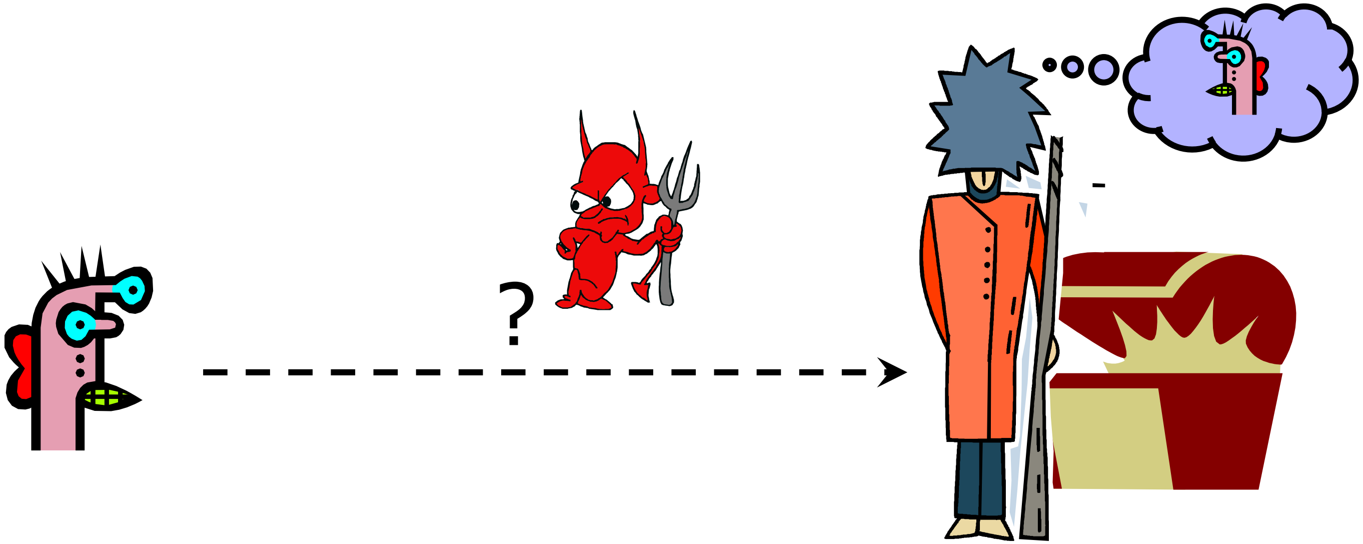
CS 5435

Authentication

Vitaly Shmatikov

Basic Problem



How do you prove to someone that you are who you claim to be?

Any system with access control must solve this problem

Many Ways to Prove Who You Are

What you know

- Passwords
- Answers to questions that only you know

Where you are

- IP address, geolocation

What you are

- Biometrics

What you have

- Secure tokens, mobile devices

Password-Based Authentication

User has a secret password.

System checks it to authenticate the user.

How is the password communicated?

- Eavesdropping risk

How is the password stored?

- In the clear? Encrypted? Hashed?

How does the system check the password?

How easy is it to guess the password?

- Easy-to-remember passwords tend to be easy to guess

Passwords and Computer Security

Phishing and use of stolen credentials are the top two hacking techniques

- Source: Verizon Data Breach Investigations Report

First step after any successful intrusion: install sniffer or keylogger to steal more passwords

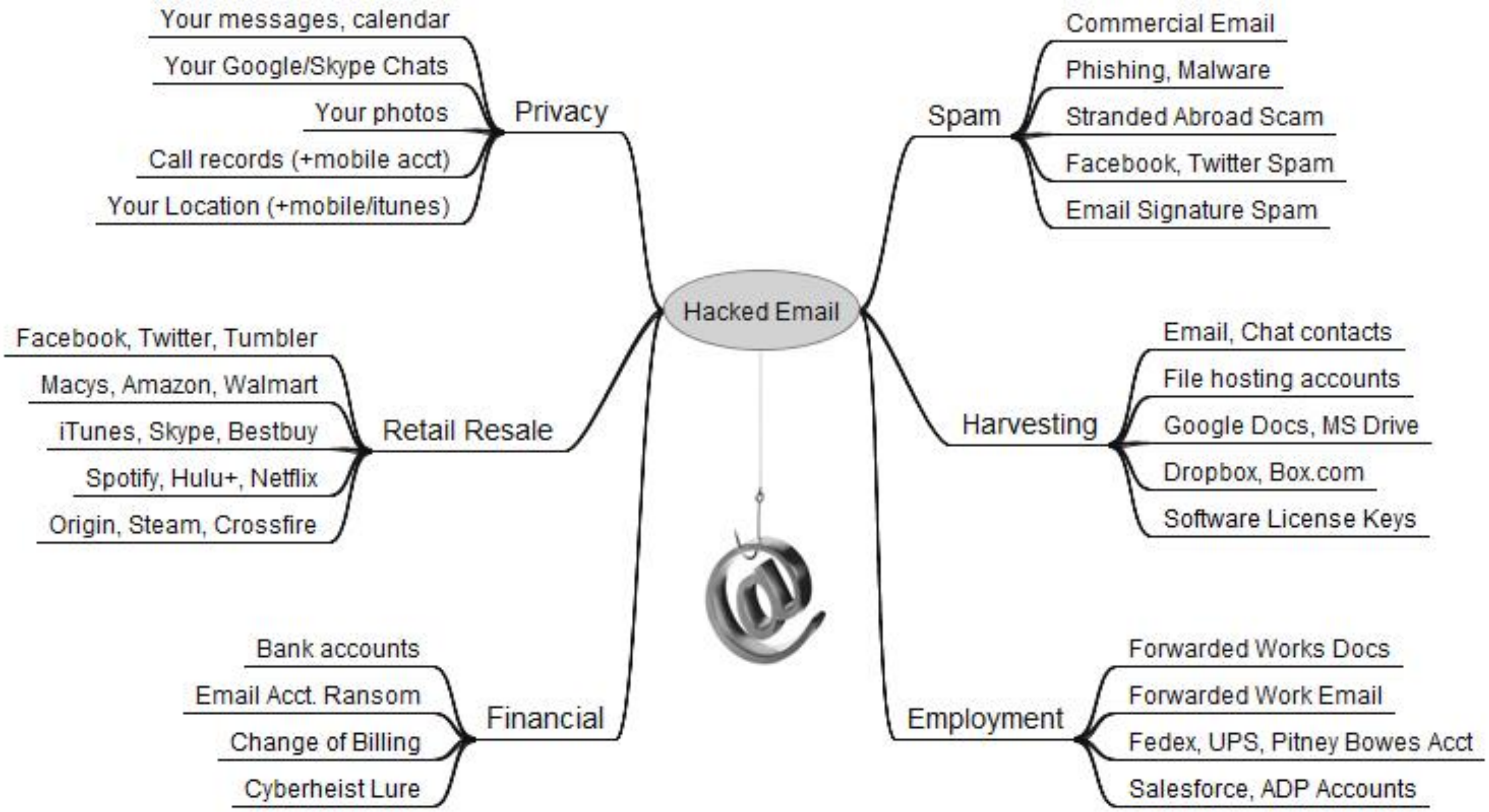
Second step: run cracking tools on password files

