

Password



Weak



RUB

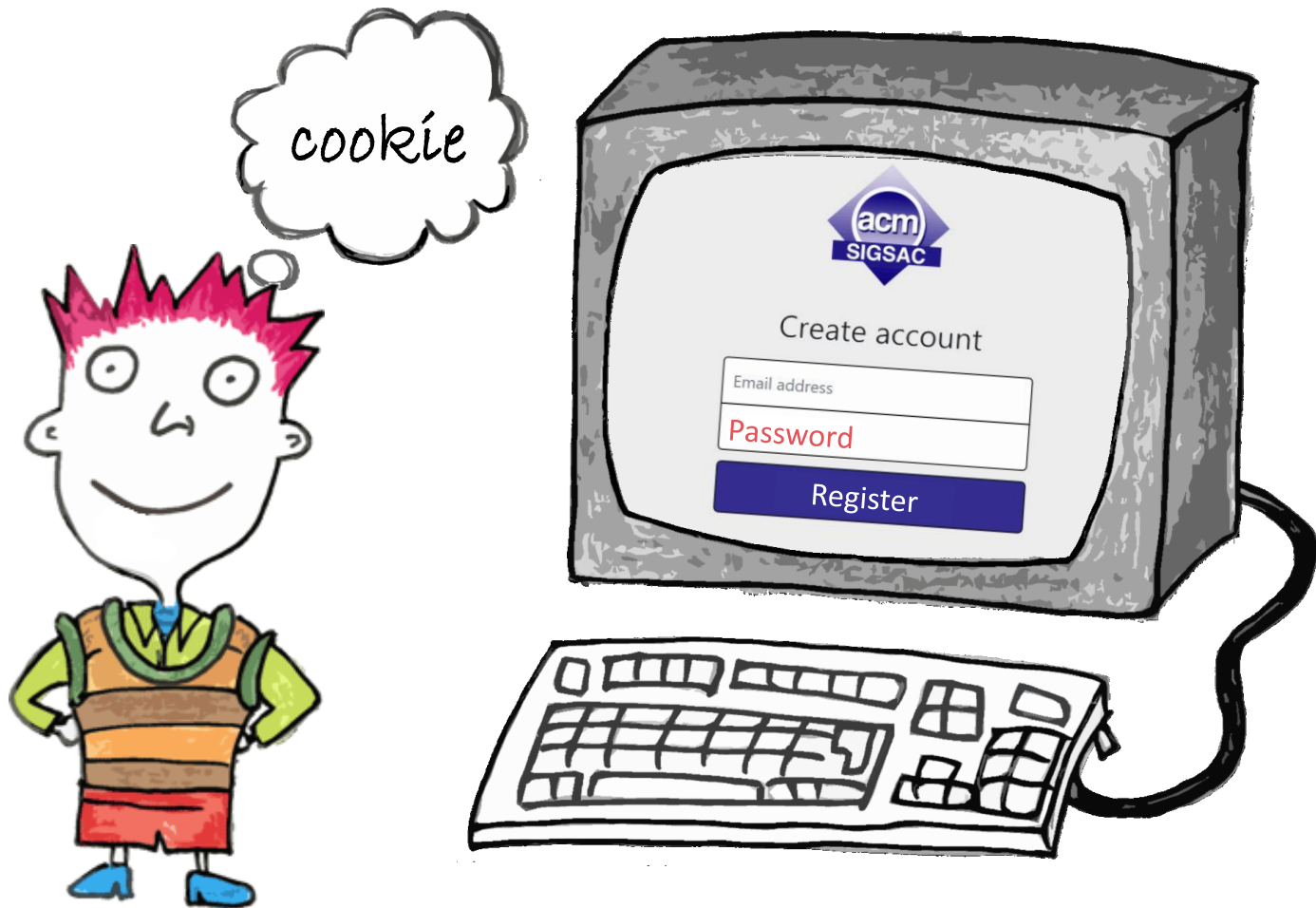
On the Accuracy of Password Strength Meters

