What's in a Name?

Evaluating Statistical Attacks on Personal Knowledge Questions

Joseph Bonneau
jcb82@cl.cam.ac.uk

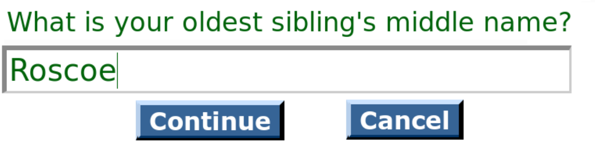
Mike Just
Greg Matthews

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

THE UNIVERSITY OF EDINBURGH

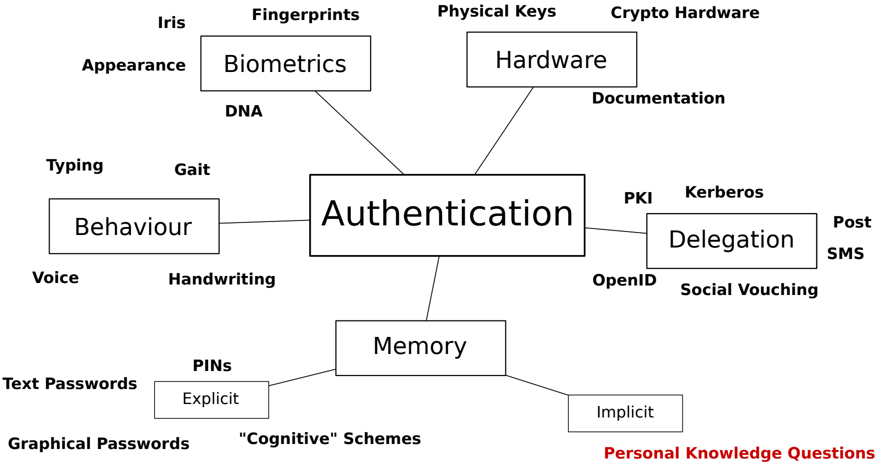
Financial Cryptography and Data Security 2010
Tenerife, Spain
January 26, 2010

Research Question



How “secure” are personal knowledge questions against guessing?

Authenticating Humans



Personal Knowledge Questions

- Pros
  - Cost
  - Memorability?
- Cons
  - Privacy
  - Security

Authentication on the Web

1. Text Passwords
2. Delegation
3. Personal Knowledge Questions

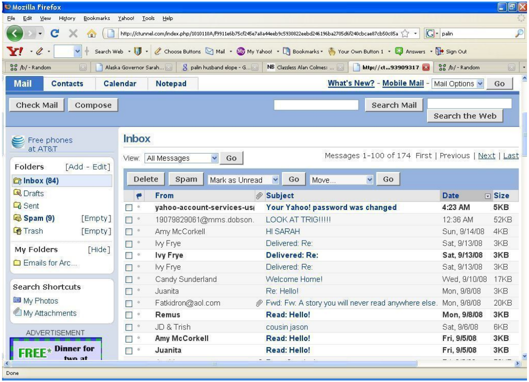
Trends:

- OpenID may make delegation preferred method
- Large webmail providers becoming the root of trust

In the News

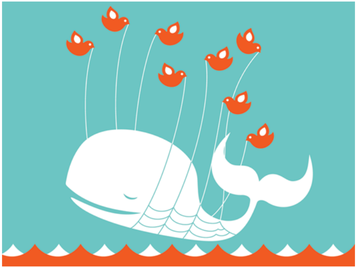
- Paris Hilton T-Mobile Sidekick, 2005-02-20
- Sarah Palin Yahoo! email, 2008-09-16
- Twitter corporate Google Docs, 2009-07-16

In the News



- Paris Hilton T-Mobile Sidekick, 2005-02-20
- Sarah Palin Yahoo! email, 2008-09-16
- Twitter corporate Google Docs, 2009-07-16

In the News



- Paris Hilton T-Mobile Sidekick, 2005-02-20
- Sarah Palin Yahoo! email, 2008-09-16
- Twitter corporate Google Docs, 2009-07-16

Protocol Model

Client		**Server**
	I am i $\longrightarrow$	
		Increment t_i
		Select $q \xleftarrow{R} Q_i$
	Please answer q $\longleftarrow$	
	The answer is x $\longrightarrow$	
		Verify x

Targeted Attacker

- Attack a **specific** i
- Real-world identity of i is **known**
- Per-target research possible

Targeted Attacker

- Web search
  - Used in Hilton, Palin compromises
- Public records
  - Griffith et. al: 30% of individual's mother's maiden names found via marriage, birth records
- Social engineering
- Dumpster diving, burglary
- Acquaintance attacks
  - Schecter et. al: $\sim$ 25% of questions guessed by friends, family

Trawling Attacker

- Attack **all** $i \in I$ from a large set I
- Real-world identities are **unknown**
- Population-wide statistics

Trawling Attacker

- **Blind** attack
  - Don't understand i or q
  - CAPTCHA-ised protocols or user-written questions
    - "What do I want to do?"
- **Statistical** attack
  - Understand q but not i
  - Guess most likely answers
  - Thought to be used in Twitter compromise

Measuring Security Against Guessing

Which is “harder” to guess:

- Surname of randomly chosen Internet user
- Randomly chosen 4-digit PIN

Mathematics of Guessing

- Answer X is drawn from a finite, known distribution $\mathcal{X}$
- $|\mathcal{X}| = N$
- $P(X = x_i) = p_i$ for each possible answer x_i
- $\mathcal{X}$ is monotonically decreasing: $p_1 \geq p_2 \geq \cdots \geq p_N$

Goal: guess X using as few queries "is $X = x_i$?"as possible.

