

The security of customer-chosen banking PINs

Joseph Bonneau, Sören Preibusch, Ross Anderson
jcb82,sdp36,rja14@cl.cam.ac.uk



UNIVERSITY OF
CAMBRIDGE

Computer Laboratory

FINANCIAL CRYPTO 2012
KRALENDIJK, BONAIRE, NETHERLANDS
FEB 27, 2012

What's a stolen wallet worth?



Do PINs resign pickpocketing to history?

PIN-like distributions

123456	290729
12345	79076
123456789	76789
password	59462
iloveyou	49952
princess	33291
1234567	21725
rockyou	20901
12345678	20553
abc123	16648

RockYou passwords

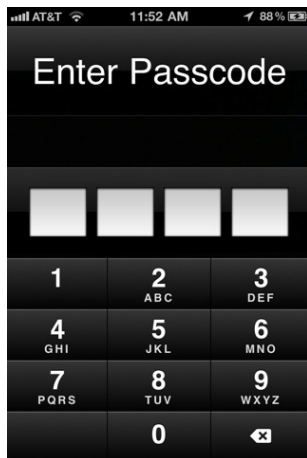
```
| grep -aEo "([0-9]|~)[0-9]{4}([0-9]|$)"
```

PIN-like distributions

1234	66193
2007	39557
2006	37229
2008	30803
2005	23683
1994	21001
1992	20126
1993	20122
1995	18761
1991	18067

1,778,095 4-digit sequences
All 10,000 possible sequences observed

PIN-like distributions



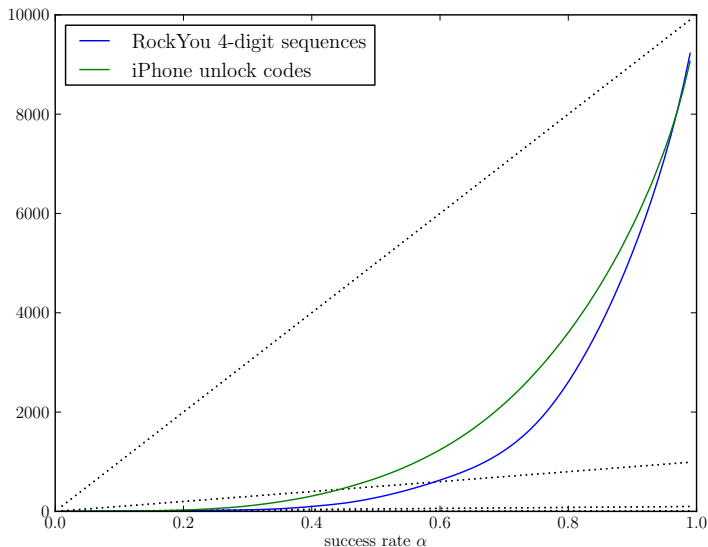
BigBrother Camera security application
Data collected by Daniel Amitay, June 2011
204,508 PINs, covering 9,954 possibilities

PIN-like distributions

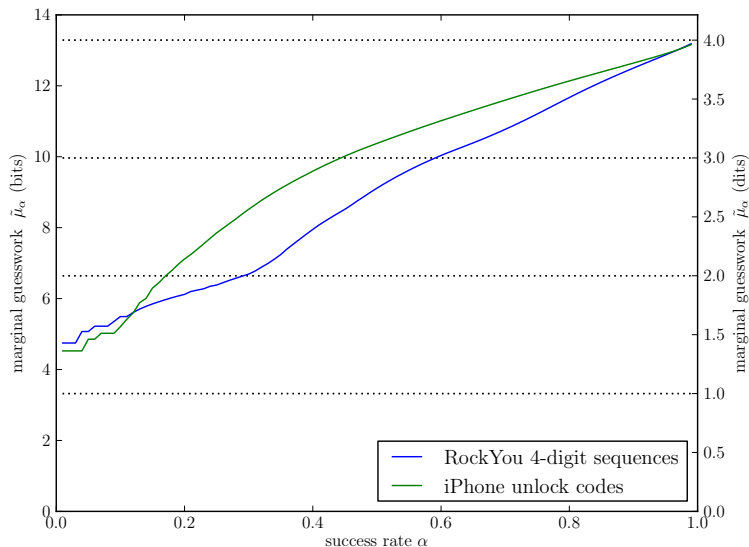
1234	8884
0000	5246
2580	4753
1111	3264
5555	1774
5683	1425
0852	1221
2222	1139
1212	944
1998	882

204,508 PINs
9,954 possibilities covered

How hard might PINs be to guess?



