universitätbonn

Technische Universität Braunschweig

CISPA Center for IT-Security, Privacy and Accountability

An Empirical Study of Textual Key-Fingerprint Representations

Sergej Dechand, Dominik Schürmann, Karoline Busse, Yasemin Acar, Sascha Fahl, Matthew Smith

universitätbonn

Technische Universität Braunschweig

CISPA Center for IT-Security, Privacy and Accountability

“

Title: *Do Users Verify SSH Keys?*

Abstract: *No*

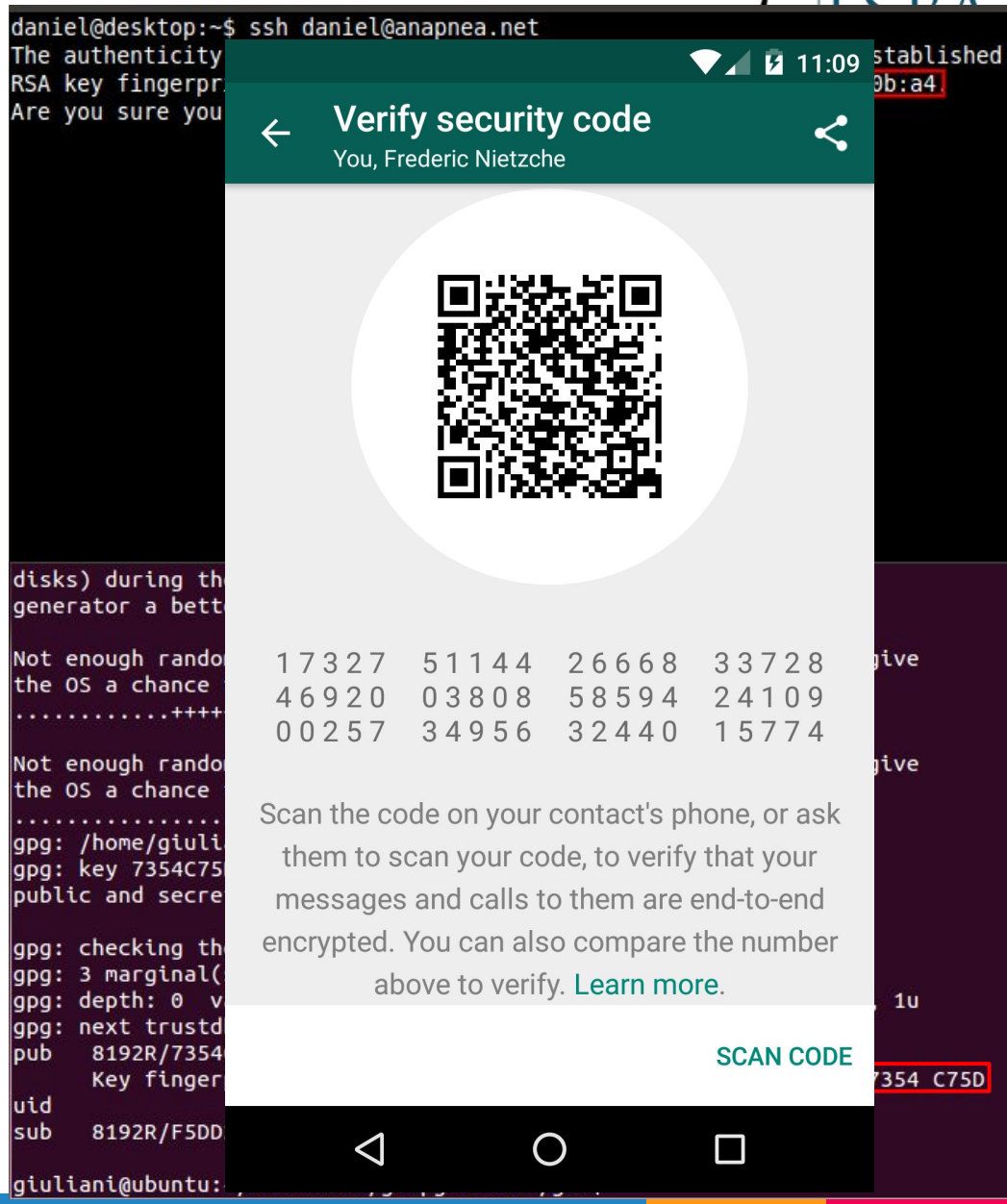
- *Peter Gutman, 2011*

Key Fingerprints

universitätbonn

Technische Universität Braunschweig

- Mostly not checked
- Error prone
  - Partial preimages
  - Hard to compare
- Meaningless?
- **Still relevant**



