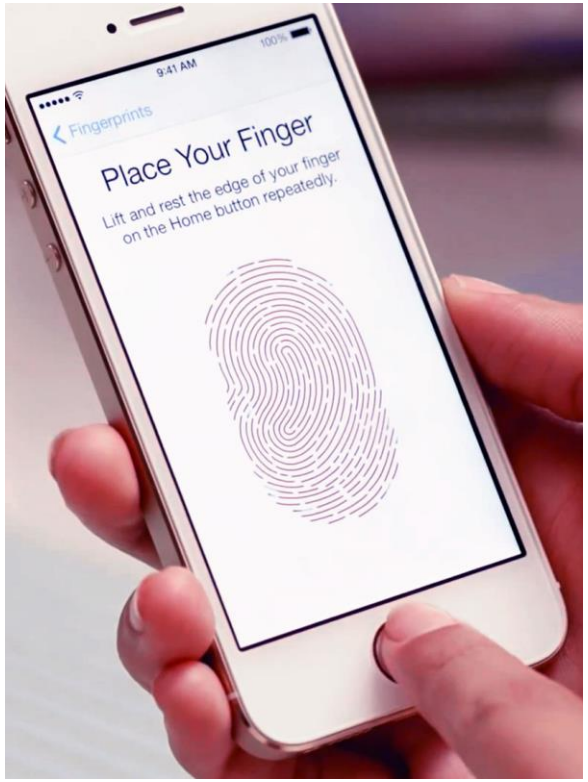
Work in Progress:
On the In-Accuracy and Influence of Android Pattern Strength Meters

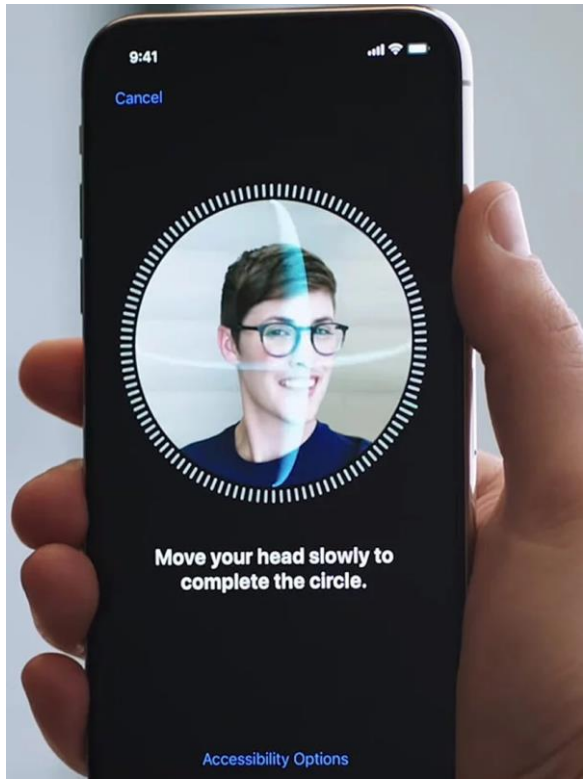
Maximilian Golla, Jan Rimkus, and Markus Dürmuth – Ruhr University Bochum
Adam J. Aviv. – United States Naval Academy

Biometric-Based Reauthentication

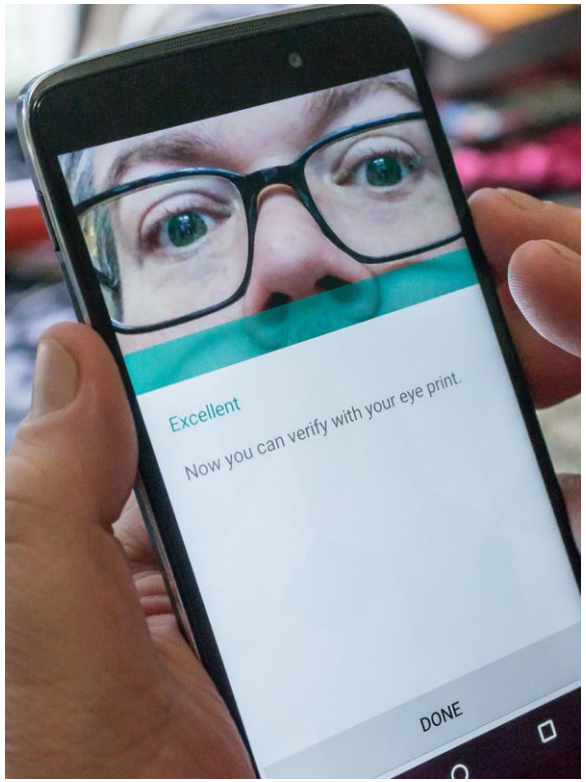
"Face unlock feels almost like not having any lock screen security."



Fingerprint



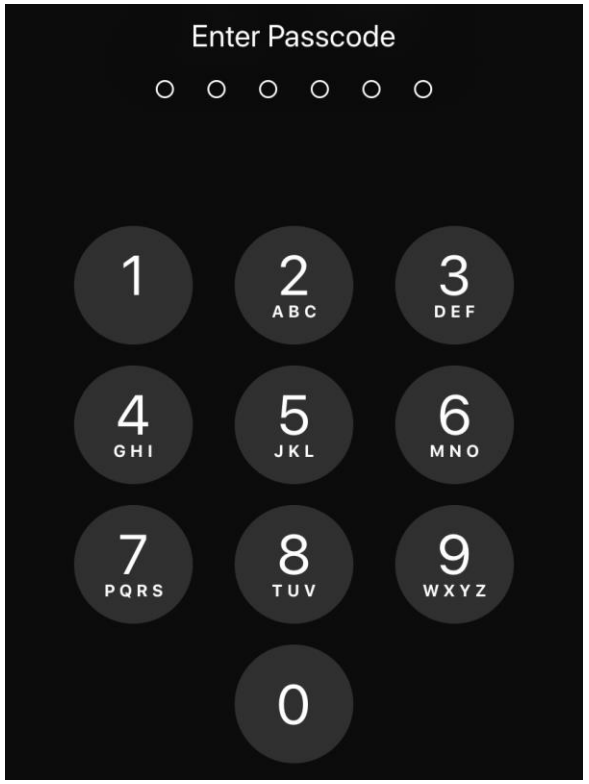
Face



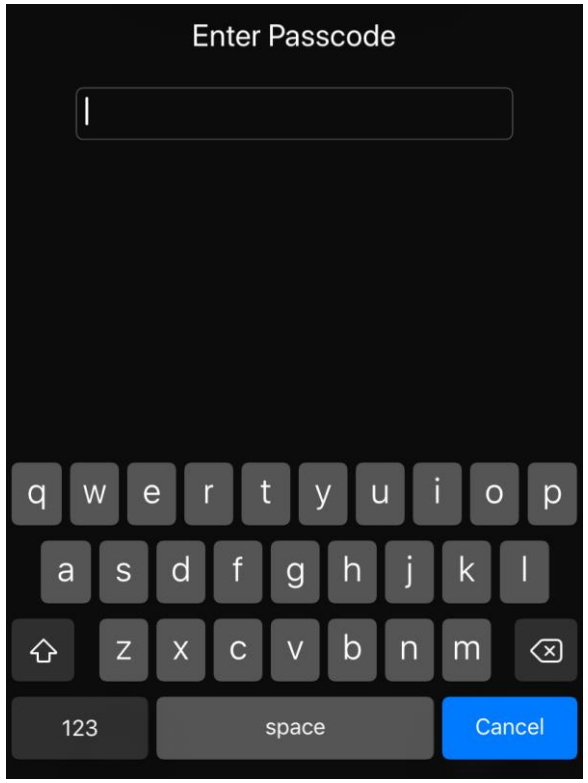
Iris

"Intelligent Scan"

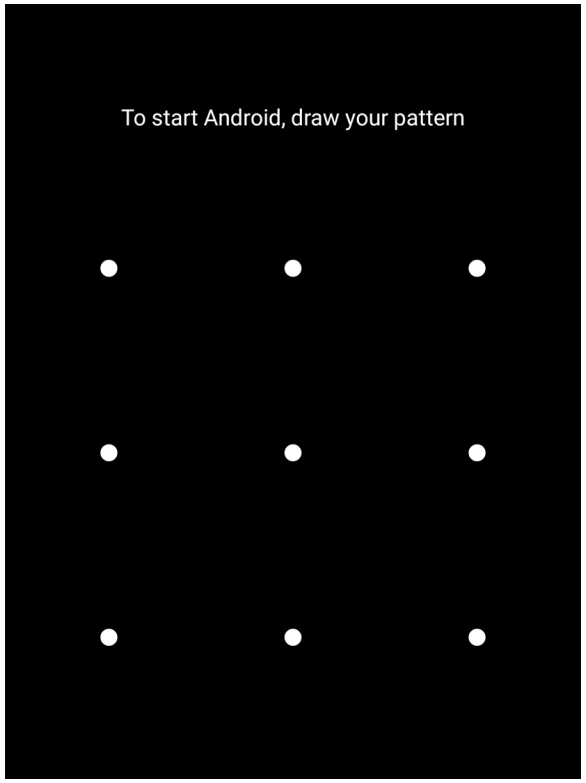
Knowledge-Based Authentication



4/6-digit PINs



Passwords



Pattern

Android Unlock Patterns

Graphical auth. scheme

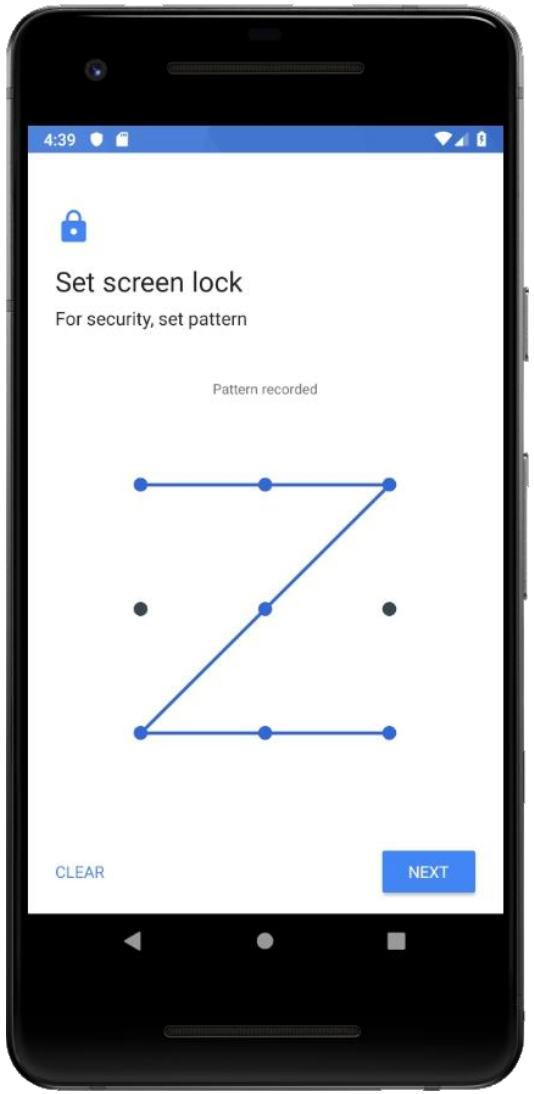
Android, 2008

Traversing nodes on a 3x3 grid

Theoretical: 389,112 patterns

Practical: Only a smaller subset is likely chosen!

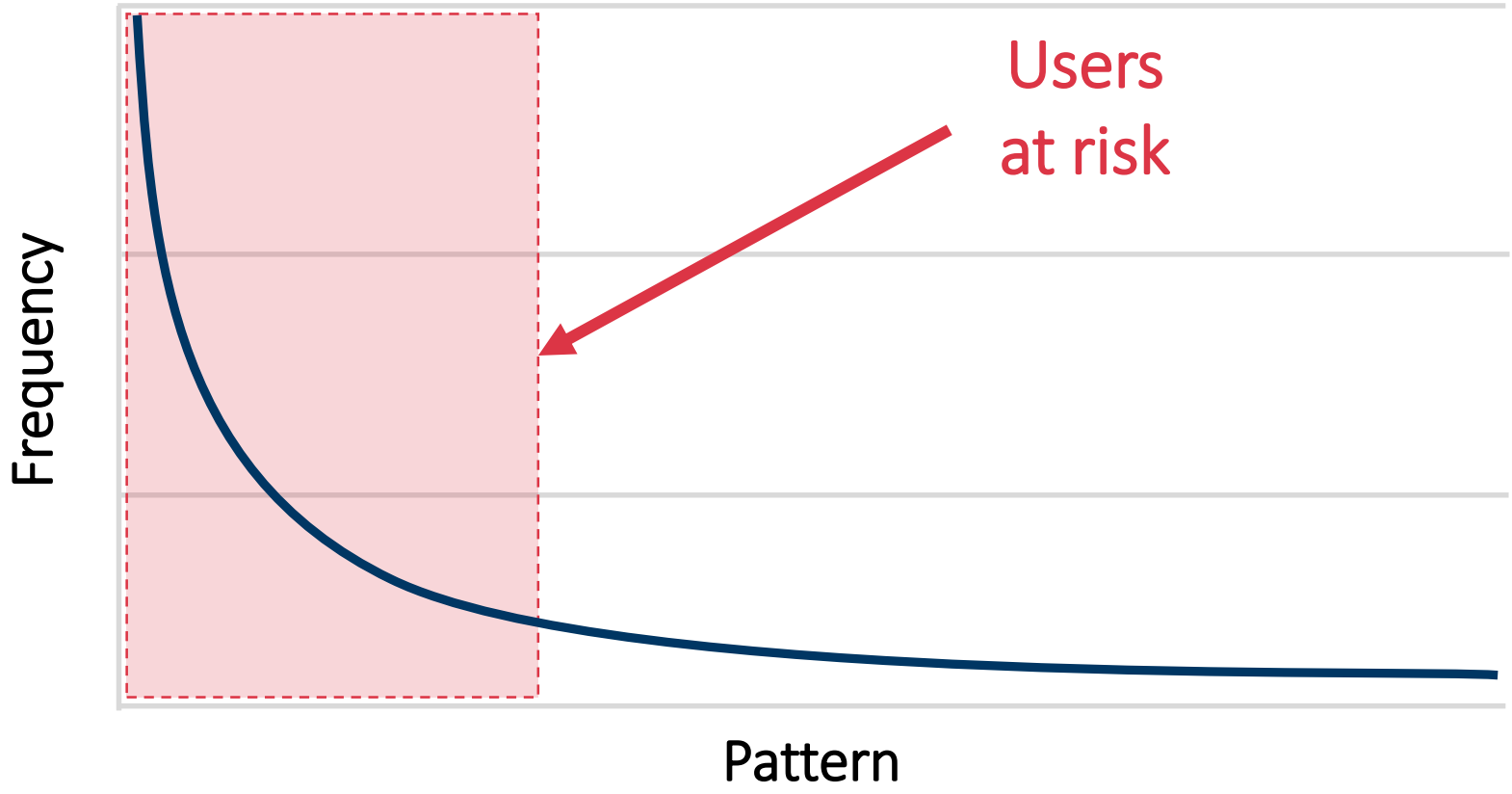
Enrollment



Authentication

Selection Bias

User-choice heavily biased.