How hard might PINs be to guess?



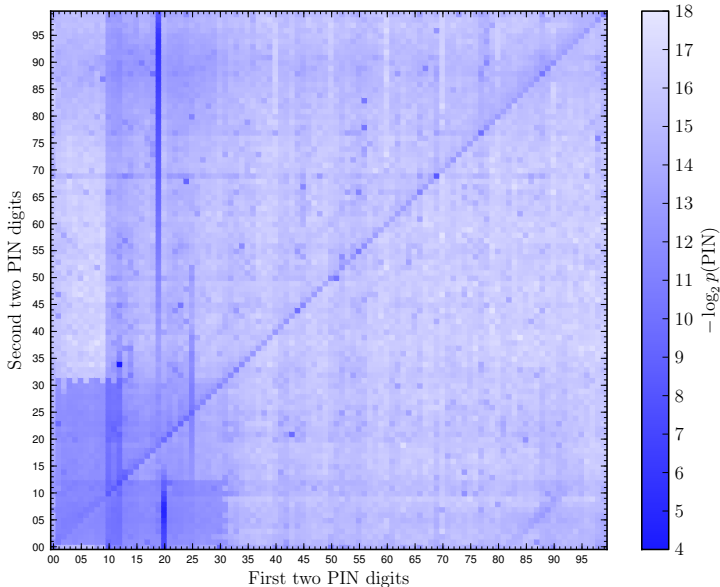
How hard might PINs be to guess?

distribution	λ_3	λ_6
RockYou	8.04%	12.29%
iPhone	9.23%	12.39%
random	0.03%	0.06%

$\lambda_\beta = \%$ of accounts covered by β optimal guesses

- $\lambda_\beta = \sum_{i=1}^{\beta} p_i$
- $\tilde{\lambda}_\beta = \lg\left(\frac{\beta}{\lambda_\beta}\right)$ (bit-converted)

Major trends in PIN selection (RockYou)



Modeling banking PINs

Linear model of PIN probability:

$$\begin{aligned} p_{1212} &= \\ &= p_{\text{date (DDMM)}} \cdot \frac{1}{365.25} \\ &+ p_{\text{date (MMDD)}} \cdot \frac{1}{365.25} \\ &+ p_{\text{repeated digit pair}} \cdot \frac{1}{100} \\ &+ \dots \\ &+ p_{\text{randomly chosen}} \cdot \frac{1}{10000} \end{aligned}$$

Modeling banking PINs

PIN selection model:

$$\begin{pmatrix} p_{0000} \\ p_{0001} \\ \vdots \\ p_{9999} \end{pmatrix} = \begin{pmatrix} f_{\text{DDMM}}(0000) & \cdots & f_{\text{rand.}}(0000) \\ f_{\text{DDMM}}(0001) & \cdots & f_{\text{rand.}}(0001) \\ \vdots & \ddots & \vdots \\ f_{\text{DDMM}}(9999) & \cdots & f_{\text{rand.}}(9999) \end{pmatrix} \cdot \begin{pmatrix} \beta_{\text{DDMM}} \\ \vdots \\ \beta_{\text{rand.}} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

- Solve for β which minimize $\sum(\varepsilon_i)^2$ with simple linear regression
- Gradually add sensible functions f
- Measure fit using $\bar{R}^2$ (avoid spurious functions)
- Sanity check: $\forall_f(\beta_f > 0)$

Modeling banking PINs

PIN selection model:

$$\begin{pmatrix} p_{0000} \\ p_{0001} \\ \vdots \\ p_{9999} \end{pmatrix} = \begin{pmatrix} f_{\text{DDMM}}(0000) & \cdots & f_{\text{rand.}}(0000) \\ f_{\text{DDMM}}(0001) & \cdots & f_{\text{rand.}}(0001) \\ \vdots & \ddots & \vdots \\ f_{\text{DDMM}}(9999) & \cdots & f_{\text{rand.}}(9999) \end{pmatrix} \cdot \begin{pmatrix} \beta_{\text{DDMM}} \\ \vdots \\ \beta_{\text{rand.}} \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

- Solve for β which minimize $\sum (\varepsilon_i)^2$ with simple linear regression
- Gradually add sensible functions f
- Measure fit using $\bar{R}^2$ (avoid spurious functions)
- Sanity check: $\forall_f (\beta_f > 0)$