Shannon Entropy

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

- H_1(surname) = **16.2 bits**
- H_1(PIN) = **13.3 bits**
- **Meaning:** Expected number of queries "Is $X \in \mathcal{S}$?" for arbitrary subsets $\mathcal{S} \subseteq \mathcal{X}$ needed to guess X. (Source-Coding Theorem)

Shannon Entropy

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

- H_1(surname) = **16.2 bits**
- H_1(PIN) = **13.3 bits**

- **Meaning:** Expected number of queries "Is $X \in \mathcal{S}$?" for arbitrary subsets $\mathcal{S} \subseteq \mathcal{X}$ needed to guess X. (Source-Coding Theorem)

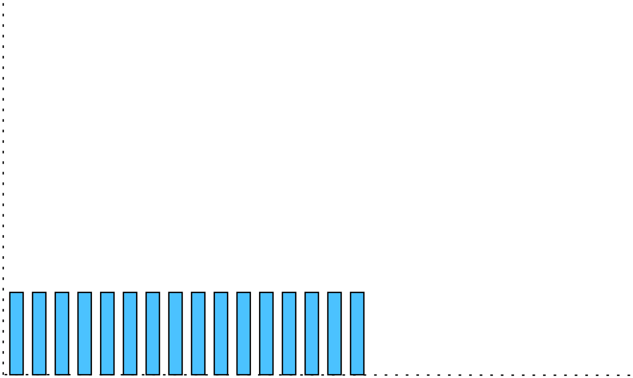
Guessing Entropy

$$G(\mathcal{X}) = E\left[\#_{\text{guesses}}(X \xleftarrow{R} \mathcal{X})\right] = \sum_{i=1}^{N} p_i \cdot i$$

- $G(\text{surname}) \approx$ **137000 guesses**
- $G(\text{PIN}) \approx$ **5000 guesses**

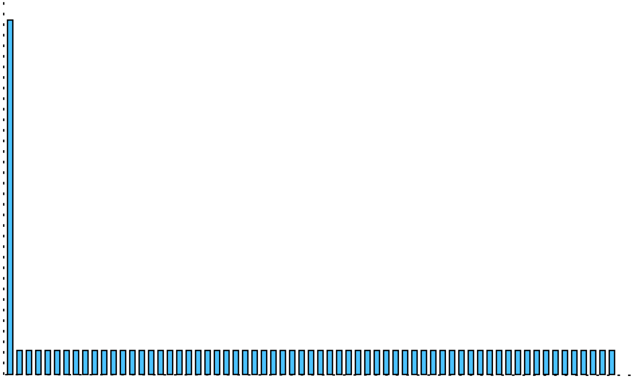
- **Meaning:** Expected number of queries "Is $X = x_i$?" for $i = 1, 2, \ldots, N$ (optimal sequential guessing)

The Trouble with Guessing



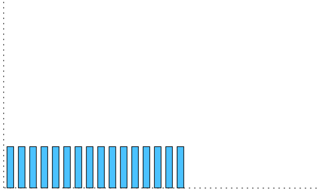
- $\mathcal{U}_{16}$ — $N = 16$, $p_1 = p_2 = \cdots = p_{16} = \frac{1}{16}$
- $H_1(\mathcal{U}_{16})$ = **4 bits**
- $G(\mathcal{U}_{16})$ = **8.5 guesses**

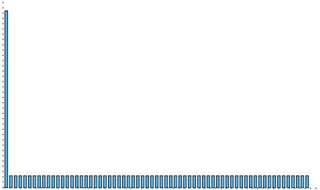
The Trouble with Guessing



- $\mathcal{X}_{65}$ — $N = 65$, $p_1 = \frac{1}{2}$, $p_2 = \cdots = p_{65} = \frac{1}{128}$
- $H_1(\mathcal{X}_{65})$ = **4 bits**
- $G(\mathcal{X}_{65})$ = **17.25 guesses**

The Trouble with Guessing





- $H_1(\mathcal{X}_{65}) = H_1(\mathcal{U}_{16})$
- $G(\mathcal{X}_{65}) > G(\mathcal{U}_{16})$
- Adversary can guess $X \xleftarrow{R} \mathcal{X}_{65}$ in 1 try half the time!

Marginal Guessing

Suppose Eve wants to guess any k out of m 4-digit PINS

PIN #1	PIN #2	PIN #3	...	PIN #m
0000	0000	0000	...	0000
0001	0001	0001	...	0001
0002	0002	0002	...	0002
...	...	...	...	...
9998	9998	9998	...	9998
9999	9999	9999	...	9999

Marginal Guessing

Suppose Eve wants to guess any k out of m 4-digit PINS

PIN #1	PIN #2	PIN #3	...	PIN #m
0000	0000	0000	...	0000
0001	0001	0001	...	0001
0002	0002	0002	...	0002
...	...	...	...	...
9998	9998	9998	...	9998
9999	9999	9999	...	9999

Marginal Guessing

Suppose Eve wants to guess any k out of m 4-digit PINS

PIN #1	PIN #2	PIN #3	...	PIN #m
0000	0000	0000	...	0000
0001	0001	0001	...	0001
0002	0002	0002	...	0002
...	...	...	...	...
9998	9998	9998	...	9998
9999	9999	9999	...	9999

Marginal Guessing

Suppose Eve wants to guess any k out of m 4-digit PINS

PIN #1	PIN #2	PIN #3	...	PIN #m
0000	0000	0000	...	0000
0001	0001	0001	...	0001
0002	0002	0002	...	0002
...	...	...	...	...
9998	9998	9998	...	9998
9999	9999	9999	...	9999

Any order of guessing is equivalent.

Marginal Guessing

Suppose Mallory wants to guess any k out of m surnames

Name #1	Name #2	Name #3	...	Name #m
Smith	Smith	Smith	...	Smith
Jones	Jones	Jones	...	Jones
Johnson	Johnson	Johnson	...	Johnson
...	...	...	...	...
Ytterock	Ytterock	Ytterock	...	Ytterock
Zdrzynski	Zdrzynski	Zdrzynski	...	Zdrzynski

Marginal Guessing

Suppose Mallory wants to guess any k out of m surnames

Name #1	Name #2	Name #3	...	Name #m
Smith	Smith	Smith	...	Smith
Jones	Jones	Jones	...	Jones
Johnson	Johnson	Johnson	...	Johnson
...	...	...	...	...
Ytterock	Ytterock	Ytterock	...	Ytterock
Zdrzynski	Zdrzynski	Zdrzynski	...	Zdrzynski

Marginal Guessing

Suppose Mallory wants to guess any k out of m surnames

Name #1	Name #2	Name #3	...	Name #m
Smith	Smith	Smith	...	Smith
Jones	Jones	Jones	...	Jones
Johnson	Johnson	Johnson	...	Johnson
...	...	...	...	...
Ytterock	Ytterock	Ytterock	...	Ytterock
Zdrzynski	Zdrzynski	Zdrzynski	...	Zdrzynski

