On Quantifying the Effective Password Space of Grid-based Unlock Gestures

Emanuel von Zezschwitz, Malin Eiband, Daniel Buschek, Sascha Oberhuber, Alexander De Luca, Florian Alt, Heinrich Hussmann
LMU Munich, Munich, Germany
{emanuel.von.zezschwitz, malin.eiband, daniel.buschek, alexander.de.luca, florian.alt, hussmann}@ifi.lmu.de, oberhuber@cip.ifi.lmu.de

ABSTRACT

We present a similarity metric for Android unlock patterns to quantify the effective password space of user-defined gestures. Our metric is the first of its kind to reflect that users choose patterns based on human intuition and interest in geometric properties of the resulting shapes. Applying our metric to a dataset of 506 user-defined patterns reveals very similar shapes that only differ by simple geometric transformations such as rotation. This shrinks the effective password space by 66% and allows informed guessing attacks. Consequently, we present an approach to subtly nudge users to create more diverse patterns by showing background images and animations during pattern creation. Results from a user study ($n = 496$) show that applying such countermeasures can significantly increase pattern diversity. We conclude with implications for pattern choices and the design of enrollment processes.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords

unlock pattern; security; similarity; metric; user selection; password space

INTRODUCTION

Today's smartphones provide access to a lot of sensitive information, motivating users to use authentication mechanisms such as PINs, patterns, Face Unlock or Touch ID.

Unlock patterns are particularly popular among users [31], since visual passwords are usually easy to remember. In addition, failure to log in comes at low cost, as users can immediately re-enter their pattern [31]. However, unlock patterns are prone to different types of attacks based on shoulder surfing [30] or the analysis of smudge stains on the screen [4]. In addition, analyses of user-defined unlock patterns indicated that users follow predictable selection strategies [28]. As a consequence of this biased pattern choice, only a subset of the theoretically possible pattern space is actually used. This makes the Android unlock system prone to dictionary attacks.

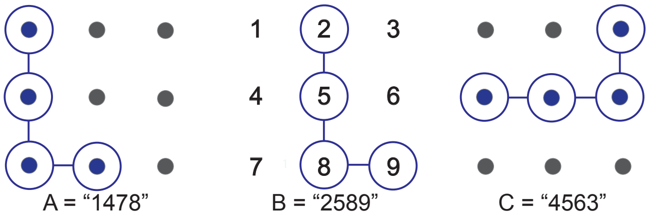


Figure 1. Unlock patterns are usually represented by digits. While such codes indicate different passwords, all three patterns are based on a simple L-shape. The similarity metric proposed in this paper assesses the distance between such patterns by analyzing the number of geometric transformations needed to convert one pattern into another.

In this paper, we present a two-fold approach to assess and address the lack of pattern diversity: First, we introduce a novel measure to assess the strength of given unlock patterns based on their geometric similarity (Figure 1). We follow Li and Vitányis's use of the Kolmogorov Similarity measure [15]. It quantifies similarity through the computational efforts required to derive a given representation of a data object A from another one B. We adopt this metric in a greedy clustering algorithm to reveal pattern groups and central patterns, which are similar to as many others as possible. The results show that applying up to two simple transformations already reduces the given practical pattern space by about two thirds. This indicates that humans indeed favor a small set of known shapes over semi-random sequences of strokes.

Consequently, the second part of our work aims at increasing practical pattern diversity. Motivated by promising results in other projects [1, 10], we present an approach that subtly nudges users to create more diverse patterns via images and animations. These images are shown in the background of the grid used for pattern creation. Results from a user study indicate that such simple changes in the user interface can lead to a significantly more diverse practical pattern space.

In summary, we contribute: 1) a user-centric metric to measure the diversity of pattern choices, applied to 506 user-defined unlock patterns; 2) a large-scale evaluation (n=496) of a method to increase pattern diversity via background images shown at the time of enrollment; 3) important insights on user-defined unlock patterns; and 4) implications for the design and the evaluation of novel authentication concepts.

MUM '16, December 12 - 15, 2016, Rovaniemi, Finland

ISBN 978-1-4503-4860-7/16/12. . . $15.00

DOI: http://dx.doi.org/10.1145/3012709.3012729

RELATED WORK

Android unlock patterns can be categorized as drawmetric graphical passwords [5, 21]. In general, drawmetric concepts are based on the recall of a previously memorized shape, sketch or gesture [12]. The first drawmetric password system, Draw-a-Secret (DAS) [14], allowed arbitrary drawings on a 5×5 grid-based canvas. Over the following years various modifications of Draw-a-Secret were proposed. For example, Passdoodles [29] resigned the visible grid or Qualitative Draw-A-Secret (QDAS) [16] relied on qualitative direction changes. Finally, Tao and Adams [26] presented Pass-Go which was based on predefined shapes which appeared whenever sensitive areas at the intersections of the grid were touched. Android unlock patterns [4] can be seen as a successor of Pass-Go as the designers followed the idea of activating predefined areas. However, the concept was further simplified by limiting the active input area to a 3×3 grid of touchable cells.

Today, Android patterns are widely accepted as a usable alternative to alphanumeric secrets [31]. However, research has shown that the input of unlock patterns is relatively easy to observe [30] and most patterns are easy to guess [28]. The predictability of drawmetric passwords has first been discussed by Nali and Thorpe [19] who analyzed Draw-a-Secret and found out that users prefer symmetric shapes with few strokes and tend to place them in the center of the grid. The analysis of Android unlock patterns indicated similar problems. Uellenbeck et al. [28] reported that most patterns are drawn from left to right and that users prefer the upper left cell as a starting point. Andriotis et al. [3] confirmed that biased selection behavior limits the practical password space of the authentication system.

Over the following years proactive pattern checking was discussed as a possible solution [2, 23, 24, 25]. Analogous to alphanumeric password meters, pattern meters calculate the strength of a given pattern based on specific composition aspects. Proposed metrics included pattern length [2, 24, 25], direction changes [2] or so called knight moves[1] [2]. However, the relative weights of these measures were not further evaluated and it remains unclear which pattern features make the secret hard to guess. Nevertheless, simulated guessing attacks indicated that the presence of pattern meters results in a harder-to-guess pattern selection. On the downside, the starting points were hardly influenced [23].

The literature review reveals that pattern strength is mostly derived from composition characteristics like length or complexity. However, user studies concerning alphanumeric passwords indicated that such metrics do not reflect user choice and therefore have limited value for predicting guessability [11, 18]. Moreover, the performance of simulated guessing attacks strongly depends on appropriate training data [32]. As unlock patterns are usually stored on the device, many patterns for training can only be collected during simulated enrollments within user studies, which questions the comparability of this method. Even though a recent field study [13] analyzed meta data of the entered secrets (e.g., pattern length), the actual pattern was not recorded.

In addition to proactive recommender systems, changes in the graphical user interface were shown to influence password choice [27]. Uellenbeck et al. [28] proposed several redesigns of the common 3×3 pattern grid. The analysis revealed that new layouts anticipated known selection patterns but at the same time introduced new predictable behavior. This indicates that simply changing from one layout to another is not effective in the long run. Instead of changing the grid layout, Dunphy et al. [10] proposed to add background images to the Draw-a-Secret concept to more subtly influence user choice. The results revealed that background images can nudge users to select more complex and less symmetric patterns. In addition, the resulting sketches were significantly longer and less centered. Por et al. [20] confirmed these positive effects with a version of Pass-Go. Finally, it was also shown in the context of authentication schemes where passwords are defined as a sequence of pass-points on an image that the underlying image has a strong effect on password choice [1, 7].

The remainder of this paper is structured as follows. Firstly, we present a new metric for grid-based patterns which is based on geometric similarity and the assumption that unpopular patterns are more secure [17, 22]. Secondly, we present a concept based on related work [10] which applies background images and animations to the Android pattern unlock to nudge users to select more diverse patterns. In addition to lab-based evaluations, we analyze the impact of this concept in a large-scale online study. We then compare the resulting patterns to the standard Android unlock patterns using the proposed similarity metric. The paper closes with a discussion on pattern predictability, possible countermeasures, and the general impact of the study design.

