

# A Large-Scale Empirical Analysis of Chinese Web Passwords

Zhigong Li, Weili Han

*Software School, Fudan University*

Wenyuan Xu

*Department of Electronic Engineering, Zhejiang University*



YAHOO!

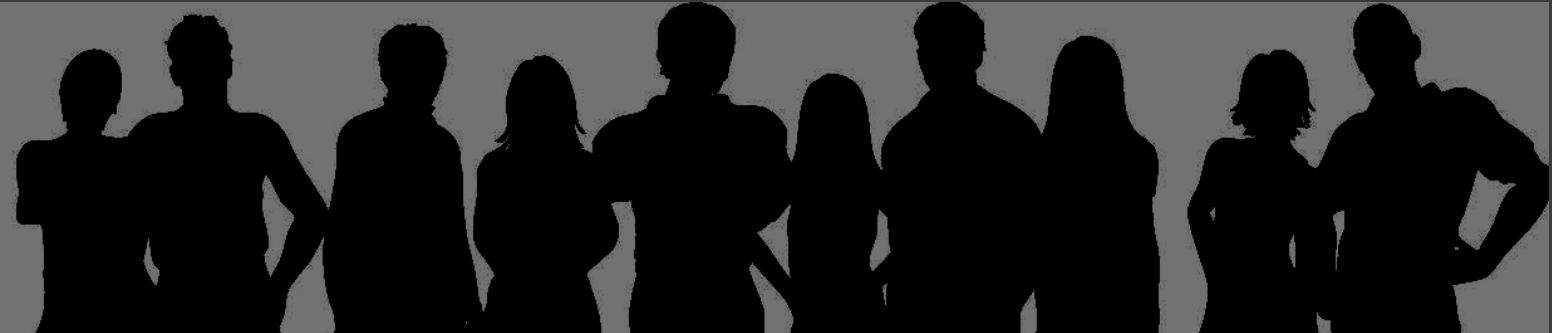
CSDN  
全球最大中文IT社区

Google

天涯社区  
www.tianya.cn

\*\*\*\*\*

Users in diff regions use diff pwd patterns



# Two Problems

1 *Does Chinese choose better passwords?*

2 *How can we guess them?*



# Why Chinese Passwords?

There are over **600 million**  
Netizens in China.



[http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201407/t20140721\\_47437.htm](http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201407/t20140721_47437.htm)

# Password Leakage



English

rockyou

YAHOO!

