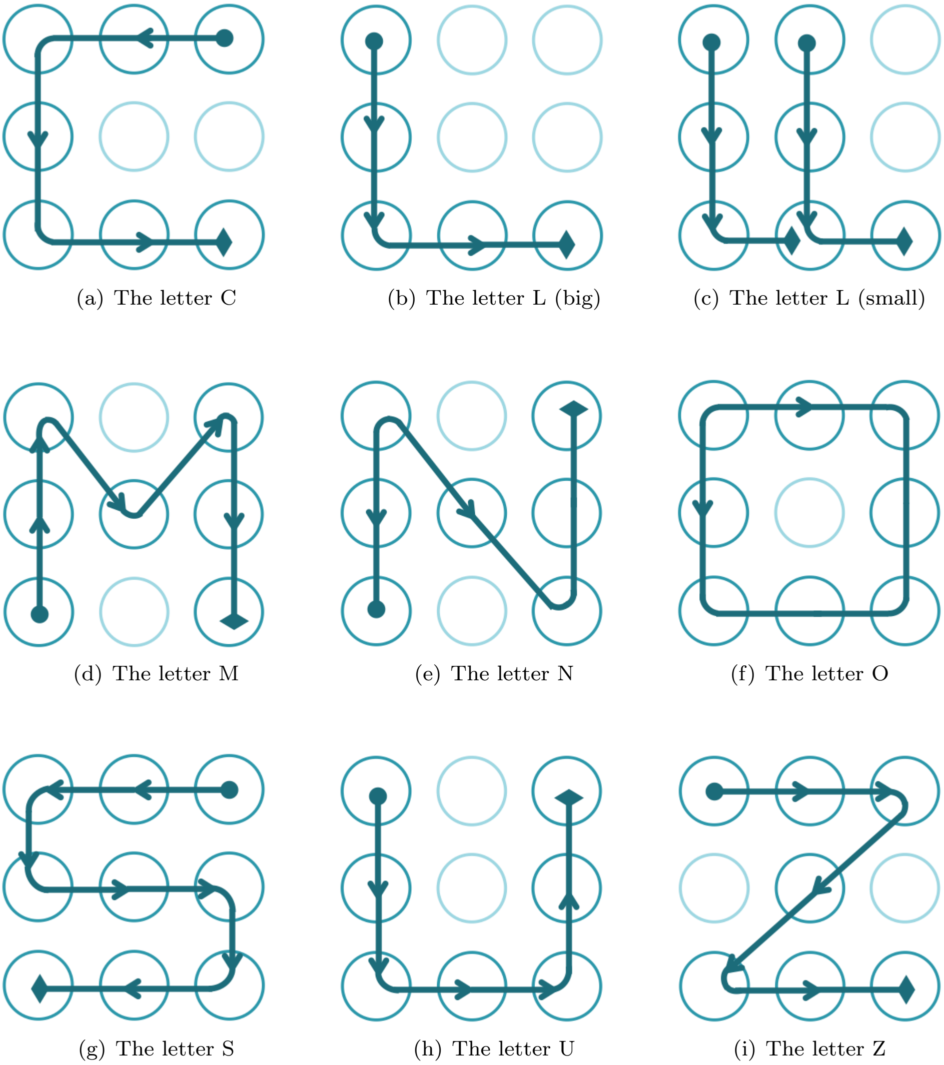
When looking at the maximum pattern strength of all patterns collected, none of the patterns obtained a maximum score of 46.8. In other words, none of the roughly 800 respondents managed to create a pattern obtaining a higher complexity score than 44.44. The set of patterns created for smartphones did include a pattern having obtained a complexity score higher than 43.187, being a lower score than for the two other types

5.2.4 Association Elements

An association element is something a person know or recognize, and can used as an element to ease the process of remembering a password. This is a known technique used by users when creating alphanumeric passwords and PIN codes. Alphanumeric passwords are often known being created containing personal information like names and dates for support the creator in remembering the password. The same strategy are being observed for PIN codes where the use of codes forming a date often occurs.

The dataset collected in this research are being scanned for patterns corresponding to association elements. By going through the alphabet, it was found 12 types of patterns corresponding to the visual representation of letters from the alphabet. Out of the 12 letters, 9 patterns had a significant number of appearances. Figure 5.6 shows the 8 most common patterns having the same visual representation as letters from the alphabet. Beside the letters C, L, M, N, O, S, U and Z, letters like G, J and W also appeared in the data set.

By iterating through the sequences corresponded to letter, 385 out of 3393 patterns in the dataset matched a letter. The number of patterns matching a letter from the alphabet constitutes 11.4% of the collected patterns.



(a) The letter C

(b) The letter L (big)

(c) The letter L (small)

(d) The letter M

(e) The letter N

(f) The letter O

(g) The letter S

(h) The letter U

(i) The letter Z

Figure 5.6: Most frequent patterns forming letters from the alphabet

5.2.5 Bias in the Selection of Start Node

The selected starting node of a pattern is crucial information when analyzing patterns. Knowing the starting node limits the number of possible patterns because a legal pattern can only visit the same node twice. When knowing the likely starting points, patterns starting with less likely starting nodes can be excluded from the theoretical password space.

Table 5.6 summarizes the likelihood of starting in a particular node for each of the four pattern types. The numbers notation used in Table 5.6, 1-3 and T, denotes a shortcut for shopping account, smartphone, banking account and training, respectively. On average, 44% of all patterns starts in node 1, e.g. the upper-left corner.

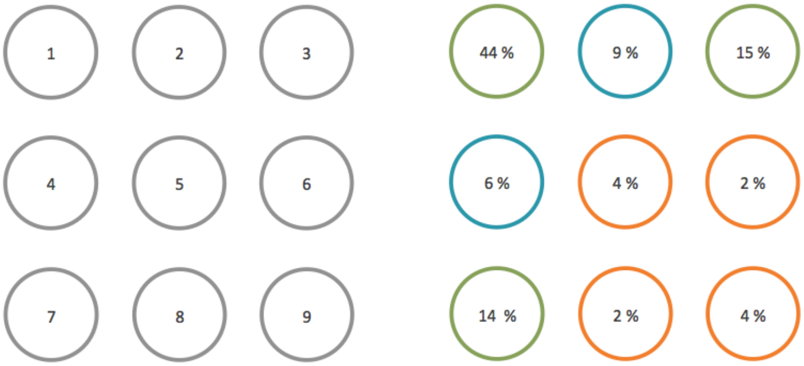
The training patterns are not included in other parts of the results because there are no control over how many times a person have entered a tarining pattern. The training pattern are still valid when looking at the starting node because it reflects where respondents starts creating patterns.

Start node	All	1,2,3	1	2	3	T
1	44%	42%	43%	41%	42%	51%
3	15%	15%	16%	14%	13%	14%
7	14%	14%	13%	15%	14%	13%
2	9%	9%	10%	9%	8%	7%
4	6%	7%	6%	7%	7%	6%
5	4%	4%	4%	4%	4%	3%
9	4%	4%	3%	3%	5%	3%
8	2%	3%	2%	3%	3%	2%
6	2%	2%	2%	2%	3%	2%

Table 5.6: Selection of starting node for all pattern types

Figure 5.7(b) is a graphical representation of the likeliehood of starting in the different nodes. Each node has a number, starting from node 1 in the upper left corner ending up with node number 9 as shown in Figure 5.7(a). All nodes in Figure 5.7(b) are colored based on the likelihood of being selected as a starting point from high to low (green - blue - orange), whereas the green nodes are the most common starting points and orange nodes is the least common starting points.

Summarizing the most common starting points, node 1, 2 and 7, they all together constitutes 73% of the patterns. The different pattern types have all over 40% of the patterns starting in the upper left corner. There are some nodes that are having a less probability of being selected as a starting point. The nodes with the lowest frequency, node 5, 6, 8 and 9, only have a total probability of 12% for being selected as the starting node.



(a) Node positions

(b) Staring node for all pattern types

Figure 5.7: Node position and likely starting point for all pattern types

5.2.6 3-gram Movement Patterns

This section includes a visualization of 3-gram movement patterns from the collected patterns. A 3-gram is a sub-sequence of a pattern consisting of 3 nodes. The notation used uses a circle when denoting the start of a 3-gram while the arrow denotes the end of the 3-gram. A 3-gram does not indicate whether a pattern starts or end. A 3-gram can appear at the start, at the end or in the middle of a pattern because a 3-gram is only a subsequence of a pattern. By iterating through all patterns by counting all possible 3-grams, a list of the commonly 3-grams are visualized as seen in Figure 5.8. The three figures contains top 20 common occurrences of 3-grams, ordered from the most commonly used (blue) to less commonly used (orange).

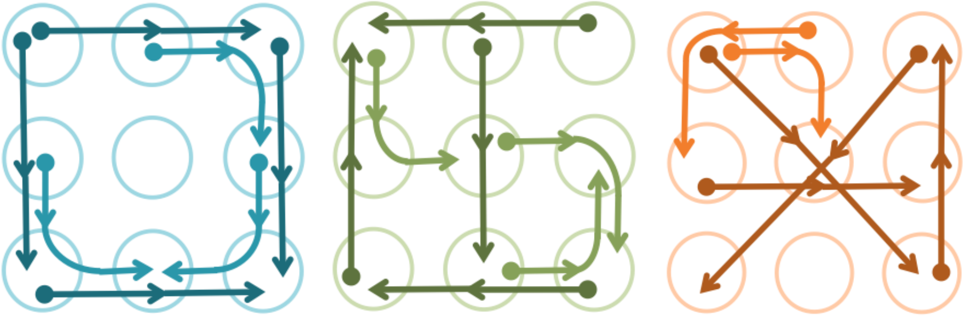


Figure 5.8: Most common 3-gram to less common 3-gram

Based on the observed bias in the selection of starting node, it is interesting to use information about the selection of starting node together with the observed 3-gram movement sequences. By looking at top 100 most commonly created patterns, the patterns top 100 patterns constitutes 42% of the collected patterns. When looking at top 120 patterns, roughly 50% of the patterns in the data set are covered. In general, when studying the movements sequences in Figure 5.8, most of the patterns are straight lines close to the edges. Diagonal lines are not used very often, whereas straight diagonal lines are the once appearing more often. 3-grams not being a straight line or curving the corners does rarly occur. An example of such 3-gram is the sequence 189.

The 3-gram with the highest frequency is the 3-gram 123 and 147, appearing in 1055 of the created patterns. Besides having a subsequence of the most frequent 3-grams, 844 of the collected patterns either started with the sequence 123 or 147.

5.3 Findings in Specific Subgroups

This section is a presentation of the results focusing on different user types. Each section is being dedicated to one user type having one of the human properties stated in the research questions in Chapter 1. Each section will include the same parts; pattern length, pattern creation time, and pattern strength. For some of the user types, extra pattern characteristics are included if some unexpected results were observed.

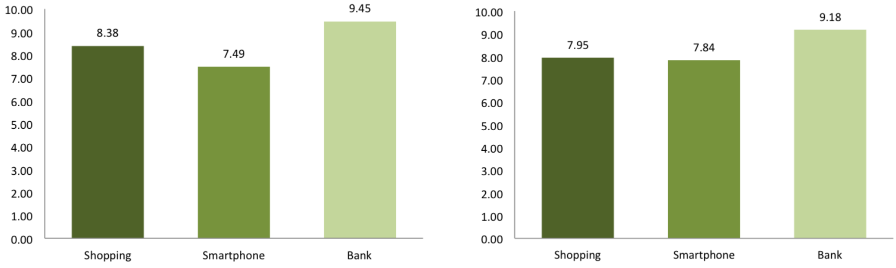
This section includes the results from a two-tailed t-test for testing the statistical significance of the results, as well as including the mean, the standard deviation, and the P-value. The t-test are being conducted on the length and visual complexity for all subgroups. A statistical significant result tells that the patterns from two samples are different, e.g. the choice in patterns are different. All tests are using a significance level of 0.05. As a result of using a two-tailed test, results getting a p-value less than 0.025 are a statistical significant result.

5.3.1 Gender

Gender are being divided into the subgroups of male and female participants. For each of the parts in presented, the results will be presented with respect to pattern type and gender.

Average pattern creation time

Figure 5.9 shows the average creation time in seconds for both genders. Male participants have in general a higher average creation time than females. The average creation time for shopping account is the only pattern type where male participants uses shorter time than the female participants for creating a pattern.



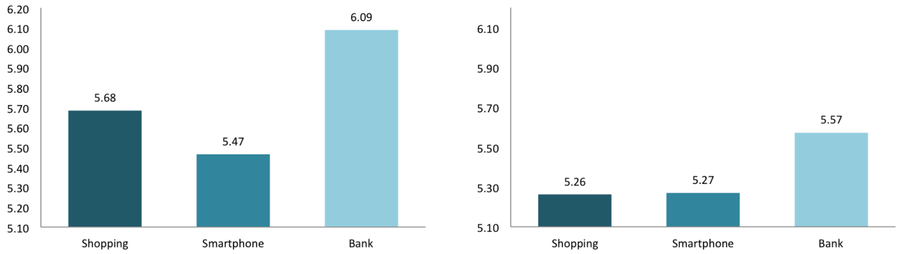
(a) Male

(b) Female

Figure 5.9: Average pattern creation time for gender

Average pattern length

Figure 5.10(a) and 5.10(b) presents the average pattern length for patterns created by male and female participants, respectively. When looking at the patterns created shopping accounts and smartphones by male participants are having a slight difference in average length. The patterns created for smartphones have the lowest average pattern length of 5.47, while patterns have the longest average pattern length of 6.09. Female participants have roughly the same average length, 5.26 and 5.27, for patterns created for both shopping accounts and smartphones, respectively. The patterns created for banking accounts have the highest average length of 5.57. Comparing the average pattern length for patterns created by both genders, both genders have the longest length for banking account and the shortest average length created for smartphones. In general, the patterns created by male participants have an longer average length than patterns created by female participants.



