Analyzing the Impact of Collection Methods and Demographics for Android's Pattern Unlock

Adam J. Aviv
United States Naval Academy
aviv@usna.edu

Justin Maguire
United States Naval Academy
justmag94@gmail.com

Jeanne Luning Prak
Carnegie Mellon University
jlprak@mimech.us

Abstract—Building upon previous work on measuring the strength of graphical passwords we report on further analysis of statistical differences between Android's graphical passwords collected in-lab via pen-and-paper surveys and those collected on-line via self-reporting on users own mobile device. Additionally, we conducted between group studies of demographic groups for online, self-reported patterns and investigated weather locale (urban vs. suburban vs. rural), handedness (left vs. right), or reported gender (male vs. female) impacts the visual features of self-reported patterns. We find that there exist subtle but statistically significant differences between data collected in-lab using pen-and-paper and data collected on-the-device via self-reporting, as well as differences between demographic groups that should be considered when recruiting participants. For example, we found subtle left/right shifting between genders in the Android pattern's grid space as well as the presence or absence of visual features, such as crosses, within the pattern. We further analyzed the guessability (a common strength metric) of self-reported and pen-and-paper as well as for different demographic groups and find that overall, the strength of patterns is very similar. Overall, these results suggest that while in-lab surveys for Android graphical passwords using pen-and-paper are a reasonable substitute for real/in-the-wild data, there are likely subtle ecological differences that need some accounting as well as consideration of the demographic sampling, particularly for gender.

I. INTRODUCTION

Studying graphical passwords in the same manner as text-based passwords is challenging because graphical passwords are not used for remote authentication, and are thus unlikely to be hacked and leaked to the public where researchers can analyze them offline. As a result, large corpora of real world graphical passwords do not exist for study. To compensate, researchers have conducted studies within the lab [10], [8], [1] to collect and analyze data, or collect data through applications distributed through the app store [7].

In a recent report [3], Aviv et. al collected Android graphical unlock patterns using an in-lab methodology that closely followed previous methods by Uellenbeck et. al [10] where participants drew patterns using pen-and-paper, and Aviv et. al

also collected data using an on-line system through Amazon Mechanical Turk where participants self-reported their patterns on their own mobile device. In this follow up communication, we wish to provide further analysis of the statistical differences between Android graphical passwords collected using *pen-and-paper* and those collected online through a *self-reporting* techniques on users own mobile device. Additionally, we conduct analysis on demographic differences based on locale, age, and gender of the self-reported dataset as it consists of a large and relatively diverse sample.

Using these data sets, it is our goal to provide insight into data collection methods and demographic differences between groups. It is important to note that we do not claim that prior work, ours or others', is invalid or wrong. To the contrary, we wish to better understand the the impact of methodlogical choices when asking participants to report or generate graphical patterns, and, in fact, this work *confirms prior results* and suggests that pen-and-paper analysis, at least for Android's graphical pattern, does provide a reasonable substitute for real-world patterns as have been self-reported. The major differences betweeen those reported online and those generated via pen-and-paper is in the pattern length (pen-and-paper use more contact point, 6.27 vs. 6.05, respectively), and the start and end location tend to be shifted more to the left and more to the center for pen-and-paper compared to self-reporting.

Additionally, diving into the demographics of the self-reported data set, we observe subtle differences between groups that the community should consider during recruitment. In particular, patterns tend to shift in a direction (towards the right vs. left) based on gender; surprisingly, females are more likely to have more right shifted patterns as compared to males. Females are also more likely to use non-adjacent strokes while males are more likely to use crosses. We hypothesize this difference might be related to hand size where females have smaller hands than males and thus, when enetering patterns with a single hand, e.g., with just the thumb while holding the device, use patterns more focused in one area of the screen instead of spread out over the entire space. Handedness also has a small affect; right-handed users have longer patterns than left-handed users, and we hypothesize that how the device is held (one vs. two handed) may underly this difference. Locale and age, however, seem to have minimial impact on features of patterns.