universitätbonn

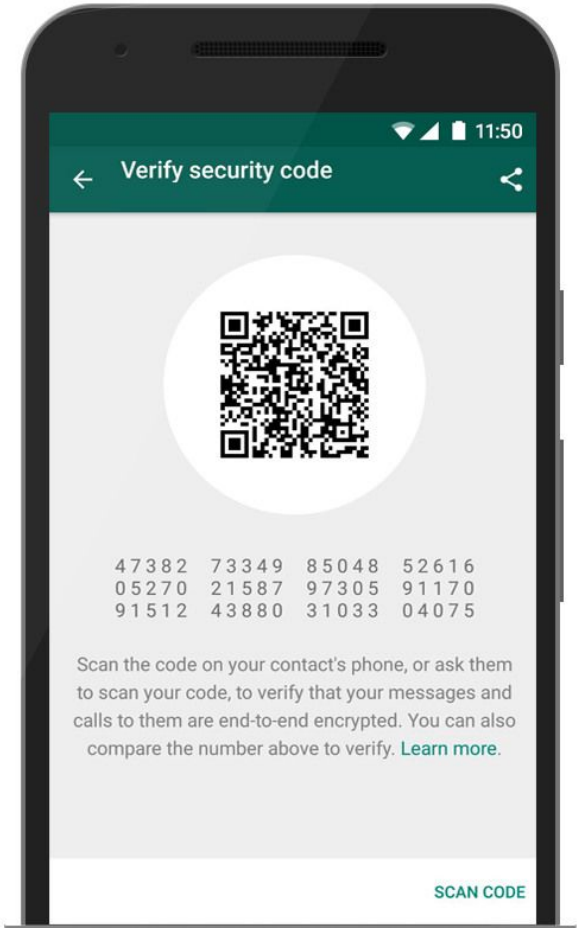
Technische Universität Braunschweig

CISPA Center for IT-Security, Privacy and Accountability

Our Goal

Which text representation is best?

- High attack detection
  - Partial preimages
  - Low false positive rate
- Efficient
  - Fast comparisons
  - Low cognitive load
- Best user perception
- Robust



Tested Representation Schemes

universitätbonn

Technische Universität Braunschweig

CISPA Center for IT-Security, Privacy and Accountability

- **Hexadecimal** `18e2 55fd b51b c808`
- **Base32** `ddrf l7nv dpea`
- **Numeric** `2016 507 6420 1070`
- **PGP List** locale voyager waffle disable
- **Peerio List** bates talking duke slurps
- **Sentences** That lazy snow agrees upon our tall offer

Threat Model

Which attacks are feasible?

universitätbonn

Technische Universität Braunschweig

CISPA Center for IT-Security, Privacy and Accountability

Attack Methods

Ideal: Preimage for an existing key fingerprint

- Expensive
- Infeasible

Workaround: Generate partial preimage

- Fingerprints almost match (except of a few chars)
- Exploit people's attention limitations

universitätbonn

Technische Universität Braunschweig

CISPA Center for IT-Security, Privacy and Accountability

Attacker Strength

- Assumptions
  - The fingerprints include key and metadata
  - New fingerprints without generating new keys
  - Only hashing needs to be performed
- 80 of 112 bits controlled by attacker
  - First and last few bits are controlled
- Still high costs to generate partial preimages
  - Although not impossible

universitätbonn

Technische Universität Braunschweig

CISPA Center for IT-Security, Privacy and Accountability

Simulated Attacks

- Inverting uncontrolled bits
- Inversions within a logical sequence
  - Characters
  - Words
  - Digits

```
18e2 55fd b51b c808
601b ee5c 2d69
```

```
18e2 55fd 4ae4 c808
601b 11a3 2d69
```

