No Single Silver Bullet: Measuring the Accuracy of Password Strength Meters

Ding Wang, Xuan Shan, Qiying Dong
Nankai University
{wangding, shanxuan, dqy}@nankai.edu.cn

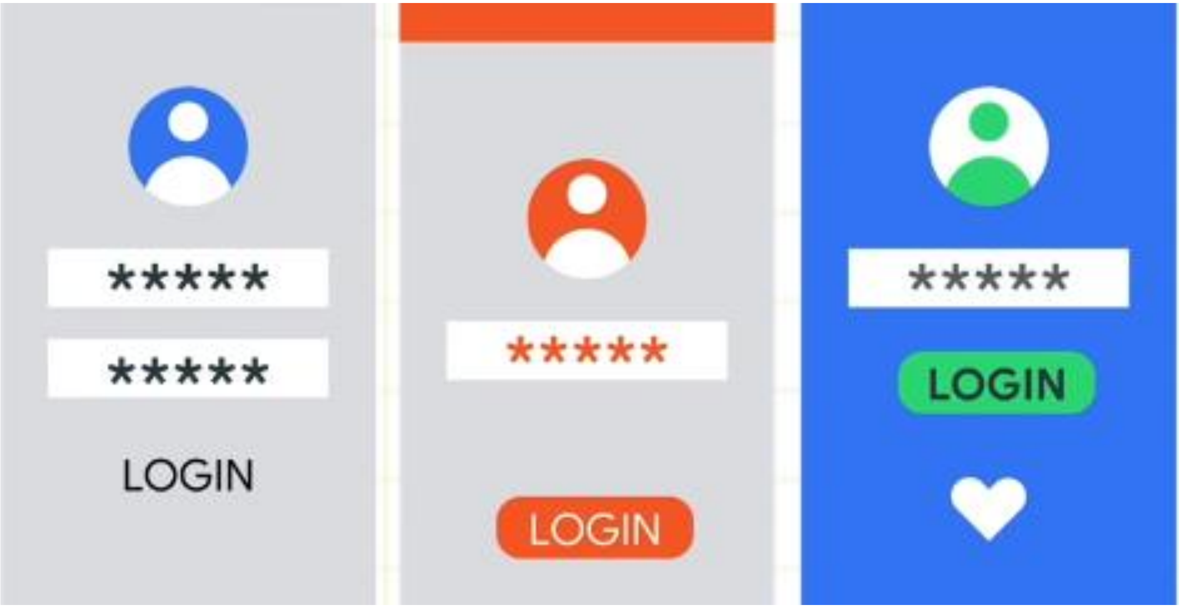
Yaoshen Shen
Peking University
ysshen@pku.edu.cn

Chunfu Jia
Nankai University
cfjia@nankai.edu.cn

The 32nd USENIX Security Symposium

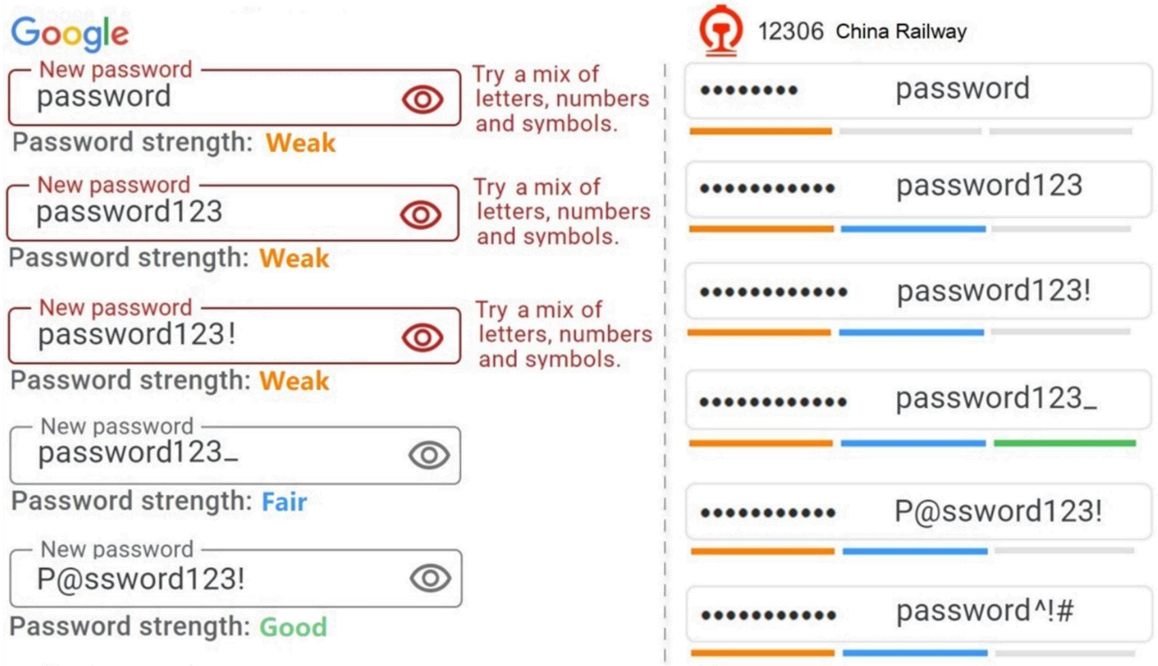
Passwords Are Not Dead

Passwords firmly remain the dominant mechanism for user access control on the Web.



Password strength meter (PSM)

- A PSM provides real-time strength feedback upon user registration and password change.
- PSMs encourage users toward secure passwords.



Academic

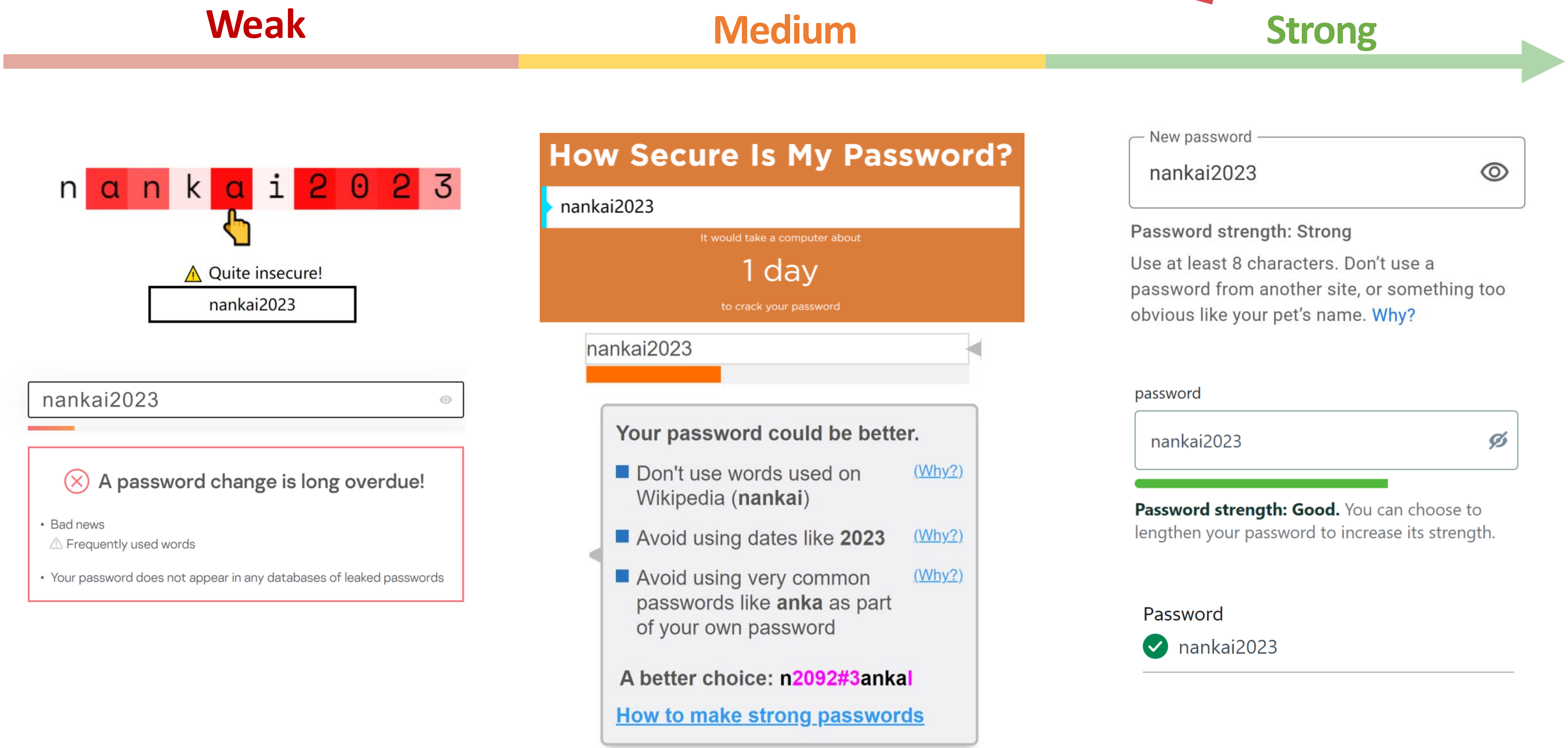
- fuzzyPSM [DSN'16]
- MultiPSM [TIFS'17]
- PCFG-PSM [ACSAC'12]
- Markov-PSM [NDSS' 12]
- RNN-PSM [Usenix SEC'16]
- LPSE [COSE'18]
- CNN-PSM [ESORICS'16]

