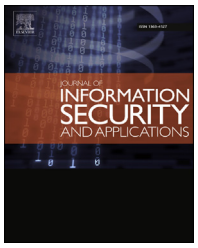
Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jisa



Dissecting pattern unlock: The effect of pattern strength meter on pattern selection

Chen Sun, Yang Wang, Jun Zheng *

Department of Computer Science and Engineering, New Mexico Institute of Mining and Technology, Socorro, NM, 87801, USA

ARTICLE INFO

Article history:
Available online 9 November 2014

Keywords:
Pattern unlock
Pattern strength meter
Mobile security
User study
Android

ABSTRACT

Pattern unlock is one of the entry protection mechanisms in Android system for unlocking the screen. By connecting 4–9 dots in a 3 × 3 grid, the user can set up an unlock pattern which is equivalent to a password or a PIN. As an alternative to the traditional password/PIN, the visual pattern has gained its popularity because of the potential advantages in memorability and convenience of input. However, the limited pattern space and existing attacks such as shoulder surfing, or smudge attack make this mechanism weak in security. In this paper, we analyzed the characteristics of all valid patterns and proposed a way to quantitatively evaluate their strengths. We then designed two types of pattern strength meters as visual indicators of pattern strength. We conducted a user study that involved 81 participants. The results of the user study showed that the presence of visual indicator of pattern strength did encourage users to create visually complex patterns, thus increasing the security of pattern unlock.

1. Introduction

The use of powerful mobile devices such as smartphones and tablets has seen tremendous growth in recent years. Google's Android is one of the most popular mobile platforms that powers hundreds of millions of mobile devices in more than 190 countries around the world (Android). These mobile devices are nearly as powerful as desktop PCs. Users can do many things on their devices including social networking, online shopping and mobile banking. Thus, tons of personal data can be accessed in these devices. Meanwhile, mobile devices can easily be lost or stolen due to their small size, so it poses a need to protect the sensitive data from unauthorized accesses.

Automatic screen lock is the most commonly used strategy in mobile devices to prevent unauthorized access. Android provides several screen lock options including slide, password, PIN (Personal Identification Number), pattern, and the latterly introduced face unlock (Android 4.0; Set Screenlock).

Among these screen lock options, slide unlock provides no protection on entry authorization. It is only used for preventing accidental touches that may trigger certain functions of the system. Password and PIN are very similar as they are also used in other contexts. The user needs to enter a predefined password or PIN to unlock the device. A password can be a combination of numbers, letters, and special symbols, while a PIN consists of only numbers. As required by Android, a password or a PIN should be no less than 4 characters (Set Screenlock). Pattern unlock is a relatively new gesture-based entry protection mechanism introduced by Google in 2008 (Android's Unlock Pattern). To unlock the device with pattern unlock, instead of typing a password/PIN into a text box, the

* *Corresponding author*. Tel.: +1 575 835 6182; fax: +1 575 835 5587.
E-mail address: zheng@nmt.edu (J. Zheng).
http://dx.doi.org/10.1016/j.jisa.2014.10.009
2214-2126/

user is asked to draw a user-defined path by connecting dots in a 3 × 3 grid. Such a path is called an unlock pattern such as the one shown in Fig. 1. The slide, password, PIN, pattern unlock options are available on all Android versions while the last option, face unlock, was introduced in Android 4.0. Using this option for screen unlocking, the system verifies the user by starting the front camera of the device to do a face recognition of the user. This sounds like a very interesting and convenient solution for entry authorization. However, according to some reviews, the face recognition may not work very well when the user angles his/her head slightly away from the camera, or when used in a low-light environment. Moreover, the face recognition can be even fooled by presenting a picture of the authorized user (Android 4.0).

Although pattern unlock is a ready-to-use entry protection mechanism available on all Android devices, it is relatively new compared to the traditional password/PIN. Due to the lack of basic knowledge of pattern unlock, users may choose some weak patterns that cannot provide enough protection against attacks like brute force attack (Botelho et al., 2012), shoulder surfing attack (Tari et al., 2006), and smudge attack (Aviv et al., 2010). In this paper, we examine the basic features of unlock patterns and let users be aware of their security strengths. The contributions of this paper are as follows:

- We studied the characteristics of all patterns allowed by Android, and proposed a way to quantitatively evaluate the complexity of a given pattern.
- We designed two types of pattern strength meters as the visual indicators of pattern strengths.
- We performed a user study involving 81 participants to investigate users' choices of unlock patterns and factors that influence their decisions.
- Based on the results from the user study, we showed that pattern strength meters were effective in encouraging users to create strong patterns.

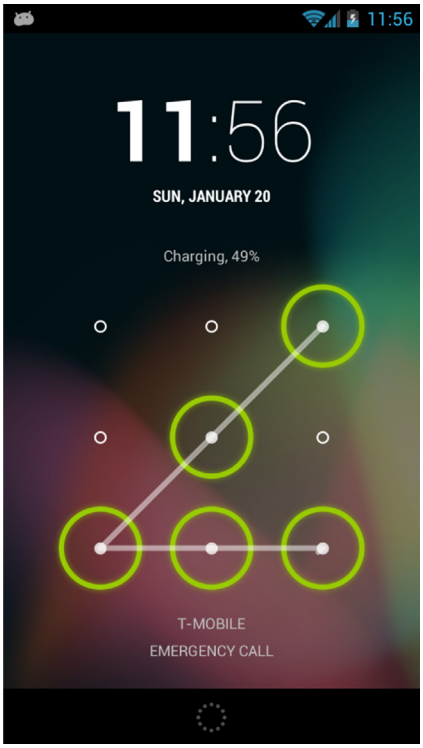


Fig. 1 – An example of unlock pattern.

2. Related work

2.1. Graphical password

Although Android's pattern unlock is relatively new, it can be viewed as a form of graphical password. The idea of the graphical password was originally introduced by Blonder in 1996: given a predetermined image, the user needs to select one or more predetermined positions in a specific order for authentication.

One of the most representative graphical password schemes is Draw A Secret (DAS) proposed by Jermyn et al. in 1999. The DAS scheme was proposed to be used as an encryption tool on Personal Digital Assistant (PDA) devices. In this scheme, the user draws one or more strokes on a $N \times N$ grid. The drawing is then mapped to a sequence of coordinate pairs served as a password.

Dunphy and Yan (2007) introduced background images to the original DAS scheme. In the BDAS (Background Draw a Secret), a user chooses a background image overlaid by the grid and then draws the strokes as in DAS. Their user study showed that users tended to draw more complex strokes when aided by background images.

Proposed by Tao and Adams (2008), Pass-Go is another successor of DAS. In this scheme, users select intersections instead of cells on a 2D grid as a way to input a password. The use of intersections instead of cells overcomes the major drawback of DAS: drawing diagonal lines is difficult. To add error tolerance to enable users to select intersections on a grid, Tao and Adams introduced sensitive areas as round circles with a radius of one-fourth of the side length of a grid cell. In this sense, Android's pattern unlock has a very similar design to Pass-Go.

Orozco et al. (2006) proposed a security enhancement to graphical password by incorporating the checking of haptic parameters such as velocity and pressure. Such a scheme is more resistive to shoulder surfing attacks than other traditional graphical password schemes.

YAGP (Yet Another Graphical Password), proposed by Gao et al. (2008) is another graphical password scheme that tries to distinguish different drawing styles due to user personality. Instead of using the physical haptic features, YAGP compares the stroke trends between two drawings.

2.2. Android's pattern unlock

For Android's pattern unlock, in 2010, Aviv et al. (2010) studied the feasibility of guessing a user's pattern from the oily residues, or smudges, left on the touch screen. In their work, the size of the pattern space, i.e. the total number of valid patterns, was calculated using a brute force method without any further study of the complexity of the patterns.

