N. Ibrahim and H. Sellahewa, "Touch gesture-based authentication: A security analysis of Pattern Unlock," presented at the 2017 International Conference on Identity, Security and Behavior Analysis, ISBA 2017, New Delhi; India, 2017. DOI: <u>10.1109/ISBA.2017.7947711</u>

© 2017 IEEE.Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Touch Gesture-Based Authentication: A Security Analysis of Pattern Unlock

Nasiru Ibrahim and Harin Sellahewa Department of Applied Computing University of Buckingham, Buckingham, United Kingdom {nasiru.ibrahim; harin.sellahewa}@buckingham.ac.uk

Abstract

Pattern Unlock is one in a family of graphical passwords used on smart mobile devices. They were proposed as an alternative to PIN and Password, as a result of a study [2] that showed better human memorability of visuals words than textual. Researchers have made several attempts to analyse and quantify the security of Pattern Unlock. However, only a few studies have been done on human characteristics and the correlation to choice of patterns.

In this paper, we investigated the Android Pattern Unlock authentication and presents a comprehensive analysis of user pattern characteristics and strength of patterns generated by users. We extended the analysis by looking in-depth into gender and also different Android operated devices were looked into to get the lock-out and wait time, to lay the foundations of a practical security framework.

Our results (125 participants) indicated how users weakly choose their patterns by simply embedding and reusing patterns that are alphabetic or numeric-like, and having a strength score less than 27bits of entropy. The results highlighted the need to make users more aware and conscious when choosing their authentication patterns.

1. Introduction

Smartphones and tablets have become an essential part of our everyday life. The main advantage of these devices is that they are not bounded by time or place. They are designed to be portable, convenient, entertaining, and support people in their daily activities. They could be used for communication, socialising, and most recently m-banking, m-payments and m-health, among numerous other purposes.

Mobile devices are ubiquitous and in most cases, are the gateway to accessing numerous user applications. The activities carried out on the devices result in storage of personal and highly sensitive data about their users and people associated with them. Thus, making it essential to restrict access to the devices by implementing a reliable and convenient access control to prevent unauthorised access and misuse. Several schemes have been used such as; PIN, Password, Pattern and most recently, Biometrics.

The devices face numerous challenges ranging from memory shortage to external threats. Pfleeger and Pfleeger in [13] categorised threats to a system into human and nonhuman. Nonhuman threats are caused by phenomena that are out of human control such as; component failure (e.g. processor chip). However, human threats are directed to cause harm to the system and the user (in some cases). Such threats include; data theft or corruption, misuse of device and unauthorised access.

Traditional methods such as; PIN and Password are the widely used authentication methods and they do not seem to go away soon. The extent to which they secure devices have been significantly studied [1, 10, 11, 12]. Two of the measures of assessing the security are a) practical assessment based on the resistance to threats and b) theoretical assessment based on the measure of password space and randomness. No system is secured if an imposter knows the password, else, they would have to keep guessing until they find the correct password and the point is to make it as difficult as possible in terms of time and effort.

Pattern Unlock were introduced as an alternative to PIN and Password because they are graphical, more usable and memorable [4]. In Pattern Unlock, the user is presented with a 3x3 grid and the password is how the nodes are connected. Grid arrangements and unlock conditions by Android restrict access to some nodes, for instance, the top-left node cannot be connected to the bottom-right node unless via the middle node. Further, this affects the number of usable patterns based on

lengths. For example, a pattern of length 5 will only have 7,152 valid patterns available. Even when the length is 9, the total patterns available is 140,704. These numbers are much smaller compared with 6-digit PIN (10^6) and 5-digit Password (36^5) .

We conducted and present in this paper, a comprehensive study on the security of pattern unlock authentication on both theoretical and practical measurement levels. Further, the characteristics of the patterns were evaluated in terms of actual nodes distribution, lengths and pattern space to understand what is viable for the users. The correlation between users and the patterns they choose was also analysed. For instance, are the patterns chosen based on association with something personal (name, alphabet or number)?