- Cracking needed because modern systems usually do not store passwords in the clear (how are they stored?)

In Mitnick's "Art of Intrusion", 8 out of 9 exploits involve password stealing and/or cracking

From Here to Eternity

[redacted] https://krebsonsecurity.com/2019/01/773m-password-megabreach-is-years-old/



- Hacked Email
  - Privacy
    - Your messages, calendar
    - Your Google/Skype Chats
    - Your photos
    - Call records (+mobile acct)
    - Your Location (+mobile/itunes)
  - Retail Resale
    - Facebook, Twitter, Tumbler
    - Macys, Amazon, Walmart
    - iTunes, Skype, Bestbuy
    - Spotify, Hulu+, Netflix
    - Origin, Steam, Crossfire
  - Financial
    - Bank accounts
    - Email Acct. Ransom
    - Change of Billing
    - Cyberheist Lure
  - Spam
    - Commercial Email
    - Phishing, Malware
    - Stranded Abroad Scam
    - Facebook, Twitter Spam
    - Email Signature Spam
  - Harvesting
    - Email, Chat contacts
    - File hosting accounts
    - Google Docs, MS Drive
    - Dropbox, Box.com
    - Software License Keys
  - Employment
    - Forwarded Works Docs
    - Forwarded Work Email
    - Fedex, UPS, Pitney Bowes Acct
    - Salesforce, ADP Accounts

Default Passwords

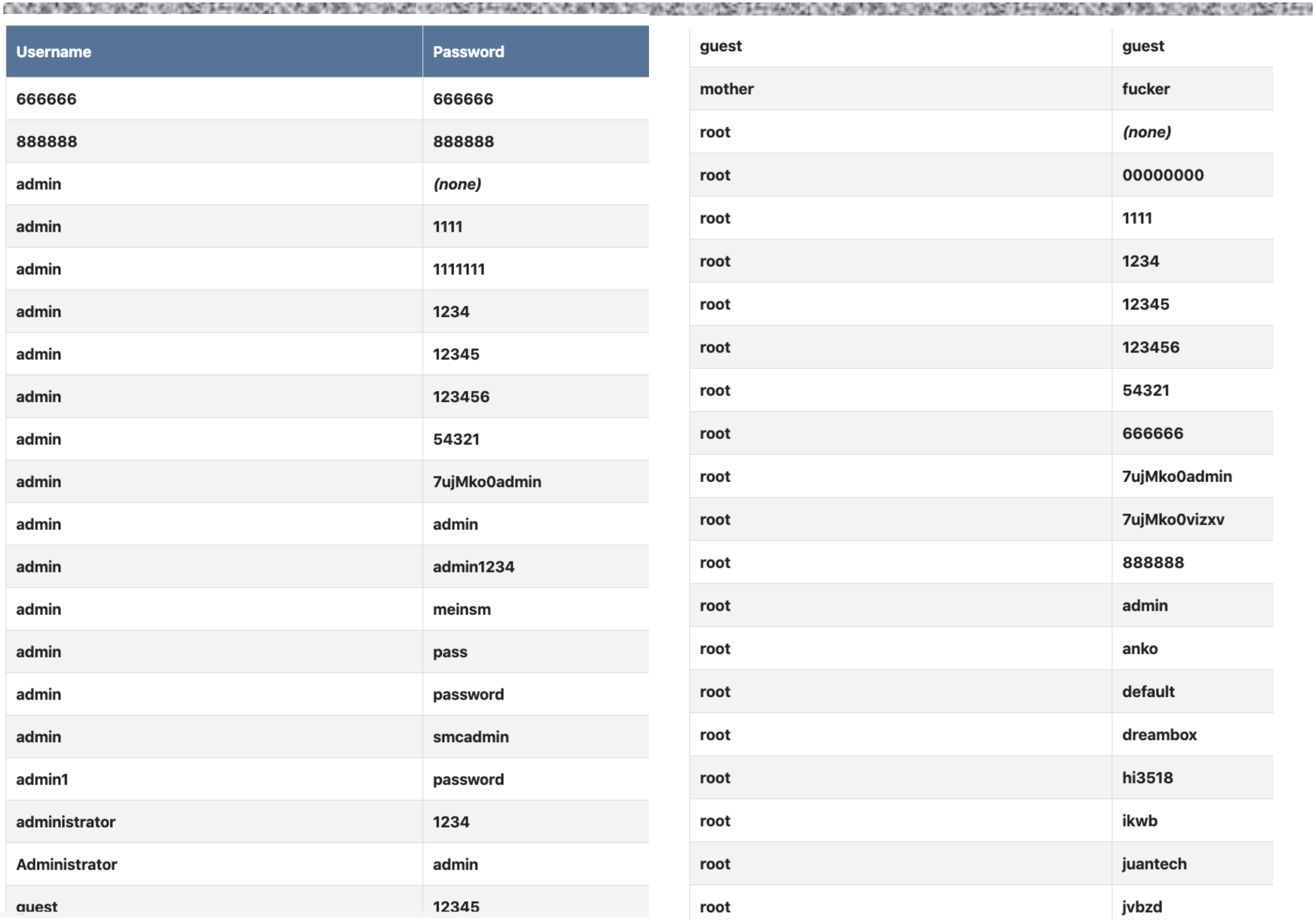
Examples from Mitnick's "Art of Intrusion"