Industrial

- Zxcvbn
- KeePSM
- 12306-PSM
- Microsoft-PSM

Strength of *nankai2023*

Which PSM is more accurate?



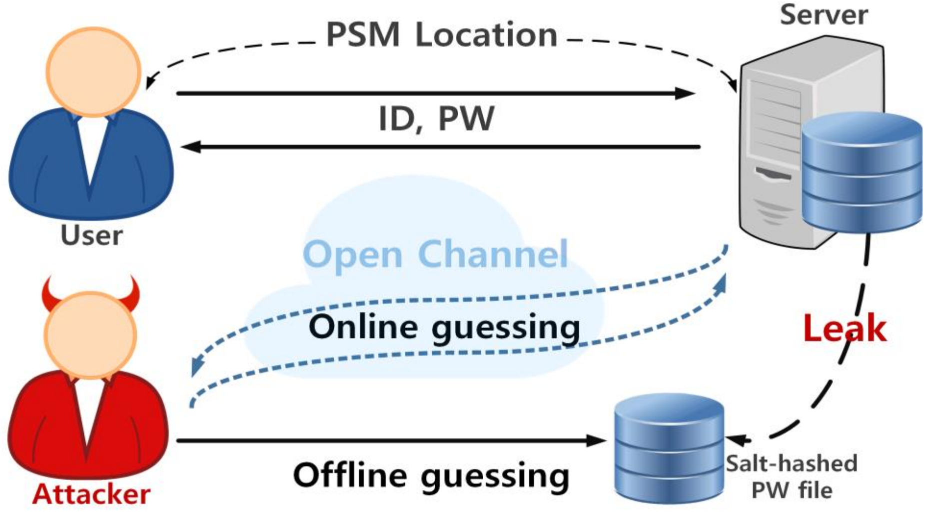
Our work

- **A systematic evaluation framework**, that
  - composed of 4 different dimensioned criteria to rate PSM accuracy under 2 guessing scenarios (and 6 guessing strategies).
- **Extensive evaluation.**
  - Our evaluation framework is utilized by rating 12 state-of-art PSMs, leveraging 14 real-world password datasets.
- **Some insights.**
  - 3 recommendations to help improve the accuracy of PSMs.

Password guessing scenarios

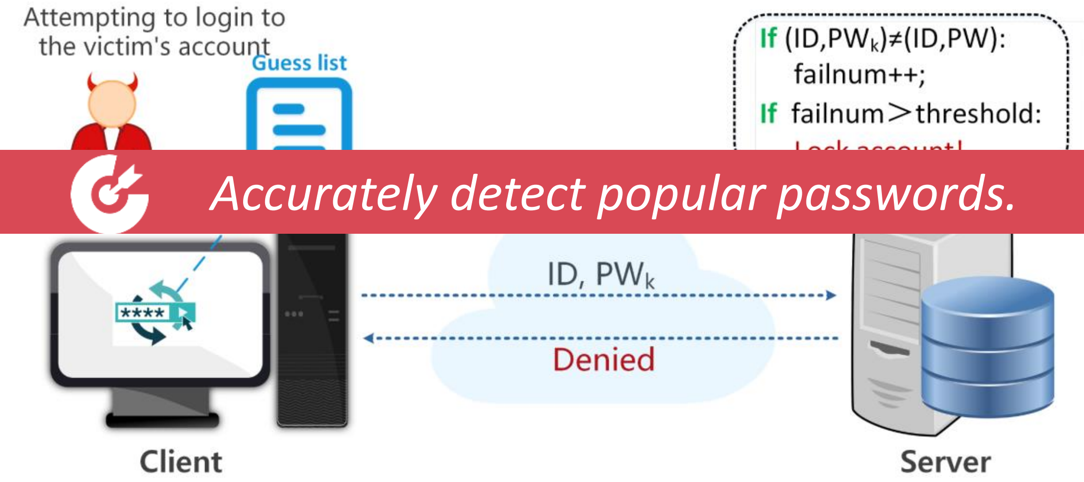
No single silver bullet metric.

▶ Password (PW) guessing threats



▶ Online guessing

⚠ Guess number ≤ 10^4



Accurately detect popular passwords.

▶ Offline guessing

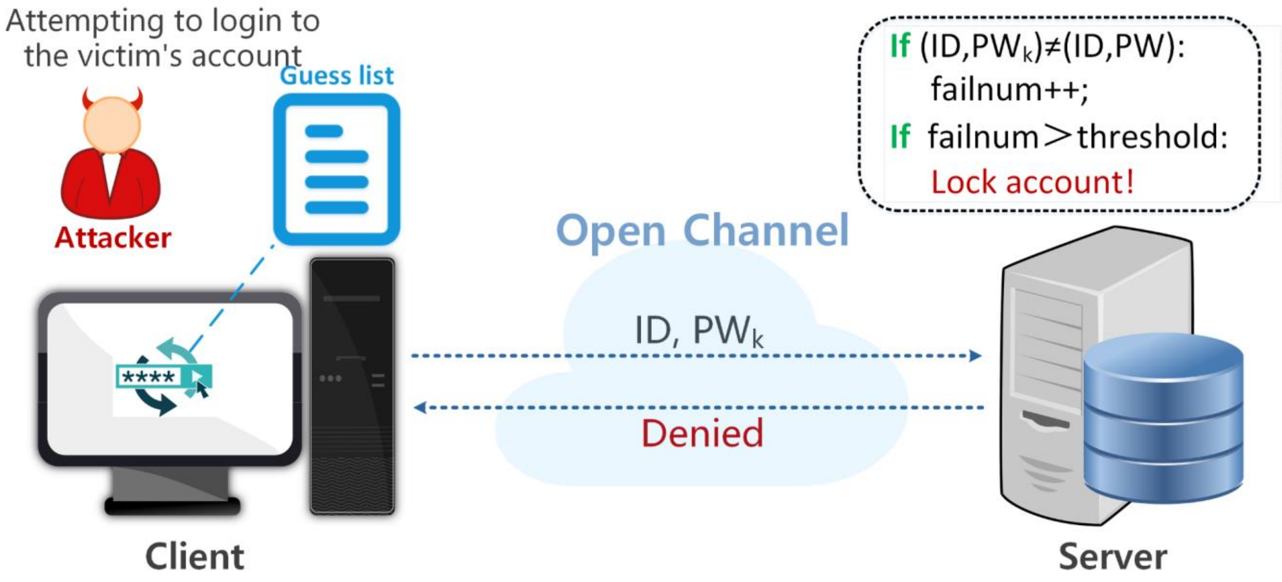
⚠ Guess number > 10^9

Prevent easy-to-crack passwords (not limited to popular passwords) from registering as much as possible.

Online guessing strategies

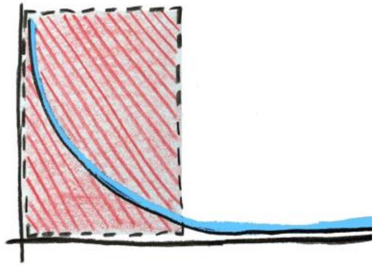
▶ Online guessing

⚠ Guess number ≤ 10^4



Knowledgeable online attacker

- Well-informed of the target password distribution.
- Prioritize trying the most popular passwords.



