

THE SCIENCE OF GUESSING

analyzing an anonymized corpus of 70 million passwords

Joseph Bonneau

jcb82@c1.cam.ac.uk



UNIVERSITY OF
CAMBRIDGE

Computer Laboratory

IEEE SYMPOSIUM ON SECURITY & PRIVACY
≈ OAKLAND, CA, USA
MAY 23, 2012

Why do password research in 2012?



Compatible Time-Sharing System, MIT 1961

Research goal

Precisely compute the guessing difficulty of a given population's password distribution

Research goal

Compare the **guessing difficulty** of password distributions chosen by different populations

Research goal

Compare the **guessing difficulty** of password distributions chosen by different populations



VS.



Research goal

Compare the **guessing difficulty** of password distributions chosen by different populations

The screenshot shows the homepage of the omg! Yahoo! website. At the top, there's a search bar and a navigation menu with links for HOME, CELEBS, PHOTOS, VIDEOS, and NEWS. Below the menu, a banner reads "Discover Yahoo! With Your Friends" with the subtext "Explore news, videos, and much more based on what your friends are reading and watching. Publish your own activity and relate full control." There are four main news cards: 1) "CELEB BIRTHDAYS: MAY 16" featuring Janet Jackson; 2) "LINDSAY MOVES INTO \$25,000-A-MONTH BEVERLY HILLS HOME" featuring Lindsay Lohan; 3) "USHER WANTS EX-WIFE OUT OF HIS GEORGIA MANSION" featuring Usher; 4) "MIRANDA COSGROVE CELEBRATES HER 19TH BIRTHDAY ON THE 'ICARLY' SET" featuring Miranda Cosgrove.

VS.

The screenshot shows the Yahoo! Shopping website for Ireland. The top navigation includes links for Home & Garden, Clothing & Accessories, Electronics, and Sports & Outdoors. A search bar is at the top right. The main content area is titled "Shopping Results" and shows a grid of television listings. Each listing includes a small image of the TV, its price range, and a "Compare Prices" button. The results are filtered by screen size, brand, price, and TV type. Some items have a "See More Like This" link. The prices range from £1,899 to £1,194.

Price Range	Items
£1,899 to £1,194	1 item
£89	1 item
£538 to £493	1 item
£220 to £200	1 item
£194 to £190	1 item
£160 to £150	1 item
£120 to £110	1 item
£80 to £70	1 item
£50 to £40	1 item
£30 to £20	1 item
£10 to £9	1 item
£5 to £4	1 item
£1 to £0	1 item

Research goal

Compare the **guessing difficulty** of password distributions chosen by different populations

Password

Retype Password

VS.

Password



Capitalization matters. Use 6 to 32 characters, and don't use your name or Yahoo! ID.

Re-type Password



For a more secure password:

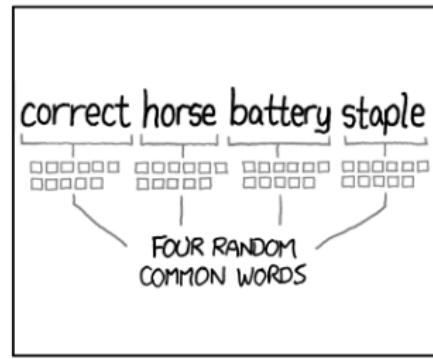
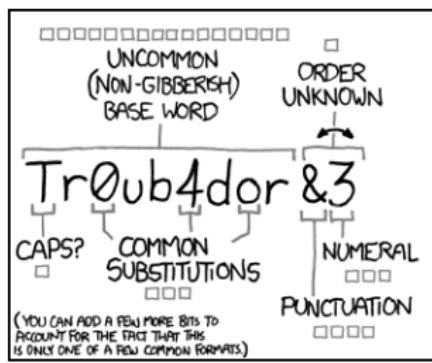
- Use both letters and numbers
- Add special characters (such as @, ?, %)
- Mix capital and lowercase letters

Research goal

Compare the **guessing difficulty** of password distributions chosen by different populations



VS.



Approach #1: Semantic password evaluation

- How long are the passwords?
- Do they look like English words?
- What kind of characters do they contain?

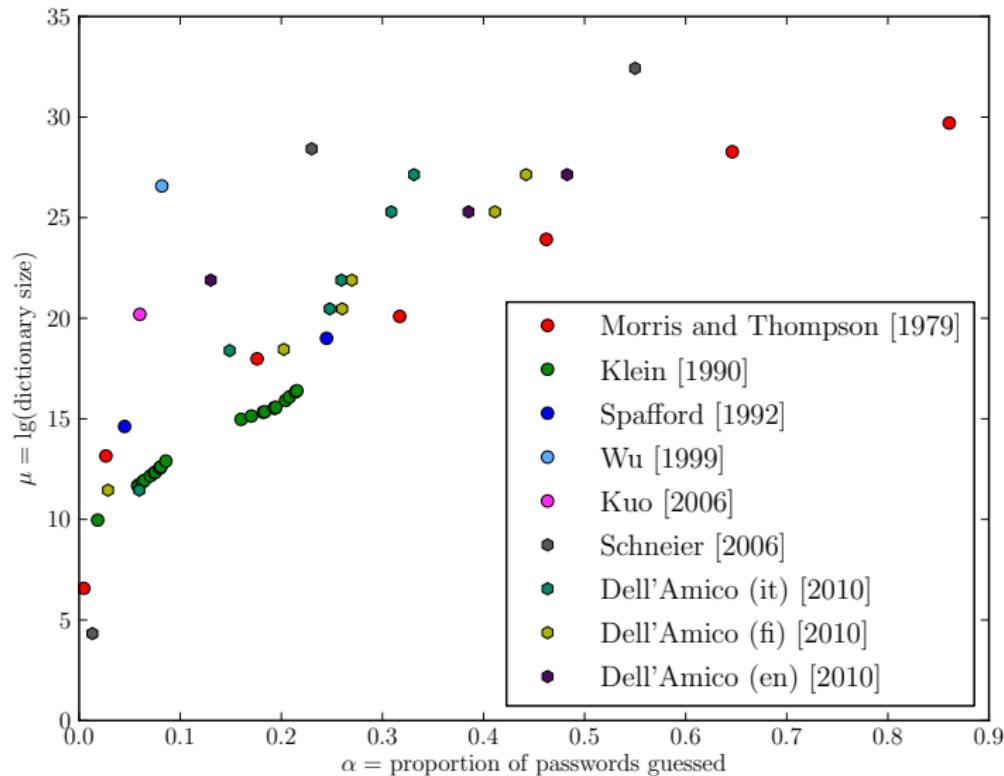
Approach #1: Semantic password evaluation

Length Char.	94 Character Alphabet			10 char. alphabet		94 char alphabet
	No Checks	Dictionary Rule	Dict. & Comp. Rule			
1	4	-	-	3	3.3	6.6
2	6	-	-	5	6.7	13.2
3	8	-	-	7	10.0	19.8
4	10	14	16	9	13.3	26.3
5	12	17	20	10	16.7	32.9
6	14	20	23	11	20.0	39.5
7	16	22	27	12	23.3	46.1
8	18	24	30	13	26.6	52.7
10	21	26	32	15	33.3	65.9
12	24	28	34	17	40.0	79.0
14	27	30	36	19	46.6	92.2
16	30	32	38	21	53.3	105.4
18	33	34	40	23	59.9	118.5
20	36	36	42	25	66.6	131.7
22	38	38	44	27	73.3	144.7
24	40	40	46	29	79.9	158.0
30	46	46	52	35	99.9	197.2
40	56	56	62	45	133.2	263.4

NIST “entropy” formula

Approach #2: Cracking experiments

Approach #2: Cracking experiments



Methodological problems with password analysis

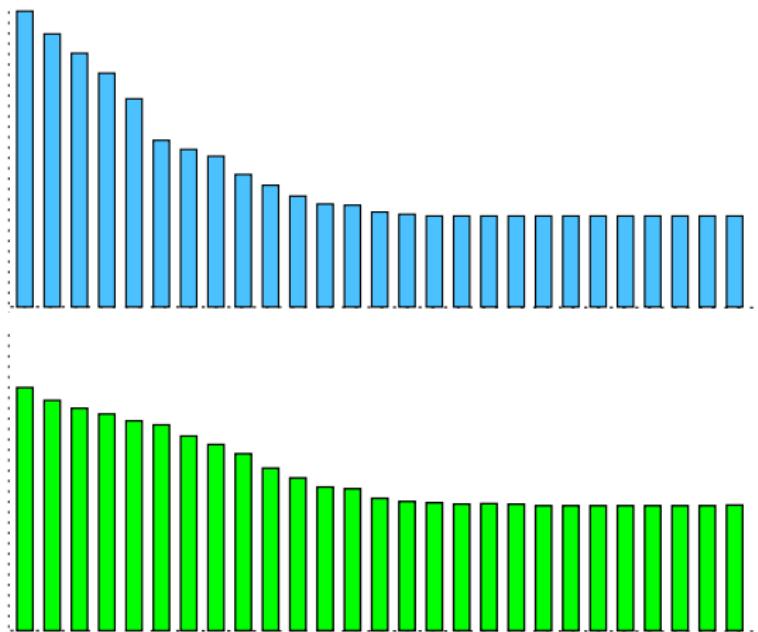
	semantic	cracking
external validity		✓
no operator bias	✓	
no demographic bias	?	
repeatable	✓	?
easy	✓	?