Dealing with word-based PINs



Keypad entry of love

Dealing with word-based PINs

love	2643
pink	747
poop	644
baby	616
sexy	529
alex	398
star	373
mike	354
blue	311
ryan	291
josh	277
nick	273
lala	270
pimp	257
john	252

four letter passwords, RockYou

Dealing with word-based PINs

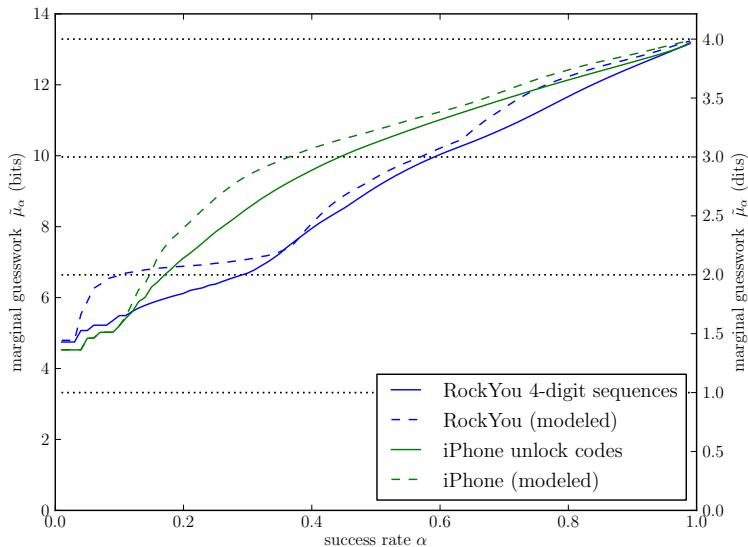
5683	2655	love, loud
7465	748	pink
2229	735	baby, abby
7667	652	poop, poms
7399	541	sexy, rexy
6453	435	mike, nike, milf, mile
2539	405	alex, blew
7827	375	star
5252	331	lala, jaja, kaka, kala
2583	318	blue, clue
5674	316	josh, lori, kori, jori
7926	297	ryan, swan
7467	289	pimp, shop, sims, rios
3825	288	fuck, duck
6425	285	nick, mick

model for word-based PINs

Results of regression model

factor	example	RockYou	iPhone
date			
DDMM	2311	5.26	1.38
DMYY	3876	9.26	6.46
MMDD	1123	10.00	9.35
MMYY	0683	0.67	0.20
YYYY	1984	33.39	7.12
<i>total</i>		<i>58.57</i>	<i>24.51</i>
keypad			
adjacent	6351	1.52	4.99
box	1425	0.01	0.58
corners	9713	0.19	1.06
cross	8246	0.17	0.88
diagonal swipe	1590	0.10	1.36
horizontal swipe	5987	0.34	1.42
spelled word	5683	0.70	8.39
vertical swipe	8520	0.06	4.28
<i>total</i>		<i>3.09</i>	<i>22.97</i>
numeric			
ending in 69	6869	0.35	0.57
digits 0-3 only	2000	3.49	2.72
digits 0-6 only	5155	4.66	5.96
repeated pair	2525	2.31	4.11
repeated quad	6666	0.40	6.67
sequential down	3210	0.13	0.29
sequential up	4567	3.83	4.52
<i>total</i>		<i>15.16</i>	<i>24.85</i>
<i>random selection</i>	3271	<i>23.17</i>	<i>27.67</i>
$\bar{R}^2$		<i>0.79</i>	<i>0.93</i>

Results of regression model



Survey of banking customers

The screenshot shows a Windows Internet Explorer browser window displaying the Amazon Mechanical Turk interface. The page title is "Amazon Mechanical Turk - Windows Internet Explorer - [InPrivate]". The URL is "https://www.mturk.com/mturk/preview?groupID=23...". The page header includes the Amazon Mechanical Turk logo and navigation links: "Your Account", "HITS", and "Qualifications". A status bar indicates "217,364 HITS available now". Below the header, there are filters for "All HITS", "HITS Available To You", and "HITS Assigned To You". A search bar shows "find HITS containing" followed by a text input field. A reward filter is set to "that pay at least \$ 0.00". A timer shows "Timer: 00:00:00 of 60 minutes". There are buttons for "Accept HIT" and "Skip HIT". A section titled "Short research survey about banking security" lists the requester as "University of Cambridge" and qualifications as "Location is US". The reward is "\$0.40 per HIT", with "HITS Available: 1" and "Duration: 60 minutes". The main content area displays the "Research Questionnaire" from the "UNIVERSITY OF CAMBRIDGE". It includes instructions for participants, a "Check now" button, and a preview of the first question: "1. Do you regularly use a PIN number with your payment cards?".

