THE JOURNAL OF DIGITAL FORENSICS, SECURITY AND LAW

Journal of Digital Forensics, Security and Law

Volume 8 | Number 4 Article 4

2013

# Technology Corner: Calculating the Number of Android Lock Patterns: An Unfinished Study in Number Theory

Gary C. Kessler
*Embry-Riddle Aeronautical University*

# TECHNOLOGY CORNER: CALCULATING THE NUMBER OF ANDROID LOCK PATTERNS: AN UNFINISHED STUDY IN NUMBER THEORY

Gary C. Kessler
Embry-Riddle Aeronautical University
Daytona, Florida
gck@garykessler.net

Although one is unlikely to ever want to brute-force an Android lock pattern, many do wonder about the relative strength of the lock pattern versus a multi-digit personal identification number (PIN). It becomes obvious pretty quickly that there are many more lock patterns than the 10,000 possible four-digit PINs.

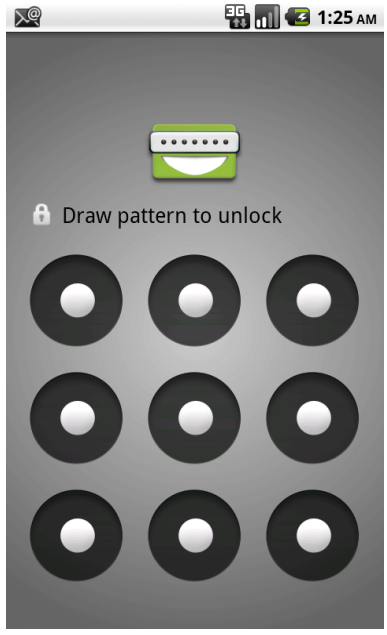


Figure 1 Android lock pattern screen

But, how many lock patterns are there? The often-cited number of Android lock patterns is 986,328, which I first found from A. Hoog (personal communication, October 2012) and S. Brothers (personal communication, October-November, 2012). Brothers went so far as to observe that this number could be found in the Android's gesture file and, more to the need for an actual formula, that the number can be derived from the formula for permutations.

The well-known formula to calculate the number of ways with which to choose r items out of a possible pool of n items–without replacement and in all possible sequences–is:

$$P(n,r) = n!/(n-r)! \quad [1]$$

According to Formula [1] and the fact that the Android screen has nine possible positions, we could find the theoretical maximum number of lock patterns (T) of length R from the formula for permutations:

$T(1) = 9!/8! = 9$

$T(2) = 9!/7! = 72$

$T(3) = 9!/6! = 504$

$T(4) = 9!/5! = 3{,}024$

$T(5) = 9!/4! = 15{,}120$

$T(6) = 9!/3! = 60{,}480$

$T(7) = 9!/2! = 181{,}440$

$T(8) = 9!/1! = 362{,}880$

$T(9) = 9!/0! = 362{,}880$

Androids only allow lock pattern lengths of three to nine, so summing up the appropriate T(r) values above yields:

$$\sum_{r=3,9} T(r) = \sum_{r=3,9} 9!/(9-r)! = 986{,}328 \quad [2]$$

For the remainder of this paper, Formula [2] will be referred to as the Theoretical Maximum Formula (TMF).

While this calculation certainly makes it clear how the value 986,328 was derived, it does not appear to accurately count the real number of lock patterns that are possible. This is because all points are not adjacent and, therefore, one cannot actually select all *theoretically* possible patterns; e.g., there is no way to go directly from one corner to another. The question then is, what is a precise formula for computing the actual number of lock patterns and, further, what is the difference between the actual value and the value derived from the TMF?

Let us label each point in the pattern with a number using the following scheme:

1 2 3
4 5 6
7 8 9

Furthermore, we will classify each point as a Corner (the set **C** = {1, 3, 7, 9}), Edge (**E** = {2, 4, 6, 8}), or Middle (**M** = {5}).