The paper contributes by presenting an in-depth analysis of user characteristics (based on gender). Further, the paper presents the lock-out time of different Android operated devices which forms the basis for a framework to measure the probability of guessing correct patterns in a certain period.

The rest of this paper is organised as follows: Section 2 presents the background and related work on primarily graphical passwords. Section 3 presents the methodology used in the study. The results and discussions are presented in Sections 4 & 5. Conclusion and future work are presented in Section 6.

2. Background and Related Work

Authentication is the method of verifying the identity of a person before granting access to a system, i.e. verifying that the person is actually who they claim to be. There are three factors in authentication. 1) Knowledge factor, 2) possession factor and 3) inherence factor [9]. The knowledge factor authenticates based on something the user knows, such as PIN, Password, and Pattern. The possession factor uses something the user has to authenticate, such as one-time password/PIN generator token or even a mobile device. The inherence factor involves the use of something the user is, i.e. using users physiological and behavioural modalities, such as fingerprints, iris, voice, face, typing rhythm and signature etc.

PINs and Passwords have for long been the popular and standard way of authentication on mobile devices and are pretty straightforward. However, their use in protecting devices is known to impose a burden on users. The burden to create and remember strong and complex passwords [12]. They make users handle the burden by writing the passwords/PINs down somewhere or choosing weak Passwords/PINs that could easily be guessed. Also, this authentication method has the limitations of susceptibility to shoulder surfing, smudge and intelligent guessing attacks.

To provide an alternative, Blonder [2] first proposed the graphical passwords based on a psychological study carried out by Calkins [3] that showed better human memory for visual words than textual words. Blonders proposal was to present an image to the user and then select certain regions based on a predefined order to grant access. The security of the proposed method was not measured in that study, hence prompting an extension. Jermyn *et al.* [8] expanded on the work and proposed Draw A Secret (DAS). DAS has a 2-Dimensional grid of M x M size and a pattern is drawn to authenticate. The coordinates at which the pattern strokes passes are recorded as the secret password. The method has an infinite password space as there is no limitation to what could be allowed. In terms of security, they argued that DAS password spaces improves the information content as users do not choose their passwords in a uniform way and attackers will have an infinite distribution of choices (for a brute force attack).

Dunphy and Yan [7] highlighted a problem with [8] in users choosing weak passwords that are susceptible to graphical dictionary attack. They proposed an alternative approach, Background Draw-A-Secret (BDAS). In BDAS, the user will first choose an image to use as the background of the grid, then draws a pattern on the grid to authenticate. They wanted to see if the background image would improve the security and memorability of DAS. They recruited 21 participants for the experiment that includes 6 female, aged 18-50 from a mixture of technical and non-technical backgrounds. The experiment was conducted using a Personal Digital Assistant (PDA). They concluded that people using BDAS created more complex passwords than DAS and background images improved the security and memorability. However, these approaches have their security flaws such as; shoulder surfing and smudge attacks. A most recent implementation of the BDAS is an alternative provided on Windows 8 PC.

The extensive usage of Passwords has resulted in many studies analyse its strength and characteristics. Researches have shown that users choose passwords from a small subset that could be cracked easily by dictionary or intelligent guessing attacks. To encourage users to choose strong passwords, password policies have been recommended and widely deployed in many respects especially websites, and some password protected systems and devices. Data-Genetics [6] conducted a study with about 3.4million PINs and found the most common to be 1234, followed by 1111 and 0000 with 11%, 6.016%, and 1.881% frequencies respectively. Similarly, Splash-Data [14] compiled more than 2million leaked passwords in 2015 and reported that '123456' and 'password' are the top two most commonly used passwords. These passwords have been top two for the last five years, suggesting that people are either not aware or just oblivious about password security.