Finally, we analyzed the security of patterns between demographic groups and collection methodology using the partial-guessability metric [5]. There apears to be no signficant
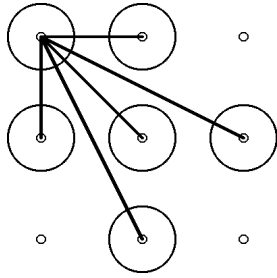
Fig. 1: The available next contact points from the contact point in the upper lef

differences between pen-and-paper and self-reported patterns, and there is also little difference between demographic groups such as gender and locale. At the extremes, the most guessable patterns, knowledge of the victim group can aid in guessing, decreasing the security, but this efffect is reduced when considering patterns in total. Overall, though, the security of all groups are consistently low as compared to other authentication techniques, on the order of a random 3-digit PIN.

II. Related Work and Background

A. Android Graphical Password

The Android graphical password pattern, is a variant of previous draw-metric schemes for graphical passwords [9], [6], [12], [11]. Password patterns are used as an authentication method on Android phones to *unlock* the device. It is not typically used for other remote authentications, and is one of many methods for device unlocking that users can choose. Other choices include PINs, passwords, and more recently, biometrics like fingerprint reading or facial recognition.

Android's graphical password system requires a user to select a password that must be recalled when unlocking the device. A password pattern is "drawn" on a 3x3 grid and must follow a set of basic rules. First the user must contact at least four of the points, and a contact points can only be selected once, no repetition. The user cannot lift their finger while drawing the pattern; it must be one continuous motion, maintaining contact with the touchscreen throughout. Finally, a user may not avoid a non-contacted point in an attempt to reach the next contact point, e.g., by going outside the grid space. Figure 1 shows which contact points are reachable from the first contact point; note that, once a contact point is selected, the user may trace over it to reach other contact points. In total, there are 389,112 possible password patterns, a smaller set than one might expect, and it is also known that users choose from a predictable subset which can impact security [3], [10].

B. Methodology Studies of Graphical Passwords

The Android pattern has been well studied by researchers using varied methodologies. The earliest analysis of user-generated patterns was conducted by Andriotis et. al [2] where users were brought into a labratory setting and asked to provide one "easy" and one "hard" pattern. Follow-up work has been conducted by Andriotis et. al [1] using a larger sample size. Song et. al [7] and Sun et. al [8] investigated the effect of password meters on user choice. Song et. al collected data by releasing an application into the Google Play store, while Sun et. al recruited participants locally.

Most related to this work is the methodology used by Uellenbeck et. al [5] and the pen-and-paper method used by Aviv et. al [3] that closely followed Uellenebeck et. al. In this method, which uses an adversarial setup, participants are asked to provide a set of *defensive patterns* and *offensive patterns*. The participant is rewarded for choosing a defensive pattern that was not guessed by another participant's offensive pattern. The adversarial, pen-and-paper collection methodology will be directly compared to another method employed by Aviv et. al in [3], namely self-reporting of patterns where participants were recurited online via Amazon Mechanical Turk and paid to optionally report their pattern on their own mobile devices.

III. Data Collection Methods

In this section, we provide more details for the data collection methodologies. Both studies were approved by our Institution Review Board and meet the ethical standards. First, we will describe the online, self-reported study, followed by the data collection methods of pen-and-paper study. Due to space considerations, we wish to refer the readers to the conference proceedings for more detailed description of the collection methodologies [3].

Self-Reporting. The online, self-reporting survey was conducted on Amazon Mechanical Turk. The survey was launched in two rounds: the first round compensated participants $0.50 and the second round compensated participants at $0.75. We required that the survey be completed on the mobile device of the participant, and this was checked using user-agent strings. The hope is that when presented with a pattern entry screen on the same device that they enter their pattern regularly, the participant, by reflex, would provide their pattern. We had further consitency checks; we required participants to enter their pattern twice, once at the beginning and once at the end, and only used data where the patterns matched.

Finally, participants have the option to *not* report their pattern and still receive compensation. At the beginning of the survey, the following message appears:

> *For this survey, you have two options. You may either provide your actual android unlock pattern or answer a series of questions about your pattern.*

A participant choosing to answer questions (as opposed to self-reporting their pattern) will be prompted for information about the start point of their pattern as well as asked to indicate common tri-grams found in their pattern (e.g., a stroke across the bottom of the grid space). The hope for this methodological design is that participants self-selecting into reporting their patterns will likely provide more truthful answers as they had the option to not do so early on. Additionally, we found consistency in the common sub-sequences and start locations for those participants self-reporting and those participants answering questions.