Chinese

CSDN

全球最大中文 IT 社区

天涯社区

www.tianya.cn



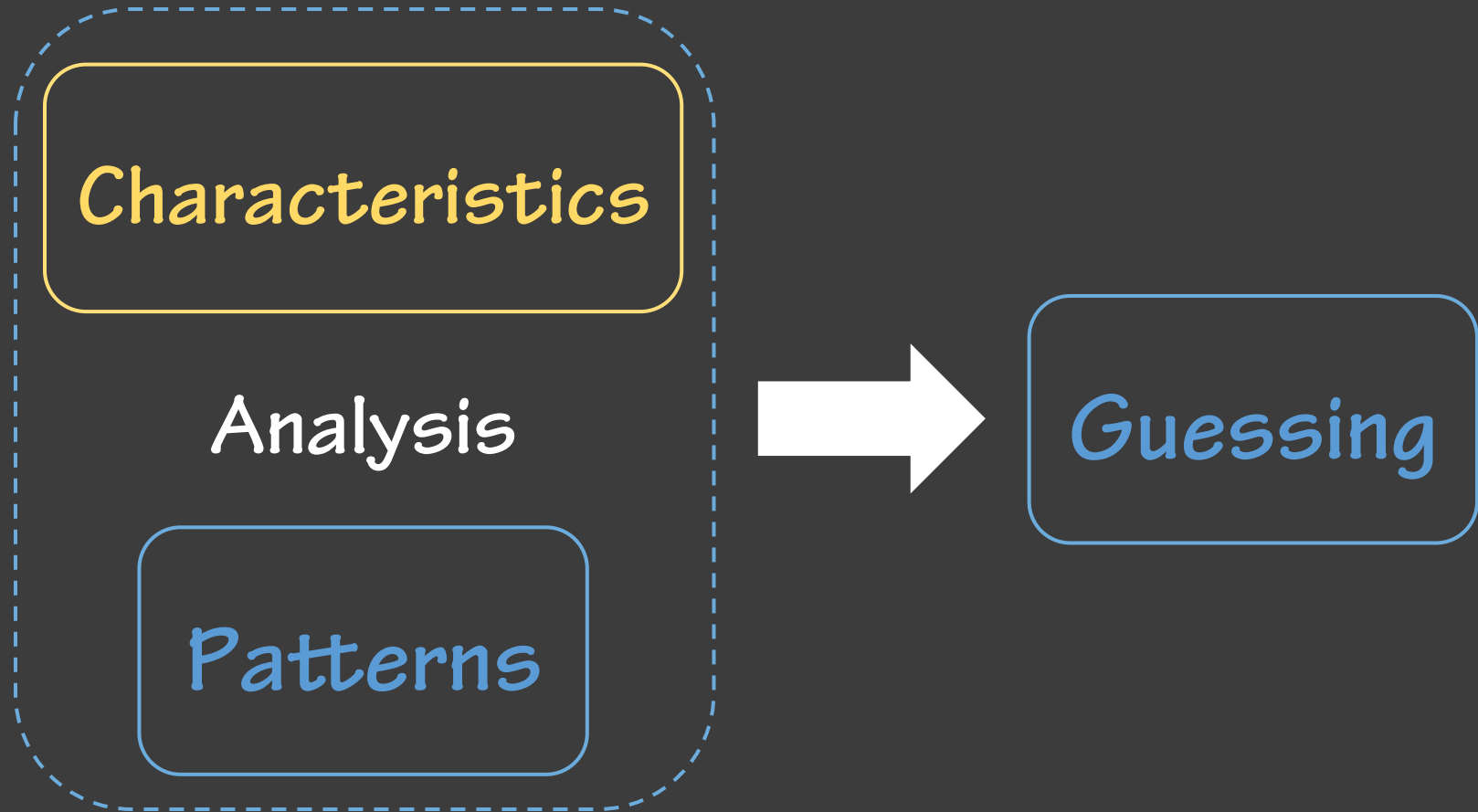
嘟嘟牛  
WWW.DODONEW.COM

7k7k

178.COM

Over 100 million plaintext passwords

# Methodologies -- Two stage analysis



# What are the Most Popular Passwords

	Chinese	English
1	123456 (2.17%)	123456 (0.88%)
2	123456789 (0.65%)	12345 (0.24%)
3	111111 (0.59%)	123456789 (0.23%)
4	12345678 (0.39%)	password (0.18%)
5	000000 (0.34%)	iloveyou (0.15%)





# PWD Strength Metrics -- alpha-work-factor

Alpha

Probability of success

Alpha-work-factor

Expected Number of guesses needed to succeed with probability alpha

[1] J. O. Pliam. *On the incomparability of entropy and marginal guesswork in brute-force attacks*, In Cryptology-INDOCRYPT 2000.

[2] J. Bonneau. *The science of guessing: analyzing an anonymized corpus of 70 million passwords*, In IEEE S&P 2012.