Andriotis et al. (2013) again focused on the smudge attack of Android's pattern unlock. Besides replicating the experiment from Aviv's work (Aviv et al., 2010), they also studied human behavioral factors on pattern setting. They used heuristics, such as which dot is most frequently used as a starting dot and which sub-pattern is most frequently drawn, to increase the effectiveness of recovering patterns when combined with smudge attack.

Using similar strategies from Orozco et al. (2006), De Luca et al. (2012) tried to enhance the security of Android's pattern unlock by adding implicit authentications based on touch screen patterns, such as XY-coordinates, pressure, and speed. Their conclusion is that it is possible to differentiate users based on the way they draw on the touch screen.

Most recently, Uellenbeck et al. (2013) performed a large-scale user study to investigate how users typically chose patterns. They collected more than 2800 patterns from 580 participants, based on which they used a Markov chain model to quantify the strength of unlock patterns. They also proposed some alternatives of pattern layout to enhance the security of pattern unlock.

In contrast to the aforementioned works, the goal of our work is to try to understand the underlying characteristics of the patterns, and to encourage users to choose strong patterns through the visual indicator of pattern strength.

2.3. *Password meter*

Our idea of calculating pattern strength and providing visual indicators is mostly inspired by the practice of password meter. A password meter, often presented as a colored bar, is a visual indicator that tells users how strong a password is. Many popular websites employ different designs of password meters to assist users in creating strong passwords (Ur et al., 2012).

Ur et al. (2012) did a large scale user study involving 2931 subjects to investigate the effect of password meters on password creation. They tested 14 password meters and found that users presented with any of those meters created passwords that were statistically longer than those created by users seeing no meter.

Egelman et al. (2013) also investigated the effectiveness of password meters. They examined two different cases: passwords used to protect sensitive accounts and password used for unimportant accounts. They concluded that the presence of password meters yielded stronger passwords when users were forced to change existing passwords on "important" accounts.

3. Analysis of pattern unlock

In this section, we dissect the inherent characteristics of pattern unlock. To make it easier when addressing a certain pattern or a path (a segment of a pattern), we label the dots in the 3 × 3 grid from 1 to 9, as shown in Fig. 2. Thus, a pattern can be represented as a sequence of connected dots. For example, the pattern shown in Fig. 2 can be described as 2 → 4 → 8 → 6.

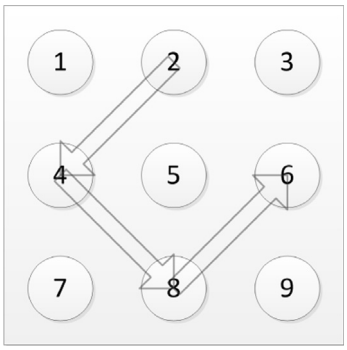


Fig. 2 – The dots in the 3 × 3 grid of the pattern screen labeled from 1 to 9.

We also consider a pattern being composed of multiple segments. For example, the pattern 2 → 4 → 8 → 6 consists of segments 2 → 4, 4 → 8, and 8 → 6. Using these notations, we explore the characteristics of valid unlock patterns.

3.1. *Pattern unlock rules*

As designed by Google, a valid unlock pattern should follow the following rules:

R1: A pattern should connect at least 4 dots.
R2: A dot can be connected only once meaning that a pattern connects no more than 9 dots.
R3: A pattern will always connect the first unconnected dot along its path. Then it may go further to connect other unconnected dots.
R4: A pattern can go through a previously connected dot along its path in order to connect an unconnected dot.

The first two rules are straightforward. To better explain the last two rules, we use the examples shown in Fig. 3. As shown in Fig. 3(a), assuming no dots are connected and we start from dot 1, we may choose to go 1 → 2, 1 → 4, 1 → 5, 1 → 6, or 1 → 8, as indicated by the green arrows (in the web version). However, we cannot directly connect 1 → 3, 1 → 7, or 1 → 9 since there are unconnected dots in the middle of the paths. In other words, in Fig. 3(a), a valid pattern may contain 1 → 2 or 1 → 2 → 3 but not 1 → 3.

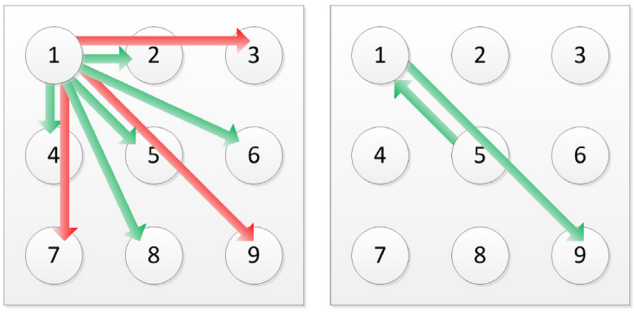


Fig. 3 – Examples of valid and invalid paths.

On the other hand, there could be a way to allow a pattern to contain paths like 1 → 3. According to Rule 4, once a dot is already connected, the pattern may go through it to connect the next unconnected dot along the path. For example, as shown in Fig. 3(b), we first connect 5 → 1. After that, since 5 is already connected, we are now able to go through 5 to connect 1 → 9. Similarly, when 2 is previously connected, we may have 1 → 3 as part of a valid pattern.

As a result of these rules, the number of valid patterns is not simply the factorial of the number of dots being connected. In Section 4.2, we enumerate all valid patterns and provide some statistical results about their characteristics. Before that, we first introduce the characteristics that could be used to describe the visual complexity of a pattern.

4. Pattern strength

When addressing the security issue of pattern unlock, we mainly focus on defending shoulder surfing attack, which happens when a bystander observes the password while the user entering it. Shoulder surfing attack is often considered as a major threat to graphical password (Tari et al., 2006; Orozco et al. (2006); Zakaria et al., 2011). To add shoulder surfing resistance without changing the authentication system itself, one straightforward way is to increase the complexity of the graphical password to limit the attacker's ability to identify and remember the correct password.

In 1957, Attneave (1957) studied the judged complexity problem of shapes, and concluded that the complexity is related to the number of turns in the contour of the shape, the symmetry of the shape, and the variability of angular change between successive turns. Dunphy and Yan (2007) indicated that the number of strokes and the password length of a DAS password (Jermyn et al., 1999) are important security metrics measuring its security strength. For Android's unlock pattern, one of the studies done by Andriotis et al. (2013) asked users to create two Android unlock patterns: the "secure" one and the "easy to remember" one. Statistics show that the "secure" patterns have longer length and more direction changes than the "easy to remember" patterns.

4.1. Patterns characteristics

Inspired by the work mentioned above, we choose the following physical characteristics as the indicators of pattern strength:

Size: The size of a pattern is defined as the number of dots the pattern connects. As constrained by the rules, the size of a valid pattern should be an integer in $\{x|x\in\mathbb{Z}\ and\ 4\leqslant x\leqslant 9\}$. The size of the pattern is considered as one of the most important metrics for pattern strength, as it is equivalent to the password length for textual password, which is essential for calculating the entropy of text password (Burr et al., 2006).

Length: The physical length of a pattern is calculated as the sum of the lengths of all its segments. We define the length of the shortest segment, such as 1 → 2 in Fig. 3(a), as 1. Thus, the length of segment 1 → 3 is 2. The length of segment 1 → 5 and the length of segment 1 → 6 can be calculated using Pythagorean Theorem as $\sqrt{2}$ and $\sqrt{5}$, respectively. The length of segment 1 → 9 is $2\sqrt{2}$ which is the longest among all segments. Empirically, we found that users tend to draw horizontal and vertical lines rather than go diagonally when creating a pattern. The same observation is also obtained by Uellenbeck et al. (2013). As a result, calculating the physical length of the pattern gives more credits to patterns contains diagonal segments, thus adding resistance to guessing attack.