General online attacker

- Unaware of the target password distribution.
- Traverse common popular passwords.

Metrics for online guessing

- Due to the limited attempts allowed in online guessing scenarios, a PSM should accurately detect popular passwords (also the preferred guessing passwords of attackers), especially the most popular passwords (e.g., 123456)
- The higher the frequency of a password, the lower its strength.
- **Ideal PSM:** $\boldsymbol{M(pw) = \mathbf{Pr}(pw), \forall pw \in \Gamma}$
- The accuracy of a PSM can be measured as its distance from the ideal PSM, by calculating the correlation between its evaluated strength rank list and the referred rank list of the ideal PSM.

Ideal PSM

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

Tested PSM

Strength	Password
7	123456
3	password
1	123456789
2	12345678
82	qwerty
1309	cookie
430	7777777

Spearman *v.s.* Wspearman

✕ Spearman Correlation Coefficient

$$\mathbf{Spearman(X,Y)} = \frac{\sum_{i=1}^{n}[(x_i-\bar{x})(y_i-\bar{y})]}{\sqrt{\sum_{i=1}^{n}[(x_i-\bar{x})^2]\sum_{i=1}^{n}[(y_i-\bar{y})^2]}}$$

Ideal PSM

Rank	Password
1	123456
2	password
3	123456789
9	12345678
19	qwerty
63	cookie
154	7777777

Tested PSM

Rank	Password
7	123456
3	password
1	123456789
2	12345678
82	qwerty
1309	cookie
430	7777777

√ **Weighted** Spearman Correlation Coefficient [CCS'20]

$$\mathbf{WSpearman(X,Y)} = \frac{\sum_{i=1}^{n}[w_i(x_i-\bar{x})(y_i-\bar{y})]}{\sqrt{\sum_{i=1}^{n}[w_i(x_i-\bar{x})^2]\sum_{i=1}^{n}[w_i(y_i-\bar{y})^2]}}$$

Weight

Frequency
176120
140630
106179
83289
19123
13076
8600

Ideal PSM

Rank	Password
1	123456
2	password
3	123456789
9	12345678
19	qwerty
63	cookie
154	7777777

Tested PSM

Rank	Password
7	123456
3	password
1	123456789
2	12345678
82	qwerty
1309	cookie
430	7777777

Evaluation results in online guessing

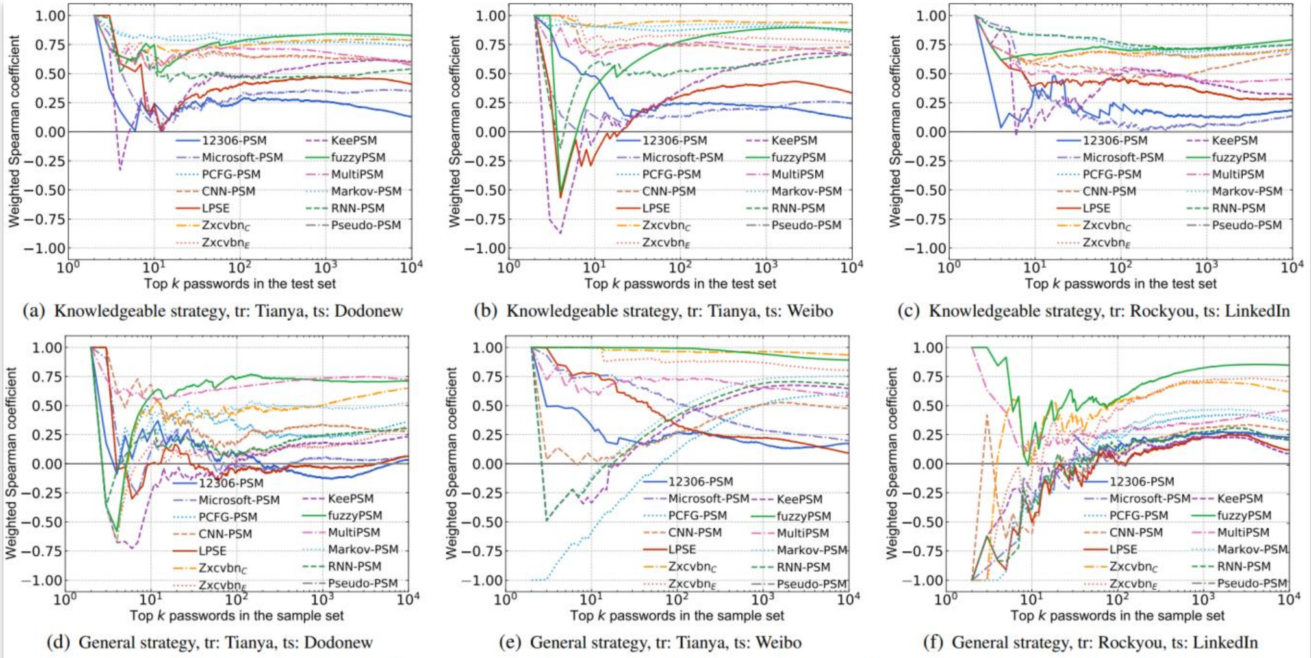
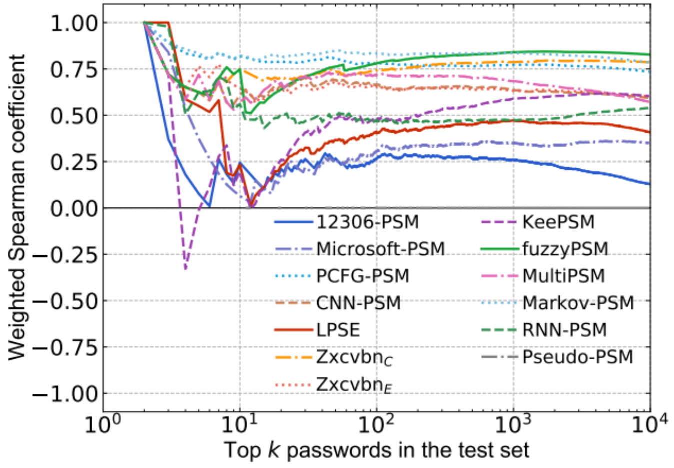
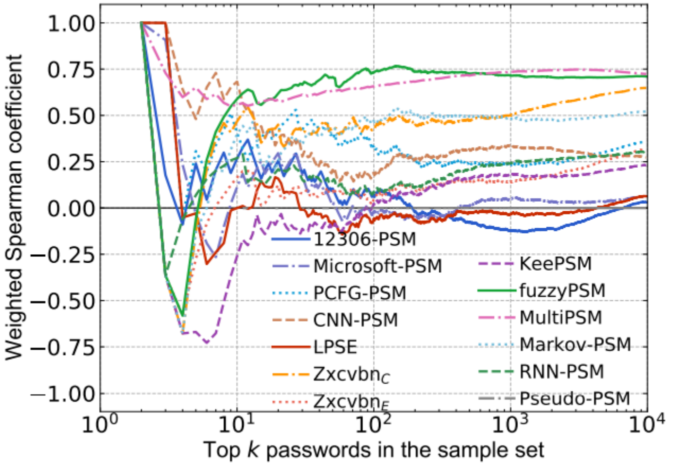


Figure 4: Weighted spearman correlation coefficient of state-of-the-art PSMs in online guessing scenarios (tr: Training set; ts: Test set).



(a) Knowledgeable strategy, tr: Tianya, ts: Dodonew



(d) General strategy, tr: Tianya, ts: Dodonew

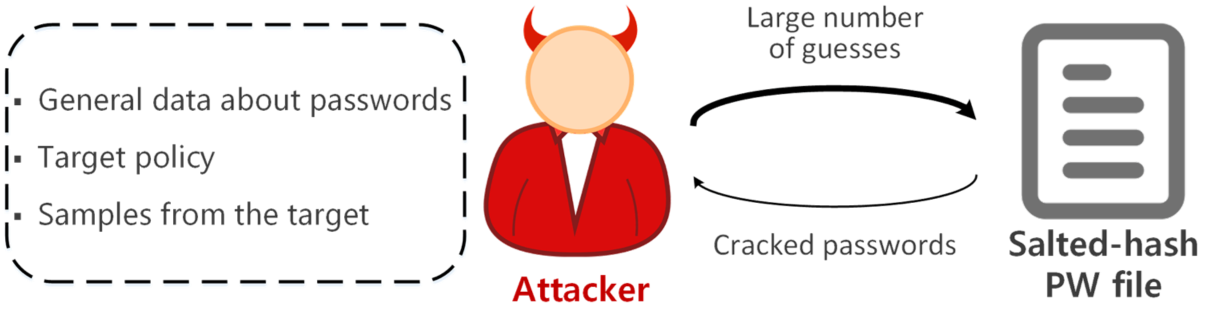
Overall conclusion

- FuzzyPSM performs the best, followed by Zxcvbn.
- Pattern-based and attack-algorithm-based PSMs need to be adapted for evaluating passwords in different languages.