My approach



- ➊ Collect password data on a huge scale
- ➋ Compare populations as probability distributions
- ➌ Test hypotheses using different populations

My approach



- ➊ Collect password data on a huge scale
- ➋ Compare populations as probability distributions
- ➌ Test hypotheses using different populations

My approach

STAND BACK



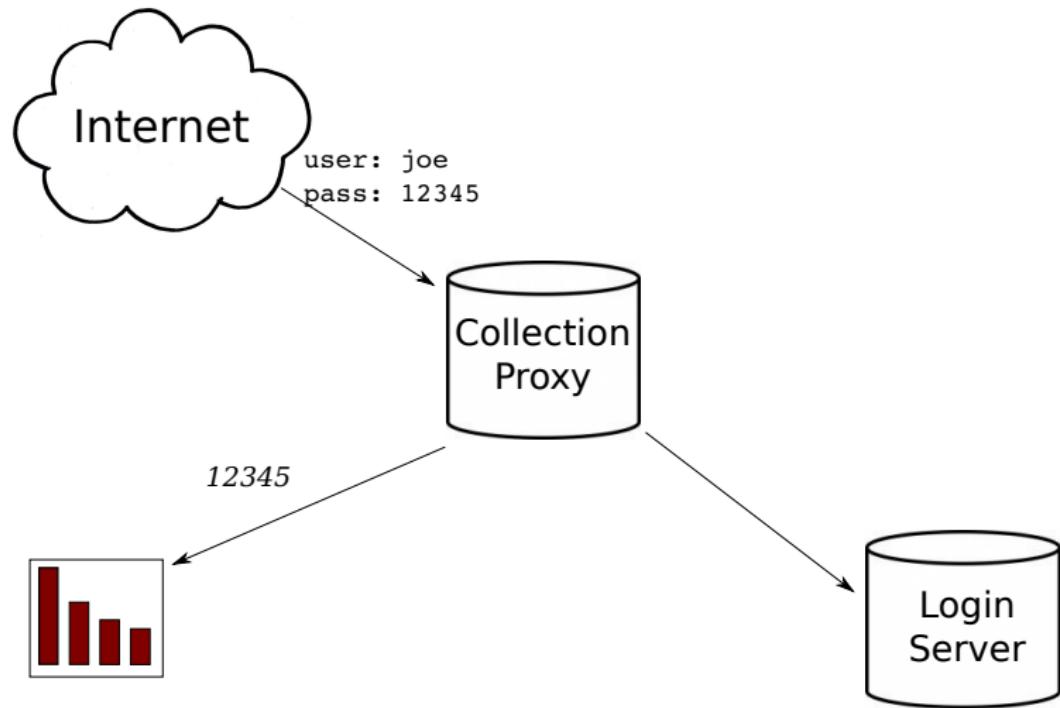
**I'M GOING TO TRY
SCIENCE**

- ➊ Collect password data on a huge scale
- ➋ Compare populations as probability distributions
- ➌ Test hypotheses using different populations

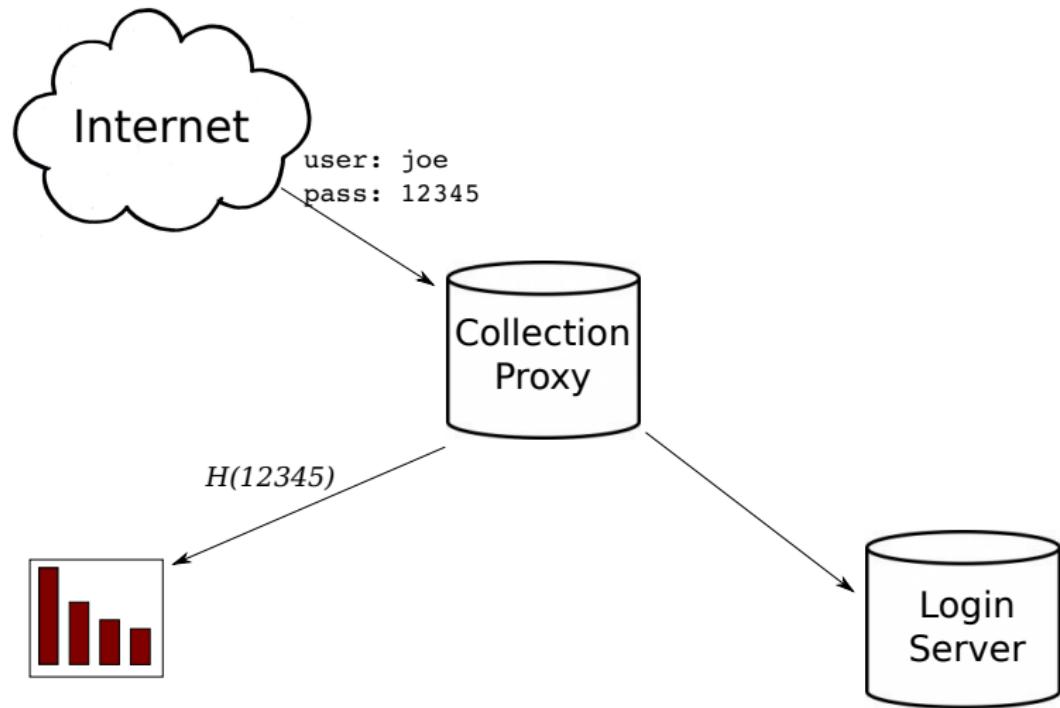
Goal #1: collect a massive data set

- with cooperation from Yahoo!
- privacy-preserving collection ☺
 - histograms only
- demographic splits collected

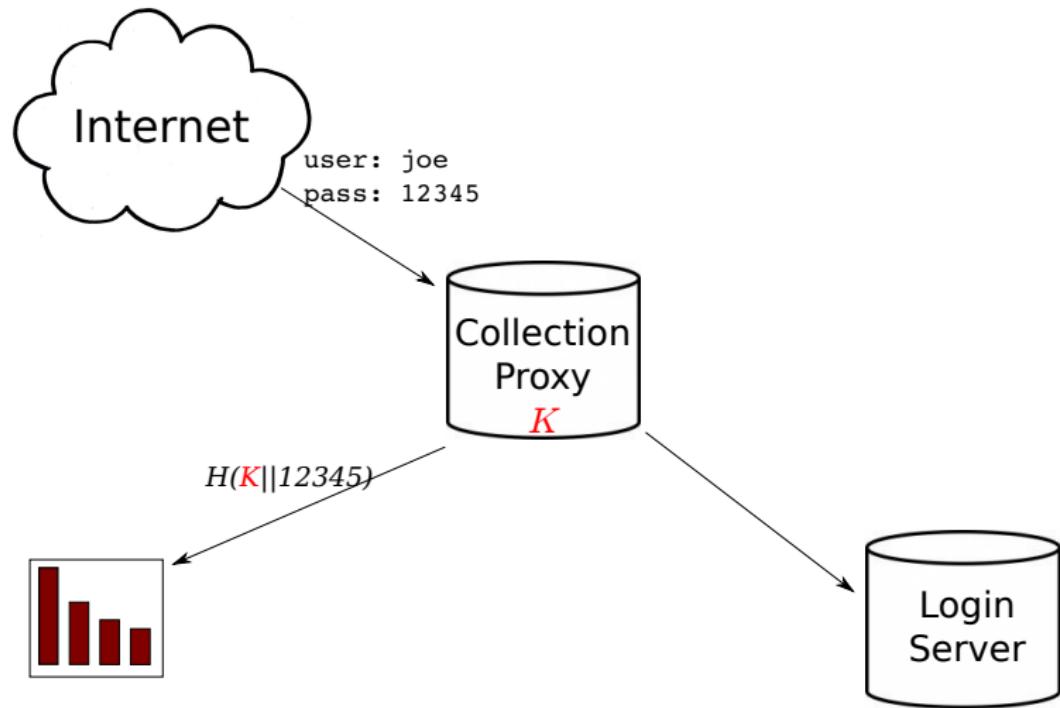
Collecting large-scale data at Yahoo!



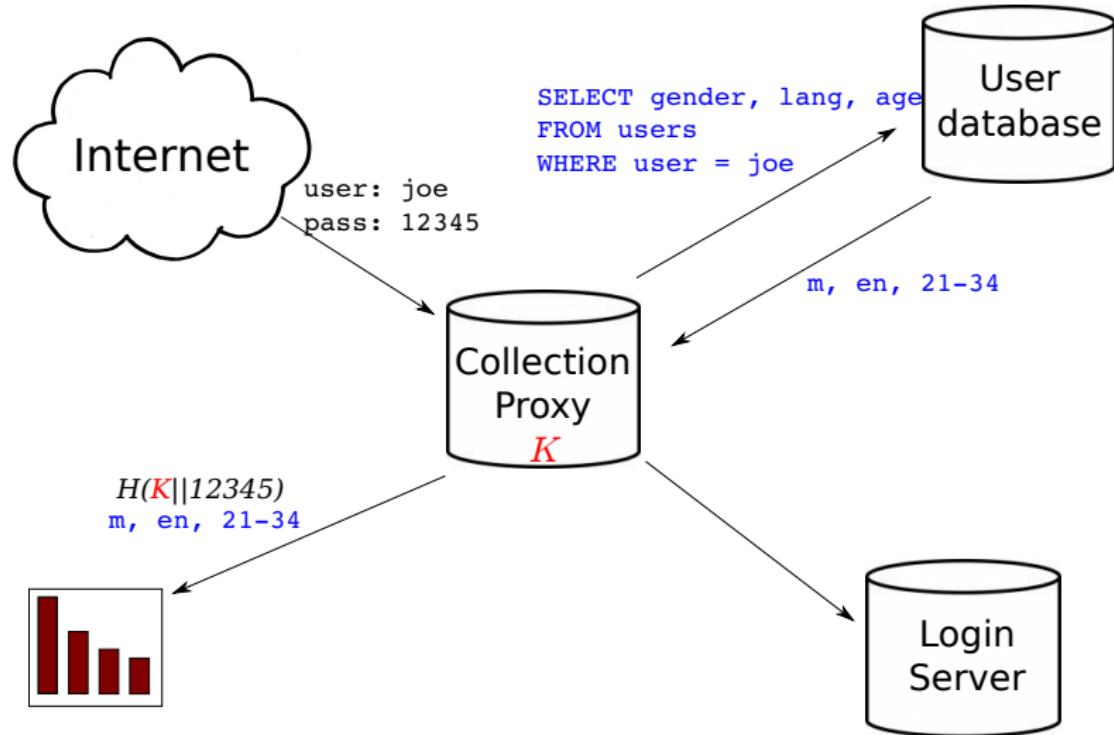
Collecting large-scale data at Yahoo!