On Pattern Unlock, Uellenbeck *et al.* [16] were of the first to study and quantify the strength of the authentication. They studied 584 participants patterns (by user study - user's patterns and pen and paper survey - statistical information of user's patterns), to which they used 2-grams and 3-grams to construct a Markov Process Model that generates likely patterns. They used the generated patterns to attack as a measure of strength and obtained the average number of guesses to find the correct password. With 10 guesses, they were able to guess 4% and 7% of defensive (user's own pattern) and offensive (another users pattern) respectively. Further increasing the guesses to 30, they successfully guessed 9% and 19% of the defensive and offensive patterns respectively. They also estimated the partial guessing entropy (i.e. the breaking of just a percentage of passwords) and obtained 9.10bits with breaking 20% of passwords.

Adam *et al.* [1] modified Uellenbecks guessing algorithm to accommodate some characteristics obtained from thier prior analysis. They found 3-gram to be more appropriate and used Markov Process Model to generate like patterns. Their guessing algorithm successfully cracked 16.7% (3x3) and 19% (4x4) patterns within 20 guesses but after 50000 guesses, the algorithm was able to correctly guess 96.2% (3x3) and 66.7% (4x4) patterns. Their methodology was similar to Uellenbeck's in that they had a pen and paper study and a self-reporting survey. They recruited 80 participants for the first study who were asked to generate 3 defensive (participant's pattern) and 10 offensive (other participant's) patterns. They collected 380 offensive, 114 defensive 3x3 patterns and 385 offensive and 119 defensive 4x4 patterns. The authors evaluated whether increasing the grid size will increase the security of human generated patterns. They found some improvements but also argue that user's pattern will remain trivial and easily guessable.

Sun *et al.* [15] used the idea of visual complexity, which includes the physical characteristics of the pattern such as; number of nodes, number of segments, intersections and overlaps. They modified the entropy equation to accommodate the characteristics and measure the strength of the patterns. Using modified entropy equation, they feedback to users the strengths of their patterns which prompted them to choose more complex and strong patterns.

3. Methodology

We collected data from 125 people, aged between 18 and 56 who were members of the University from various backgrounds. They were selected on voluntary terms and recruitment methods were; wide email announcement and face to face contact. Of the 125 people, 42 were female, 11 left handed, all with touch screen experience.

An Android application was developed that runs on a Sony Xperia Z3 with Android 6.0.1 OS and 5.2 touchscreen display. Participants were asked to draw predefined patterns first, so as to be comfortable with the 3x3 grid and Android Pattern Unlock rules some might not have used it before. Thereafter, they were asked to create patterns that they could possibly use in securing their devices (and repeat 7 folds).

For simplicity of reference, we refer to the grid nodes using letters (A to I) as shown in Figure 1.



Figure 1. Grid Nodes Identification

4. Experimental Results

This section reports the findings from the various evaluations conducted on the collected data sets. The findings include; node and length distributions, pattern frequency, association and strength evaluations.

4.1. Start Node Distribution

We found and analysed all 389,112 possible Android Unlock Patterns and obtained the number of patterns usable from each node of the grid. The pattern space reduce immensely from 389,112 depending on where the pattern is started from. For instance, starting a pattern from node A will give 38.042 usable patterns. Table 1 outlines the number of possible patterns from each node.

Node	Patterns
А	38,042
В	43,176
С	38,042
D	43,176
E	64,240
F	43,176
G	38,042
Н	43,176
Ι	38,042
Total	389,112

	Table 1. Start N	Node and	Number o	of Possible	Patterns
--	------------------	----------	----------	-------------	----------

One could assume that most users would start their patterns from node E (since it gives more possibilities), but surprisingly, we found that 48.8% of users start their patterns at node A (with least pattern space) and 37.6% finish their patterns at node I. The findings were compared with existing studies on start nodes where similar results were reported. Table. 2 shows the comparison.