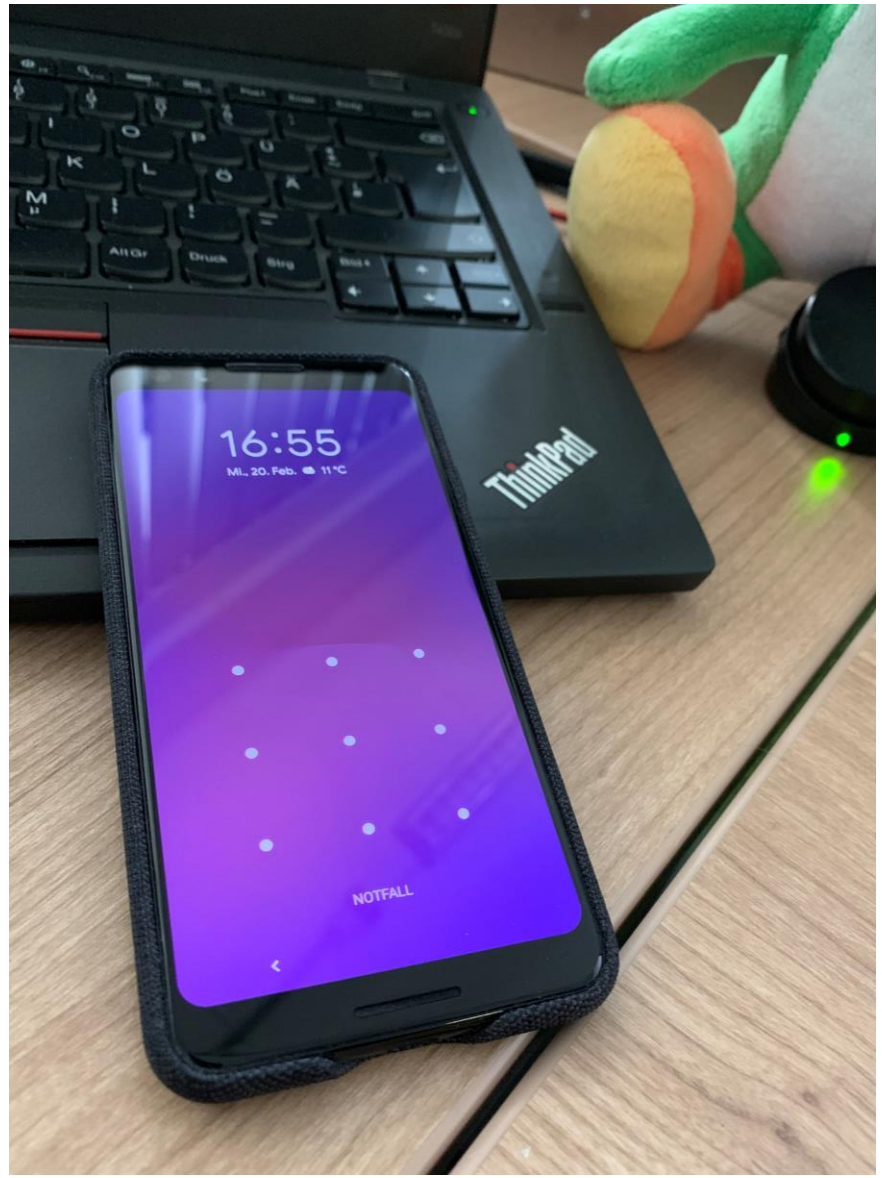


Threat Model

Attacker guesses the ***n*** most common secrets in decreasing order of success.

Throttled Guessing Attack:

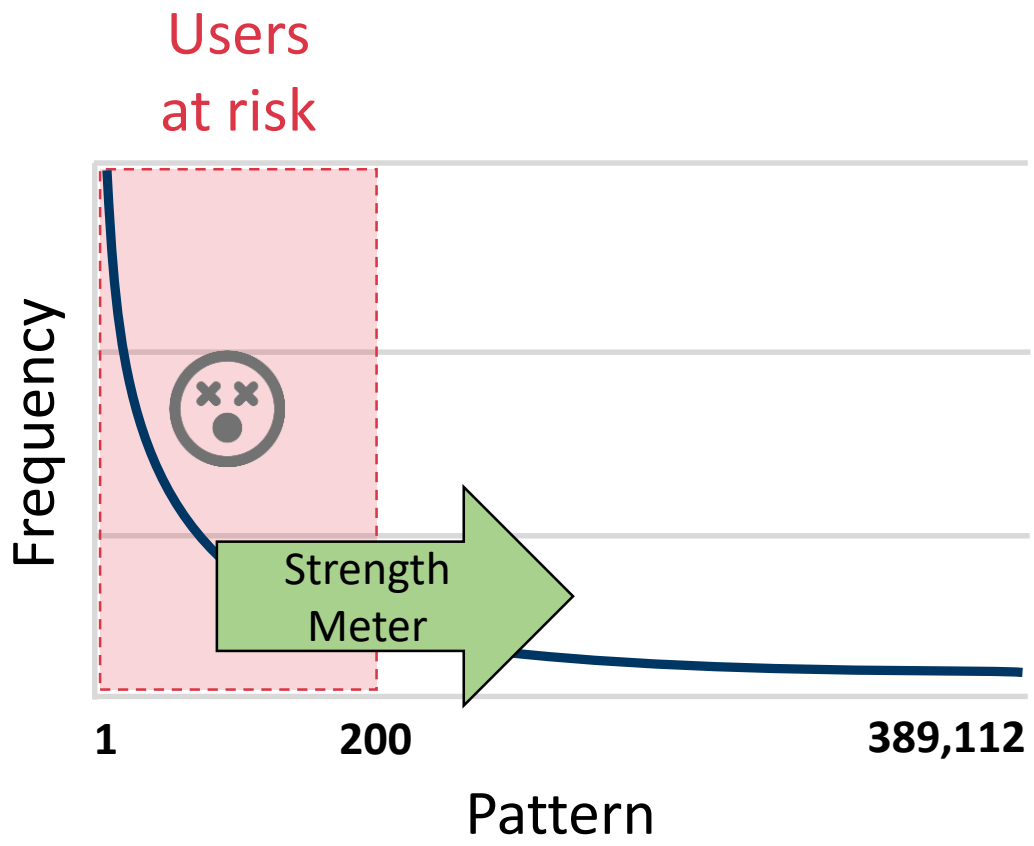
	Android 6	Android 7-9
100 Guesses	45m	10h 45m
200 Guesses	**1h 35m**	**67d 2h 45m**
300 Guesses	2h 25m	167d 2h 45m

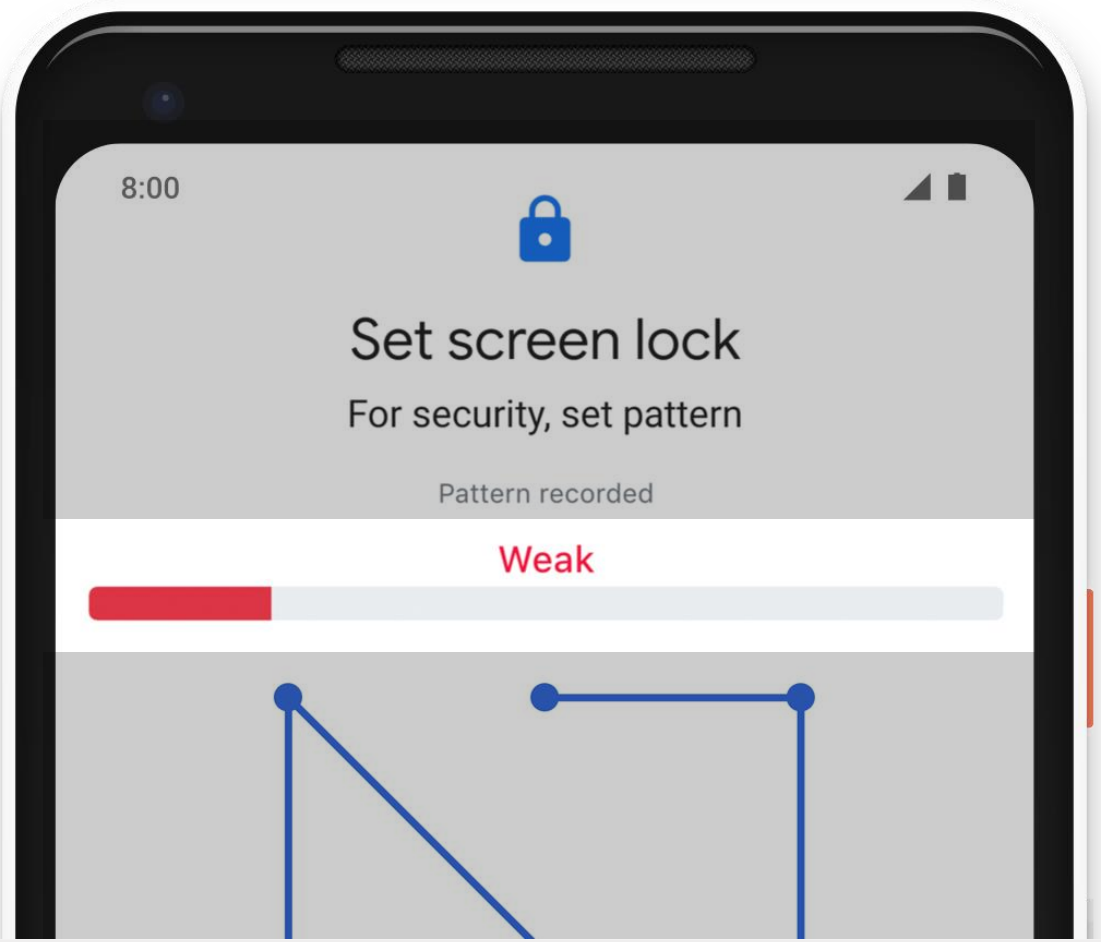


*Apple iOS only allows 10 guesses!

Strength Meter

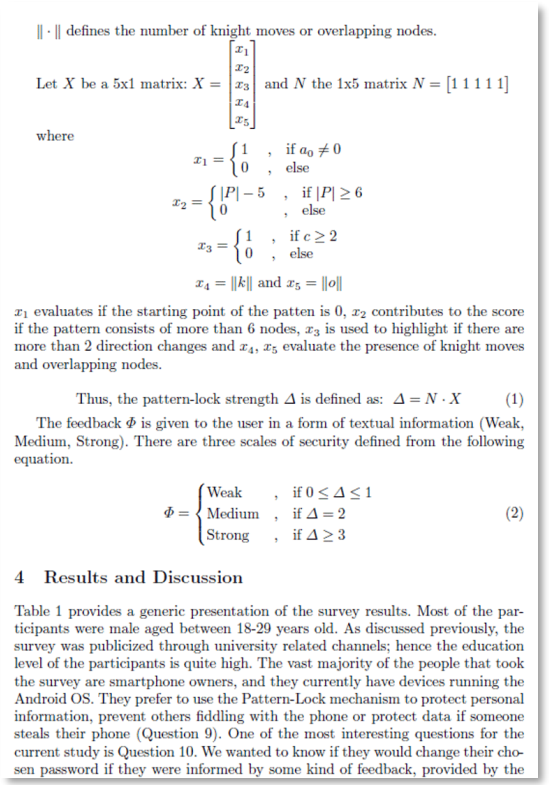
Support users while choosing their secret





Pattern Meter Proposals

2014



$\|\cdot\|$ defines the number of knight moves or overlapping nodes.

Let X be a 5x1 matrix: $X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$ and N the 1x5 matrix $N = [1\ 1\ 1\ 1\ 1]$

where

$$x_1 = \begin{cases} 1 & , \text{ if } a_0 \neq 0 \\ 0 & , \text{ else} \end{cases}$$

$$x_2 = \begin{cases} |P| - 5 & , \text{ if } |P| \geq 6 \\ 0 & , \text{ else} \end{cases}$$

$$x_3 = \begin{cases} 1 & , \text{ if } c \geq 2 \\ 0 & , \text{ else} \end{cases}$$

$$x_4 = \|k\| \text{ and } x_5 = \|o\|$$

x_1 evaluates if the starting point of the patten is 0, x_2 contributes to the score if the pattern consists of more than 6 nodes, x_3 is used to highlight if there are more than 2 direction changes and x_4, x_5 evaluate the presence of knight moves and overlapping nodes.

Thus, the pattern-lock strength Δ is defined as: $\Delta = N \cdot X$ (1)

The feedback Φ is given to the user in a form of textual information (Weak, Medium, Strong). There are three scales of security defined from the following equation.

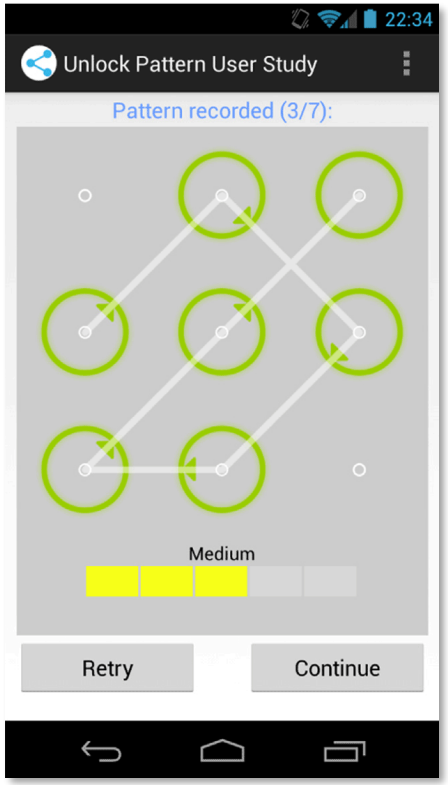
$$\Phi = \begin{cases} \text{Weak} & , \text{ if } 0 \leq \Delta \leq 1 \\ \text{Medium} & , \text{ if } \Delta = 2 \\ \text{Strong} & , \text{ if } \Delta \geq 3 \end{cases} \quad (2)$$

4 Results and Discussion

Table 1 provides a generic presentation of the survey results. Most of the participants were male aged between 18-29 years old. As discussed previously, the survey was publicized through university related channels; hence the education level of the participants is quite high. The vast majority of the people that took the survey are smartphone owners, and they currently have devices running the Android OS. They prefer to use the Pattern-Lock mechanism to protect personal information, prevent others fiddling with the phone or protect data if someone steals their phone (Question 9). One of the most interesting questions for the current study is Question 10. We wanted to know if they would change their chosen password if they were informed by some kind of feedback, provided by the

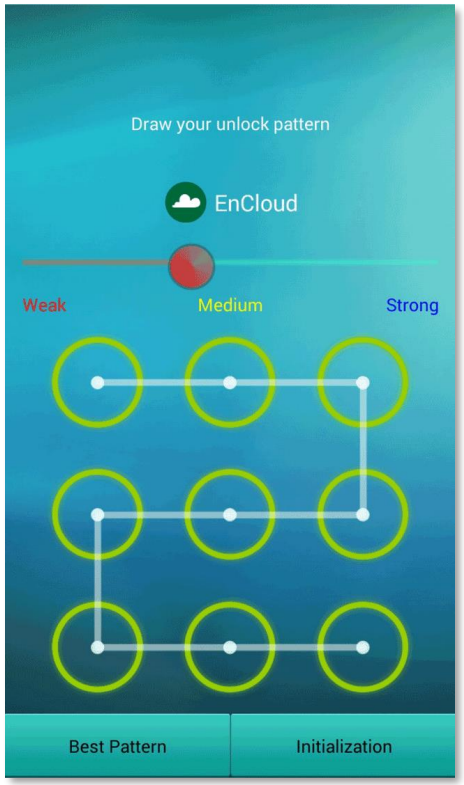
Andriotis et al.