When a user selects a lock pattern, there are certain rules that must be followed:

1) The lock pattern run length must be between three and nine unique positions.
2) A point in the lock pattern can be repeated but it does not get counted more than once (e.g., the sequence 1, 2, 3, 2, 3, 6 is a legal pattern of length four).
3) For counting purposes, a point in the lock pattern must be adjacent to some point already in the pattern (e.g., if the pattern so far is 1, 2, 3, the next point can be 4, 5, 6, 7, 8, or 9).
4) A corner point is adjacent to five other points (e.g., 1 is adjacent to 2, 4, 5, 6, and 8). An edge point is adjacent to seven other points (e.g., 2 is adjacent to 1, 3, 4, 5, 6, 7, and 8). The middle point is adjacent to all other eight points.

The difference between the actual number of patterns versus the theoretical maximum number comes about, as suggested earlier, because it is against the rules to just select a random set of points from the pool. The key appears to be that M *is* adjacent to every other point and, therefore, the TMF and general permutation formula applies to the lock pattern only *after* the middle point has been selected. Therefore, we need to know how many patterns we can have *before* M is selected. I attribute this logical leap to discussion with D. Velleman (personal communication, November 2012), who produced a paper by Ponstein (1966). While Ponstein did not directly address this problem, the paper provided some new insights.

We will define:

> R = lock pattern run length (which can be between 3 and 9)
> L(i) = the number of lock patterns of length i
> NoM(i) = the number of patterns of length i that do **not** include the middle point

The "NoM(i)" concept is important because after the lock pattern includes the middle point, every other point is adjacent and we can rely on the general formula for permutations to determine the number of choices for the rest of the lock pattern. Before that, we need a Modified Maximum Formula (MMF) for counting lock patterns that accounts for all of the possible patterns **without** the middle point and then merely applies the TMF (which is, of course, the permutation formula).

Before going further, let us take a look at how this "NoM" concept works. Suppose we want to calculate the number of lock patterns of length three

(R=3), which is the minimum legal lock pattern. There are, then, four possible generic sequences:

- *Case I:* If the first position is the middle point, then there are eight possible second positions and seven possible third positions. Put another way, after the middle, there are two more points in the sequence to choose out of a remaining pool of eight possible choices.
- *Case II:* If the first position is a corner or edge (of which there are 8 choices), and the second position is M, then there are seven possible third positions. Put another way, after the middle, there is one more point in the sequence to choose out of a remaining pool of seven possible choices.
- *Case III:* If the first two positions are a corner or edge, the third position might be M.
- *Case IV:* None of the positions is M.

In the table below, M = middle position, x = either a corner or edge, and a number represents the number of remaining legal choices in that place in the sequence. The number of possible patterns for each case can be found by multiplying:

| | |
|---|---|
| Case I: M 8 7 | NoM(0) x P(8,2) = NoM(0) x 8!/6! |
| Case II: x M 7 | NoM(1) x P(7,1) = NoM(1) x 7!/6! |
| Case III: x x M | NoM(2) |
| Case V: x x x | NoM(3) |

Note that for any value of R, there will always be R+1 cases, which can be generalized from the notes above. From this case, we can generalize that the MMF with which to accurately count the number of possible lock patterns can be given by:

$$L(R) = NoM(R) + NoM(R-1) + \sum_{i=1,R-1} NoM(i-1) \times (9-i)!/(9-R)! \qquad \text{for } 2 \leq R \leq 9 \quad [3]$$

For completeness, we can define:

L(0) = 0 (i.e., no runs of length 0)
L(1) = 9 (i.e., nine possible runs of length one)

To calculate L(R), then, we need a table of NoM(i) values where i = 0, 8.

NoM(0) is a trivial case; there is a single pattern where the middle point is first, hence, NoM(0) = 1. NoM(9) is another trivial case; there are no sequences of nine points without a middle so NoM(9)=0.

For the remaining NoM values, one just needs to count the various possibilities. A number of interesting properties appear, however, which

requires additional nomenclature. NoM(1) is a rather simple case but provides a way to generalize the rest. First, note that it is obvious that there are eight possible patterns that have a single point and no middle; thus, NoM(1) = 8.