Study Design

Controlled experiment followed by a survey
Conducted on MTurk

universitätbonn

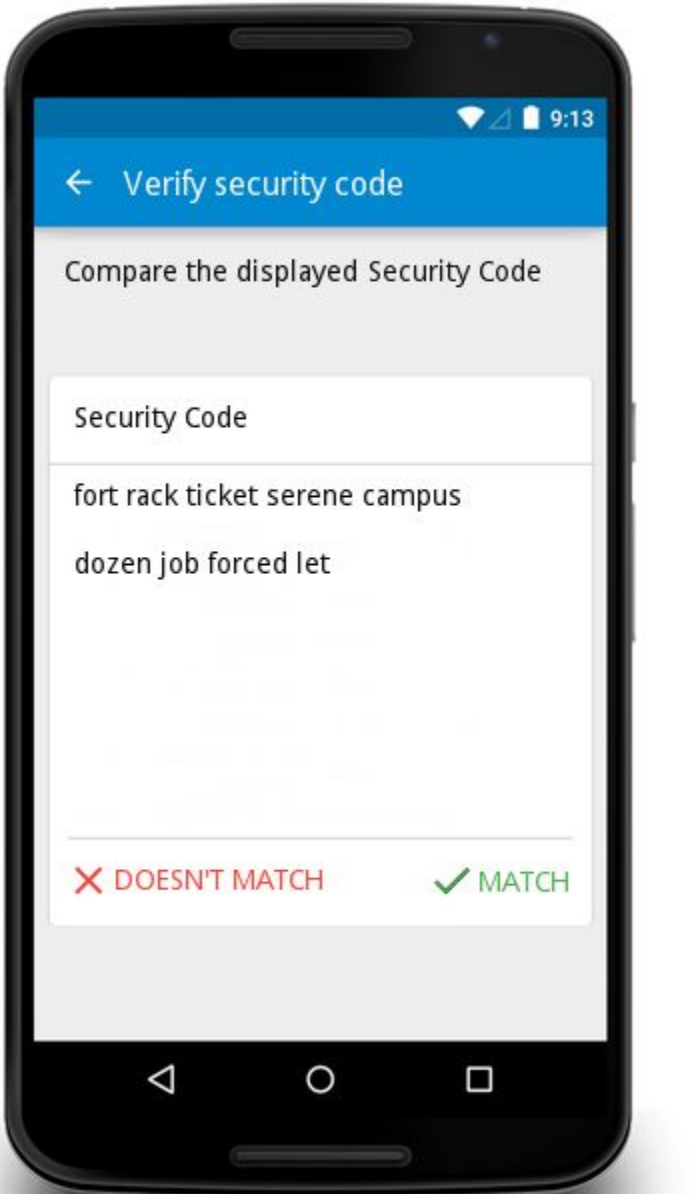
Technische Universität Braunschweig

CISPA Center for IT-Security, Privacy and Accountability

Study Design

- Users compare fingerprints
  - Match vs. Doesn't match
- Survey with usability questions
- Pre-study before setting study parameters
- 4 tested schemes, factorial design (mixing within and between groups)
  - Hex **or** Base32
  - Numeric
  - PGP **or** Peerio word list
  - Sentences

Experiment Task



universitätbonn

Technische Universität Braunschweig

CISPA Center for IT-Security, Privacy and Accountability

Study Design

- 40 comparisons in randomized order
  - Avoids fatigue and learning effect
  - Each scheme attacked once (randomized order)
  - Higher attack rate leads to higher detection rate
- Attention tests with obvious mismatches
  - Users failing the attention tests are excluded
- Training sets for each scheme
  - Reported typo search in language-based schemes
  - Not considered in the results

universitätbonn

Technische Universität Braunschweig

CISPA Center for IT-Security, Privacy and Accountability

Survey

- Survey after finishing all tasks
  - Rating the schemes
  - Demographics

The comparisons were easy for me with this method.

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
Alphanumeric (e.g., "e512 94f2 e9a2 a4be...")	○	○	○	○	○
Numeric (e.g., "2156 12 5325 7999...")	○	○	○	○	○
Unrelated words (e.g., "topmost treadmill Pacific dictator...")	○	○	○	○	○
Generated sentences (e.g., "My blue house runs our of time...")	○	○	○	○	○

universitätbonn

Technische Universität Braunschweig

CISPA Center for IT-Security, Privacy and Accountability

Challenges

- High number of participants required
  - High attack detection rate
  - Low differences between some approaches
- No parameter testing
  - Condition explosion if parameters are tested
  - Font settings
  - Chunking
  - Colors
- Additional experiment testing the chunking

Results

Controlled experiment and survey

universitätbonn

Technische Universität Braunschweig

CISPA Center for IT-Security, Privacy and Accountability

Results

- 1047 participants from MTurk
  - 46 excluded due to failed attention tests
  - Mixed demographics
  - No performance differences based on age, gender, education
- Relatively high attack detection rate for all schemes

universitätbonn

Technische Universität Braunschweig

CISPA Center for IT-Security, Privacy and Accountability

Experiment Results

	Speed (median)	Undetected Attacks	False Positives
Hexadecimal	10s	10.44%	0.5%
Base32	8.9s	8.50%	2.6%
Numeric	9.5s	6.34%	0.3%
PGP Word List	11.2s	8.78%	0.5%
Peerio Word List	7.3s	5.75%	0.4%
Sentences	10.7s	2.99%	1.5%