- U.S. District Courthouse server: "public" / "public"
- NY Times employee database: pwd = last 4 SSN digits
- "Dixie bank": break into router (pwd="administrator"), then into IBM AS/400 server (pwd="administrator"), install keylogger to snarf other passwords
  - "99% of people there used password123 as their password"

Mirai botnet (2016)

- Used default passwords in IoT devices (Internet cameras, home routers, etc.) to stage a massive distributed denial-of-service flooding attack

From Mirai’s Source Code



Username	Password
666666	666666
888888	888888
admin	*(none)*
admin	1111
admin	1111111
admin	1234
admin	12345
admin	123456
admin	54321
admin	7ujMko0admin
admin	admin
admin	admin1234
admin	meinsm
admin	pass
admin	password
admin	smcadmin
admin1	password
administrator	1234
Administrator	admin
guest	12345
guest	guest
mother	fucker
root	*(none)*
root	00000000
root	1111
root	1234
root	12345
root	123456
root	54321
root	666666
root	7ujMko0admin
root	7ujMko0vizxv
root	888888
root	admin
root	anko
root	default
root	dreambox
root	hi3518
root	ikwb
root	juantech
root	jvbzd

Old Password Surveys

Klein (1990) and Spafford (1992)

- 2.7% guessed in 15 minutes, 21% in a week
- Much more computing power is available now!

U. of Michigan: 5% of passwords were "goblue"

- How many passwords in this university involve "cornell", "big red", etc.?

Zviran and Haga (1999)

- Password usage at a DoD facility in California
- 80% of passwords were 4-7 characters in length, 80% used alphabetic characters only, 80% of the users had never changed their password

rockyou™ Hack (2009)