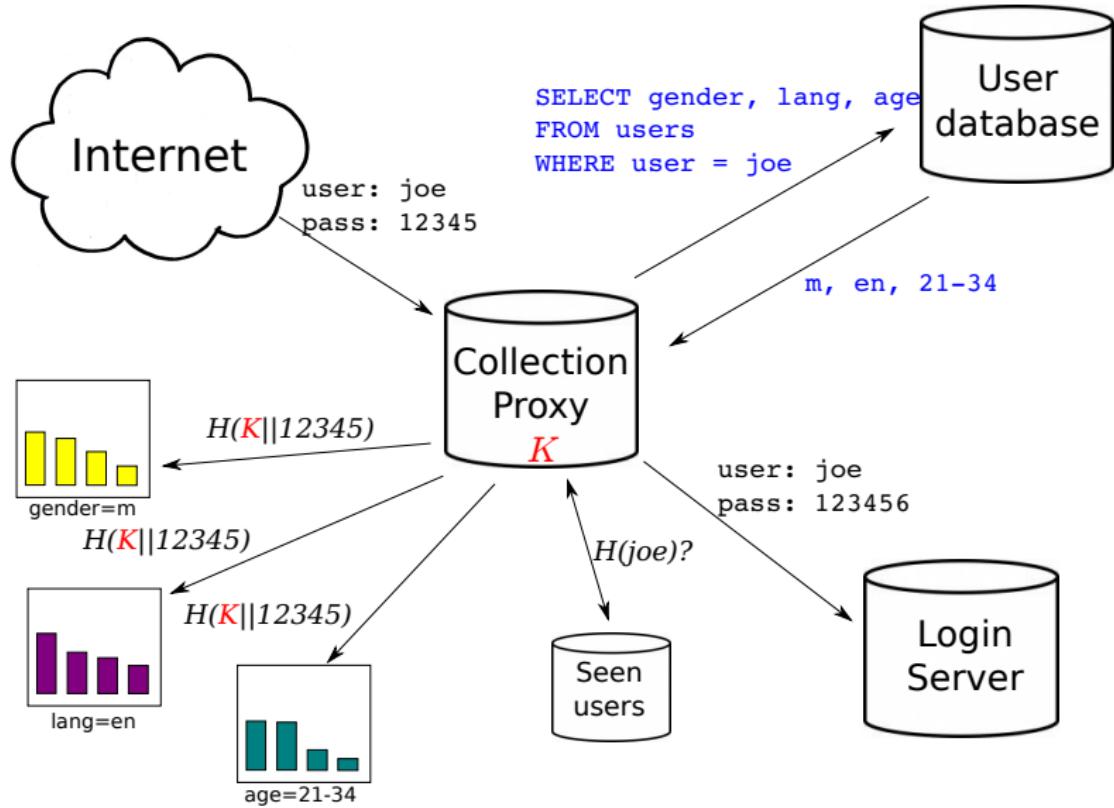
Collecting large-scale data at Yahoo!



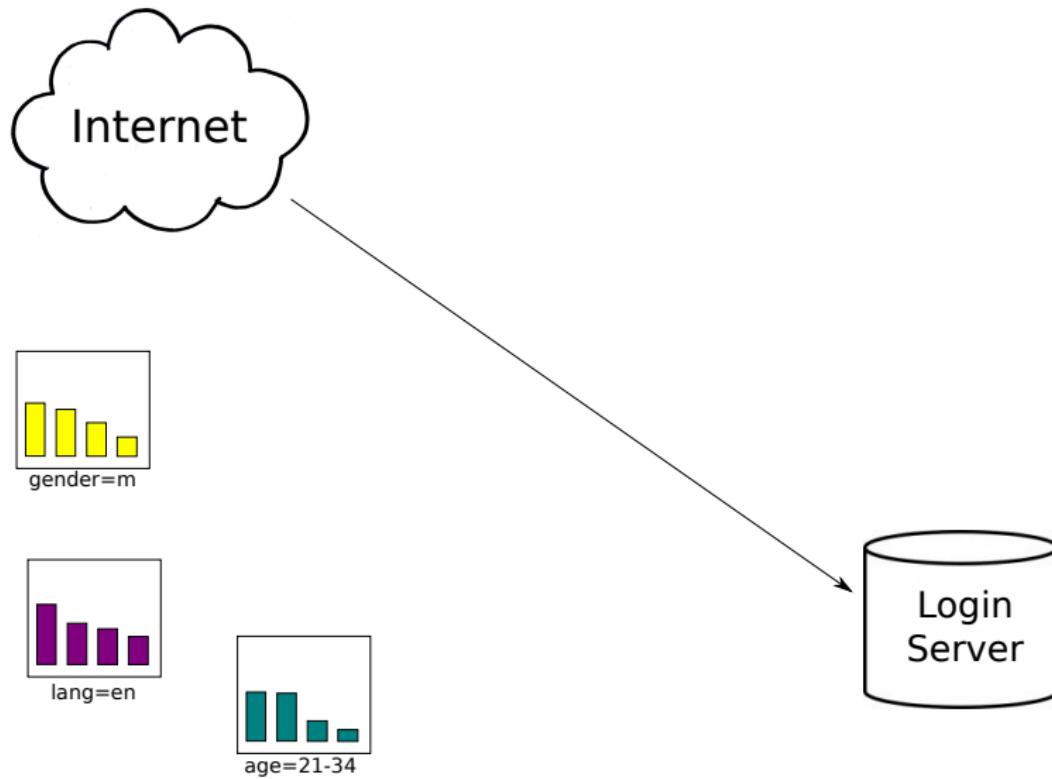
Collecting large-scale data at Yahoo!



Collecting large-scale data at Yahoo!



Collecting large-scale data at Yahoo!



Collecting large-scale data at Yahoo!

- Experiment run May 23–25, 2011
- 69,301,337 unique users
- 42.5% unique
- 328 different predicate functions

Goal #2: model guessing as a probability problem

- Assume perfect knowledge of the distribution \mathcal{X}
- \mathcal{X} has N events (passwords) x_1, x_2, \dots
- Events have probability $p_1 \geq p_2 \geq \dots \geq p_N \geq 0$
- Each user chooses at random $X \xleftarrow{R} \mathcal{X}$

Question: How hard is it to guess X ?

Shannon entropy

$$H_1(\mathcal{X}) = - \sum_{i=1}^N p_i \lg p_i$$

Interpretation: Expected number of queries “Is $X \in \mathcal{S}$?” for arbitrary subsets $\mathcal{S} \subseteq \mathcal{X}$ needed to guess X . ([Source-Coding Theorem](#))

Guesswork (guessing entropy)

$$G_1(\mathcal{X}) = E [\#\text{guesses}] = \sum_{i=1}^N p_i \cdot i$$

Intepretation: Expected number of queries “Is $X = x_i?$ ” for $i = 1, 2, \dots, N$ (optimal sequential guessing)

G_1 fails badly for real password distributions

Random 128-bit passwords in the wild at RockYou ($\sim 2^{-20}$)

```
ed65e09b98bcd70576d6c5f5e2ee38a9  
e54d409c55499851aeb25713c1358484  
dee489981220f2646eb8b3f412c456d9  
c4df8d8e225232227c84d0ed8439428a  
bd9059497b4af2bb913a8522747af2de  
b25d6118ffc44b12b014feb81ea68e49  
aac71eb7307f4c54b12c92d9bd45575f  
9475d62e1f8b13676deab3824492367a  
92965710534a9ec4b30f27b1e7f6062a  
80f5a0267920942a73693596fe181fb7  
76882fb85a1a8c6a83486aba03c031c9  
6a60e0e51a3eb2e9fed6a546705de1bf ...
```

$$\Rightarrow G_1(\text{RockYou}) > 2^{107}$$

Attackers might be happy ignoring the hard values



α -work-factor

$$\mu_\alpha(\mathcal{X}) = \min \left\{ \mu \in [1, N] \middle| \sum_{i=1}^{\mu} p_i \geq \alpha \right\}$$

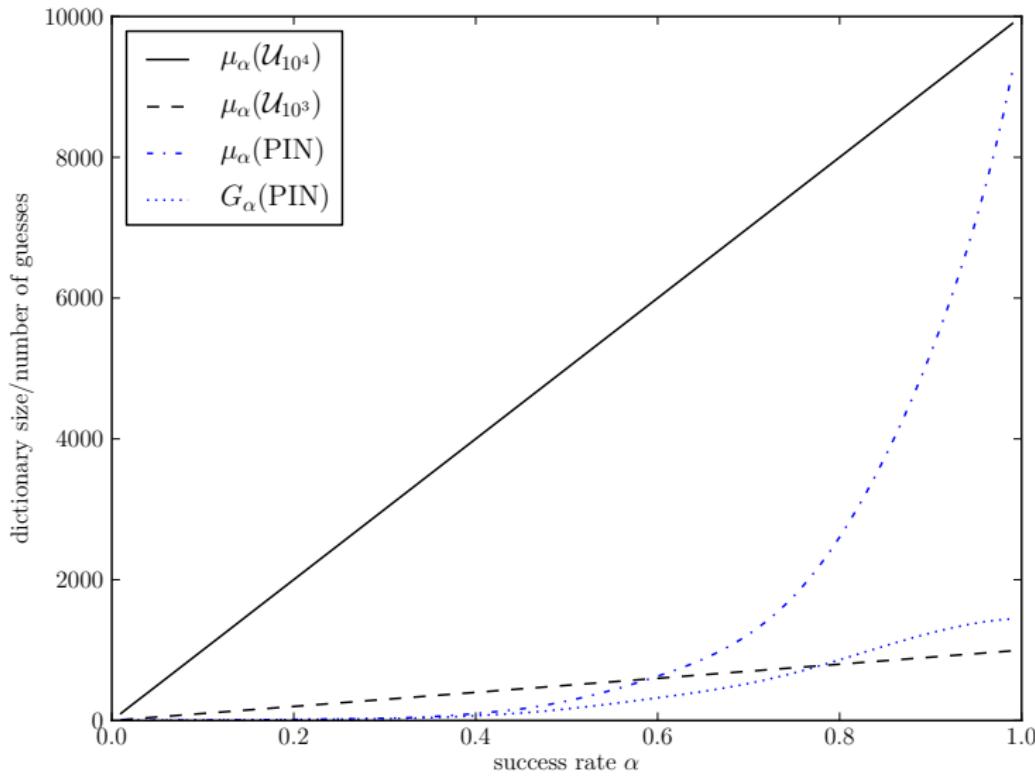
Interpretation: Minimal dictionary size to succeed with probability α