amazonmechanicalturk
Artificial Artificial Intelligence

Your Account HITS Qualifications 217,364 HITS available now

All HITS | HITS Available To You | HITS Assigned To You

find HITS containing that pay at least \$ 0.00 for which you are qualified require Master Qualification 60

Timer: 00:00:00 of 60 minutes

Want to work on this HIT? Accept HIT Want to see other HITs? Skip HIT

Short research survey about banking security
Requester: University of Cambridge
Qualifications Required: Location is US

Reward: \$0.40 per HIT HITS Available: 1 Duration: 60 minutes

Research Questionnaire

UNIVERSITY OF CAMBRIDGE

For our research on computer interfaces, we would like to ask you to complete a short questionnaire. Please read all instructions carefully and may reject your task if you do not agree to participate. This project is conducted by the University of Cambridge. Your responses are sent directly to the researchers and not recorded by mTurk. Your answers are sent directly to the researchers and not recorded by mTurk.

Any questions? Contact the requester.

1. Do you regularly use a PIN number with your payment cards?

By payment cards, we mean ATM cards, credit cards, debit cards or other cards issued by your bank or a financial institution.

PIN survey released to 1,351 mTurk users, Sept 2011
(1,337 valid responses)

Survey of banking customers

- Overall, how often do you type your PIN when making a purchase in a shop? And how often do you type your PIN at an ATM/cash machine? ($N = 1177$)

	shop		ATM	
Multiple times per day	81	(6.9%)	14	(1.2%)
About once per day	117	(9.9%)	19	(1.6%)
Several times a week	342	(29.1%)	118	(10.0%)
About once per week	241	(20.5%)	384	(32.6%)
About once per month	113	(9.6%)	418	(35.5%)
Rarely or never	283	(24.0%)	224	(19.0%)

Survey of banking customers

- How many payment cards with a PIN do you use?($N = 1177$)

1	2	3	4
708 (60.2%)	344 (29.2%)	89 (7.6%)	23 (2.0%)

Median: 1, Mean: 1.5

- If you have more than one payment card which requires a PIN, do you use the same PIN for several cards?($N = 469$)

yes	no
161 (34.3%)	308 (65.7%)

Survey of banking customers

- *Have you ever changed the PIN associated with a payment card?(N = 1177)*

Never	Yes, initially	Yes, periodically
591 (50.2%)	376 (31.9%)	210 (17.8%)

- *Have you ever forgotten your PIN and had to have your financial institution remind you or reset your card?(N = 1177)*

yes	no
186 (15.8%)	991 (84.2%)

Survey of banking customers

- *Have you ever shared your PIN with another person so that they could borrow your payment card?(N = 1177)*

spouse or significant other	475	(40.4%)
child, parent, sibling, or other family member	204	(17.3%)
friend or acquaintance	40	(3.4%)
secretary or personal assistant	1	(0.1%)
any	621	(52.8%)

Survey of banking customers

- *Have you ever used a PIN from a payment card for something other than making a payment or retrieving money?(N = 1177)*

password for an Internet account	180	(15.3%)
password for my computer	94	(8.0%)
code for my voicemail	242	(20.6%)
to unlock the screen for mobile phone	104	(8.8%)
to unlock my SIM card	29	(2.5%)
entry code for a building	74	(6.3%)
any	399	(33.9%)

Modeling banking distribution using surveyed data

7. Does your PIN...

- ☐ include a straight horizontal line on the keypad
(for example, 3219 or 8456)?

- ☐ include a straight vertical line on the keypad
(for example, 0852 or 1476)?

- ☐ include a diagonal line on the keypad
(for example, 1597 or 7534)?

- ☐ use the 4 corners of the keypad
(for example, 1397 or 7319)?

- ☐ use the 4 points of the cross on the keypad
(for example, 2684 or 8246)?

- ☐ use 4 numbers which make a 2 by 2 box on the keypad
(for example, 7548 or 1254)?

- ☐ represent some other contiguous path on the keypad
(for example, 2698 or 1486)?

- ☐ spell a memorable word using the letters commonly written on the keypad
(for example, 5683 for L-O-V-E)?