In total, 757 participants were recruited to take the survey. After attention tests were taking into account, we accepted 440 patterns as part of the data set for self-reported pattern (often referred to as the *SR* acronym in figures) . The demographics for the self-reported data is presented in Table I.

		Male	Female	Right	Left	Total
Locale	Urban	104	54	136	22	**158**
	Suburban	111	84	161	34	**195**
	Rural	38	49	77	10	**87**
Age Range	18-24	91	55	125	21	**146**
	25-34	131	95	195	31	**226**
	35-44	24	29	43	10	**53**
	45-54	6	7	9	4	**13**
	55-64	1	1	2	0	**2**
	Total	**253**	**187**	**374**	**66**	**440**

		Male	Female	Right	Left	Total
Locale	Urban	9	2	11	0	**11**
	Suburban	15	6	21	0	**21**
	Rural	4	2	5	1	**6**
Age Range	18-24	20	10	29	1	**30**
	25-34	7	0	7	0	**7**
	35-44	1	0	1	0	**1**
	45-54	0	0	0	0	**0**
	55-64	0	0	0	0	**0**
	Total	**28**	**10**	**37**	**1**	**38**

TABLE I: Demographic Breakdown of Self-Reported (left) and Pen-and-Paper (right)

Pen-And-Paper Study. The second data set consists of participants generating patterns using a pen-and-paper survey that closely follows the methodology of Uellenbeck et. al [10]. The procedure applies an adversarial model where participants are asked to create patterns of their own (so called *defensive patterns*) that must be defended from attack from other participants who are allowed to make a series of guesses (so called *offensive patterns*). The survey was conducted over three sessions, two at the United States Naval Academy and at the University of Maryland, Baltimore County. The demographics of the participants, again, can be be found in Table I.

The session proceeded by first having participants select three defensive patterns. Then, participants choose ten offensive patterns that are attempts to guess others' defensive patterns in their session. Participants are rewarded for correct guesses (in the form of a candy treet) and also for being able to recall their offensive pattern at the end of the study. All patterns were drawn by hand using a marker on sheets of paper containing a set of 3x3 grids (three 3x3 grids for offensive and ten 3x3 grids for defensive). During the selection, the participants were informed that they will be rewarded (in the form of a candy treat) if their patterns were not guessed by other participants, and participants were also infomred that they would have the opportunity to earn more treats by guessing other participants patterns.

In total, we collected 494 3x3 patterns (380 offensive and 114 defensive) from 38 partcicipants. In our analysis, as we are concerned with the behavior of individual partipants, we analyze the *average pattern statisitcs* from all the 13 patterns created by the participant (three offensive and ten defensive). That is, for example, we analyze the average length of all the patterns created by a single participant and do not treat each 13 patterns seperately. While there are clear differences between offensive and defensive patterns (we refer the reader to our other report for details [3]), we argue that this average pattern statistics form a good representation for the kind of pattern a user is likely to use on their device.

IV. PATTERN FEATURES

It is not the intention of this paper to review prior published results regarding the general features of the data sets being analyzed. Instead, we are interested in analyzing features as they differ in the collection methodology and the demographics of the data. As such, we focus on the set of visual features of
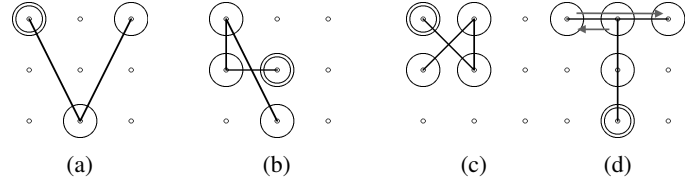

Fig. 2: Non-Ordinal Visual Features Used in Analysis: (a) Kinght Move, (b) Cross, (c) Ex, and (d) Non-Adjacent

patterns as well as the partial-guessability metric of patterns as a measure of security.