A SIMILARITY METRIC FOR UNLOCK PATTERNS

Our approach compares patterns according to their geometric similarity. While popular patterns are assumed to be easy to guess, patterns which significantly differ from such popular patterns are assumed to provide a higher level of security. In contrast to simulated guessing attacks the proposed metric supports the detailed comparison of multiple pattern sets without the need for training data.

Definition

The goal of our work is to determine how many similar shapes a given set of user-defined unlock patterns contains. To this end, we first define a similarity metric inspired by Euclidean plane isometries [6]: Two patterns A and B are n-similar for $n \in \mathbb{N}_0$ resp. a set of transformations T, if A can be transformed into B with exactly n geometrical or logical transformations from T.

For example, if we choose $n = 1$, this means that patterns in the same group differ by one transformation. We define the following transformations T:

- *Rotation*: Rotate a pattern by 90, 180 or 270 degrees
- *Translation*: Translate a pattern by 1 point in either north, east, south or west direction
- *Mirror*: Mirror a pattern on the x-axis or y-axis
- *Inversion*: Traverse a pattern in inverted order

[1] A connection of two nodes which are not neighboured (e.g., 2-7).

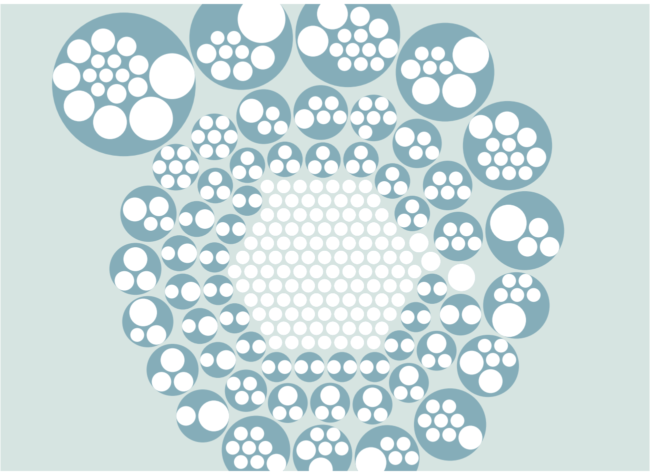


Figure 2. "Circle-packing" chart visualizing the results for $n \leq 2$ (based on 506 samples). Each white circle is a pattern; its size grows with its total occurrence count in the database. Dark circles enclose groups of similar patterns. The small white circles in the center represent patterns that remained unique after applying the grouping algorithm.

If A and B are equal, they are assigned a distance of 0, thus matching the Kolmogorov distance [15]: A higher number of operations translates to higher computational complexity. We limit our analysis to groups of congruent patterns (with equal length), since we consider shapes to be an important property in the pattern creation process, and a change in length often alters the shape of a pattern.

Grouping Patterns

We cluster patterns based on similarity, as defined above. Each group contains all patterns n-similar to a "central" pattern within the group. Each pattern is assigned to exactly one group. Choosing the central patterns in a way that minimizes the total number of groups presents an optimization problem.

Example Problem

To motivate and explain our approach to solving this problem, consider the following example:

$$A \xrightarrow{1} B \xrightarrow{1} C$$

with $n = 1$ for the pairs $\{(A,B),(B,C)\}$. This is a common situation – imagine A being an "L" aligned to the left (⣇⠅), B the same "L" translated right (⠅⣇), and C being B rotated by 90 degrees (⠭). This means that A has a distance of two to C, while B has a distance of one to both (see Figure 1). The optimization problem is evident here: If we choose $n = 1$ (i.e. the maximum distance within groups is 1) and start with A as a central pattern, this would result in two groups $\{A,B\}$ and $\{C\}$. If we used B as the central pattern instead, we would only need a single group $\{A,B,C\}$.

Approach - Greedy Clustering

In general, finding the "best" set of central patterns is equivalent to the Minimum Set Covering problem [9], and thus NP-hard. Therefore, an optimal solution cannot be computed for problems of interesting size in a feasible amount of time. As a consequence, we use a greedy algorithm instead. In each step, we add one ungrouped pattern to the set of central patterns – the one which is n-similar to the largest number of yet ungrouped patterns. This procedure is repeated until each pattern is part of one group, possibly a group containing only one pattern (i.e. a unique one, which cannot be derived with n transformations from any other pattern).

In this work, we consider a maximum of n=2, because we are particularly interested in closely related patterns. Figure 2 visualizes[2] the results of the greedy clustering applied to a dataset, collected under standard Android conditions. It shows that despite a number of unique patterns (in the center), the vast majority of patterns belongs to groups, meaning that they can be derived from each other with simple transformations. The next section discusses the results in detail.

ASSESSING THE SIMILARITY OF UNLOCK PATTERNS

The similarity metric is now applied to a set of 506 user-defined Android unlock patterns.

Data Collection

To quickly collect a huge set of real-life unlock patterns, we utilized Amazon Mechanical Turk (MTurk).

Procedure

After the task was accepted, the MTurk users were asked to open an external URL on their smartphone, linking to our web application. We used PHP Mobile Detect[3] to ensure that participants actually used their mobile devices to draw the unlock patterns. The pattern creation process followed the standard Android enrollment procedure[4]. This means that the patterns had to be conform to the standard Android rules and that selected patterns had to be confirmed once. Participants were allowed to reset their pattern and start again. After the input was recorded, we collected demographical data and gathered information on the technical experience. Finally, we provided a secret code which confirmed the completion of the study. The whole procedure took 102 seconds on average (SD=53). Each participant was compensated with US$ 0.5.

Participants

We checked that all participants provided the correct secret code. In addition, we validated the given answers and excluded participants who did not fulfill the requirements. For example, we excluded participants who stated not to use mobile devices. Finally, 506 participants contributed to the data set. All participants were US citizens, 334 were male and 172 were female. The average age was 28 years (18-67, SD=8). 50.2% reported to use Android smartphones, 49.4% used iPhones.

Results

Basic Statistics

As each participant contributed one unlock pattern, the final data set comprised 506 patterns. The average pattern length was 5.03 points (SD=1.4). The favored starting point was at the top left with 41.1%, and 20.0% of the patterns finished at the bottom right. In terms of occurrence of complexity aspects, the results of our dataset confirm what was found in previous research [3, 28]. Only 7.3% of the patterns included overlapping nodes, only 5.9% had knight moves.

[2] Generated with http://d3js.org Library, BSD license. Copyright 2015 Mike Bostock.

[3] http://mobiledetect.net/, MITLicense. Accessed: 2016/08/25.

[4] http://phandroid.com/2014/03/20/android-101-lock-screen/; Accessed: 2016/08/25.

Rank	Top	# Permutations	# Occurrences	% Dataset
1	(1478)	17	56	11.1
2	(1596)	9	25	4.9
3	(5896)	13	24	4.7
4	(1256)	8	23	4.5
5	(1235)	12	18	3.6

Table 1. The five largest groups for $n \leq 2$. The table shows the most frequent (top) pattern of each group, the number of different patterns covered in the group, the accumulated absolute number of occurrences for all patterns in the group, and the covered ratio of the whole dataset.