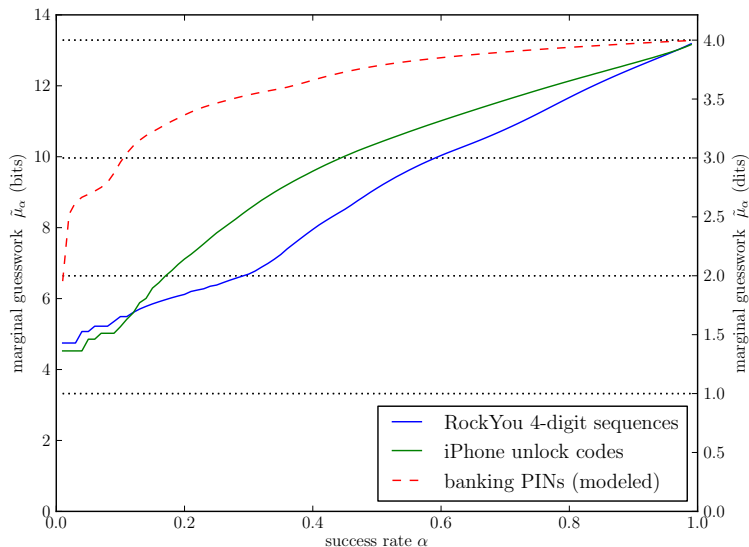
- ☐ represent another pattern on the keypad which is meaningful to you?

73% of respondents were willing to classify their PIN

Modeling banking distribution using surveyed data

factor	example	RockYou date	iPhone	surveyed
DDMM	2311	5.26	1.38	3.07
DMYY	3876	9.26	6.46	5.54
MMDD	1123	10.00	9.35	3.66
MMYY	0683	0.67	0.20	0.94
YYYY	1984	33.39	7.12	4.95
<i>total</i>		58.57	24.51	22.76
keypad				
adjacent	6351	1.52	4.99	—
box	1425	0.01	0.58	—
corners	9713	0.19	1.06	—
cross	8246	0.17	0.88	—
diagonal swipe	1590	0.10	1.36	—
horizontal swipe	5987	0.34	1.42	—
spelled word	5683	0.70	8.39	—
vertical swipe	8520	0.06	4.28	—
<i>total</i>		3.09	22.97	8.96
numeric				
ending in 69	6869	0.35	0.57	—
digits 0-3 only	2000	3.49	2.72	—
digits 0-6 only	5155	4.66	5.96	—
repeated pair	2525	2.31	4.11	—
repeated quad	6666	0.40	6.67	—
sequential down	3210	0.13	0.29	—
sequential up	4567	3.83	4.52	—
<i>total</i>		15.16	24.85	4.60
<i>random selection</i>	3271	23.17	27.67	63.68

Modeling banking distribution using surveyed data

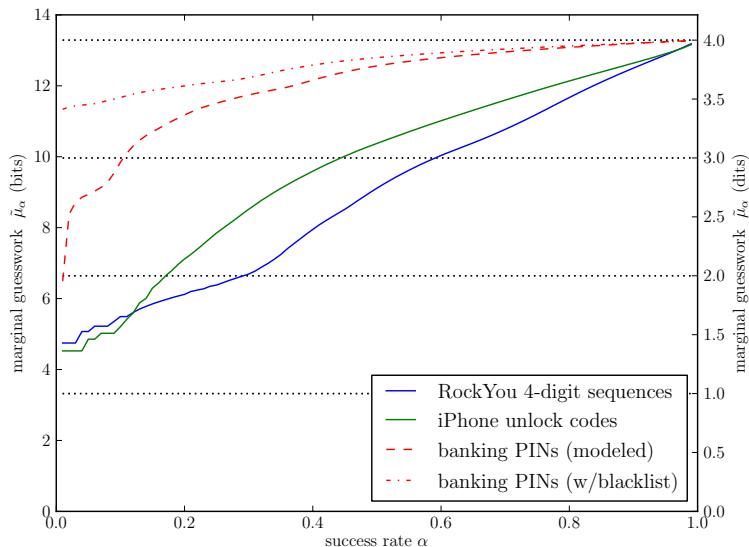


Modeling banking distribution using surveyed data

What if banks employed a blacklist?