(a) Male

(b) Female

Figure 5.10: Average pattern length for gender

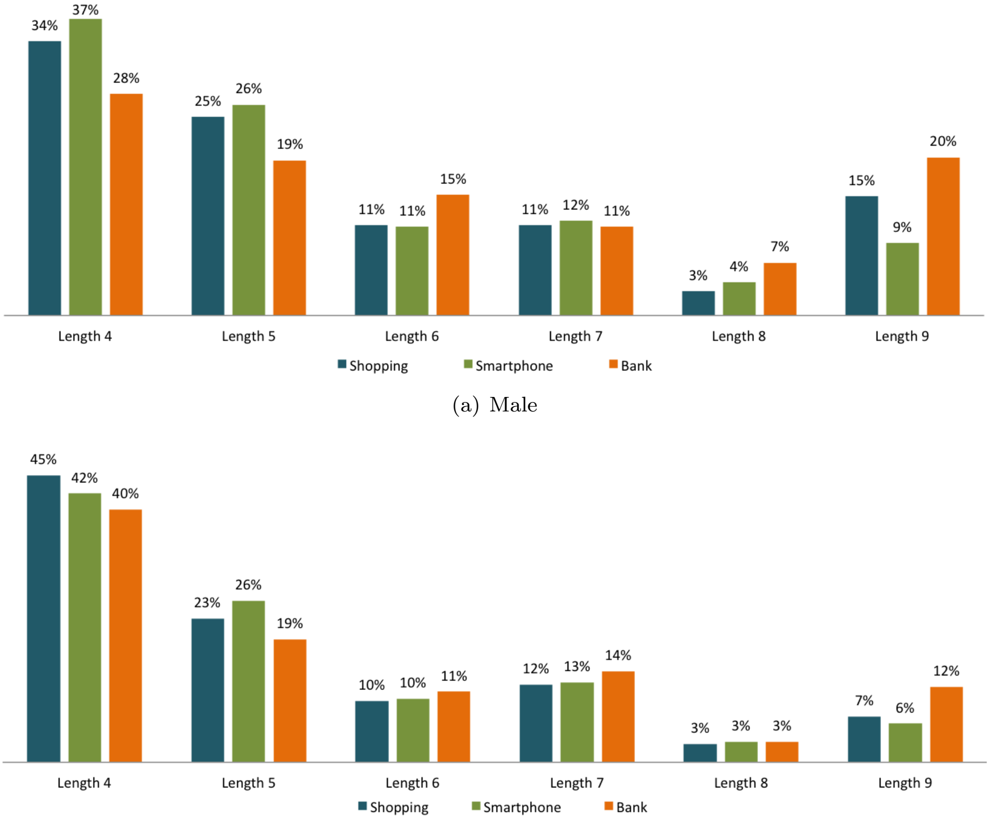
Table 5.7 summarizes the results when performing a two-tailed t-test for testing the difference in pattern length for patterns created by male and female respondents. The results reveal a significant difference in patterns created for shopping accounts and banking accounts.

Pattern type	Type	Mean	SD	P-value	Result
Shopping	Male Female	5.68 5.26	1.76 1.55	0.00048841	Significant
Smartphone	Male Female	5.47 5.27	1.60 1.49	0.08616198	Not significant
Bank	Male Female	6.09 5.57	1.86 1.72	0.00009069	Significant

Table 5.7: Statistical significance for pattern length and gender

Pattern length distribution

Figure 5.11 shows the pattern length distribution for both genders. The differences between the genders are noticeable in the endpoints, e.g. patterns with length 4 and 9. The male participants have a higher frequency of patterns with a longer length while the female participants have a higher frequency of patterns with a short length. The average number of patterns created with a length between 5 and 8 are about the same for both genders.



(a) Male

(b) Female

Figure 5.11: Average pattern length distribution for gender

Pattern complexity

Table 5.8 summarizes the patterns strength for the three pattern types for both genders. Patterns created by female participants have the characteristics of infrequent use of both intersections and overlaps for all types. The patterns created for smartphones by female respondents did not have a single pattern including an overlap, while only three patterns created by female respondents included an overlap.

None of the patters created by female respondents reached the maximum strength score. The highest strength score reached for patterns created by female respondents was 40.072. The the patterns with the lowest pattern strength are patterns created for smartphones by female participants, only reaching a total strength score of 32.078. The patterns created by male respondents had a higher frequency of both intersections and overlaps, hence a higher average strength score for patterns created by male respondents.

	Shopping		**Smartphone**		**Bank**	
Parameters	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
#Patterns	529	278	529	278	529	278
Avg. Size	5.684	5.263	5.465	5.270	6.089	5.572
Avg. Length	5.225	4.687	5.034	4.720	5.927	5.154
#Intersections	147	20	120	23	284	69
Avg. Intersections	0.278	0.072	0.227	0.082	0.537	0.248
#Overlaps	13	1	12	0	16	3
Avg. Overlaps	0.025	0.04	0.023	0	0.030	0.011
Avg. Strength	14.127	12.062	13.221	12.122	16.398	13.744
Min strength	6.340	6.340	6.340	6.340	6.340	6.340
Max strength	44.442	32.950	44.442	32.078	44.442	40.072

Table 5.8: Pattern strength and gender

Table 5.9 summarizes the results when performing a two-tailed t-test for testing the difference in pattern strength for patterns created by male and female respondents. The results reveal a significant difference in patterns created for all pattern types.

Pattern type	**Type**	**Mean**	**SD**	**P-value**	**Result**
Shopping	Male Female	14.13 12.06	8.02 6.48	0.00008596	**Significant**
Smartphone	Male Female	13.27 12.12	7.46 6.31	0.02153682	**Significant**
Bank	Male Female	16.40 13.74	9.11 7.74	0.00001588	**Significant**

Table 5.9: Statistical significance for visual complexity and gender

5.3.2 Age

The participants are being divided into divided into seven age intervals; under 20, 20-24, 25-29, 30-34, 35-39, 40-49, and over 50.

Average pattern creation time

Figure 5.12 is the graph visualizing the average creation time for each pattern type created by the different age groups. In general, patterns created for banking accounts have the highest average creation time across the various age groups. The average creation time seem to increase slightly as the age increases. It is noticeable that the creation time for smartphone patterns differs from the two other pattern types, especially for the younger participants where pattern creation time for smartphone are significantly lower.

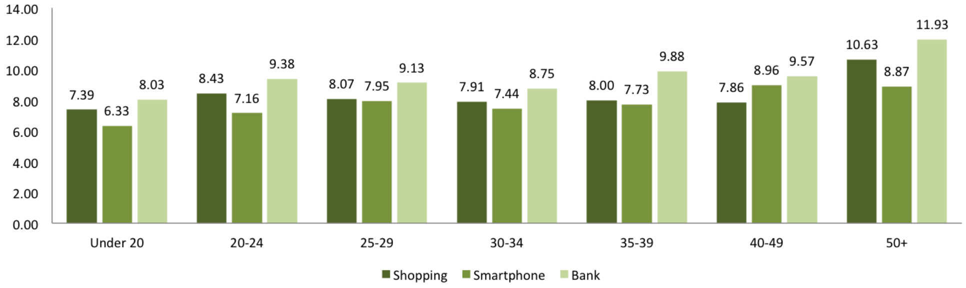


Figure 5.12: Pattern creation time by age

Average pattern length

Figure 5.13 shows the pattern length for the three pattern types by the different age groups. In the graph, the average pattern length goes slightly down as the age increases. As the age increases, the average length of the all pattern types evens out. The age groups having the highest difference in pattern length is the patterns created by the youngest and the oldest respondents.

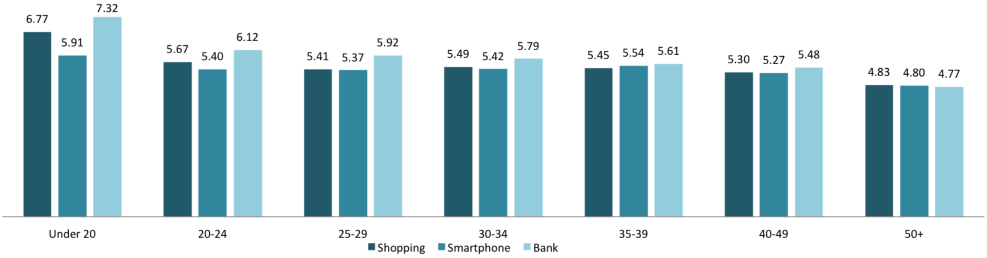


Figure 5.13: Pattern length by age

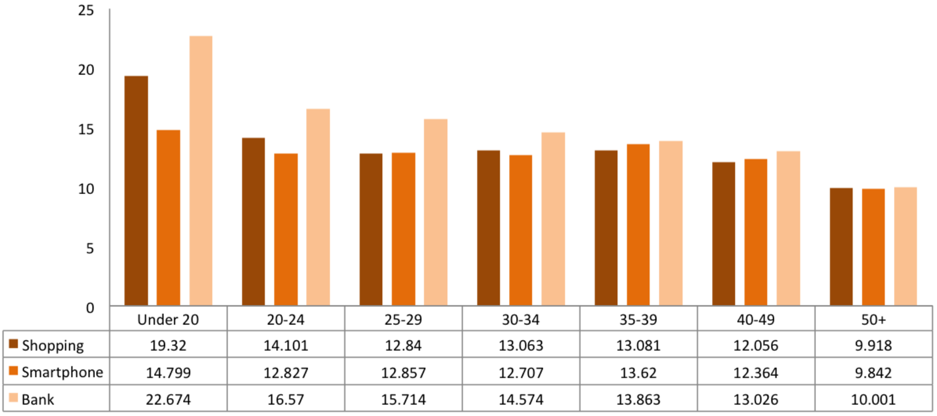
Table 5.10 summarizes the results when performing a two-tailed t-test for testing the difference in pattern length for patterns created by respondents with an age under 25 and respondents with a age of 25 years or older. The results reveal a significant difference in patterns created for shopping accounts and banking accounts.

Pattern type	Type	Mean	SD	P-value	Result
Shopping	Under 25 25+	5.74 5.38	1.76 1.64	0.00318325	**Significant**
Smartphone	Under 25 25+	5.43 5.36	1.56 1.56	0.496001184	**Not significant**
Bank	Under 25 25+	6.19 5.69	1.91 1.74	0.0001225	**Significant**

Table 5.10: Statistical significance for pattern length and age

Pattern complexity

The visual complexity of the patterns created by the different age groups are found in Table 5.14. To be able to illustrate the pattern strength for all age groups the data are shorten down to just visualize the pattern strength of each pattern type created by the different age groups. As mentioned earlier, the younger age respondents the longer patterns are created. The same is with the strength. The youngest age group have about twice as hight average pattern strength than the oldest age group. As the age increases, the strength of the different age groups also evens out. In general, bank have the highest average strength score while patterns created for smartphones have the lowest average strength score across the different age groups.



	Under 20	20-24	25-29	30-34	35-39	40-49	50+
■ Shopping	19.32	14.101	12.84	13.063	13.081	12.056	9.918
■ Smartphone	14.799	12.827	12.857	12.707	13.62	12.364	9.842
■ Bank	22.674	16.57	15.714	14.574	13.863	13.026	10.001

Figure 5.14: Pattern strength and age distribution

Table 5.11 summarizes the results when performing a two-tailed t-test for testing the difference in pattern strength for patterns created by respondents with an age under 25 and respondents with a age of 25 years or older. The results reveal a significant difference in patterns created for shopping accounts and banking accounts.

Pattern type	Type	Mean	SD	P-value	Result
Shopping	14.43 12.61	8.14 7.00		0.00089041	Significant
Smartphone	12.95 12.68	6.88 7.10		0.59266007	Not significant
Bank	16.95 14.32	9.36 8.04		0.00002946	Significant

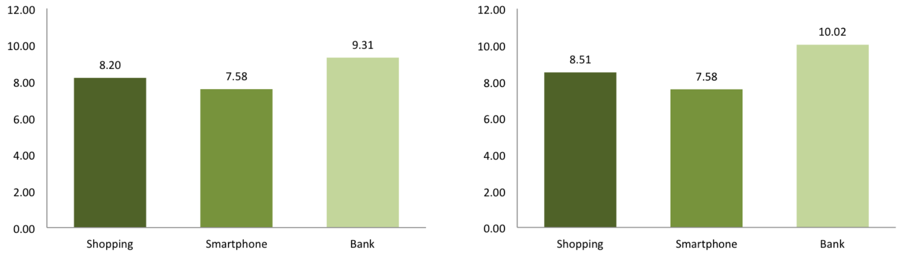
Table 5.11: Statistical significance for visual complexity and age

5.3.3 Handedness

This section includes the results from studying left- and right handed respondents and their choice of patterns.