But, in a slightly more formal fashion, let us quantify the possibilities. (While a lock pattern of length one is invalid, it does get us started.) There are two possible initial point types; i.e., either corner or edge. So, we can denote the general sequence pattern as either:

C or
E

That tells us that there are two general patterns that we might see.

To actually count the number of legal sequence, we need to return to the possible patterns and determine how many points are available for each position in the pattern. In this case, of course, there are four possible corners or four possible edges. We will denote the number of valid points in a given position of the sequence by use of a superscript; thus, our possible patterns now emerge as:

$C^4$ ==> 4
$E^4$ ==> 4

Adding these values together yields a possible eight points, which we already knew from NoM(1) = 8.

We can extend this reasoning to determine NoM(2), which is the pattern that includes two points and no middle. First, start with the two patterns that we had above, namely:

C
E

Next, note that a C position in a sequence *must* be followed by an E, whereas an E position in the sequence can be followed by a C or another E. From this, we find that there are three possible patterns for a run of two that does not include the middle:

C E
E C
E E

At this point, we can quantify the number of valid points in each position. If the first point is a corner, then there are four legally adjacent edge points; since there are four corners, there are 16 possible two-position patterns starting at a corner. Similarly, if the first point is an edge, then there are four legally adjacent corner points for another 16 possibilities. Finally, if the first point is

an edge, there are only two legally adjacent edges, yielding eight more possibilities. Using our new nomenclature:

$C^4\ E^4$ ==> 16

$E^4\ C^4$ ==> 16
$E^4\ E^2$ ==> 8

Summing up, we find NoM(2) = 40.

For NoM(3), we again build on the prior set of sequences. In building these sequences, we know that if a C point can only be followed by an E, and an E can be followed by by either a C or an E. Our new patterns are:

C E C
C E E

E C E

E E C
E E E

With the patterns intact, we can now apply superscripts to represent the number of legal positions remaining in the sequence. Note that an edge only has two legal adjacent edges (e.g., 2 can move directly to 4 or 6 but cannot move to 8).

$C^4\ E^4\ C^3$ ==> 48
$C^4\ E^4\ E^2$ ==> 32

$E^4\ C^4\ E^3$ ==> 48

$E^4\ E^2\ C^4$ ==> 32
$E^4\ E^2\ E^2$ ==> 16

Thus, NoM(3) = 176.

Unfortunately, everything changes at NoM(4) because the ability to repeat a point upsets the orderly rules from above. Consider this example.

Under "ordinary" circumstances, there should be no way to have two C points in a row. Indeed, if the first point is a C and the second point an adjacent E, then the next two points must be a CE, EE, or EC. However, if the first point is a C and the second point a non-adjacent E, then it can be followed by two C points (e.g., 1, 6, 3, 9 is a valid CECC sequence). Indeed, a legal eight-position pattern could be EEEECCCC (e.g., 2, 4, 8, 6, 9, 7, 1, 3).

This work is, obviously, not yet complete; NoM(4) through NoM(8) values have not yet been calculated and may well need a brute-force program. Nevertheless, an observation can be made about the theoretical and actual number of lock patterns, which is really what this work is all about. Although only NoM(0) through NoM(3) values are supplied here, we can calculate the

number of lock patterns of length three with what we already know. From Formulas [1] and [3], we find:

| R | P(9,R) | L(R) | L(R)/P(9,R) |
|---|---|---|---|
| 3 | 504 | 328 | 0.6508 |

At first blush, then, it appears that L(R) will be noticeably smaller than the theoretical maximum. The remaining work is to determine NoM(R) for R=4,8; L(R) for R = 4,9; and:

$$\sum_{r=3,9} L(r)$$

**ACKNOWLEDGEMENTS**

My thanks are given to Sam Brothers (DHS), Andrew Hoog (viaForensics), and Dr. Daniel Velleman (University of Massachusetts, Amherst).

**REFERENCES**

Ponstein, J. (1966, May). Self-avoiding paths and the adjacency matrix of a graph. *SIAM Journal on Applied Mathematics*, *14*(3), 600-609.