Table 2. Start Would Comparison					
Node	Uellenbeck[16]	Adam[1]	Sun[10]	Our	
				Result	
A	43%	42.3%	44%	48.8%	
В	6%	8.9%	9%	7.2%	
C	9%	12.5%	15%	17.6%	
D	6%	4.1%	6%	4.8%	
E	2%	5.0%	4%	4.1%	
F	4%	2.5%	2%	1.7%	
G	18%	20.0%	14%	13.6%	
Н	4%	2.5%	2%	0.0%	
Ι	8%	2.3%	4%	2.4%	

Table 2. Start Nodes Comparison

4.2. Length Distribution

In the same way that starting nodes impact the pattern space, so as the pattern lengths. The longer a password is, the larger the password space and consequently the higher the potential security. It was found in our data set that users used an average length of 6.8 nodes (± 1.42 with a mode of 7). Table. 3 outlines the number of patterns from the various lengths and the frequency of usage by users.

Length	Frequency	Patterns
4	4.8%	1,624
5	16.0%	7,152
6	19.2%	26,016
7	32.0%	72,912
8	10.4%	140,704
9	17.6%	140,704

Table 3. Lengths, Frequency and Number of Patterns

The most used length (7) has the second largest pattern space of 72,912 (with 8 & 9 having the highest 140,704). For the least used length, it is quite expected to be 4 as many would deem that to be insecure and hence the low usage. Surprisingly, users do not use length 8 that much (second least used) even-though it shares the largest patterns space with 9. Similar analysis was performed by [10] but the results differ quite significantly. They obtained an average pattern length of 5.40. A comparison of the distribution is shown in Table 4.

Length	Our Result	Lorge [10]
4	4.8%	36.0
5	16.0%	23.0
6	19.2%	12.0
7	32.0%	12.0
8	10.4%	4.0
9	17.6%	12.0

Table 4. Length Distribution Result Comparison

4.3. Frequent Patterns and Association Elements

Most frequent and association patterns provide a starting place for adversaries when trying to guess a password. Adversaries could try the patterns first as they are more probable, in the hope to guess the correct pattern easily. We found that 'ABCEGHI' was the most common pattern with a frequency of 8.0%, followed by 'CBADEFIHG' with 2.4% frequency. Figure 2 shows the 2 patterns.



Figure 2. ABCEGHI & CBADEFIHG

People often use information that are personal to them as their PINs or Passwords such as; names, sports clubs, anniversary dates etc., largely due to convenience and easy memorability [10]. But, the consequence of that is weakness in device security as the Password could be easily guessed.

The patterns we collected were analysed for similar strategies. In this case, it was not expected to have full dates or names but only single characters and numbers (as the authentication is single touch). Therefore, it would be expected to have patterns that correspond to alphabets and numbers (0-9). Upon analysis, it was found that 39 (32.5%) of patterns correspond to an English alphabet and 13 (10.83%) to a number. Some of the alphabets and numbers are; M, N, Z, 3, 7, 9 in various variations. It is not clear if the patterns have to do with something personal to the users. An in-depth analysis has to be conducted to ascertain reasons for the pattern choice. In total, 43.33% of the patterns were associated with either an alphabet or a number. Some of the common patterns and the association elements (alphabets) were reported in [10].

4.4. Strength Evaluation

Strength could be seen as a measure of the effectiveness of a system in resisting various attacks. Password strength is measured based on characters composition, complexity, and unpredictability and also measured on both theoretical and practical measurement levels. Practical analysis of strength is often done with regards to the time it takes to crack the password using various attacks, such as; shoulder surfing, smudge, brute force and dictionary attacks. The longer it takes, the more secure the Password is. On the other hand, the theoretical measure is based on statistics and probability models and distribution of passwords and password spaces.

Our approach to quantifying the strength of Patterns is to use Entropy (a statistical model) and simulate attacks with contributions from the various evaluations earlier conducted. Entropy is one of the standard metrics used in measuring the strength

of password distribution and individual passwords. It is a probability-based measure and measures the average uncertainty in predicting the outcome of a random variable [5]. It provides a way of calculating and comparing the potential strengths of patterns based on their lengths and composing nodes. Maximum entropy is obtained when the random distribution is uniform and composing items are equi-probable.