Intersections: When two non-consecutive line segments have a common point, it counts as one intersection. As shown in Fig. 4(a), the pattern 4 → 2 → 5 → 7 → 8 → 1 contains two intersections: one between 8 → 1 and 5 → 7, and another one between 8 → 1 and 4 → 2. In Fig. 4(b), the pattern also has one intersection as the two non-consecutive line segments 2 → 5 and 6 → 4 have a common point at dot 5. Intersections increase visual complexity of the pattern, thus making it more resistant to shoulder surfing attack.

Overlaps: When a line segment of a pattern is covered by another segment, it counts as one overlap. For example, in Fig. 4(b), the pattern 2 → 5 → 6 → 4 contains one overlap between 5 → 6 and 6 → 4. (Note that we separate the two lines in Fig. 4(b) to make it easier for readers to recognize the pattern. In reality, 5 → 6 and 6 → 4 are in the same line.) Overlaps sometimes introduce "hidden complexity" to the pattern. For example, the pattern 2 → 1 → 3 → 6 would highly likely be recognized as 1 → 2 → 3 → 6 by a shoulder surfing attacker if the attacker observed only the final path of the pattern or the partial process of pattern drawing action.

4.2. Statistics of pattern characteristics

To obtain the statistics of all valid patterns, we wrote a brute force program to generate all patterns that follow the rules described in Section 3.1. We found that there are 389,112 valid patterns in total, which is the same as the number obtained by Aviv et al. (2010). In addition to the size of the pattern space, the program also calculates and records the characteristics of all valid patterns. Table 1 presents the statistics of the sizes of all valid patterns. Fig. 5 shows the distributions of three other characteristics including length, intersections, and overlaps. In Fig. 5, different colors (in the web version) are used to represent the distributions of different pattern sizes. Obviously, when a pattern connects more dots, it tends to have a longer length and more intersections and overlaps, which makes the pattern visually more complex. The observations of

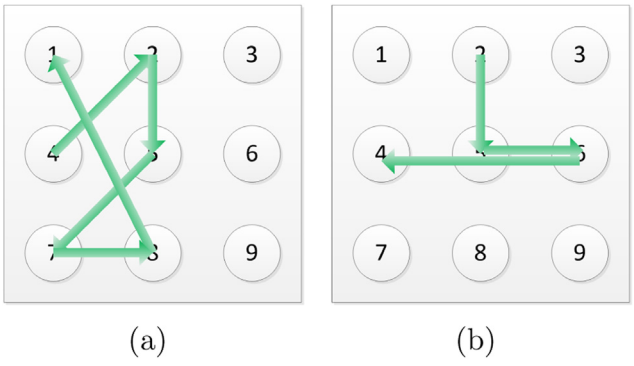


Fig. 4 – Examples of intersections and overlap.

Table 1 – Size statistics of all valid patterns.

# of dots	# of valid patterns
4	1624
5	7152
6	26,016
7	72,912
8	140,704
9	140,704
Total	389,112

pattern characteristics from Table 1 and Fig. 5 are used as the basis for quantitatively evaluating pattern strength.

4.3. *Calculating pattern strength*

For the text-based password, one popular way to determine whether a password is strong or weak is using entropy (Burr et al., 2006). For a password that consists of L randomly selected symbols from a set of N possible symbols, the entropy H is defined as $\log_2$ of the number of possible passwords as shown in Equation (1):

$$H = \log_2 N^L = L \times \log_2 N \quad (1)$$

For example, a password with length 8 and uses only lower-case alphabet characters has an entropy of $8 \times \log_2 26 \approx 37.6$.

For the graphical patterns, we consider the security strength of a pattern is largely determined by its visual complexity. Based on this assumption, we modified the entropy calculation in Equation (1) and came up with the equation to calculate the strength score of a pattern, P, as follows:

$$PS_P = S_P \times \log_2(L_P + I_P + O_P) \quad (2)$$

where PS_P is the strength score of pattern P. S_P, L_P, I_P and O_P are the size, the physical length, the number of intersections, and the number of overlaps of P, respectively.

Using Equation (2), we calculated the strength scores of all valid patterns. The scores vary from 6.340 to 46.807 and the distribution is shown in Fig. 6. It can be seen that connecting more dots could potentially increase the visual complexity of the pattern, thus resulting in higher strength score.

It should be noted that there could be other ways to describe pattern security strength. Fig. 7 shows the relation of

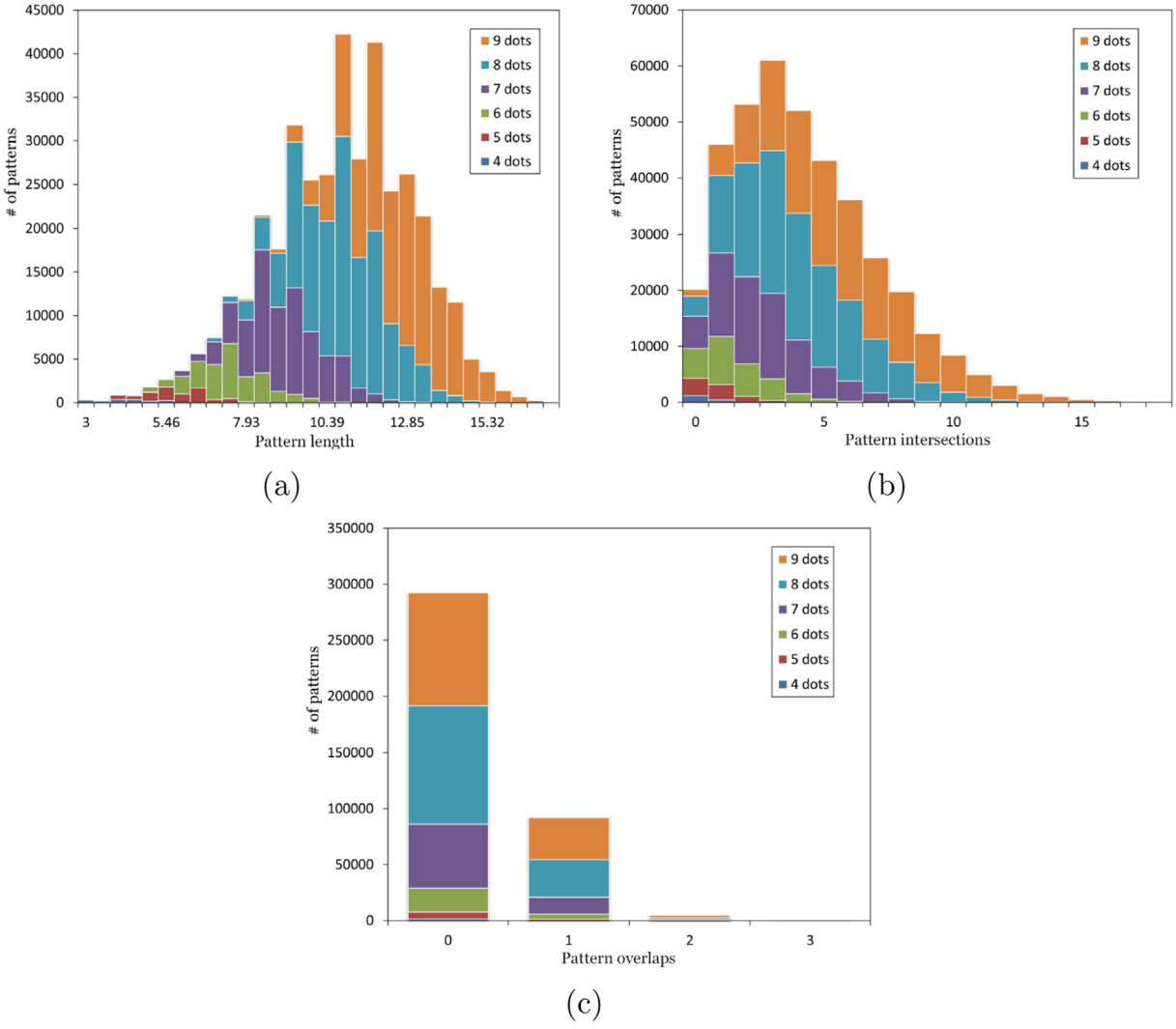


Fig. 5 – Distributions of pattern characteristics: (a) length, (b) intersections, (c) overlaps.