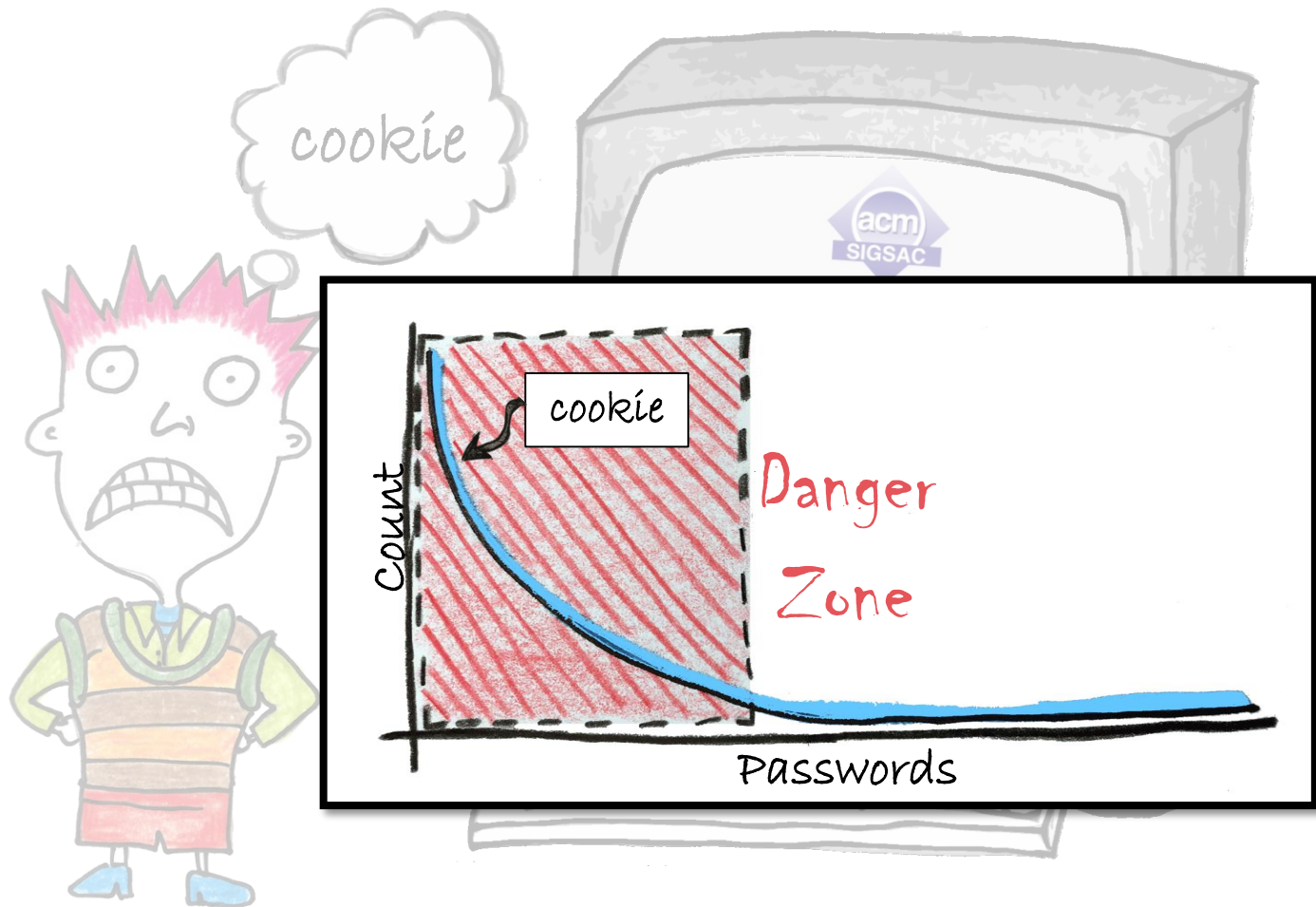
Maximilian Golla and Markus Dürmuth

Horst Görtz Institute for IT-Security
Ruhr University Bochum

Rockst@r
123654
cookie



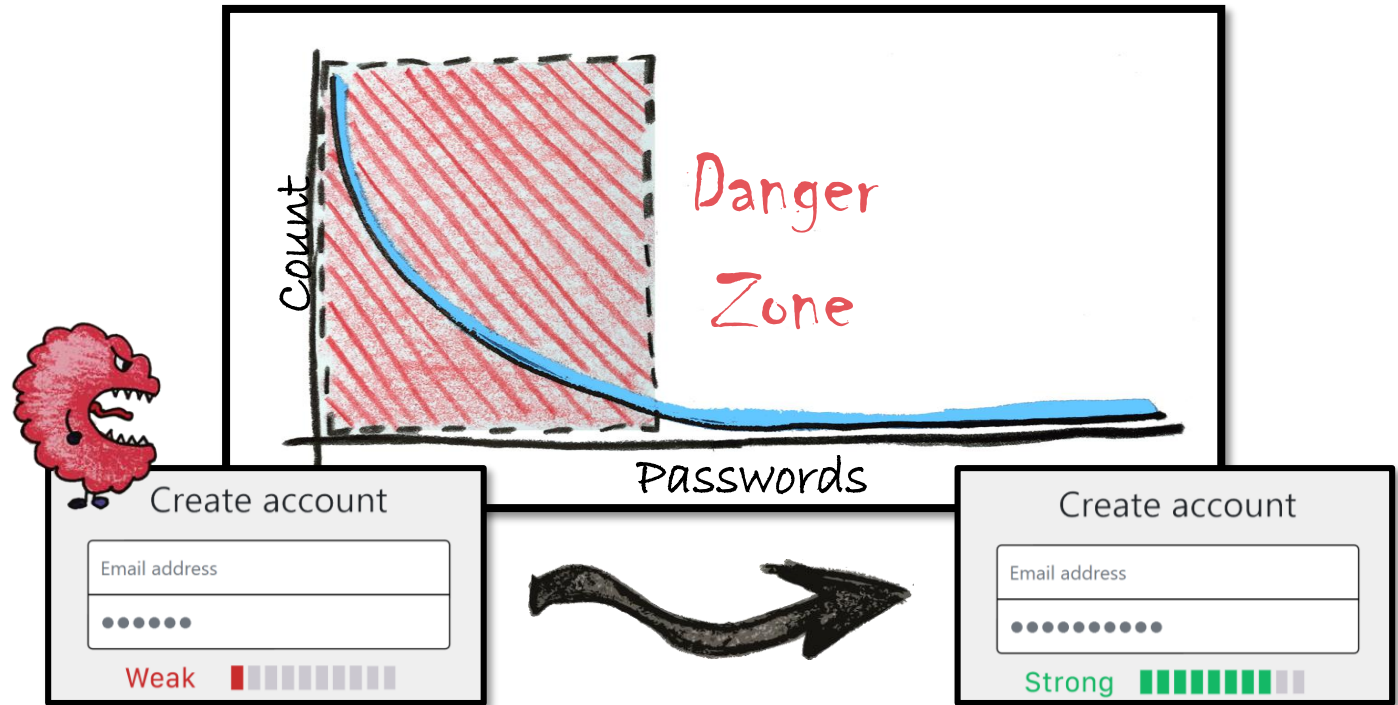






Password Strength Meter

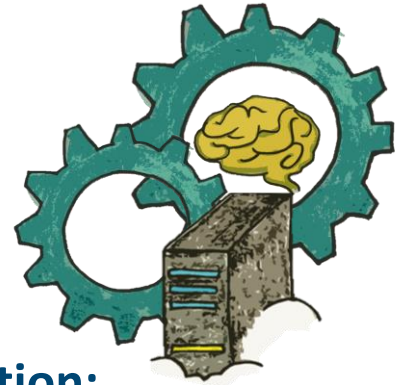
Nudge users toward more secure passwords



Meter Implementations

Strength Value:

- Text [*Weak, Medium, Strong*]
- Color (**Red**, **Orange**, **Green**)
- Percentages (**42%**)
- Values/Scores (**67**)
- Time (**12 days**)
- “Bits” (**82 bits**)
- **Guesses (1,018,291)**



Implementation:

- Client-/server-side
- Heuristics
- Probabilistic models

acmccs18



Create account

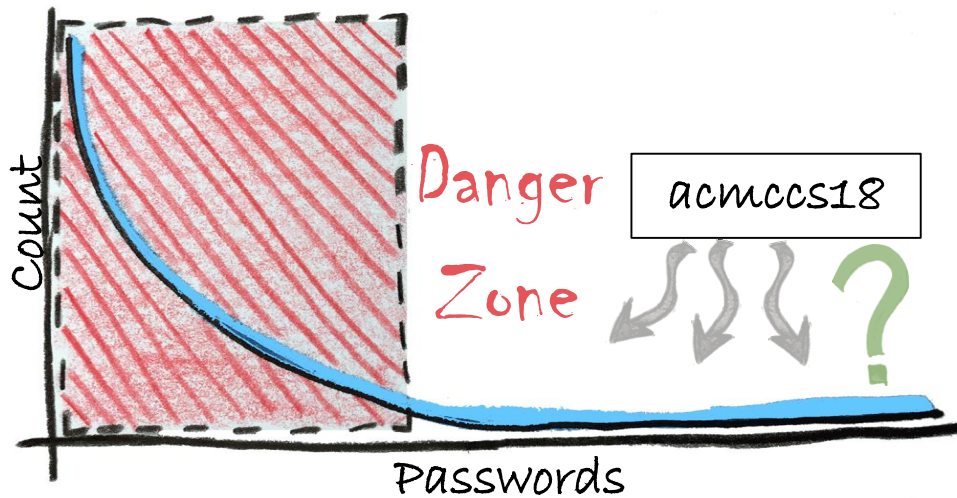
Email address

acmccs18

Strong 

Sign in

acmccs18



How secure is acmccs18 ?

Weak

Medium

Strong

HOW SECURE IS MY PASSWORD?

acmccs18

It would take a computer about

1 MINUTE

to crack your password

acmccs18

confirm

Security

answer

Security

Your password must have:

- ✓ 8 or more characters
- ✓ Upper & lowercase letters
- ✓ At least one number

Strength:

Avoid using a password that you use with other websites or that might be easy for someone else to guess.

Have a longer password and prosper - Spock

acmccs18

Password

acmccs18

Show Password & Detailed Feedback

acmccs18

strength

length 20

Regenerate

words

digits

symbols

Enter a password

acmccs18

Confirm password

Strong password

Great! That one works. Just don't forget it

acmccs18

Your password will be bruteforced with an average home computer in approximately

12 DAYS

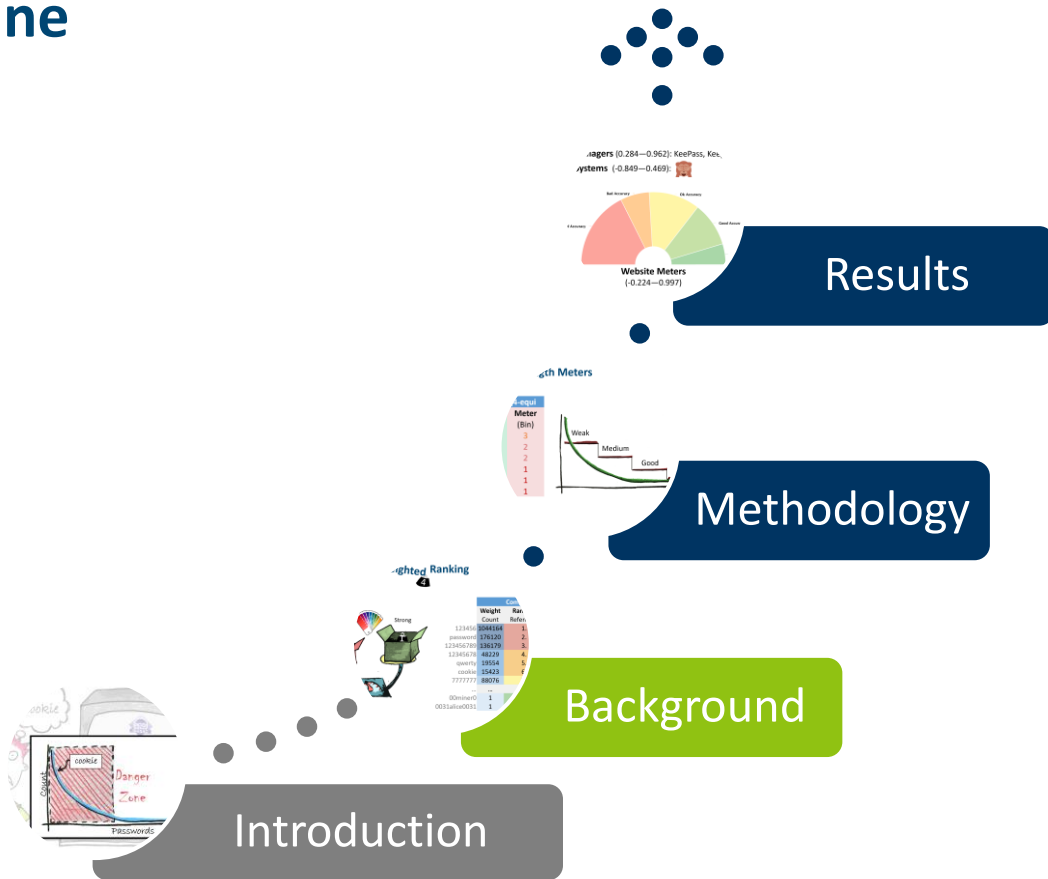
acmccs18

pwned?

Good news — no pwnage found!

This password wasn't found in any of the Pwned Passwords loaded into Have I Been Pwned. That doesn't necessarily mean it's a good password, merely that it's not indexed on this site. If you're not already using a password manager, go and download 1Password and change all your passwords to be strong and unique.