The visual features of patterns have been analyzed in prior work [4] in the context of understanding user perceptions, and we apply the same visual features here. We divide these features into two classes, ordinal and non-ordinal features:

Ordinal Features. These are features that can be measured in a continuous manner and assigned a value, and we used the following ordinal features in our analysis:

- Pattern *height* refers to the shifting up or down of the pattern in the grid space. This is calculated by assigning a positive value to the contact points in the top row of the grid space, a negative value to the contact points in the bottom row, and zero to the middle row of contact points. A height value is determined by taking the average of the assigned values for each contact point used in the pattern. For pen-and-paper, we use the average height for all patterns created by a signal user.

- Pattern *side* refers to the shifting left or right in the pattern grid space. It is calculated in a similar manner as height with positive and negative values assigned to the left and right columns, respectively, and a neutral value for the middle column. The side value is determined by taking the average of the assigned value for the contact points in the pattern, and, again, for pen-and-paper, the average side value is use for all the patterns created by a single user.

- Pattern *length* refers to the total number of contact points used in the pattern. This value can range

	PP	SR	Right (SR)	Left (SR)	Male (SR)	Female (SR)	Urban (SR)	Suburban (SR)	Rural (SR)
Height	-0.17	0.13	0.13	0.17	0.16	0.10	0.20	0.20	-0.14
Length	6.27	6.05	6.11	5.77	6.10	6.01	6.11	6.06	5.95
Stroke Length	5.91	5.82	5.90	5.36	5.85	5.78	5.90	5.81	5.71
Side-Shift	-0.09	-0.04	-0.05	0.02	0.10	0.23	0.20	-0.12	-0.28

TABLE II: Ordinal Attribute Measurements

between 4 and 9 as a pattern must use at least four contact points. For pen-and-paper, the average length of the thirteen patterns created by a participant is considered.

- Pattern *stroke length* is similar to length but is instead concerned with the total length of all the lines in the pattern. This is calculated by mapping each of the contact points into a Cartesian plain with (0,0) as the center contact point and one step is used to reach the contact point to the immediate right, left, up, and down. The sum of the Gaussian distances of each stroke is used as the statistic. For pen-and-paper, the average stroke-length across all generated patterns is created, per participant.

- Pattern *Start-X*, *Start-Y*, *End-X*, and *End-Y* features are a measure of the right/left and up/down shift of the starting and ending contact point. To measure this, we again mapped each of the contact points into a Cartesian plain with (0,0) at the center contact point. A negative X or Y value indicates a tendency to the left and down, respectively, while a positive X or Y value indicates a tendency to the right and up, respectively. The average of Start-X, Start-Y, End-X and End-Y, as calculated across all patterns generated by the participant, is used in the analysis for pen-and-paper.

Non-Ordinal Features. These are features where we are only concerned with the presence or absence of the feature, and we are primarily focus on complexity features as used in previous work [4].

- *Knight moves* is a stroke on the grid space that connects two contact points that is over one and down two, much like a knight moves in chess. An example of a knight move is in Figure 2a.

- *Crosses* is a stroke that crosses over a prior stroke. Figure 2b shows an example of a cross with a knight move, and Figure 2c shows an example that forms of 'X'.

- *Exes* is a subset of crosses and only refers to the case where a perfect 'X' is formed with 90 degree angles.

- *Non-adjacency* is a stroke that double backs on itself such that a previously contact point is traced over. Figure 2d shows an example of a non-adjacency that would occur in a pattern shaped like a 'T'.

V. ANALYSIS AND RESULTS

For significance testing of ordinal data for between group comparisons (e.g., handedness, gender), we applied the standard Student's t-test if the data was normally distributed or