Offline guessing strategies

▶ Offline guessing

⚠ Guess number > 10^9



Brute-force offline attacker

- Performs an exhaustive search over all possible passwords in a given search space.

Dictionary-based offline attacker

- Generates a guess list that contains several wordlists and candidate passwords extended by mangling rules.

Probability-based offline attacker

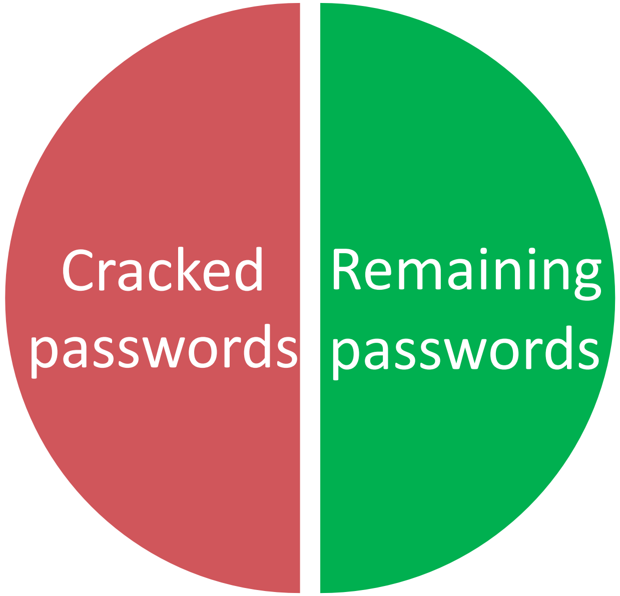
- Describe the target password distribution by parametric probability models (e.g., Markov), and produce guessing in the descending order of probability.

Combined offline attacker

- Try all these three guessing strategies to achieve higher cracking rates.

Metrics for offline guessing

- An attacker can perform large-scale (usually>10^9) guesses under offline guessing scenarios.
- She does not care about whether any specific guess is hit or not, but pursues a higher cracking rate under the constraints of computing resources and time.



A good PSM shall give differentiated strength ratings between the cracked and remaining (uncracked) passwords.

KL-divergence

$$\boldsymbol{KL}(P \parallel Q) = \sum_i P(i) \cdot \log \frac{P(i)}{Q(i)}$$

Evaluation results in offline guessing

Metric: KL-divergence

- MultiPSM generally performs the best.
- Markov-PSM and Zxcvbn are the most accurate PSMs under probability-based and dictionary-based guessing, respectively.

Table 4: KL-divergence of leading PSMs in offline guessing scenarios. Multi-PSM [19] performs best under brute-force and combined guessing strategies.†

PSM	KL	Chinese			English		
		Dodonew	CSDN	Weibo	LinkedIn	000web.	Yahoo
fuzzyPSM [17]	KL_{Bfa}	1.0788	2.5179	3.1533	0.8308	0.9268	0.7865
	KL_{Dic}	2.1009	6.7767	0.9021	3.0155	4.3028	3.7928
	KL_{Prob}	1.2467	1.1164	0.4478	1.3486	2.4993	1.5537
	KL_{All}	1.6193	3.1912	3.9749	2.2784	2.1186	3.3429
MultiPSM [19]	KL_{Bfa}	7.8006	2.0598	1.1105	8.5515	10.8753	6.4370
	KL_{Dic}	0.3519	0.2727	0.1161	0.5221	1.0597	0.2488
	KL_{Prob}	1.6762	1.8614	1.1436	5.4648	2.0630	2.0235
	KL_{All}	8.8994	7.2047	5.3692	7.5933	4.6160	3.9984
PCFG-PSM [25]	KL_{Bfa}	3.0046	1.4362	2.3564	2.2791	2.3951	1.2541
	KL_{Dic}	5.5556	5.9258	1.1966	3.4309	3.0581	1.8182
	KL_{Prob}	0.5047	1.1080	1.0369	1.2626	1.4452	0.9729
	KL_{All}	4.4277	2.3542	5.2805	6.0787	6.5533	2.9720
Markov-PSM [16]	KL_{Bfa}	1.2793	2.7560	3.8787	1.1933	0.9148	0.9439
	KL_{Dic}	3.0063	6.6061	3.8439	1.1163	1.6300	1.0145
	KL_{Prob}	2.0506	2.0341	0.7793	6.9454	6.9880	6.7100
	KL_{All}	2.2533	3.4217	3.8988	5.8993	1.3360	2.3893
RNN-PSM [18]	KL_{Bfa}	0.6410	0.9677	1.1538	1.0628	0.9133	0.6623
	KL_{Dic}	1.8715	2.9528	1.5388	3.4172	5.6578	2.0859
	KL_{Prob}	0.9765	1.6174	0.9335	2.6606	3.0942	1.7423
	KL_{All}	1.1992	1.5318	2.8269	4.1227	3.2999	6.3708
LPSE [13]	KL_{Bfa}	8.6483	6.7929	3.6468	4.0480	1.7946	1.3838
	KL_{Dic}	0.6029	0.6291	0.3974	0.3833	0.5417	0.3120
	KL_{Prob}	0.3492	0.4390	0.3814	1.8581	1.1519	1.2526
	KL_{All}	6.7505	1.8783	0.9242	6.7797	6.5693	4.6571
CNN-PSM [20]	KL_{Bfa}	1.2415	1.7885	1.3490	0.3299	0.7236	0.5631
	KL_{Dic}	4.1313	6.3760	4.1089	2.2732	2.6128	2.6266
	KL_{Prob}	0.3446	0.6661	0.5040	1.1542	2.5521	1.7725
	KL_{All}	3.0218	3.3403	2.8551	3.5520	2.1229	2.4907
$Zxcvbn_C$	KL_{Bfa}	2.1916	3.6398	2.9347	1.1732	1.7442	1.1001
	KL_{Dic}	2.6813	5.8958	2.7036	1.0440	2.1223	2.0496
	KL_{Prob}	0.7428	1.1549	0.7512	0.9996	1.7949	1.2733
	KL_{All}	2.8681	3.8673	4.4639	1.6443	2.0144	2.3132
$Zxcvbn_E$ [14]	KL_{Bfa}	2.8831	4.2145	2.7957	1.4231	1.5154	1.2219
	KL_{Dic}	4.3938	7.6722	4.0757	2.0382	3.7266	2.8444
	KL_{Prob}	0.9770	1.5533	1.1632	1.3931	1.8223	1.4169
	KL_{All}	4.3930	4.6782	5.4262	4.7708	2.6660	2.6944
KeePSM [15]	KL_{Bfa}	2.5697	2.2666	3.9500	2.8319	2.2526	2.8474
	KL_{Dic}	1.0812	2.0236	0.8438	0.1939	0.5237	0.3673
	KL_{Prob}	0.2886	0.4965	0.9081	1.0141	0.9805	0.9177
	KL_{All}	2.2322	2.6261	3.9691	1.6662	0.8765	1.3227
12306-PSM	KL_{Bfa}	1.6920	1.6607	0.3271	0.6190	1.6402	0.3176
	KL_{Dic}	0.2456	0.3219	0.1479	0.1504	0.1692	0.2607
	KL_{Prob}	0.0058	0.0127	0.0080	0.1873	0.4803	0.0319
	KL_{All}	1.4990	1.3592	0.2232	0.3698	0.8558	0.2026
Microsoft-PSM	KL_{Bfa}	3.7802	0.2433	0.3576	3.6622	5.6669	3.2280
	KL_{Dic}	0.6652	0.0574	0.1811	0.4127	0.4552	0.0112
	KL_{Prob}	0.0545	0.0364	0.0688	0.9957	0.8549	0.1223
	KL_{All}	3.3667	0.2370	0.2617	2.7974	3.0522	2.6380

†KL=KL-divergence; Bfa=Brute-force attack; Dic=Dictionary-based guessing; Prob=Probability-based guessing; All=The combined guessing. A line with background color means the corresponding PSM is the best under the given strategy.