2014



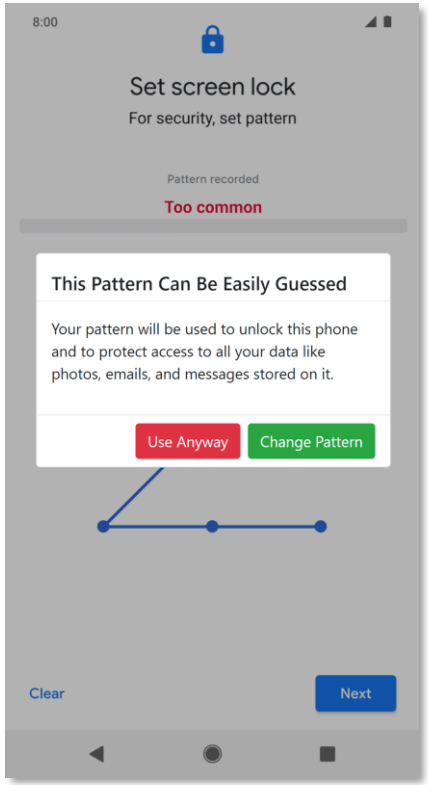
Sun et al.

2015



Song et al.

2019



Our work

What we have done!

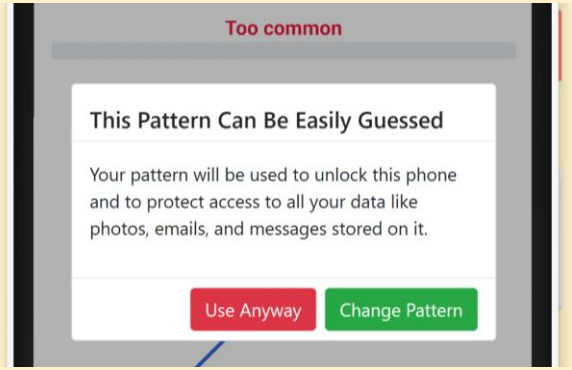
Completed:

- Implemented Android pattern strength meters proposed in the literature.
- Analyzed strength estimates for their accuracy.

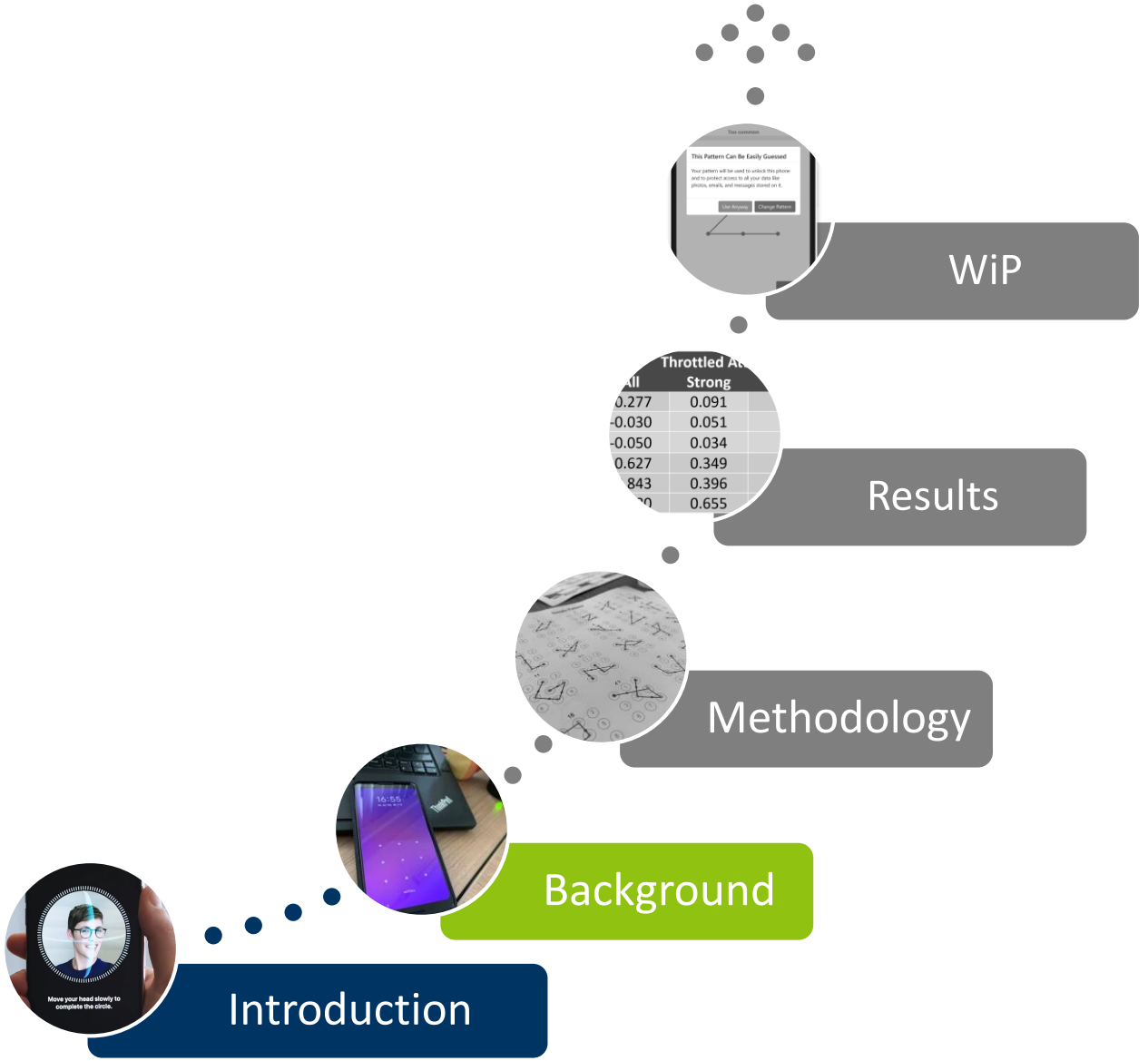
			Throttled Attacker – Top 200		
	Meter	Feature	All	Strong	Medium	Weak
	Andriotis et al.	V	0.3	0.1	0.2	0.4
	Sun et al.	V	0.0	0.1	0.0	0.0
	Song et al.	V	-0.1	0.0	0.0	0.0
Our work	Markov – Generic	P	0.6	0.3	0.6	0.7
Our work	Markov – Multi	P	0.8	0.4	0.8	0.8
Our work	Markov – Top 20	P+T	0.9	0.7	0.9	0.9

Work in Progress:

- Run user study to evaluate a *trained placebo* meter.
  - **Q1**: Importance of accuracy?
  - **Q2**: How to not waste effort in the “don’t care” region?



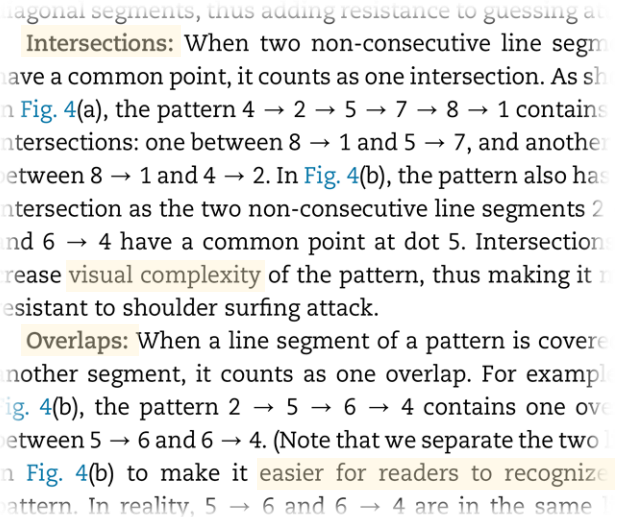
Outline


This Pattern Can Be Easily Guessed
WiP
Strong
0.091
0.051
0.034
0.349
0.396
0.655
Results
Methodology
Move your head slowly to complete the circle.
Background
Introduction

Pattern Strength Estimation

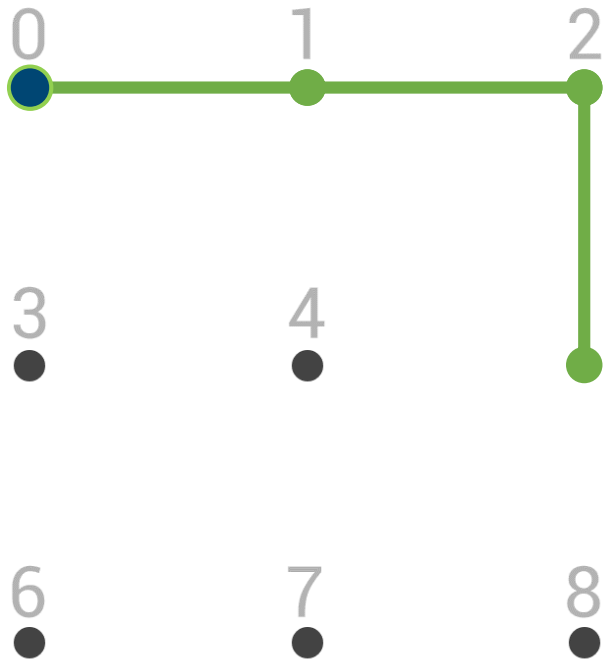
1. Based on visual features
   - Basic heuristic rules; At best, based on a user study
   - Approach known from LUDS meters

Andriotis/Sun/Song


agonal segments, thus adding resistance to guessing a
Intersections: When two non-consecutive line segm
ave a common point, it counts as one intersection. As sh
n Fig. 4(a), the pattern 4 → 2 → 5 → 7 → 8 → 1 contains
ntersections: one between 8 → 1 and 5 → 7, and anothe
etween 8 → 1 and 4 → 2. In Fig. 4(b), the pattern also has
ntersection as the two non-consecutive line segments 2
nd 6 → 4 have a common point at dot 5. Intersection
rease visual complexity of the pattern, thus making it
esistant to shoulder surfing attack.
Overlaps: When a line segment of a pattern is covere
nother segment, it counts as one overlap. For exampl
ig. 4(b), the pattern 2 → 5 → 6 → 4 contains one ove
etween 5 → 6 and 6 → 4. (Note that we separate the two
n Fig. 4(b) to make it easier for readers to recognize
attern. In reality, 5 → 6 and 6 → 4 are in the same

Visual Features

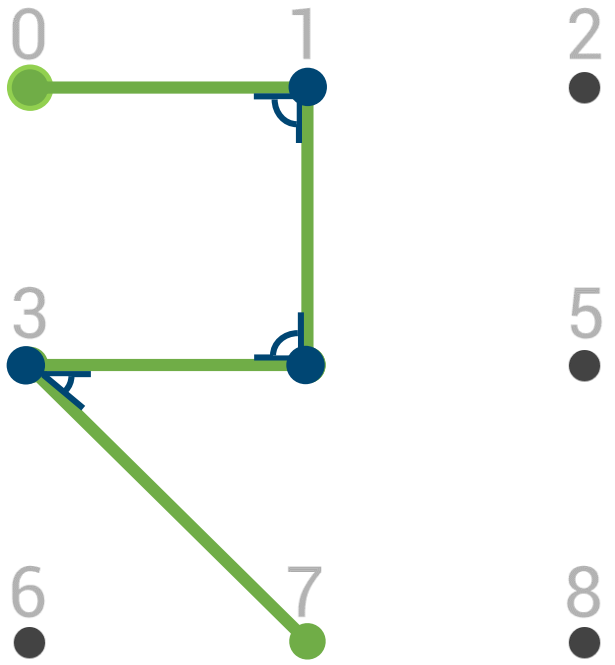
- Length
- **Starting Node**
- Direction Changes
- Knight Moves
- Overlapping Nodes
- Overlapping Segments
- Intersections
- Intersection (Restricted)
- …



0.1.2.5 → Node “0”

Visual Features

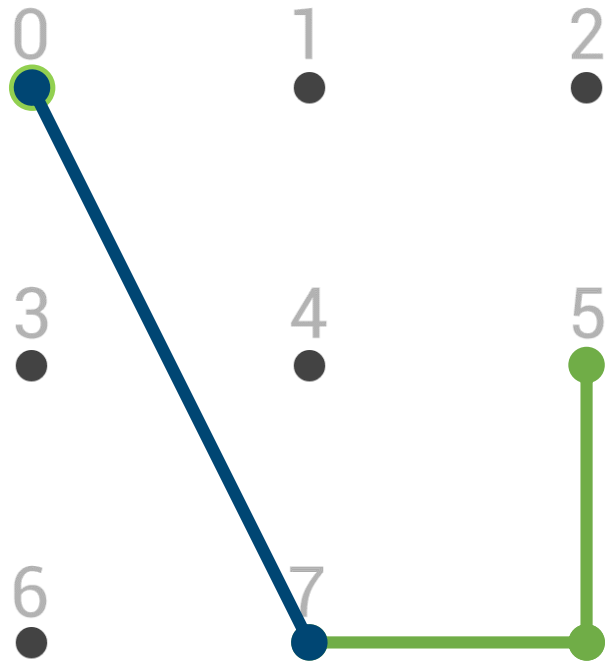
- Length
- Starting Node
- **Direction Changes**
- Knight Moves
- Overlapping Nodes
- Overlapping Segments
- Intersections
- Intersection (Restricted)
- …



0.1.4.3.7 → 3x Changes

Visual Features

- Length
- Starting Node
- Direction Changes
- **Knight Moves**
- Overlapping Nodes
- Overlapping Segments
- Intersections
- Intersection (Restricted)
- …