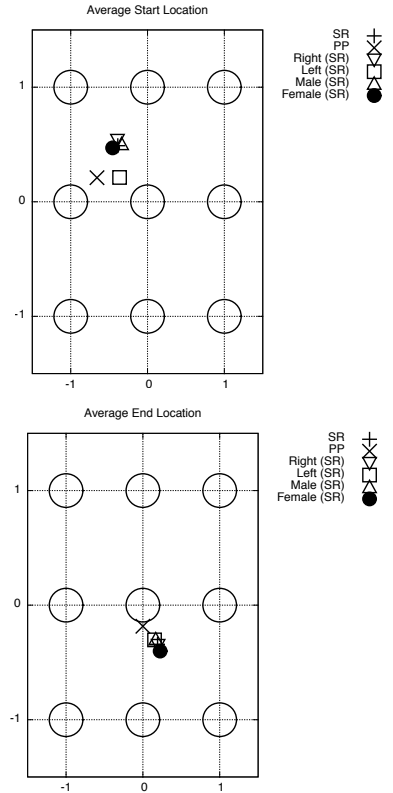|	Start-X	Start-Y	End-X	End-Y
SR v. PP	$p =0.099*$	$p =2.663\text{e-}08**$	$p =0.0235**$	$p =0.142$
Hand (SR-only)	$p =0.36$	$p =0.001**$	$p =0.28$	$p =0.329$
Gender (SR-only)	$p =0.075*$	$p =0.333$	$p =0.209$	$p =0.083*$

TABLE III: Start and End Shifting

Fig. 3: Average Start and End Location

Mann-Whitney U-Test if the data was not normally distributed. We used Anderson-Darling test to determine if the data was normal. When analyzing multiple groups to test significance (e.g., location and age), we used ANOVA testing. Finally, for comparing non-ordinal data between groups we used the χ^2 test. As we are conducting many tests over the same data, we performed p-corrections using Bonferonni, and significance was claimed when $p < 0.05$. We also indicate possible significant interest when $p < 0.10$. These two settings are indicated in all tables with $**$ for $p < 0.05$ and $*$ for $p < 0.10$.

Ordinal Feature Comparisons. The first set of comparisons we consider is ordinal features comparing the following groups: self-reported patterns to pen-and-paper patterns (*SR v PP*); right handed self-reporters to left-handed ones (*Hand*); and male self-reporters to female self-repoters (*Gender*). Ta-

	Height	Length	Strk Length	Side
SR v. PP	$p = 0.237$	$p = 0.049$**	$p = 0.115$	$p = 0.130$
Hand (SR-only)	$p = 0.405$	$p = 0.027$**	$p = 0.027$**	$p = 0.455$
Gender (SR-Only)	$p = 0.327$	$p = 0.261$	$p = 0.227$	$p = 0.019$**

TABLE IV: Testing of Ordinal Attributes of Patterns

	Kmoves	Crosses	Non-Adj	Exes
SR vs. PP	$p = 0.883$	$p = 0.732$	$p = 0.354$	$p = 0.781$
Hand (SR-only)	$p = 0.802$	$p = 0.555$	$p = 0.871$	$p = 0.689$
Gender (SR-Only)	$p = 0.088$*	$p = 0.004$**	$p = 0.071$*	$p = 0.866$

TABLE V: Testing of Non-Ordinal Attributes of Patterns

Fig. 4: Frequency Occurrences of Non-Ordinal Features

ble II provides the measurement results for the ordinal data, e.g., differences in the calculated means, and Table IV provides the primary statistical results for ordinal features.

Foremost, comparing patterns generated using pen-and-paper and self-reported patterns, we see that there exists a significant difference in the length of the patterns. Namely, pen-and-paper patterns use 6.27 contact points on average as compared to 6.05 for self-reported ones (these figures can be found in Table II). Interestingly, the stroke length for the two sets are not significantly different, 5.99 vs. 5.82. This suggests that although using pen-and-paper may encourage participants to select patterns with one or two more contact points it does not necessarily encourage participants to select more complex patterns with longer stroke lengths, which would occur when adding visual features with more diagonals like crosses or exes.

Additionally, in Table IV, we see that there exist three other significant differences when comparing between groups for the self-reported data. First, for handedness, we find that there exists a difference in both the length and stroke-length when comparing left and right handed participants. Namely, right handed individuals choose longer patterns in both the number of contact points and the length of the individual strokes (see Table II). We hypothesize that this difference may be related to the way that right-handed and left-handed individuals hold their phone and are more inclined to use one versus two hands. Unfortunately, we do not have additional information to test this hypothesis, but the fact that side and height shifting are not significant for handedness suggest that it is not related to which hand being used (e.g., entering a pattern with your left-handed should shift left) but rather some other factor with holding the device.