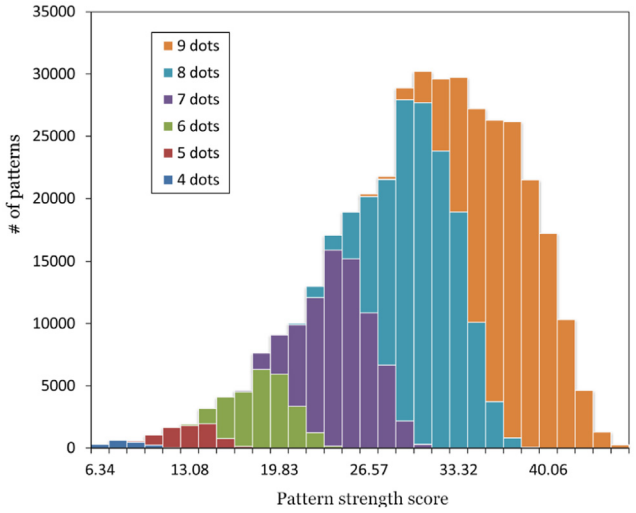


Fig. 6 – Pattern strength score distribution.

the strength score calculated from Equation (2) and the probability obtained from the Markov n-gram probability model which was used in Uellenbeck et al. (2013). The Markov n-gram model describes the probability of a sequence of a given tokens $c_1,c_2,\ldots,c_m$ as

$$P(c_1,\ldots,c_m) = P(c_1)P(c_2|c_1)\ldots P(c_m|c_1,\ldots,c_{m-1}) \quad (3)$$

Uellenbeck et al. (2013) used this model to learn the 2-gram and 3-gram probabilities from their collected data (training set), and then estimated the probabilities of all the remaining patterns. However, it is hard to obtain a reasonable Markov n-gram model from a training set with a small number of patterns (e.g. the Markov n-gram model of password meter obtained by Castelluccia et al. (2012) was trained with the RockYou password list consisting of over 32.6 million passwords). Thus, we calculate the Markov n-gram probability by considering each available dot with equal opportunity to be selected. Although this treatment does not reflect user's preference, it gives some insight into the pattern's strength from another aspect.

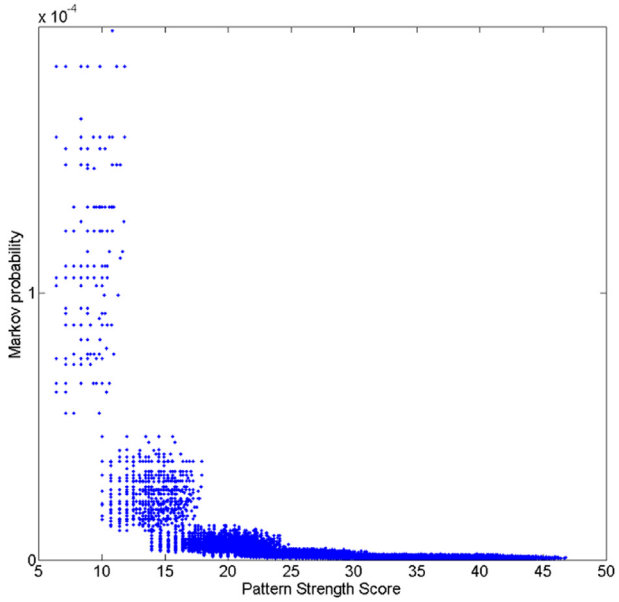


Fig. 7 – Relationship of pattern strength score and Markov n-gram probability.

Fig. 7 shows that the pattern strength score obtained by Equation (2) has significant negative correlation with Markov n-gram probability (Pearson correlation coefficient, $R = -0.4156$, $p < 0.001$). This proves that Equation (2) is reasonable for evaluating pattern strength.

4.4. *Pattern strength meter*

With the quantitative strength score of a pattern, we can enhance the unlock pattern setting interface with a pattern strength meter (PSM) as a visual indicator of pattern strength. Similar to password meter, we expect that a pattern strength meter can help users to create strong patterns.

Following the designs of typical password meters shown by Ur et al. (2012), we designed two types of pattern strength meters as follows:

- PSM1: a 5-segment color changing bar (in the web version) with text description of the pattern strength as shown in Fig. 8(a).
- PSM2: a gradient color ratio bar (in the web version) with percentage strength score as shown in Fig. 8(b).

Both meters use normalized pattern strength scores. PSM1 divides the scores into 5 equal intervals. It displays segments of the bar according to the score interval, and also gives a text description based on the score level as "Very Weak", "Weak", "Medium", "Strong", and "Very Strong". PSM2 directly uses the normalized strength score and displays a colored ratio bar with the score in percentage format.

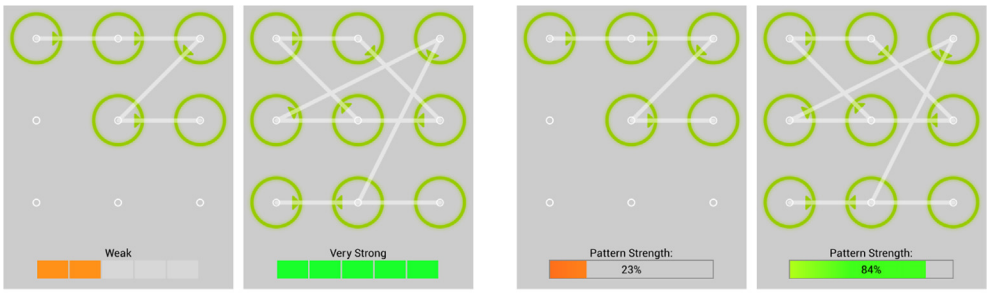
5. User study

We performed a user study to test whether the pattern strength meter can help users to create strong patterns. This study was approved by New Mexico Tech's Institutional Review Board (IRB). Statistical tests were used to evaluate the results of the user study. We used Wilcoxon rank sum test for pair-wise statistical analysis of quantitative group data since the data collected is not normally distributed (One-sample Komogorov–Smirnov test). The χ^2 test was used to analyze categorical data for equality of proportions. The significance level was set to $\alpha = .05$ for all statistical tests.

5.1. *Hypotheses*

The user study was designed to test the following hypotheses:

H1 Users will create stronger patterns with PSM1 compared to without pattern strength meter.
H2 Users will create stronger patterns with PSM2 compared to without pattern strength meter.



(a) PSM1 (b) PSM2

Fig. 8 – Pattern strength meters.

H3 Users will create stronger patterns with PSM1 compared to with PSM2.

5.2. *Participants*

The only requirement was that participants must be over 18 years old. In total, we recruited 81 participants. Since we performed the user study on campus, the majority of the participants were college students. All of them at least had, or were, pursuing a Bachelor's degree. The participants were a wide variety of majors, including Computer Science, Physics, Chemical Engineering, Petroleum Engineering, Biology, Math, Geology, and Management.

