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!
Selection Bias

User-choice heavily biased.

Pattern

Frequency

Users at risk
Threat Model

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

**Throttled Guessing Attack:**

<table>
<thead>
<tr>
<th></th>
<th>Android 6</th>
<th>Android 7-9</th>
</tr>
</thead>
<tbody>
<tr>
<td>100 Guesses</td>
<td>45m</td>
<td>10h 45m</td>
</tr>
<tr>
<td>200 Guesses</td>
<td>1h 35m</td>
<td>67d 2h 45m</td>
</tr>
<tr>
<td>300 Guesses</td>
<td>2h 25m</td>
<td>167d 2h 45m</td>
</tr>
</tbody>
</table>

*Apple iOS only allows 10 guesses!*
Strength Meter

Support users while choosing their secret

Users at risk

Frequency

Strength Meter

Pattern

1 200 389,112

Set screen lock
For security, set pattern
Pattern recorded
Weak
Pattern Meter Proposals

2014

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.

**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?
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
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 $\rightarrow$ 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

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

Andriotis/Sun/Song

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

Measuring Accuracy

Reference

Strength | Pattern
--- | ---
392 | 0.1.2.5.8
218 | 0.1.2.4.6.7.8
198 | 0.3.6.7.8
82 | 0.3.6.7
47 | 0.1.2.5
46 | 1.4.7.8
45 | 0.4.8.5

Strength Meter

Strength | Pattern
--- | ---
32 | 0.1.2.5.8
15 | 0.1.2.4.6.7.8
191 | 0.3.6.7.8
37 | 0.3.6.7
294 | 0.1.2.5
117 | 1.4.7.8
51 | 0.4.8.5

## Measuring Accuracy

<table>
<thead>
<tr>
<th>Strength</th>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>0.1.2.5.8</td>
</tr>
<tr>
<td>37</td>
<td>0.1.2.4.6.7.8</td>
</tr>
<tr>
<td>41</td>
<td>0.3.6.7.8</td>
</tr>
<tr>
<td>59</td>
<td>0.3.6.7</td>
</tr>
<tr>
<td>118</td>
<td>0.1.2.5</td>
</tr>
<tr>
<td>240</td>
<td>1.4.7.8</td>
</tr>
<tr>
<td>392</td>
<td>0.4.8.5</td>
</tr>
</tbody>
</table>

### Reference

<table>
<thead>
<tr>
<th>Strength</th>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>0.1.2.5.8</td>
</tr>
<tr>
<td>15</td>
<td>0.1.2.4.6.7.8</td>
</tr>
<tr>
<td>191</td>
<td>0.3.6.7.8</td>
</tr>
<tr>
<td>37</td>
<td>0.3.6.7</td>
</tr>
<tr>
<td>294</td>
<td>0.1.2.5</td>
</tr>
<tr>
<td>114</td>
<td>1.4.7.8</td>
</tr>
<tr>
<td>51</td>
<td>0.4.8.5</td>
</tr>
</tbody>
</table>
Measuring Accuracy

<table>
<thead>
<tr>
<th>Rank</th>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>0.1.2.5.8</td>
</tr>
<tr>
<td>2.</td>
<td>0.1.2.4.6.7.8</td>
</tr>
<tr>
<td>3.</td>
<td>0.3.6.7.8</td>
</tr>
<tr>
<td>4.</td>
<td>0.3.6.7</td>
</tr>
<tr>
<td>5.</td>
<td>0.1.2.5</td>
</tr>
<tr>
<td>6.</td>
<td>1.4.7.8</td>
</tr>
<tr>
<td>7.</td>
<td>0.4.8.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rank</th>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.</td>
<td>0.1.2.5.8</td>
</tr>
<tr>
<td>1.</td>
<td>0.1.2.4.6.7.8</td>
</tr>
<tr>
<td>6.</td>
<td>0.3.6.7.8</td>
</tr>
<tr>
<td>3.</td>
<td>0.3.6.7</td>
</tr>
<tr>
<td>7.</td>
<td>0.1.2.5</td>
</tr>
<tr>
<td>5.</td>
<td>1.4.7.8</td>
</tr>
<tr>
<td>4.</td>
<td>0.4.8.5</td>
</tr>
</tbody>
</table>
## Measuring Accuracy

### Reference

<table>
<thead>
<tr>
<th>Weight</th>
<th>Rank</th>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>95</td>
<td>1.</td>
<td>0.1.2.5.8</td>
</tr>
<tr>
<td>91</td>
<td>2.</td>
<td>0.1.2.4.6.7.8</td>
</tr>
<tr>
<td>90</td>
<td>3.</td>
<td>0.3.6.7.8</td>
</tr>
<tr>
<td>86</td>
<td>4.</td>
<td>0.3.6.7</td>
</tr>
<tr>
<td>70</td>
<td>5.</td>
<td>0.1.2.5</td>
</tr>
<tr>
<td>41</td>
<td>6.</td>
<td>1.4.7.8</td>
</tr>
<tr>
<td>40</td>
<td>7.</td>
<td>0.4.8.5</td>
</tr>
</tbody>
</table>