0.7.8.5 → 1x Knight Move (“0.7”)

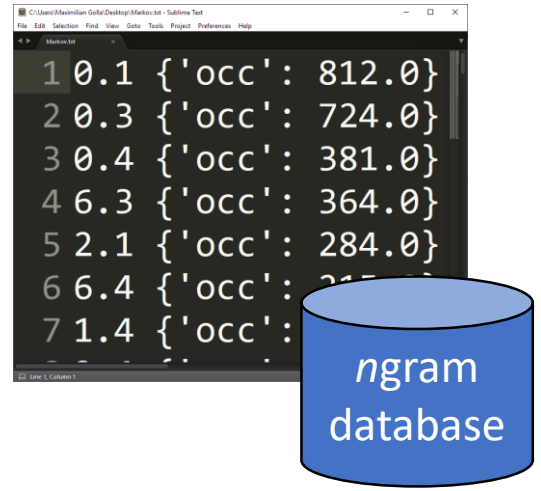
Pattern Strength Estimation

1. Based on visual features
   - Basic heuristic rules; At best, based on a user study
   - Approach known from LUDS meters

Andriotis/Sun/Song

2. Based on probabilistic model
   - Some transitions occur more often than others (based on prior state)
   - Requires large enough training corpus

Our work



Datasets

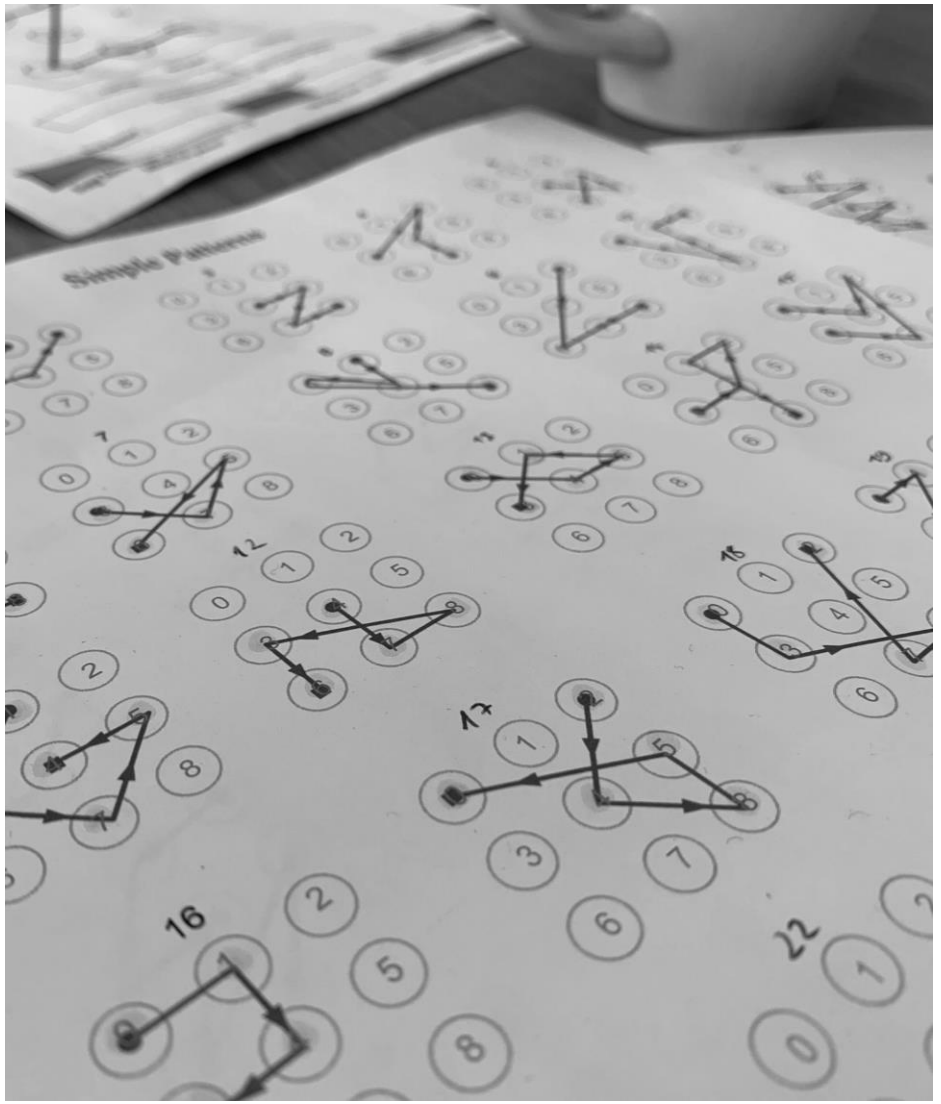
4,637 patterns

Merged from 4 different user studies:

- Aviv et. al [3] in 2015
- Løge et. al [28] in 2016
- Uellenbeck et. al [37] in 2013
- Von Zezschwitz et al. [45] in 2016

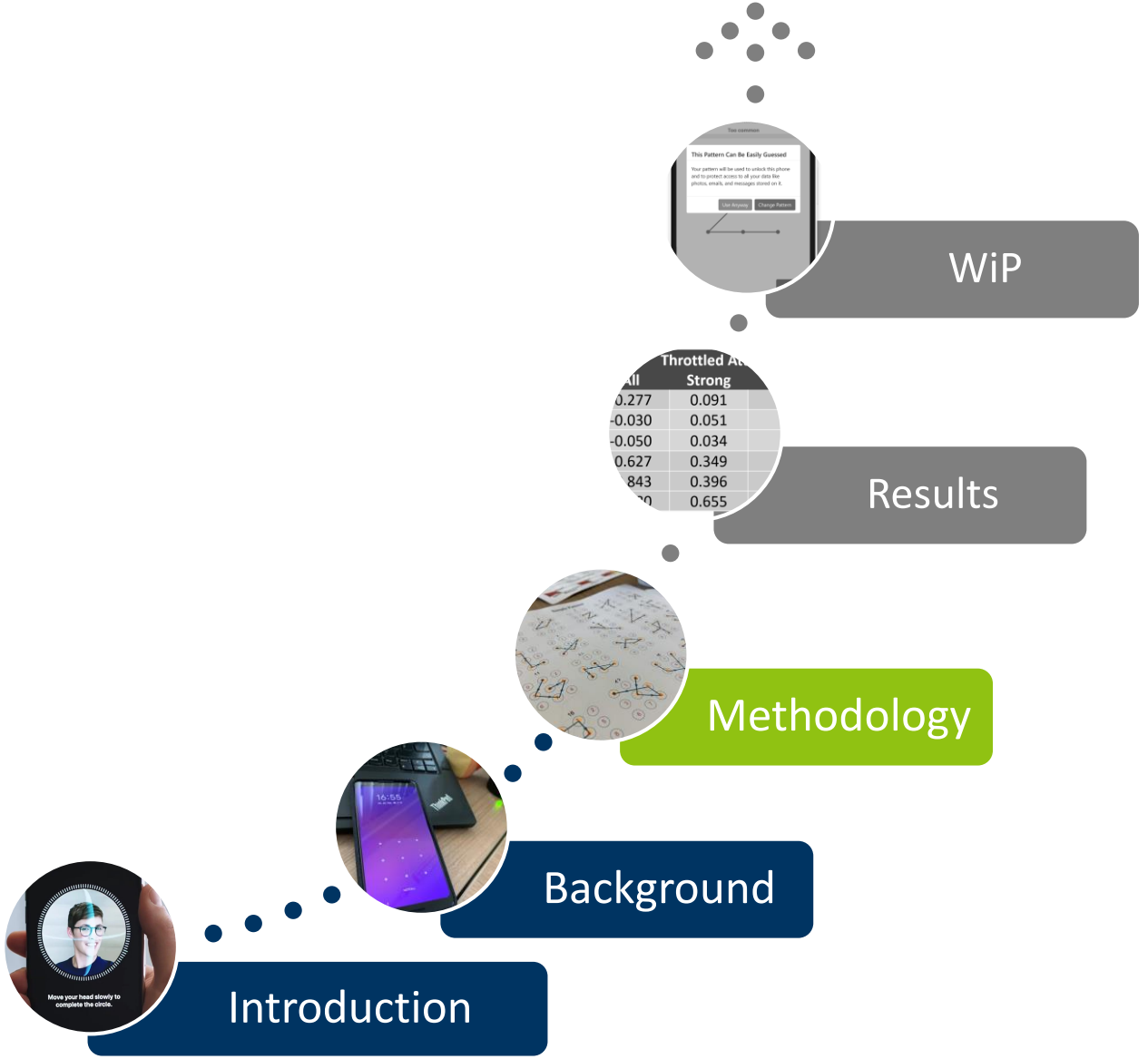
Divided into three groups:

- Weak, Medium, and Strong



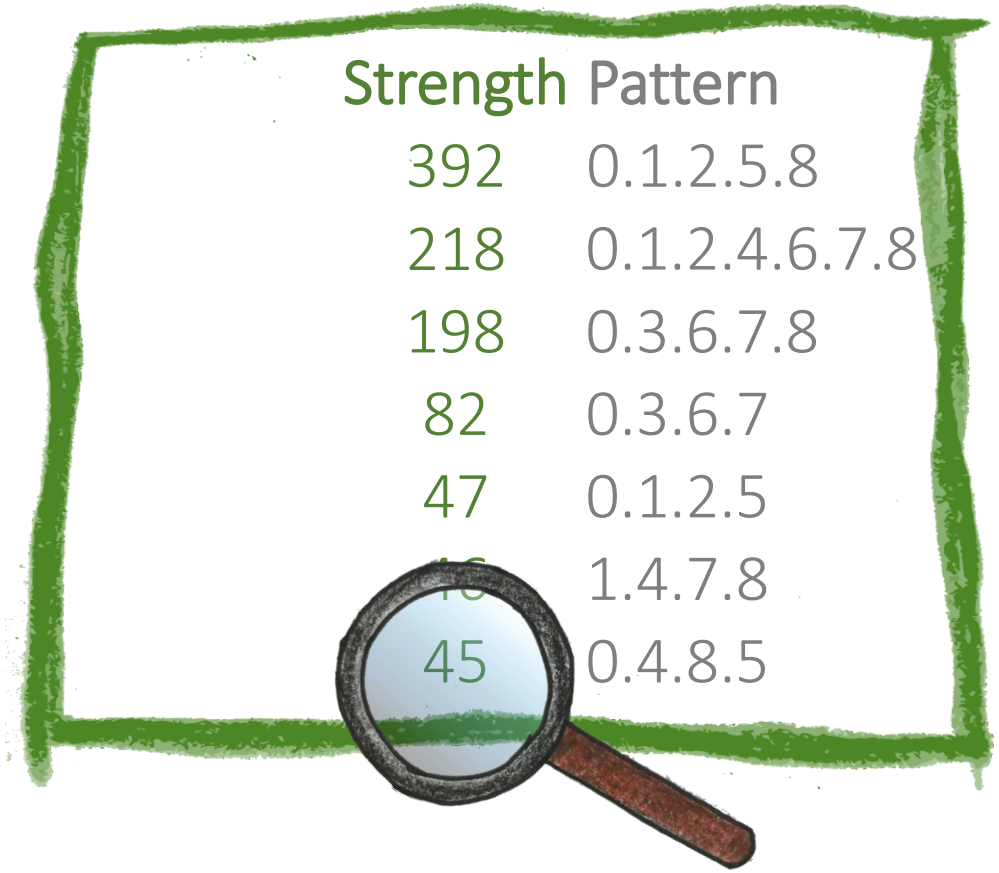
More Info: Adam J. Aviv and Markus Dürmuth, "A Survey of Collection Methods and Cross-Data Set Comparison of Android Unlock Patterns", CoRR, vol. abs/1811.10548, Nov. 2018.

Outline

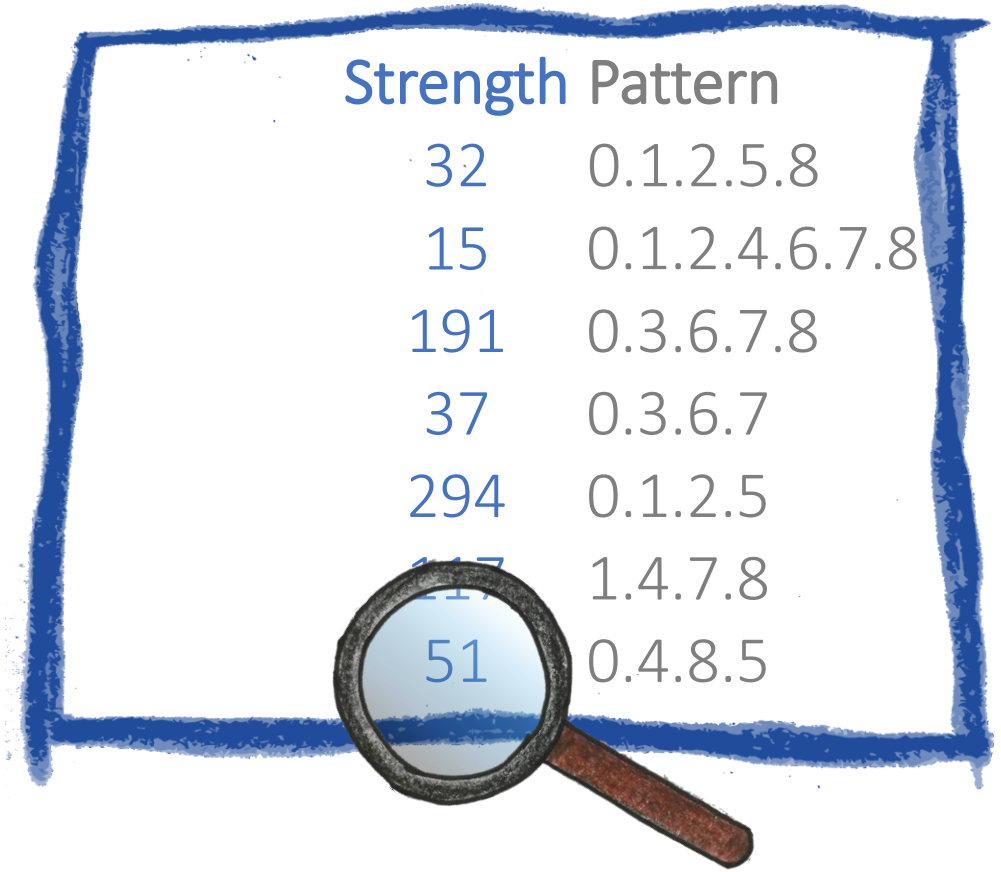

This Pattern Can Be Easily Guessed
WiP
Strong
0.091
0.051
0.034
0.349
0.396
0.655
Results
Methodology
Background
Move your head slowly to complete the circle.
Introduction