# PWD Strength Metrics -- alpha-work-factor

passwords	40%
123456	30%
iloveyou	20%
123456789	10%

Example Dataset

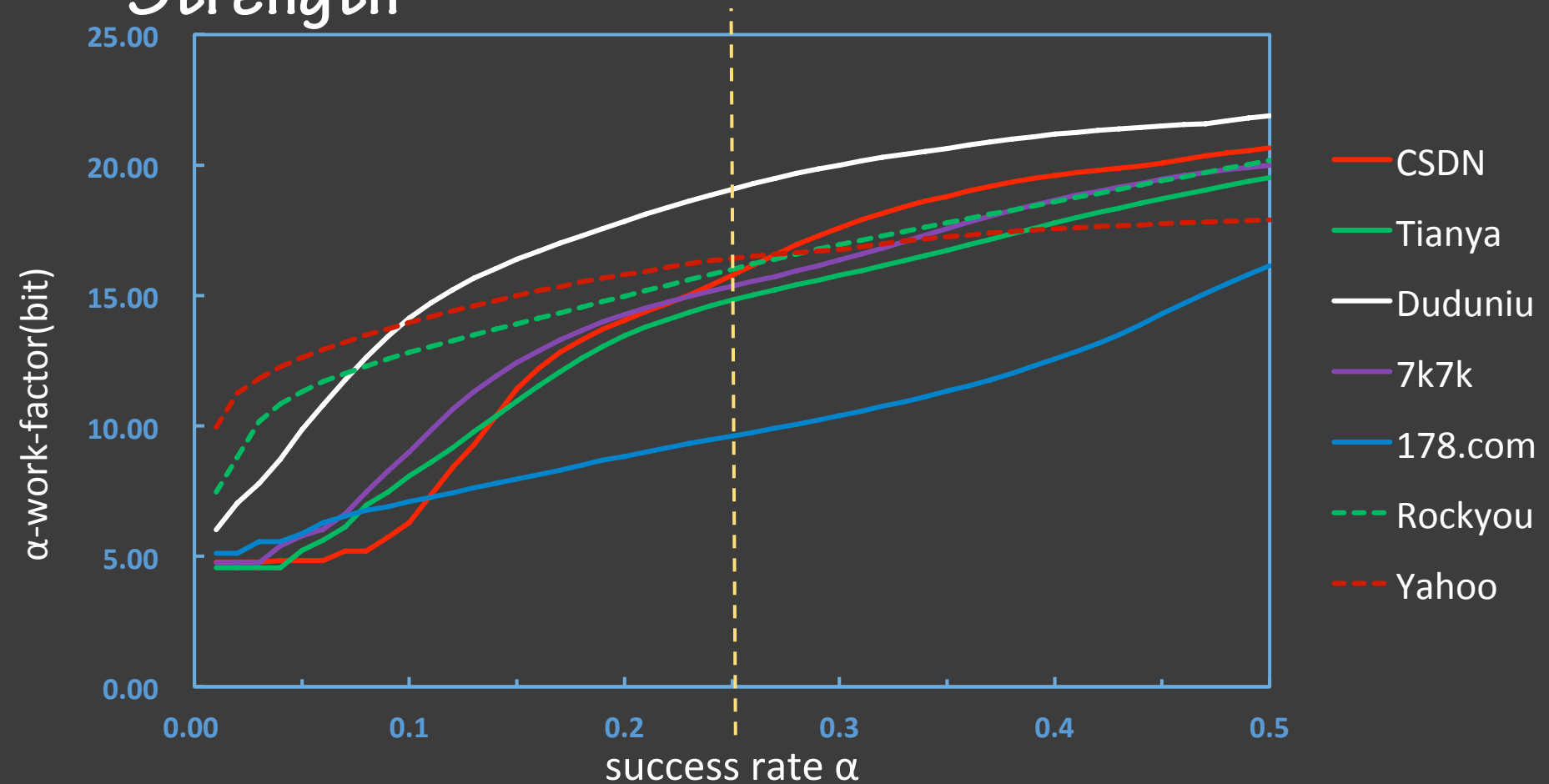
alpha=0.4

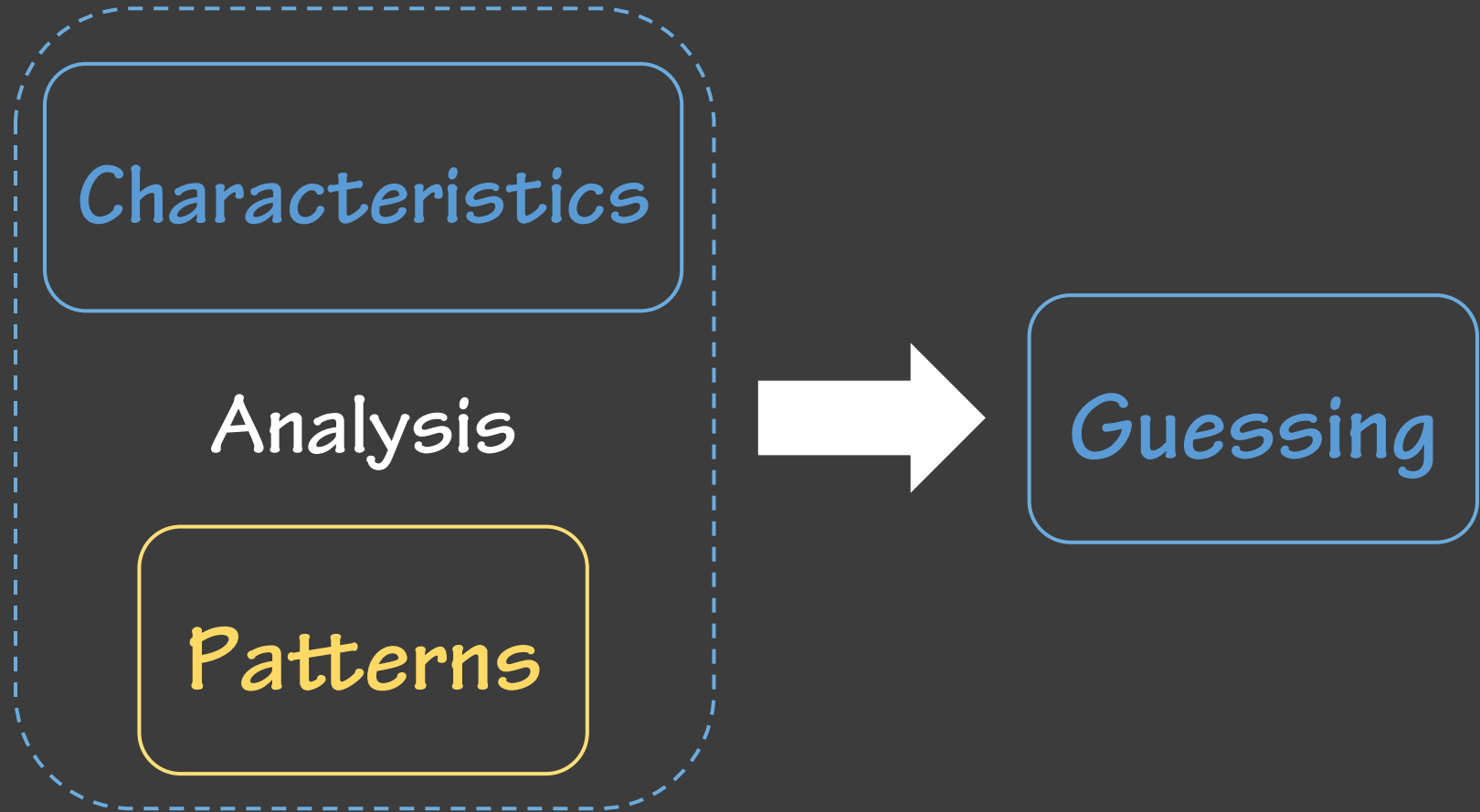
alpha-work-factor = 1

alpha=0.9

alpha-work-factor = 3

# PWD Strength Result -- Similar Strength





# PWD Patterns of Chinese -- Pinyins

nǐ



hǎo



# Pinyins/Words -- Letter only

Letter-Only Passwords	Chinese Pinyins	English Words
CSDN	41.61 % (5.15%)	15.59% (1.93%)
Tianya	40.63% (4.15%)	10.39% (1.06%)
Duduniu	33.28% (3.91%)	15.35% (1.80%)
7k7k	44.70% (4.97%)	10.04% (1.02%)
178.com	57.31% (5.25%)	2.20% (0.20%)
Rockyou	6.94% (2.99%)	25.47% (10.98%)
Yahoo	4.31% (1.46%)	34.92% (11.86%)