### Strength Meter

<table>
<thead>
<tr>
<th>Weight</th>
<th>Rank</th>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>95</td>
<td>2.</td>
<td>0.1.2.5.8</td>
</tr>
<tr>
<td>91</td>
<td>1.</td>
<td>0.1.2.4.6.7.8</td>
</tr>
<tr>
<td>90</td>
<td>6.</td>
<td>0.3.6.7.8</td>
</tr>
<tr>
<td>86</td>
<td>3.</td>
<td>0.3.6.7</td>
</tr>
<tr>
<td>70</td>
<td>7.</td>
<td>0.1.2.5</td>
</tr>
<tr>
<td>41</td>
<td>5.</td>
<td>1.4.7.8</td>
</tr>
<tr>
<td>40</td>
<td>4.</td>
<td>0.4.8.5</td>
</tr>
</tbody>
</table>

### Recommendation

- weighted and ranked metrics → weighted Spearman correlation
# Weighted Spearman Correlation – Strength Meters

<table>
<thead>
<tr>
<th>Meter</th>
<th>Feature</th>
<th>All</th>
<th>Strong</th>
<th>Medium</th>
<th>Weak</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andriotis et al.</td>
<td>V</td>
<td>0.3</td>
<td>0.1</td>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>Sun et al.</td>
<td>V</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Song et al.</td>
<td>V</td>
<td>-0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Markov – Generic</td>
<td>P</td>
<td>0.6</td>
<td>0.3</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>Markov – Multi</td>
<td>P</td>
<td>0.8</td>
<td>0.4</td>
<td>0.8</td>
<td>0.8</td>
</tr>
<tr>
<td>Markov – Top 20</td>
<td>P+T</td>
<td>0.9</td>
<td>0.7</td>
<td>0.9</td>
<td>0.9</td>
</tr>
</tbody>
</table>

- **Our work**
- **Throttled Attacker – Top 200**

1.0 High positive correlation
0.0 No correlation
Outline

- Introduction
- Background
- Methodology
- Results
- WiP
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

Your password could be better.
- Don't use words used on Wikipedia (SanDiego) (Why?)
- Avoid using dates like 2019 (Why?)
- Consider inserting digits into the middle, not just at the end (Why?)

A better choice: S2019anQiego

How to make strong passwords
The Mobile Setting Is Different

 Propel:
 ~200 guesses

 Unpropel:
 Exhaustive search feasible

 Idea: Build meter based on
 • perception of security (length driven)
 • non-enforcing blacklisting (top N)
Accurate Meter

Trained Placebo Meter

Pattern: 1.2.5.8.7.6.3.0.4
Rank: 240
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)
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.
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

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
### Statistical - Guessability

<table>
<thead>
<tr>
<th>Datasets</th>
<th>Samples</th>
<th>$\varnothing$ Length</th>
<th>$H_\infty$</th>
<th>$\lambda_3$</th>
<th>$\lambda_{30}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strong</td>
<td>219</td>
<td>6.3</td>
<td>6.2</td>
<td>3.7</td>
<td>22.8</td>
</tr>
<tr>
<td>Medium</td>
<td>2,162</td>
<td>5.8</td>
<td>5.8</td>
<td>4.9</td>
<td>21.9</td>
</tr>
<tr>
<td>Weak</td>
<td>2,256</td>
<td>5.7</td>
<td>5.1</td>
<td>7.8</td>
<td>30.9</td>
</tr>
<tr>
<td>All</td>
<td>4,637</td>
<td>5.8</td>
<td>5.6</td>
<td>6.0</td>
<td>24.9</td>
</tr>
</tbody>
</table>

**Comparison**

<table>
<thead>
<tr>
<th>PIN Type</th>
<th>Samples</th>
<th>Bits</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-digit PIN (Amitay)</td>
<td>204,432</td>
<td>4.0</td>
<td>4.5</td>
</tr>
<tr>
<td>6-digit PIN (Wang)</td>
<td>2,758,490</td>
<td>6.0</td>
<td>3.1</td>
</tr>
<tr>
<td>PW (Melicher)</td>
<td>273</td>
<td>10.4</td>
<td>8.1</td>
</tr>
</tbody>
</table>
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)
- ...
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)
- ...

$7.4.1.0.2 \rightarrow 1x \text{ Segment ("1.0")}$
Visual Features

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

1.4.5.3 $\rightarrow$ 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 $\rightarrow$ 1x Intersection ("4.0 vs. 1.3")