Length	Total		$n = 0$		$n = 1$		$n = 2$		$n \leq 2$	
any	506	*100%*	350	*69.2%*	213	*42.1%*	179	*35.4%*	169	*33.4%*
4	262	*51.8%*	156	*30.8%*	68	*13.4%*	50	*9.9%*	44	*8.7%*
5	119	*23.5%*	94	*18.6%*	64	*12.6%*	56	*11.1%*	52	*10.3%*
6	48	*9.5%*	43	*8.5%*	38	*7.5%*	32	*6.3%*	32	*6.3%*
7	32	*6.3%*	22	*4.3%*	15	*3.0%*	15	*3.0%*	15	*3.0%*
8	15	*3.0%*	10	*2.0%*	10	*2.0%*	9	*1.8%*	9	*1.8%*
9	30	*5.9%*	25	*4.9%*	18	*3.6%*	17	*3.4%*	17	*3.4%*

Table 2. Absolute number of groups (unique patterns) by pattern length for $n \in \{0,1,2\}$ and their percentage of the dataset. Patterns with four cells are exceptionally similar while patterns with six or eight cells are more diverse.

Popular User Patterns

The similarity analysis confirms that users select their unlock patterns from a rather limited pool of similar shapes. The groups for $n \leq 2$ with pattern length four deserve special attention, as they include more than half of the patterns in the dataset. The largest observed group is formed around the -shape (Table 1). It covers 17 different permutations.

Hence, attackers brute-forcing their way through all these 17 permutations of will get a hit for 56 of the dataset's 506 patterns – that is 11.1%. The second largest group comprises -forms and covers nine different patterns, whose occurrences account for almost 5% of our dataset. The group ranked three includes and covers 13 different patterns with 24 occurrences in our data (4.7%). Together, the five largest groups presented in Table 1 comprise 59 different patterns. Their occurrences account for roughly 29% of the dataset.

Regarding patterns of lengths other than four, we highlight the following results: The largest group contains patterns of length seven and was the sixth largest group in total. The group covers 13 -shapes in four different permutations. This means that over 40% of all length-seven-patterns can be derived from this single form. Most (8.4%) of the length-five-patterns were based on -forms, 23.3% of the length-nine-patterns formed -shapes. Patterns with six or eight cells showed most diversity.

Pattern Similarity

Figure 3 and Table 2 summarize the results: The total number of gestures shrinks from 506 to 350 when removing duplicates (i.e. $n = 0$). For a distance of 1, the total number of groups is 213. Hence, considering those patterns as duplicates that differ only by a single transformation already reduces the number of unique patterns by about 57%. If we set $n \leq 2$, only 169 groups are left, as determined by the greedy grouping algorithm. That is, the effective gesture space shrinks to one third of its size when we consider all gestures with a distance ≤ 2 as duplicates.

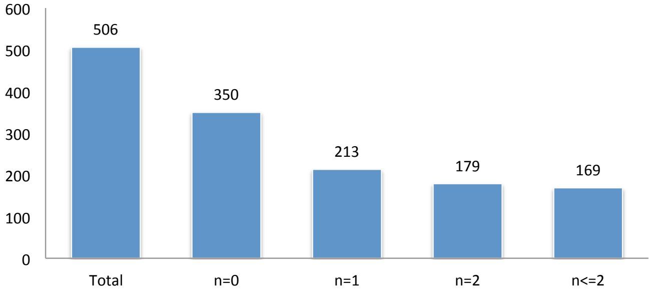


Figure 3. Total number of unique gestures (groups) for $n \in \{0,1,2\}$. Considering $n \leq 2$ reduces the effective gesture space by 66%.

If we focus on the more than 50% of participants who used patterns with a length of four, we find exceptionally high similarity: As shown in Table 2, the grouping algorithm here reduces the effective space from 262 patterns to 44 similarity groups (-83%) with $n \leq 2$.

Patterns with length ≥ 5 show less similarity than patterns of length four. However, the results in Table 2 imply that the ratio between total pattern count and number of groups does not shrink linearly with pattern length. Instead, it reaches a minimum for length six, where the number of effective patterns is reduced by just 33%. Length eight is close second with a reduction by 40%, and the unique pattern count for length nine is reduced by 43%, indicating very predictable selection strategies. For most pattern lengths, the unique pattern count is reduced by approximately 50%.

Implications

The analysis of the collected unlock patterns confirmed that users indeed follow predictable strategies when selecting such patterns. We confirmed that user-defined patterns are usually short (avg. 5 points), start on the top left of the matrix and end on the bottom right. In addition to these basic insights, the application of our proposed metric revealed that the most used patterns are made up of very similar shapes.

This was especially the case for short patterns. Here, over 80% could be traced back to the popular -shape. Overall, considering a similarity distance of two reduced the effective pattern space by approximately 66%. Thus, we conclude that pattern composition strategies are even more predictable than previous analyses of complexity aspects indicate [3, 28].

The insights of this analysis enable an attacker to cover more than a quarter of the unlock patterns by memorizing the center pattern of the five largest groups and applying simple geometrical transformations. As a consequence, we argue that novel concepts which increase the diversity in the effective pattern space need to be designed and evaluated.

We assume that applying the same explicit feedback to all users (e.g. pattern meters) may encourage them to select more cells, but it does not necessarily facilitate pattern diversity. Hence, the remainder of this paper investigates a more personalized solution which aims to subtly nudge users to select a more diverse set of patterns.

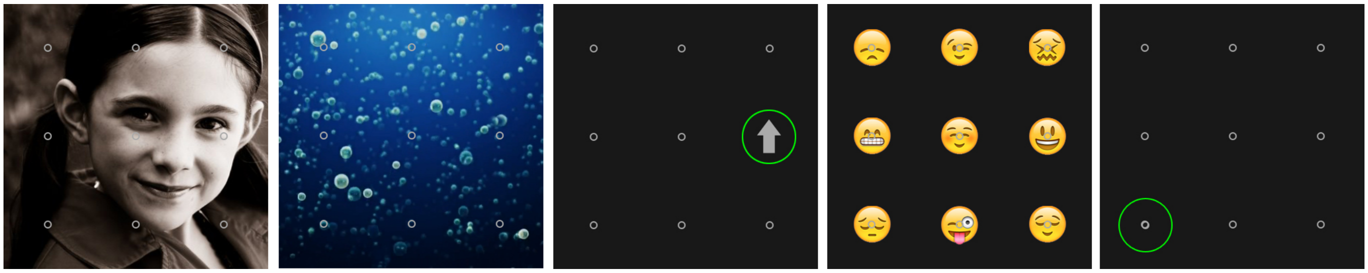


Figure 4. The five designs of the pre-study, from left to right: *BackgroundStatic*, *BackgroundDynamic*, *NodeArrow*, *NodeSmiley* and *Forced*.

INCREASING THE DIVERSITY OF UNLOCK PATTERNS

Previous work revealed that a promising approach to improve the user choice of graphical passwords is based on changing the design of a user interface [7, 27, 28]. For drawmetric concepts, the use of background images was specifically advocated [10, 20].

Concept Development and Pre-Study

The concepts were developed following a thorough design process, based on brainstorming sessions, preliminary user studies and multiple iterations. After the literature review and a first round of brainstorming, we came up with five different design candidates.

Design Candidates

The primary goal of the concepts was to influence the starting position of the patterns, as we assumed that different starting positions would most likely result in different shapes. Figure 4 illustrates the five design candidates.

For the *BackgroundStatic* scheme we chose a greyscaled image of a girl[5]. The image contained several hotspots, such as the eyes or the mouth which are likely to attract visual attention [8]. This scheme is close to the original idea by Dunphy and Yan [10]. The *BackgroundDynamic* condition was based on a looped video of air bubbles[6] floating in water. The bubbles had different sizes and moved from bottom to top with individual speed.

In addition, we designed two concepts which focused on modifying single cells. The *NodeArrow* concept highlighted one of the nodes in the grid with an upward pointing arrow. For the *NodeSmiley* scheme, we decided to utilize the smileys[7] of a messenger app. We assumed that the majority of the users would be familiar with the smileys and would thus be less confused by them. Furthermore, we expected that users would opt for selecting positive faces. Therefore, we distributed the smileys in such a way that the nodes that are usually rarely selected were emphasized by the happy smileys, whereas we positioned the sad and angry smileys towards the upper left corner.