“Social gaming” company

Database with 32 million user passwords from partner social networks

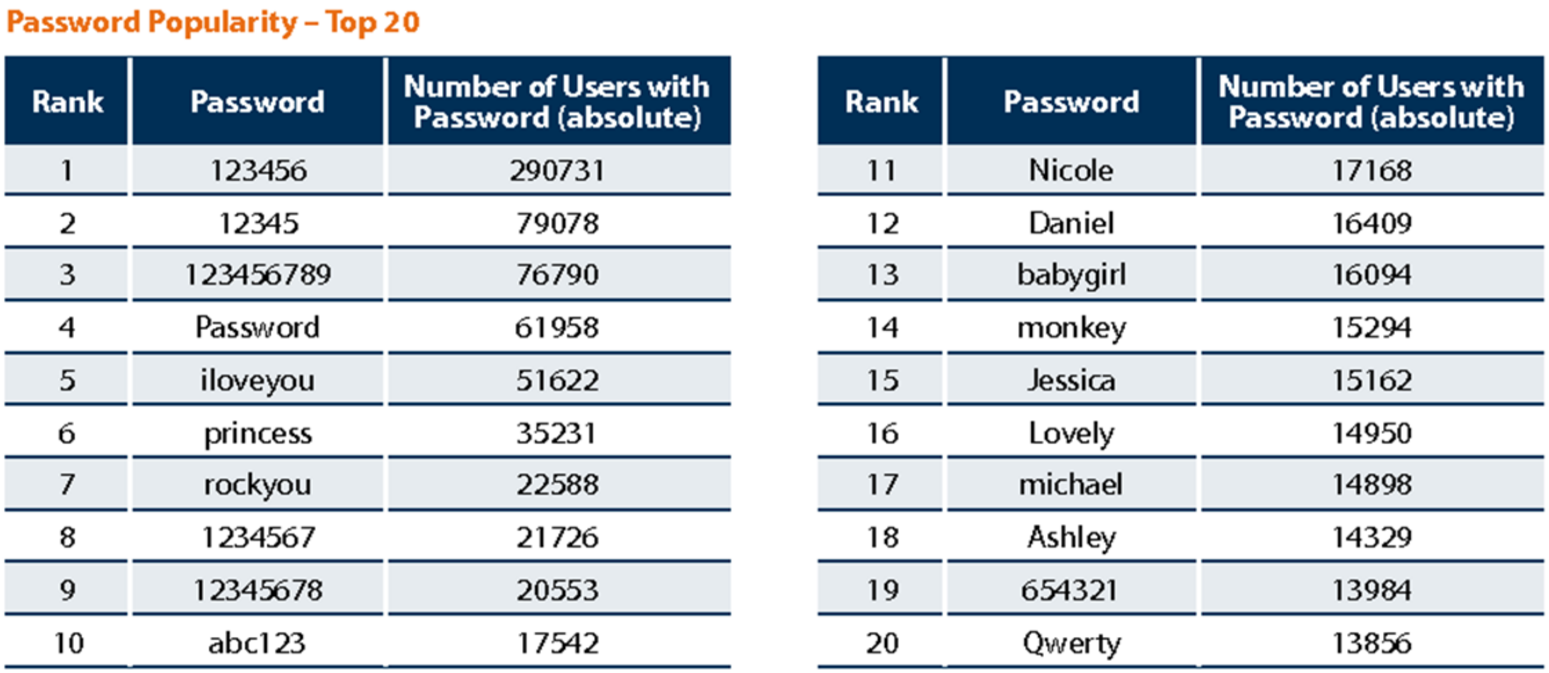
Passwords stored in the clear

December 2009: entire database hacked using a SQL injection attack and posted on the Internet

- More about SQL injection attacks later

Passwords in RockYou Database

[Imperva]

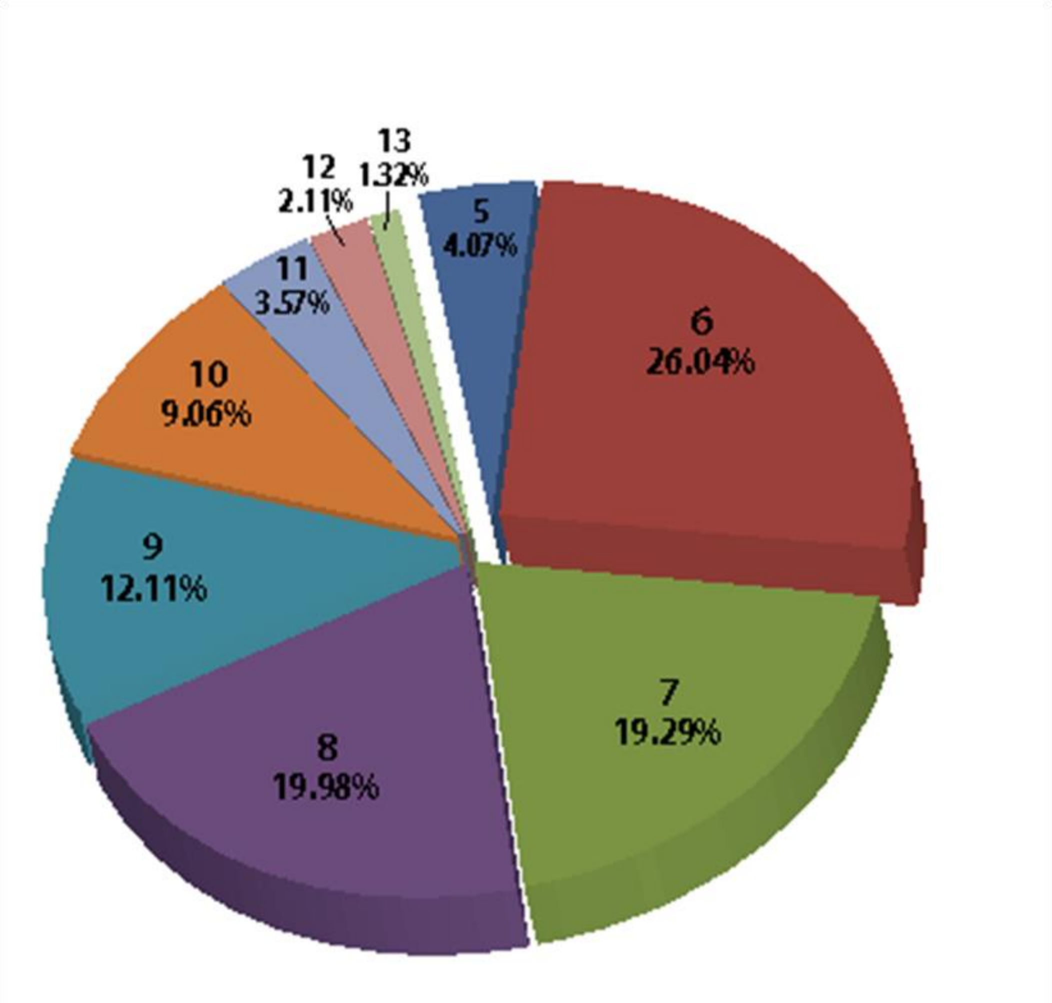


Password Popularity – Top 20

Rank	Password	Number of Users with Password (absolute)
1	123456	290731
2	12345	79078
3	123456789	76790
4	Password	61958
5	iloveyou	51622
6	princess	35231
7	rockyou	22588
8	1234567	21726
9	12345678	20553
10	abc123	17542
11	Nicole	17168
12	Daniel	16409
13	babygirl	16094
14	monkey	15294
15	Jessica	15162
16	Lovely	14950
17	michael	14898
18	Ashley	14329
19	654321	13984
20	Qwerty	13856