Obvious optimal strategy

Measuring Security Against Guessing

Given 100 accounts:

- PIN: 50% chance of success after **5000 guesses**
- Surname: 50% chance of success after **168 guesses**

Marginal Guessing

- Neither H_1 nor G model an adversary who can **give up**
- **Marginal Guesswork**
  Give up after reaching probability α of success:

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

- **Marginal Success Rate**
  Give up after β guesses:

$$\lambda_\beta(\mathcal{X}) = \sum_{i=1}^{\beta} p_i$$

Marginal Guessing

- Neither H_1 nor G model an adversary who can **give up**
- **Marginal Guesswork**
  Give up after reaching probability α of success:

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

- **Marginal Success Rate**
  Give up after β guesses:

$$\lambda_\beta(\mathcal{X}) = \sum_{i=1}^{\beta} p_i$$

Marginal Guessing

- Neither H_1 nor G model an adversary who can **give up**
- **Marginal Guesswork**
  Give up after reaching probability α of success:

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

- **Marginal Success Rate**
  Give up after β guesses:

$$\lambda_\beta(\mathcal{X}) = \sum_{i=1}^{\beta} p_i$$

Conversion to Bits

- H_1, G, μ_α, λ_β all have different units
- To convert $G(\mathcal{X})$ to bits
  1. Find discrete uniform $\mathcal{U}_N$ with $G(\mathcal{U}_N) = G(\mathcal{X})$
  2. "Effective key length" $\tilde{G}(\mathcal{X}) = \lg N$
- In general:

$$\boxed{\tilde{G}(\mathcal{X}) = \lg[2 \cdot G(\mathcal{X}) - 1]}$$

- Similarly:

$$\boxed{\tilde{\mu}_\alpha(\mathcal{X}) = \lg\left(\frac{\mu_\alpha(\mathcal{X})}{\alpha}\right)} \qquad \boxed{\tilde{\lambda}_\beta(\mathcal{X}) = \lg\left(\frac{\beta}{\lambda_\beta(\mathcal{X})}\right)}$$

- Nice property: $\tilde{\lambda}_1$ is the min-entropy H_∞

Conversion to Bits

- H_1, G, μ_α, λ_β all have different units
- To convert $G(\mathcal{X})$ to bits
  1. Find discrete uniform $\mathcal{U}_N$ with $G(\mathcal{U}_N) = G(\mathcal{X})$
  2. "Effective key length" $\tilde{G}(\mathcal{X}) = \lg N$
- In general:

$$\boxed{\tilde{G}(\mathcal{X}) = \lg[2 \cdot G(\mathcal{X}) - 1]}$$

- Similarly:

$$\boxed{\tilde{\mu}_\alpha(\mathcal{X}) = \lg\left(\frac{\mu_\alpha(\mathcal{X})}{\alpha}\right)} \qquad \boxed{\tilde{\lambda}_\beta(\mathcal{X}) = \lg\left(\frac{\beta}{\lambda_\beta(\mathcal{X})}\right)}$$

- Nice property: $\tilde{\lambda}_1$ is the min-entropy H_∞

Conversion to Bits

- H_1, G, μ_α, λ_β all have different units
- To convert $G(\mathcal{X})$ to bits
  1. Find discrete uniform $\mathcal{U}_N$ with $G(\mathcal{U}_N) = G(\mathcal{X})$
  2. "Effective key length" $\tilde{G}(\mathcal{X}) = \lg N$
- In general:

$$\boxed{\tilde{G}(\mathcal{X}) = \lg[2 \cdot G(\mathcal{X}) - 1]}$$

- Similarly:

$$\boxed{\tilde{\mu}_\alpha(\mathcal{X}) = \lg\left(\frac{\mu_\alpha(\mathcal{X})}{\alpha}\right)} \qquad \boxed{\tilde{\lambda}_\beta(\mathcal{X}) = \lg\left(\frac{\beta}{\lambda_\beta(\mathcal{X})}\right)}$$

- Nice property: $\tilde{\lambda}_1$ is the min-entropy H_∞

Conversion to Bits

- H_1, G, μ_α, λ_β all have different units
- To convert $G(\mathcal{X})$ to bits
  1. Find discrete uniform $\mathcal{U}_N$ with $G(\mathcal{U}_N) = G(\mathcal{X})$
  2. "Effective key length" $\tilde{G}(\mathcal{X}) = \lg N$
- In general:

$$\boxed{\tilde{G}(\mathcal{X}) = \lg[2 \cdot G(\mathcal{X}) - 1]}$$

- Similarly:

$$\boxed{\tilde{\mu}_\alpha(\mathcal{X}) = \lg\left(\frac{\mu_\alpha(\mathcal{X})}{\alpha}\right)} \qquad \boxed{\tilde{\lambda}_\beta(\mathcal{X}) = \lg\left(\frac{\beta}{\lambda_\beta(\mathcal{X})}\right)}$$

- Nice property: $\tilde{\lambda}_1$ is the min-entropy H_∞

Conversion to Bits

- H_1, G, μ_α, λ_β all have different units
- To convert $G(\mathcal{X})$ to bits
  1. Find discrete uniform $\mathcal{U}_N$ with $G(\mathcal{U}_N) = G(\mathcal{X})$
  2. "Effective key length" $\tilde{G}(\mathcal{X}) = \lg N$
- In general:

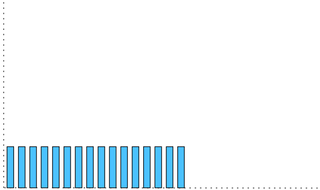
$$\boxed{\tilde{G}(\mathcal{X}) = \lg[2 \cdot G(\mathcal{X}) - 1]}$$

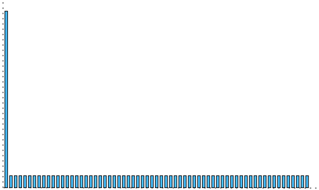
- Similarly:

$$\boxed{\tilde{\mu}_\alpha(\mathcal{X}) = \lg\left(\frac{\mu_\alpha(\mathcal{X})}{\alpha}\right)} \qquad \boxed{\tilde{\lambda}_\beta(\mathcal{X}) = \lg\left(\frac{\beta}{\lambda_\beta(\mathcal{X})}\right)}$$