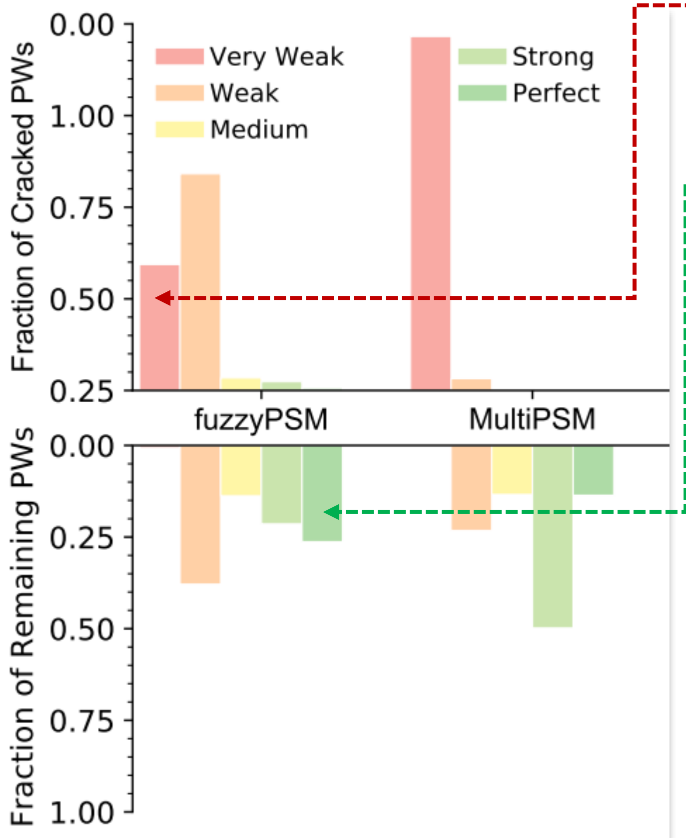
How will the accuracy of a PSM change?

Fine-grained strength feedback

e.g., probabilities, guess number…

Coarse-grained strength feedback

e.g., [weak, medium, strong] ….



L: Passwords with the lowest score;

$\boldsymbol{NC_L}$ and $\boldsymbol{NR_L}$: The number of the cracked passwords and the remaining ones.

H: Passwords with the highest score;

$\boldsymbol{NC_H}$ and $\boldsymbol{NR_H}$: The number of the cracked passwords and the remaining ones.

$$\boldsymbol{Precision} = W_L \times \frac{NC_L}{NR_L + NC_L} + W_H \times \frac{NR_H}{NR_H + NC_H}$$

$$W_L = \frac{NR_L + NC_L}{NR_L + NC_L + NR_H + NC_H} \qquad W_H = \frac{NR_H + NC_H}{NR_L + NC_L + NR_H + NC_H}$$

strong→weak

weak→strong

$$\boldsymbol{Precision}^{\boldsymbol{Security}} = \beta \times W_L \times \frac{NC_L}{NR_L + NC_L} + (1-\beta) \times W_H \times \frac{NR_H}{NR_H + NC_H}$$

Evaluation results in offline guessing

Metric: Precision & Precision^{Security}

- MultiPSM has the highest Precision and Precision^{Security} under the combined guessing attacks.
- Zxcvbn also shows its advantage under dictionary-based and probability-based guessing strategies.

Table 5: Precision of leading PSMs with the feedback of (transformed) bins/scores in offline guessing scenarios. LPSE [13], Zxcvbn$_E$ [14], Markov-PSM [16] and MultiPSM [19] perform the best under brute-force, dictionary-based, probability-based and combined guessing strategies, respectively.†

PSM	Prec	Chinese			English		
		Dodonew	CSDN	Weibo	LinkedIn	000web.	Yahoo
fuzzyPSM [17]	$Prec_{Bfa}$	0.6593	0.7356	0.9622	0.8099	0.9136	0.9182
	$Prec_{Dic}$	0.8320	0.8828	0.7896	0.8118	0.9722	0.9941
	$Prec_{Prob}$	0.7636	0.7042	0.1239	0.8738	0.9575	0.9750
	$Prec_{All}$	0.6842	0.7631	0.9811	0.7544	0.9604	0.9569
MultiPSM [19]	$Prec_{Bfa}$	0.8738	0.8608	0.9582	0.9676	0.6753	0.8721
	$Prec_{Dic}$	0.6134	0.6435	0.6653	0.9598	0.9358	0.9791
	$Prec_{Prob}$	0.3248	0.1462	0.1736	0.8842	0.9189	0.8433
	$Prec_{All}$	0.9971	0.9917	0.9978	0.9962	0.9918	0.9975
PCFG-PSM [25]	$Prec_{Bfa}$	0.9331	0.8492	0.8829	0.9122	0.9507	0.9057
	$Prec_{Dic}$	0.9706	0.9793	0.7531	0.9523	0.9792	0.9460
	$Prec_{Prob}$	0.3522	0.9895	0.2680	0.9886	0.9891	0.9945
	$Prec_{All}$	0.9241	0.8305	0.8136	0.8747	0.9344	0.8594
Markov-PSM [16]	$Prec_{Bfa}$	0.5706	0.5967	0.9214	0.6284	0.8954	0.7478
	$Prec_{Dic}$	0.7933	0.8058	0.8520	0.5314	0.9519	0.6039
	$Prec_{Prob}$	0.9333	0.9534	0.6133	0.9900	0.9991	0.9799
	$Prec_{All}$	0.5647	0.6038	0.9181	0.4347	0.8707	0.5056
RNN-PSM [18]	$Prec_{Bfa}$	0.6514	0.6425	0.5042	0.8778	0.9527	0.9180
	$Prec_{Dic}$	0.9535	0.9359	0.7356	0.9682	0.9961	0.9667
	$Prec_{Prob}$	0.9050	0.8540	0.7709	0.8609	0.9580	0.8537
	$Prec_{All}$	0.6270	0.6301	0.4788	0.8715	0.9503	0.9796
LPSE [13]	$Prec_{Bfa}$	1.0000	1.0000	0.5467	0.9999	0.9998	0.9992
	$Prec_{Dic}$	0.3505	0.6817	0.3272	0.5878	0.7074	0.6786
	$Prec_{Prob}$	0.2349	0.6326	0.0986	0.8030	0.7956	0.8243
	$Prec_{All}$	0.9993	0.9866	0.5465	0.9974	0.9968	0.9755
CNN-PSM [20]	$Prec_{Bfa}$	0.6975	0.7272	0.7046	0.7142	0.2628	0.5030
	$Prec_{Dic}$	0.4363	0.4505	0.4425	0.7951	0.2505	0.4415
	$Prec_{Prob}$	0.1963	0.0656	0.1274	0.6746	0.2675	0.4129
	$Prec_{All}$	0.7966	0.7869	0.7475	0.9231	0.4199	0.6666
Zxcvbn$_C$	$Prec_{Bfa}$	0.8926	0.8912	0.9586	0.8251	0.8707	0.9162
	$Prec_{Dic}$	0.8485	0.9206	0.8159	0.8404	0.9827	0.9711
	$Prec_{Prob}$	0.3210	0.2329	0.1856	0.8268	0.9728	0.9367
	$Prec_{All}$	0.9866	0.9794	0.9938	0.8831	0.9866	0.9943
Zxcvbn$_E$ [14]	$Prec_{Bfa}$	0.9561	0.9326	0.9752	0.9191	0.6990	0.9206
	$Prec_{Dic}$	0.9254	0.9325	0.8770	0.9609	0.9866	0.9796
	$Prec_{Prob}$	0.2639	0.1055	0.1942	0.9160	0.9385	0.9216
	$Prec_{All}$	0.9881	0.9802	0.9970	0.9962	0.9976	0.9986
KeePSM [15]	$Prec_{Bfa}$	0.6622	0.6939	0.9085	0.6967	0.2375	0.4374
	$Prec_{Dic}$	0.4281	0.4853	0.5685	0.7807	0.3105	0.3785
	$Prec_{Prob}$	0.2232	0.0879	0.1679	0.6598	0.2953	0.3510
	$Prec_{All}$	0.7805	0.7753	0.9678	0.9158	0.4442	0.5766
12306-PSM	$Prec_{Bfa}$	0.9105	0.8295	0.4205	0.6799	0.9406	0.9978
	$Prec_{Dic}$	0.2400	0.2859	0.2519	0.4339	0.3958	0.8217
	$Prec_{Prob}$	0.1208	0.0657	0.0655	0.4389	0.5917	0.9210
	$Prec_{All}$	0.9341	0.8643	0.4320	0.7403	0.9399	0.9995
Microsoft-PSM	$Prec_{Bfa}$	0.9979	0.9985	0.5161	0.9952	0.9822	0.9937
	$Prec_{Dic}$	0.5596	0.3134	0.3082	0.6488	0.4385	0.7556
	$Prec_{Prob}$	0.1531	0.1345	0.0837	0.7743	0.5624	0.8476
	$Prec_{All}$	0.9982	0.9929	0.5164	0.9948	0.9806	0.9937