Finally, we added a *Forced* condition, which predefined a starting point and highlighted it by a green circle.

Pre-study

The five designs were implemented using HTML5, CSS and JavaScript and evaluated in a lab-based user study. The study was conducted using a within-participants repeated-measures design. The dependent variable was the *starting point*, the independent variable was the *background image scheme* with five levels. The order of the schemes was randomized. We collected the participants' patterns as well as qualitative data via interviews and a questionnaire.

We invited ten participants (25–34 years), seven of them female, via email and personal contacts. We informed the participants that we tested new authentication concepts but we did not reveal the purpose of the background images. For each scheme, we then asked them to spontaneously choose an unlock pattern, using the web-based prototype. Each session took about 20 minutes and was recorded for later analysis. As an incentive for participating in our study, a 20 Euro shopping voucher was raffled among all participants.

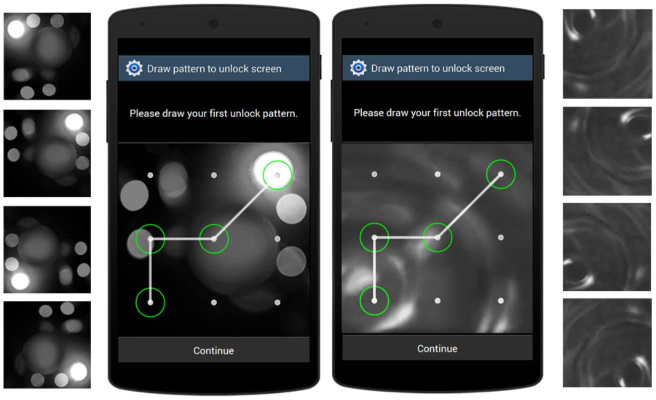
The results of the pre-study indicated that all five design candidates had high potential to influence the starting points. In the *BackgroundStatic* condition, the majority (80%) of the participants started their pattern above or underneath the girl's face, that is from node two, three or nine, in order to frame her face clockwise or counter-clockwise. In addition, participants generally avoided crossing the girl's face. The *BackgroundDynamic* scheme nudged participants to draw their pattern from bottom to top along with the floating bubbles. Half of the participants started in the lower left corner where the movement was strongest. Using the *NodeArrow* scheme, 90% of the users started from or above the node highlighted by the arrow and 80% continued their patterns in the direction the arrow pointed. Finally, the *NodeSmiley* scheme led to the most evenly distributed starting points compared to the other schemes. Participants used primarily happy smileys and smileys whose facial expression was more salient.

After the debriefing, the users' feedback revealed that the *NodeArrow* concept was by far the most unpopular scheme. Four participants perceived the arrow in the background as confusing since it "pretended a limited pattern choice". One participant stated that "it almost felt like a regulation". The same was true for the *Forced* starting point. However, two participants mentioned that a predetermined starting point might lead to more secure patterns. The *BackgroundDynamic* was more polarizing. Half of the users mentioned it was annoying as it was "too busy" or "made it hard to see the actual nodes".

[5] "love this face" by Jack Fussell, licensed under CC BY-NC-ND 2.0 (https://www.flickr.com/photos/travelingtribe/3844008664); accessed: 2016/08/25.

[6] "Air Bubbles Live Wallpaper", reproduced with permission by Eugene Pestov (https://www.youtube.com/watch?v=fLbhOILIEcs); accessed: 2016/08/25.

[7] Based on Apple Color Emoji font



Please draw your first unlock pattern.

Please draw your first unlock pattern.

Figure 5. The final Background-Pattern Concept: The *Static* scheme (left) and the *Dynamic* scheme (right). The smaller images illustrate the different states (rotations) of the respective scheme.

On the other hand, the rest of the participants gave mostly positive feedback: They stated that they liked the look of the animation and thought it was a more subtle approach than the *Arrow* scheme. *BackgroundStatic* and *NodeSmiley* were rated best. However, two participants said they did not really notice the background image and suggested more salient images. Overall, the *NodeSmiley* scheme was similarly popular. However, one participant mentioned to be generally annoyed by smileys, another user found the concept distracting.

Design Implications and Final Concept

The results and user feedback we received in the pre-study were very promising, as all designs had an impact on the pattern choice. However, the data did not immediately bring forth a clear favorite for further investigation, but rather suggested that there is a trade-off between the users' acceptance and the effectiveness of a scheme. Effectiveness was considered as the strength of the impact on the starting points that was qualitatively observed in the pre-study.

Due to their unpopularity, we decided to rule out the *NodeArrow* and the *Forced* schemes. The *NodeSmiley* scheme and the *BackgroundStatic* scheme were the most popular concepts and had shown a similar impact on pattern choice. The *BackgroundDynamic* scheme had received mixed feedback, but the hotspot in the lower left corner had a very strong impact on the starting points. For better comparability to related work, we finally decided to evaluate the *BackgroundStatic* and the *BackgroundDynamic* versions as they were close to the BDAS concept by Dunphy and Yan [10].

The final prototypes were improved based on the data gathered during the pre-study. As the face-based background image in the pre-study was considered to be not salient enough, we opted for images with stronger contrasts and clearer hotspots. On the other hand, the animation used in the pre-study was perceived disturbing due to the strong movement of the bubbles. We therefore looked for a calmer animation in the final prototype. Additionally, we decided to rule out the impact of colors by using grayscale images for both concepts.

Furthermore, a feasible concept would need to apply visual hotspots at random positions to induce a counterbalanced distribution of starting points. We opted for background images which were neutral enough to allow rotation in different directions. Each rotation should have the effect that the main hot spot would match one of the corner cells.

Figure 5 illustrates the final prototype comprising the used visualizations and the respective rotations. For the *Static* scheme, we decided to use an abstract image[8] showing several spotlights on a dark background, among which one spotlight is particularly bright. For the *Dynamic* version, we chose to utilize a looped video[9] of a drop falling into water. The water ripples spread concentrically, thus not suggesting a certain drawing direction.

Main User Study

We conducted an online user study to evaluate the impact of the concept on chosen starting points and on pattern similarity. The independent variable was the background image with two levels (*Static* and *Dynamic*). The study followed a between groups design. Each participant was randomly assigned to one specific rotation of one of the schemes.

Prototype

Since we wanted to collect a large set of unlock patterns, we decided to perform the main user study online. In order to avoid possible side effects arising from the utilization of different input techniques, we restricted the study to laptops and desktop computers using Mobile Detect. Please note that we had to opt for desktop computers as preliminary tests revealed that some popular mobile devices were not able to correctly play the required animations.

The implementation was based on the prototypes from the pre-study. Patterns could be drawn with the mouse while pressing the button, releasing it to finish. Entered unlock patterns had to comply with the Android pattern selection rules. Moreover, the interface design was enhanced in order to make the context more clear to the participants, and we implemented a progress bar indicating the different steps of the study. The used prototypes are shown in Figure 5. User interaction was logged using PHP and a MySQL database.

Procedure

As we assumed that some users might enter familiar standard patterns, we decided to collect two patterns per participant. However, participants were allowed to enter the same pattern twice. Each participant was presented with either the *Static* or the *Dynamic* scheme, rotated in one of the four directions according to a round-robin-wise assignment.

After they had entered two patterns, they were forwarded to a questionnaire inquiring aspects of the patterns they had selected, whether their choices were influenced by the background image and about the general perception of the concept. In addition, we collected basic demographic data. As an incentive for participating in our study, one E-book reader and five 10 Euro shopping vouchers were raffled among all participants.

[8] "Untitled" by I Love Trees, licensed under CC BY 2.0 (https://www.flickr.com/photos/ilovetrees/2770624201); accessed: 2016/08/25.

[9] "Water ripples" by ChoiceSlides bought on http://choiceslides.com/products/water-ripples; accessed: 2016/08/25.