Measuring Accuracy

Reference



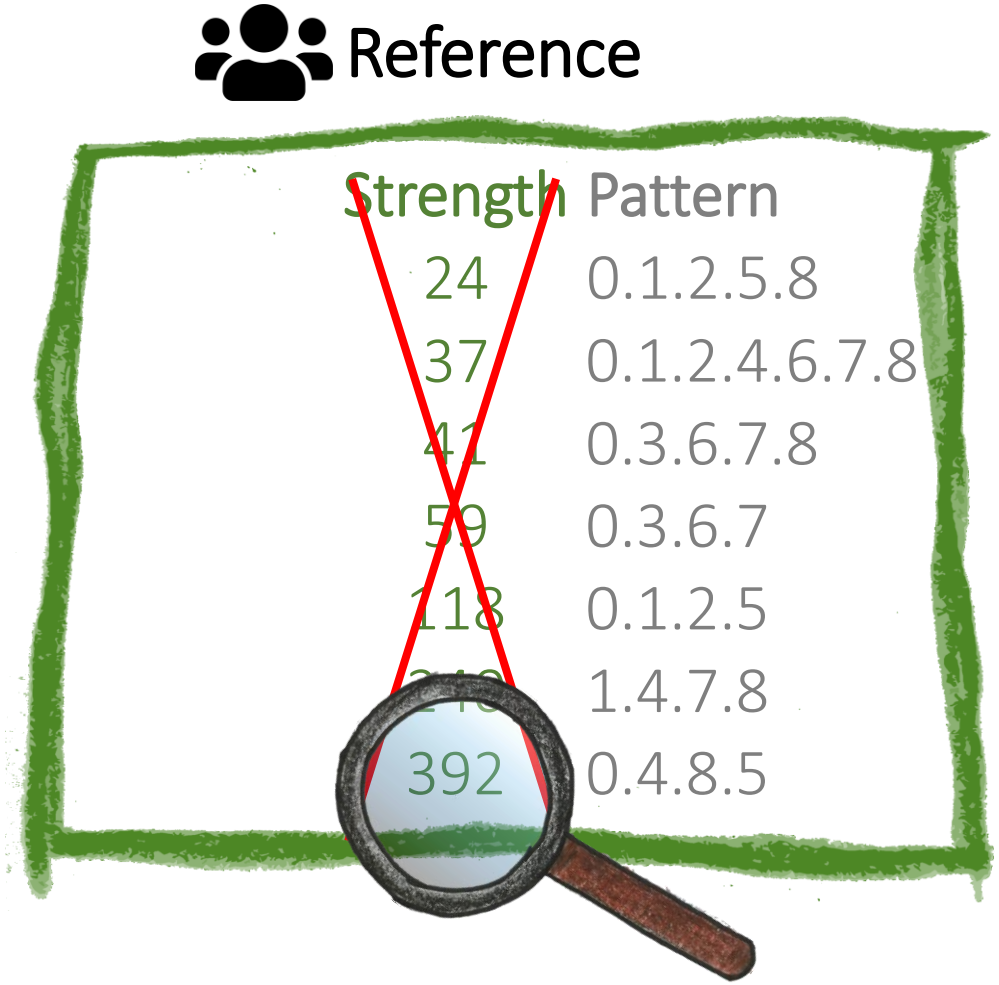
Strength Meter

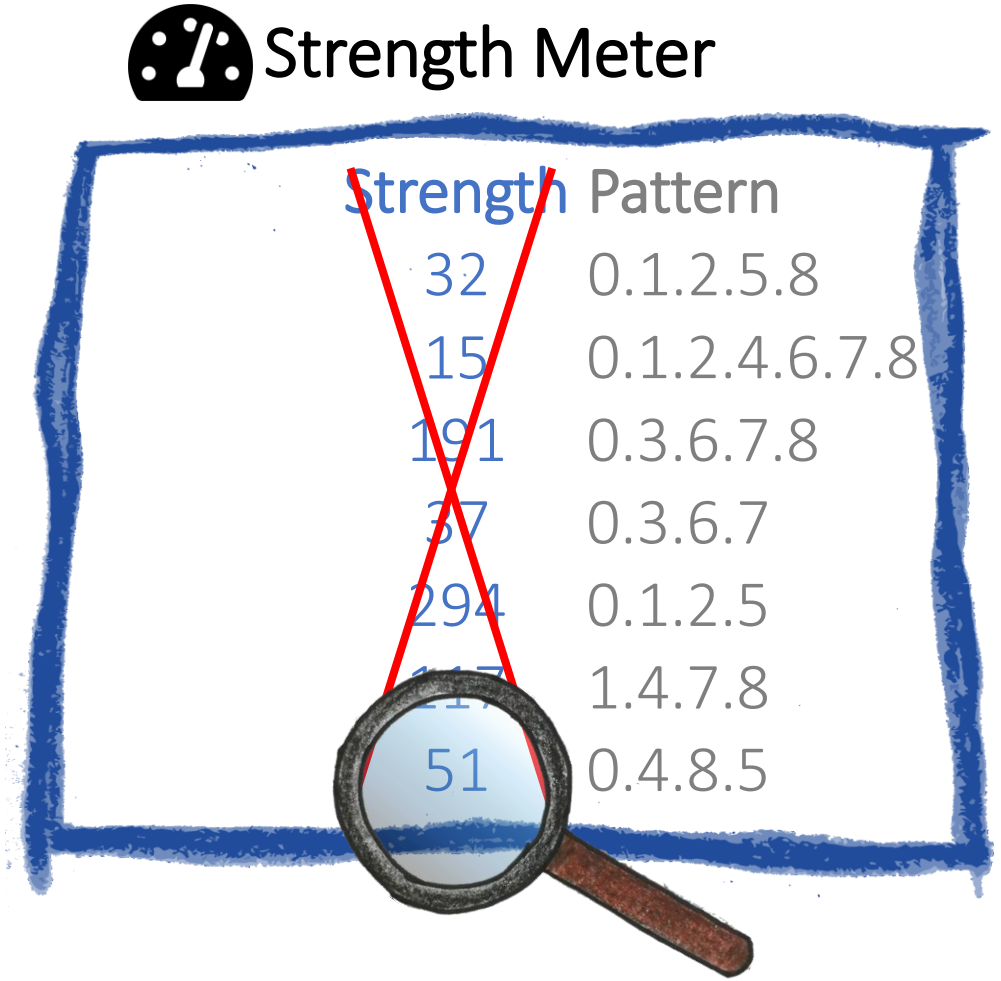


More Info: Maximilian Golla and Markus Dürmuth, "On the Accuracy of Password Strength Meters", ACM Conference on Computer and Communications Security, Oct. 2018.

RUB

Measuring Accuracy





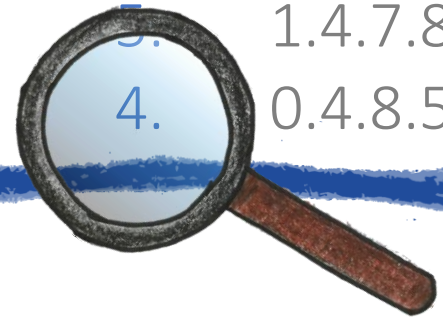
Measuring Accuracy

Reference

Rank	Pattern
1.	0.1.2.5.8
2.	0.1.2.4.6.7.8
3.	0.3.6.7.8
4.	0.3.6.7
5.	0.1.2.5
6.	1.4.7.8
7.	0.4.8.5

Strength Meter

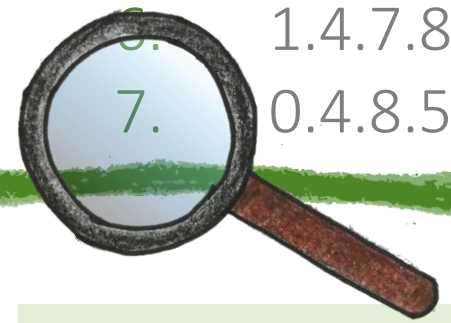
Rank	Pattern
2.	0.1.2.5.8
1.	0.1.2.4.6.7.8
6.	0.3.6.7.8
3.	0.3.6.7
7.	0.1.2.5
5.	1.4.7.8
4.	0.4.8.5



Measuring Accuracy

Reference

Weight	Rank	Pattern
95	1.	0.1.2.5.8
91	2.	0.1.2.4.6.7.8
90	3.	0.3.6.7.8
86	4.	0.3.6.7
70	5.	0.1.2.5
41	6.	1.4.7.8
40	7.	0.4.8.5



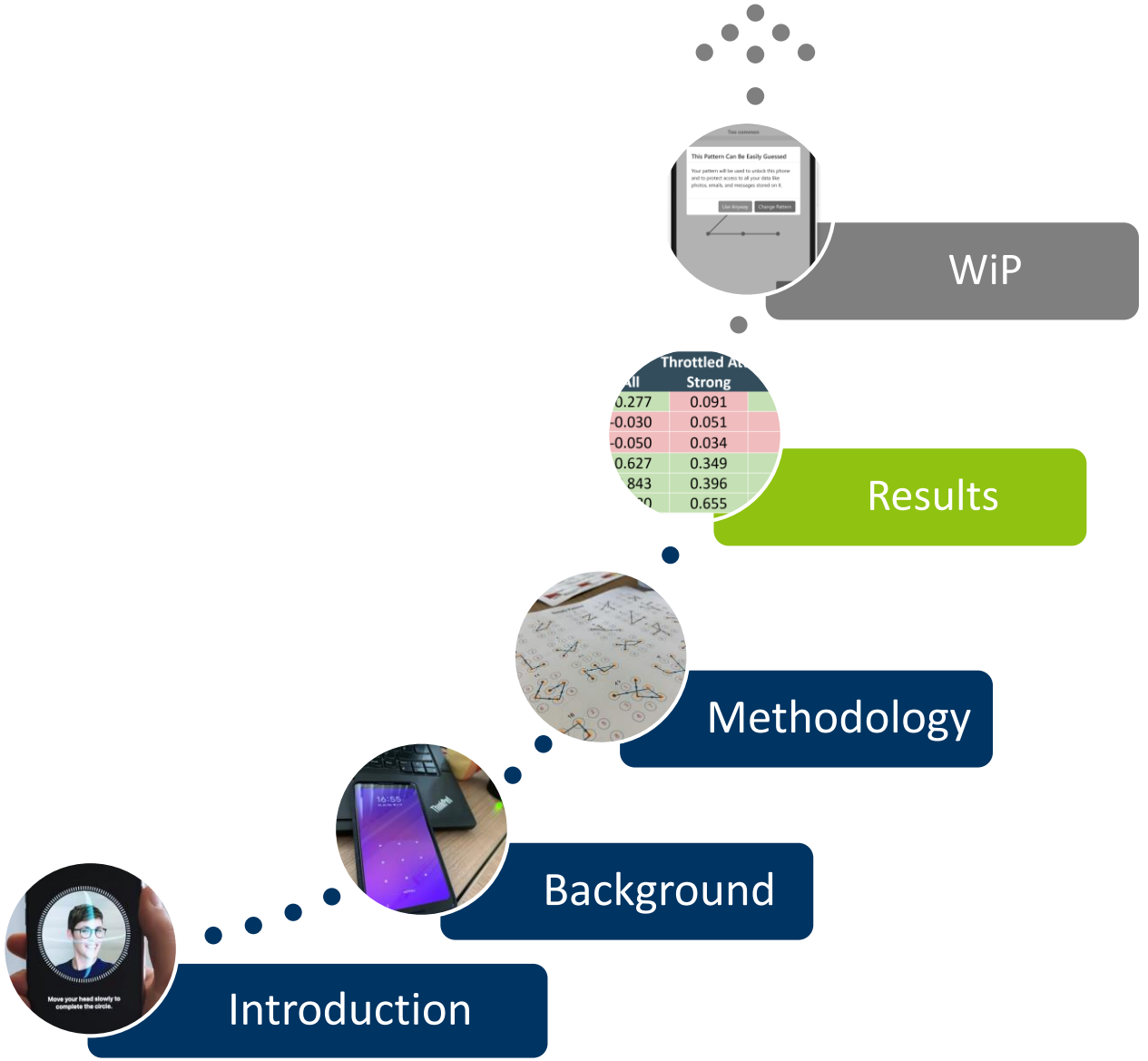
Strength Meter

Weight	Rank	Pattern
95	2.	0.1.2.5.8
91	1.	0.1.2.4.6.7.8
90	6.	0.3.6.7.8
86	3.	0.3.6.7
70	7.	0.1.2.5
41	5.	1.4.7.8
40	4.	0.4.8.5

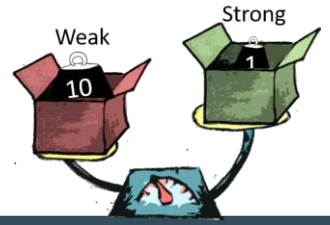
Recommendation

- weighted and ranked metrics → weighted Spearman correlation

Outline


This Pattern Can Be Easily Guessed
Strong
0.091
0.051
0.034
0.349
0.396
0.655
WiP
Results
Methodology
Background
Introduction