- Nice property: $\tilde{\lambda}_1$ is the min-entropy H_∞

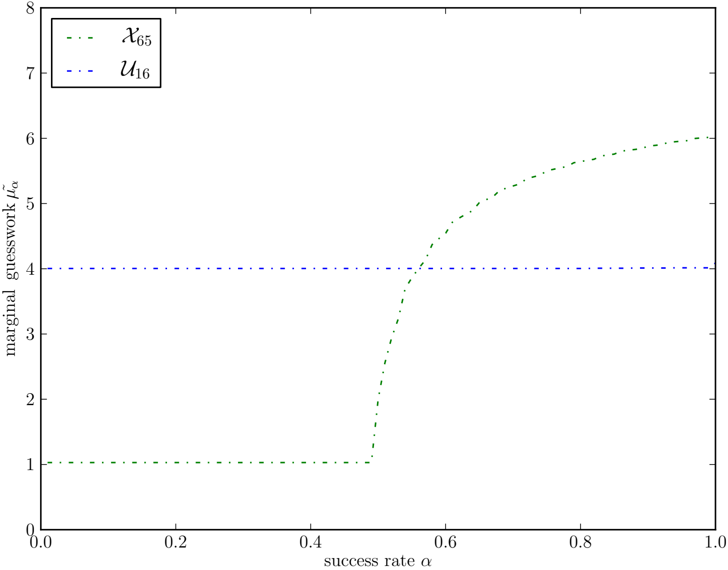
Examples



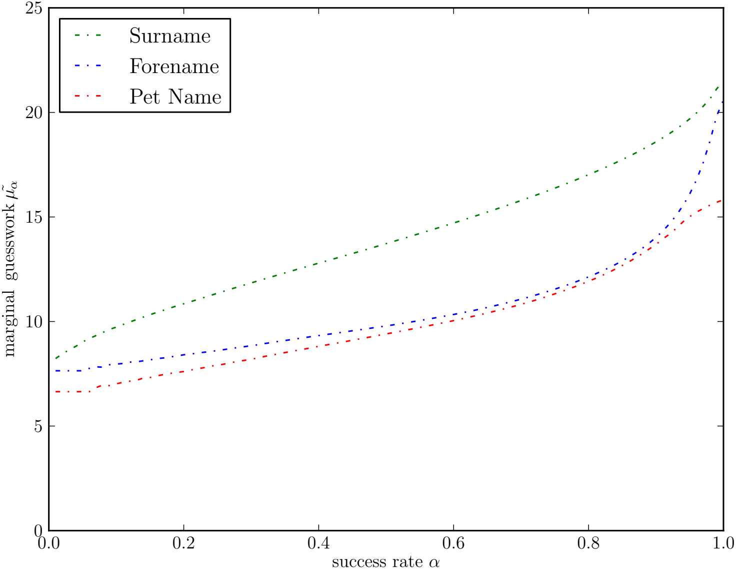


	$\mathcal{U}_{16}$	$\mathcal{X}_{65}$
H_1	4	4
$\tilde{G}$	4	5.1
$\tilde{\mu}_{\frac{1}{2}}$	4	1
$\tilde{\lambda}_8$	4	3.8

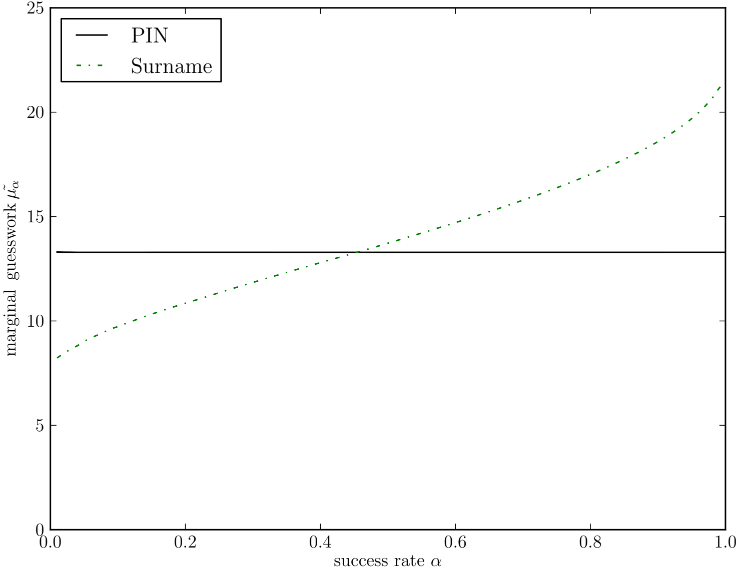
The Complete View



The Complete View



The Complete View



Incomparability Theorems

Theorem (adapted from Pliam)

Given any $m > 0$, $\beta > 0$ *and* $0 < \alpha < 1$, *there exists a distribution* $\mathcal{X}$ *such that* $\tilde{\mu}_\alpha(\mathcal{X}) < H_1(\mathcal{X}) - m$ *and* $\tilde{\lambda}_\beta(\mathcal{X}) < H_1(\mathcal{X}) - m$.

Theorem (adapted from Boztaş)

Given any $m > 0$, $\beta > 0$ *and* $0 < \alpha < 1$, *there exists a distribution* $\mathcal{X}$ *such that* $\tilde{\mu}_\alpha(\mathcal{X}) < \tilde{G}(\mathcal{X}) - m$ *and* $\tilde{\lambda}_\beta(\mathcal{X}) < \tilde{G}(\mathcal{X}) - m$.

Theorem (new)

Given any $m > 0, \alpha_1 > 0$, *and* $\alpha_2 > 0$ *with* $0 < \alpha_1 < \alpha_2 < 1$, *there exists a distribution* $\mathcal{X}$ *such that* $\tilde{\mu}_{\alpha_1}(\mathcal{X}) < \tilde{\mu}_{\alpha_1}(\mathcal{X}) - m$.