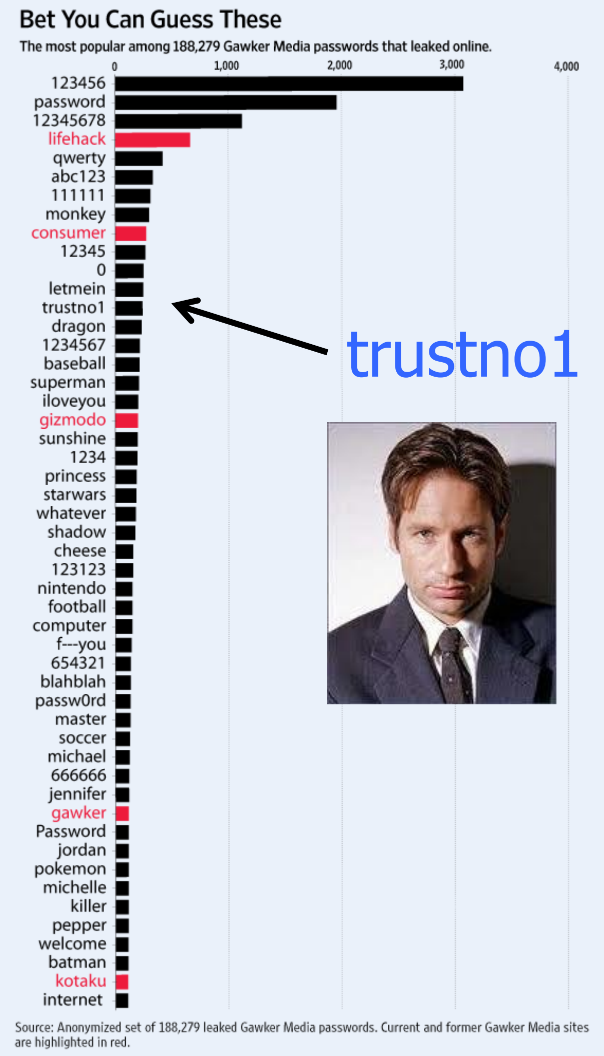
Password Length Distribution

[Imperva]



Gawker Passwords (2010)

[WSJ]

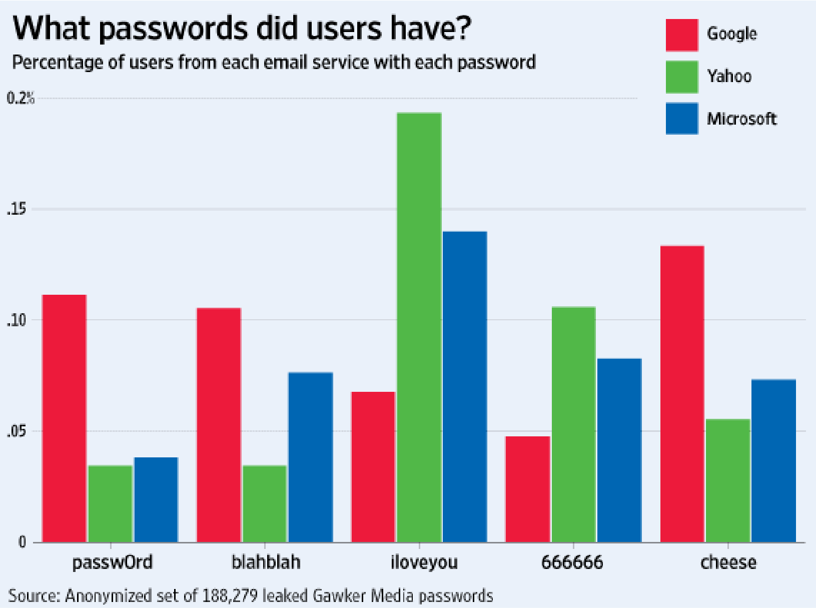


Bet You Can Guess These

The most popular among 188,279 Gawker Media passwords that leaked online.

trustno1

Source: Anonymized set of 188,279 leaked Gawker Media passwords. Current and former Gawker Media sites are highlighted in red.



What passwords did users have?