# Pinyins/Words -- Mixed

Mixed Passwords	Chinese Pinyins	English Words
CSDN	25.49% (10.68%)	7.97% (3.34%)
Tianya	23.59% ( 5.78%)	6.05% (1.48%)
Duduniu	25.17% (13.87%)	6.48% (3.57%)
7k7k	21.09% ( 5.84%)	7.02% (1.94%)
178.com	23.49% ( 9.97%)	4.58% (1.94%)
Rockyou	6.88% ( 2.61%)	28.11% (10.65%)
Yahoo	4.53% ( 2.59%)	27.99% (16.01%)

# Special Passwords -- Love

	Top Chinese Pinyins	Top English Words
1	<b>woaini</b> (1.47%)	password (1.28%)
2	li (1.06%)	<b>iloveyou</b> (0.98%)
3	wang (0.97%)	<b>love</b> (0.76%)
4	tianya (0.89%)	angel (0.59%)
5	zhang (0.84%)	monkey (0.45%)



# Dates -- Extraction

a**123456**b



a1234567b



a**12345678**b



a123456789b



a1234b56789



# Dates -- Eight digits

YYYYMMDD MMDDYYYY DDMMYYYY

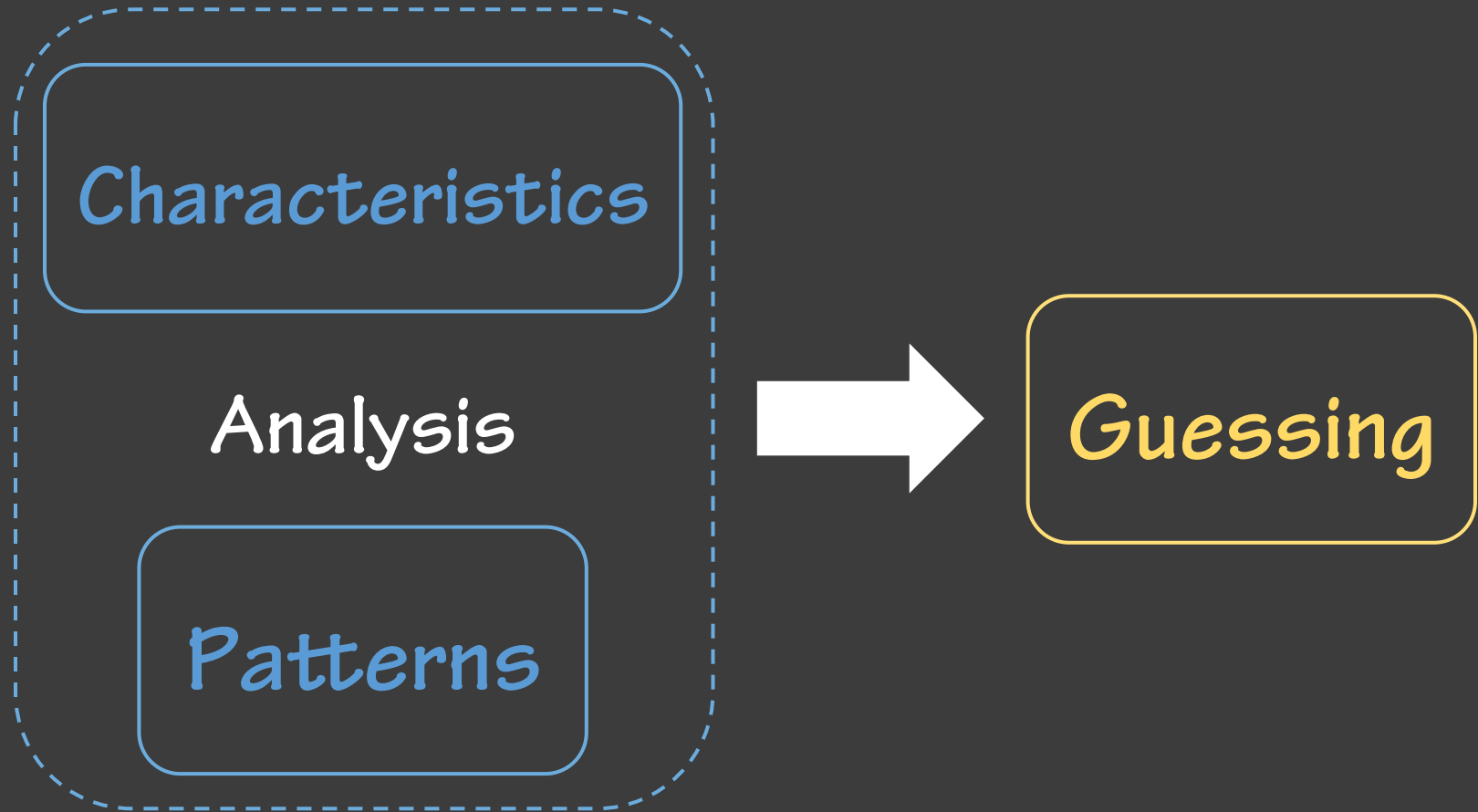
CSDN	29.24%	0.25%	0.43%
Tianya	36.26%	0.35%	0.60%
Duduniu	28.87%	0.28%	0.84%
7k7k	32.41%	0.18%	0.37%
178.com	30.46%	0.13%	0.19%
Rockyou	2.64%	7.70%	17.66%
Yahoo	2.78%	12.00%	11.17%