0000, 0101-0103, 0110, 0111, 0123, 0202,
0303, 0404, 0505, 0606, 0707, 0808, 0909,
1010, 1101-1103, 1110-1112, 1123, 1201-1203,
1210-1212, 1234, 1956-2015, 2222, 2229, 2580,
3333, 4444, 5252, 5683, 6666, 7465, 7667

Modeling banking distribution using surveyed data



The cardinal sin of PIN selection



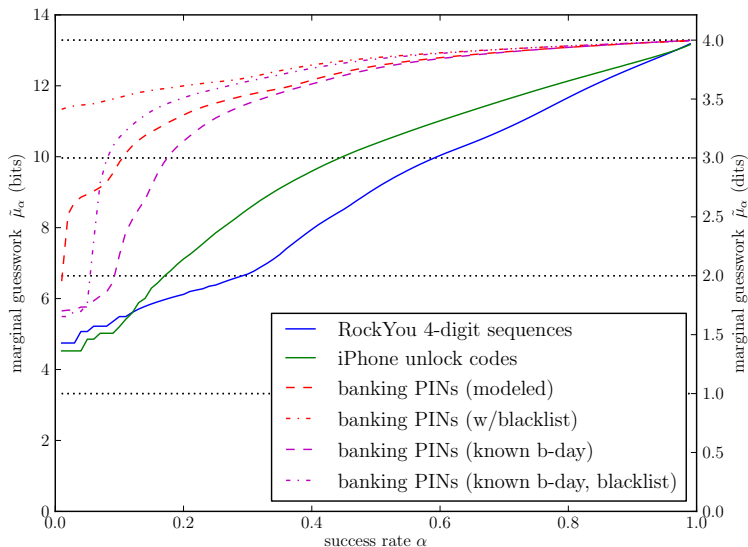
JESUS KEYS IN HIS PIN NUMBER

Courtesy of Chris Madden

The cardinal sin of PIN selection

- 7% of users use a variation of their own birthday as their PIN...
 - 22% YYYY
 - 19% DMYY
 - 18% MMDD
 - 14% DDMM
 - 12% DDYY
 - ...
- 99% of users indicate they carry their DOB in their wallet or purse!

The cardinal sin of PIN selection



Practical implications



Attackers can try at least 6 guesses (3 ATM, 3 CAP)

- General case: 1234, 1990, 1989, 1988, 1987, 1986
- Born 1983-06-03: 1983, 6383, 0306, 0603, 1234, 0383

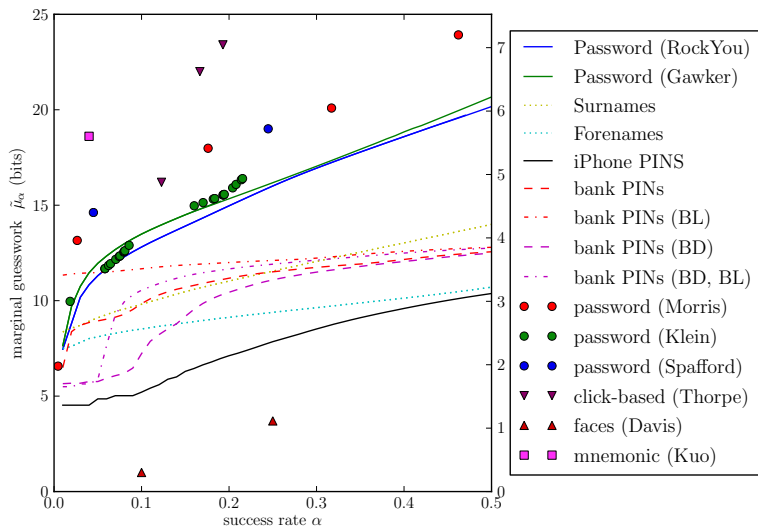
Practical implications

scenario	λ_3	λ_6
baseline	1.44%	1.94%
w/blacklist	0.12%	0.24%
known DOB	5.52%	8.23%
blackl., DOB	5.11%	5.63%
random PIN	0.03%	0.06%

Practical implications

scenario	number of stolen cards				exp.
	1	2	3	4	
baseline	1.9%	2.9%	3.9%	4.9%	2.5%
w/blacklist	0.2%	0.5%	0.7%	0.9%	0.4%
known DOB	8.2%	9.7%	10.3%	10.9%	8.9%
blackl., DOB	5.6%	6.0%	6.2%	6.4%	5.8%
random PIN	0.1%	0.1%	0.2%	0.2%	0.1%

Putting PINs into context



How did we get stuck with PINs?