Outline



We need a Reference

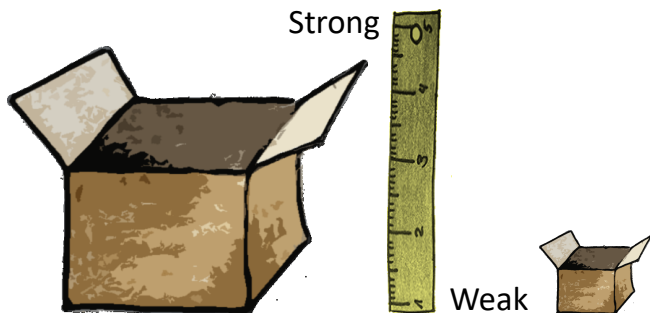
Reference

Strength	Password
1	123456
2	password
3	123456789
9	12345678
19	qwerty
85	cookie
154	7777777

Strength Meter

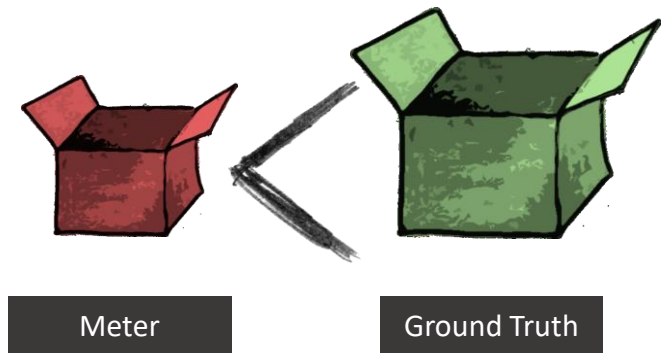
Strength	Password
7	123456
3	password
1	123456789
2	12345678
82	qwerty
1300	cookie
430	7777777

1) Compare Strength Values



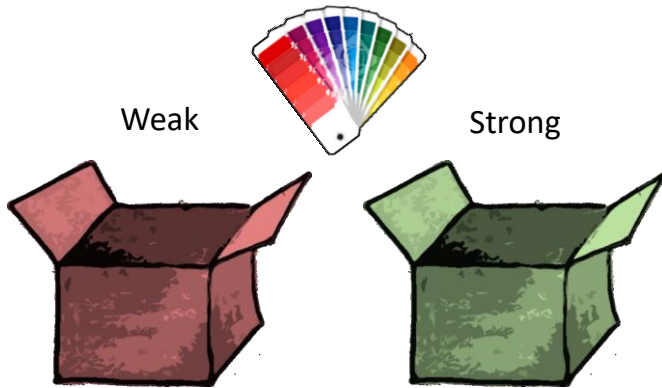
	Comparison	
	Strength Reference	Strength Meter
123456	1	7
password	2	3
123456789	3	1
12345678	9	2
qwerty	19	82
cookie	63	1309
7777777	154	430
...
00miner0	2.18E+8	1.12E+11
0031alice0031	1.69E+14	2.94E+15

2) Count Over- and Underestimates



Comparison		
	Strength Reference	Strength Meter
123456	1	7
password	2	3
123456789	3	1
12345678	9	2
qwerty	19	82
cookie	63	1309
7777777	154	430
...
00miner0	2.18E+8	1.12E+11
0031alice0031	1.69E+14	2.94E+15

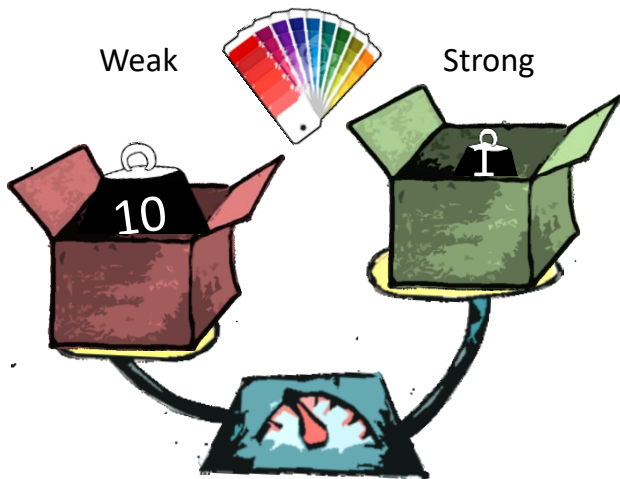
3) Compare Ranking



	Comparison	
	Ranking Reference	Ranking Meter
123456	1.	4.
password	2.	3.
123456789	3.	1.
12345678	4.	2.
qwerty	5.	5.
cookie	6.	7.
7777777	7.	6.
...
00miner0	9999.	9999.
0031alice0031	10000.	10000.