The participants were randomly divided into three equal sized groups with 27 participants in each group. The three groups are:

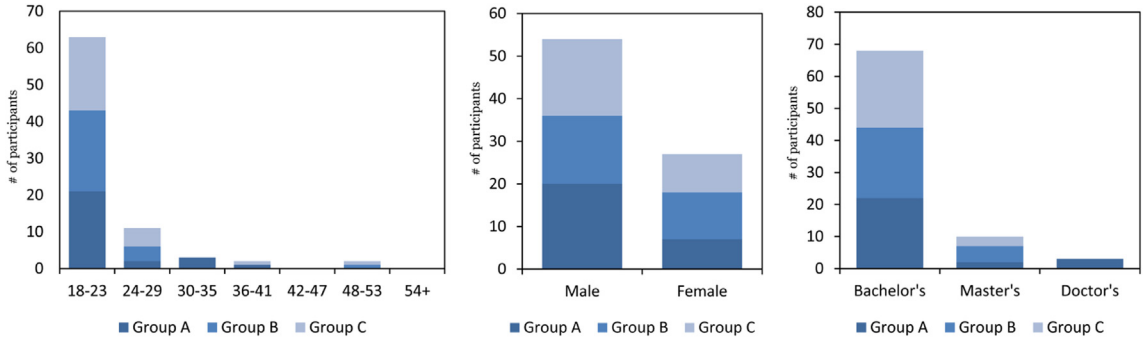
- Group A: The control group in which the participants saw no pattern strength meter when creating patterns;
- Group B: The experimental group 1 in which the participants saw PSM1 when creating patterns;
- Group C: The experimental group 2 in which the participants saw PSM2 when creating patterns.

Fig. 9 presents detailed statistics about participants' basic background information including age (Fig. 9(a)), gender (Fig. 9(b)), and education background (Fig. 9(c)). There is no statistically significant difference across groups on the basic background (χ^2, $p > 0.1$).

We also asked whether the participants owned a smartphone, and whether they used any screen lock mechanism on the phone. The statistics are shown in Fig. 10(a) and (b), respectively. Statistical tests show no significant difference on the participants' smartphone and screen lock experiences across groups (χ^2, $p > 0.67$). Note that when answering the screen lock selection question, some participants checked multiple options. One reason is that some of the participants used more than one screen lock options over times. Another reason is that in iPhone, users can combine slide and password/PIN (called Passcode by Apple (iOS)) in sequence. Excluding those who used no lock, slide unlock or button unlock which provides no entry authorization protection, a total of 38 participants used pattern/password/PIN to protect their phones.

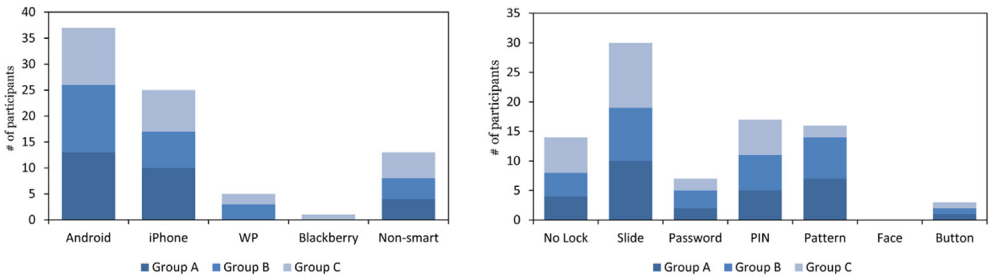
5.3. *Procedure*

We performed our user study in April 2013 on the campus of New Mexico Tech. The participants were told that we were doing some research investigation on user's preference towards Android's unlock pattern. Being opposite to the user study conducted by Uellenbeck et al. (2013), we did not



(a) Age (b) Gender (c) Education

Fig. 9 – Participants' basic background.



(a) Smartphone platforms (b) Screen lock options

Fig. 10 – Participants' smartphone and screen lock experiences.

mention that the specific goal of this study is security-related, since we intended not to give participants an incentive to compete for stronger patterns. We asked our participants to draw patterns that they would like to use on their own devices. Thus, we were able to see the impact of pattern meters on uses' pattern choices.

To facilitate the user study process, we developed an Android app that randomly assigns a participant into one of the three groups and leads the participant to set up unlock patterns. The app simulates the native unlock pattern setting process on Android 4.2 system. When a participant creates a pattern, he/she needs to draw the same pattern twice to confirm the pattern. This requirement is similar to that of setting up a password. If a participant cannot redraw the same pattern for confirmation, he/she can either click the "Cancel" button to start over with a new pattern, or the app will automatically reset to let the participant start over after 5 mismatched drawings. Fig. 11 shows a screenshot of the user study app interface when the app is used by a participant of Group B.

During the user study, we first explained the Android pattern unlock option to the participants. Next, we asked them to create patterns like they had an Android smartphone and used Pattern unlock to protect it.

The participants were then guided by the user study app to set up seven unlock patterns. The first two patterns were just for practice purpose so that the participants could get familiar with the process of setting up an unlock pattern in case they did not have such experiences before. The rest five patterns were recorded by the app for later comparison and analysis. A Samsung Galaxy Nexus phone running Android 4.2 system was used as the experimental device in the user study.

5.4. *Result*

To evaluate the effectiveness of the pattern strength meters on encouraging users to create strong patterns, we first compare characteristics of unlock patterns created by the participants across groups. We then examine the strengths of these patterns.

5.4.1. *Pattern characteristics*

Table 2 compares the characteristics of patterns created by the participants across groups. The mean, the standard deviation, and the median of each metric are presented. Compared with Group A, the participants of Group B and Group C created patterns with significantly more dots, longer length, and more interactions (Wilcoxon rank sum test). However, Wilcoxon rank sum test also shows that the patterns created by Group A did not have significantly different number of overlaps from those created by Group B and C.

5.4.2. *Pattern strength*

Fig. 12 shows the boxplots of the strength scores of 5 patterns created by the participants in the three groups. We found that the participants from Group B and Group C created significant stronger patterns than those from Group A (Wilcoxon rank sum test, $p < 0.001$). The results confirm our hypotheses H1 and H2 that users with pattern strength meter will create

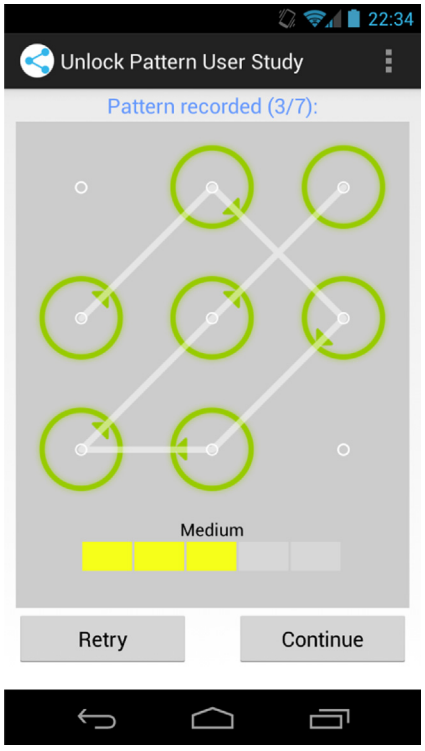


Fig. 11 – Screenshot of the user study Android app interface when used by participants of Group B.

Table 2 – Comparison of the characteristics of created patterns. A symbol '★' indicates that the metric of the group is significantly larger than that of Group A without meter.

Metric	Group A	Group B	Group C
Size		★	★
		$p < 0.001$	$p < 0.001$
Mean	6.948	8.156	7.815
STD	1.774	1.264	1.709
Median	7	9	9
Length		★	★
		$p < 0.001$	$p < 0.001$
Mean	7.180	8.697	8.343
STD	2.428	2.074	2.434
Median	7.414	8.657	8.828
Overlaps			
		$p = 0.151$	$p = 0.253$
Mean	0.015	0.052	0.037
STD	0.121	0.253	0.189
Median	0	0	0
Intersections		★	★
		$p = 0.039$	$p = 0.017$
Mean	1.030	1.444	1.437
STD	1.646	2.128	1.848
Median	0	1	1

stronger patterns compared with those without meter. On the other hand, we found that the patterns created by participants of Group B were not stronger than those created by Group C (Wilcoxon rank sum test, $p = 0.255$). The result rejects our hypothesis H3 which means that our two types of pattern strength meters do not have significant difference in helping users to create strong patterns.