Average pattern creation time

Figure 5.15 is the average creation time for patterns with respect to the handedness of the respondents. A participant can either be left or right handed. By looking at the graph, right-handed respondents has a lower average creation time than left-handed respondents. The only exception is the patterns created for smartphones where left-handed respondents have a slightly higher average creation time. For both left-handed and right-handed respondents, the average creation time for the banking account has the highest average creation among the three pattern types.



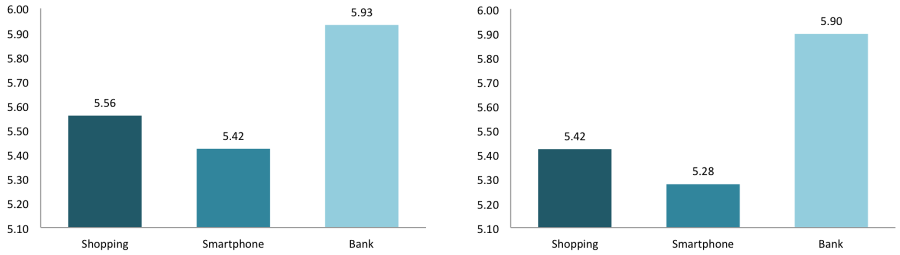
(a) Right-handed

(b) Left-handed

Figure 5.15: Average pattern creation time for handedness

Average pattern length

Figure 5.16(a) and 5.16 presents the average pattern length for patterns created by left- and right-handed, respectively. In general, both left- and right-handed respondents have about the same average pattern length. One exception is the patterns created for smartphones where there is a slight indication of left-handed people creating shorter patterns than right-handed people. The average pattern length for patterns created for smartphones created by right- and left-handed respondents are 5.42 and 5.28 respectively. The patterns that are getting the highest average length is the patterns created for banking accounts with an average pattern length of 5.93 and 5.90 for right- and left-handed respondents respectiveley.



(a) Right-handed (b) Left-handed

Figure 5.16: Average pattern length for handedness

Table 5.12 summarizes the results when performing a two-tailed t-test for testing the difference in pattern length for patterns created by left- and right-handed respondents. The results reveal no significant differences for any of the pattern types.

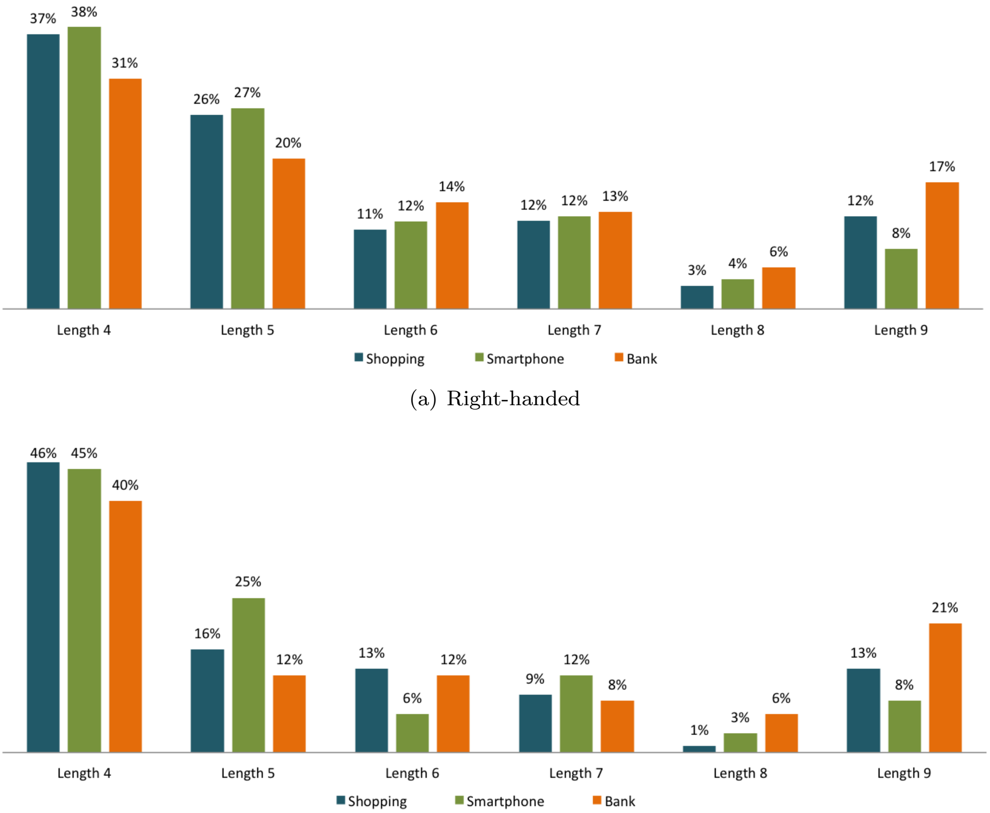
Pattern type	Type	Mean	SD	P-value	Result
Shopping	Left Right	5.42 5.56	1.75 1.70	0.47066079	**Not significant**
Smartphone	Left Right	5.28 5.42	1.60 1.56	0.40381003	**Not significant**
Bank	Left Right	5.90 6.93	1.99 1.81	0.86996995	**Not significant**

Table 5.12: Statistical significance for pattern length and handedness

Pattern length distribution

Figure 5.17 shows the pattern length distribution for both right- and left-handed respondents. The selection of pattern length seems to have no significant difference beside left-handed respondents having a tendency of having more patterns of length 4 than than right-handed respondents. It is also noticeable that the frequency of patterns tends to go down as the pattern length goes up. The only exception are patterns with length 8 where patterns with length 7 and 9 have a higher frequency.

A pattern length of 4 and 5 can be considered as a short pattern, whereas the minimum pattern length for Android Pattern Lock is 4. By looking at the number of patterns with the minimum length, 45% of the left-handed respondents created a pattern of length 4 for smartphones. In total, 70% of all left-handed respondent created a pattern of length 4 or 5 for smartphones while 65% of the patterns created for smartphones by Right-handed respondents is of length 4 or 5. Compared to patterns created for banking accounts, there are only 51% and 52% of the patterns having a length of 4 or 5 for right- and left-handed respondents, respectively. This resulr in a difference of 15% and 20% for right- and left-handed respondents between patterns created for smartphone and banking accounts of length 4 and 5.



(a) Right-handed

(b) Left-handed

Figure 5.17: Average pattern length distribution for handedness

Pattern complexity

Table 5.13 is the pattern strength for each pattern type with respect to the handedness of the respondents. The table does also include an overview of the parameters used for calculating the pattern strength.

The number of intersections differs on both shopping account and smartphone when comparing the patterns created by left- and right-handed respondents. The only pattern type having an equal number of intersections between right- and left-handed respondents are patterns created for banking accounts, resulting in 0.443 and 0.433 average number of intersections. Left-handed respondents have approximately an equal number of intersections for both shopping and bank, but only half the number half as many intersections for patterns created for smartphones. Patterns created by right-handed respondents has an approximately equal distribution of intersections on patterns created for shopping account and smartphones, but about twice as many occurrences of intersections on patterns created for banking accounts.

Comparing the number of overlaps, left-handed respondents have only one occurrence of overlap in one of the patterns created for banking accounts. Right-handed respondents have a somewhat higher number of overlaps, but overlaps still rarely occur. The pattern type reaching the highest number of overlaps and intersections are the patterns created for banking accounts.

By comparing average strength for both right- and left-handed respondents, the strength is roughly the same. In general, pattern strength has an indication to be higher for patterns created for banking accounts, but there are no significant differences caused by handedness. Looking at the maximum pattern strength reached, both right- and left-handed respondents had reached a maximum pattern strength of 44.441 that is the same as the maximum pattern strength in the entire dataset. For patterns created for smartphones, the maximum strengths for both right- and left-handed respondents were 43.187 and 37.280 respectiveley.

	Shopping		**Smartphone**		**Bank**	
Parameters	**Right**	**Left**	**Right**	**Left**	**Right**	**Left**
#Patterns	690	97	690	97	690	97
Avg. Size	5.560	5.423	5.423	5.280	5.932	5.897
Avg. Length	5.036	5.134	4.966	4.781	5.703	5.566
#Intersections	126	42	124	18	306	42
Avg. Intersections	0.183	0.433	0.180	0.186	0.443	0.433
#Overlaps	14	0	12	0	18	1
Avg. Overlaps	0.020	0.0	0.017	0.0	0.026	0.010
Avg. Strength	13.455	13.360	12.966	12.339	15.601	15.339
Min strength	6.340	6.340	6.340	6.340	6.340	6.340
Max strength	39.827	44.441	43.187	37.280	44.441	44.441

Table 5.13: Pattern strength and handedness

Table 5.14 summarizes the results when performing a two-tailed t-test for testing the difference in pattern strength for patterns created by left- and right-handed respondents. The results reveal no significant differences for any of the pattern types.

Pattern type	**Type**	**Mean**	**SD**	**P-value**	**Result**
Shopping	Left Right	13.36 13.46	8.64 7.46	0.9176142	**Not significant**
Smartphone	Left Right	12.34 12.97	7.29 7.03	0.4271685	**Not significant**
Bank	Left Right	15.34 15.60	9.50 8.67	0.7970912	**Not significant**

Table 5.14: Statistical significance for visual complexity and handedness

Typing habits

Handedness is a human characteristic influencing the physical interaction with a smartphone, hence might having an impact on the way a person create a pattern. The only way a person are interacting with a smartphone is by using their hands, making this an important behavior to study. The hand used are also being defined by handedness, making it interesting to look at the typing behavior of both.

Table 5.15 and 5.16 summarizes the typing habits of the respondents focusing on the physical interaction with the touch screen. The parameters included are handedness, hand used to hold the smartphone, and the finger used to type the pattern. Table 5.15 are summarizing the typing habits of right-handed respondents while Table 5.16 are summarizing the typing habits of left-handed respondents.

The two main options for interacting with a smartphone are described in Section 5.1.4. By starting looking at the right-handed respondents, the majority (53%) are interacting with the screen as described in category 1. This means that the 53% of the right-handed respondents used their right hand and their thumb, e.g. used only one hand, for interacting with the screen. About 32% of the right-handed respondents interacted with the screen as described in category 2. In other words, 32% of the right-handed respondents used their left hand to hold the smartphone while using their forefinger on their right hand for interacting with the screen. The last 15% of the right-handed respondents had other typing habits than the two most common ways of interacting with the screen.

The left-handed respondents had a more indefinite typing habit because none of the alternatives appear to be more common than others. By summarizing the numbers, it seems that the frequency of category 1 and 2 for left-handed respondents have no significant difference in number of respondents. 26 respondents are in category 1 and 26 respondents are in category 2. Another observation is that 22 of the left-handed respondents interact in the same way as Category 1 for right-handed respondents. Only 5% of the right-handed respondents acts in the same way as Category 1 for left-handed respondents. In other words, the typing habits of right-handed respondents seems to be more predictable than as for the typing habits of left-handed respondents.

Hand used	**Finger used**	#
Right hand	Thumb	366
	Forefinger	41
	Other	8
Left hand	Thumb	33
	Forefinger	217
	Other	23

Table 5.15: **Right-handed** typing habits

Hand used	**Finger used**	#
Right hand	Thumb	22
	Forefinger	26
	Other	6
Left hand	Thumb	26
	Forefinger	10
	Other	4