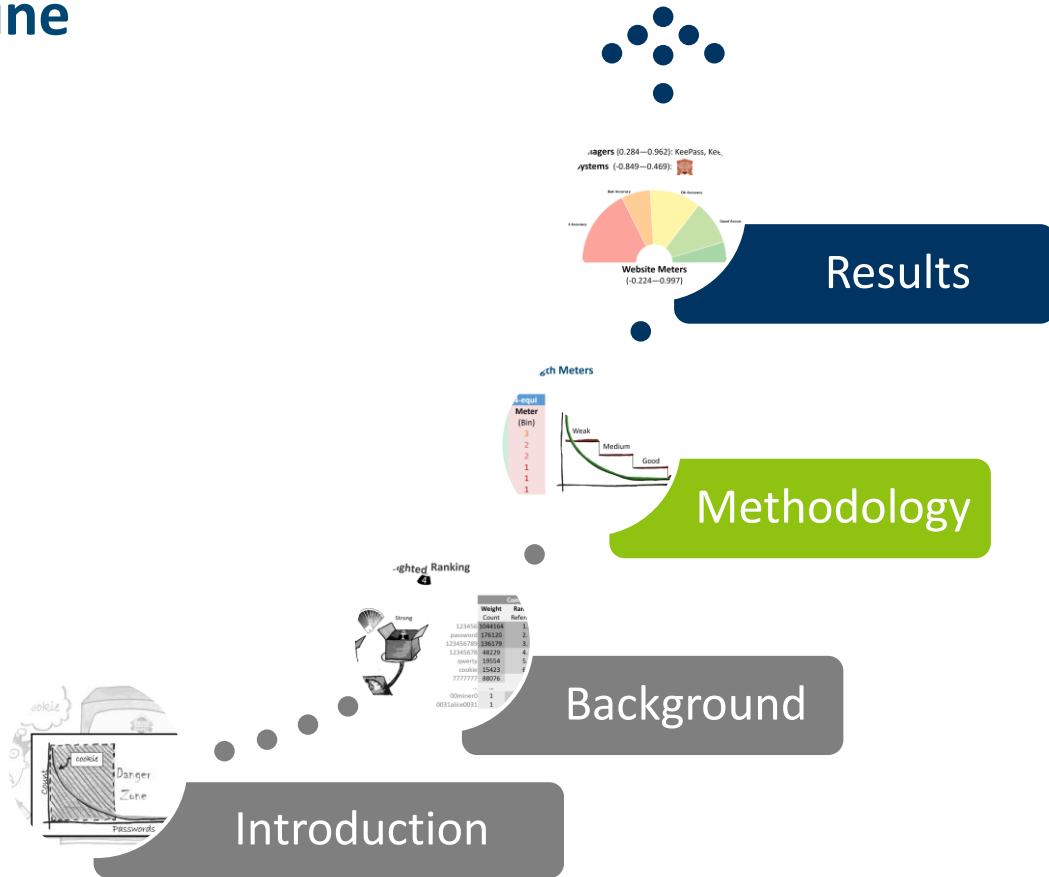
4) Compare *Weighted* Ranking

4



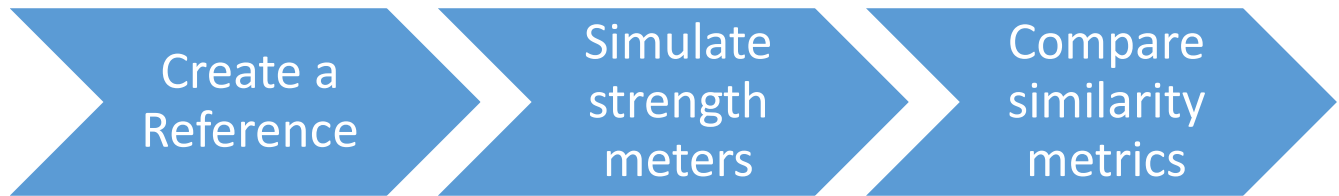
Comparison			
	Weight Count	Ranking Reference	Ranking Meter
123456	1044164	1.	4.
password	176120	2.	3.
123456789	136179	3.	1.
12345678	48229	4.	2.
qwerty	19554	5.	5.
cookie	15423	6.	7.
7777777	88076	7.	6.
...
00miner0	1	9999.	9999.
0031alice0031	1	10000.	10000.

Outline



Our Approach

Goal: Identify the most appropriate metric to measure the accuracy of strength meters