Another interesting result is comparisons between genders. We find that females are more inclined for their patterns to be right shifted; although, these patterns do not show significant differences in other ways, including security. We hypothesize this may be related to hand size as female participants are right-handed and that females, generally, have smaller hand sizes than males. As a result, it is more natural and comfortable for them to focus their pattern more on the right side of the screen if entering a pattern using one hand, e.g., by holding the phone in their right hand and entering the pattern with the thumb on their right hand.

Start and End Point. We also observe significant differences in the start and end locations for certain groups. These results can be found in Table III and Figure 3. Foremost, we find that there is a significant difference in both the start and end
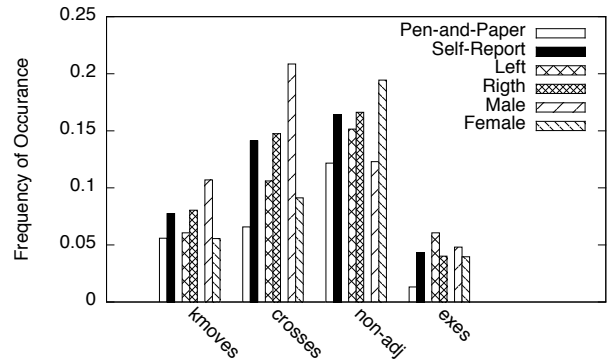
locations for pen-and-paper patterns and self-reported patterns. Pen-and-paper patterns are shifted more down and to the left for their start location and more towards the center for their end location as compared to self-reported patterns. We hypothesize that this is likely due to the act of drawing on paper and the methodology used, i.e., having to generate many patterns and guess the patterns of others. Participants were encourage to explore more of the grid space and less likely to always select the common starting point in the upper left. Similarly, this likely led to the change in the end location, however, with slightly less effect.

Additionally, there exists a difference in the up/down shifting of the start location for differently handed individuals. Right-handed participants were more likely to be shifted to the lower contact points as compared to their left-handed counter parts. Again, this is likely due to subtle differences in how these two groups handle their devices. Finally, we do observe some differences in the Start-X and End-Y locations for gender, but these results are not strongly significant, and we would need further analysis to make strong claims here.

Non-Ordinal Features. Figure 4 presents the frequency results for presence or absence of non-ordinal features, and Table V presents the statistical results. We observe no significant differences when comparing the self-reported to the pen-and-paper patterns or when comparing differently handed individuals. However, we do observe a difference with gender; namely, that male participants are more likely to use crosses while female participants are more likely to use non-adjacent strokes (see Figure 4). We hypothesize, again, that this might be related to hand size.

Age and Locale. We also analyzed both ordinal and non-ordinal features with respect to the impact of locale (i.e., urban, suburban, and rural) and age range. As there are multiple factors at play we used ANOVA as our statistical test, and we found that there was no statistically significant differences between these groups except for side shifting and location. Urban users indicated a right shift in their patterns as compared to suburban and rural participants. We hypothesize that this might be related to using the device with single vice multi-handed. For example, an urban user may be more inclined to authenticate with one hand as they might be walking at the time while suburban and rural users may have more opportunity to

	height	length	strklen	side	kmoves	crosses	non-adj	exes
Locale	$p = 0.069$	$p = 0.765$	$p = 0.803$	$p = 0.016^{**}$	$p = 0.666$	$p = 0.704$	$p = 0.682$	$p = 0.565$
Age	$p = 0.698$	$p = 0.256$	$p = 0.047$	$p = 0.33$	$p = 0.153$	$p = 0.101$	$p = 0.077$	$p = 0.31$

TABLE VI: Variance Among Demographic Groups (ANOVA on SR-only)

Distribution	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.5$	$\alpha = 0.7$
Pen-and-Paper	6.59	6.99	8.92	10.12
Self-Reporting	6.62	6.95	9.49	10.74
Men (SR)	5.86	7.00	9.59	10.60
Women (SR)	5.57	7.33	9.80	10.64
Urban (SR)	6.09	7.22	9.95	10.69
Suburban (SR)	6.58	8.70	9.58	10.60
Rural (SR)	6.08	8.22	9.60	10.42
Random 4-Digit PIN	13.28	13.28	13.28	13.28
Random 3-Digit PIN	9.97	9.97	9.97	9.97
Random 3x3 Pattern	18.57	18.57	18.57	18.57

TABLE VII: Comparing Partial Guessing Entropy

use two hands, and thus more inclined to use more of the contact points.