Table 5.16: **Left-handed** typing habits

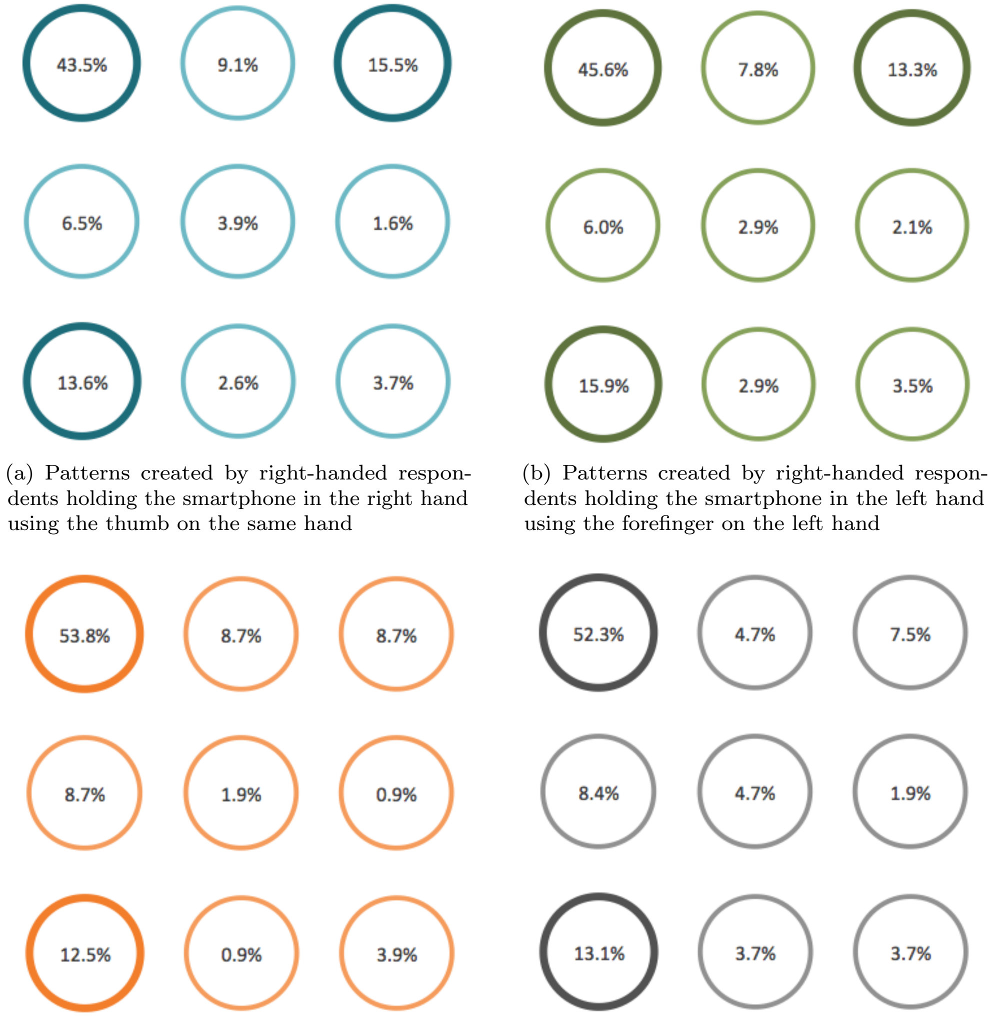
Bias in the Selection of Start Node

How a person are holding a smartphone might impact the nodes being reachable. It is defined two main types of typing a pattern, e.g. either use one or both hands. When using both hands, one hand are used to hold the phone while the other hand are used for interacting with the screen. When holding the smartphone in one hand, the same hand are used for both holding and interacting with the screen, whereas the thumb are the only finger available when using one hand. Figure 5.18 are showing likelihood of starting nodes from patterns created by respondents with different dominant hand, combined with the two ways of holding a smartphone and finger used. Figure 5.18(a) and 5.18(b) are looking at right-handed respondents' selection in starting node for patterns created by using either one or two hands, respectively.

Figure 5.18(a) are showing the choice of starting nodes for patterns created by right-handed respondents holding the smartphone in the right hand using the thumb on the same hand for pattern creation. As seen in Figure 5.7, node 1, 3, 7 are the most common choice for selecting starting node. Figure 5.18(b) are the selection of starting node by right-handed respondents using their left hand holding the smartphone using their right forefinger for pattern creation. The three main starting nodes are 1, 3, and 7, similarly for the right-handed respondents using one hand.

Figure 5.18(c) are showing the selection of starting nodes for patterns created by left-handed respondents holding the smartphone in the left hand using the thumb on the same hand for pattern creation. About 54% of the left-handed respondents using one hand starting their pattern in node 1. The majority of the left-handed respondents using one hand starts their pattern on the left side of the grid. The nodes in Figure 5.18(d) are starting points selected by left-handed respondents using the right hand for holding the smartphone while interacting with the screen using the forefinger on the left hand.

By comparing handedness and way of creating the patterns, the way of holding the smartphone, either using one or two hands, do not seem to affect the node starting node. For all ways of creating a pattern with respect to handedness, the majority of the patterns are created by starting on the left side of the grid.



(a) Patterns created by right-handed respondents holding the smartphone in the right hand using the thumb on the same hand

(b) Patterns created by right-handed respondents holding the smartphone in the left hand using the forefinger on the left hand

(c) Patterns created by left-handed respondents holding the smartphone in the left hand using the thumb on the same hand

(d) Patterns created by left-handed respondents holding the smartphone in the right hand using the forefinger on the left hand

Figure 5.18: Starting node based on handedness, hand used to hold smartphone, and finger used used when creating patterns

5.3.4 Experience with IT and Security

This section takes a more detailed review of the patterns created by people with a difference in experience with IT and security. The level of interest, and experience with IT and security, can cause people to create different passwords due to risk perception and security awareness.

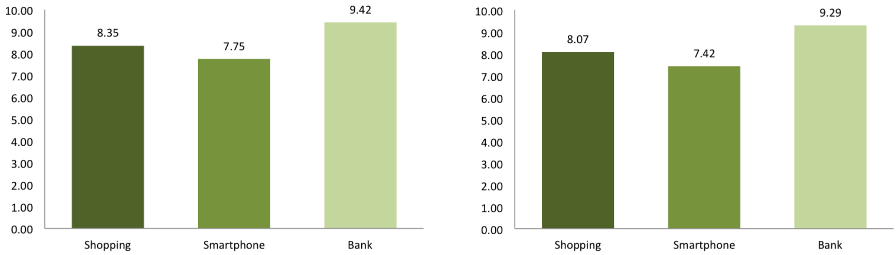
Average pattern creation time

Figure 5.19 is showing the average creation time measured in seconds for both respondents experienced and not experienced with IT and security.

Figure 5.19(a) are showing the average creation time for the three pattern types created by respondents experienced with IT and security. The patterns with the highest creation time are the patterns created for banking accounts with an average creation time of 9.42 seconds. The patterns with the lowest average creation time are being created for smartphones with an average creation time of 7.75 seconds.

Figure 5.19(b) are showing the average creation time for the three pattern types created by respondents inexperienced with IT and security. The patterns with the highest creation time are the patterns created for banking accounts with an average creation time of 9.29 seconds. The patterns with the lowest average creation time are being created for smartphones with an average creation time of 7.42 seconds.

Comparing the respondents experienced and inexperienced with IT and security, the inexperienced respondents has a slightly slower response time for pattern creation for all three pattern types.



(a) Experience with IT and security

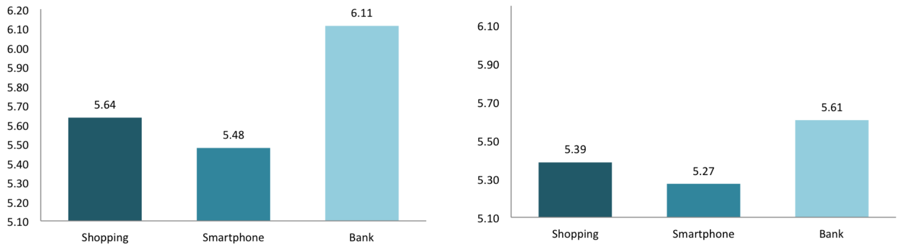
(b) No experience with IT and security

Figure 5.19: Average pattern creation time for experience with IT and security

Average pattern length

Figure 5.20(a) are showing the average pattern length for patterns created by respondents experienced with IT and security. The patterns with the highest average pattern length are being created for banking accounts with a total average pattern length of 6.11. The pattern type with the shortest average pattern length are patterns created for smartphones with an average pattern length of 5.48. Figure 5.20(b) are showing the average pattern length for patterns created by respondents inexperienced with IT and security. The patterns with the highest average pattern length are being created for banking accounts with a total average pattern length of 5.61. The pattern type with the shortest average pattern length are patterns created for smartphones with an average pattern length of 5.27.

The trends in the graphs show that respondents with experience in IT and security create longer patterns than people inexperienced with IT and security for all three pattern types. For both graphs, patterns created for banking accounts has the highest average length while patterns created for smartphones have the shortest average pattern length.



(a) Experience with IT and security (b) No experience with IT and security

Figure 5.20: Average pattern length for experience with IT and security

Table 5.17 summarizes the results when performing a two-tailed t-test for testing the difference in pattern length for patterns created by respondents being experienced and inexperienced with IT and security. The results reveal a significant difference in patterns created for banking accounts.

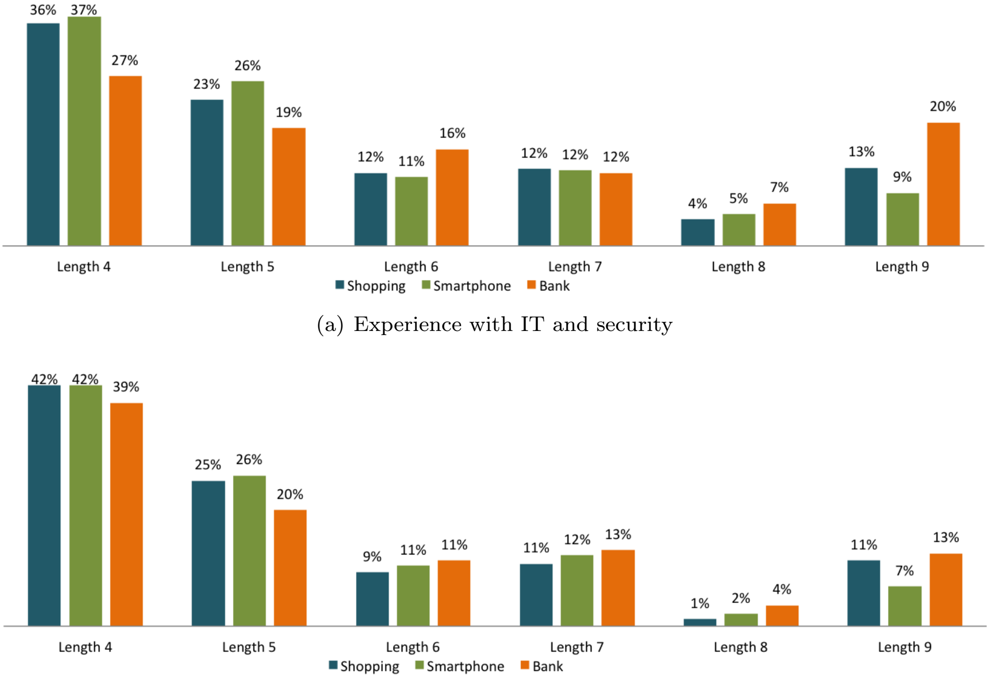
Pattern type	Type	Mean	SD	P-value	Result
Shopping	Experienced Inexperienced	5.64 5.39	1.72 1.66	0.038204664	**Not significant**
Smartphone	experienced Inexperienced	5.48 5.27	1.60 1.50	0.064560991	**Not significant**
Bank	Experienced Inexperienced	6.11 5.61	1.86 1.74	0.000083188	**Significant**

Table 5.17: Statistical significance for pattern length and experience with IT and security

Pattern length distribution

Figure 5.21(a) is the distribution of pattern length for patterns created by respondents experienced with IT and security. For the patterns created for smartphones, 70% of the patterns constitutes of patterns with a lower length, e.g. patterns with a length between 4 and 6. For patterns created for banking accounts, 62% of the patterns are created having a low length. The created patterns having a length of 9 nodes are the patterns created for banking accounts. Figure 5.21(b) is the pattern length distribution of patterns created by respondents inexperienced with IT and security. For the patterns created for smartphones, 79% of the patterns are patterns having a low length.

By comparing the pattern length distribution for experienced and inexperienced respondents, the patterns created by inexperienced respondents have a higher frequency of patterns with lower pattern length than patterns created by experienced respondents. The difference observed is that experienced respondents create fewer patterns of low length for banking accounts compared to inexperienced respondents. Instead of having a high frequency of short patterns, patterns created by experienced respondents have a higher frequency of patterns with length 9 for banking accounts. Experienced users do only have 27% of their patterns for banking accounts created with the length 4, while inexperienced users have 39% of their patterns for banking accounts created with the length 4. Both graphs show that the lower the length a lower frequency of the pattern length occurs.



(a) Experience with IT and security

(b) No experience with IT and security

Figure 5.21: Average pattern length distribution for experience with IT and security

Pattern complexity

Table 5.18 is a summary of the parameters used for calculating the pattern strength for each pattern type with respect to the respondents experience with IT and security.

Patterns created by experienced users seems to have a higher frequency of both overlaps and intersections, hence reaching a higher average strength score for all patterns created by experienced respondents. The patterns created for smartphones, by both experienced and inexperienced respondents, was the patterns getting the lowest strength score.

By comparing average strength for both experienced and inexperienced respondents, the patterns created by experienced respondents have obtained the highest average strength. Looking at the maximum pattern strength, neither the experienced or the inexperienced respondents manage to create a pattern reaching the maximum pattern score for the dataset.

	Shopping		**Smartphone**		**Bank**	
Parameters	**Yes**	**No**	**Yes**	**No**	**Yes**	**No**
#Patterns	470	332	470	332	470	332
Avg. Size	5.636	5.386	5.479	5.274	6.113	5.605
Avg. Length	5.182	4.818	5.045	4.753	5.919	5.281
#Intersections	123	41	108	34	231	121
Avg. Intersections	0.262	0.123	0.230	0.102	0.491	0.364
#Overlaps	10	4	9	3	13	6
Avg. Overlaps	0.021	0.012	0.019	0.009	0.028	0.018
Avg. Strength	13.911	12.635	13.271	12.202	16.417	14.089
Min strength	6.340	6.340	6.340	6.340	6.340	6.340
Max strength	44.441	39.827	43.187	38.870	44.441	44.441

Table 5.18: Password strength and Experience with IT/Security

Table 5.19 summarizes the results when performing a two-tailed t-test for testing the difference in pattern strength for patterns created by respondents being experienced and inexperienced with IT and security. The results reveal a significant difference in patterns created for all three pattern types.