We also compared the pattern strengths from a pattern to the next created pattern for the three groups. As shown in Table 3, no statistical significance can be seen for Group A and Group C. For Group B, there are statistical significance between pattern 2 and pattern 3 and between pattern 4 and pattern 5. However, pattern 3 of Group C has significant lower strength than pattern 2 as shown in Fig. 12. The results

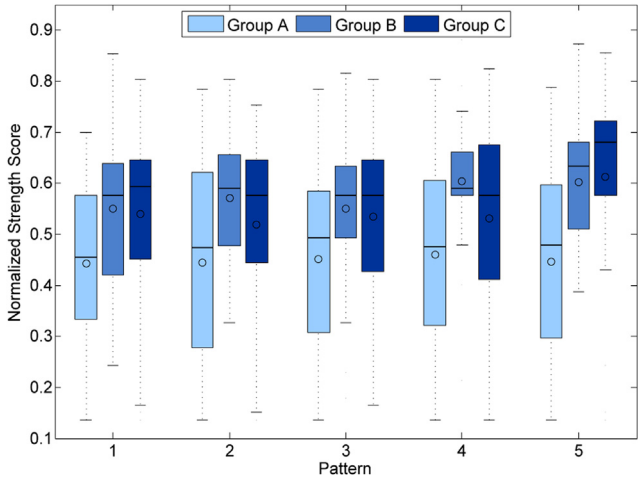


Fig. 12 – Comparison of pattern strength scores across the three groups. The bar and the circle within a box indicate median score and mean score, respectively.

Table 3 – Comparison of pattern strengths from a pattern to the next created pattern for the three groups. A symbol '★' indicates statistical significance (Wilcoxon rank sum test).

Group	1 vs 2	2 vs 3	3 vs 4	4 vs 5
A	$p = 0.924$	$p = 0.931$	$p = 0.876$	$p = 0.802$
B	$p = 0.083$	$p = 0.023^{\star}$	$p = 0.071$	$p = 0.003^{\star}$
C	$p = 0.591$	$p = 0.736$	$p = 1.000$	$p = 0.115$

demonstrate that users did not try to game the meters to achieve higher strength scores.

5.5. *Exit survey*

The last step of our user study for the participants was to fill an exit survey questionnaire to share their thoughts about pattern unlock and pattern strength meter.

The first question of the exit survey asked the participants whether they would like to use pattern unlock assuming they have an Android phone. Fig. 13 shows the result of this question. The figure shows that 33 participants checked "Highly Likely" and 23 participants checked "Somewhat Likely" accounting for 69% of all participants. This indicates that users intend to adopt this relatively new form of screen unlock mechanism.

Among the 9 participants (11%) who answered "Somewhat Unlikely" and "Highly Unlikely" to use pattern unlock, two of them left some comments. One participant prefers PIN to pattern because he/she is fast with a PIN, thus reducing the chance to be seen by other people. The reason of another participant that doesn't like pattern unlock is something about the lack of contact with screen due to hand tremors.

We also asked the participants whether they think pattern strength meter is helpful or not. For those who were in the control group thus hadn't seen our pattern strength meters, we gave them a demo of a randomly selected pattern strength meter from our two designs so that they could also provide opinions. The results are shown in Fig. 14.

As we can see from Fig. 14, in Group A, 13 participants strongly agreed and 9 participants somewhat agreed that pattern strength meter could be helpful. There were 5 participants choosing "Neither agree nor disagree" and no participants chose "Disagree". In Group B, a total of 25 participants

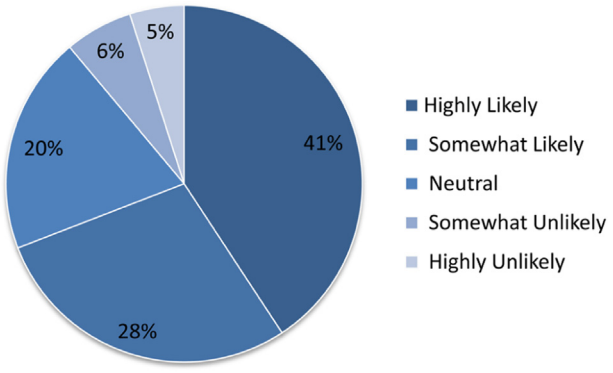


Fig. 13 – Opinions on pattern unlock.

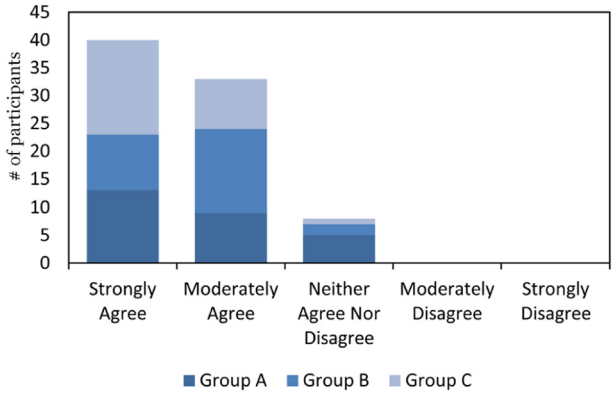


Fig. 14 – Opinions on pattern strength meter.

agreed that PSM1 was helpful, while the rest 2 participants were neutral. In Group C, a total of 26 participants agreed that PSM2 was helpful. Only 1 participant was neutral in Group C. In summary, a majority of participants agreed that pattern strength meter was useful in helping them to create strong patterns. There was no significant difference across groups on the proportion of participants who were positive about the pattern strength meter (χ^2, $p = 0.165$).

In the last question of the exit survey, we let the participants choose from the factors shown below which may prevent them from using strong patterns:

- Pattern memorability (Hard to remember strong patterns).
- Input convenience (Not easy to enter a strong pattern on the touch screen).
- Security need (Not necessary to use strong patterns to protect data on my phone).
- Other (Please specify).
- None of the above (I will use patterns as strong as possible).

We allowed participants to choose multiple factors. The answers from the participants are shown in Fig. 15. The proportion of answers in each option did not differ significantly across groups (χ^2, $p = 0.953$). Among all factors, the majority of the participants checked "Input convenience" and "Memorability" (40 times and 36 times, respectively). "Security need" and "Other" were checked 9 times and 8 times, respectively. Besides, 10 participants reported that they would not care any of these factors and they would use patterns as strong as possible. For those who answered "Other", most of them described something similar to "Memorability", "Input convenience", or "Security need". In next section, we will discuss these three major factors that could affect pattern creation.

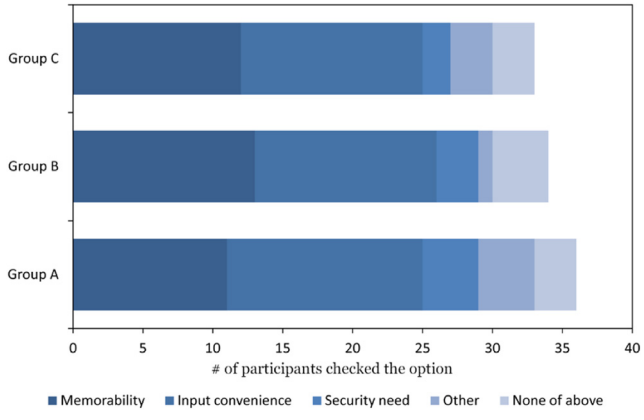


Fig. 15 – Factors that prevent the users from using strong patterns.