Application to Personal Knowledge Questions

- λ_3 models the usual cutoff of 3 guesses
- $\lambda_1 = H_\infty$ models an attacker with infinite accounts
- $\mu_{\frac{1}{2}}$ is reasonable for offline attacks

Common Answer Categories

Category	Example Questions
Forename	What is your grandfather's first name? What is your father's middle name?
Surname	What is your mother's maiden name? Who was your favourite school teacher?
Pet Name	What was your first pet's name?
Place	In what city were you born? Where did you go for your honeymoon? What is the name of your high school?
Other	What was your grandfather's occupation? What is your favourite movie?

Common Answer Categories

- Just and Aspinall: 70% of answers are proper names
  - 25% surname
  - 10% forename
  - 15% pet name
  - 20% place name
- Most others are trivially insecure
  - What is my favourite colour?
  - What is the worst day of the week?

Our Data Sources

- Collected name data from published government sources
  - Most census statistics suppress uncommon names
  - Doesn't impact $\tilde{\mu}_\alpha$, $\tilde{\lambda}_\beta$
  - Can still get lower bounds on H_1, $\tilde{G}$
- Crawled Facebook for 65 M full names

Overview

Source	H_0	H_1	$\tilde{G}$	H_2	$\tilde{\mu}_{\frac{1}{2}}$	$\tilde{\lambda}_3$	H_∞	x_1
UK City	9.2	8.5	8.8	5.9	8.7	4.4	3.0	London
Pet Name	15.8	11.7	13.1	9.2	9.4	6.5	6.4	Lucky
UK High School	8.7	8.5	8.2	8.3	8.0	7.4	7.3	Holyrood
Forename	20.6	12.4	15.7	9.9	9.8	7.4	7.3	David
Surname	21.5	16.2	18.1	12.1	13.7	8.1	7.7	Smith
Full Name	25.1	24.0	24.4	20.8	23.3	14.4	14.4	Maria Gonzalez

Surnames

Source	H_0	H_1	$\tilde{G}$	H_2	$\tilde{\mu}_{\frac{1}{2}}$	$\tilde{\lambda}_3$	H_∞	x_1
South Korea	7.5	4.6	4.5	3.5	3.3	2.7	2.2	Kim
Chile	6.8	6.6	6.3	6.3	6.0	4.9	4.5	González
Spain	9.6	8.9	9.1	7.6	8.8	5.4	5.0	Garcia
Japan	14.5	11.3	12.0	9.0	9.2	6.2	6.0	Satō
Finland	13.8	12.2	12.3	10.5	10.5	7.9	7.8	Virtanen
England	17.4	13.3	14.6	10.2	11.0	6.7	6.4	Smith
Estonia	11.9	11.7	11.7	11.3	11.6	7.9	7.6	Ivanov
Australia	18.6	14.1	15.3	10.9	11.8	7.4	6.8	Smith
Norway	13.7	12.5	13.0	9.9	11.9	6.5	6.4	Hansen
USA	19.1	14.9	16.9	10.9	12.3	7.2	6.9	Smith
Facebook	21.5	16.2	18.1	12.1	13.7	8.1	7.7	Smith

Forenames

Source	H_0	H_1	$\tilde{G}$	H_2	$\tilde{\mu}_{\frac{1}{2}}$	$\tilde{\lambda}_3$	H_∞	x_1
Iceland (♀)	7.9	7.5	7.3	6.9	6.8	5.1	4.9	Guðrún
Spain (♀)	8.3	7.9	7.8	7.3	7.1	5.3	5.1	Maria
Belgium (♀)	15.2	10.1	10.9	8.1	8.2	5.5	4.9	Maria
USA (♀)	15.1	10.9	12.9	8.7	8.3	6.5	6.3	Jennifer
Spain (♂)	8.6	7.8	7.8	6.9	6.6	4.9	4.8	Jose
Iceland (♂)	7.9	7.5	7.3	6.9	6.8	5.0	4.8	Jón
USA (♂)	15.2	9.4	12.0	7.2	6.9	5.2	5.0	Michael
Belgium (♂)	15.0	9.7	10.4	8.2	7.8	6.1	5.7	Jean
Iceland	8.9	8.5	8.3	7.9	7.7	5.9	5.8	Jón
Spain	9.7	9.0	8.9	8.1	7.9	6.0	5.9	Jose
Belgium	15.0	10.2	10.3	8.8	8.7	6.1	5.7	Maria
USA	16.7	11.2	14.0	8.7	8.6	6.2	5.9	Michael
Facebook	20.6	12.4	15.7	9.9	9.8	7.4	7.3	David

Forenames over time

Source	H_0	H_1	$\tilde{G}$	H_2	$\tilde{\mu}_{\frac{1}{2}}$	$\tilde{\lambda}_3$	H_∞	x_1
USA, 1950 (♀)	11.8	8.6	9.1	7.1	6.8	5.2	5.0	Mary
USA, 1950 (♂)	11.7	7.7	8.3	6.2	5.8	4.6	4.6	James
USA, 1960 (♀)	11.9	9.1	9.5	7.6	7.1	5.6	5.2	Lisa
USA, 1960 (♂)	11.9	7.9	8.6	6.4	5.9	4.7	4.6	Michael
USA, 1970 (♀)	12.1	9.7	10.3	7.7	7.6	5.5	4.8	Jennifer
USA, 1970 (♂)	12.1	8.4	9.3	6.7	6.3	5.0	4.6	Michael
USA, 1980 (♀)	12.2	9.7	10.4	7.7	7.6	5.4	5.3	Jessica
USA, 1980 (♂)	12.2	8.6	9.6	6.9	6.4	5.1	4.9	Michael
USA, 1990 (♀)	12.3	10.3	10.8	8.4	8.3	6.1	6.0	Jessica
USA, 1990 (♂)	12.3	9.3	10.0	7.5	7.1	5.7	5.5	Michael
USA, 2000 (♀)	12.4	10.8	11.1	9.1	9.0	6.6	6.5	Emily
USA, 2000 (♂)	12.2	9.9	10.4	8.2	7.8	6.4	6.2	Jacob