Percentage of users from each email service with each password

Source: Anonymized set of 188,279 leaked Gawker Media passwords

More Password Datasets

PlentyOfFish®
Free Online Dating

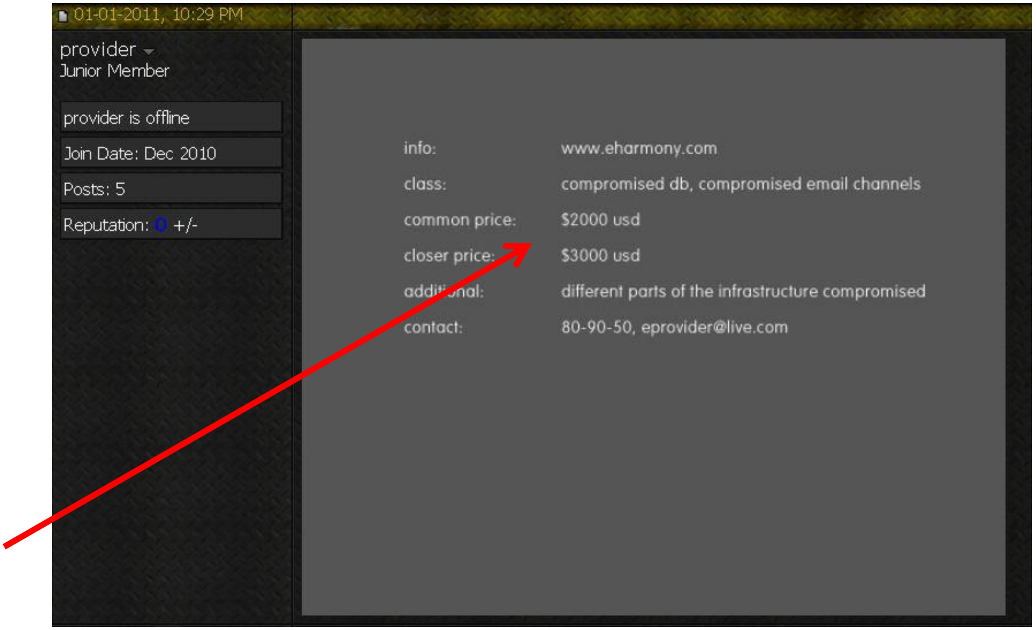
More than 30 million passwords

eHarmony®

"#1 Most Trusted Online Dating Site"

SQL injection attack

For sale for $3000



01-01-2011, 10:29 PM

provider
Junior Member

provider is offline

Join Date: Dec 2010

Posts: 5

Reputation: 0 +/-

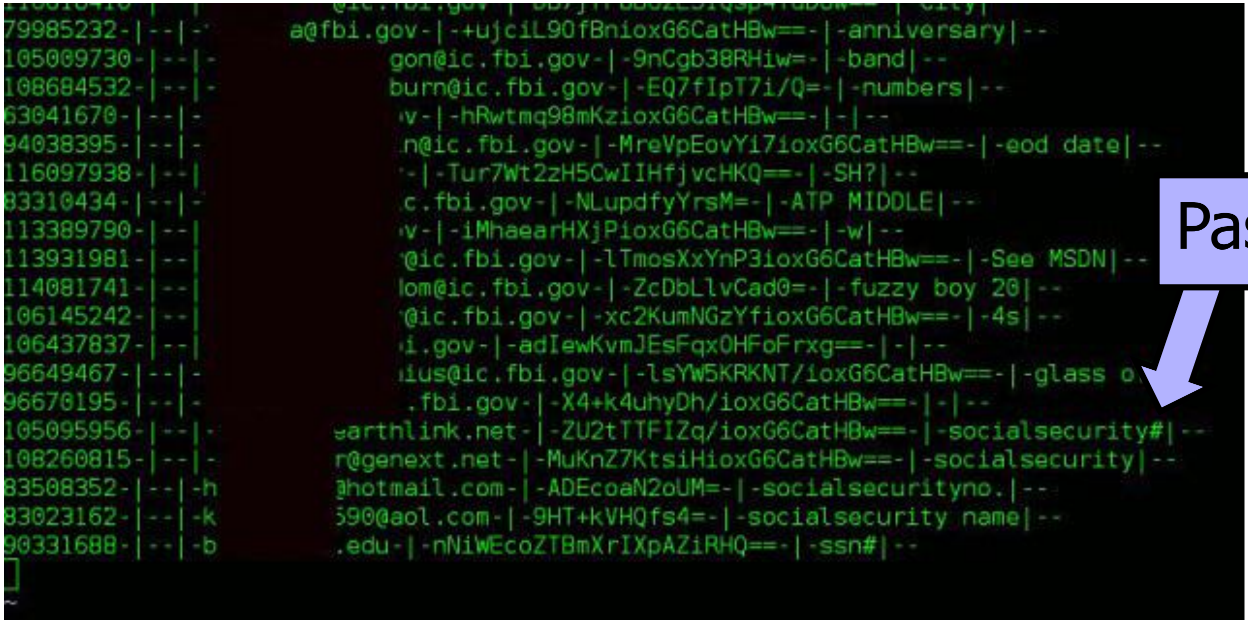
info:	www.eharmony.com
class:	compromised db, compromised email channels
common price:	$2000 usd
closer price:	$3000 usd
additional:	different parts of the infrastructure compromised
contact:	80-90-50, eprovider@live.com

Adobe Passwords (2013)