Pattern type	**Type**	**Mean**	**SD**	**P-value**	**Result**
Shopping	Experienced Inexperienced	13.91 12.64	7.78 7.19	0.01699014	**Significant**
Smartphone	Experienced Inexperienced	13.32 12.20	7.40 6.52	0.02375298	**Significant**
Bank	Experienced Inexperienced	16.42 14.09	8.94 8.26	0.00015740	**Significant**

Table 5.19: Statistical significance for visual complexity and experience with IT and security

5.3.5 Hand Size and Screen Size

To be able to analyze if there is a correlation between hand size and choice of graphical patterns, the data collected needs to be validated due to the subjective form. Along with the hand size, the screen size and finger can also be an interesting factor to look at. When using a survey, there is no good way of asking respondents for their classification of their hand size, hence a need for validation to decide whether the data are reliable for further use.

Figure 5.22 are visualizing the frequency of each hand size for all respondents. Table 5.20 are summarizing the respondents classification of the size of their hands with respect to the gender of the respondents. Based on own experiences, there is a difference in how the two genders will classify the size of their hand. In society, men are supposed to be masculine where big hands are often being associated with masculinity. The opposite of masculine, femininity, can be associated with smaller hands. It is therefore, believed that the majority of the male respondents will classify their hand as medium or large while the majority of the female respondents will classify their hand as small or medium. The question asked in the survey were asking the participants to classify the size of their hands based on their gender. Table 5.20 confirms that the majority of the male respondents classified their hand as a medium size or higher, while female respondents classified the size of their hand as medium or lower.

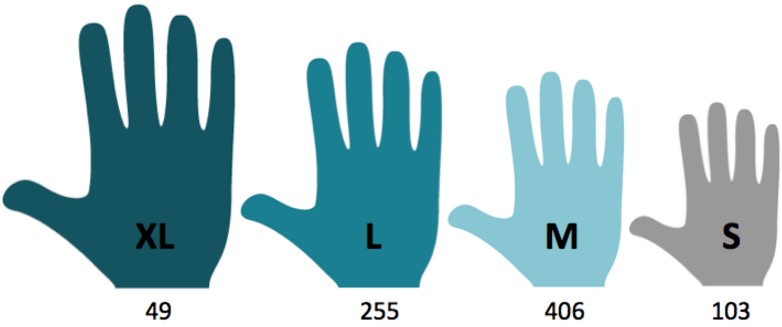


Figure 5.22: Handsize of all participants

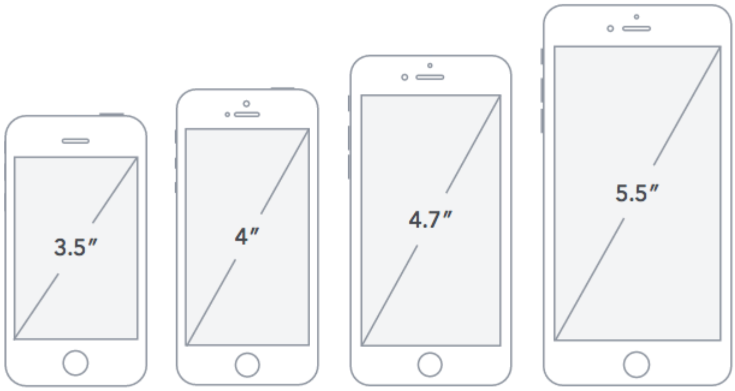
	Xtra Large	Large	Medium	Small	Total
Male	45 (9%)	200 (38%)	246 (47%)	38 (7%)	529
Female	3 (1%)	54 (19%)	157 (56%)	64 (23%)	278

Table 5.20: Handsize and gender

The problem with using data being a subjective classification by the respondents, it is not known if the classification is correct. On the other side, what is a correct classification? The hand size of the respondents are not used for further analysis because of the difficulties of comparing the size selected by both genders against each other. If used, any results can not be guaranteed to be correct if including the hand size. The hand size property will are not being used further in this research.

In the survey, both the subjective classification of the screen size as well as the height and width in pixels were collected. Classification of screen sizes is troublesome due to the various screen sizes and resolutions across different devices. Smartphones operating on Android are known to have a high variation in screen resolutions while Apple have created their devised with a smaller and fixed number of different screens. The physical size of a mobile screen is not correlated with the pixel size because a pixel is not a physical measure due to resolution, hence being troublesome to obtain the correct screen size.

Besides screen size and pixels, the respondents were asked for the mobile operating system on the device used. As a result of *Apple* having standardized the screen of their devices makes it possible to obtain the physical size of the screen for respondents using a iPhone for answering the survey. Figure 5.23 is the complete list of the devices provided by Apple and the corresponding size in inches.



(a) iPhone 3.5" (b) iPhone 4" (c) iPhone 4.7" (d) iPhone 5.5"

Figure 5.23: iPhone screen resolutions

There are only four sized used by Apple mobile devices as illustrated in Figure 5.23. In the dataset, there are only obtained the pixel sizes of the devices. For the devices running on iOS, four different types were observed in the dataset. The respondents' classification of the screen sizes, as well as the pixel size, are summarized in table 5.24 for all respondents using an iPhone.

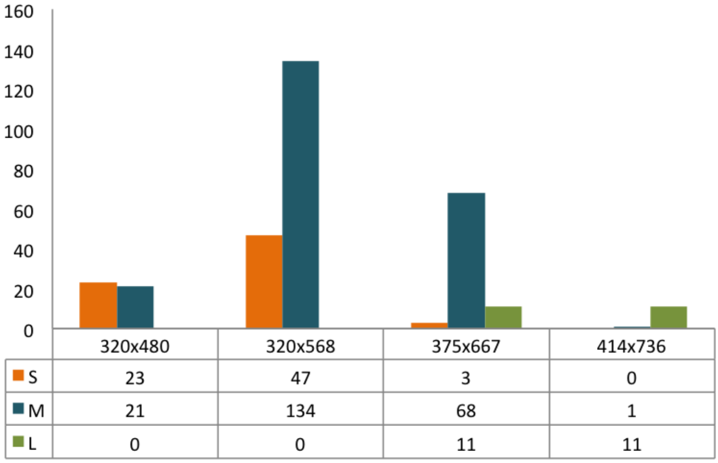


Figure 5.24: iPhone screensize distribution

To be able to evaluate the provided classification from the respondents, the Apple devices are classified according to what this research defines as a small, medium and big screen. The classifications are listed in Table 5.21 with the corresponding device and pixel sizes. To be able to evaluate the provided classification from the respondents, the Apple devices are classified according to what this research defines as a small, medium and big screen. The classifications are listed in Table 5.21 with the corresponding device and pixel sizes. When comparing Figure and 5.24 and Table 5.21, the respondents agree on the size to some extent. When looking at the number of respondents, only 40% of the respondents used an iPhone, making the provided data for each screen type hard to use for further analysis.

Iphone Model	Resolution (px)	Inches	Size
iPhone 2G, 3G, 3GS, 4, 4s	320 × 480	3.5”	S
iPhone 5, 5s	320 × 568	4”	M
iPhone 6	320 × 568, 375 × 667	4.7”	M/L
iPhone 6 Plus	375 × 667, 414 × 736	5.5”	L

Table 5.21: iPhone screensizes

Based on the classification of the screen size on Apple devices, it can be used for evaluating respondent’s ability to classify what this research defines as a small, medium or big screen. The problem with the rest of the respondents not using Apple products is that there is no fixed standard for screen sizes, making it impossible to predict the physical size by knowing the pixels. Since the majority of the respondents were using an Android device (58%), making it unlikely of being able to classify the screen sizes correctly. The 40% of the devices being an iPhone was possible with some uncertainty to classify, but the size of the dataset are too small for getting any results. Either being an Android or iOS device, it can neither be guaranteed to be correctly classified. It is therefore decided to not include any results using the screen size.

5.4 Preprocessing and Validation of Collected Data

When collecting data, it is important to preprocess the data before analyzing the data. One of the reasons for preprocessing the data is, for example, to find an outlier in the data. An outlier can impact the data to not show the correct results. For example, when going through the pattern creating time, some of the respondents had used over 10 hours for creating a pattern. The high response time causes noise in the data and is declared as outliers, meaning that the response will be removed from the dataset. Figure 5.25 is a plot of the pattern creation time, using time intervals of 1 seconds and number of occurrences. When time goes up to 30 seconds, the frequency goes down. In a normal situation, a pattern does not take a long time to type. It was therefore decided only to use the patterns being created within 30 seconds, as a pattern time higher do not represent a normal situation. Besides the time used, there was no need for preprocessing the other data and were used as is.

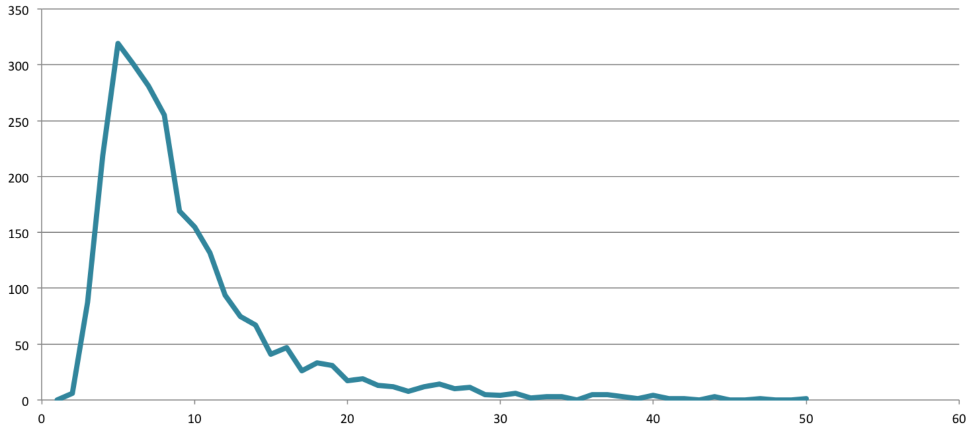


Figure 5.25: Defining max pattern creation time

It was mentioned that is was used different pattern orders for different respondents. The different orderings were applied to ensure that an ordering was not impacting the choice in patterns. When comparing the time used, the length of the patterns, and visual complexity, there were no significant difference between the patterns created when creating patterns in a different order. Figure 5.26 visualises the number of respondents having the different pattern orderings. The specific orders are specified using the numbers 1 to 3 corresponding to patterns created for shopping accounts, smartphones, and banking accounts, respectively.

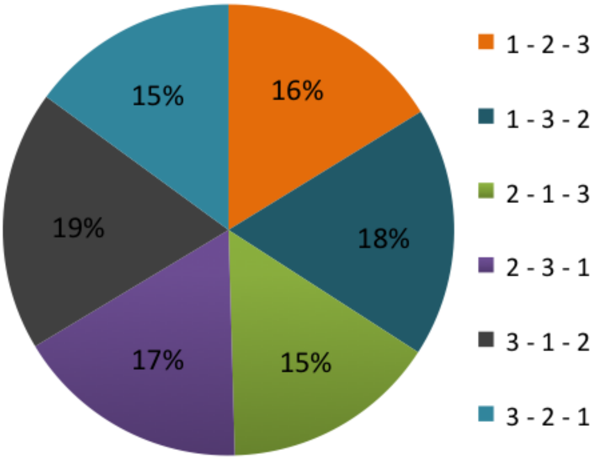


Figure 5.26: Percentage of times the pattern orders occurred

6 Discussion

This chapter is a discussion of the results presented in Chapter 5. The first six sections corresponds to the first six research questions from Chapter 1 focusing on choice in passwords and human properties. Section 6.7 are discussing the results when looking at the entire population. Section 6.8 discusses the results from a context of use perspecitive. The last section is a disussion of the limitations of this research.

6.1 Age and Choice of Graphical Passwords

When testing the significance of the length and complexity for different age groups, the test distinguished two groups; under and over 25 years. The tests revealed that there is a significant difference in pattern length and visual complexity for patterns created for shopping accounts and banking accounts. The respondents under 25 created longer patterns as well as patterns with a higher visual complexity for banking accounts and shopping accounts. There was no significant difference in length and the complexity of patterns created for smartphones.

Before starting to analyze the results, it was not expected that the respondents under 25 years would create longer and more complex patterns than the respondents over 25. It was rather being expected that the respondents over 25 years created longer and more complex patterns as they might, for example, use the mobile device for work purposes where a strict security policy is often required. A factor that may cause the difference in complexity and pattern length is that the respondents under 25 have owned a mobile device when growing up. Knowing how to use a smartphone might make younger respondents more prepared for creating stronger passwords as they might be more familiar with the utilization of a smartphone. As far as this research is familiar with, no other research has been found regarding the selection of password particularly looking at age.

6.2 Gender and Choice of Graphical Passwords

By comparing the patterns created by male and female participants, there was a significant difference in patterns created. The results revealed a significant difference in patterns created for shopping account and banking accounts when testing the pattern length. When testing the visual complexity of the patterns, there was a significant difference in the created patterns for all pattern types. The results prove as evidence that male participants create longer patterns with higher visual complexity compared to the patterns created by female respondents.

By studying the choice of length, female participants had a higher frequency of patterns of length four and at the same time had a lower frequency of patterns of length nine as opposed to male participants. When looking at patterns with length five up to eight, both genders have the same frequency. There is therefore a difference in how frequently male and female participants select the minimum and the maximum pattern length. One factor that may be involved is that many of the male participants are experienced with IT and security while fewer of the female respondents have the same experience.