Rank	First Pattern				Second Pattern			
	Top [A ∣ S ∣ D]	# Permutations [A ∣ S ∣ D]	# Occurrences [A ∣ S ∣ D]	% Dataset [A ∣ S ∣ D]	Top [A ∣ S ∣ D]	# Permutations [A ∣ S ∣ D]	# Occurrences [A ∣ S ∣ D]	% Dataset [A ∣ S ∣ D]
1	Z ∣ Z ∣ L	05 ∣ 08 ∣ 10	27 ∣ 13 ∣ 17	5.4 ∣ 5.4 ∣ 6.6	L ∣ L ∣ ⊓	06 ∣ 06 ∣ 08	21 ∣ 16 ∣ 13	4.2 ∣ 6.6 ∣ 5.1
2	L ∣ Z ∣ Z	11 ∣ 04 ∣ 04	27 ∣ 12 ∣ 15	5.4 ∣ 5.0 ∣ 5.9	L ∣ L ∣ Z	14 ∣ 09 ∣ 06	20 ∣ 11 ∣ 11	4.0 ∣ 4.6 ∣ 4.3
3	N ∣ ≡ ∣ N	08 ∣ 06 ∣ 11	20 ∣ 10 ∣ 11	4.0 ∣ 4.1 ∣ 4.3	Z ∣ Z ∣ Z	05 ∣ 03 ∣ 04	19 ∣ 09 ∣ 10	3.8 ∣ 3.8 ∣ 3.9
4	5 ∣ 5 ∣ 5	08 ∣ 02 ∣ 09	13 ∣ 07 ∣ 09	2.6 ∣ 2.9 ∣ 3.5	N ∣ ⊔ ∣ ⊏	09 ∣ 04 ∣ 07	19 ∣ 09 ∣ 07	3.8 ∣ 3.8 ∣ 2.7
5	5 ∣ M ∣ Z	03 ∣ 03 ∣ 05	11 ∣ 06 ∣ 05	2.2 ∣ 2.5 ∣ 2.0	⊔ ∣ N ∣ L	05 ∣ 06 ∣ 04	12 ∣ 08 ∣ 07	2.4 ∣ 3.3 ∣ 2.7

Table 3. The five largest groups for $n \leq 2$. For each condition, the table shows the most frequent (top) pattern of the group, the number of different patterns covered in each group, the accumulated absolute number of occurrences for all patterns in the group, and the number of occurrences as a ratio of the respective dataset. We report the whole set ($\equiv$ A), the *Static* set ($\equiv$ S) and the *Dynamic* set ($\equiv$ D) for both the first and second pattern set.

Participants

Participants were invited through a university-wide mailing list. After seven participants were removed from the data due to incomplete data sets, a total of 496 people contributed to the user study. Due to our recruiting method, the study was primarily distributed among younger people. The average age was 27 years (17–72 years). The gender was nearly counterbalanced with 51.2% of the participants being female. The majority (86%) of the participants reported to use a smartphone or a tablet on a daily basis, with Android being the most popular operating system (64%).

Results

This section presents the results of the background image study and compares them to the baseline data.

Basic Statistics

As each participant contributed two patterns, the final data set comprised two sets of 496 patterns each. Overall, the average pattern length of the first set was 5.95 points (SD = 1.58). Patterns of the *Dynamic* scheme were marginally shorter on average (5.79, SD = 1.52), compared to those of the *Static* scheme (6.12, SD = 1.62). Compared to the first set, patterns of the second set were slightly longer (6.08, SD = 1.59). Again, patterns of the *Static* scheme (6.18, SD = 1.61) were longer than those of the *Dynamic* scheme (5.98, SD = 1.57).

A one-way mixed ANOVA was performed to compare the collected data with the baseline data from the previous study. The results revealed that the difference in average pattern lengths of background image schemes and baseline patterns is significant ($F(2, 999) = 61.72$, $p < .0001$). Bonferroni-corrected post-hoc tests revealed that this difference was significant between both the patterns of the *Dynamic* scheme and the baseline ($p < .0001$), and between the patterns of the *Static* scheme and the baseline ($p < .0001$). However, there was no significant difference between the average pattern lengths of the two background image schemes ($p > .05$).

In the first pattern set, 47.8% of the users chose the upper left corner as their starting point, in the second set this point was prioritized by 36.1% of all users. In both sets, the lower right corner was the most common endpoint with 28.0% and 25.2%, respectively. The complexity analysis reveals that, compared to the baseline condition, more users decided to use overlapping nodes or knight moves. 17.9% of the patterns in the first set and 17.1% of the second set included overlaps. In addition, 8.1% of the first patterns and 10.1% of the second patterns comprised knight moves.

The Impact on Starting Points

In order to run statistical tests, we defined the respective node highlighted by each background image rotation as the *focused node* of this rotation. The focused node for both of the concepts was either node one (top left), node three (top right), node nine (bottom right) or node seven (bottom left). Furthermore, we defined the participants' choice of starting points as binary events: Either the starting point of a pattern does match the focused node of the respective rotation, or it does not.

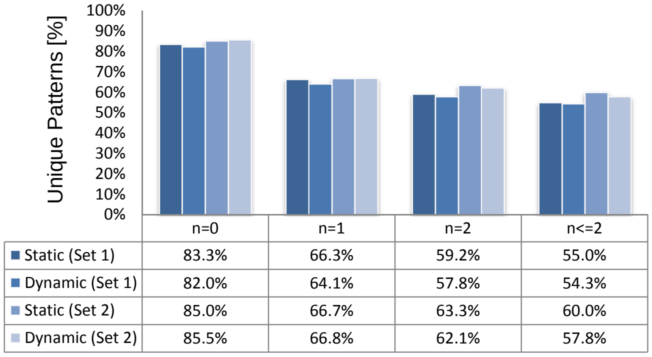
We analyzed this binary classification data using a two-tailed binomial test. The test calculates the probability with which the number of matches for the respective focused node would have been the same in the baseline condition. For the first pattern set, we found a significant association between the *Static* condition focusing the bottom left and the distribution of node seven with $p = .004$. For the second pattern set, the test indicated that the association between the *Static* scheme focusing the top right and the distribution of node three was significant ($p = .033$). For all other rotations, the results of the binomial test were not significant ($p > .05$).

Popular User Patterns

The analysis of popular unlock patterns reveals that Z was chosen most often. We found 17 instances of this exact pattern in the first set and 12 instances in the second pattern choice. Overall, all patterns based on the Z-shape cover 4.6% of the whole data set. In the baseline condition, the pattern was chosen by seven users which makes it the most popular pattern with length greater than four.

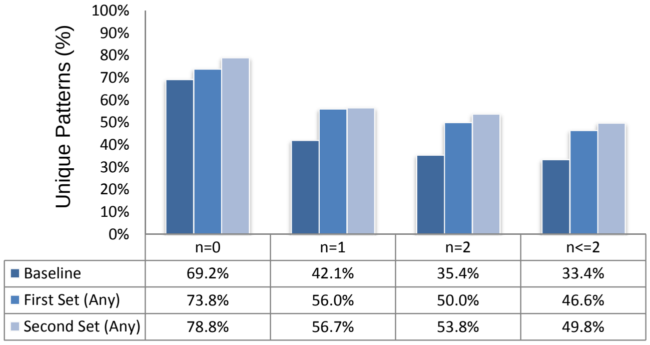
Table 3 illustrates the five largest pattern groups for $n \leq 2$. The data is split into the first and the second pattern input. In addition to the overall statistics, we report the popular patterns of the *Static* and *Dynamic* condition. The results indicate that all conditions lead to a very similar set of shapes which show only slight differences in form and distribution. In addition, the patterns show high similarity to the most popular patterns selected under standard conditions. Three groups of the background schemes comprise shapes (L, ⊏, 5) which we had already seen in the baseline condition. The other two popular shapes (N, ⊓) of the baseline study are found in a slightly modified longer version (N).