Guessability Metric. To measure the guessability of the data, we followed the guessing algorithm as described in [3] and used the average of five iterations of a randomized five-fold cross validation. We consider each pen-and-paper pattern individually when guessing, as in previous work. For the security metric, we use partial-guessing entropy [5] which is a calculation based on the number of guesses required to recover the given fraction of the data set. For example, with $\alpha = 0.5$, measured in bits, the value indicates the amount of work for an attacker to guess 50% of the data set. This metric has been used previously [3], [10], [7] as a measure of graphical pattern strength.

Table VII contains the primary result. Here we find that the overall entropy is low for all the groups, on the order of a 3-digit random PIN for cracking 50% of the dataset. This was known previously [3], [10]; however, comparing between groups, we find that there are similarities in strength for all groups. No paricular group seems to have more or less secure patterns. Interestingly, observe $\alpha = 0.1$ the entropy for both genders are lower suggesting that if the attacker had knowledge of the victims gender and can train on that gender, the most easily guessed password become more easily guessed as compared to other groups. This advantage quickly diminishes, though, and by the time the attacker is interested in cracking 70% of the patterns, all groups are indistinguishable in difficulty.

VI. CONCLUSION

In this paper, we have shown that there exists subtle but important demographic differences between groups for Android's graphical password. In particular, there seems to be differences between genders and right-handed and left-handed user. Additionally, we analyzed pattern collection methodologies between patterns drawn using pen-and-paper and patterns that are collected via a self-report study using the participants own mobile device. While there exist differences in the number of contact points, in other respects, such as stroke length, visual features, and guessability, it would seem that pen-and-paper can form a good proximate for real world patterns as have been self-reported.

REFERENCES

[1] Panagiotis Andriotis, Theo Tryfonas, and George Oikonomou. Complexity metrics and user strength perceptions of the pattern-lock graphical authentication method. In *Human Aspects of Information Security, Privacy, and Trust*, pages 115–126. Springer, 2014.

[2] Panagiotis Andriotis, Theo Tryfonas, George Oikonomou, and Can Yildiz. 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, pages 1–6, 2013.

[3] Adam J Aviv, Devon Budzitwoski, and Ravi Kuber. Is bigger better? comparing user-generated passwords on 3x3 vs. 4x4 grid sizes for android's pattern unlock. In *Proceedings of the 31th Annual Computer Security Applications Conference*, ACSAC'15, 2015.

[4] Adam J Aviv and Dane Fichter. Understanding visual perceptions of usability and security of Android's graphical password pattern. In *Proceedings of the 30th Annual Computer Security Applications Conference*, ACSAC'14, pages 286–295, 2014.

[5] Joseph Bonneau. The science of guessing: analyzing an anonymized corpus of 70 million passwords. In *Proceedings of the 2012 IEEE Symposium on Security and Privacy*, SP'12, pages 538–552, 2012.

[6] Ian Jermyn, Alain Mayer, Fabian Monrose, Michael K. Reiter, and Aviel D. Rubin. The design and analysis of graphical passwords. In *Proceedings of the 8th Conference on USENIX Security Symposium - Volume 8*, SSYM'99, 1999.

[7] Youngbae Song, Geumhwan Cho, Seongyeol Oh, Hyoungshick Kim, and Jun Ho Huh. 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, 2015.

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

[9] H. Tao and C. Adams. Pass-go: A proposal to improve the usability of graphical passwords. *International journal of Network Security*, 7 (2):273–292, 2008.

[10] Sebastian Uellenbeck, Markus Dürmuth, Christopher Wolf, and Thorsten Holz. Quantifying the security of graphical passwords: The case of Android unlock patterns. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, CCS'13, 2013.

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

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