153 million account passwords

- 56 million of them unique

Encrypted using 3DES in ECB mode rather than hashed (why is this important?)



```
79985232-|--|-  [redacted]  a@fbi.gov-|-+ujciL90fBnioxG6CatHBw==-|-anniversary|--
105009730-|--|-        gon@ic.fbi.gov-|-9nCgb38RHiw=-|-band|--
108684532-|--|-        burn@ic.fbi.gov-|-EQ7fIpT7i/Q=-|-numbers|--
63041670-|--|-          v-|-hRwtmq98mKzioxG6CatHBw==-|-|--
94038395-|--|-          n@ic.fbi.gov-|-MreVpEovYi7ioxG6CatHBw==-|-eod date|--
116097938-|--|          -|-Tur7Wt2zH5CwIIHfjvcHKQ==-|-SH?|--
83310434-|--|-          c.fbi.gov-|-NLupdfyYrsM=-|-ATP MIDDLE|--
113389790-|--|          v-|-iMhaearHXjPioxG6CatHBw==-|-w|--
113931981-|--|          @ic.fbi.gov-|-lTmosXxYnP3ioxG6CatHBw==-|-See MSDN|--
114081741-|--|          lom@ic.fbi.gov-|-ZcDbLlvCad0=-|-fuzzy boy 20|--
106145242-|--|          @ic.fbi.gov-|-xc2KumNGzYfioxG6CatHBw==-|-4s|--
106437837-|--|          i.gov-|-adIewKvmJEsFqxOHFoFrxg==-|-|--
96649467-|--|-          ius@ic.fbi.gov-|-lsYW5KRKNT/ioxG6CatHBw==-|-glass o
96670195-|--|-           .fbi.gov-|-X4+k4uhyDh/ioxG6CatHBw==-|-|--
105095956-|--|-      earthlink.net-|-ZU2tTTFIZq/ioxG6CatHBw==-|-socialsecurity#|--
108260815-|--|-      r@genext.net-|-MuKnZ7KtsiHioxG6CatHBw==-|-socialsecurity|--
83508352-|--|-h      @hotmail.com-|-ADEcoaN2oUM=-|-socialsecurityno.|--
83023162-|--|-k      590@aol.com-|-9HT+kVHQfs4=-|-socialsecurity name|--
90331688-|--|-b      .edu-|-nNiWEcoZTBmXrIXpAZiRHQ==-|-ssn#|--
```

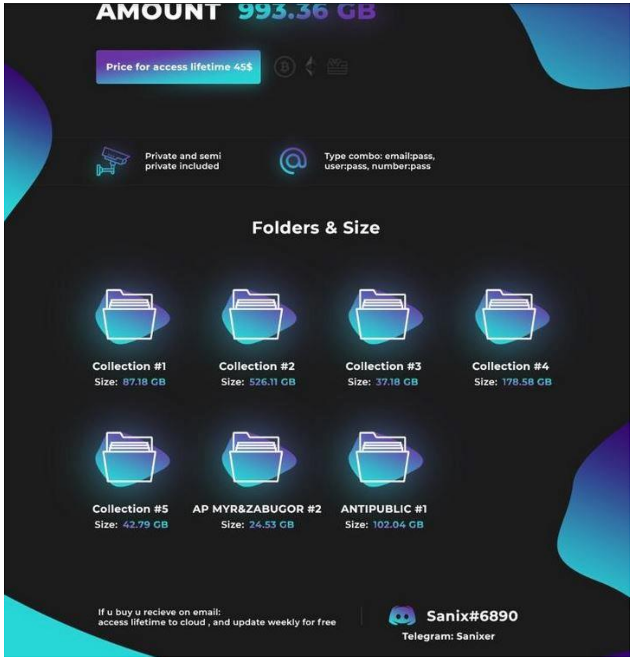
Password hints

“Collection #1” (2018-2019)

[redacted] https://krebsonsecurity.com/2019/01/773m-password-megabreach-is-years-old/

Mother of All Breaches Exposes 773 Million Emails, 21 Million Passwords

… just a subset of the seller’s offerings



AMOUNT 993.36 GB

Price for access lifetime 45$

Private and semi private included

Type combo: email:pass, user:pass, number:pass

Folders & Size

Collection #1
Size: 87.18 GB

Collection #2
Size: 526.11 GB

Collection #3
Size: 37.18 GB

Collection #4
Size: 178.58 GB

Collection #5
Size: 42.79 GB

AP MYR&ZABUGOR #2
Size: 24.53 GB

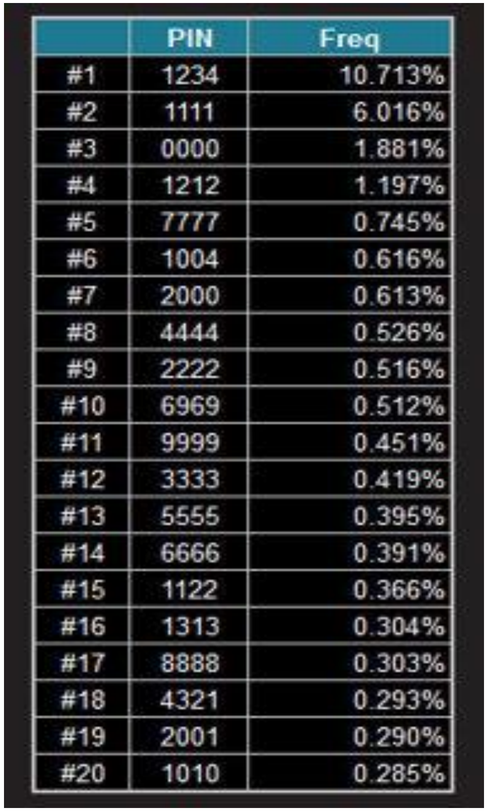
ANTIPUBLIC #1
Size: 102.04 GB

If u buy u recieve on email:
access lifetime to cloud , and update weekly for free