By studying the visual complexity of the patterns created by the two genders, it is observed that none of the female participants managed to create a pattern with the maximum score in the dataset. The cause of female participants choosing less secure patterns might be biased by the number of male participants with a background in IT and Security or that there are more male participants in the dataset. As far as this research is familiar with, no other research has been found focusing on the selection of passwords based on gender. One research group was investigating the security of the PassFace scheme where it was observed that users tended to choose faces that they liked or could compare themselves to [11]. By knowing the gender, they managed to perform a dictionary attack to guess user selected passwords. If the gender of the user was known as male, then 10% of the passwords could easily be guessed on the first or second attempt.

6.3 Handedness and Choice of Graphical Passwords

Handedness is a biological characteristic that relates to how people interact with physical objects. In general, the majority of the population have the property of being right-handed while only 12% of the population are left-handed. The results do not show any significant results indicating any differences in length and visual complexity for patterns created by left- and right-handed respondents. Thus, there is no statistical evidence for graphical passwords being influenced by the handedness of the creator.

When selecting handedness as a human property in this study, it was believed that handedness and the way that respondents interacted with the touch screen would impact the selection of starting node. Based on the result, two main ways of interacting with a touch screen for unlocking a smartphone were observed. The first, and most common way, is to use one hand for holding and interacting with the phone. The other way is using one hand for holding the phone while using the forefinger on the other hand for interacting with the screen. 85% and 53% of the right- and left-handed respondents, respectively, used the two described methods for interaction. The results confirm that left-handed respondents do not have a main way of interacting with a smartphone as for right-handed respondents. An explanation for the difference in the physical behavior could be a historical impact where left-handed children were forced to write with their right hand.

44% of the respondents started their patterns in the upper left corner, wheras 73% of the patterns either started in the the upper left corner, upper right corner or the bottom left corner. Because of the biased selection of starting node, and the different ways of interacting with the touch screen, it was believed that there would be a difference in the selection of starting node by looking at handedness. Since the majority of respondents was right-handed, it was assumed that this was the factor causing the high frequency of patterns starting in the upper left corner. Therefore, a higher frequency of patterns starting in the upper right corner as a cause of handedness and physical interaction was expected when regarding left-handed respondents. However, the results revealed the opposite, where the majority of the patterns started on the left side of the grid. The majority of the left-handed respondents would rather start on the left side than the right side as originally assumed.

An explanation for the unexpected behavior is that right-handed and left-handed respondents prefer to start the patterns at the same nodes. The cause of the same behavior need to be looked further into, but studies exploring how memory retrieval and storage works can be a good start. Researchers have found that people prefer to scan information and store information in the short memory in a better way when scanning according to preferred reading and writing direction [10]. As a result of the way we scan information, it may cause users to start creating the patterns in the same way we read and write to be able to remember the patterns we create. The results do not imply that handedness have any impact on the way we create out patterns, rather a stronger indication that reading and writing orientation have a stronger impact on the way we create our patterns. In this study, less than 2% of the population had a different reading and writing orientation than the majority, e.g. left-to-right. Thus, it is not possible to use the collected data to make any conclusions about the impact on the choice of patterns based on the reading and writing orientation.

6.4 Experience with IT and Security and Choice of Graphical Passwords

The result in this research shows that there is a significant difference in the patterns created by experienced and inexperienced respondents. There is a significant difference in the length of patterns created for banking accounts, as well as a significant difference in the visual complexity for all pattern types.

If a person acknowledges that they have experience with IT and security, it is likely that they know the risks of selecting short passwords with low complexity that are easily guessed. The results indicated that the respondents with experience with IT and security are better with creating patterns, but there is still a low average complexity score and a high frequency of patterns created with the minimum length. Not a single respondent created a pattern with the highest complexity score for smartphones; the patterns created for smartphones were the patterns with the lowest average complexity score and length, regardless of the IT experience of the creator.

The results shows that even experienced respondents do not create strong patterns. This study does not know why experienced respondents create short passwords when they probably know the consequence. One explanation is that users in general are not able to create and remember long and complex patterns.

6.5 Reading and Writing Orientation and Choice in Graphical Passwords

As stated, the number of participants with a different reading and writing orientation than from left-to-right were too low, and thus cannot be used for making any conclusions. When looking at choice in graphical passwords and handedness, it was observed that there was no significant difference in the patterns created by left- and right-handed respondents. This observation gave a stronger indication of reading and writing orientation having an impact on the choice of graphical password. The reading and writing direction is proposed for future reasearch.

6.6 Hand Size and Choice in Graphical Passwords

The data collected for the hand size were not used as a cause of being difficult to correctly verify the selected hand size. From the start, the property was known to have a subjective form, meaning that there was a risk of this happening. It is not known how the collection of hand size could be done differently, as this study selected the best-known approach. If wanting to look further into the choice of patterns and the size of the hand, it is probably a better idea to set up an experiment for being able to correctly measure the size.

6.7 Similarities in the Choice of Graphical Passwords in the Entire Population

This section looks at the results found when looking at the entire population.

6.7.1 Pattern Creation Time and Length

The average pattern length for the entire population varied according to the pattern type. Patterns created for banking accounts had the highest average length of 5.92 while the patterns created for smartphones had the lowest average length of 5.40. Research have reported that the average length of patterns created for smartphones is 5.63 [39], which is close to the average length for smartphones observed in this study.

There are a different number of combinations of patterns with different length, but patterns of length eight seem to occur less often. For all pattern types, the probability of a pattern of length 8 to be selected is only 4-5%. The cause of patterns with length 8 occurring less often then than patterns having a length of 7 or 9 nodes is not found. Patterns having a long length do not occur at the same frequency as the short patterns, but the patterns of length 8 are almost absent in some cases. There are 140.000 patterns of length eight, making the probability of selecting a password of length 8 high given a uniform selection. From a security perspective, this can reduce the number of combinations with roughly 140.000 combinations because of the low probability of users creating a pattern of length eight. Any reduction in the number of likely combinations is a violation of the security of the password scheme.

Pattern creation time can tell a lot about the validity of the dataset. Respondents experienced with the Android Unlock Pattern had different reaction time when creating the patterns for a smartphone. The time used for creating patterns for a shopping account and bank account was about the same. The difference in creation time between respondents experienced and inexperienced with ALP was 1.2 seconds. The different reaction time can be a result of many respondents having shared their actual lock pattern, or an another pattern known to the respondent. This result indicates that the dataset possessed includes patterns representative for patterns used in "the wild".

As the result shows, people creates on average short patterns, especially short patterns are observed created for smartphones. From a security perspective, a smartphone could cause loss of huge amounts of sensitive information compared to a shopping account. As reported by researchers, getting access to a smartphone automatically logged into an email account can give access to sensitive information like SSN, Bank Account Number, Email Password and Home Address [16] Selecting short patterns can be seen as a form of bad risk assessment because it is a risk the possibility of losing sensitive information. The cause low priority of a secure password mechanism on mobile devices may be influenced by the trade-off between security and usability. Many users do not want to spend more time than needed for typing a password. Because of the rapid use of our mobile device, it makes users spend a lot of the when interacting with the phone to unlocking the screen. Research have reported that an average user will use about 2.9% of the spend on interacting with their mobile phone to gaining access by unlocking the screen [20].

The average pattern length, regardless of pattern type, seems to be low. One factor might be that the respondents had to retype the pattern they selected, indicating that people are not capable of remembering longer and more complex patterns. To be able to remember a password, regular use are required for permanently store the password in long-term memory. A pattern created for one-time use are stored in short-term memory, making it hard to recall a complex pattern just created due to how our short-term memory works [12].

6.7.2 Visual Complexity

On average, the patterns collected for all pattern types were given a low complexity score. In the entire population, none of the participants managed to produce a pattern receiving the maximum complexity score of 46.8. The patterns receiving the highest average complexity score was patterns created for banking applications while patterns created for smartphones received the lowest complexity score.

Graphical passwords, especially the Android Pattern Lock, need to have their security evaluated on an extra dimension because of their graphical characteristics. Pattern locks can, for example, be easily captured by someone, accidentally or intentionally, by looking over someone's shoulder. Typing a long pattern will not help when the visual complexity is low. The length will help in terms of possible combinations, but will not necessarily avoid the possibility of someone capturing the pattern if the visual complexity is low. For avoiding such capturing attacks, there is a variety of new schemes purposed for avoiding such capturing attacks [49, 3].

Overlaps and intersections are some of the parameters used when evaluating the visual complexity of a pattern. Only 46 overlaps are registered in the dataset. The total number of patterns including overlaps may be lower because one pattern can have more than one overlap. One explanation for users not utilizing overlaps in their patterns is because it might not be very intuitive as a result of how Android explains how the scheme works. One rule commonly known is the rule of a node only being able to be selected one time. This rule does not make an overlap very intuitive to select because the pattern needs to go through an already selected node. The supported feature for creating overlaps should have been communicated to the users in a better way. The use of overlaps can assist users to create more visually complex and secure patterns. The average number of intersections are also low compared to the number of possible intersections in a pattern [34]. When looking at all patterns collected, the patterns in the dataset had an average of only 0.27 intersections.

The observation of the low frequency of intersections and overlaps can be an indication of people finding it hard to remember visually complex patterns. This result can be used to reduce the likely password space because there is a low probability that people would create a pattern containing many intersections or any overlaps at all. This observation can be seen as a violation of the security of the password scheme because the number of possible combinations can be reduced.

6.7.3 Association Elements

A unique aspect of graphical patterns is that they are visual and not just text or numbers with a semantic representation as in PIN codes and alphanumeric passwords. When creating a password, it is common to create a password that is associated with something you know to be able to remember and recall the password. In the dataset, 11.4% of the collected patterns corresponded to a letter. The behavior of selecting Android Pattern Locks corresponding to association elements were also recognized in another user study [34]. When looking at user-selected passwords, studies show that many users make use of graphical shapes, or objects, to support the process of remembering [45].

In this project, only letters were considered as an association element. Other studies have found that users also uses numbers to form a pattern, indicating that there might be other association elements than letters used as association elements. An another research reordered the nodes in the Android Pattern scheme, whereas one of the rearrangements ended up having the same shape as a *Delta* [39]. The participants recognized the element, making the majority of the participants creating a pattern corresponding to a *Delta*.

Whether some of the patterns are used as an association element, or just looking like one as a coincidence, is not possible to know. To be able to answer such question, a qualitative study are recommended to be able to ask people about their password selection strategy and whether the use of association elements are a used.

6.7.4 Selection of Starting Node and Movement Patterns

One restriction of the Android Pattern Lock is that each node can only be selected once, making the selection of starting node crucial in terms of being able to guess a pattern. The result of this study shows an indication of bias towards the choice of starting node, whereas 44% of the respondents started creating their patterns in the upper left corner. The second and third most common starting point was the upper-right corner and the bottom-left corner, which was selected as a starting node 15% and 14% of the time, respectively. The top three starting nodes thus constitutes 73% of all patterns in the dataset. Given a uniform distribution, the probability of starting in any node should be 11%, and the total probability of starting in the top three starting nodes would then be 33%. The results show that users do not select their patterns uniformly.

The predictable behavior when selecting the starting node corresponds to the results from another user study having the same amount of patterns starting in the upper-left corner [39]. The study also reported having 73% of all patterns starting in either the upper-left corner, upper-right corner or bottom-left corner, matching the observations in this study.

The results do not only reveal that patterns are likely to start in the upper-left noder, patterns also include predictable movement sequences. Most of the patterns are straight lines close to the edges, whereas the subsequence `123` and `147` often occurs. It does not only often occur as a subsequence, but it is also one of the most commonly used ways of starting a pattern. This study reveals that 42% of the created patterns are being found in the top 100 list of most frequently created patterns. In other words, by selecting a pattern from the top 100 list, there is a 42% chance of a match. In the same list, the subsequences `123` and `147` are subsequence appearing most commonly. There is not found a specific explanation for this behaviour, but it occurs that the respondents preferred creating patterns close to the edges. Since there is a low frequency of patterns having intersections and overlaps, an explanation can probably be found by studying Gestalt and cognitive psychology. The study of psychology is outside the scope of this research, but the results strongly indicate that user-selected password are being biased as a cause of the visual appearance of the graphical scheme. The visual appearance is might causing users to select lines close to the edge as well as starting patterns in the corners.

6.8 Choice in Graphical Passwords and Context of Use

When looking at the selection of pattern it seems like the participants have performed a type of risk assessment. The patterns created do not appear to be created by chance because there is a difference in both average pattern length, creation time, and visual complexity for each pattern types.

Sun et al. [34] investigated the effect of using password meters for Android Lock Patterns to observe if it would assist users in creating stronger patterns. The result showed that the strength of the created patterns when using pattern meters resulted in higher complexity and strength, but also introduced a higher error rate when retyping the pattern. As seen in the result, the patterns created for smartphones have a lower length and complexity. Engelman et al. [16] reported that 33% of the smartphone users were thinking about the locking mechanisms as too much of a hassle. Since higher complexity often can introduce a higher overhead in time used for unlocking the smartphone, it might be easier for users to select a pattern that are less complex to retype and remember.