It is important to highlight that there are two ways to calculate entropy, (i) entropy of all possible outcomes (patterns) and (ii) entropy of particular outcome (pattern). Equations (1) and (2) shows both mathematical representations.

$$H(X) = \log_2(total number of patterns); \tag{1}$$

$$H(X) = -\sum P(x_i) * \log(P(x_i));$$
⁽²⁾

where i is the node position, $p(x_i)$ is the probability of the node at position i.

We utilised these methods in measuring the overall strength of all the patterns and individual patterns. Using equation (1), we calculated the strength of the entire 389,112 patterns and obtained 18.57bits. It was shown in subsection 4.1 and 4.2 that starting a pattern in different nodes or using different lengths result in variable usable patterns. The strength obtained from both distributions was between 10.66bits and 15.97bits (inclusive). Figure 3 and Figure 4 show both strengths.



Figure 3. Strength Scores of Starting Nodes



Figure 4. Strength Scores of Pattern Lengths

Combining start nodes with lengths gives a microscopic view of the patterns. Table 5 shows the number of patterns that are usable when each node is chosen as the starting and having variable length (4-9).

It is quite explicit (from Table 5) that users have no more than 23,280 patterns to choose from (best case scenario) at any point and that is about 1/17 of 389,112. Looking at the table in light of the start node bias, node A has a maximum of

Nodes	4	5	6	7	8	9
A	154	684	2516	7104	13792	13792
B	188	816	2926	8118	15564	15564
С	154	684	2516	7104	13792	13792
D	188	816	2926	8118	15564	15564
E	256	1152	4248	12024	23280	23280
F	188	816	2926	8118	15564	15564
G	154	684	2516	7104	13792	13792
H	188	816	2926	8118	15564	15564
Ι	154	684	2516	7104	13792	13792

Table 5. Nodes and Length Distribution

13,792 valid patterns. Furthermore, the strength of the distribution (Table 5) was measured using equation (1) and found to be between 7.27bits 14.507bits (inclusive). The strength and the pattern space gives an indication of how weak (in terms of pattern space and strength) the Pattern Unlock could be.

For the individual patterns, we first calculated the conditional probability (due to nodes dependencies) of the individual nodes and computed the entropy using equation (2). We obtained strength scores between 6.34bits - 34.24bits (prior strength). The scores were obtained under the assumption that the probability of selecting all nodes is uniform. Theoretically, the score and the assumption are valid, however, earlier evaluations such as; starting node, show that choosing the nodes are far from being uniform. Hence, the strength should be re-evaluated to reflect the evaluation results.

We used the concept of Bayes' rule to update our prior belief in the face of the evidence (evaluation results) and obtained posterior probabilities with which we calculated the posterior strength. The posterior probability used was obtained from the collected patterns. The transition probability of selecting a node based on its previous node or nodes was calculated. For instance, the probability of selecting the second node based on the first or the third based on the first and second nodes. The strength scores thereafter obtained was between 2.858bits - 26.140bits. The mean and standard deviation for the prior and posterior strengths are outlined in Table 6.

Table 6.	Pattern Stree	ngth in Entropy (M	lean and Standard De	viation)
		Prior (bits)	Posterior (bits)	

	Prior (bits)	Posterior (bits)
Strength	21.05 ± 7.70	12.79 ± 5.63

It should be noted that there could be other ways of describing pattern strength. Sun *et al.* [15] used a variation of the entropy to calculate the strength on all 389,112 valid patterns. They obtained scores between 6.340bits - 46.807bits. Some of our results were compared with the strength obtained by Sun *et al.* [15]. Table 7 shows the comparison of the strengths.

Table 7. Strength Score Comparison						
Pottorn	Strength Scores (bits)					
1 attern	Sun [12]	Our Prior	Our Posterior			
Z	19.40	22.88	14.63			
7	6.34	11.24	9.35			
Ν	19.40	22.88	12.19			

4.5. Gender Evaluation

To get an in-depth insight into user characteristics and correlations, we present evaluations based on gender. The data set was divided into male and female and same evaluations were ran. We found small difference between the male and female results.