# Dates -- Six digits

	YYMMDD	MMDDYY	DDMMYY
CSDN	27.21%	4.04%	1.24%
Tianya	23.93%	3.05%	1.19%
Duduniu	17.84%	2.97%	1.78%
7k7k	24.34%	2.63%	0.88%
178.com	13.96%	1.72%	1.30%
Rockyou	5.63%	21.90%	18.42%
Yahoo	4.66%	25.99%	7.77%

# Dates at the End

	Beginning	Middle	End
CSDN	21.68%	4.32%	74.00%
Tianya	27.33%	4.75%	67.07%
Duduniu	24.76%	1.36%	73.88%
7k7k	32.17%	2.70%	65.13%
178.com	22.30%	1.03%	76.67%
Rockyou	27.40%	3.91%	68.69%
Yahoo	22.66%	5.00%	72.34%

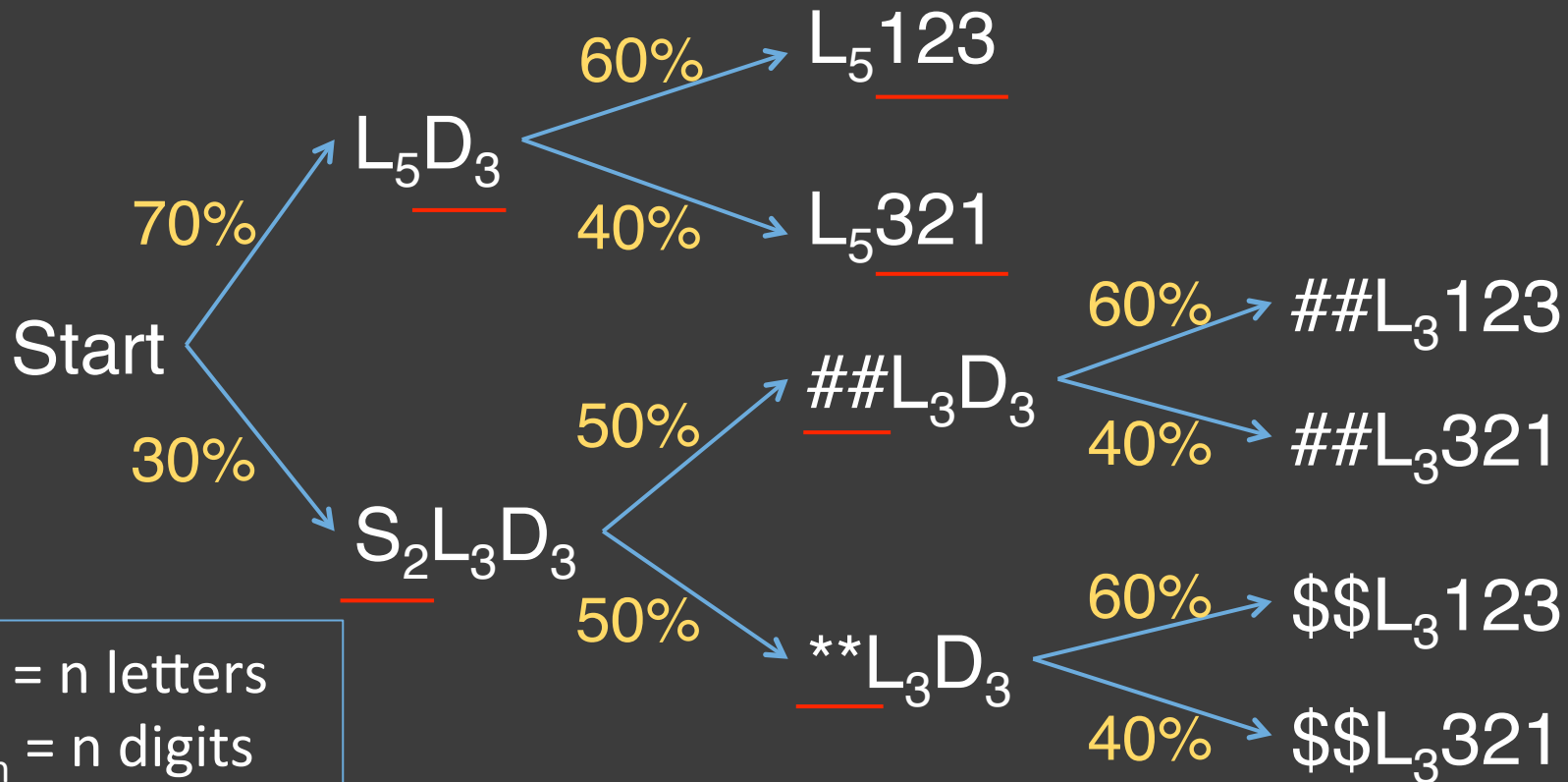


# Guessing based on Probabilistic Context-Free Grammar

- Describe the structures of passwords using a set of **rules**.
- **Dictionary** guessing

[1]Matt Weir et al. *Password Cracking Using Probabilistic Context-Free Grammar*, In IEEE Symposium on Security and Privacy, 2009.