Weighted Spearman Correlation – Strength Meters

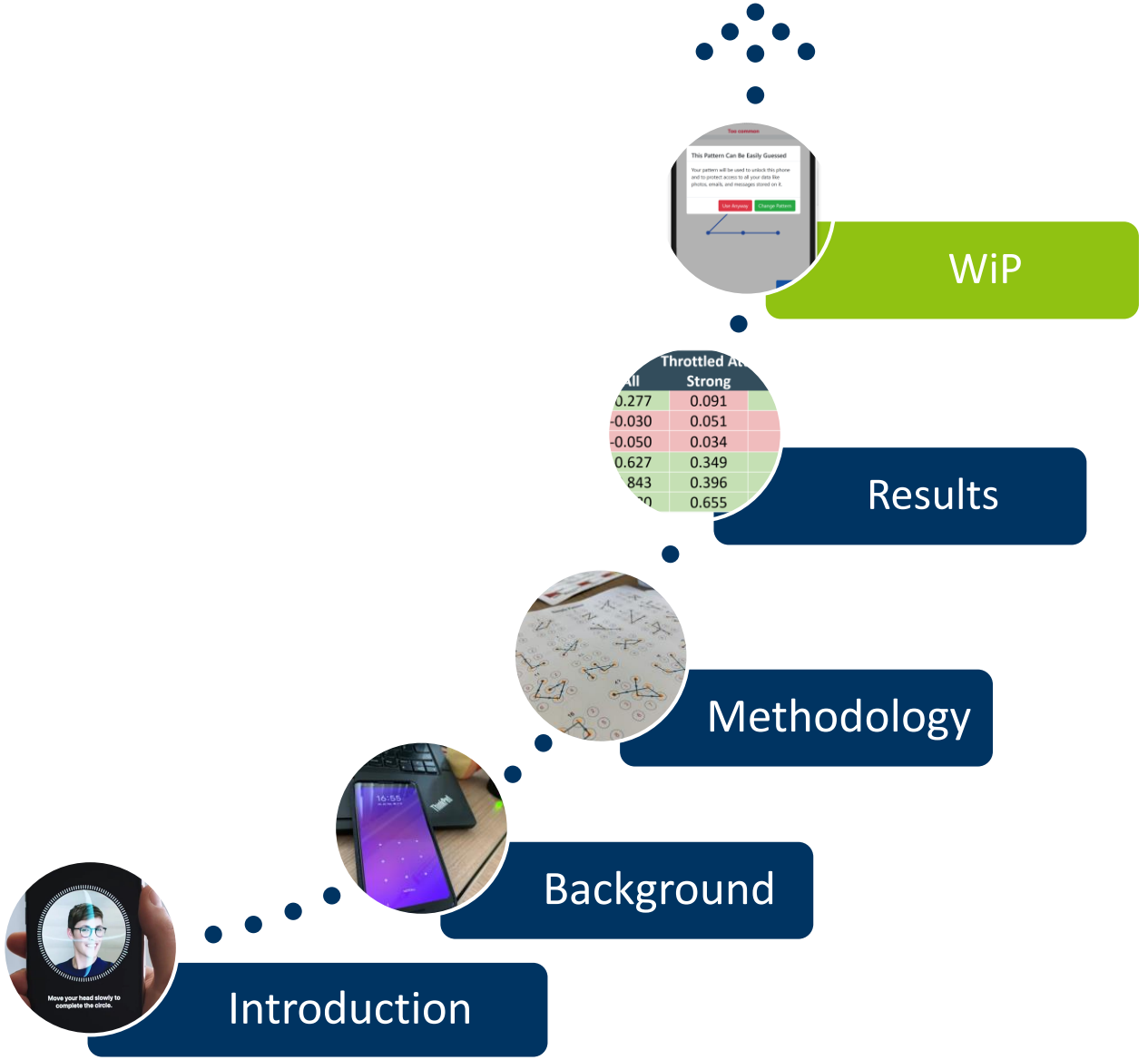


	Meter	Feature	Throttled Attacker – Top 200			
			All	Strong	Medium	Weak
	Andriotis et al.	V	0.3	0.1	0.2	0.4
	Sun et al.	V	0.0	0.1	0.0	0.0
	Song et al.	V	-0.1	0.0	0.0	0.0
Our work	Markov – Generic	P	0.6	0.3	0.6	0.7
Our work	Markov – Multi	P	0.8	0.4	0.8	0.8
Our work	Markov – Top 20	P+T	0.9	0.7	0.9	0.9

1.0 High positive correlation

0.0 No correlation

Outline


This Pattern Can Be Easily Guessed
Strong
0.091
0.051
0.034
0.349
0.396
0.655
Move your head slowly to complete the circle.
WiP
Results
Methodology
Background
Introduction

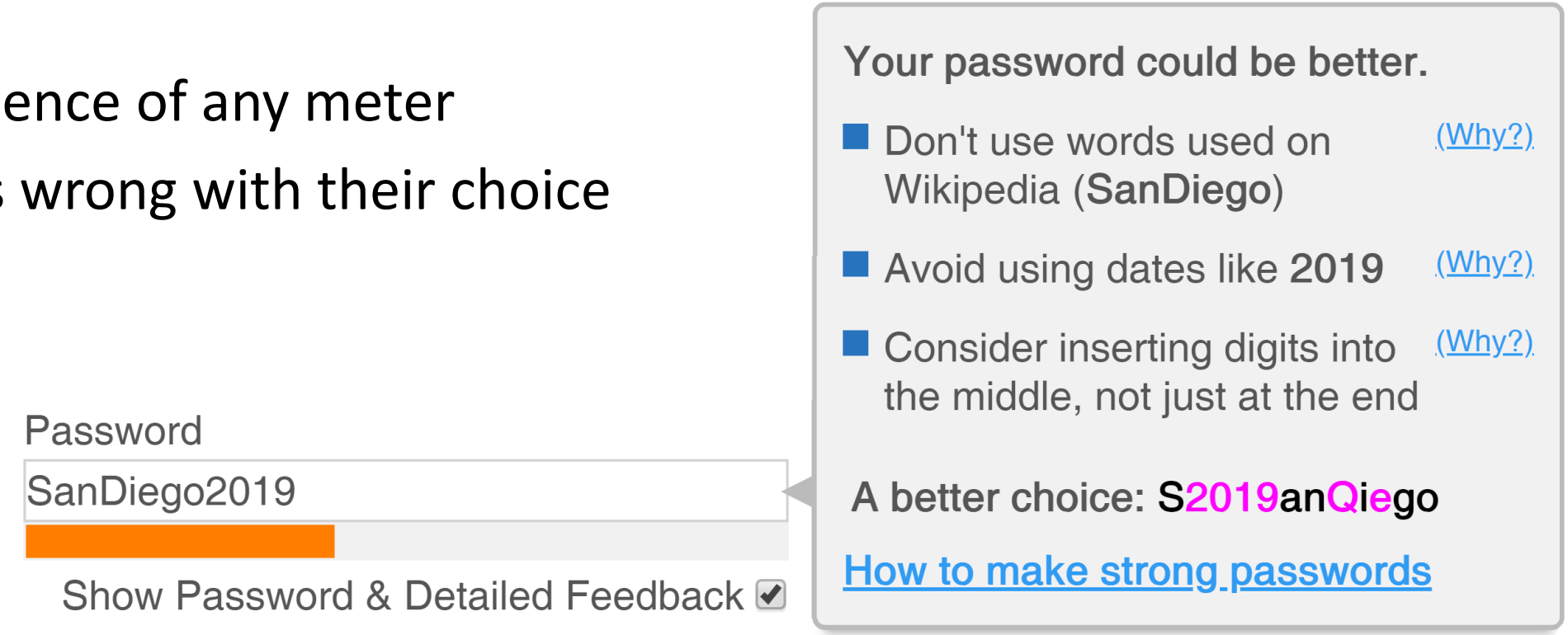
Does Accuracy Really Matter?

Accurate Meter:

- Data-driven approach (requires training)

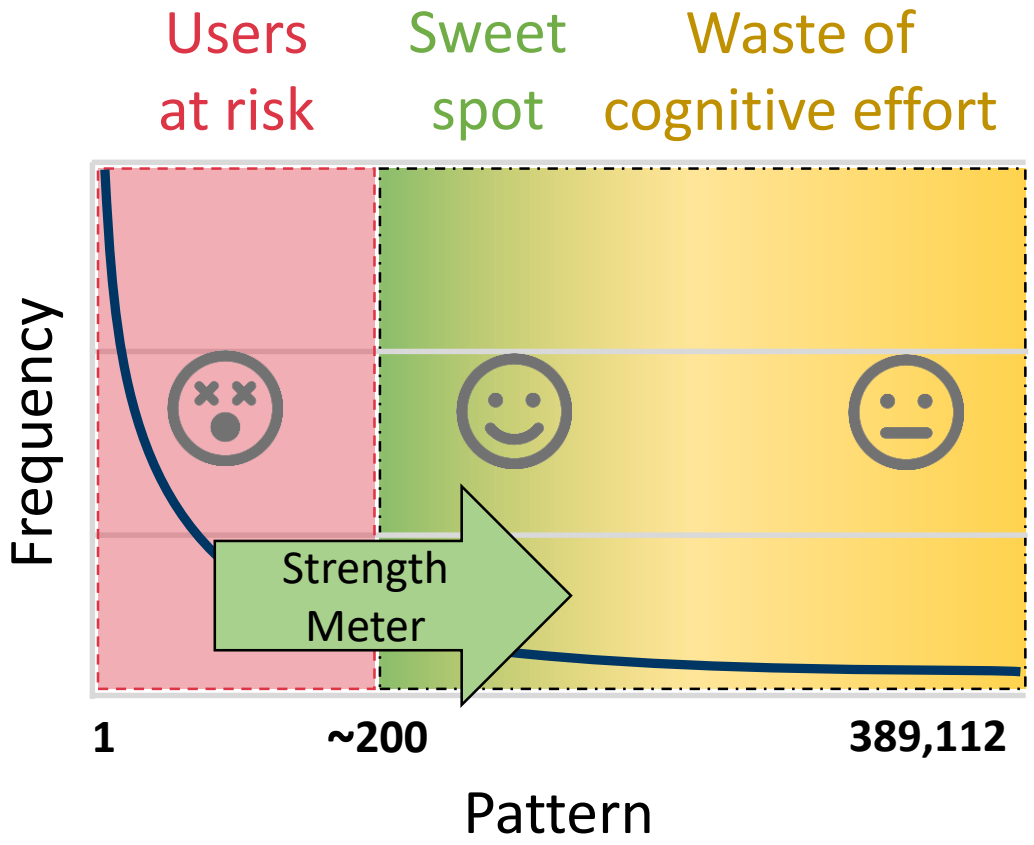
Influence Users:

- The sheer presence of any meter
- Explain what is wrong with their choice



The Mobile Setting Is Different

Throttled:
~200 guesses
Unthrottled:
Exhaustive search **feasible**

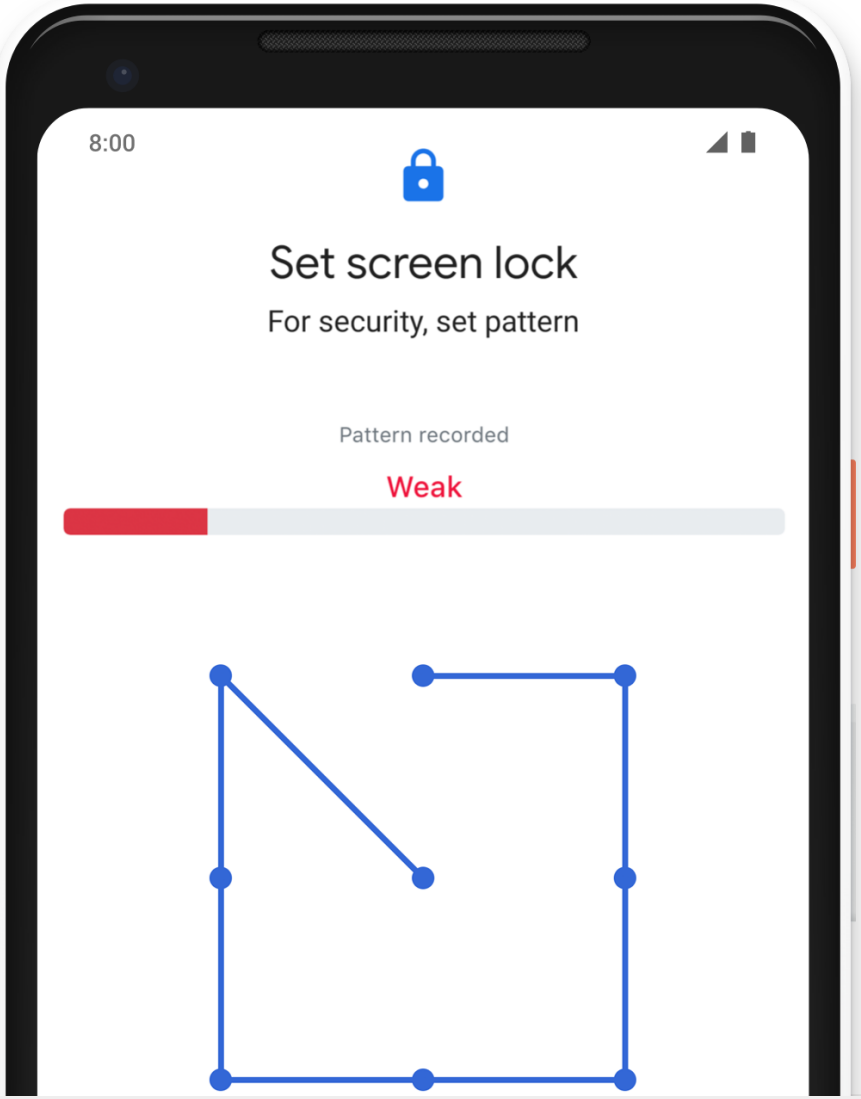


Idea: Build meter based on

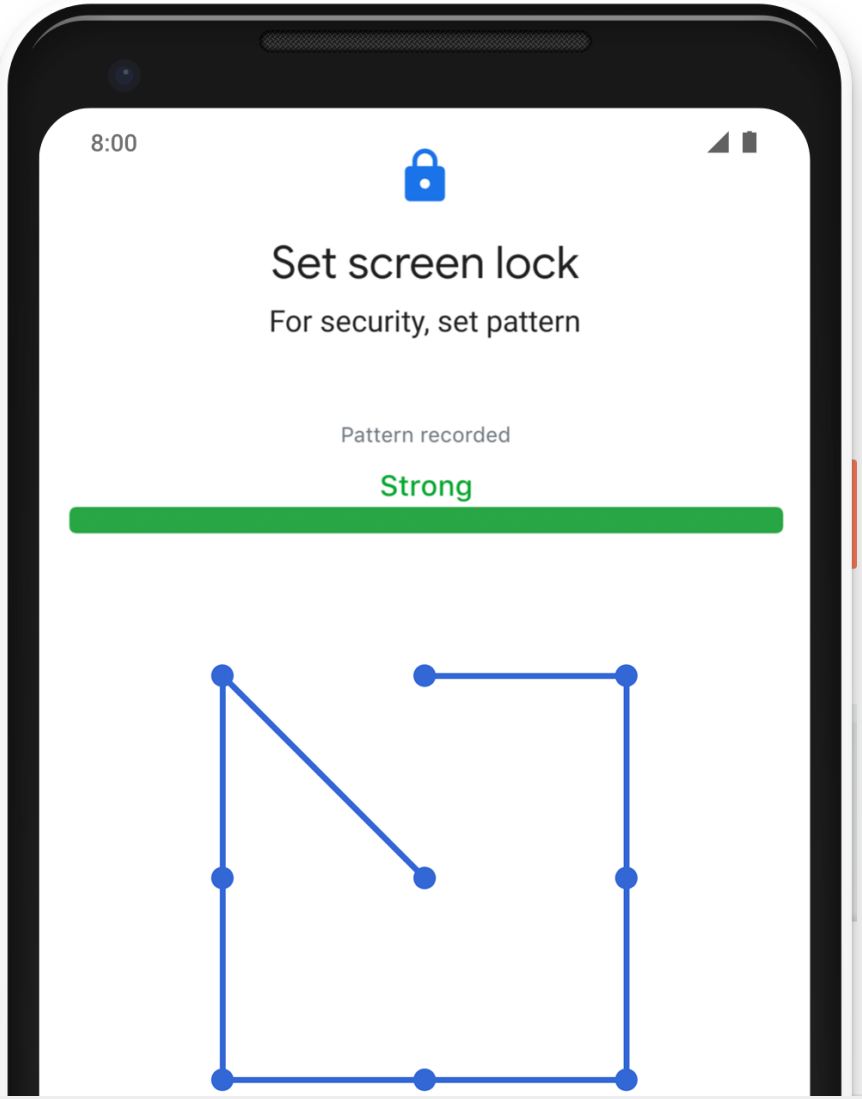
- perception of security (length driven)
- non-enforcing blacklisting (top *N*)