universitätbonn

Technische Universität Braunschweig

CISPA Center for IT-Security, Privacy and Accountability

Chunking Results

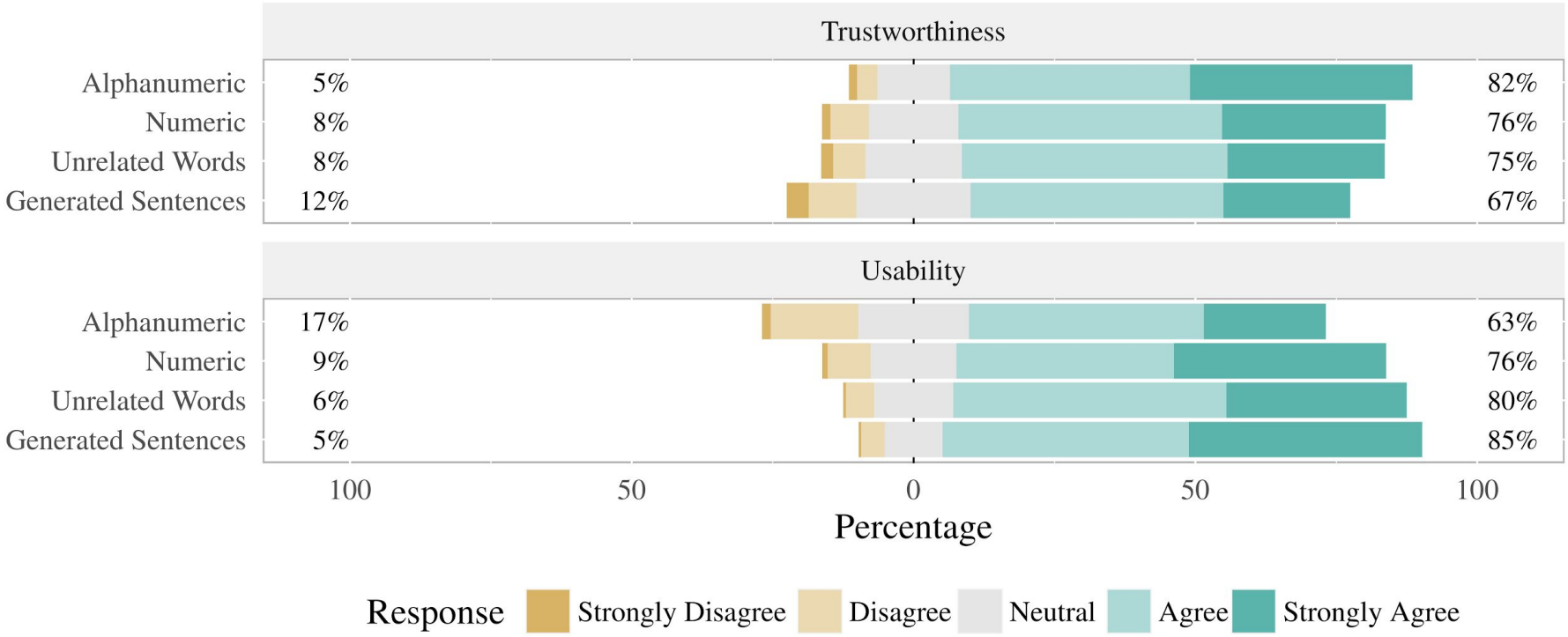
	Speed (median)	Undetected Attacks	False Positives
Hex 2	11.3s	8.15%	0.38%
Hex 3	**10.3s**	**6.14%**	**0.29%**
Hex 4	10.4s	6.78%	0.38%
Hex 5	11.6s	7.89%	0.78%
Hex 8	13.6	8.13%	0.5%

universitätbonn

Technische Universität Braunschweig

CISPA Center for IT-Security, Privacy and Accountability

Survey Results



universitätbonn

Technische Universität Braunschweig

CISPA Center for IT-Security, Privacy and Accountability

Limitations

- No guarantee if verification is performed
- Validity of MTurk (as with any MTurk study)
  - More tech-savvy
  - Younger
  - Used to textual and visual tasks
- No tests for additional parameters due to condition explosion
  - Font settings (type, size, etc.)
  - Use of colors
  - Line break settings

Conclusion

Takeaways?

universitätbonn

Technische Universität Braunschweig

CISPA Center for IT-Security, Privacy and Accountability

Conclusion

- Hex has shown the worst performance
  - Lower attack detection rate
  - Slower than most approaches
  - Perceived to be more annoying
- Generated sentences with best results
  - Highest attack detection rate
  - Best results regarding usability
- Numeric best non language-based scheme