# Example



$L_n = n$  letters  
 $D_n = n$  digits  
 $S_n = n$  symbols

# Guessing CSDN

## Training Sets

RockyouTS

MRockyouTS

RockyouDuduTS

DuduTS





# Guessing CSDN -- Dictionaries

{ **EDict**: *Dic-0294* and *English-lower*  
20,000 Most Popular **Pinyins** } **CDict**

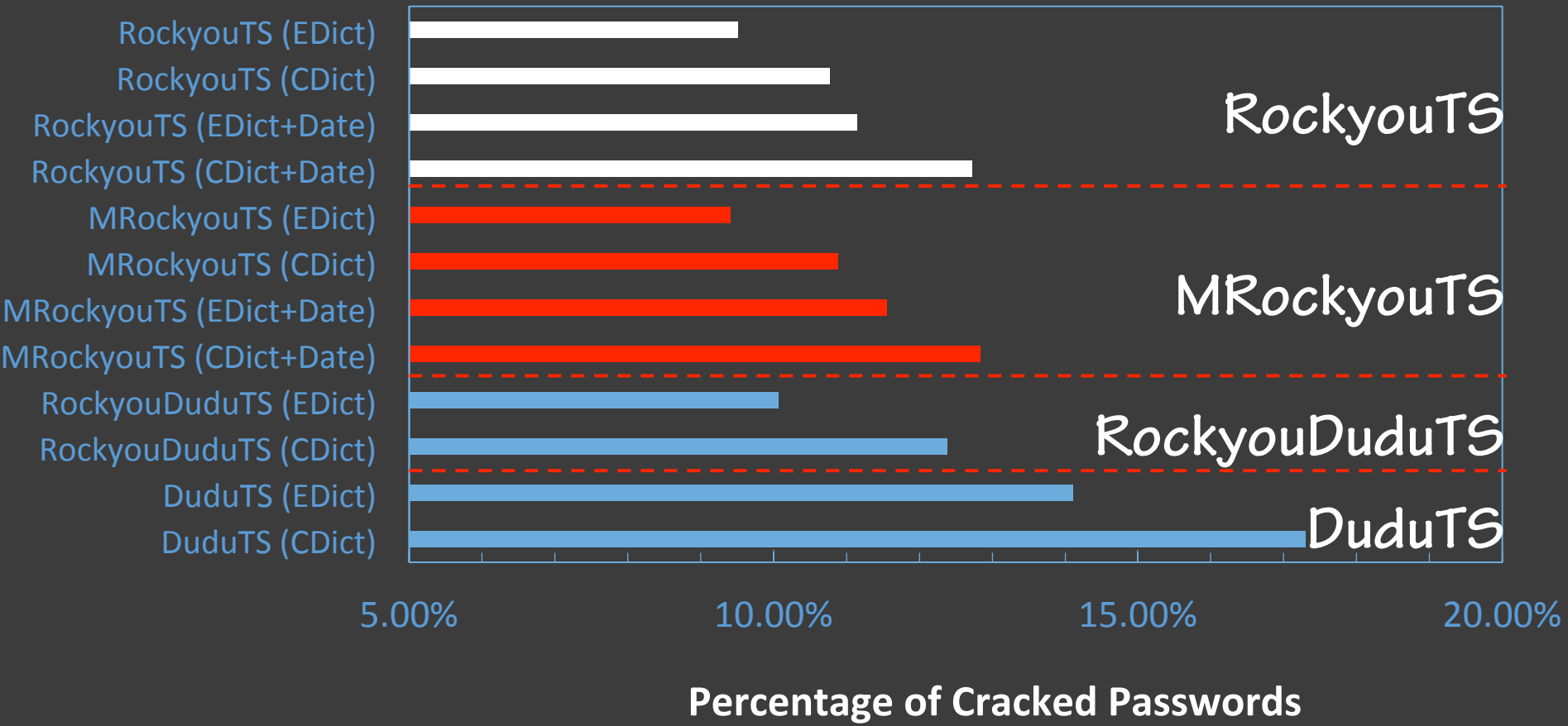
20,000 Most Popular Six-digit **Dates**

20,000 Most Popular Eight-digit **Dates**

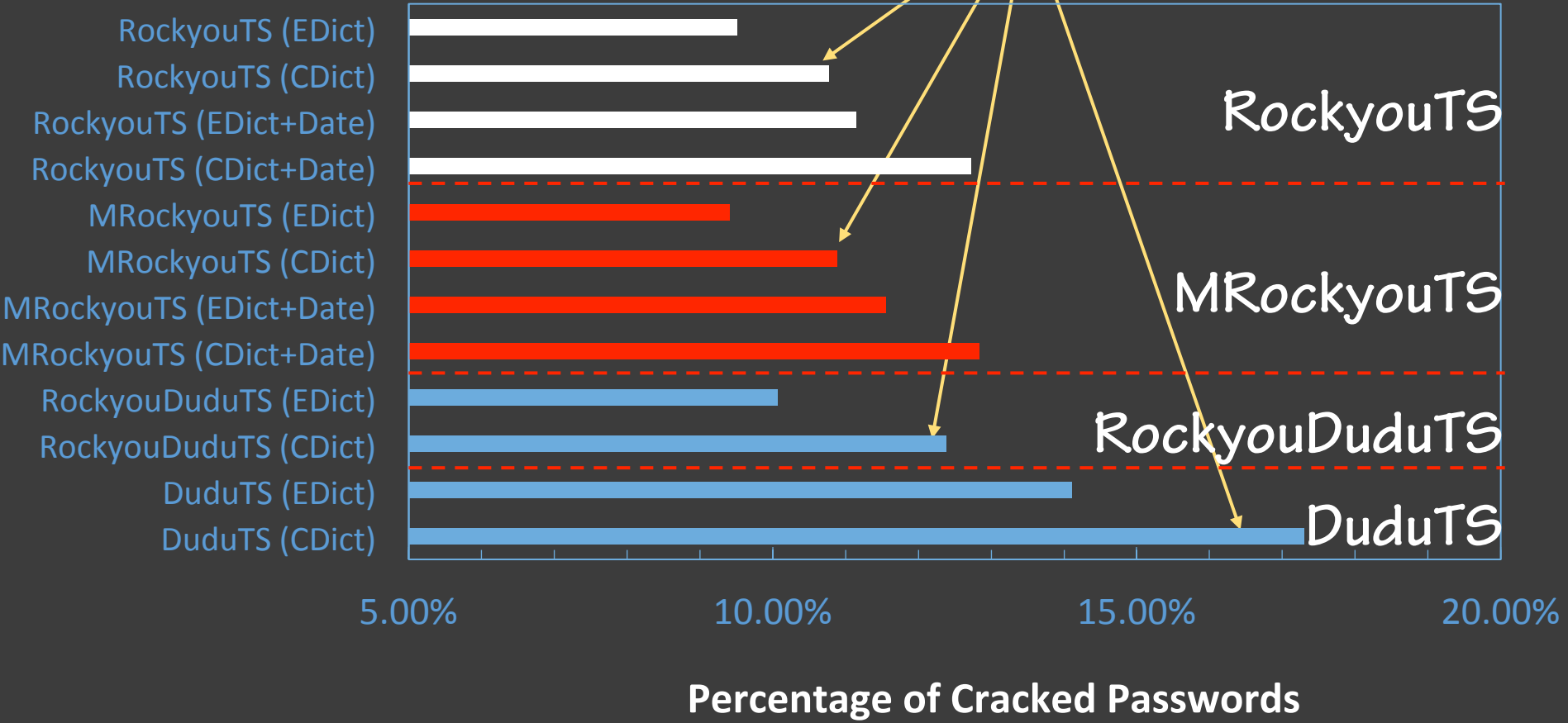
*Dic-0294* can be downloaded from <http://www.outpost9.com/files/WordLists.html>

*English-lower* can be downloaded from <http://download.openwall.net/pub/passwords/wordlists/>