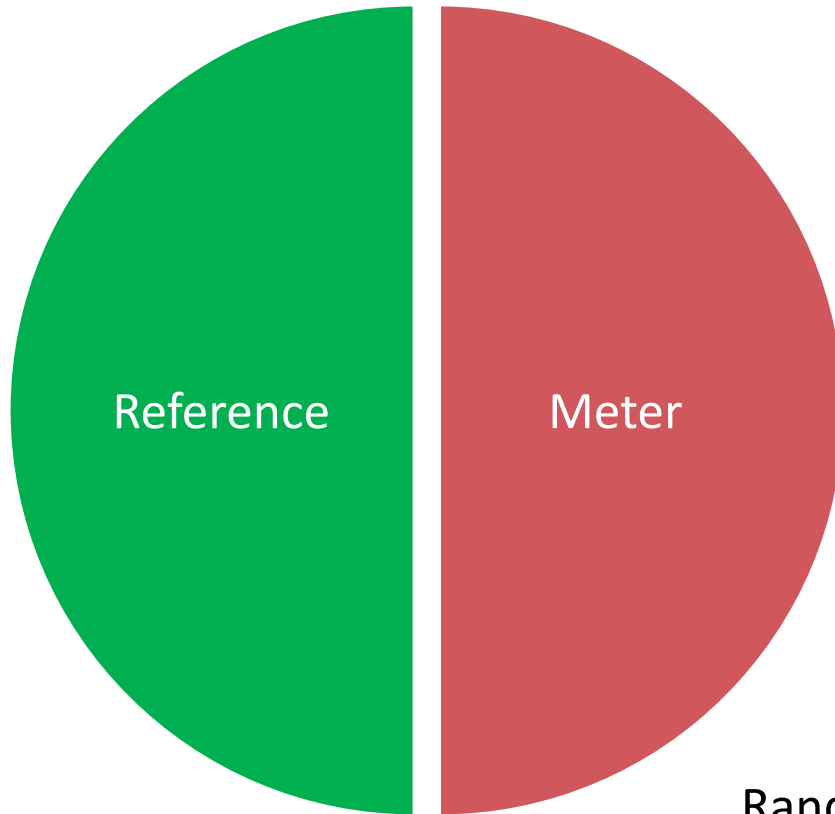


Create a Reference



Strength: Frequency

Create a Reference



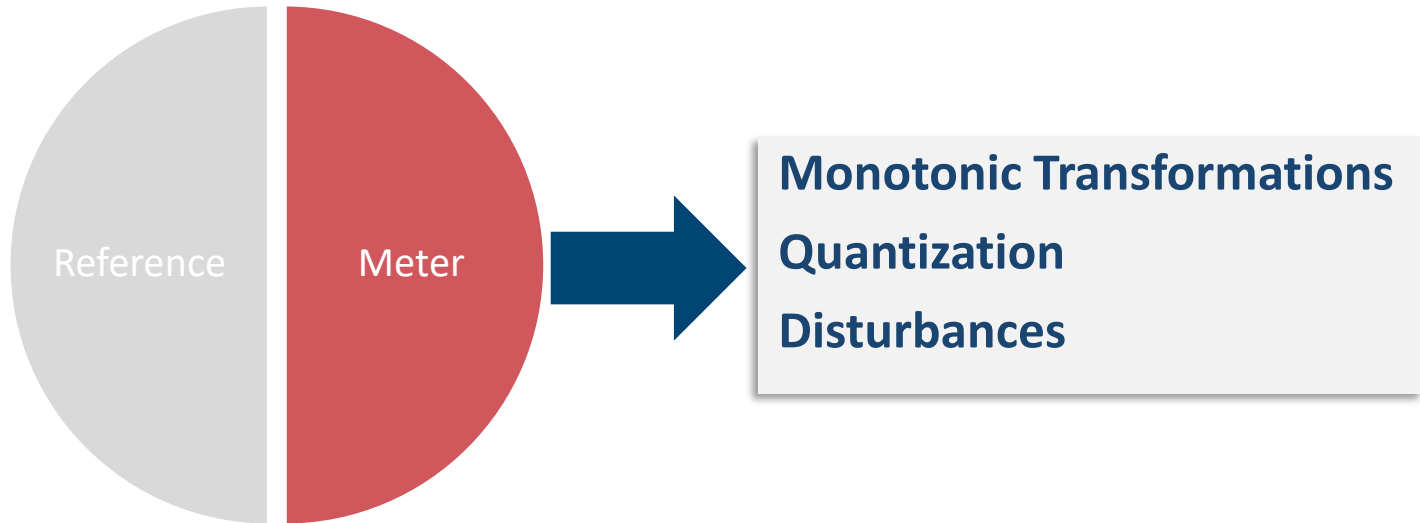
Random sampling

Create a Reference

REF-A vs. REF-B	
Reference (Count)	Meter (Count)
63	64
19	19
9	7
3	3
2	2
1	1
...	...

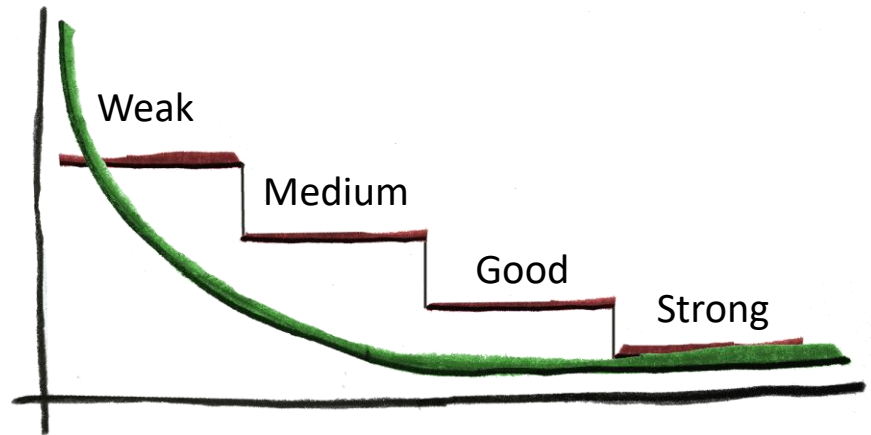


Simulate Strength Meters



Simulate Strength Meters (Example)

REF-A vs. Q4-equi	
Reference (Count)	Meter (Bin)
63	40
19	30
9	20
3	20
2	10
1	10
...	...



Compare Similarity Metrics



Correlation

- Pearson Correlation
- Spearman Rank Correlation
- Kendall Rank Correlation



Mean Error Metrics

- Mean Absolute Error
- Mean Square Error



Weighted variants



One-Sided/Pairwise Error Metrics

- Mean Abs./Squared One-Sided Lower Error
- Pairwise Error/Utility Rate

Tested 19
metrics!



Recommendation

- weighted and ranked metrics
(e.g., weighted Spearman correlation)

Evaluated Datasets

PW List	Year	Service
RockYou	2009	Social Games
LinkedIn	2012	Social Network
000Webhost	2016	Web Hosting

Why multiple datasets?

- Service-specific passwords
- “Important” vs. “Don’t care” accounts
- Composition Policies
- Tuned meters

Evaluation

Use Case 1: Online Attacker

Sample: Rand. 1k from top 10k

Strength: Count value

Use Case 2: Offline Attacker

Sample: Random 10k

Strength: PGS min_auto^[1]

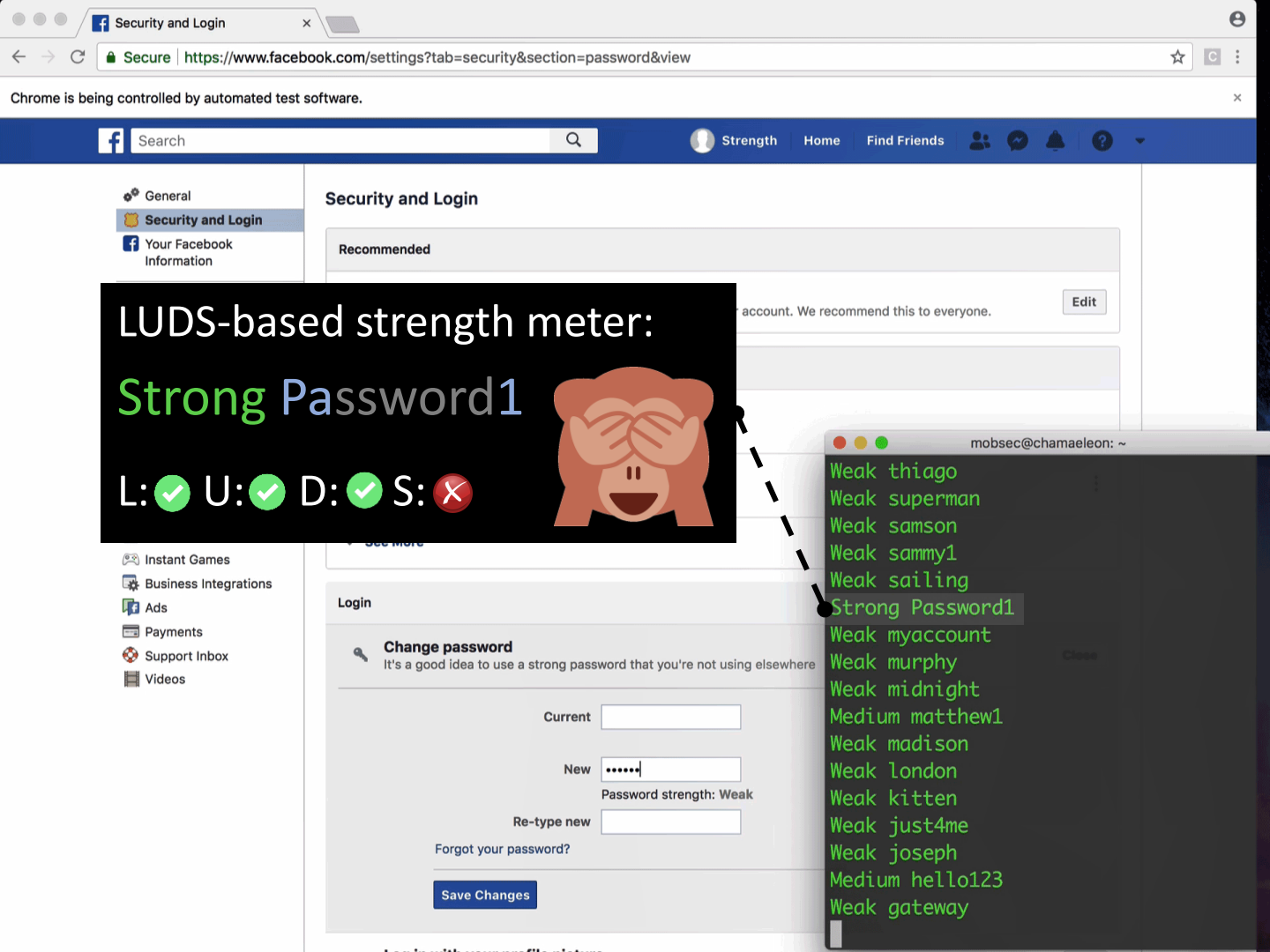
Querying the Meters:

Selenium (Appium)

Contacting the developers

Train/Patch/Reimplement





LUDS-based strength meter:

Strong Password1

L: [green checkmark] U: [green checkmark] D: [green checkmark] S: [red X]



```
Weak thiago
Weak superman
Weak samson
Weak sammy1
Weak sailing
Strong Password1
Weak myaccount
Weak murphy
Weak midnight
Medium matthew1
Weak madison
Weak london
Weak kitten
Weak just4me
Weak joseph
Medium hello123
Weak gateway
```

Large Scale Comparison

81

Implementations tested



A. METER COMPARISON

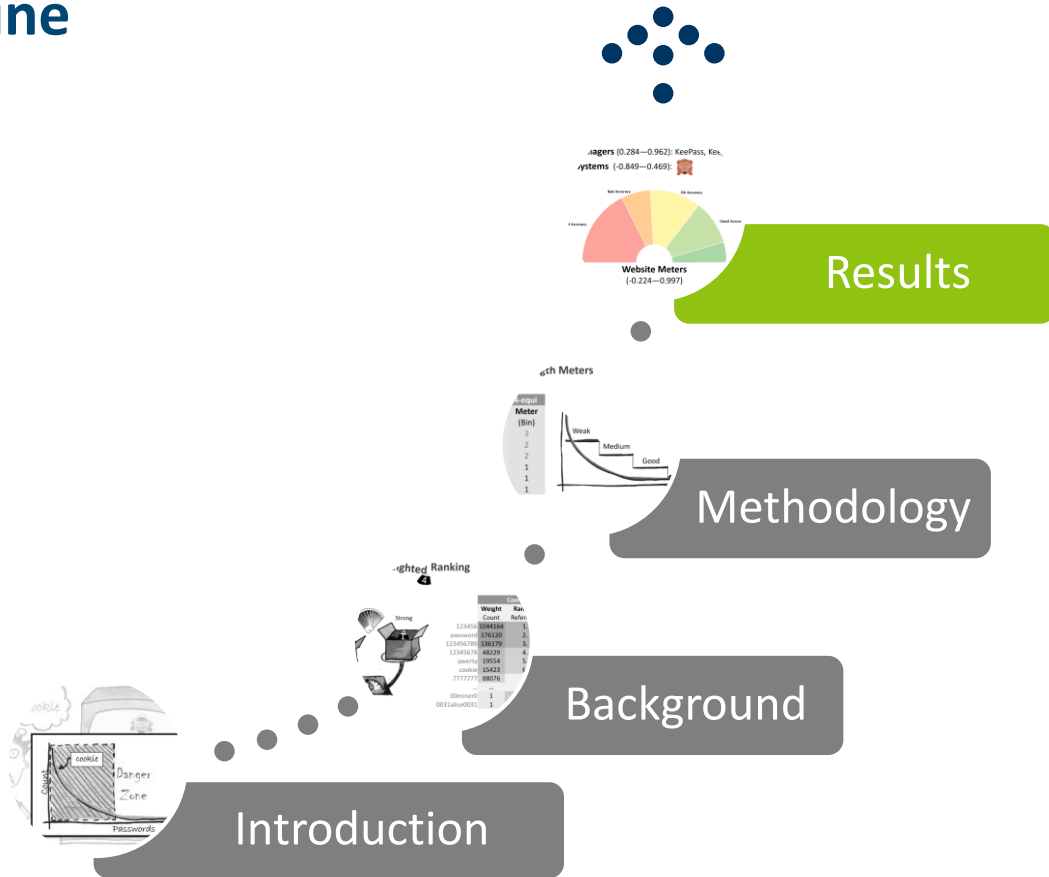
In the following, we list the full results of our data collection. We separated the five categories Academic Proposals, Password Managers, Operating Systems, Websites, and Previous Work into two tables. A colorful version that allows easier comparison can be found online [29].

A.1 Academic Proposals, Password Managers, and Operating Systems

Table 5: We computed the weighted Spearman correlation as the best similarity score (cf. Section 5). The table lists the online use case on the left, the offline use case on the right. We highlighted if and on how many times a meter was equal to its output. Additionally, we list whether a meter runs on client- or server-side and how the meter visualizes the strength to the user.