Pattern:
1.2.5.8.7.6.3.0.4
Rank:
240

Accurate Meter

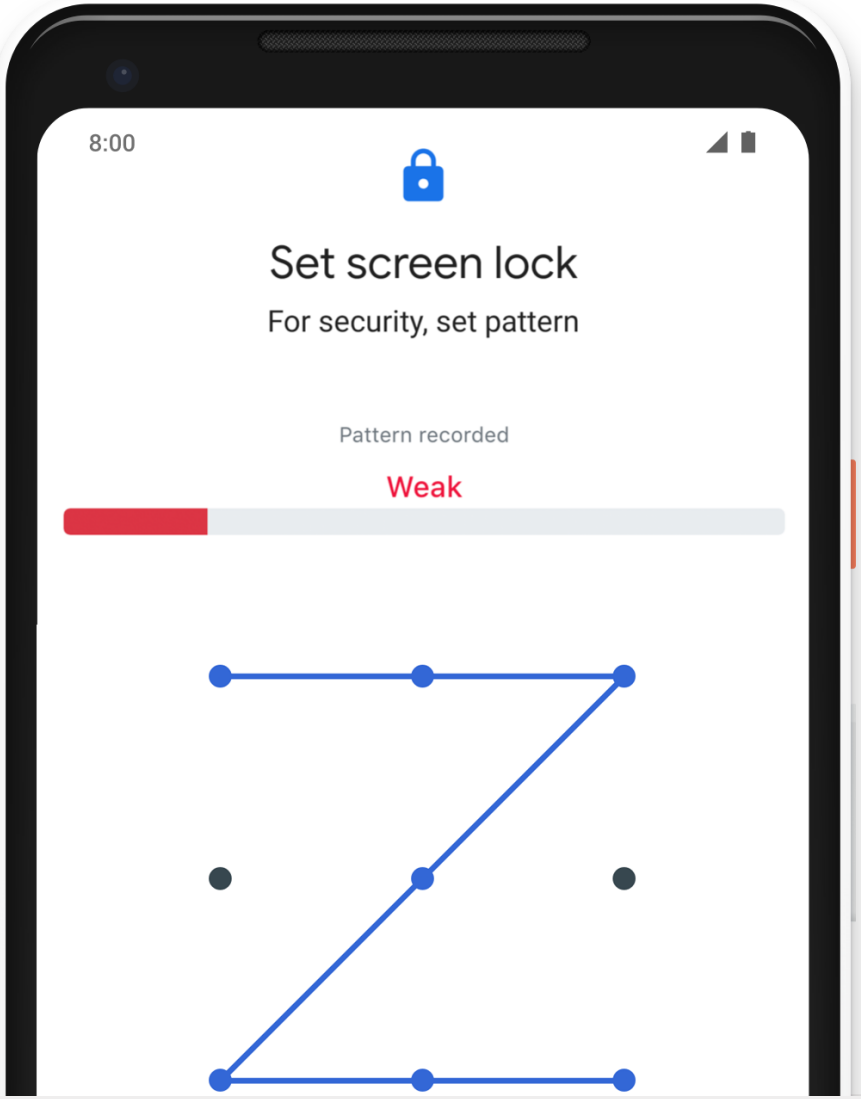


Trained Placebo Meter

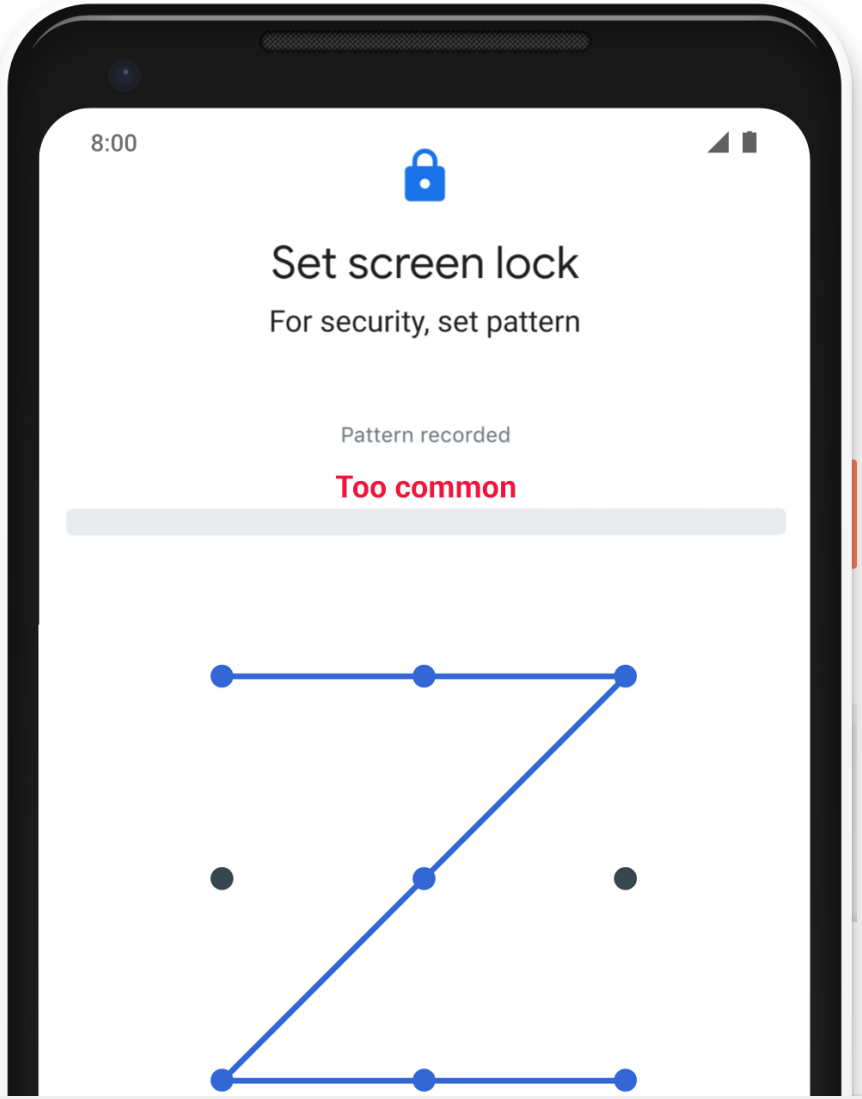


Pattern:
0.1.2.4.6.7.8
Rank:
2

Accurate Meter

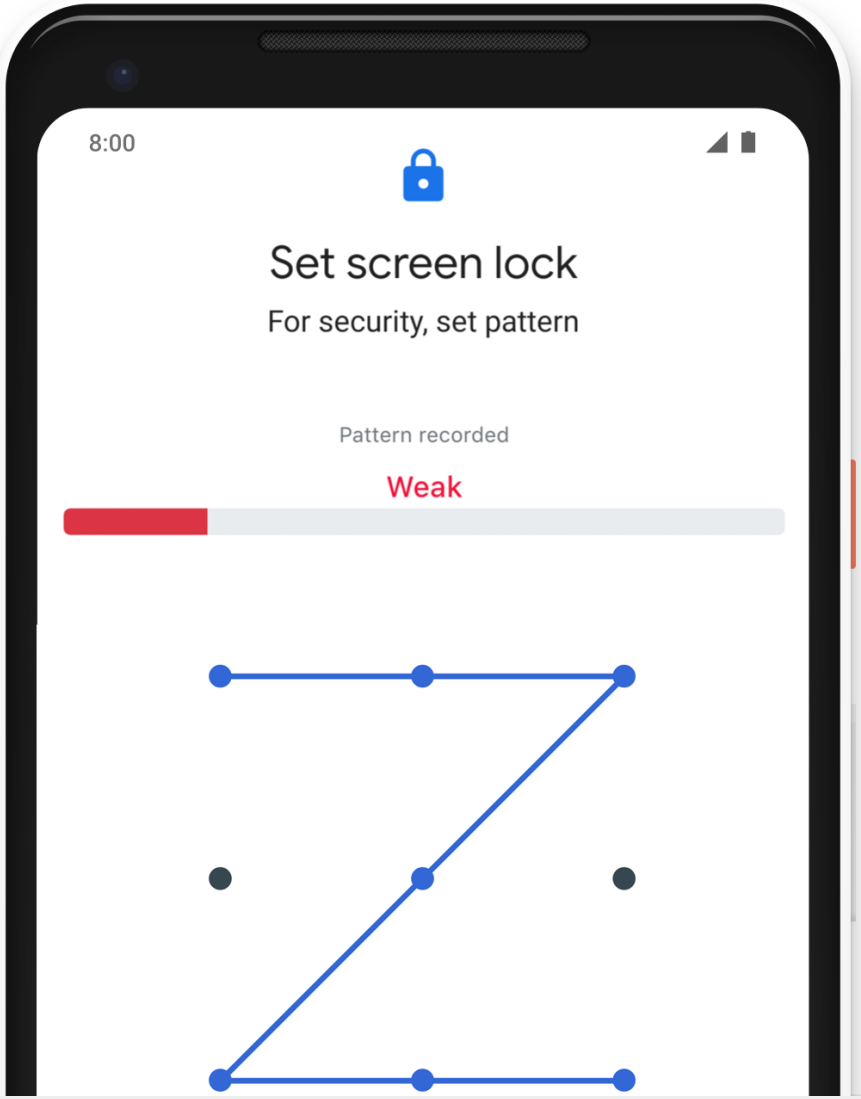


Trained Placebo Meter

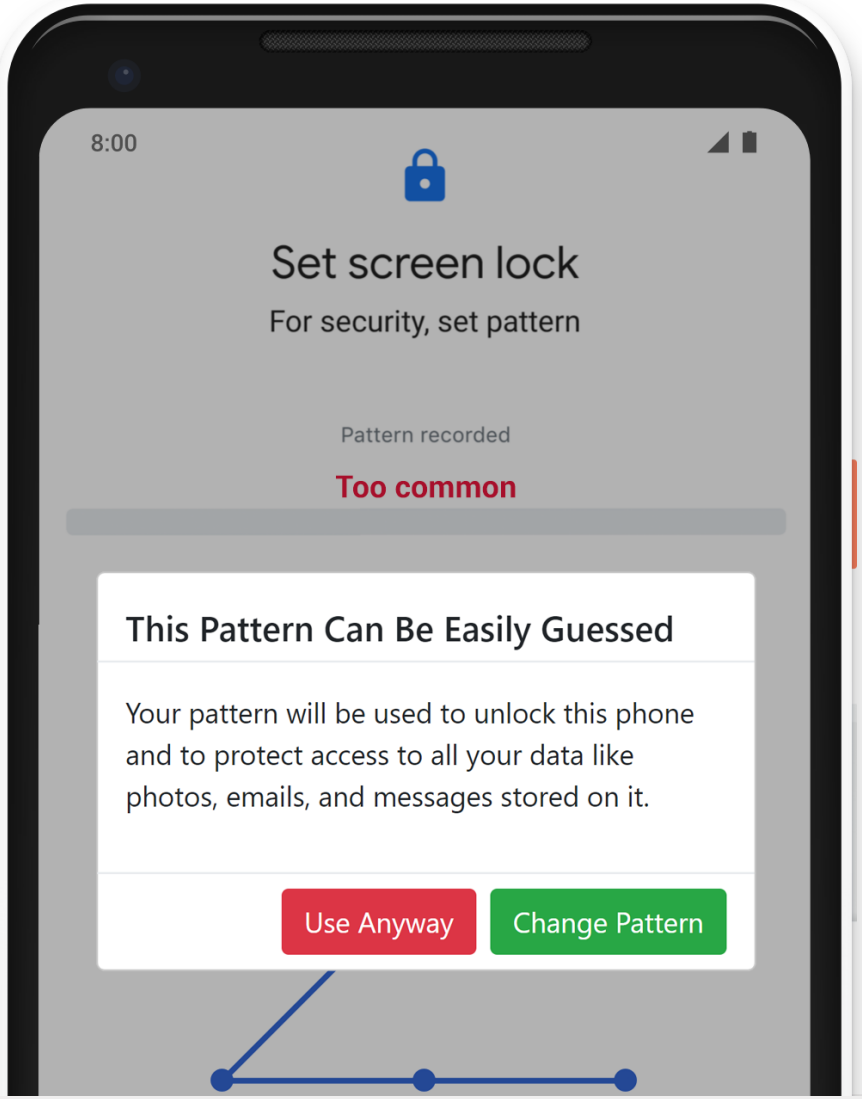


Pattern:
0.1.2.4.6.7.8
Rank:
2

Accurate Meter



Trained Placebo Meter



User Study

(Not started yet)

Play again?

In your opinion, which Bar belongs to the pattern displayed above?

Note: A strong pattern is difficult to guess by a stranger, thus protects your personal data like photos and messages.

Bar A

Weak

Bar B

Terrible

Play again?

Note: A strong pattern is difficult to guess by a stranger, thus protects your personal data like photos and messages.

Bar A

Weak

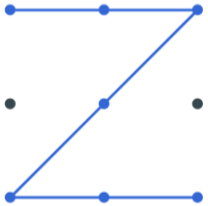
Bar B

Terrible

Please explain why you believe Bar B belongs to the pattern:

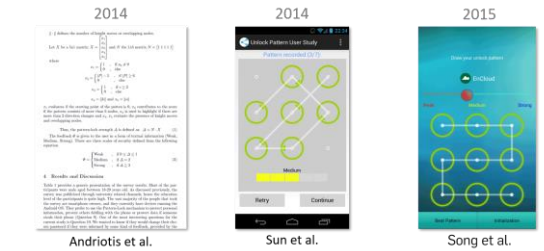
Because it is really short and somehow reminds me of the Starfleet Insignia from Star Trek.