For the starting node, 51.8% of male patterns were found to have started at node A and 43.4% finished at node I. While 42.9% of female patterns also started at node A and 26.2% finished at node I. See Table 8 for more details.

It is quite surprising to see from Table 8 that both males and females did not start their patterns at node H even with it having the second highest pattern space.

The male and female participants chose their patterns with variable lengths. Males had patterns with an average length of 6.9 (\pm 1.42) while females had an average length of 6.59 (\pm 1.45). The male participants utilised more of the higher lengths

Nodo	M	Male		nale
INDUE	Start	End	Start	End
Α	51.8%	1.2%	42.9%	2.4%
В	7.2%	8.4%	7.1%	0.0%
С	15.7%	10.8%	21.4%	21.4%
D	7.2%	2.4%	0.0%	4.8%
E	3.6%	2.4%	4.8%	2.4%
F	1.2%	8.4%	2.4%	14.3%
G	12.0%	14.5%	16.7%	16.7%
Н	0.0%	8.4%	0.0	11.9%
Ι	1.2%	43.4%	4.8%	26.2%

Table 8. Start and End Nodes (Gender)

(5-9), with only 3.6% having a length 4. On the other hand, females utilised more of lengths 5-7. Further, lengths 4 and 8 have the lowest frequency of usage. Table 9 highlights the distribution of the lengths.

Length	Male	Female
4	3.6%	7.1%
5	14.5%	19.0%
6	20.5%	16.7%
7	30.1%	35.7%
8	12.0%	7.1%
9	19.3%	14.3%

Table 9. Pattern Length Distribution (Gender)

To summarise subsection 4.5, male participants were found to have used patterns which are more common than the female participants. The top 8 male patterns were repeated more than ones and had a combined frequency of 21 (25.2%) while only 2 of the female patterns were repeated more than ones with a combined frequency of 6 (14.3%).

The pattern strength for both gender ranges from 6.34bits to 34.24bits (prior) and 0.39bits 26.06bits (posterior). The males had an average prior strength of 21.5bits (\pm 7.7bits) and 12.4bits (\pm 5.7bits) posterior strength. The females had average strength of 20.0bits (\pm 8.0bits) prior and 10.2bits (\pm 4.9) posterior strengths.

5. Discussion

We have studied Pattern Unlock authentication by analysing user characteristics and correlation between users and their choice of patterns. It was shown that the conditions by Android (pattern policy) makes the starting node and length crucial as it determines the number of valid moves and patterns. Our study in conformity with existing studies revealed that majority of users' start their patterns from the top left node and end their patterns at the bottom right node. Starting at the top left reduces the pattern space to just 38,042 (fewest). Three possible reasons for the starting point bias could be made, first, due to the writing sequence i.e. English-language writing style from left-right, second, due to convenience and arrangement of the nodes and last, the environment of the participants (UK). The circumstance that all the participants are studying in the UK and proficient in English could be a factor for the bias. Additionally, modifying the grid arrangements by [16] did not significantly improve things as users simply used the node to the right of top-left. In the dataset, there were 31 participants who were Europeans, 18 Asians, 50 Africans/Caribbean, 23 Middle Eastern and 3 Americans. Hence, assessing the role language plays in choosing patterns would be worthwhile.

The average pattern length of the entire data set was 6.7 (with mode of 7). Length 7 has the second highest pattern space (72,912) after 8 and 9 (140,704). Unsurprisingly, length 4 was the least used and that could be attributed to the users security consciousness. However, length 8 is the second least used and there is not a possible explanation for that, perhaps it warrants further investigation. Lengths 6, 7, 8 and 9 were used a total of 79.2% with a cumulative pattern space of 380,336.