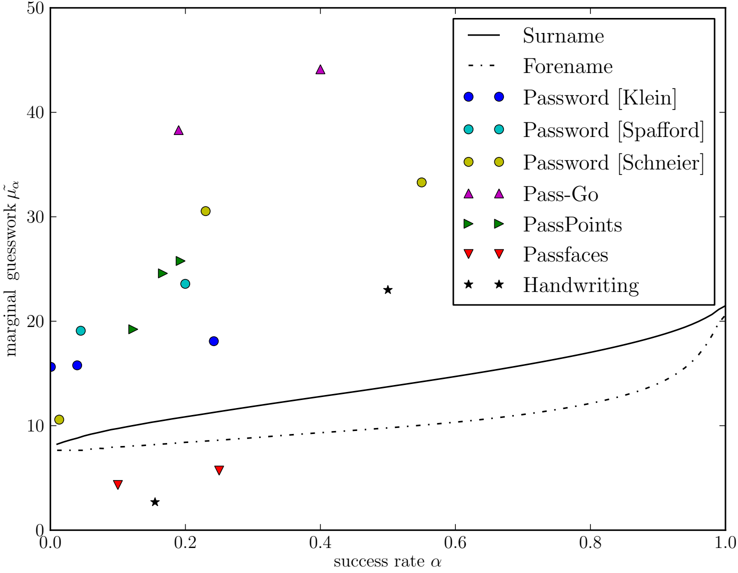
Pets

Source	H_0	H_1	$\tilde{G}$	H_2	$\tilde{\mu}_{\frac{1}{2}}$	$\tilde{\lambda}_3$	H_∞	x_1
Los Angeles	15.8	11.7	13.1	9.2	9.4	6.5	6.4	Lucky
Des Moines	13.6	11.6	12.4	9.4	9.7	6.5	6.2	Buddy
San Francisco	13.7	11.6	12.0	9.6	9.8	6.7	6.7	Buddy

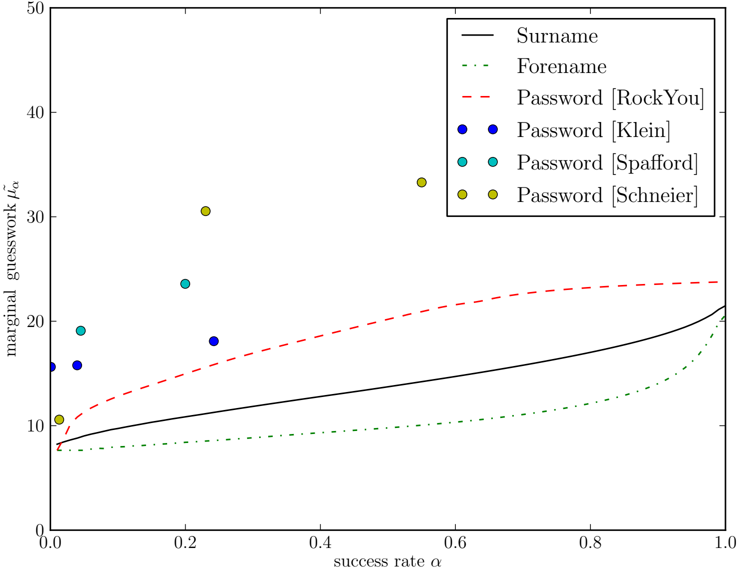
Places

Source	H_0	H_1	$\tilde{G}$	H_2	$\tilde{\mu}_{\frac{1}{2}}$	$\tilde{\lambda}_3$	H_∞	x_1
School Mascots (US)	11.8	8.1	9.3	6.2	5.7	4.5	4.1	Eagles
UK High Schools	8.7	8.5	8.2	8.3	8.0	7.4	7.3	Holyrood
UK Cities	9.2	8.5	8.8	5.9	8.7	4.4	3.0	London
Tourist Destinations	13.0	12.0	12.5	9.5	12.4	6.3	5.9	London
UK Primary Schools	14.0	13.8	13.5	13.6	13.3	12.1	12.1	Essex

Comparison to Other Authentication Schemes



Comparison to Other Authentication Schemes



Remarks

- Security even lower than expected!
- Against online attack: $\boxed{\tilde{\lambda}_3 \lessapprox \textbf{8 bits}}$
  - Compromise 1 of every 80 accounts . . .
- Against offline attack: $\boxed{\tilde{\mu}_{\frac{1}{2}} \lessapprox \textbf{12 bits}}$
  - A few thousand guesses per account . . .
- Interesting: $\tilde{\mu}_{\frac{1}{2}}$ well-approximated by H_2

Name Correlations

Dubious model: forenames chosen independently from surnames

Name Correlations

Erik Anderson 28.5000027
Scott Anderson 26.2240310808
Eric Anderson 25.7454870714
Ryan Anderson 24.9834030274
Kyle Anderson 22.59694489
Tyler Anderson 20.7791328141
Ashley Anderson 20.1428280702
...
Nicolas Anderson -10.658058566
Claudia Anderson -10.827656673
Luis Anderson -11.8887183582
Marco Anderson -12.0011017638
Ana Anderson -12.0950091322
Carlos Anderson -12.7907931815
Jose Anderson -14.4516505046
Juan Anderson -15.411686568
Maria Anderson -18.6010320036

Name Correlations

Jose Garcia 98.5011019005
Juan Garcia 82.5912299727
Carlos Garcia 79.5644630229
Luis Garcia 78.9805405513
Ana Garcia 71.4654714218
Javier Garcia 68.1730545731
Maria Garcia 65.5565931662
Miguel Garcia 59.2541621707
. . .
Scott Garcia -16.6967016634
Michael Garcia -16.781135422
Amy Garcia -17.0189476524
Ryan Garcia -18.2193592941
James Garcia -18.628543594
Matt Garcia -18.9610296901
Chris Garcia -20.1867129035
Sarah Garcia -22.3262090845

Ethnic Correlations

- Most frequently-paired names: **Maria Gonzalez**
- Least frequently-paired names: **Juan Khan**

- Knowing a target's ethnicity can **double** attack efficiency

Source	H_0	H_1	$\tilde{G}$	H_2	$\tilde{\mu}_{\frac{1}{2}}$	$\tilde{\lambda}_3$	H_∞	x_1
				Surnames				
Spanish Forenames	19.8	14.9	16.8	11.0	12.4	7.3	7.2	Gonzalez
All Forenames	21.5	16.2	18.1	12.1	13.7	8.1	7.7	Smith
				Forenames				
Spanish Surnames	17.5	11.0	13.4	8.6	8.4	6.0	5.8	Maria
All Surnames	20.6	12.4	15.7	9.9	9.8	7.4	7.3	David