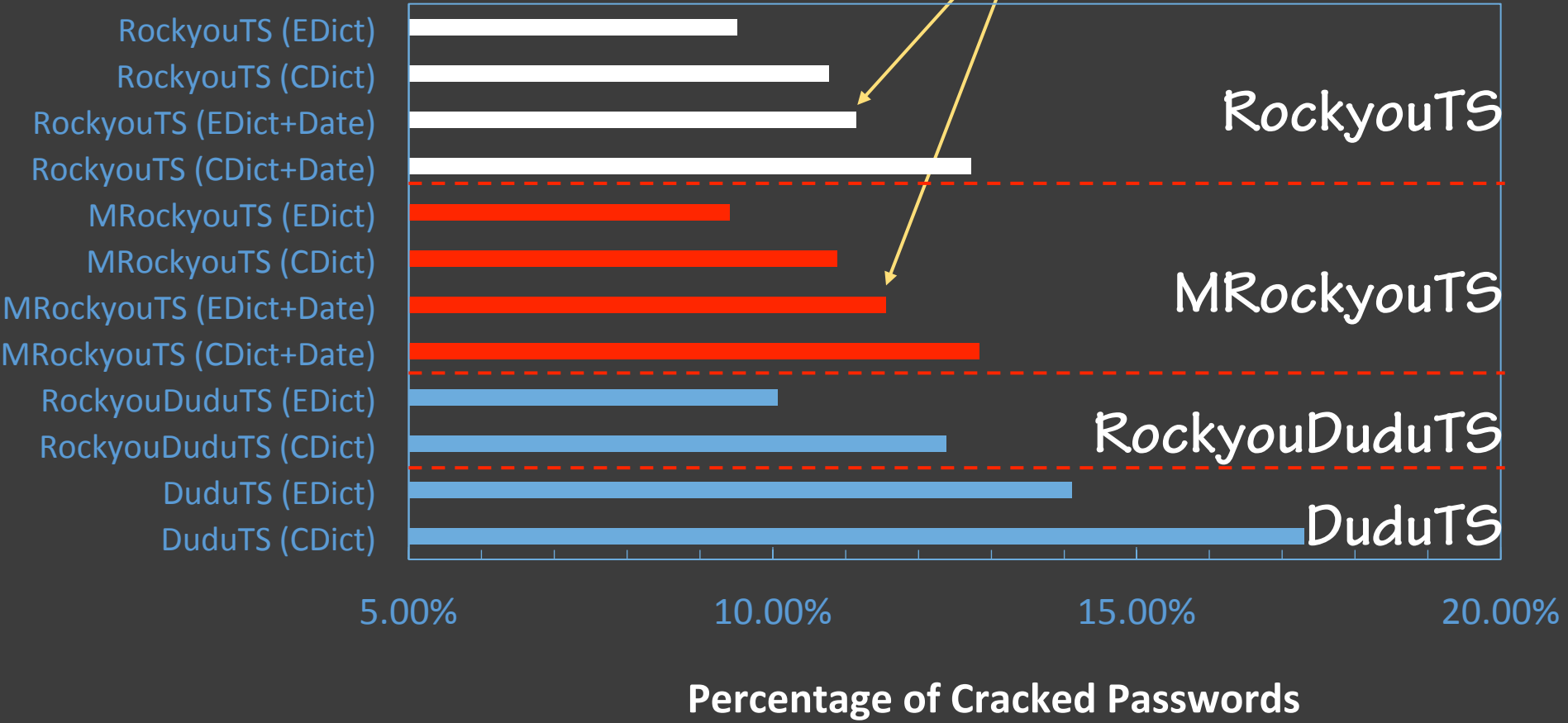
# Guessing CSDN -- Result



# Pinyin

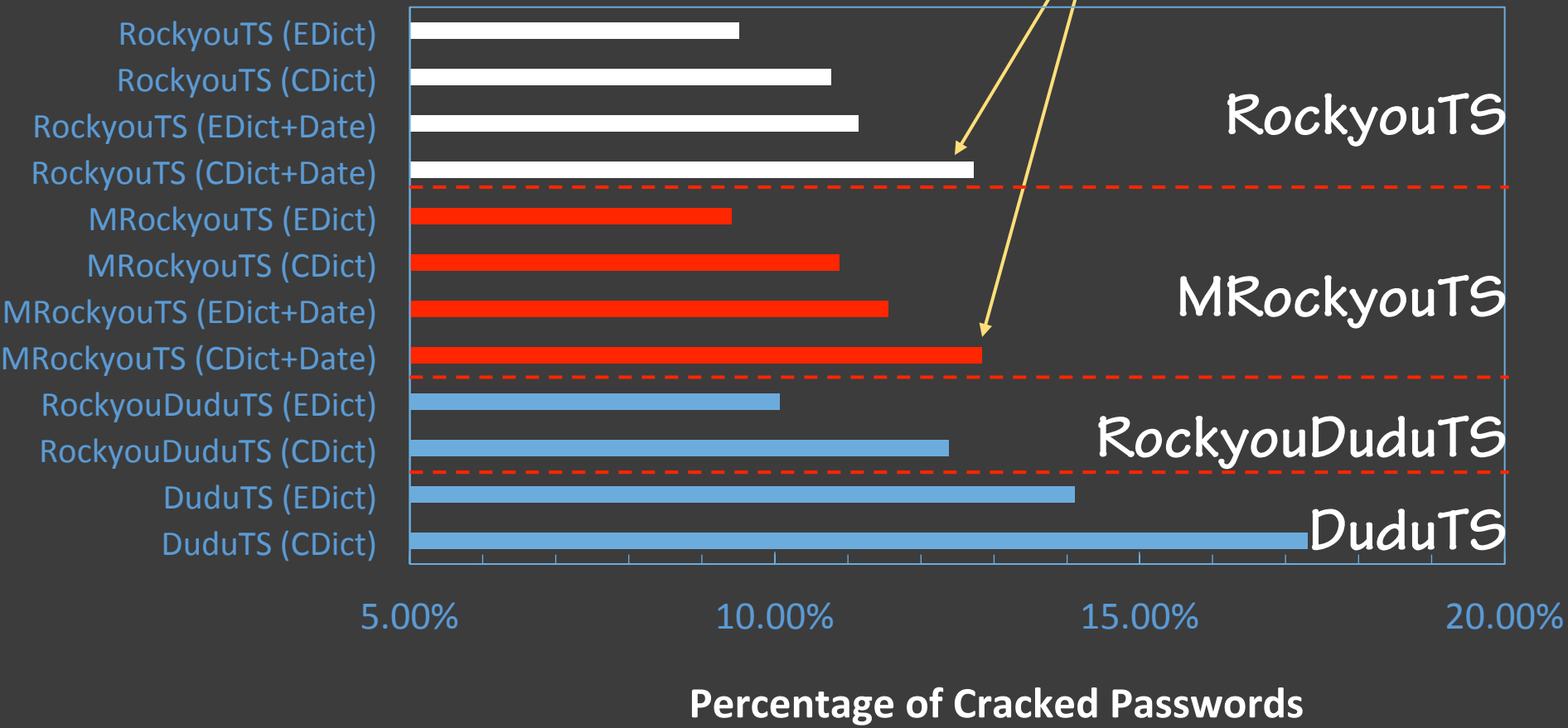


Dates





# Pinyin+Dates



# Guessing CSDN -- Improvement

We increase the guessing  
efficiency by **34%**

# Conclusion

- With similar strength, Chinese passwords contain more **DIGIT-ONLY (>50%)** passwords.
- Just like English words in English passwords, Chinese **PINIYINs (10%-15%)** appear in Chinese passwords.
- Using Pinyin/Dates, we can improve the **EFFICIENCY (34%)** of guessing Chinese passwords.

# Q&A