α -guesswork

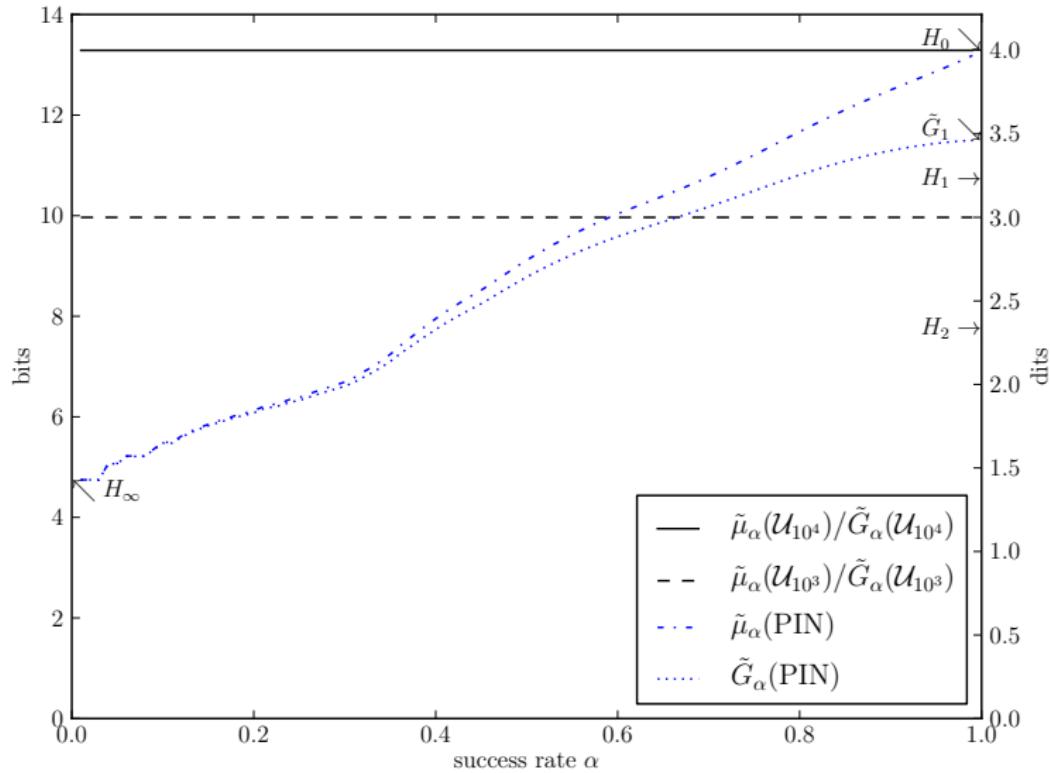
$$G_\alpha(\mathcal{X}) = (1 - \lceil \alpha \rceil) \cdot \mu_\alpha(\mathcal{X}) + \sum_{i=1}^{\mu_\alpha(\mathcal{X})} p_i \cdot i$$

Intepretation: Mean number of guesses to succeed with probability α

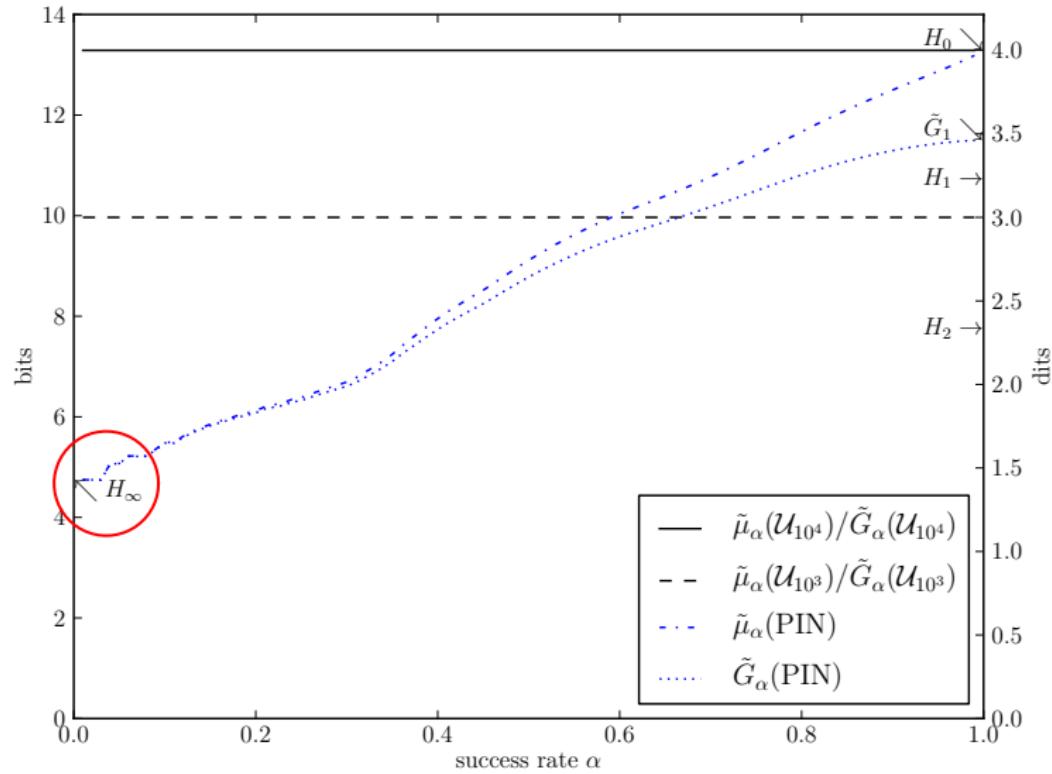
Guessing curves visualise all possible attacks