Countermeasures

- If we know $\mathcal{X}$, we can actively *shape* it
  - Respond with $\perp$ for some enrolment attempts
- Naive approach: Always reject most common answers
- Better: Probabilistically reject common answers
  - For any $\mathcal{X}$, find optimal $r_1, r_2, \ldots, r_N$
  - Subject to a constraint on overall rejection rate r_*

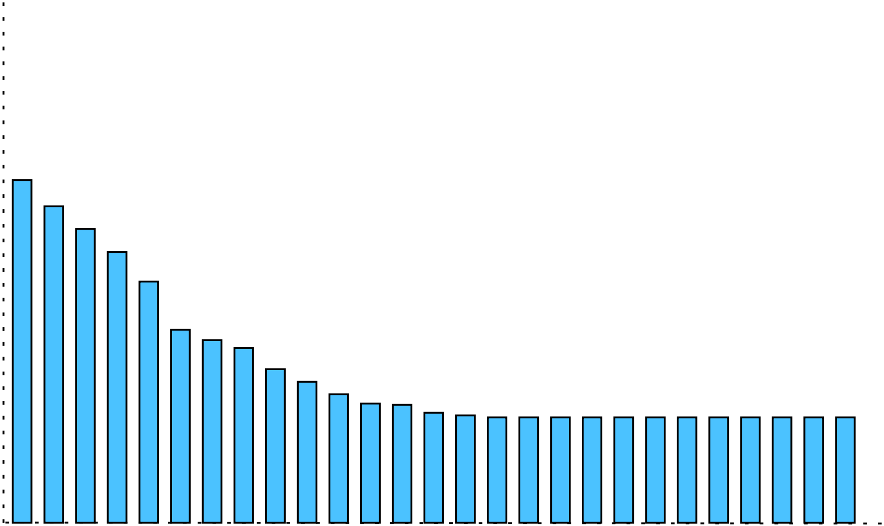
Countermeasures

- If we know $\mathcal{X}$, we can actively *shape* it
  - Respond with $\perp$ for some enrolment attempts
- Naive approach: Always reject most common answers
- Better: Probabilistically reject common answers
  - For any $\mathcal{X}$, find optimal $r_1, r_2, \ldots, r_N$
  - Subject to a constraint on overall rejection rate r_*

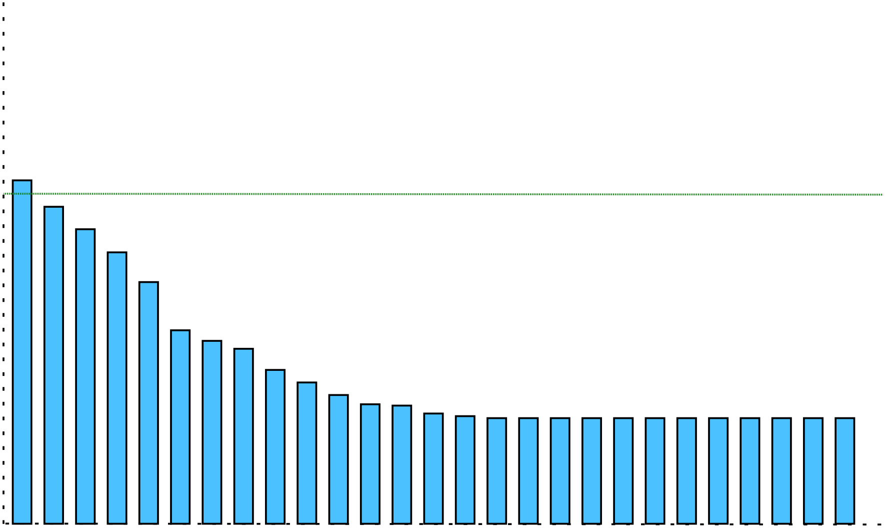
Countermeasures

- If we know $\mathcal{X}$, we can actively *shape* it
  - Respond with $\perp$ for some enrolment attempts
- Naive approach: Always reject most common answers
- Better: Probabilistically reject common answers
  - For any $\mathcal{X}$, find optimal $r_1, r_2, \ldots, r_N$
  - Subject to a constraint on overall rejection rate r_*