Takeaway



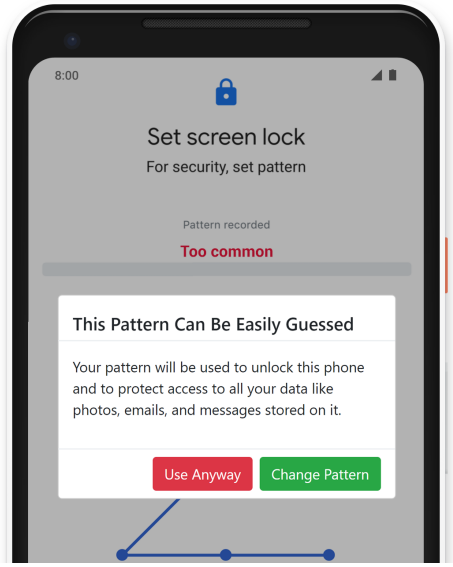
Unlock Issues

2014 Andriotis et al. | 2014 Sun et al. | 2015 Song et al.



Meter	Throttled Attacker – Top 200			
	All	Strong	Medium	Weak
Andriotis et al.	0.277	0.091	0.236	0.355
Sun et al.	-0.030	0.051	0.014	0.041
Song et al.	-0.050	0.034	0.001	0.012
Markov – Generic	0.627	0.349	0.595	0.663
Markov – Multi	0.843	0.396	0.818	0.824
Markov – Top 20	0.880	0.655	0.870	0.916

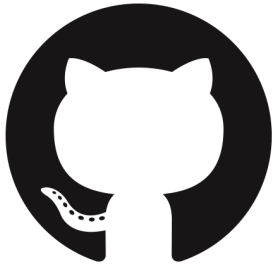
Meter Accuracy



Trained Placebo Meter

Research Artifact: Android (Unlock) Pattern Classifier

- Verifies patterns (detect data collection issues)
- Calculates scores for 3 Android strength meters
- Calculates scores for various visual features
- Support for larger grid sizes than 3x3
- Written in Kotlin (runs on Android, PC, Web)



```
$ java -jar apc.jar -p 0.1.2.5.8
```

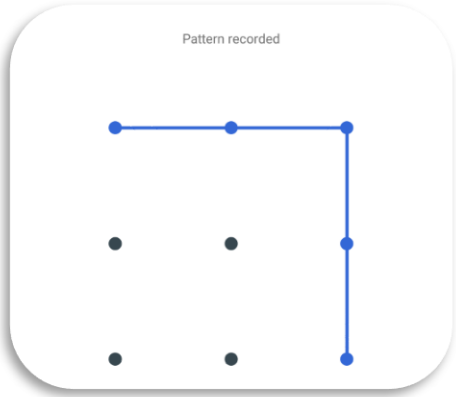
github.com/RUB-SysSec/APC

Statistical - Guessability

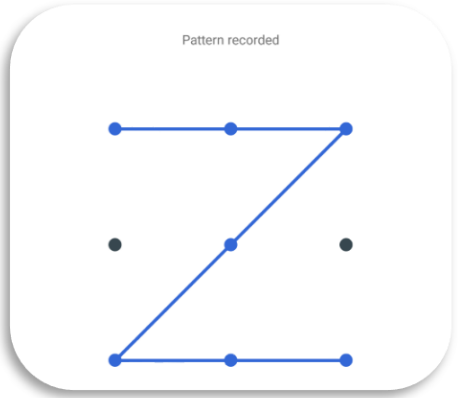
			Higher is better	Lower is better	
			Bits	Percentage	
Datasets	Samples	∅ Length	H_∞	λ_3	λ_{30}
Strong	219	6.3	6.2	3.7	22.8
Medium	2,162	5.8	5.8	4.9	21.9
Weak	2,256	5.7	5.1	7.8	30.9 👎
All	4,637	5.8	5.6	6.0	24.9
Comparison					
4-digit PIN (Amitay)	204,432	4.0	4.5	9.2	20.6
6-digit PIN (Wang)	2,758,490	6.0	3.1	12.8	17.3
PW (Melicher)	273	10.4	8.1	1.1	11.0 👍

Top 10: Android Unlock Patterns

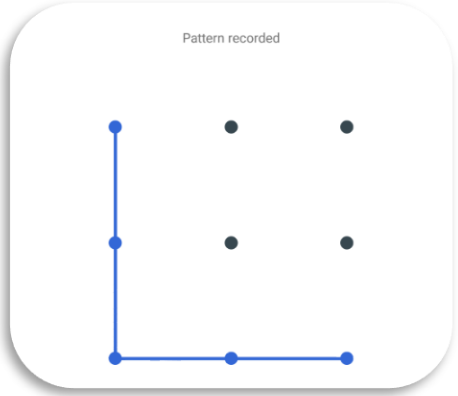
1



2



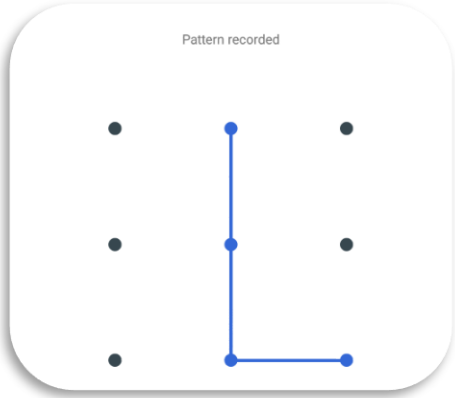
3



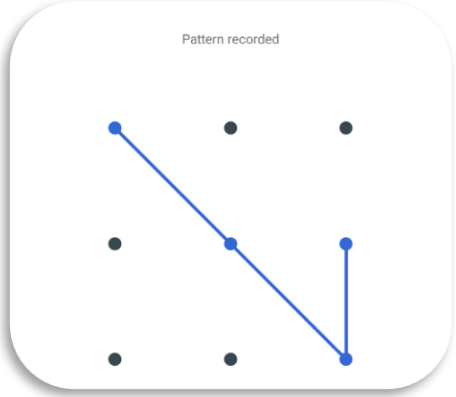
4

5

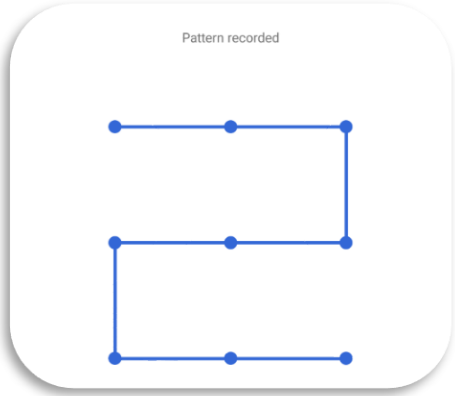
6



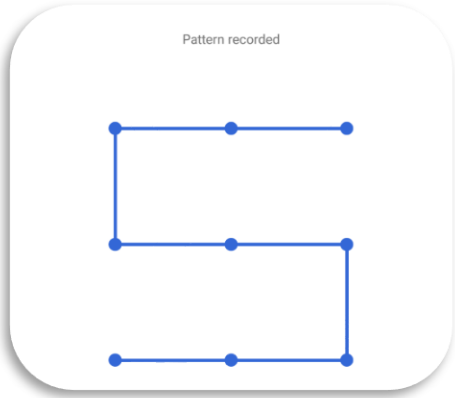
7



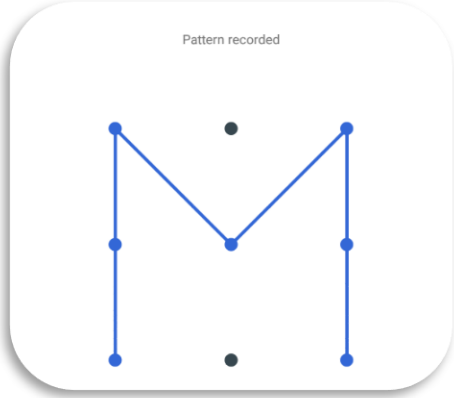
8



9

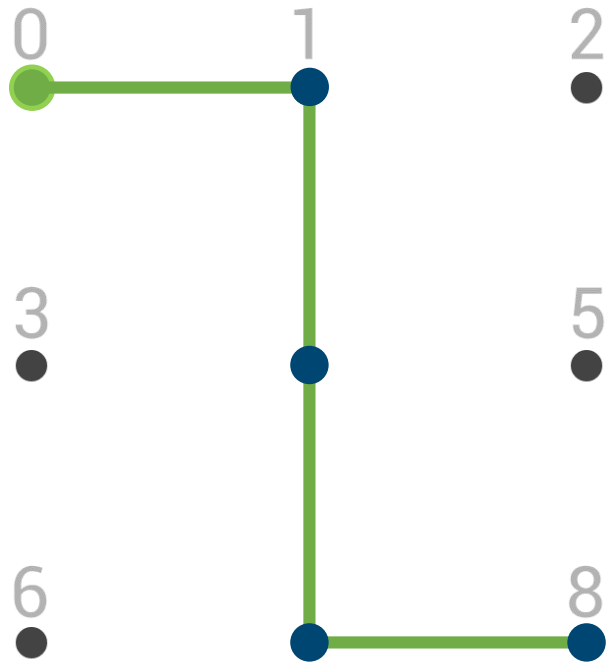


10



Visual Features

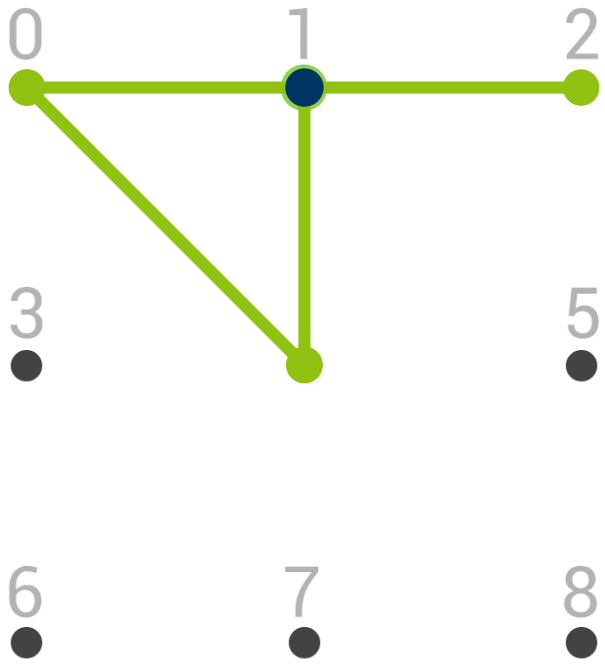
- **Length**
- Starting Node
- Direction Changes
- Knight Moves
- Overlapping Nodes
- Overlapping Segments
- Intersections
- Intersection (Restricted)
- …



0.1.4.7.8 → Length 4

Visual Features

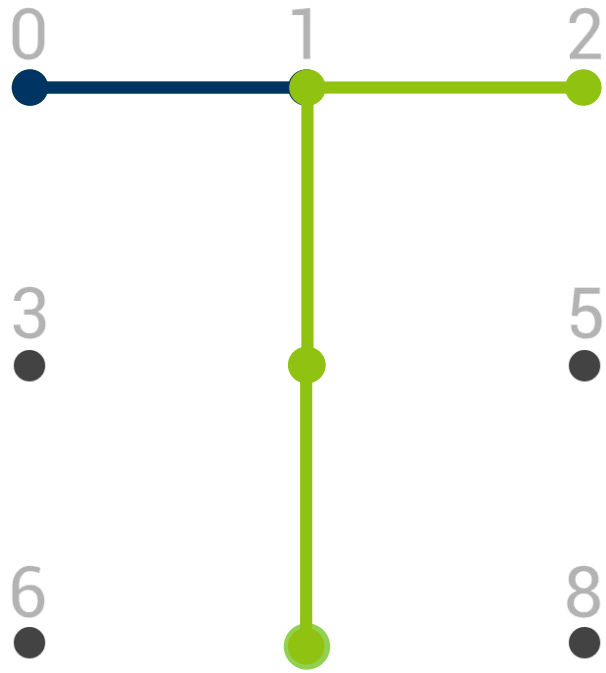
- Length
- Starting Node
- Direction Changes
- Knight Moves
- **Overlapping Nodes**
- Overlapping Segments
- Intersections
- Intersection (Restricted)
- …



1.4.0.2 → 1x Node

Visual Features

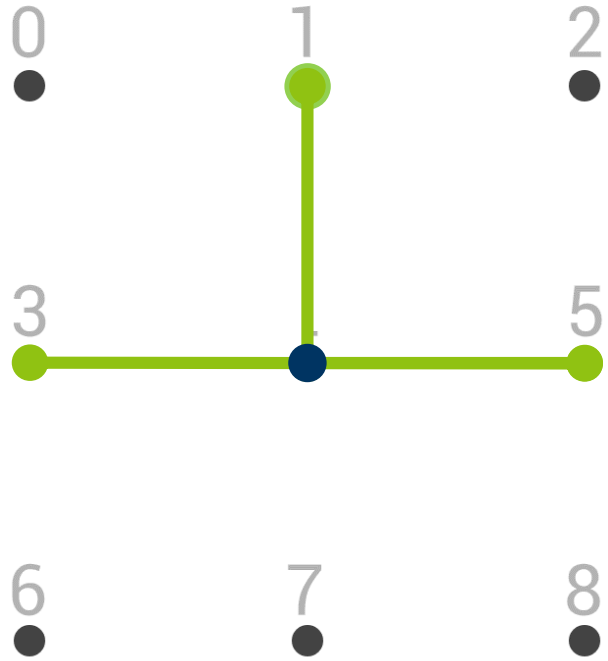
- Length
- Starting Node
- Direction Changes
- Knight Moves
- Overlapping Nodes
- **Overlapping Segments**
- Intersections
- Intersection (Restricted)
- …



7.4.1.0.2 → 1x Segment (“1.0”)

Visual Features

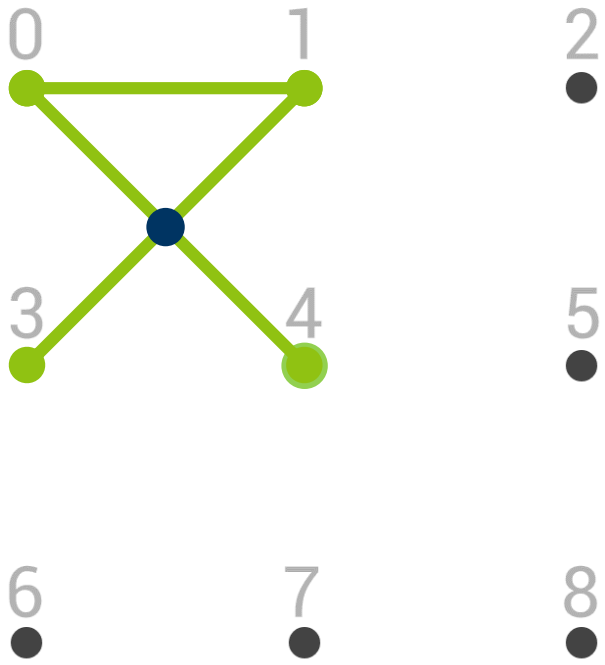
- Length
- Starting Node
- Direction Changes
- Knight Moves
- Overlapping Nodes
- Overlapping Segments
- **Intersections**
- Intersection (Restricted)
- …



1.4.5.3 → 1x Intersection (“1.4 vs. 5.3”)

Visual Features

- Length
- Starting Node
- Direction Changes
- Knight Moves
- Overlapping Nodes
- Overlapping Segments
- Intersections
- **Intersection (Restricted)**
- …



4.0.1.3 → 1x Intersection (“4.0 vs. 1.3”)