Money is something all respondents are familiar with and understand the consequence of losing. When looking at the patterns created for smartphones, the results from this study indicated that users are not being aware of the consequences of choosing weak passwords on mobile devices. The respondents in this study on average selected short patterns for smartphones, with a low visual complexity. The described behavior might indicate that users do not understand the consequence if someone could easily guess the password and gain access to the device. A study reported that 26% of the users did not think that someone would care about the information stored on their smartphone [16]. Other studies report otherwise, where only gaining access to for example an email exposed huge amounts of sensitive information [16].

When setting up the survey with the different pattern types, it requires the respondents to make a risk assessment and prioritize what the most important; usability or security. Users might create a more complex pattern for a bank account because losing money is a tangible risk they might be familiar with. A study reported that users are typically more security conscious when they are aware of the need for such behavior [2]. At the same time, a study revealed that 48% of the respondents thought that locking mechanisms were annoying. In the same study, 95% of them agreed or fully agreed that they liked the idea of their smartphone being secured [20].

6.9 Limitations

One of the limitations by utilizing a survey is that the accuracy and honesty of people's responses cannot be verified. The survey was selected for avoiding manual work, as well as for ethical concerns. When needing to keep respondents anonymous, there is no better or easier way for performing data collection and at the same time being able to verify the honesty of the responses.

Another limitation of using an online survey is that it is not possible to have control of who participates due to ethical concerns. When conducting research, it is often desired to have a control group to validate the quality of the collected data. To be able to use a control group one would have to ask people about their real patterns, which is not desirable for both ethical and security reasons. This experiment will only be able to see the choice in patterns based on the properties collected from the survey.

Some of the data properties were subjective, meaning that it is hard to verify the correctness of the data. Som examples are hand size, screen size, and the question asking whether the respondent have a background in IT and security. The two first properties were not used further because the quality was too poor, but the experience of the users was used. There is no way of verifying that the answer is correct because the respondents are the ones deciding what "having a background in IT and security" means.

The research looks at the hypothesis to see whether human properties of the user impacts the choice in graphical passwords. It is important to mention that the result from this research is not valid for all graphical passwords. This research does not test for all human properties, making it possible to do further research on the choice of graphical password including other human properties than the ones contained in this research.

7 Conclusion and Future Work

Section 7.1 presents the conclusion for the hypothesis ond the research questions. Section 7.2 proposes ideas for future work based on the results from this research.

7.1 Conclusion

The hypotheses in this research tests whether users' choice of graphical passwords is influenced by the human properties of the user. The hypotheses are the following:

> H_0: `Human properties have no influence on a user's choice of graphical passwords`
>
> H_1: `A user's choice of graphical passwords is influenced by the human properties of the user`

The human properties included in this research are age, gender, handedness, the user's experience with IT and security, reading/writing direction and hand size.

Unfortunately, the alternate hypothesis was not possible to either accept or reject based on the data collected. First, enough human properties was not collected, which means that the hypothesis may have been too broad for the experiment. To be able to answer this hypothesis, more data and properties need to be analyzed. Second, some properties had to be ignored due to poor data quality and challenges with the population composition.

Even though the hypothesis cannot be accepted or rejected, the results show a significant difference in patterns created by various user types. A conclusion of the research questions is enumerated below.

RQ1 - The choice of graphical passwords and age

The results confirm that the respondents under 25 years creates longer and more complex patterns than people older than 25.

RQ2 - The choice of graphical passwords and gender

There was a significant difference in the patterns created by male and female respondents, where male respondents created longer and more complex patterns.

RQ3 - The choice of graphical passwords and handedness

The results provide no evidence of a correlation between handedness and choice of graphical passwords. The length and visual complexity of the patterns created by left- and right-handed respondent were not significantly different.

RQ4 - The choice of graphical passwords and experience with IT and security

The data shows a significant difference in pattern length and visual complexity between people with high IT and security experience, and those with little experience. The patterns created by experienced respondents were longer and had a higher visual complexity than the patterns created by inexperienced respondents.

RQ5 - The choice of graphical passwords and reading/writing orientation

The collected data was not sufficient to make any conclusions if choice in graphical passwords is influenced by reading and writing orientation.

RQ6 - The choice of graphical passwords and size of hand

The collected hand size data could not be used because the size classification could not be verified. Therefore,no conclusions about hand size and choice of patterns were made.

RQ7 - The choice of graphical passwords of the entire population

The results show predictable behaviour when looking at the entire population. There is a bias towards the selection of starting node. In addition, a significant number of patterns correspond to a letter in the alphabet. The top 100 patterns in the dataset constituted 42% of the collected patterns, indicating that users select similar patterns.

RQ8 - The choice of graphical passwords and context of use

The results show that there is a difference in patterns created for the various pattern types. The respondents create longer and visually more complex patterns for banking accounts. The patterns created for smartphones had the lowest average length as well as being less visually complex.

7.2 Future Work

This section provides a list of three suggestions for future research based on the results in this research.

7.2.1 Reading and Writing Orientation and Choice of Graphical Passwords

Instead of observing a correspondence between handedness and selection of graphical password, this study found that handedness have any impact. Since both left- and right-handed started their patterns on the left side instead of starting at a different side as first predicted, reading and writing orientation looks even more promising than previously thought. Unfortunately, this project did not manage to collect enough data from respondents having an another reading and writing direction than from left-to-right. Reading and writing orientation looks like a human property having a potential for giving positive results when looking at people's choice in graphical passwords.

7.2.2 The Use of Statistical Attack Models in Forensics

This study started with the far-fetched idea of "tell me who you are and I will tell you your lock pattern". This idea is difficult to solve in practice. For being able to achieve the described idea, an attack model needs to be built. Previous research have created a statistical attack model, also known as a *Markov Model*, only using the patterns [39]. By applying knowledge of how patterns are created by the different user types, it is believed that the guessing rate of the model can be improved.

The described model are requested from a security intelligence group in a country whereas the name need to stay unmentioned. Such attack model can assist security intelligences in forensic cases, as mobile devices can provide important data for solving such cases. When needing to catch a criminal, time and information is crucial. Android Pattern Lock have in particular been reported as a problem where it currently lacked research. For future research, it is possible to use the results of this research for building models to predict patterns with a higher success rate than is being possible today.

7.2.3 Exploring Other Graphical Password Schemes

The Android Lock Pattern are one of the few locking mechanisms available on mobile devises. When reviewing the literature, there are lacking studies looking at human properties and choice of graphical passwords. This study looked in particular at the Android Lock Pattern. It would be interesting to see if the results in this research were found across different locking mechanisms. Such knowledge can be used to get a better understanding of how people handle security on mobile devices. How people manage security can provide insightful information for building locking mechanism assisting users to create more secure passwords.

Bibliography

[1] ACM. Acm digital library. `http://dl.acm.org/`. Last accessed: 02.06.2015.

[2] Anne Adams and Martina Angela Sasse. Users are not the enemy. *Commun. ACM*, 42(12):40–46, December 1999.

[3] S Almuairfi, P Veeraraghavan, and N Chilamkurti. IPAS: Implicit password authentication system. In *Advanced Information Networking and Applications (WAINA), 2011 IEEE Workshops of International Conference on*, pages 430–435, March 2011.

[4] F Attneave. Symmetry, information, and memory for patterns. *Am. J. Psychol.*, 68(2):209–222, June 1955.

[5] Robert Biddle, Sonia Chiasson, and P.C. Van Oorschot. Graphical passwords: Learning from the first twelve years. *ACM Comput. Surv.*, 44(4):19:1–19:41, sep 2012.

[6] Greg Blonder. Graphical password, 1996. Patient.

[7] J Bonneau. The science of guessing: Analyzing an anonymized corpus of 70 million passwords. In *Security and Privacy (SP), 2012 IEEE Symposium on*, pages 538–552, may 2012.

[8] Yellow Bridge. Chinese chat codes. `http://www.yellowbridge.com/chinese/pagercodes.php`. Last accessed: 11.05.2015.

[9] James P Byrnes, David C Miller, and William D Schafer. Gender differences in risk taking: A meta-analysis. *Psychol. Bull.*, 1999.

[10] T T Chan and B Bergen. Writing direction influences spatial cognition. *Proceedings of the 27th annual conference of the*, 2005.

[11] D Davis, F Monrose, and M K Reiter. On user choice in graphical password schemes. *USENIX Security Symposium*, 2004.

[12] Antonella De Angeli, Lynne Coventry, Graham Johnson, and Karen Renaud. Is a picture really worth a thousand words? exploring the feasibility of graphical authentication systems. *Int. J. Hum.-Comput. Stud.*, 63(1-2):128–152, July 2005.

[13] R Dhamija and A Perrig. Deja Vu-A user study: Using images for authentication. *USENIX Security Symposium*, 2000.

[14] Ahmet Emir Dirik, Nasir Memon, and Jean-Camille Birget. Modeling user choice in the PassPoints graphical password scheme. In *Proceedings of the 3rd symposium on Usable privacy and security*, SOUPS '07, pages 20–28, New York, NY, USA, 2007. ACM.

[15] P Dunphy and J Yan. Do background images improve draw a secret graphical passwords? *Proceedings of the 14th ACM conference on*, 2007.

[16] S Egelman, S Jain, R S Portnoff, K Liao, S Consolvo, and others. Are you ready to lock? *blues.cs.berkeley.edu*, 2014.

[17] Dinei Florencio and Cormac Herley. A large-scale study of web password habits. In *Proceedings of the 16th International Conference on World Wide Web*, WWW '07, pages 657–666, New York, NY, USA, 2007. ACM.

[18] R S French. Identification of dot patterns from memory as a function of complexity. *J. Exp. Psychol.*, 47(1):22–26, January 1954.

[19] J Goldberg, J Hagman, and V Sazawal. Doodling our way to better authentication. *CHI'02 extended abstracts on*, 2002.

[20] M Harbach, E Von, A Fichtner, and others. It'sa hard lock life: A field study of smartphone (un) locking behavior and risk perception. *Symposium on Usable*, 2014.

[21] C Hardyck and L F Petrinovich. Left-handedness. *Psychol. Bull.*, 84(3):385–404, may 1977.

[22] S Hoober and E Berkman. *Designing Mobile Interfaces*. O'Reilly Media, 2011.

[23] IEEE. Ieee digital library. `http://ieeexplore.ieee.org/Xplore/home.jsp`. Last accessed: 02.06.2015.

[24] I Jermyn, A J Mayer, F Monrose, M K Reiter, and A D Rubin. The design and analysis of graphical passwords. *Usenix Security*, 1999.

[25] D V Klein. Foiling the cracker: A survey of, and improvements to, password security. *Proceedings of the 2nd USENIX Security Workshop*, 1990.

[26] Marte Loege. Graphical passwords - literature review and research design. Norwegian University of Science and Technology, 2014.

[27] NSD. Personversnombudet for forskning. `http://www.nsd.uib.no/personvern/`. Last accessed: 02.07.2015.

[28] Briony J Oates. *Researching Information Systems and Computing*. SAGE Publications, 2006.

[29] A Perrig and D Song. Hash visualization: A new technique to improve real-world security. *Workshop on Cryptographic Techniques*, 1999.

[30] RealUser. `www.realuser.com`. last accessed 15.06.2014.

[31] Pew Internet Project's research. Privacy and data management on mobile devices. `http://www.pewinternet.org/2012/09/05/privacy-and-data-management-on-mobile-devices/`, September 2012. Last accessed: 11.06.2015.

[32] Pew Internet Project's research. Mobile technology fact sheet. `http://www.pewinternet.org/fact-sheets/mobile-technology-fact-sheet/`, January 2014. Last accessed: 11.06.2015.

[33] Springer. Springer digital library. `http://www.springer.com/`. Last accessed: 02.6.2015.

[34] Chen Sun, Yang Wang, and Jun Zheng. Dissecting pattern unlock: The effect of pattern strength meter on pattern selection. *Journal of Information Security and Applications*, 19(4–5):308–320, November 2014.

[35] H Tao and C Adams. Pass-Go: A proposal to improve the usability of graphical passwords. *IJ Network Security*, 2008.

[36] J Thorpe and P C Van. Graphical dictionaries and the memorable space of graphical passwords. *USENIX Security Symposium*, 2004.

[37] J Thorpe and P C Van. Human-Seeded attacks and exploiting Hot-Spots in graphical passwords. *USENIX Security*, 2007.

[38] Julie Thorpe, Brent MacRae, and Amirali Salehi-Abari. Usability and security evaluation of GeoPass: a geographic location-password scheme. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, SOUPS '13, page 14, New York, NY, USA, 24 July 2013. ACM.

[39] Sebastian Uellenbeck, Markus Dürmuth, Christopher Wolf, and Thorsten Holz. Quantifying the security of graphical passwords: The case of android unlock patterns. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer &#38; Communications Security*, pages 161–172, New York, NY, USA, 2013. ACM.

[40] Dirk Van Bruggen, Shu Liu, Mitch Kajzer, Aaron Striegel, Charles R Crowell, and John D'Arcy. Modifying smartphone user locking behavior. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, SOUPS '13, pages 10:1–10:14, New York, NY, USA, 2013. ACM.

[41] Wouter A J van Eekelen, John van den Elst, and Vassilis-Javed Khan. Picassopass: a password scheme using a dynamically layered combination of graphical elements. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 1857–1862, New York, NY, USA, 27 April 2013. ACM.

[42] Christopher Varenhorst, M V Kleek, and Larry Rudolph. Passdoodles: A lightweight authentication method. *Research Science Institute*, 2004.