Looking into the gender data, the male participants used an average length of 6.9 while the female used 6.59. The frequency of usage varies significantly as the male tended to use lengths 6 9 (with exception of 8 which was the second least used). However, the female most used lengths were between 5 7 with 8 & 9 being the least and second least used respectively. This shows that female participants chose patterns that are relatively shorter than their male counterparts.

Attomate			
Attempts	Android	Android 6.0+	Android 6.0+
	5.0+	(Xperia Z3)	(Samsung
			Galaxy S6)
1 - 4	None	None	None
5	30 seconds	30 seconds	30 seconds
6 - 9	None	None	None
10	30 seconds	30 seconds	1 minute
11 - 14	None	30 seconds	5, 10, 30, 60
			minutes
15	30 seconds	30 seconds	Wipe data if Auto Factory Reset is on else wait 60 minutes and same for every further attempt
16 - ~	Repeat last 2	30 seconds and for ev- ery further attempt	

Table 10. Android OS Waiting and Lock-Out Time Comparison (by device)

For easy memorability, most users chose numbers or alphabets as their patterns. The study revealed that 26.67% and 11.67% of the entire patterns correspond to an alphabet and number respectively. In the entire data set, 43.33% of the patterns were either alphabets or numbers. When users were asked whether they deliberately based their patterns on something personal, alphabet, number or just random, 32.70% of them said they based their patterns on a number/alphabet, 23.08% on something personal and 34.62% on something random. However, some patterns that users thought were random turned out to be associated with an alphabet or number - showing that some users did not deliberately choose letters or numbers.

Of the male (37) and female (15) participants that gave what they based their patterns on, 35.14% of the male patterns were based on a number or alphabet, 24.32% on something personal, 27.03% on something random while 26.67% of the female patterns were based on a number or alphabet, 20.00% on something personal, 46.67% of something random. This information gives an informed consent of the way not just all participants chose their patterns but specifically how males and females chose their patterns. However, at this juncture, no solid/concrete conclusion can be drawn as further investigation could reveal pattern relation to other entities (e.g. language, nationality or ethnicity).

This study has highlighted the need for users to be aware of not only the kinds of patterns to avoid but also how to choose potentially more secure patterns. More specific analysis could be made for better insights. For instance; analyses based on users handedness, age groups and possibly language.

5.1. Practical Security Framework

The theoretical measure of the amount of work that an attacker needs to break a pattern was measured in subsection 4.4. In this section, we provide a framework for a practical measurement of the number of guesses that would crack a pattern within the valid attempts (before phone lock-out). We propose an online (on phone) attack simulation since an attacker will have to make attempts physically on a device and it is only practical to see how many patterns are cracked before the phone locks-out or data wipes off.

The frameworks would use the result of the evaluations to generate likely patterns and try them on a phone and record the number of attempts, timing and success rate. We looked into lock-out times for different versions of Android OS and found variations in the implementation, as some devices lock-out while others do not (even for the same OS version). Table 10 show variations between 3 Android operated devices (2 with same OS version). However, the variations could possibly be due to different phone manufacturers.

With total valid patterns being 389,112, theoretically, the probability of guessing a correct pattern would be 1 in 389,112

(0.00000257) without taking any biases or evaluations into account. But how much time would it take to guess a pattern? In the worst-case scenario (the pattern being the last), it would take 2,334,672 secs (38911.2mins) for Android 5.0+, 11,673,120 secs (194,552 mins) for Android 6.0+ (Xperia Z3) and 1,400,759,190 secs (23345986.5 mins) for Android 6.0+ (Samsung Galaxy S6).

However, previous evaluations and results (Section 4) have shown that certain patterns and nodes have high probability of usage with many associated with alphabets and numbers. Hence, adversaries could use those information to exploit the limitations of the phone lock-out or data wipe off. They could generate more likely patterns and keep trying until they guess the correct pattern.

To measure the probability of guessing the correct password for each phone and during a certain period, an empirical study would need to be conducted. Therefore, it is our hope to conduct that empirical study in future work.