ID	Meter	Type	Quant.	Vis.	Online Attacker BackOff LockInfo	Offline Attacker LockInfo	Online Attacker BackOff LockInfo	Offline Attacker LockInfo
Academic Proposals								
1A	Comprehensive [41]	C	Test	-	-4.02	-4.09	4.21	-4.45
1B	Comprehensive [41]	C	Test	-	-3.93	-3.84	4.00	-4.28
2	Flow [12]	C	Test	-	4.09	4.02	4.40	4.30
3	TPM [35]	C	Test	-	4.26	4.00	4.00	4.44
4A	Metasploit [40]	C	Test	-	4.10	4.00	4.40	4.40
4B	Metasploit [40]	C	Test	-	4.10	4.00	4.40	4.40
4C	Metasploit [40]	C	Test	-	4.10	4.00	4.40	4.40
5A	NIST [11]	C	Test	-	4.09	4.02	4.40	4.30
5B	NIST [11]	C	Test	-	4.09	4.02	4.40	4.30
6	PCRF (PassGuard) [36]	C	Test	-	4.00	4.00	4.40	4.40
7A	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7B	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7C	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7D	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7E	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7F	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7G	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7H	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7I	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7J	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7K	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7L	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7M	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7N	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7O	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7P	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7Q	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7R	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7S	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7T	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7U	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7V	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7W	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7X	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7Y	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7Z	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AA	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AB	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AC	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AD	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AE	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AF	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AG	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AH	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AI	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AJ	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AK	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AL	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AM	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AN	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AO	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AP	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AQ	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AR	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AS	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AT	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AU	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AV	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AW	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AX	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AY	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7AZ	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BA	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BB	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BC	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BD	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BE	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BF	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BG	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BH	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BI	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BJ	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BK	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BL	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BM	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BN	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BO	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BP	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BQ	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BR	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BS	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BT	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BU	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BV	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BW	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BX	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BY	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7BZ	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CA	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CB	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CC	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CD	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CE	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CF	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CG	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CH	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CI	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CJ	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CK	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CL	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CM	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CN	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CO	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CP	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CQ	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CR	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CS	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CT	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CU	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CV	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CW	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CX	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CY	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7CZ	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DA	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DB	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DC	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DD	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DE	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DF	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DG	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DH	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DI	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DJ	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DK	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DL	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DM	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DN	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DO	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DP	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DQ	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DR	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DS	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DT	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DU	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DV	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DW	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DX	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DY	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7DZ	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7EA	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7EB	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7EC	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7ED	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7EE	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7EF	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7EG	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7EH	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7EI	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7EJ	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7EK	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7EL	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7EM	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7EN	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7EO	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7EP	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7EQ	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40
7ER	BNV (Benevo) [37]	C	Test	-	4.00	4.00	4.40	4.40

Outline



Accuracy

Markov

Markov Model
NDSS '12 [13]

0.721 – 0.998

PCFG

Probabilistic Context-Free Grammar
ACSAC '12 [34]
DSN '16 [65]

0.963 – 1.000

RNN

Recurrent Neural Network
USENIX '16 [46]
CHI '17 [59]

0.913 – 0.965

zxcvbn

Advanced Heuristic
USENIX '16 [71]

0.554 – 0.990

1.0 High positive correlation

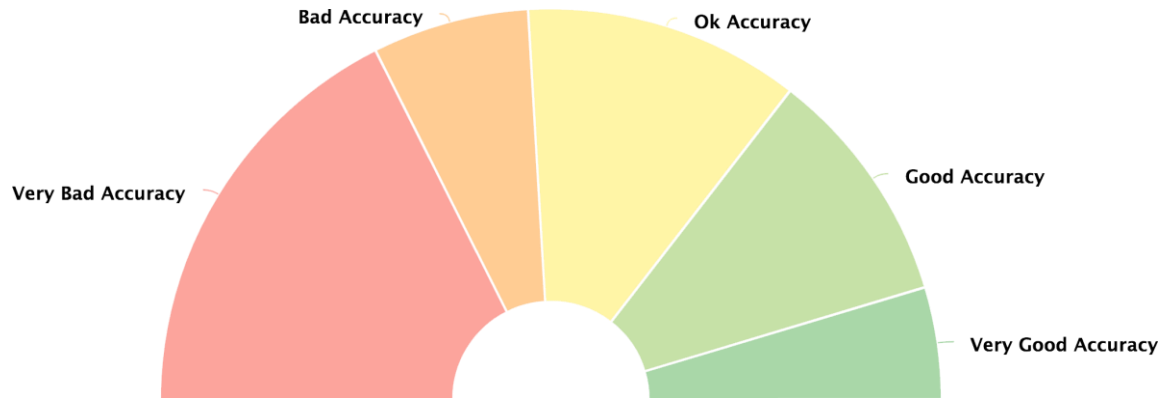
0.0 No correlation

-1.0 High negative correlation

Accuracy

Password Managers (0.284—0.962): KeePass, Keeper, zxcvbn-based

Operating Systems (-0.849—0.469):



Website Meters
(-0.224—0.997)

Effect of Quantization

It's important to carefully choose the quantization thresholds.



KeePass

- Bad thresholds:

Abs.: 0.393—0.884

Quant.: 0.000—0.321

1Password

- Good thresholds:

Abs.: 0.276—0.807

Quant.: 0.276—0.813

Cookie: 21 vs. Cookie: "Weak"

```
def quantize(strength):  
    if strength >= 90:  
        return "Fantastic"  
    if strength >= 60:  
        return "Excellent"  
    if strength >= 35:  
        return "Good"  
    if strength >= 20:  
        return "Weak"  
    return "Terrible"
```


Future Directions & Limitations

Next steps:

- Dependency on pw distribution
 - ID: 7B RNN Generic (Web) (0.421—0.777)
 - ID: 7C RNN Target (0.860—0.965)
- Model size
- Understand and mitigate the negative effects of quantization
- Work on the deployability, too! (zxcvbn's success)

Limitation

Usability and deployability aspects are vital for a complete assessment but are not presented in this work!

password-meter-comparison.org