[43] J Wagemans. Detection of visual symmetries. *Spat. Vis.*, 9(1):9–32, 1995.

[44] Johan Wagemans, James H Elder, Michael Kubovy, Stephen E Palmer, Mary A Peterson, Manish Singh, and von der Heydt. A century of gestalt psychology in visual perception: I. perceptual grouping and figure-ground organization. *Psychol. Bull.*, 138(6):1172, November 2012.

[45] Roman Weiss and Alexander De Luca. PassShapes: Utilizing stroke based authentication to increase password memorability. In *Proceedings of the 5th Nordic Conference on Human-computer Interaction: Building Bridges*, NordiCHI '08, pages 383–392, New York, NY, USA, 2008. ACM.

[46] Susan Wiedenbeck, Jean-Camille Birget, Alex Brodskiy Jim Waters, and Nasir Memon. Authentication using graphical passwords: Basic results. *11th international conference on Human-Computer Interaction (HCI International)*, 2005.

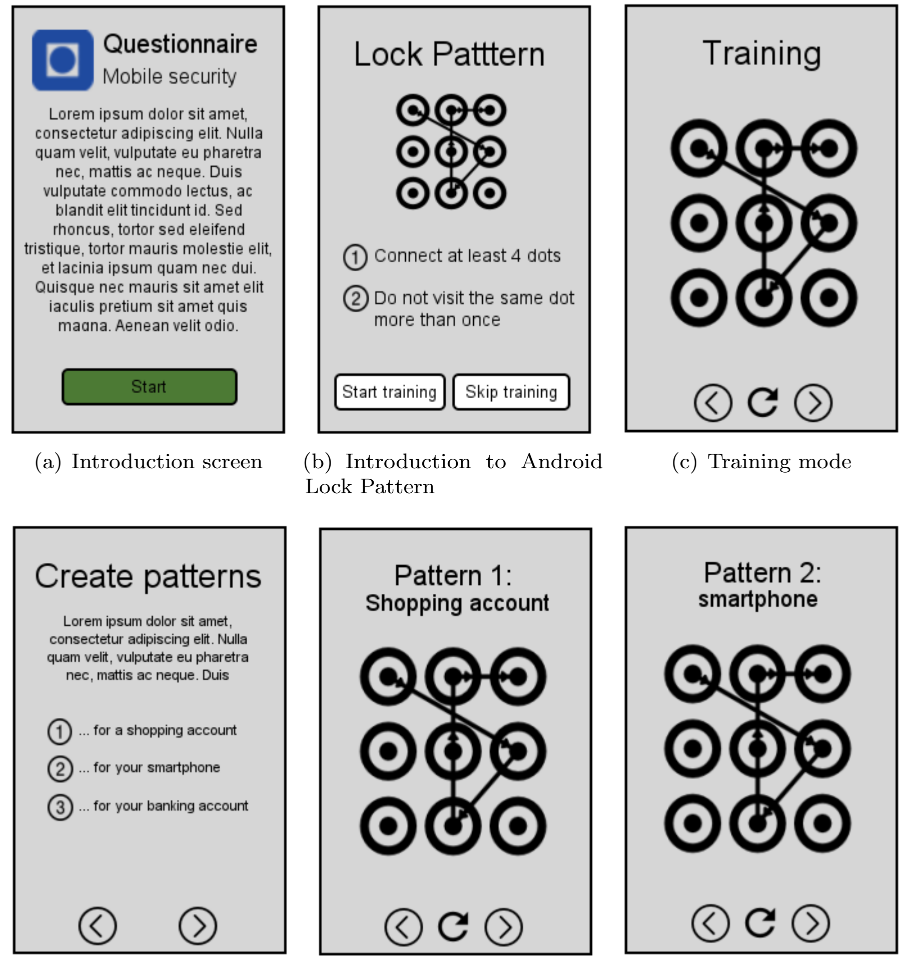
[47] Susan Wiedenbeck, Jim Waters, Jean-Camille Birget, Alex Brodskiy, and Nasir Memon. Authentication using graphical passwords: effects of tolerance and image choice. In *Proceedings of the 2005 symposium on Usable privacy and security*, SOUPS '05, pages 1–12, New York, NY, USA, 6 July 2005. ACM.

[48] Susan Wiedenbeck, Jim Waters, Jean-Camille Birget, Alex Brodskiy, and Nasir Memon. PassPoints: Design and longitudinal evaluation of a graphical password system. *Int. J. Hum. Comput. Stud.*, 63(1-2):102–127, July 2005.

[49] Susan Wiedenbeck, Jim Waters, Leonardo Sobrado, and Jean-Camille Birget. Design and evaluation of a shoulder-surfing resistant graphical password scheme. In *Proceedings of the working conference on Advanced visual interfaces*, AVI '06, pages 177–184. ACM, 23 May 2006.

Appendices

A Wireframes of Survey Aplication



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nulla quam velit, vulputate eu pharetra nec, mattis ac neque. Duis vulputate commodo lectus, ac blandit elit tincidunt id. Sed rhoncus, tortor sed eleifend tristique, tortor mauris molestie elit, et lacinia ipsum quam nec dui. Quisque nec mauris sit amet elit iaculis pretium sit amet quis magna. Aenean velit odio.

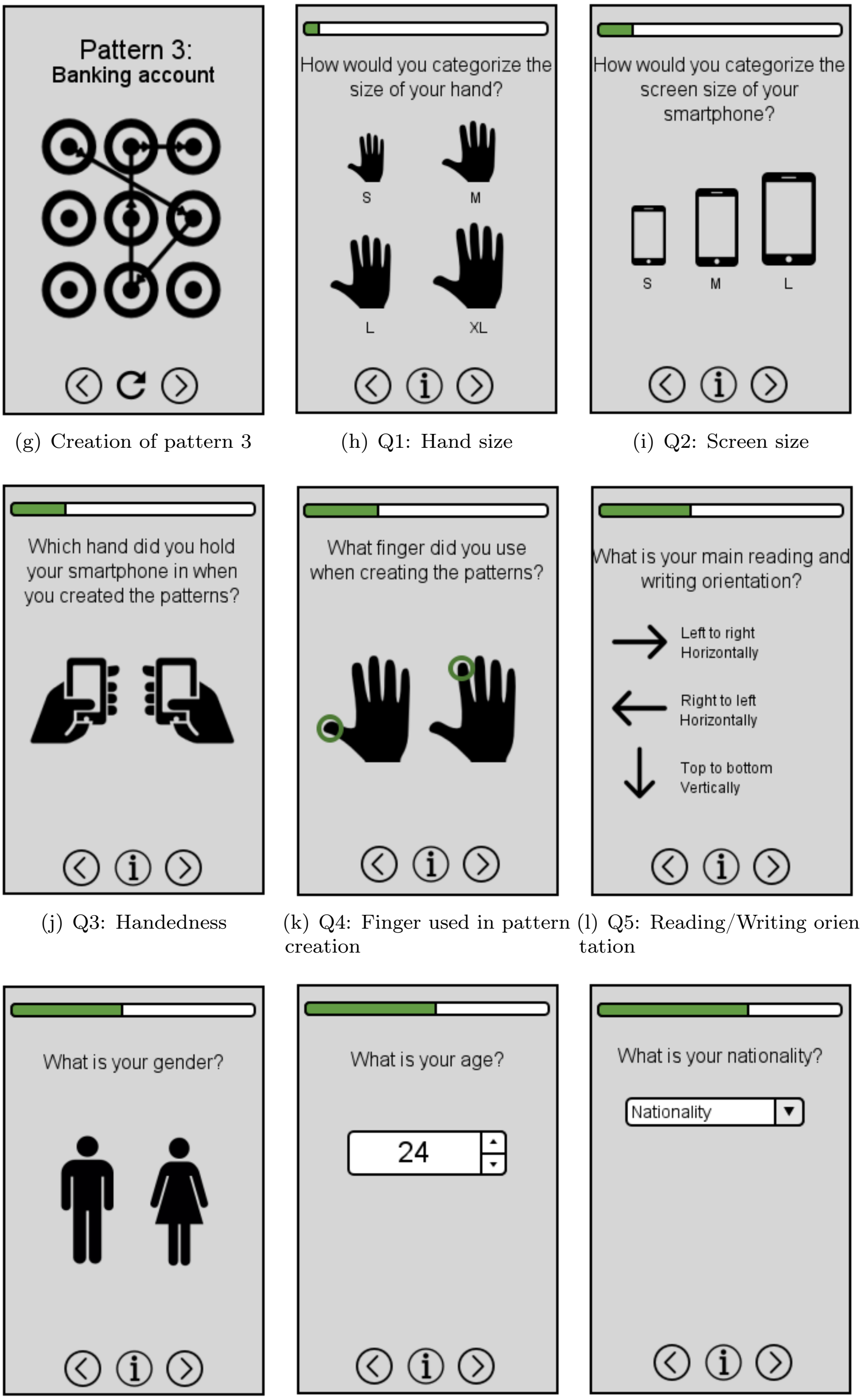
1. Connect at least 4 dots
2. Do not visit the same dot more than once

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nulla quam velit, vulputate eu pharetra nec, mattis ac neque. Duis

1. ... for a shopping account
2. ... for your smartphone
3. ... for your banking account

(a) Introduction screen (b) Introduction to Android Lock Pattern (c) Training mode

(d) Introduction to pattern creation (e) Creation of pattern 1 (f) Creation of pattern 2



Pattern 3:
Banking account

How would you categorize the size of your hand?

How would you categorize the screen size of your smartphone?

(g) Creation of pattern 3

(h) Q1: Hand size

(i) Q2: Screen size

Which hand did you hold your smartphone in when you created the patterns?

What finger did you use when creating the patterns?

What is your main reading and writing orientation?

(j) Q3: Handedness

(k) Q4: Finger used in pattern creation

(l) Q5: Reading/Writing orientation

What is your gender?

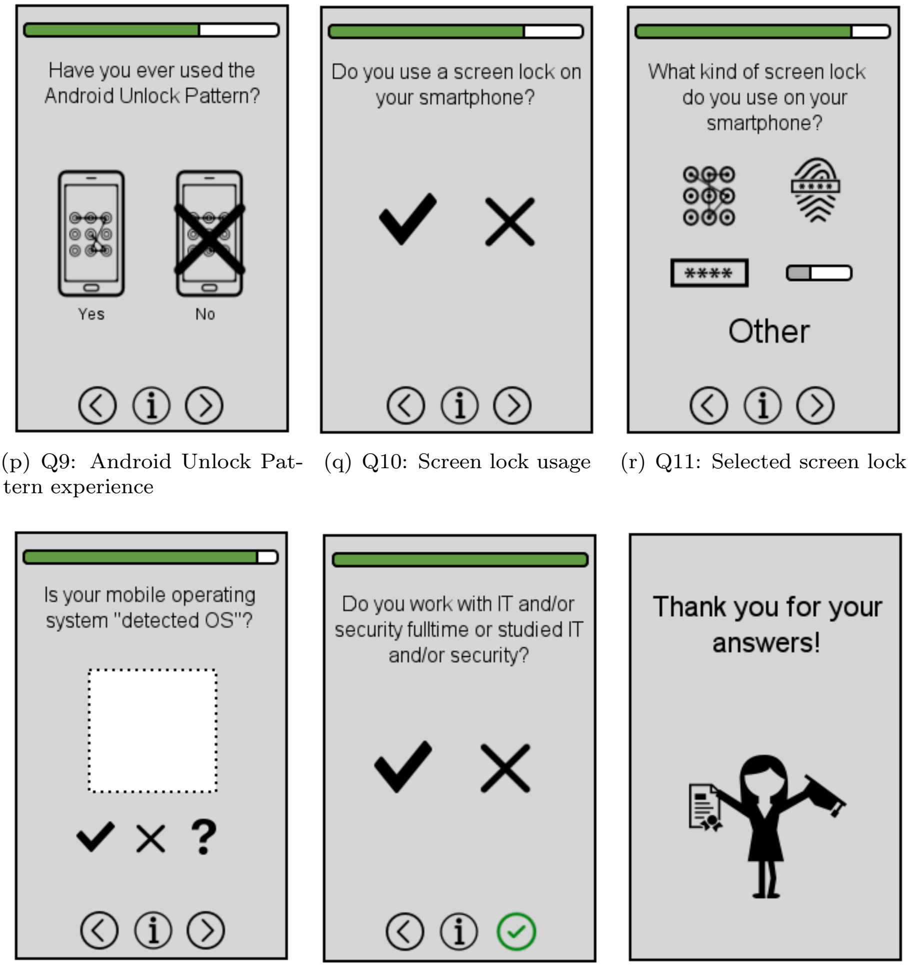
What is your age?

What is your nationality?

(m) Q6: Gender

(n) Q7: Age

(o) Q8: Nationality


Have you ever used the Android Unlock Pattern?
Yes
No
Do you use a screen lock on your smartphone?
What kind of screen lock do you use on your smartphone?
****
Other
Is your mobile operating system "detected OS"?
?
Do you work with IT and/or security fulltime or studied IT and/or security?
Thank you for your answers!

(p) Q9: Android Unlock Pattern experience

(q) Q10: Screen lock usage

(r) Q11: Selected screen lock

(s) Q12: Mobile OS

(t) Q13: Experience with IT and security

(u) Questionnaire completed

Figure A.1: Wireframes