While the top five popular shapes in the baseline condition were all based on only four cells, the use of the background schemes resulted in longer patterns. In the first pattern set, only two of these popular shapes are made up of four cells, two groups are based on five-cell-patterns and one pattern



	n=0	n=1	n=2	n<=2
■ Static (Set 1)	83.3%	66.3%	59.2%	55.0%
■ Dynamic (Set 1)	82.0%	64.1%	57.8%	54.3%
■ Static (Set 2)	85.0%	66.7%	63.3%	60.0%
■ Dynamic (Set 2)	85.5%	66.8%	62.1%	57.8%

Figure 6. The portion of unique patterns for the *Static* and the *Dynamic* condition ($n \in \{0,1,2\}$). Both schemes lead to similar distributions. The second input tends to comprise a larger variety of patterns.



	n=0	n=1	n=2	n<=2
■ Baseline	69.2%	42.1%	35.4%	33.4%
■ First Set (Any)	73.8%	56.0%	50.0%	46.6%
■ Second Set (Any)	78.8%	56.7%	53.8%	49.8%

Figure 7. The portion of unique patterns in the first and the second set of the background condition and in the baseline condition ($n \in \{0,1,2\}$). While the effective pattern space in the baseline condition shrinks to 33%, the background condition is less affected.

group comprises length-seven-patterns. The second pattern choice resulted in even longer popular patterns. Only one group comprises length-four-patterns, three groups are based on five or seven cells and one popular shape has the maximum length nine (⊔⊓).

While the L⁚-shape covered over 11% of the whole data set in the baseline condition, none of the shapes selected with background images covers more than 6.6%. Overall, the five most popular shapes in the baseline study covered 29% of the data set. In contrast, shapes selected with background images seem more evenly distributed as the top five groups of the first set cover about 20% of the data, and respectively 18% for the second set. Finally, Table 3 indicates that the selection of longer shapes results in less permutations.

Pattern Similiarity

Figure 6 illustrates the unique pattern groups selected under the *Static* and *Dynamic* conditions ($n \in \{0,1,2\}$). The results indicate no significant difference between both concepts. However, user choice was more diverse when users selected their second pattern. While the effective password space (i.e. number of unique patterns) was reduced to 55% (*Static*) and 54% (*Dynamic*) in the first round, the second set still comprised 60% and 58% unique patterns, respecively after $n \leq 2$ transformations.

Figure 7 compares the effective pattern space of the first set and the second set of both background schemes with the data collected under standard conditions. The results reveal that users chose a more diverse set of patterns when background images were shown. While the baseline set comprises 31% duplicates ($n = 0$), the first pattern set selected with the background schemes included 26% duplicates. The second pattern set chosen under these conditions resulted in 21% duplicates. Applying up to two transformations ($n \leq 2$) reduces the effective pattern space of the baseline set to 33%. The effective pattern space of the first and the second set of the background schemes remain larger. While the first pattern choice shrinks to 47% of its size, half of the patterns chosen in the second set stay unique.

The analysis of the sets according to their length confirms the results of the baseline study. Patterns with a length of four, five and seven show the highest similarity, while patterns with six or eight cells comprise more diverse shapes. Considering up to two transformations shrinks the effective pattern space of the first length-four pattern set to 31% of its size, the second pattern set (length=4) is reduced to 33%. Under the baseline condition, patterns of the same size were reduced to 17%. In contrast, 67% of the firstly selected length-eight patterns remain unique. Length-eight patterns in the second set are even more diverse as the effective pattern space after two transformations is 79%. In the baseline study, 60% of the size-eight patterns remained unique (for $n \leq 2$).

User Feedback

The qualitative feedback revealed that dynamic background images are more conspicuous than static ones. While only 54% of the participants assigned to the *Static* scheme reported to have noticed the background image, 68% of the *Dynamic* scheme users had consciously perceived the effect. At the same time, 16% of the participants agreed or strongly agreed that the *Static* background image affected their first pattern choice and 6% indicated an impact for the second pattern. For the *Dynamic* condition, only 8% of the participants indicated an influence on their first pattern and 5% agreed with the impact on the second pattern.

We collected both positive and negative feedback on our two schemes as open-ended questions. Inductive coding comprising two coders resulted in 362 and 318 code instances with which we tagged answers, respectively. While the majority of participants referred their feedback to unlock patterns in general, only 18% of the code instances were related to our background image schemes. Participants who indicated positive effects mostly said that they wanted to start from the node highlighted by the background image ("*I chose my starting point to be at the brightest spot in the image*", "*The waves were mostly in the upper left corner. I started my pattern there.*"). In addition, more general impact was indicated. One participant said the "*[..] big shining points left and right gave [her] an idea of [the] pattern [she] could use*". Another user reported "*[she] simulated the movement of the background*". The facilitation of the selection process and the "*rethinking*" of the first idea for the pattern were also mentioned as positive psychological effects.

In contrast to such reinforcing effects, several users reported negative influences. Some avoided the visual elements of the background images in their pattern. For example, one participant reported "*[she] took the four dots where no white circles were [...]*". Some participants did not like the visual appearance in general and said the background image was "*not very inviting*". Moreover, several participants claimed that the background interfered with the nodes in the grid or with their pattern. In addition, some participants reported that the background was confusing, distracting or irritating.

Other answers revealed interesting insights on how users understood the concept. For example, one participant stated that she "*thought of [the image] as a plain background without further function*" and that "*the system could [have been] more obvious*". Another participant thought that the image was meant to "*simulate the reflection of a real display*", and a third one said that it was "*a nice idea*", but suggested "*to use a better picture*". When asked how often they would like the background images to be displayed in practice, 37% stated that they should be displayed every time they unlock their device. A third said that the background images should never be displayed, followed by 20% who wanted the background images to be displayed only during pattern selection.

DISCUSSION

The application of the similarity metric and the evaluation of background images in combination with Android unlock patterns revealed several interesting insights into user preferences and pattern similarity and shows the impact of the chosen evaluation strategy in experimentation. This section links the main results and discusses their implications.

Pattern Length is an Important Security Feature

The analysis of user-defined unlock patterns revealed that primarily short patterns are more similar. Under baseline conditions, 21% of all patterns with the length of four could be traced back to simple L-shapes. This is a serious security issue as over 50% of the participants in this group chose patterns of this length. Overall, the effective pattern space for such short patterns was reduced to 17% of its actual size. In the background image condition, only 20% of participants decided to base their pattern on four cells and chose a more diverse set. The effective space still was reduced to one third of its actual size.

In contrast, longer patterns generally comprised less similarity. We therefore conclude that pattern length is indeed a fundamental security measure. However, it is important to note that pattern length cannot be used as a *linear* security feature as the maximum diversity was found for six and eight cells but dropped for patterns using seven cells or the whole grid (length = 9). This aspect was found in both the baseline and the background concept and results from the users' preferences for specific shapes.

There is More to Strength than Length

In contrast to others, the proposed metric reflects human interest in geometric properties. Analyses with our metric thus consider similarity to other patterns as more important than properties of their composition. This revealed that only a certain sub-space of the theoretically possible set of patterns is actually used: For example, knowing a single pattern, namely the center of the largest group in our baseline dataset, one can deduce more than a tenth of our whole dataset by applying only up to two simple geometric transformations.

This knowledge renders patterns much more susceptible to informed guessing attacks than what is usually anticipated, when just looking at the theoretical number of combinations. In addition, the metric identifies popular shapes which are not necessarily found in the top ranks of unique patterns but still cluster a significant portion of the pattern space. We thus argue that pattern strength is better assessed by measures that consider human factors, such as geometric similarity, compared to measures that are purely based on obvious properties of pattern composition, such as their length.