6. Discussion

Users are typically security conscious, as long as they are aware of the need for such behaviors (Adams and Sasse, 1999). Most users should be very familiar with password or PIN. For example, an Android user is required to log in with a Google account in order to use the full functionality of the device. Also, users may have the opportunities being educated on password security either by various password strength meters applied by many popular websites (Egelman et al., 2013) or by some password security tips available in online articles or research work (Secure Your Passwords; Gehringer, 2002). Compared with password/PIN, pattern unlock is relatively new to users. Currently there are very few studies addressing the security of this screen unlock mechanism. Therefore, users may not have enough awareness of how secure a pattern is. In this work, we provide a way to calculate the pattern strength and demonstrate that a pattern strength meter could help users to create stronger patterns which are visually more complex.

The security of pattern unlock could be enhanced with visually complex patterns since they are more difficult to recognize and guess, thus making it harder for shoulder surfing attack (Tari et al., 2006), and other attacks like brute force attack (Botelho et al., 2012) and smudge attack (Aviv et al., 2010). However, complex patterns may also incur certain troubles that prevent users from using them. In the following, we discuss the three main factors obtained from our user study.

6.1. *Memorability*

Memorability is certainly one of the major issues when using strong patterns. In our user study, the participants were required to draw each pattern twice to confirm it. If a participant could not redraw the pattern in the second time, either the participant could click a "Cancel" button to start over, or our user study app would automatically reset after 5 mismatched tries. The app kept track of the number of times a participant failed to repeat a pattern for confirmation besides recording the patterns drawn by the participant.

As shown in Fig. 16, the number of users who failed to confirm at least one pattern and the number of patterns failed to be confirmed in Group B were significantly higher than those of Group A (χ^2, $p < 0.001$). For Group C, although the number of users who failed to confirm at least one pattern did not differ from that of Group A (χ^2, $p = 0.214$), they did fail to confirm more patterns (χ^2, $p = 0.026$). This means that when participants were encouraged by the pattern strength meters

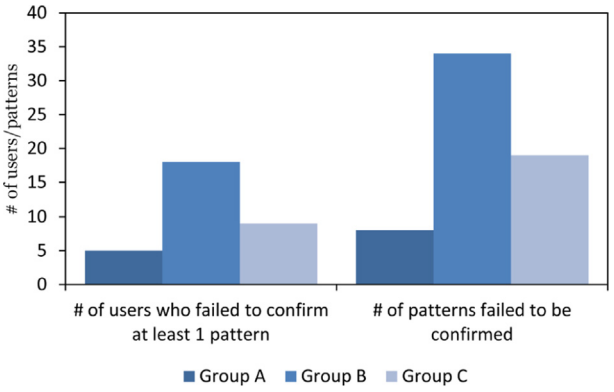


Fig. 16 – Statistics of failed pattern confirmations.

to create patterns with higher strengths, the chances that they could not recall those patterns were also higher.

We also studied the statistics of characteristics and strength scores of failed patterns for each group. The results are shown in Table 4. It can be seen that for each group the pattern characteristics, the strength scores of failed patterns were significantly different from those of successfully created patterns except the size of failed patterns of Group A and Group C (Wilcoxon rank sum test). Wilcoxon rank sum test also revealed that the failed patterns of Group B and Group C did not have significantly higher strength scores than those of Group A ($p = 0.249$ and $p = 0.254$, respectively). This proves that across groups those visually complex patterns with longer length and more intersections and overlaps indeed caused problems to memorize them, even in a very short time period. This result again backs up our assumption that increasing complexity of the pattern could add resistance to shoulder surfing attack since complex pattern is harder to memorize. Shoulder surfing attackers often have a very short time to peep at it which makes it hard for them to get the correct pattern.

Although complex patterns cause the memorability problem just like those complex passwords, it may not be an issue for certain users. Actually, some users may even prefer a graphical pattern than a text password for memorizing. For instance, one of our participants told us that he used pattern unlock on his iPhone. As iPhone does not even provide pattern unlock, the participant explained that he actually used a PIN, and memorized it using the graphical path on the PIN keyboard. So he recognized himself as using kind of pattern unlock. This is a typical example of using graphical shape to support memory (Weiss and De Luca, 2008).

Empirically, we also found that some users tended to use patterns which look like letters or numbers to help memorability. For example, letters like C, L, N, Z, and numbers like 2, 7, 9, can be easily recreated in the 3×3 grid. Such patterns often suffer from guessing attack and shoulder surfing attack due to relatively poor visual complexity. However, these patterns do

Table 4 – Statistics of failed patterns. A symbol '★' indicates that the metric of failed patterns differs significantly from that of successfully created patterns for the corresponding group.

Metric	Group A	Group B	Group C
Size		★	
	$p = 0.094$	$p = 0.037$	$p = 0.130$
Mean	8.000	8.647	8.526
STD	1.414	0.774	0.905
Median	8.5	9	9
Length	★	★	★
	$p = 0.003$	$p < 0.001$	$p < 0.001$
Mean	9.861	10.962	10.846
STD	2.334	2.142	2.075
Median	10.275	11.065	10.828
Overlaps	★	★	★
	$p = 0.037$	$p = 0.033$	$p = 0.027$
Mean	0.125	0.147	0.158
STD	0.354	0.359	0.375
Median	0	0	0
Intersections	★	★	★
	$p = 0.001$	$p < 0.001$	$p < 0.001$
Mean	3	3.794	3.842
STD	1.690	3.427	2.522
Median	3	3	4
Strength score	★	★	★
	$p = 0.007$	$p < 0.001$	$p < 0.001$
Mean	0.631	0.707	0.702
STD	0.175	0.133	0.137
Median	0.675	0.731	0.748

Table 5 – Strengths of sample letter and number patterns.

Letter/ Number	Pattern	Strength score	Score level	Ratio
C		11.358	Weak	24%
L		6.340	Very weak	14%
N		19.401	Medium	41%
Z		19.401	Medium	41%
2		13.932	Weak	30%
7		6.340	Very weak	14%
9		16.844	Weak	36%

not have high strength scores as shown in Table 5. Thus, our pattern strength meters could encourage users to avoid such patterns.

6.2. Input convenience

Input convenience was the reason that caused the highest number of participants in our user study not using strong patterns. As a pattern connects more dots, it takes more time for the user to draw it on the screen. Besides, for a complex pattern which usually contains more intersections and overlaps, the chance of accidentally hitting a wrong dot during the drawing is also increased.

During the user study, we also recorded those successfully created patterns after recovering from failed confirmations. It is reasonable to assume that the failed confirmations of those patterns were mainly due to input errors. We found that 8 Group A participants, 12 Group B participants and 10 Group C participants recovered 10, 13 and 11 patterns, respectively. The statistics of characteristics and strength scores of those recovered patterns are shown in Table 6.

We can observe some statistically significant differences in length and intersections but not in size and overlaps using Wilcoxon rank sum test. The strength scores of each group's recovered patterns did not differ significantly from those of successfully created patterns (Wilcoxon rank sum test, $p > 0.15$). For Group C, the recovered patterns had significantly lower strength scores than those failed patterns (Wilcoxon rank sum test, $p = 0.033$). Thus, the patterns were recovered mainly because they were visually less complex than the failed patterns. We could state that the failed confirmations of them were largely because of accidentally hitting the wrong dots.

Table 6 – Statistics of recovered patterns. A symbol '★' and a symbol '†' indicate that the metric of recovered patterns differs significantly from that of all successfully created patterns and that of failed patterns for the corresponding group, respectively.

Metric	Group A	Group B	Group C
Size			
Mean	7.300	8.231	7.727
STD	1.494	1.013	2.195
Median	7.5	9	9
Length	†	★	†
Mean	7.940	9.866	8.369
STD	2.302	1.778	3.155
Median	7.621	10.071	9.243
Overlaps			
Mean	0	0.077	0
STD	0	0.277	0
Median	0	0	0
Intersections	†	★	†
Mean	1.2	2.692	1.455
STD	2.201	2.097	2.012
Median	0	2	1
Strength score			†
Mean	0.489	0.637	0.544
STD	0.158	0.130	0.243
Median	0.494	0.634	0.645

Note that the input problem could be more severe when the user tries to unlock a phone in a moving state or with one hand. Unfortunately, it seems that there is no way to overcome this drawback when using strong patterns, at least with the current pattern unlock system.