† Prec=Precision; Other abbreviations are the same with Table 4. A line with background color means the corresponding PSM is the best under the given strategy.

Table 6: Precision^{Security} of leading PSMs with the feedback of (transformed) bins/scores in offline guessing scenarios. MultiPSM [19] is still the most accurate PSM under brute-force and combined guessing strategies.†

PSM	$Prec^{Sec}$	Chinese			English		
		Dodonew	CSDN	Weibo	LinkedIn	000web.	Yahoo
fuzzyPSM [17]	$Prec^{Sec}_{Bfa}$	0.2993	0.3226	0.7526	0.2772	0.2139	0.3551
	$Prec^{Sec}_{Dic}$	0.3228	0.3576	0.6121	0.2861	0.2586	0.3907
	$Prec^{Sec}_{Prob}$	0.2066	0.1582	0.0736	0.2915	0.2452	0.3767
	$Prec^{Sec}_{All}$	0.3249	0.3462	0.7702	0.2791	0.2571	0.3868
MultiPSM [19]	$Prec^{Sec}_{Bfa}$	0.7978	0.7519	0.4118	0.7809	0.7303	0.7930
	$Prec^{Sec}_{Dic}$	0.4884	0.5142	0.5270	0.7666	0.7235	0.7832
	$Prec^{Sec}_{Prob}$	0.2574	0.1163	0.1336	0.7061	0.7100	0.6745
	$Prec^{Sec}_{All}$	0.7883	0.7836	0.7923	0.7987	0.7729	0.7988
PCFG-PSM [25]	$Prec^{Sec}_{Bfa}$	0.5760	0.1749	0.6153	0.3078	0.1913	0.1878
	$Prec^{Sec}_{Dic}$	0.5751	0.2008	0.5245	0.3134	0.1970	0.1958
	$Prec^{Sec}_{Prob}$	0.0717	0.1979	0.0582	0.3183	0.1986	0.2055
	$Prec^{Sec}_{All}$	0.5742	0.1712	0.6014	0.3003	0.1881	0.1785
Markov-PSM [16]	$Prec^{Sec}_{Bfa}$	0.1732	0.2026	0.3295	0.2845	0.2009	0.3108
	$Prec^{Sec}_{Dic}$	0.2047	0.2407	0.2686	0.2672	0.2203	0.2922
	$Prec^{Sec}_{Prob}$	0.2017	0.1833	0.1732	0.3375	0.2209	0.3425
	$Prec^{Sec}_{All}$	0.1796	0.2156	0.3331	0.2523	0.2056	0.2764
RNN-PSM [18]	$Prec^{Sec}_{Bfa}$	0.1588	0.1543	0.3561	0.2476	0.1959	0.7231
	$Prec^{Sec}_{Dic}$	0.2174	0.2118	0.3675	0.2679	0.2051	0.7651
	$Prec^{Sec}_{Prob}$	0.1919	0.1964	0.1554	0.2738	0.2057	0.7682
	$Prec^{Sec}_{All}$	0.1543	0.1539	0.3533	0.2514	0.1960	0.7780
LPSE [13]	$Prec^{Sec}_{Bfa}$	0.7748	0.4265	0.4191	0.7554	0.4563	0.6329
	$Prec^{Sec}_{Dic}$	0.2555	0.1779	0.2437	0.4272	0.2240	0.3905
	$Prec^{Sec}_{Prob}$	0.1630	0.1377	0.0608	0.5980	0.2932	0.4942
	$Prec^{Sec}_{All}$	0.7747	0.4239	0.4191	0.7549	0.4557	0.6284
CNN-PSM [20]	$Prec^{Sec}_{Bfa}$	0.5580	0.5817	0.5636	0.5713	0.1951	0.4023
	$Prec^{Sec}_{Dic}$	0.3490	0.3604	0.3540	0.6361	0.1852	0.3532
	$Prec^{Sec}_{Prob}$	0.1570	0.0524	0.1019	0.5397	0.1989	0.3303
	$Prec^{Sec}_{All}$	0.6373	0.6295	0.5980	0.7385	0.3208	0.5333
Zxcvbn$_C$	$Prec^{Sec}_{Bfa}$	0.6983	0.6399	0.7586	0.6306	0.2743	0.6610
	$Prec^{Sec}_{Dic}$	0.6599	0.6580	0.6442	0.6534	0.3687	0.7073
	$Prec^{Sec}_{Prob}$	0.2376	0.1064	0.1398	0.6316	0.3560	0.6773
	$Prec^{Sec}_{All}$	0.7739	0.7107	0.7869	0.6876	0.3719	0.7260
Zxcvbn$_E$ [14]	$Prec^{Sec}_{Bfa}$	0.7560	0.7253	0.7679	0.7314	0.3725	0.7027
	$Prec^{Sec}_{Dic}$	0.7314	0.7237	0.6890	0.7670	0.6053	0.7506
	$Prec^{Sec}_{Prob}$	0.2030	0.4632	0.1431	0.7290	0.5645	0.7035
	$Prec^{Sec}_{All}$	0.7900	0.7742	0.7867	0.7953	0.6095	0.7649
KeePSM [15]	$Prec^{Sec}_{Bfa}$	0.5297	0.5551	0.7268	0.5573	0.1900	0.3499
	$Prec^{Sec}_{Dic}$	0.3425	0.3883	0.4548	0.6245	0.2484	0.3028
	$Prec^{Sec}_{Prob}$	0.1785	0.0703	0.1343	0.5278	0.2362	0.2808
	$Prec^{Sec}_{All}$	0.6244	0.6202	0.7742	0.7326	0.3553	0.4613
12306-PSM	$Prec^{Sec}_{Bfa}$	0.7250	0.6603	0.3360	0.5419	0.6981	0.7981
	$Prec^{Sec}_{Dic}$	0.1886	0.2255	0.2011	0.3451	0.2627	0.6573
	$Prec^{Sec}_{Prob}$	0.0934	0.0493	0.0519	0.3492	0.4200	0.7367
	$Prec^{Sec}_{All}$	0.7440	0.6882	0.3451	0.5903	0.6988	0.7995
Microsoft-PSM	$Prec^{Sec}_{Bfa}$	0.6968	0.6880	0.7613	0.7728	0.5150	0.6976
	$Prec^{Sec}_{Dic}$	0.4471	0.2071	0.2455	0.5041	0.2976	0.6025
	$Prec^{Sec}_{Prob}$	0.1219	0.0622	0.0659	0.6044	0.3963	0.6762
	$Prec^{Sec}_{All}$	0.7981	0.7508	0.4121	0.7812	0.7320	0.7930

†$Prec^{Sec}$=Precision^{Security}; Other abbreviations are the same with Table 4. Background color indicates the corresponding PSM is the best under the given strategy.