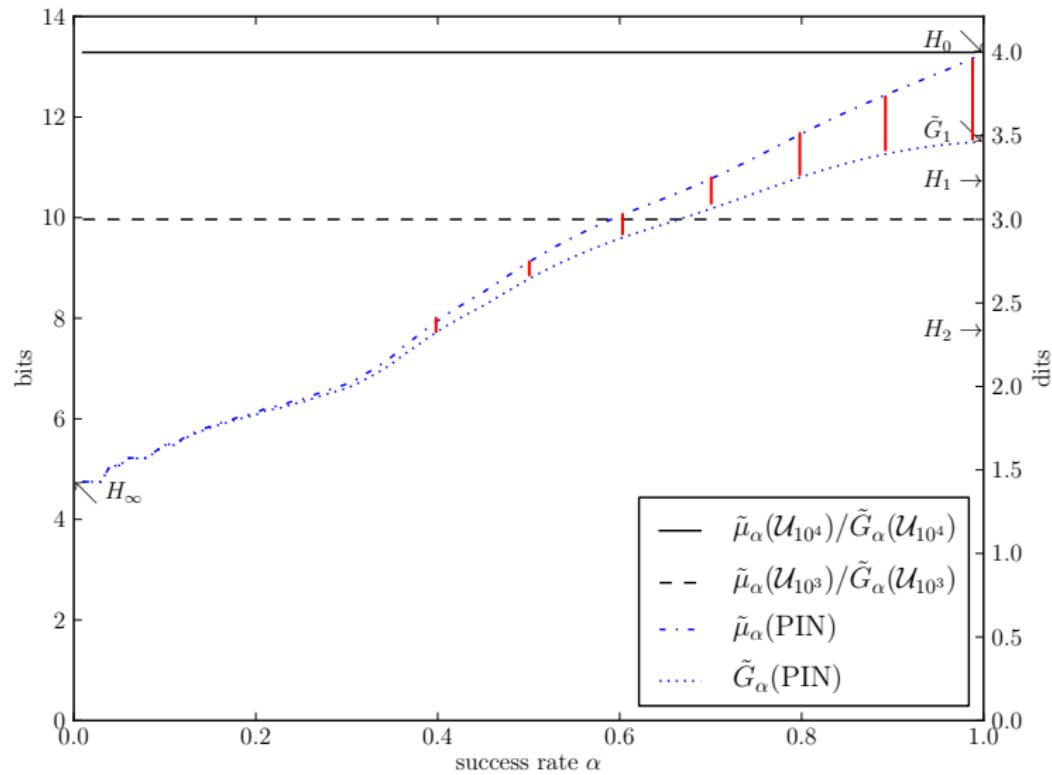
More intuitive after converting to bits



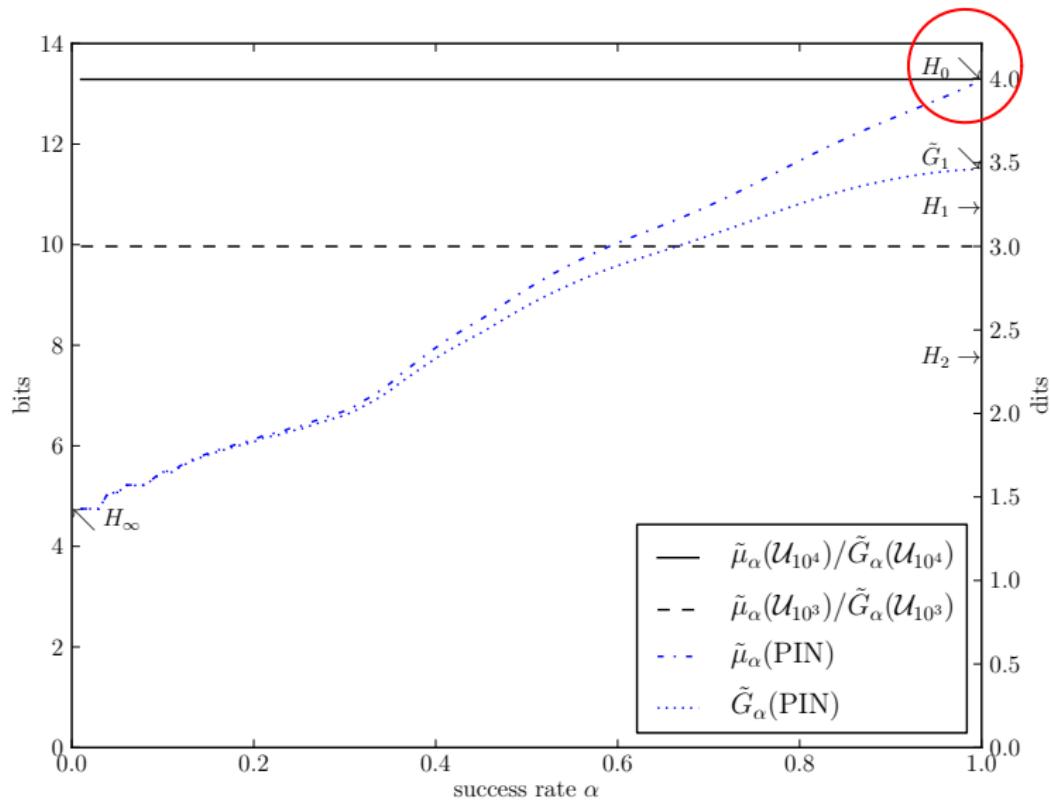
More intuitive after converting to bits



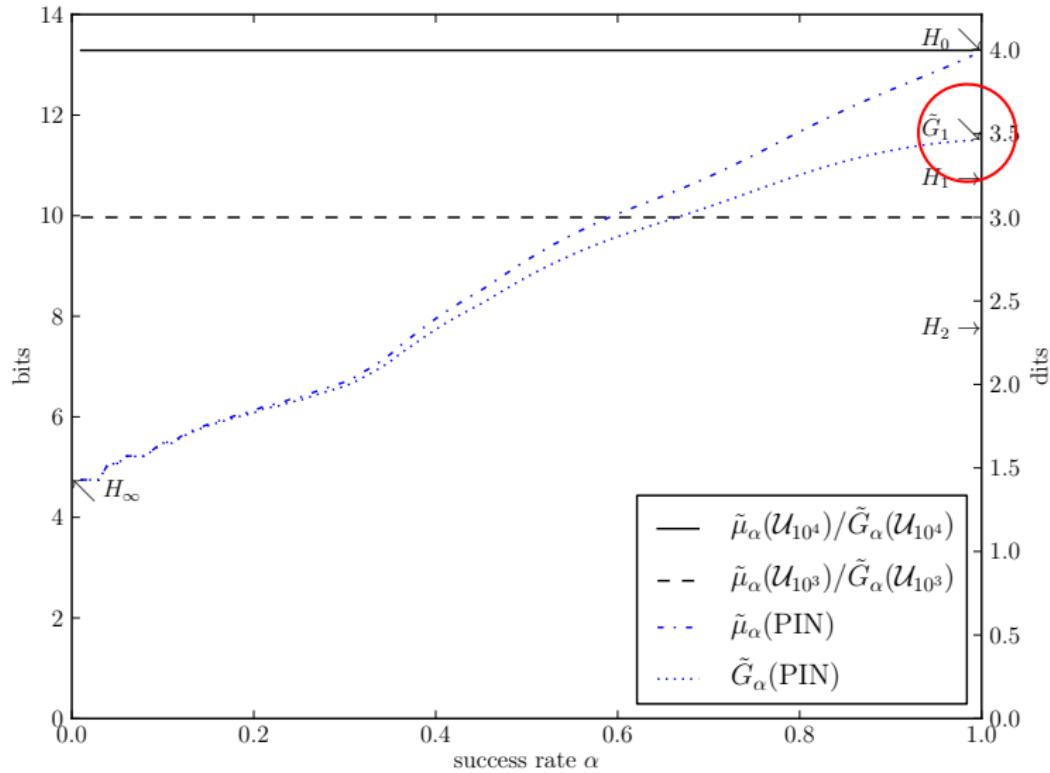
More intuitive after converting to bits



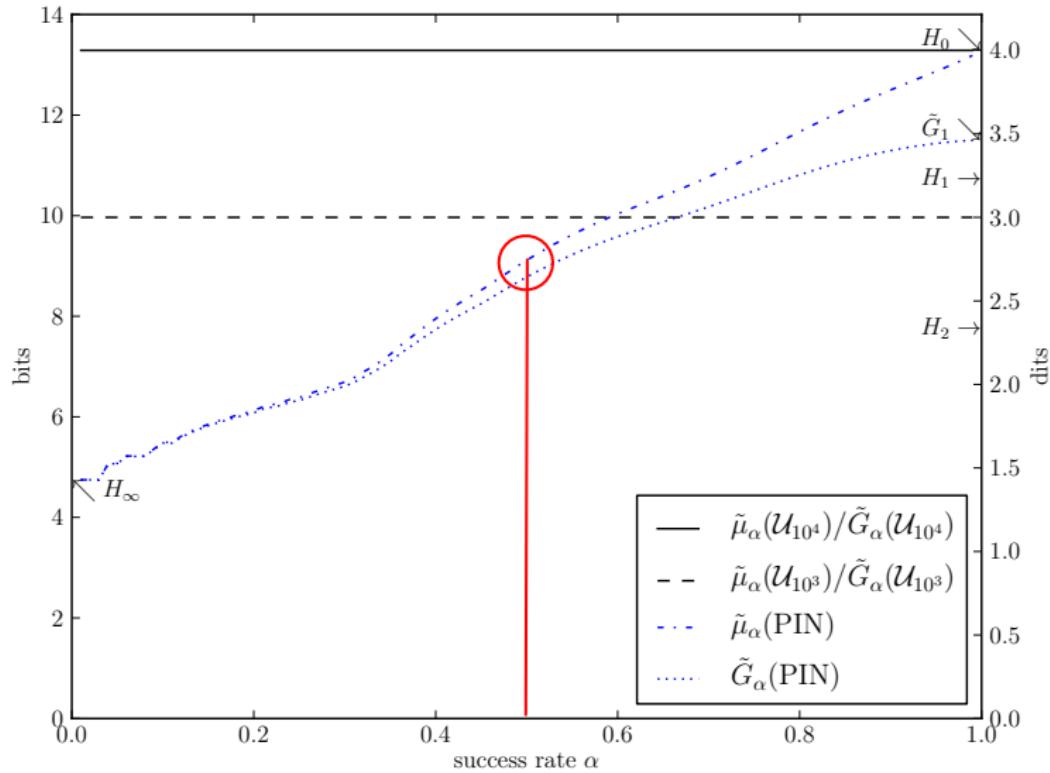