6.3. Security need

There were 9 participants choosing "Security need" as the reason why not using stronger patterns. This may be due to those users not having a strong need of protecting access and data on their phones. Unlike passwords used for protecting online accounts, patterns used for unlocking a phone have a relatively weak security requirement mainly due to the limited pattern space. For users who don't worry much about phone getting lost or being acquired by others, they will not be very serious about screen unlock mechanisms. Thus, they would avoid drawing complex patterns when using pattern unlock.

7. Conclusions

In this paper, we dissected Android's pattern unlock by traversing all valid patterns and analyzing their characteristics. Based on the analysis, we proposed a formula to calculate pattern strength that assigns a higher score to visually more complex patterns. In addition, we designed two types of pattern strength meters to give users a visual indicator of pattern strength. Our user study showed that the presence of visual indicators was effective in helping users to create strong patterns, thus enhancing the security of pattern unlock. However, due to the limited pattern space, the overall security of pattern unlock is still relatively weak compared with PIN/password. In future, we would like to investigate additional ways to enhance the security of pattern unlock such as using biometrics information (De Luca et al., 2012). Another issue worthy of investigation is to see how the proposed pattern strength score formula correlates with the Markov n-gram model obtained from a large training set of patterns created by users once the set is available.

REFERENCES

Adams A, Sasse MA. Users are not the enemy. Commun ACM 1999;42(12):40–6. doi: URL, http://doi.acm.org/10.1145/322796.322806.

Andriotis P, Tryfonas T, Oikonomou G, Yildiz C. A pilot study on the security of pattern screen-lock methods and soft side channel attacks. In: Proceedings of the sixth ACM conference on Security and privacy in wireless and mobile networks, WiSec '13. New York, NY, USA: ACM; 2013. p. 1–6. doi: URL, http://doi.acm.org/10.1145/2462096.2462098.

Android 4.0 ice cream sandwich review, http://www.engadget.com/2011/12/01/android-4-0-ice-cream-sandwich-review/.

Android, the world's most popular mobile platform, http://developer.android.com/about/index.html.

Android's unlock pattern, http://www.touchusability.com/blog/2008/9/23/androids-unlock-pattern.html.

Attneave F. Physical determinants of the judged complexity of shapes. J Exp Psychol April, 1957;53(4):221–7.

Aviv AJ, Gibson K, Mossop E, Blaze M, Smith JM. Smudge attacks on smartphone touch screens. In: Proceedings of the 4th USENIX conference on Offensive technologies, WOOT'10. Berkeley, CA, USA: USENIX Association; 2010. p. 1–7. URL, http://dl.acm.org/citation.cfm?id=1925004.1925009.

Blonder GE. Graphical password. U.S. Patent 5559961. Sep. 1996. URL, http://www.freepatentsonline.com/5559961.html.

Botelho B, Nakamura E, Uto N. Implementation of tools for brute forcing touch inputted passwords. In: Internet Technology And Secured Transactions, 2012 International Conference For; 2012. p. 807–8.

Burr WE, Dodson DF, Polk WT. Electronic authentication guideline. nIST Special Publication 800–63. 2006.

Castelluccia C, Drmuth M, Perito D. Adaptive password-strength meters from Markov models. In: NDSS, the internet society; 2012. URL, http://dblp.uni-trier.de/db/conf/ndss/ndss2012.html#CastelluccioDP12.

De Luca A, Hang A, Brudy F, Lindner C, Hussmann H. Touch me once and i know it's you!: implicit authentication based on touch screen patterns. In: Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems, CHI '12. New York, NY, USA: ACM; 2012. p. 987–96. doi: URL, http://doi.acm.org/10.1145/2208516.2208544.

Dunphy P, Yan J. Do background images improve "draw a secret" graphical passwords?. In: Proceedings of the 14th ACM Conference on Computer and Communications Security, CCS '07. New York, NY, USA: ACM; 2007. p. 36–47. doi: URL, http://doi.acm.org/10.1145/1315245.1315252.

Egelman S, Sotirakopoulos A, Muslukhov I, Beznosov K, Herley C. Does my password go up to eleven?: the impact of password meters on password selection. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '13. New York, NY, USA: ACM; 2013. p. 2379–88. doi: URL, http://doi.acm.org/10.1145/2470654.2481329.

Gao H, Guo X, Chen X, Wang L, Liu X. YAGP: yet another graphical password strategy. In: Proceedings of the 2008 Annual Computer Security Applications Conference, ACSAC '08. Anaheim, CA, USA: IEEE Computer Society; 2008. p. 121–9. doi: URL, http://dx.doi.org/10.1109/ACSAC.2008.19.

Gehringer E. Choosing passwords: security and human factors. In: Technology and Society, editor. (ISTAS'02). 2002 International Symposium on, 2002; 2002. p. 369–73. http://dx.doi.org/10.1109/ISTAS.2002.1013839.

Introducing Android 4.0, http://www.android.com/about/ice-cream-sandwich/.

iOS passcodes: understanding passcodes, http://support.apple.com/kb/HT4113.

Jermyn I, Mayer A, Monrose F, Reiter MK, Rubin AD. The design and analysis of graphical passwords. In: Proceedings of the 8th Conference on USENIX Security Symposium. SSYM'99, vol. 8. Berkeley, CA, USA: USENIX Association; 1999. p. 1. URL, http://dl.acm.org/citation.cfm?id=1251421.1251422.

Orozco M, Malek B, Eid M, Saddik AE. Haptic-based sensible graphical password. Proceedings of Virtual Concept. 2006. URL http://citeseerx.ist.psu.edu/viewdoc/download.

Secure your passwords, http://www.google.com/goodtoknow/online-safety/passwords/.

Set screen lock, http://support.google.com/android/bin/answer.py?hl=en&answer=2677611.

Tao H, Adams C. Pass-go: a proposal to improve the usability of graphical passwords. Int J Netw Secur 2008;7(2):273–92.

Tari F, Ozok AA, Holden SH. A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords. In: Proceedings of the second symposium on Usable privacy and security, SOUPS '06. New York, NY, USA: ACM; 2006. p. 56–66. doi: URL, http://doi.acm.org/10.1145/1143120.1143128.

Uellenbeck S, Dürmuth M, Wolf C, Holz T. Quantifying the security of graphical passwords: the case of android unlock patterns. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security, CCS '13. New York, NY, USA: ACM; 2013. p. 161–72. doi: URL, http://doi.acm.org/10.1145/2508859.2516700.

Ur B, Kelley PG, Komanduri S, Lee J, Maass M, Mazurek ML, et al. How does your password measure up? The effect of strength meters on password creation. In: Proceedings of the 21st USENIX conference on Security symposium, Security'12. Berkeley, CA, USA: USENIX Association; 2012. p. 5. URL, http://dl.acm.org/citation.cfm?id=2362793.2362798.

Weiss R, De Luca A. PassShapes: utilizing stroke based authentication to increase password memorability. In: Proceedings of the 5th Nordic Conference on Human-computer Interaction: Building Bridges, NordiCHI '08. New York, NY, USA: ACM; 2008. p. 383–92. doi: URL, http://doi.acm.org/10.1145/1463160.1463202.

Zakaria NH, Griffiths D, Brostoff S, Yan J. Shoulder surfing defence for recall-based graphical passwords. In: Proceedings of the Seventh Symposium on Usable Privacy and Security, SOUPS '11. New York, NY, USA: ACM; 2011. 6:1–6:12. doi: URL, http://doi.acm.org/10.1145/2078827.2078835.