Users Prefer a Small Set of Similar Shapes

The analysis confirmed that users indeed prefer short patterns based on simple shapes. The similarity analysis revealed that the selection of Android unlock patterns is even more restricted than previous work assumed. Previous studies already analyzed most frequently used patterns and reported the preferred use of L-shapes and Z-shapes. Our analysis showed that most patterns which seem unique at first glance are close relatives of these shapes.

Combined with the knowledge that over 40% of patterns started at the upper left node, this significantly reduces the theoretical security of the systems. Based on the results, we argue that proactive pattern meters need to consider such shapes in form of a dictionary when assessing the security of an entered pattern. This is important, for example, as a composition-based analysis of the ⌛-pattern would indicate the use of special moves (i.e. the overlap) and medium length. At the same time this pattern was the fifth popular pattern in the first set collected with background images. The same is true for the analysis of the Z-shape or the ᒉᑎ-pattern which would lead to high length scores. Nevertheless, such strength values would not reflect the practical security of these patterns as such popular shapes are more prone to dictionary attacks.

Some Users are Resistant to Background Images

Only slightly more than half of the participants assigned to the *Static* scheme and two-thirds of the participants assigned to the *Dynamic* scheme stated to have noticed the background images. Only a minority of the participants agreed that the background image did actually influence their pattern choice.

However, the evaluation of the background image concept indicated that users in this group selected a more diverse set of longer patterns. The effect was present for both the *Static* condition and the *Dynamic* condition. This indicates that the concept has an unconscious influence in general, which affects only a subset of users. We assume that the characteristics of the used background image significantly influence its perceptibility. In addition, the response to different image types might vary between users. We therefore conclude that the interplay between specific features of the visualization and the impact on particular user groups needs further investigation.

The Force of Habit is Hard to Break

Our analysis indicated that pattern selection habits can indeed be changed slightly with background images. However, it is presumably very difficult to entirely change the existing behavior, which is already heavily biased towards starting in the top left corner. In particular, it seems infeasible to nudge users to start at the opposite cell, meaning at their most common end point (i.e. bottom right). Our results confirmed that participants prefer Android unlock patterns that follow the general direction of the reading process (left to right; top to bottom). Moreover, considering the popularity of L∷-like patterns, squares and other known shapes, users tend to stick to what they already know.

Still, we found significant influences of static images on starting at the top right and bottom left corners, presumably because these cells are still more in line with starting point and direction of reading habits, compared to completely flipping this flow around (i.e. starting at bottom right). In addition, when background images were present, users chose (slightly) more complex modifications of known shapes (e.g. Z, L∵). We thus conclude that suitable concepts should not try to change existing behavior completely, but can find useful opportunities in aiming at slight changes within these general habits.

Repeated Measures Design is Unfeasible

During the development process, we evaluated the background schemes and three other designs in a lab-based pre-study. The outcome of the pre-study was very promising as all concepts had a significant impact on the chosen starting point. This impact was confirmed by the participants' qualitative feedback. In contrast to such promising results, the online study indicated only minor impact on pattern choice and only a small fraction of the users chose the intended starting points. We therefore conclude that the study design had a significant impact on the outcome. Since the lab study was designed following a repeated measures design, participants were exposed to different designs and adapted their behavior accordingly. In the online study, each participant was exposed to only one condition. We argue that repeated measure designs are unfeasible whenever behavior changes shall be observed.

LIMITATIONS

Our metric reflects that users do not create patterns at random, but rather guided by geometric properties. To achieve this, we currently employ a simple pattern distance, namely counting geometric transformations. However, we do not know how different geometric transformations compare to one another in terms of the users' *perceived similarity*. Hence, different transformations should possibly be counted with different relative weights, not all contributing to the total distance with the same weight, as we assumed here. Such weights should be determined by future studies.

Nevertheless, our greedy algorithm itself is flexible and can be adapted to work with any similarity metric. On the other hand, pattern groups found with our greedy approach may not match the global optimum. This is unavoidable for NP-hard problems. However, as our insights are derived from analyses of the largest groups and the most popular patterns, we expect them to be rather stable.

Considering our evaluation strategies, not all confounding factors could be ruled out. The baseline dataset was collected via Amazon Mechanical Turk, mostly completed by US citizens. In contrast, the image study was conducted in Europe. This could have had an impact on aspects of pattern selection. In addition, the system setup was slightly different as participants used indirect mouse input in the latter study.

However, both samples come from cultures with left-to-right reading/writing and all participants received the same instructions. As we hypothesize that these are the major influencing factors, we believe that they are indeed comparable with respect to our analyses. Still, it is important to note that our data might not be representative of other age groups and cultures. Finally, as the study task was artificial and required no memorization, the collected datasets might differ from actually used real-world patterns. Nevertheless, the collected data is consistent with the results reported by related work [13, 28].

CONCLUSION AND FUTURE WORK

In this work, we have presented a similarity metric for Android unlock patterns. The metric is based on simple geometric transformations (e.g. rotation) and identifies patterns which are based on similar shapes. We utilized this metric to analyze 506 user-defined unlock patterns. We were able to show that considering similarities of up to two simple transformations reduces the effective pattern space of unlock patterns in our set by approximately 66%. This result indicated that the effective pattern space of Android unlock patterns is significantly smaller in practical use. Therefore, we argue that solutions to make user-selections more diverse are required.

Consequently, we adapted the idea of Dunphy and Yan [10] and designed an unlock pattern concept based on background images. While the lab-based evaluation of the system was very promising, the field study involving 496 users yielded only small effects. Nevertheless, our analysis revealed that users selected a more diverse set of longer patterns when background images were present.

In summary, we claim that the results presented in this paper are important for both, researchers who study user behavior related to authentication systems, and users who want to strengthen their choice of unlock patterns.

Future work should extend the presented metric by considering more transformations and measuring similarity across patterns of different length. Furthermore, one could investigate whether overlaps and knight moves indeed increase security. Another area worth studying is the perceived similarity of patterns. A rotation might be more difficult to perform for human attackers than a translation or inversion. Finally, we would like to motivate other researchers to find countermeasures to avoid predictable pattern selection and to propose systems which have the potential to diversify pattern choice.

ACKNOWLEDGEMENTS

Special thanks go to Iris Maurer and Peter Arnold for their valuable help with the data collection and the data analysis.

REFERENCES

1. Florian Alt, Stefan Schneegass, Alireza Sahami Shirazi, Mariam Hassib, and Andreas Bulling. 2015. Graphical Passwords in the Wild: Understanding How Users Choose Pictures and Passwords in Image-based Authentication Schemes. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '15)*. ACM, New York, NY, USA, 316–322. DOI:http://dx.doi.org/10.1145/2785830.2785882

2. Panagiotis Andriotis, Theo Tryfonas, and George Oikonomou. 2014. Complexity Metrics and User Strength Perceptions of the Pattern-Lock Graphical Authentication Method. In *Human Aspects of Information Security, Privacy, and Trust*, Theo Tryfonas and Ioannis Askoxylakis (Eds.). Lecture Notes in Computer Science, Vol. 8533. Springer International Publishing, 115–126. DOI:http://dx.doi.org/10.1007/978-3-319-07620-1_11

3. Panagiotis Andriotis, Theo Tryfonas, George Oikonomou, and Can Yildiz. 2013. A Pilot Study on the Security of Pattern Screen-lock Methods and Soft Side Channel Attacks. In *Proceedings of the Sixth ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec '13)*. ACM, New York, NY, USA, 1–6. DOI: http://dx.doi.org/10.1145/2462096.2462098

4. Adam J. Aviv, Katherine Gibson, Evan Mossop, Matt Blaze, and Jonathan M. Smith. 2010. Smudge Attacks on Smartphone Touch Screens. In *Proceedings of the 4th USENIX Conference on Offensive Technologies (WOOT'10)*. USENIX Association, Berkeley, CA, USA, 1–7. https://www.usenix.org/legacy/event/woot10/tech/full_papers/Aviv.pdf