Insights and suggestions

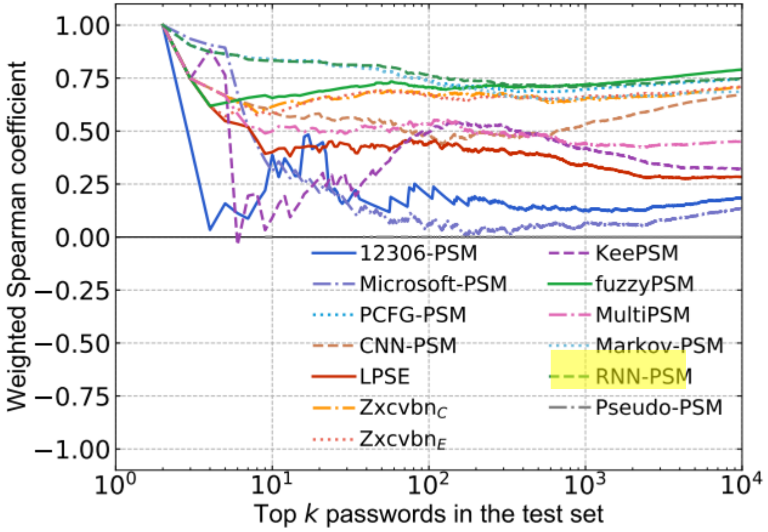
Adaptive score conversion methods can be used to facilitate PSM application.

Fine-grained strength feedback

e.g., probabilities, guess number…

Coarse-grained strength feedback

e.g., [weak, medium, strong] ….



(c) Knowledgeable strategy, tr: Rockyou, ts: LinkedIn

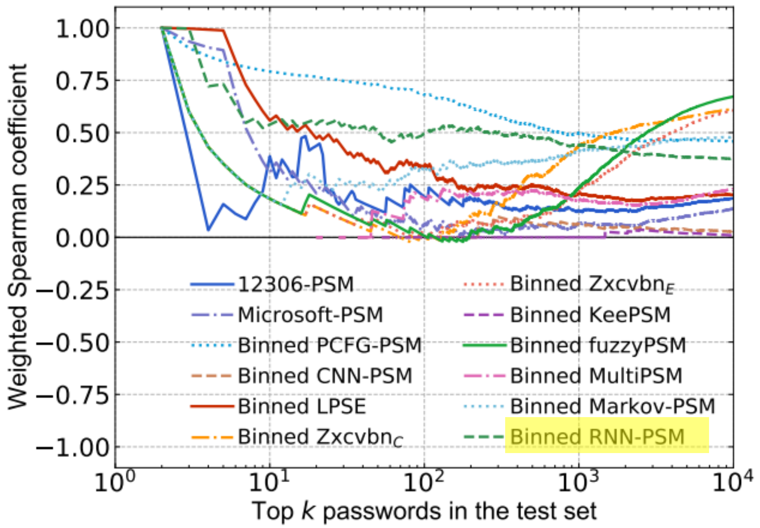


Figure 6: WSpearman of binned-PSMs under the knowledgeable online guessing strategy (tr: Rockyou, ts: LinkedIn). Binning degrades accurate PSMs while improving inaccurate ones. See Appendix F for more results.

Insights and suggestions

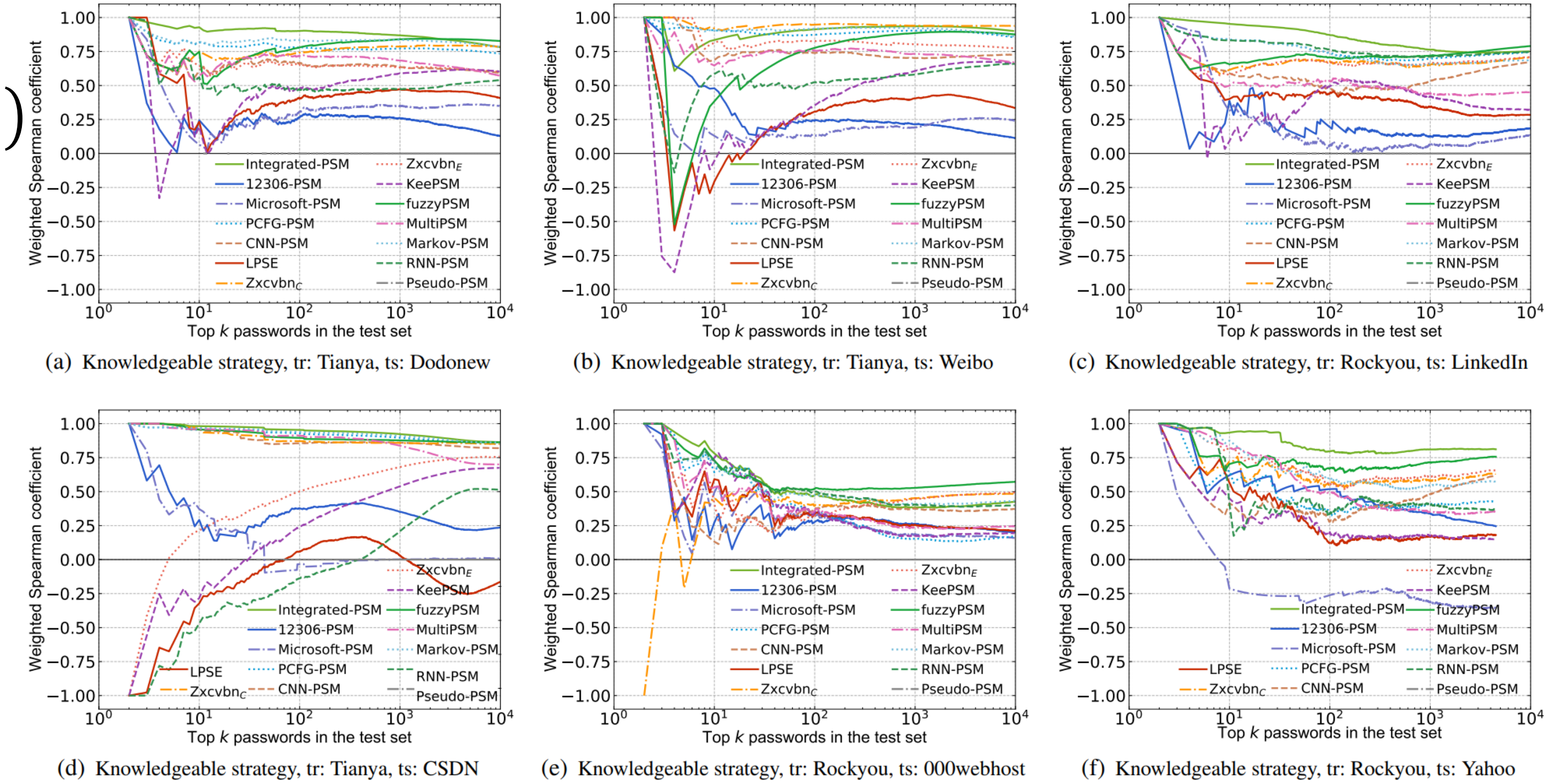
PSMs can be effectively integrated to perform better.

$$\mathbf{Strength}_{Integrated} = \mathbf{Strength}_{Multi} - \lg\left(\mathbf{Strength}_{fuzzy}\right)$$

Table 13: KL-divergence of fuzzyPSM [17], MultiPSM [19] and our proposed Integrated-PSM. Integrated-PSM performs better than its components.†

PSM	KL	Chinese			English		
		Dodonew	CSDN	Weibo	LinkedIn	000web.	Yahoo
fuzzyPSM [17]	KL_{Bfa}	1.0788	2.5179	3.1533	0.8308	0.9268	0.7865
	KL_{Dic}	2.1009	6.7767	0.9021	3.0155	4.3028	3.7928
	KL_{Prob}	1.2467	1.1164	0.4478	1.3486	2.4993	1.5537
	KL_{All}	1.6193	3.1912	3.9749	2.2784	2.1186	3.3429
MultiPSM [19]	KL_{Bfa}	7.8006	2.0598	1.1105	8.5515	10.8753	6.4370
	KL_{Dic}	0.3519	0.2727	0.1161	0.5221	1.0597	0.2488
	KL_{Prob}	1.6762	1.8614	1.1436	5.4648	2.0630	2.0235
	KL_{All}	8.8994	7.2047	5.3692	7.5933	4.6160	3.9984
Integrated-PSM	KL_{Bfa}	**8.2858**	**3.3781**	**4.0845**	6.3886	**11.4628**	**13.7336**
	KL_{Dic}	0.5479	0.2455	**2.1118**	1.7112	1.2068	0.3340
	KL_{Prob}	**2.3865**	**3.9533**	**1.3915**	**8.0426**	1.2102	**2.1179**
	KL_{All}	**9.1274**	2.4857	**5.7832**	2.5356	**6.9474**	2.4715

† A **bold** value indicates that the KL-divergence of Integrated-PSM is *higher* than fuzzyPSM [17] and MultiPSM [19] under the given strategy.