More intuitive after converting to bits



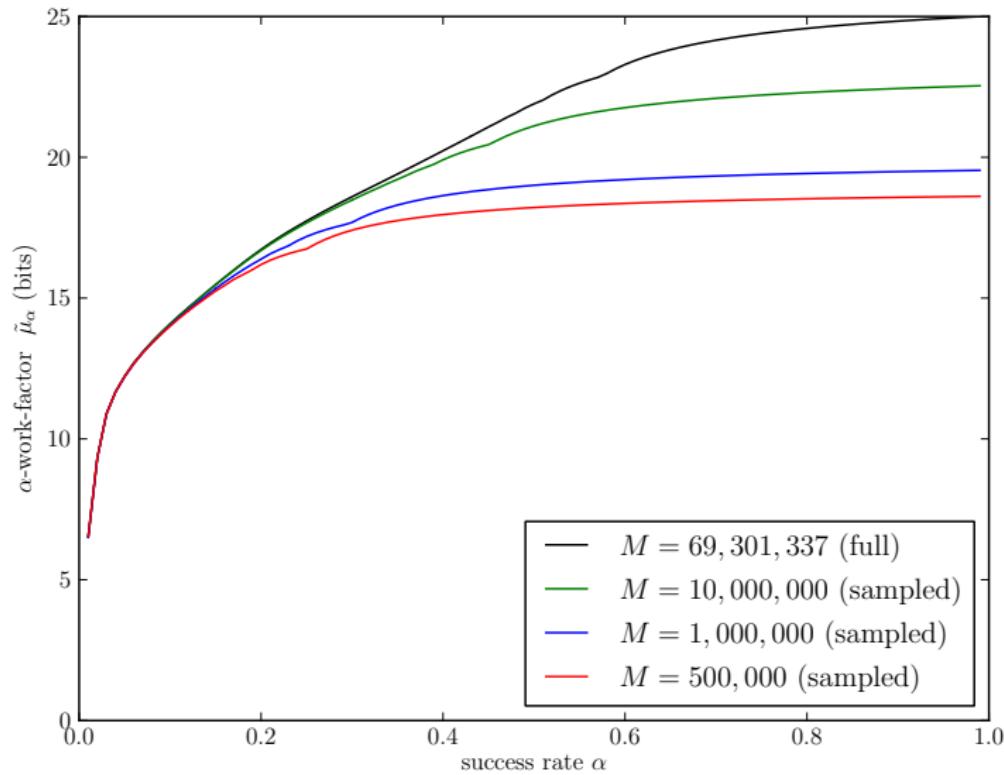
More intuitive after converting to bits



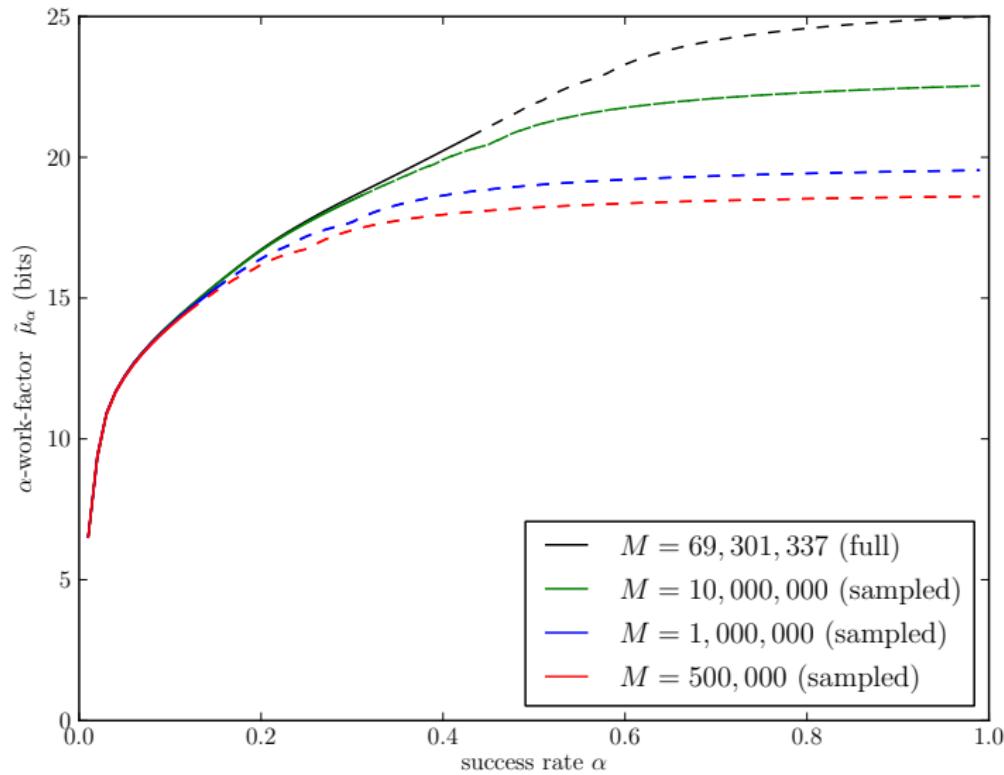
More intuitive after converting to bits



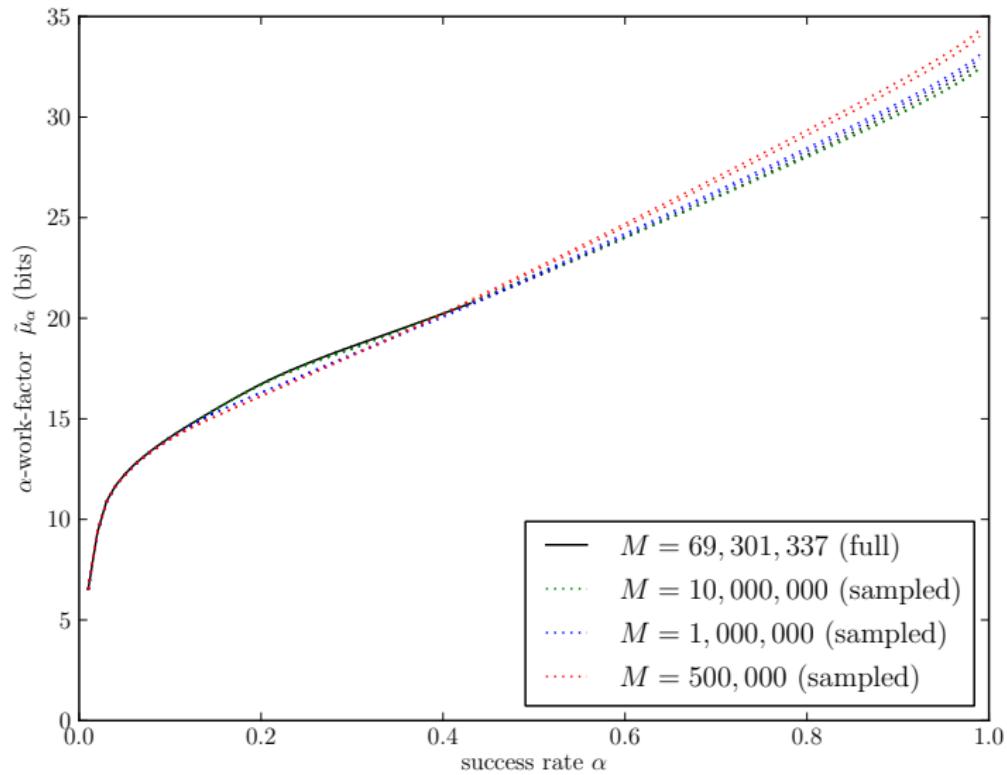
Sample size is a major problem for passwords...