BARCLAYCASH system, late 1960's

Thank you

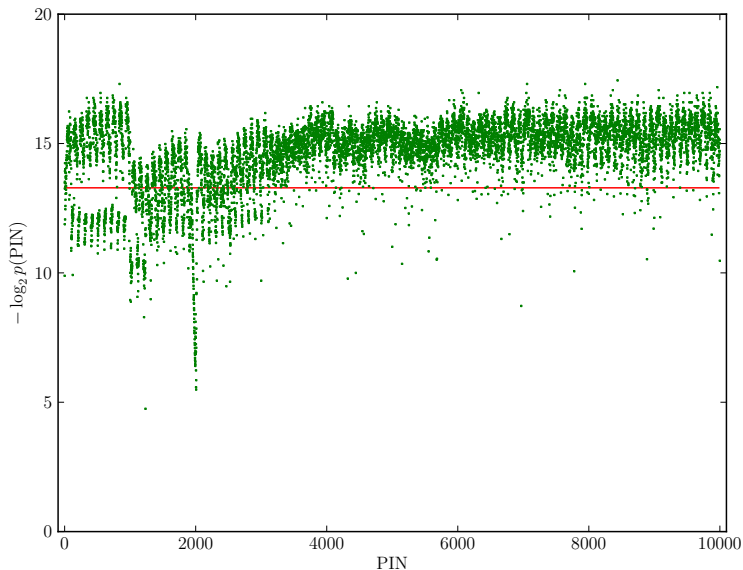
jcb82@cl.cam.ac.uk

Full guessing statistics

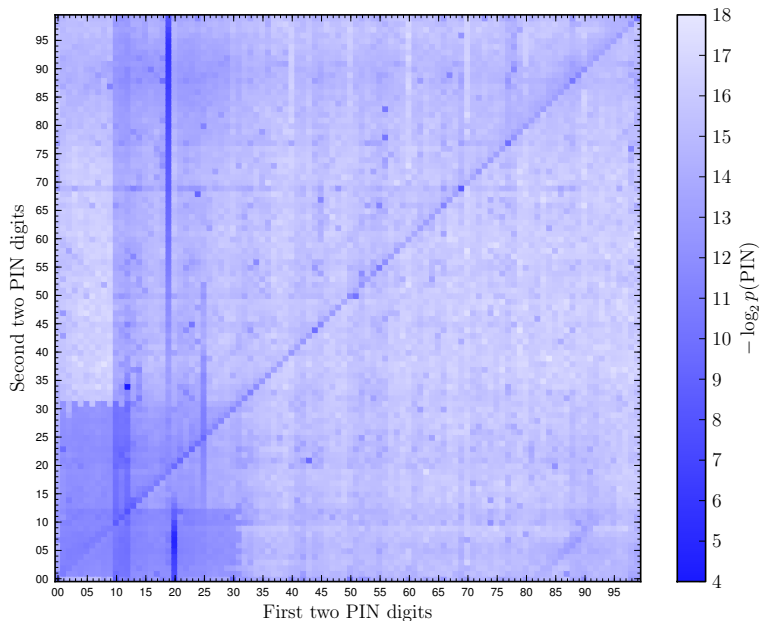
	H_1	$\tilde{G}$	$\tilde{\mu}_{0.5}$	λ_3	λ_6
RockYou	10.74	11.50	9.11	8.04%	12.29%
RockYou M	11.01	11.79	9.39	5.06%	7.24%
iPhone	11.42	11.83	10.37	9.23%	12.39%
iPhone M	11.70	12.06	10.73	9.21%	11.74%
baseline	12.90	12.83	12.56	1.44%	1.94%
w/blacklist	13.13	12.95	12.79	0.12%	0.24%
DOB	12.57	12.80	12.49	5.52%	8.23%
DOB, BL	12.85	12.92	12.75	5.11%	5.63%
random	13.29	13.29	13.29	0.03%	0.06%

- H_1 = Shannon entropy
- $\tilde{G}$ = Guesswork (bit-converted)
- $\tilde{\mu}_{0.5}$ = Marginal guesswork (bit-converted)
- λ_β = % of accounts covered by β optimal guesses