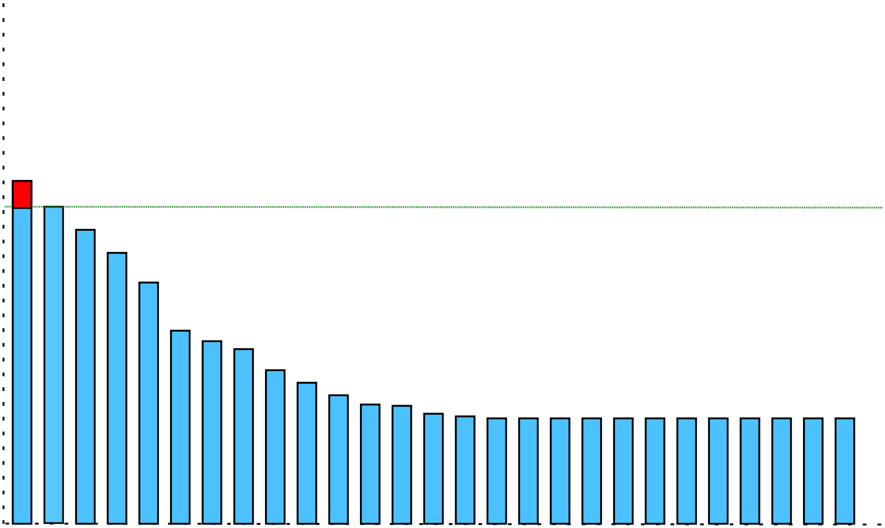
Optimal Shaping Algorithm



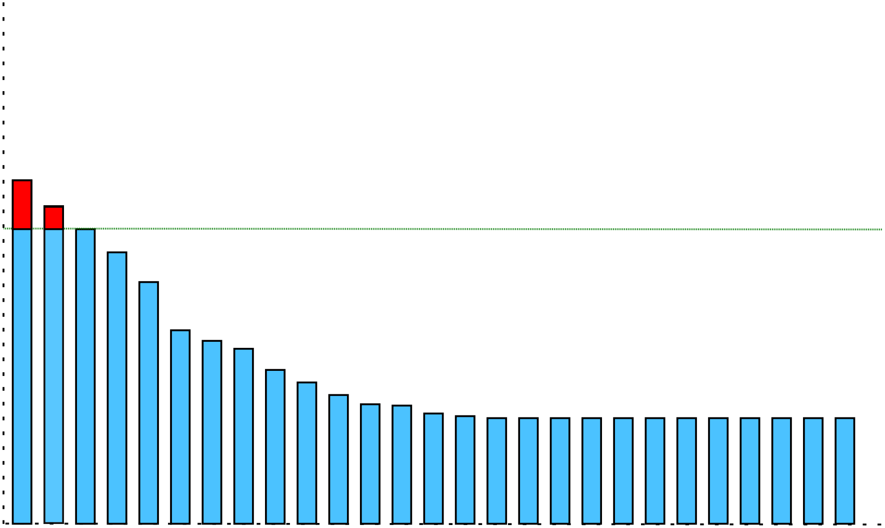
Optimal Shaping Algorithm



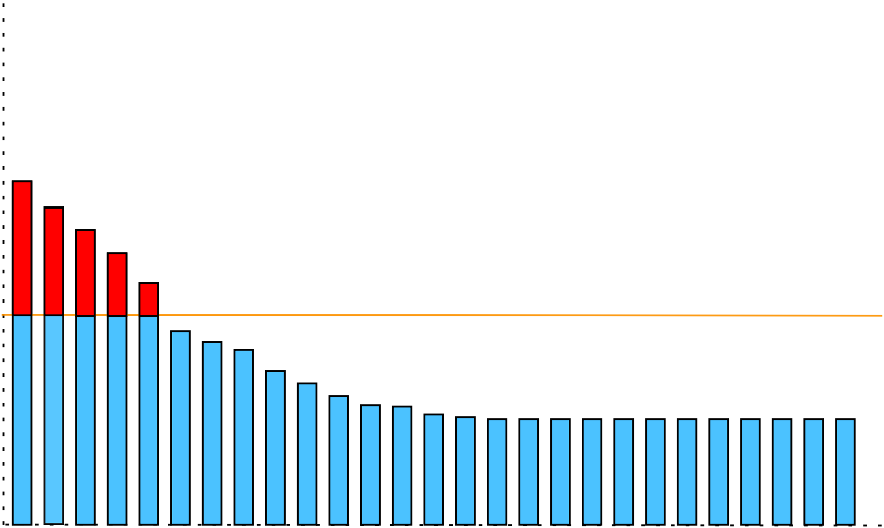
Optimal Shaping Algorithm



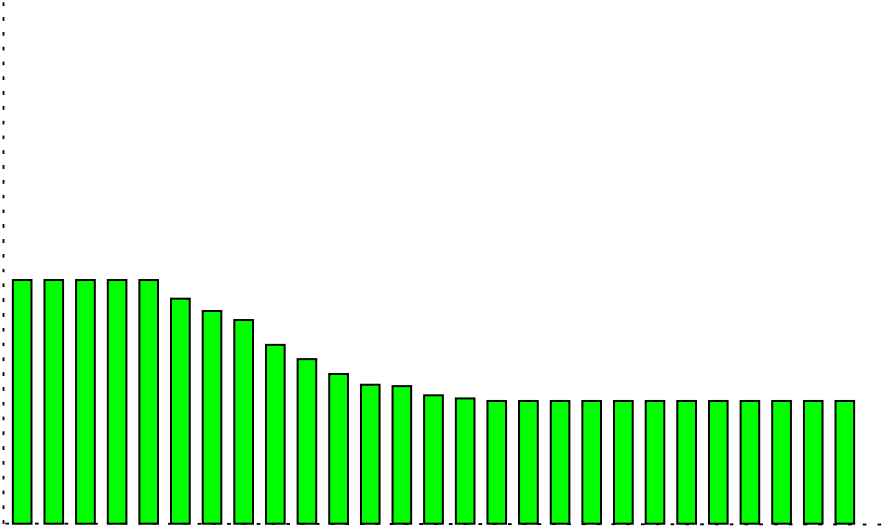
Optimal Shaping Algorithm



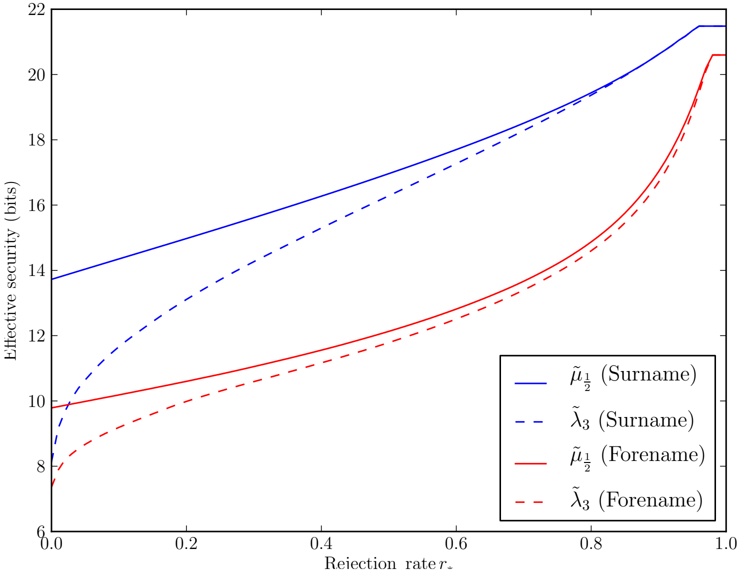
Optimal Shaping Algorithm



Optimal Shaping Algorithm



Effectiveness of Shaping



Conclusions

- Need new metrics to reason about guessing attacks
- Most deployed questions insecure against statistical attack
- Human-generated names inherently lack sufficient diversity
    - Approximated well by Zipf distribution!
- Systems should use alternate channels whenever possible