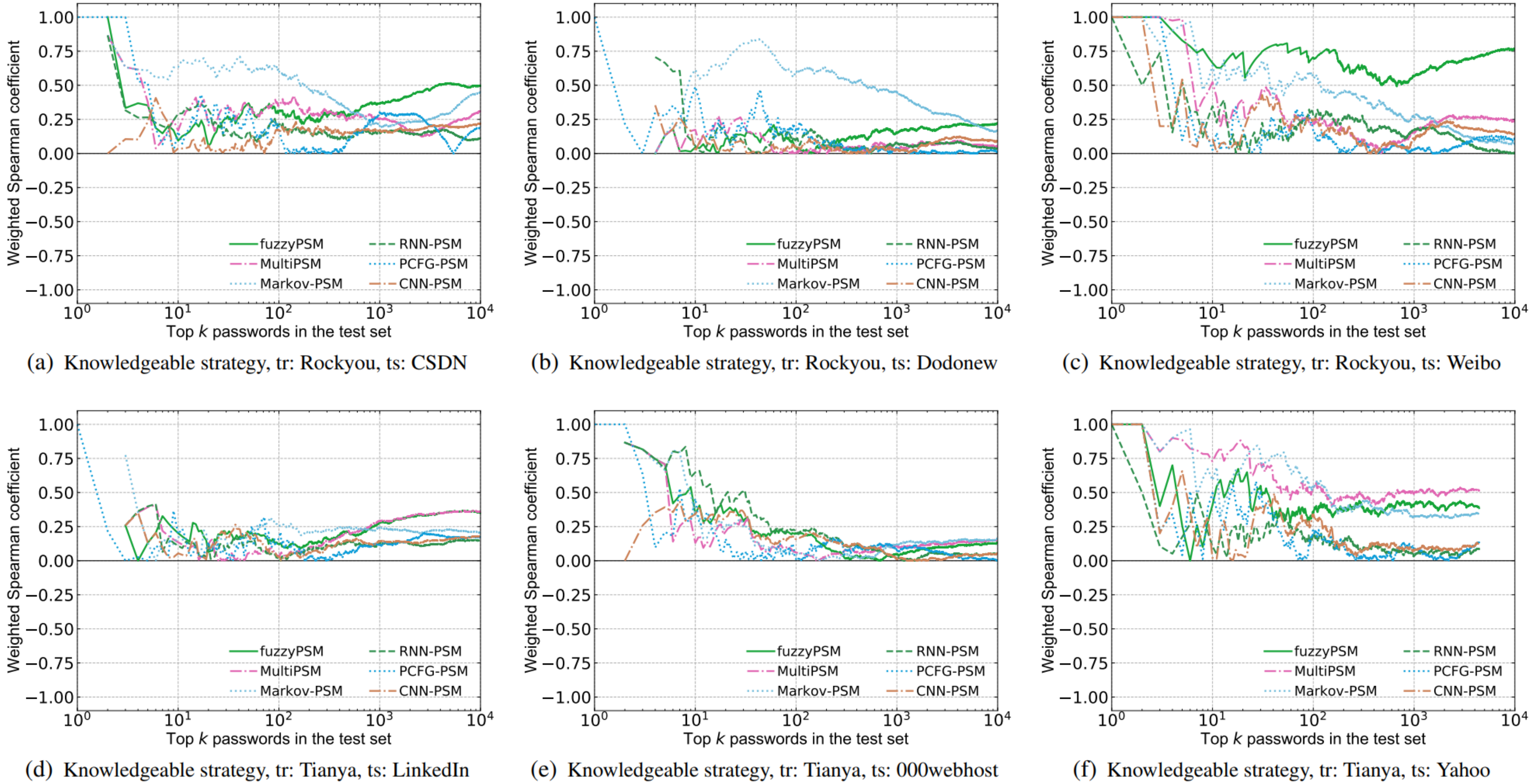


(a) Knowledgeable strategy, tr: Tianya, ts: Dodonew
(b) Knowledgeable strategy, tr: Tianya, ts: Weibo
(c) Knowledgeable strategy, tr: Rockyou, ts: LinkedIn
(d) Knowledgeable strategy, tr: Tianya, ts: CSDN
(e) Knowledgeable strategy, tr: Rockyou, ts: 000webhost
(f) Knowledgeable strategy, tr: Rockyou, ts: Yahoo

Figure 11: WSpearman of Integrated-PSM and other 12 PSMs under the knowledgeable online guessing strategy (tr: Training set; ts: Test set).

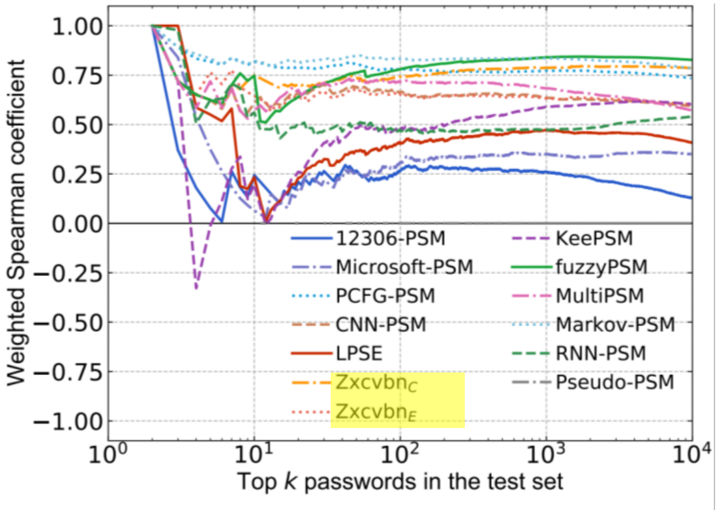
Insights and suggestions

PSMs need to be modified and adapted to accommodate different languages.

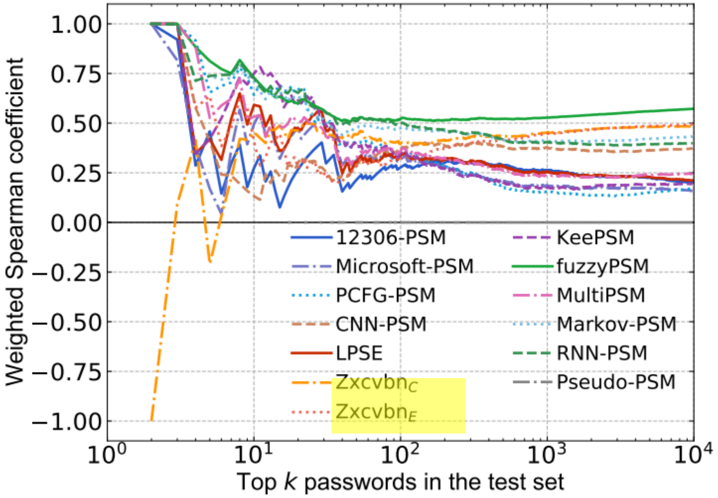


(a) Knowledgeable strategy, tr: Rockyou, ts: CSDN
(b) Knowledgeable strategy, tr: Rockyou, ts: Dodonew
(c) Knowledgeable strategy, tr: Rockyou, ts: Weibo
(d) Knowledgeable strategy, tr: Tianya, ts: LinkedIn
(e) Knowledgeable strategy, tr: Tianya, ts: 000webhost
(f) Knowledgeable strategy, tr: Tianya, ts: Yahoo

Figure 12: Influence of training set language on the accuracy of six attack algorithm-based PSMs under the knowledgeable online guessing strategy.



(a) Knowledgeable strategy, tr: Tianya, ts: Dodonew



(b) Knowledgeable strategy, tr: Rockyou, ts: 000webhost

Thank you!

No Single Silver Bullet: Measuring the Accuracy of Password Strength Meters

Ding Wang, Xuan Shan, Qiying Dong
Nankai University
{wangding, shanxuan, dqy}@nankai.edu.cn

Yaoshen Shen
Peking University
ysshen@pku.edu.cn

Chunfu Jia
Nankai University
cfjia@nankai.edu.cn