5. Robert Biddle, Sonia Chiasson, and P.C. Van Oorschot. 2012. Graphical Passwords: Learning from the First Twelve Years. *ACM Computing Surveys (CSUR)* 44, 4, Article 19 (Sept. 2012), 41 pages. DOI: http://dx.doi.org/10.1145/2333112.2333114

6. Judith Cederberg. 2013. *A course in modern geometries*. Springer Science & Business Media.

7. Sonia Chiasson, Alain Forget, Robert Biddle, and P.C. van Oorschot. 2009a. User interface design affects security: patterns in click-based graphical passwords. *International Journal of Information Security* 8, 6 (2009), 387–398. DOI: http://dx.doi.org/10.1007/s10207-009-0080-7

8. Sonia Chiasson, Alain Forget, Robert Biddle, and P. C. van Oorschot. 2009b. User Interface Design Affects Security: Patterns in Click-based Graphical Passwords. *Int. J. Inf. Secur.* 8, 6 (Oct. 2009), 387–398. DOI: http://dx.doi.org/10.1007/s10207-009-0080-7

9. V. Chvatal. 1979. A Greedy Heuristic for the Set-Covering Problem. *Mathematics of Operations Research* 4, 3 (1979), 233–235. DOI: http://dx.doi.org/10.1287/moor.4.3.233

10. Paul Dunphy and Jeff Yan. 2007. Do Background Images Improve "Draw a Secret" Graphical Passwords?. In *Proceedings of the 14th ACM Conference on Computer and Communications Security (CCS '07)*. ACM, New York, NY, USA, 36–47. DOI: http://dx.doi.org/10.1145/1315245.1315252

11. Dinei Florêncio, Cormac Herley, and Paul C. Van Oorschot. 2014. An Administrator's Guide to Internet Password Research. In *Proceedings of the 28th USENIX Conference on Large Installation System Administration (LISA'14)*. USENIX Association, Berkeley, CA, USA, 35–52. https://www.usenix.org/system/files/conference/lisa14/lisa14-paper-florencio.pdf

12. Haichang Gao, Wei Jia, Fei Ye, and Licheng Ma. 2013. A Survey on the Use of Graphical Passwords in Security. *Journal of Software* 8, 7 (Jul 2013), 1678–1698. DOI: http://dx.doi.org/10.4304/jsw.8.7.1678-1698

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

14. Ian Jermyn, Alain Mayer, Fabian Monrose, Michael K. Reiter, and Aviel D. Rubin. 1999. The Design and Analysis of Graphical Passwords. In *Proceedings of the 8th Conference on USENIX Security Symposium - Volume 8 (SSYM'99)*. USENIX Association, Berkeley, CA, USA, 1–14. https://www.usenix.org/legacy/events/sec99/full_papers/jermyn/jermyn.pdf

15. Ming Li and Paul MB Vitányi. 2008. *An Introduction to Kolmogorov Complexity and Its Applications* (3 ed.). Springer New York. DOI: http://dx.doi.org/10.1007/978-0-387-49820-1

16. Di Lin, Paul Dunphy, Patrick Olivier, and Jeff Yan. 2007. Graphical Passwords & Qualitative Spatial Relations. In *Proceedings of the 3rd Symposium on Usable Privacy and Security (SOUPS '07)*. ACM, New York, NY, USA, 161–162. DOI: http://dx.doi.org/10.1145/1280680.1280708

17. Wanli Ma, John Campbell, Dat Tran, and Dale Kleeman. 2010. Password Entropy and Password Quality. In *4th International Conference on Network and System Security (NSS'10)*. IEEE, 583–587. DOI: http://dx.doi.org/10.1109/NSS.2010.18

18. Michelle L. Mazurek, Saranga Komanduri, Timothy Vidas, Lujo Bauer, Nicolas Christin, Lorrie Faith Cranor, Patrick Gage Kelley, Richard Shay, and Blase Ur. 2013. Measuring Password Guessability for an Entire University. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security (CCS '13)*. ACM, New York, NY, USA, 173–186. DOI: http://dx.doi.org/10.1145/2508859.2516726

19. Deholo Nali and Julie Thorpe. 2004. *Analyzing User Choice in Graphical Passwords*. Technical Report. School of Information Technology and Engineering, University of Ottawa, Canada. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.85.998&rep=rep1&type=pdf

20. L. Y. Por, X. T. Lim, M. T. Su, and F. Kianoush. 2008. The Design and Implementation of Background Pass-Go Scheme Towards Security Threats. *WSEAS Transactions on Information Science & Applications* 5, 6 (June 2008), 943–952. http://www.wseas.us/e-library/transactions/information/2008/27-356.pdf

21. Karen Renaud and Antonella De Angeli. 2009. Visual Passwords: Cure-all or Snake-oil? *Commun. ACM* 52, 12 (Dec. 2009), 135–140. DOI: http://dx.doi.org/10.1145/1610252.1610287

22. Stuart Schechter, Cormac Herley, and Michael Mitzenmacher. 2010. Popularity is Everything: A new approach to protecting passwords from statistical-guessing attacks. In *The 5th USENIX Workshop on Hot Topics in Security (HotSec'10)*. USENIX Association, Berkeley, CA, USA, 1–8. http://research.microsoft.com/apps/pubs/default.aspx?id=132859

23. Hossein Siadati and Nasir Memon. 2015. *Fortifying Android Patterns using Persuasive Security Framework*. Technical Report. New York University. http://isis.poly.edu/~hossein/publications/hossein-qualexam.pdf

24. Youngbae Song, Geumhwan Cho, Seongyeol Oh, Hyoungshick Kim, and Jun Ho Huh. 2015. On the Effectiveness of Pattern Lock Strength Meters: Measuring the Strength of Real World Pattern Locks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2343–2352. DOI: http://dx.doi.org/10.1145/2702123.2702365

25. Chen Sun, Yang Wang, and Jun Zheng. 2014. Dissecting pattern unlock: The effect of pattern strength meter on pattern selection. *Journal of Information Security and Applications* 19, 4-5 (2014), 308 – 320. DOI: http://dx.doi.org/10.1016/j.jisa.2014.10.009

26. Hai Tao and Carlisle Adams. 2008. Pass-Go: A Proposal to Improve the Usability of Graphical Passwords. *International Journal of Network Security* 7, 2 (Sep 2008), 273–292. http://ijns.jalaxy.com.tw/contents/ijns-v7-n2/ijns-2008-v7-n2-p273-292.pdf

27. Julie Thorpe, Muath Al-Badawi, Brent MacRae, and Amirali Salehi-Abari. 2014. The Presentation Effect on Graphical Passwords. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2947–2950. DOI: http://dx.doi.org/10.1145/2556288.2557212

28. Sebastian Uellenbeck, Markus Dürmuth, Christopher Wolf, and Thorsten Holz. 2013. Quantifying the Security of Graphical Passwords: The Case of Android Unlock Patterns. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security (CCS '13)*. ACM, New York, NY, USA, 161–172. DOI: http://dx.doi.org/10.1145/2508859.2516700

29. Christopher Varenhorst, Max Van Kleek, and Larry Rudolph. 2004. *Passdoodles: A lightweight authentication method*. Technical Report. Massachusetts Institute of Technology. http://people.csail.mit.edu/emax/public_html/papers/varenhorst.pdf

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

31. Emanuel von Zezschwitz, Paul Dunphy, and Alexander De Luca. 2013. Patterns in the Wild: A Field Study of the Usability of Pattern and Pin-based Authentication on Mobile Devices. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '13)*. ACM, New York, NY, USA, 261–270. DOI: http://dx.doi.org/10.1145/2493190.2493231

32. Matt Weir, Sudhir Aggarwal, Breno de Medeiros, and Bill Glodek. 2009. Password Cracking Using Probabilistic Context-Free Grammars. In *30th IEEE Symposium on Security and Privacy*. 391–405. DOI: http://dx.doi.org/10.1109/SP.2009.8