6. Conclusion

In this paper, we dissected the Pattern Unlock by analysing the characteristics empirically with user collected data. Our analysis revealed interesting results, some of which agrees with existing research findings. We found bias in the starting node of patterns among users and users mostly chose patterns with lengths 6, 7 and 9, surprisingly, the length of 8 were not well utilised. There was a high percentage of patterns that were associated with alphabets and numbers. For the security strength, the entropy scores were measured to be between 2.858bits and 26.140bits. Furthermore, significant disparities were found in the choice and usage of patterns between male and female participants.

Some limitations are highlighted and largely due to the data collection methodology. The data collection was centred on one city and the participants were either staff or students of the University. This is a limitation as it only focused on one location and a certain population of people. In the whole data sets, the number of left-handed participants was very small that no significant evaluation could be conducted on handedness. However, more left-handers would be sought after and data would be collected in a second location to see if the results would still hold. There may be experimental bias where participants created patterns for the study knowing the patterns would not be used regularly.

The analysis conducted in Section 4 revealed interesting outcomes and indicated weaknesses in users pattern choice consciousness and techniques. The section highlighted the need for users to be more aware of not only the kinds of patterns to avoid but also how to choose potentially more secure patterns. However, telling users the kind of patterns to chose or not might not be sufficient as some users would still use weak, simple and easy patterns, perhaps they might choose strong patterns if they see a visual indication of the strength of their patterns. In that case, the various evaluation results should be factored in when calculating the strength and that would be more precise indication which would make users choose better and more secure patterns.

Further study will explore more specific analysis for better insight. For instance, analysis based on user's handedness and language. Online attack would be simulated and the viability of enhancing the security of Pattern Unlock by incorporating biometric modalities to provide multi-layer security would be evaluated.

References

- [1] A. J. Aviv, D. Budzitowski, and R. kuber. Is bigger better? comparing user-generated passwords on 3x3 vs. 4x4 grid sizes for android's pattern unlock. *Annual Computer Security Applications Conference (ACSAC), Los Angeles, CA. USA*, 2015.
- [2] G. Blonder. Graphical passwords. (US Patent 5559961), 1996.
- [3] M. W. Calkin. Short studies in memory and association from the wellesley college laboratory. In: Psychological Review, 1898.
- [4] D. G. Clark and J. Lindqvist. Engineering gesture-based authentication systems. *Pervasive Computing*, 2015.
- [5] T. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, New York, USA, 1991.
- [6] DataGenetics. Pin analysis, 2015.
- [7] P. Dunphy and J. Yan. Do background images improve draw a secret graphical passwords? In ACM, 2007.
- [8] J. I. et al. The design and analysis of graphical passwords. 8th USENIX Security Symposium, 1999.
- [9] N. Jeanjaitrong and P. Bhattarakosol. Feasibility study on authentication based keystroke dynamic over touch-screen devices. 13th International Symposium on Communications and Information Technology (ISCIT), 2013.
- [10] M. D. Lrge. Tell me who you are and i will tell you your unlock pattern, 2015.
- [11] D. Matteo, P. Michiardi, and Y. Roudier. Password strength: An empirical analysis. INFOCOM, 2010 Proceedings.
- [12] G. K. Patrick. Guess again (and again and again): Measuring password strength by simulating password-cracking algorithms. *Symposium on Security and Privacy.*, 2012.
- [13] P. C. Pfleeger and L. S. Pfleeger. Analyzing Computer Security: A Threat/Vulnerability/Countermeasure Approach. Pearson Education, New Jersey. USA, 2011.

- [14] Splash-Data. Worst password of 2015, 2015.
- [15] C. Sun, Y. Wang, and J. Zheng. Dissecting pattern unlock: The effect of pattern strength meter on pattern selection. *Journal of Information Security and Applications*, 2014.
- [16] S. Uellenbeck, M. Durmuth, C. Wolf, and T. Holz. Quantifying the security of graphical passwords: the case of android unlock patterns. *Conference on Computer & Communications Security (CCS)*, 2013.