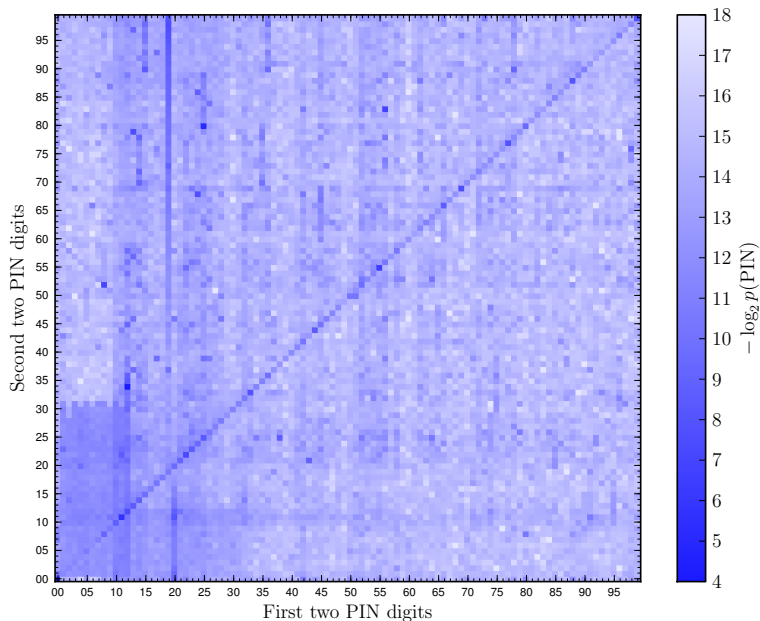
Other PIN visualisations (RockYou)



Other PIN visualisations (RockYou)



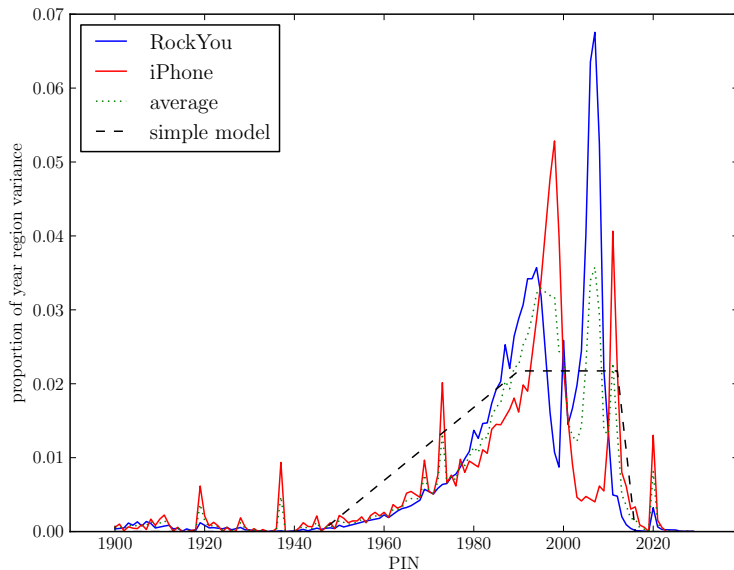
Other PIN visualisations (iPhone)



Regression details...

- need to avoid *omitted variable bias*
 - singleton functions added: 0000, 1111, 1234, 2580
 - intentionally weakened model of years
- non-binary functions:
 - years
 - keypad words
 - February 29th

Regression details...



Regression details...



Keypad entry of love

Regression details...

love	2643
pink	747
poop	644
baby	616
sexy	529
alex	398
star	373
mike	354
blue	311
ryan	291
josh	277
nick	273
lala	270
pimp	257
john	252

four letter passwords, RockYou

Regression details...

5683	2655	love, loud
7465	748	pink
2229	735	baby, abby
7667	652	poop, poms
7399	541	sexy, rexy
6453	435	mike, nike, milf, mile
2539	405	alex, blew
7827	375	star
5252	331	lala, jaja, kaka, kala
2583	318	blue, clue
5674	316	josh, lori, kori, jori
7926	297	ryan, swan
7467	289	pimp, shop, sims, rios
3825	288	fuck, duck
6425	285	nick, mick

model for word-based PINs

Survey of banking customers

- *Do you regularly use a PIN number with your payment cards?* ($N = 1337$)

yes, a 4-digit PIN	yes, a PIN of 5+ digits	no
1108 (82.9%)	69 (5.2%)	160 (12.0%)

Survey of banking customers

- *When making purchases in a shop, how do you typically pay?* (N = 1177)

I use my payment card and key in my PIN	477 (40.5%)
I use my payment card and sign a receipt	357 (30.3%)
I use my payment card with my PIN or my signature equally often	184 (15.6%)
I normally use cash or cheque payments and rarely use payment cards	159 (13.5%)