Predict our confidence range by *bootstrapping*

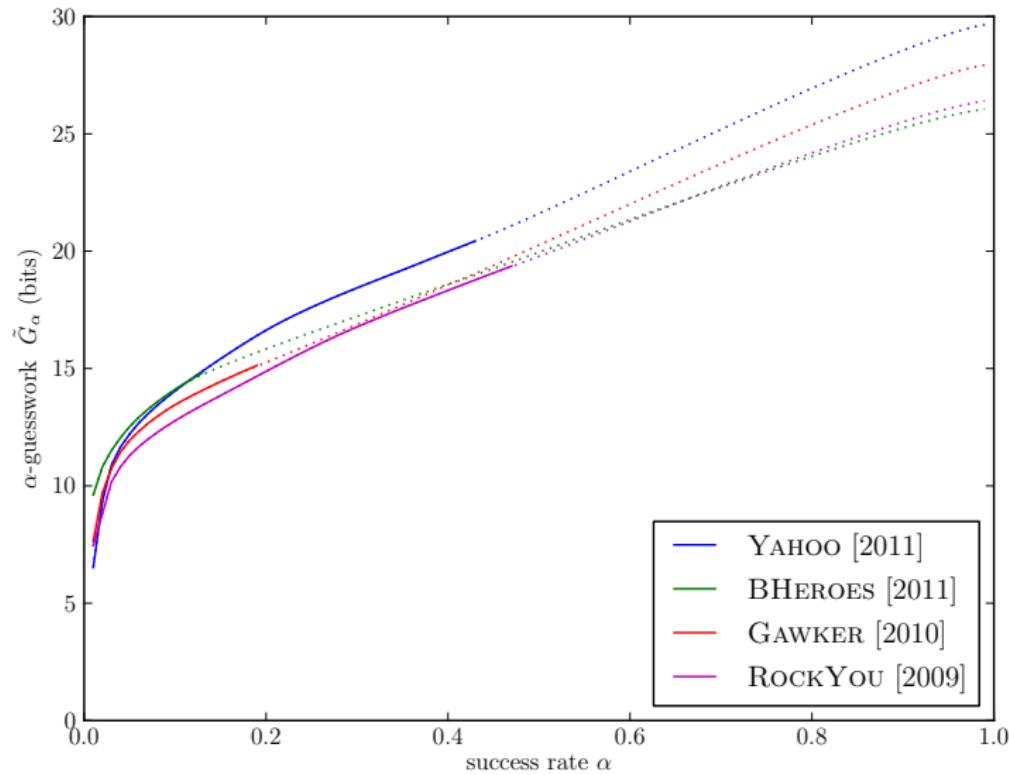


Extrapolation w/ truncated Sichel-Poisson distribution

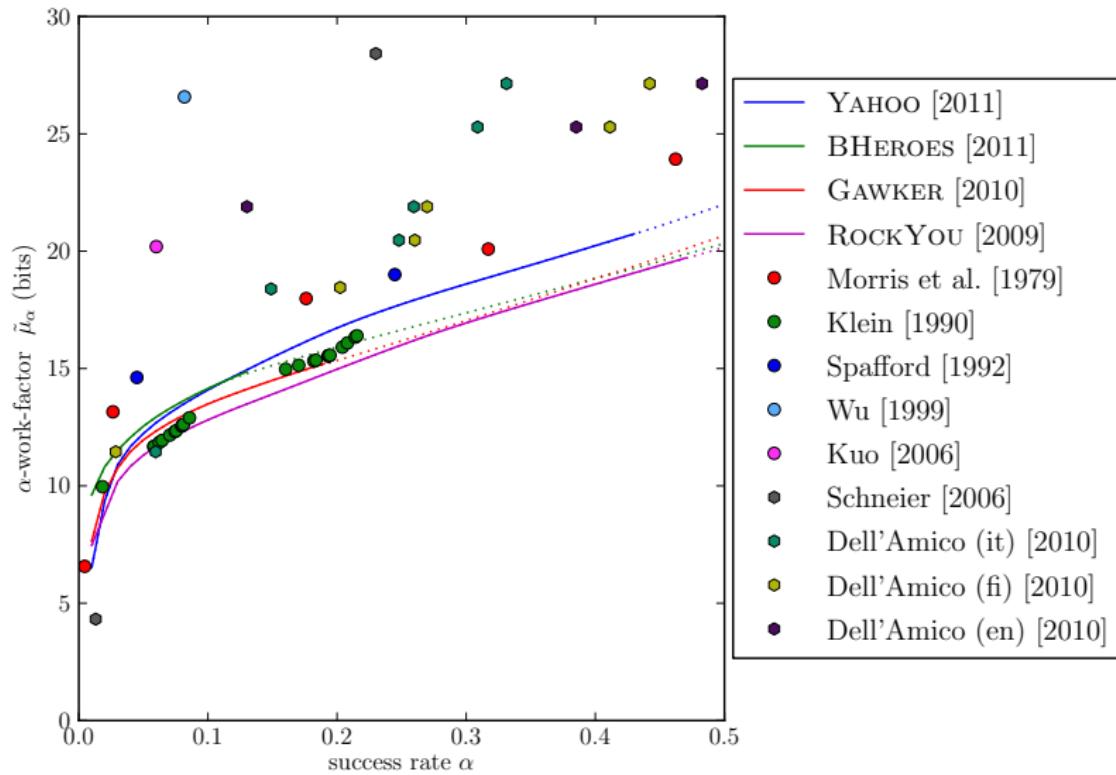


Goal #3: Analyze Yahoo! passwords

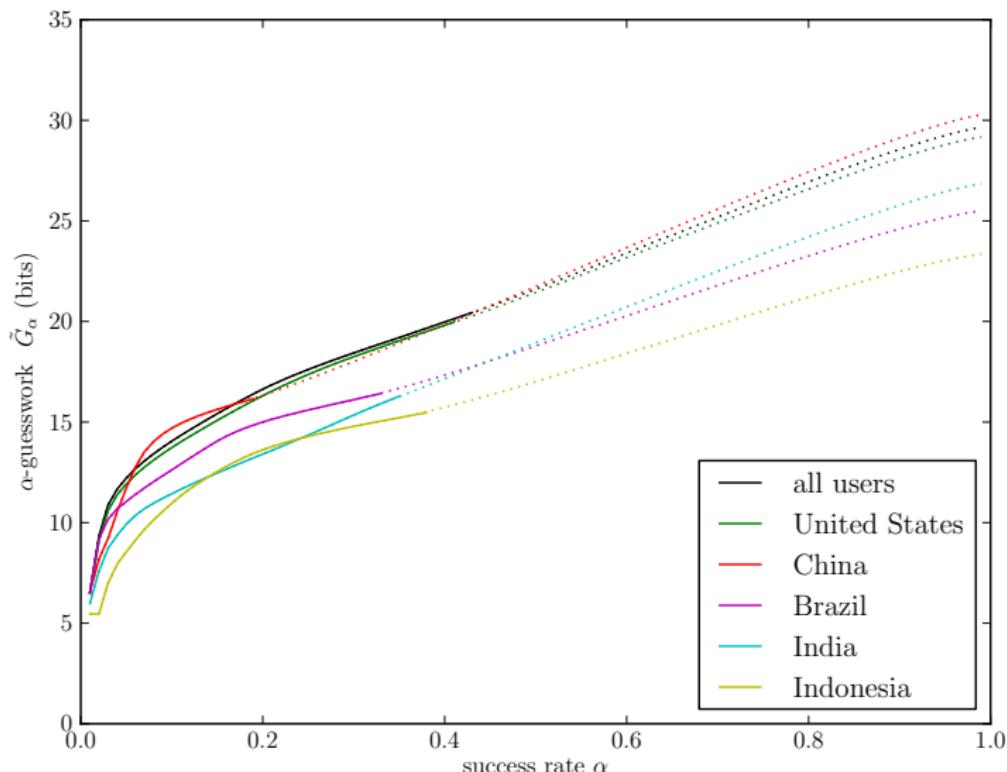
Goal #3: Analyze Yahoo! passwords



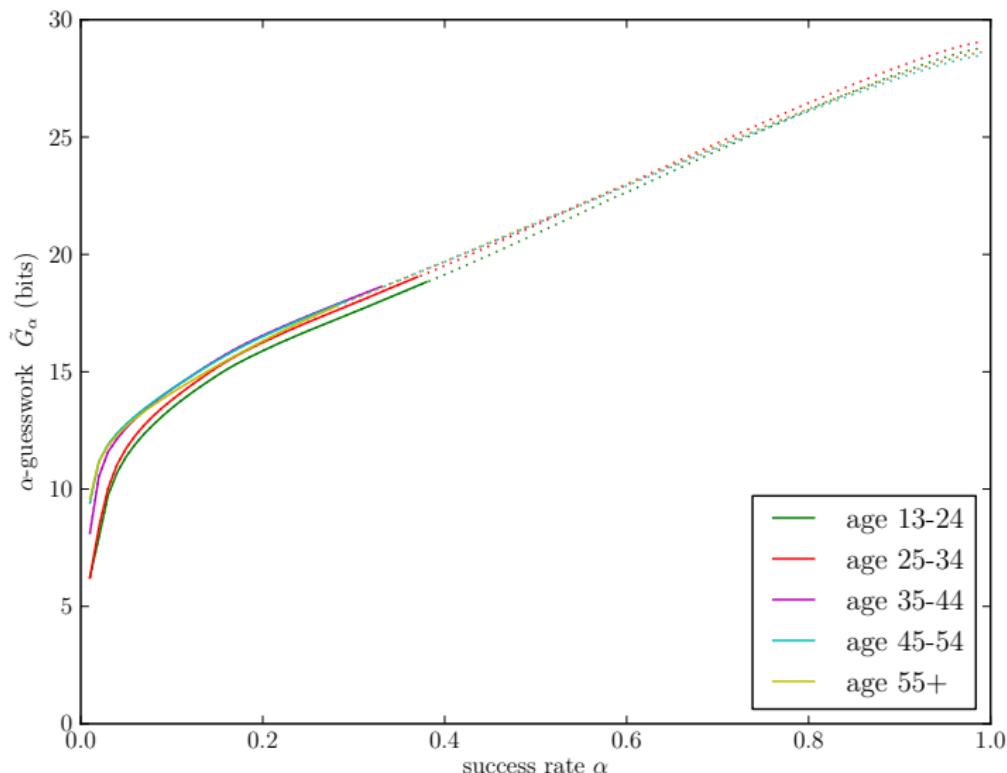
Goal #3: Analyze Yahoo! passwords



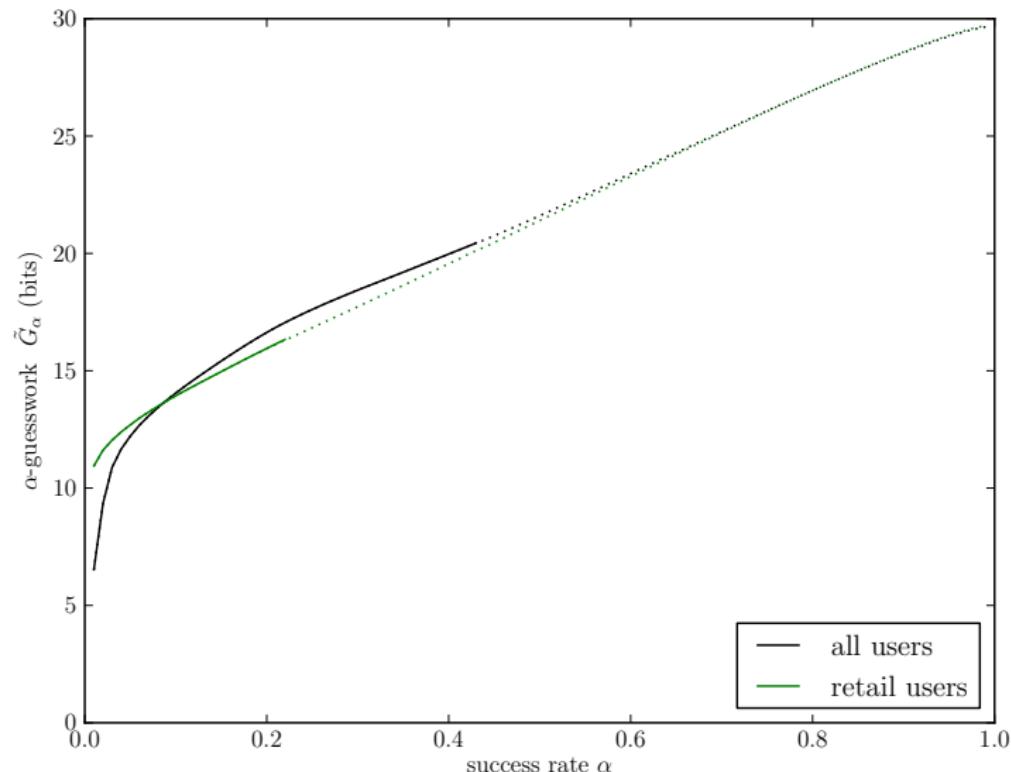
Demographic trends: nationality



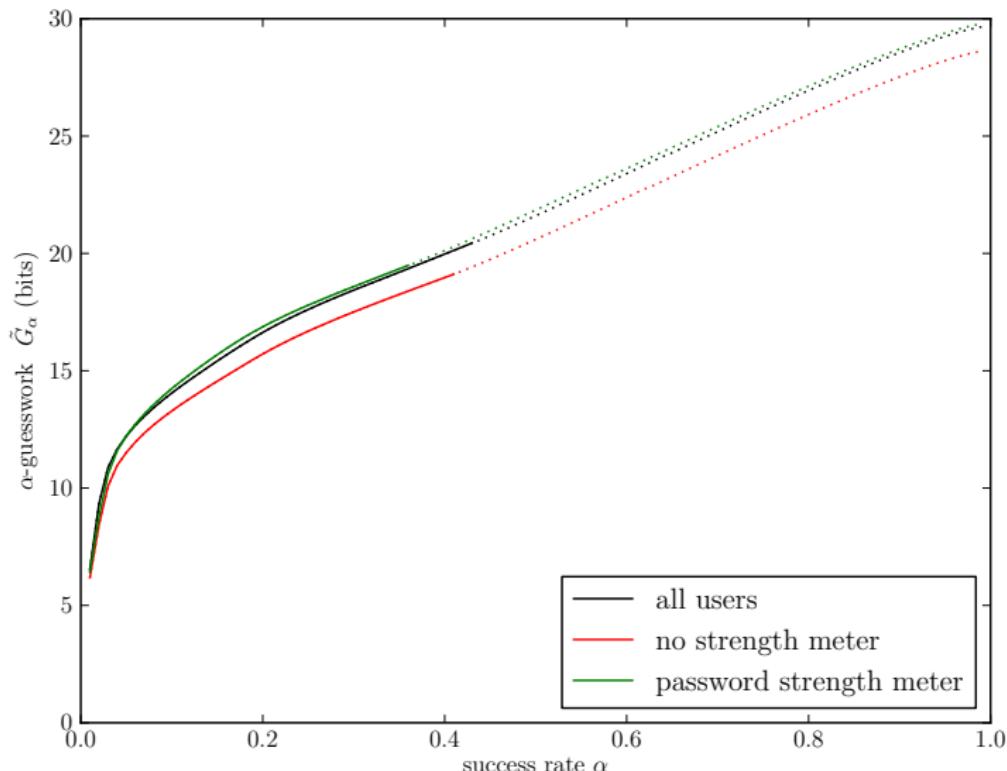
Demographic trends: age



Credit card details make little difference



Password strength meter makes little difference



Demographic summary

- there is no “good group” of users
- differences small but statistically significant
- online attack 6–9 bits ($\tilde{\lambda}_{10}$)
- offline attack 15–25 bits ($\tilde{G}_{0.5}$)

Surprisingly little language variation

	dictionary										
	de	en	es	fr	id	it	ko	pt	zh	vi	global
target	6.5%	3.3%	2.6%	2.9%	2.2%	2.8%	1.6%	2.1%	2.0%	1.6%	3.5%
de	4.6%	8.0%	4.2%	4.3%	4.5%	4.3%	3.4%	3.5%	4.4%	3.5%	7.9%
en	5.0%	5.6%	12.1%	4.6%	4.1%	6.1%	3.1%	6.3%	3.6%	2.9%	6.9%
es	4.0%	4.2%	3.4%	10.0%	2.9%	3.2%	2.2%	3.1%	2.7%	2.1%	5.0%
fr	6.3%	8.7%	6.2%	6.3%	14.9%	6.2%	5.8%	6.0%	6.7%	5.9%	9.3%
id	6.0%	6.3%	6.8%	5.3%	4.6%	14.6%	3.3%	5.7%	4.0%	3.2%	7.2%
it	2.0%	2.6%	1.9%	1.8%	2.3%	2.0%	5.8%	2.4%	3.7%	2.2%	2.8%
ko	3.9%	4.3%	5.8%	3.8%	3.9%	4.4%	3.5%	11.1%	3.9%	2.9%	5.1%
pt	1.9%	2.4%	1.7%	1.7%	2.0%	2.0%	2.9%	1.8%	4.4%	2.0%	2.9%
zh	5.7%	7.7%	5.5%	5.8%	6.3%	5.7%	6.0%	5.8%	7.0%	14.3%	7.8%

With 1000 guesses, greatest efficiency loss is only 4.8 (fr/vi)

Joseph Bonneau and Rubin Xu.

Of contraseñas, סיסמאות and 密码: Character encoding issues for web passwords *Web 2.0 Security & Privacy*, 2012.

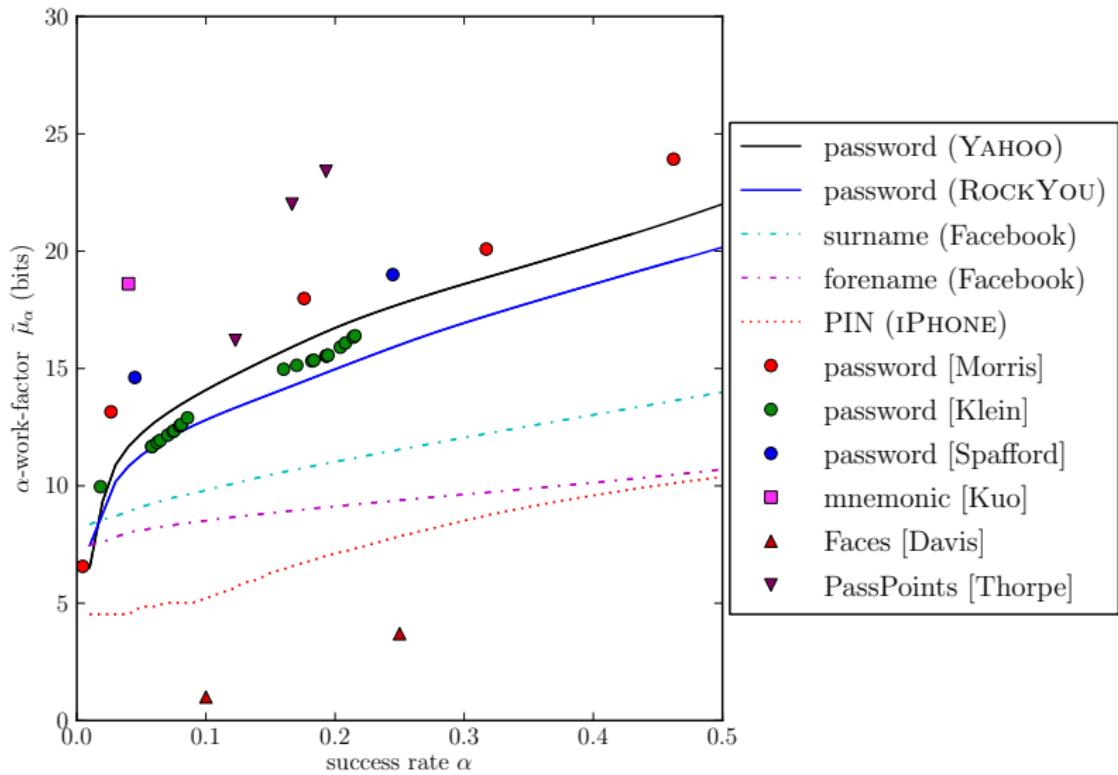
Comparing password analysis methods

	semantic	cracking	statistical
external validity		✓	?
no operator bias	✓		✓
no demographic bias	?		✓
repeatable	✓	?	✓
easy	✓	?	✓

Comparing password analysis methods

	semantic	cracking	statistical
external validity		✓	?
no operator bias	✓		✓
no demographic bias	?		✓
repeatable	✓	?	✓
easy	✓	?	✓
works w/small data	✓	✓	

The picture so far



For more information

my email

jcb82@cl.cam.ac.uk

my dissertation

Guessing human-chosen secrets

Acknowledgements



UNIVERSITY OF
CAMBRIDGE
Computer Laboratory

Elizabeth Zwicky
Henry Watts
Ram Marti
Clarence Chung
Christopher Harris
Ross Anderson
Richard Clayton
Frank Stajano
Markus Kuhn
Saar Drimer
Andrew Lewis
Paul van Oorschot
Cormac Herley
Arvind Narayanan

Converting metrics to bits

- Find the size of a uniform distribution \mathcal{U}_N with equivalent security
- Easy case:

$$\tilde{\mu}_\alpha(\mathcal{X}) = \lg \left(\frac{\mu_\alpha(\mathcal{X})}{\lceil \alpha \rceil} \right)$$

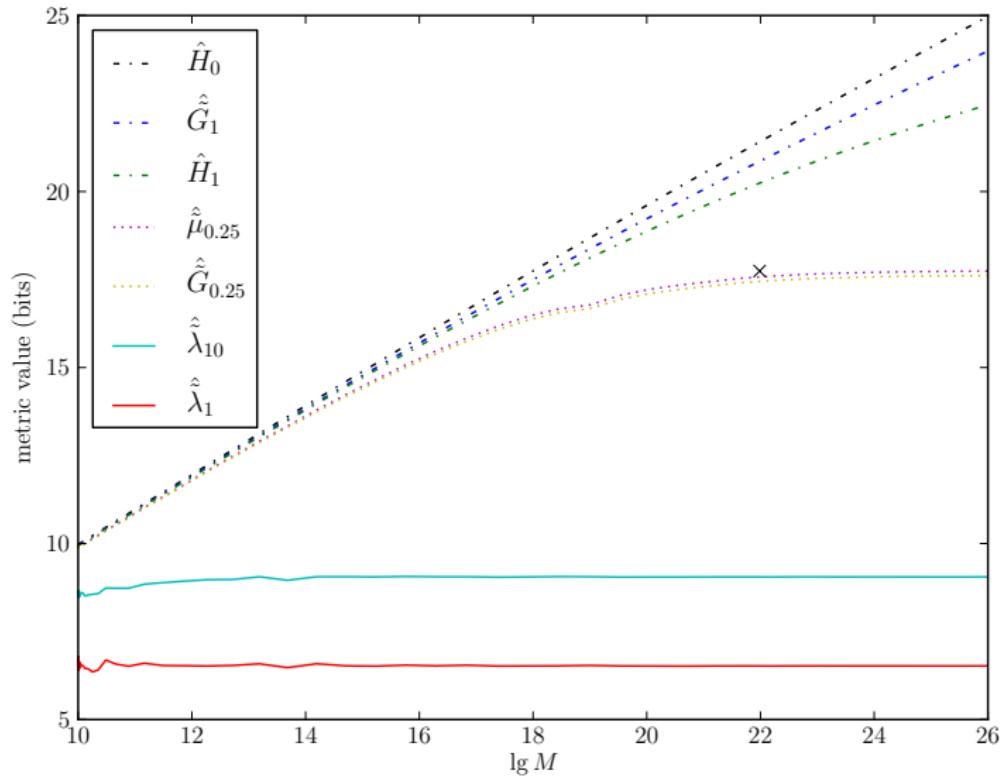
- More complicated:

$$\tilde{G}_\alpha(\mathcal{X}) = \lg \left[\frac{2 \cdot G_\alpha(\mathcal{X})}{\lceil \alpha \rceil} - 1 \right] - \lg(2 - \lceil \alpha \rceil)$$

- Sanity check:

$$\tilde{\lambda}_\beta(\mathcal{U}_N) = \tilde{\mu}_\alpha(\mathcal{U}_N) = \tilde{G}_\alpha(\mathcal{U}_N) = \lg N$$

Sample size is a major problem for passwords...



Poor password implementations

Results from a study of password authentication in the wild:

- 29–40% of websites don't hash passwords during storage
- 41% of websites don't use any encryption for password submission
 - 22% do so incompletely
- 84% of websites don't rate-limit against guessing attacks
- 97% of websites leak usernames to simple

Joseph Bonneau and Sören Preibusch.

The password thicket: technical and market failures in human authentication on the web.

Workshop on the Economics of Information Security, 2010.