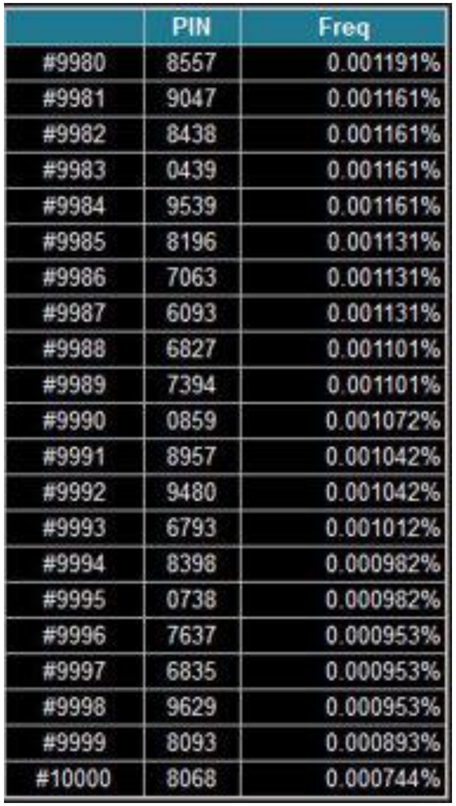
Sanix#6890
Telegram: Sanixer

How About PINs?

In 2012, Nick Berry analyzed all four-digit passwords from previous leaks

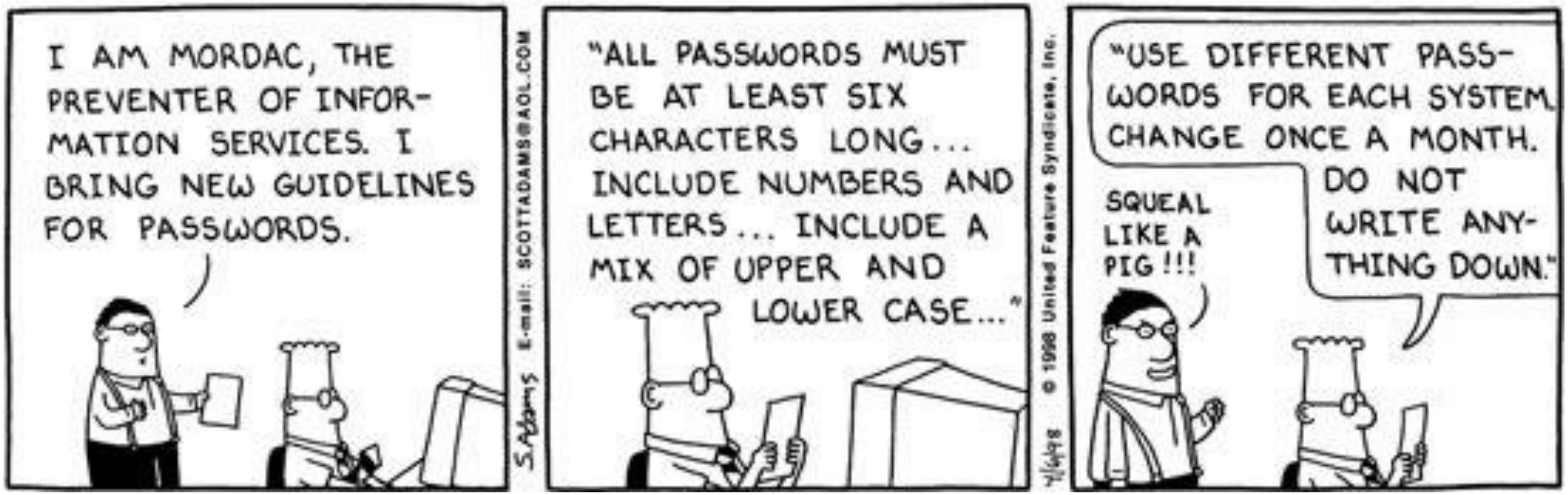


	PIN	Freq
#1	1234	10.713%
#2	1111	6.016%
#3	0000	1.881%
#4	1212	1.197%
#5	7777	0.745%
#6	1004	0.616%
#7	2000	0.613%
#8	4444	0.526%
#9	2222	0.516%
#10	6969	0.512%
#11	9999	0.451%
#12	3333	0.419%
#13	5555	0.395%
#14	6666	0.391%
#15	1122	0.366%
#16	1313	0.304%
#17	8888	0.303%
#18	4321	0.293%
#19	2001	0.290%
#20	1010	0.285%



	PIN	Freq
#9980	8557	0.001191%
#9981	9047	0.001161%
#9982	8438	0.001161%
#9983	0439	0.001161%
#9984	9539	0.001161%
#9985	8196	0.001131%
#9986	7063	0.001131%
#9987	6093	0.001131%
#9988	6827	0.001101%
#9989	7394	0.001101%
#9990	0859	0.001072%
#9991	8957	0.001042%
#9992	9480	0.001042%
#9993	6793	0.001012%
#9994	8398	0.000982%
#9995	0738	0.000982%
#9996	7637	0.000953%
#9997	6835	0.000953%
#9998	9629	0.000953%
#9999	8093	0.000893%
#10000	8068	0.000744%

Password Usability

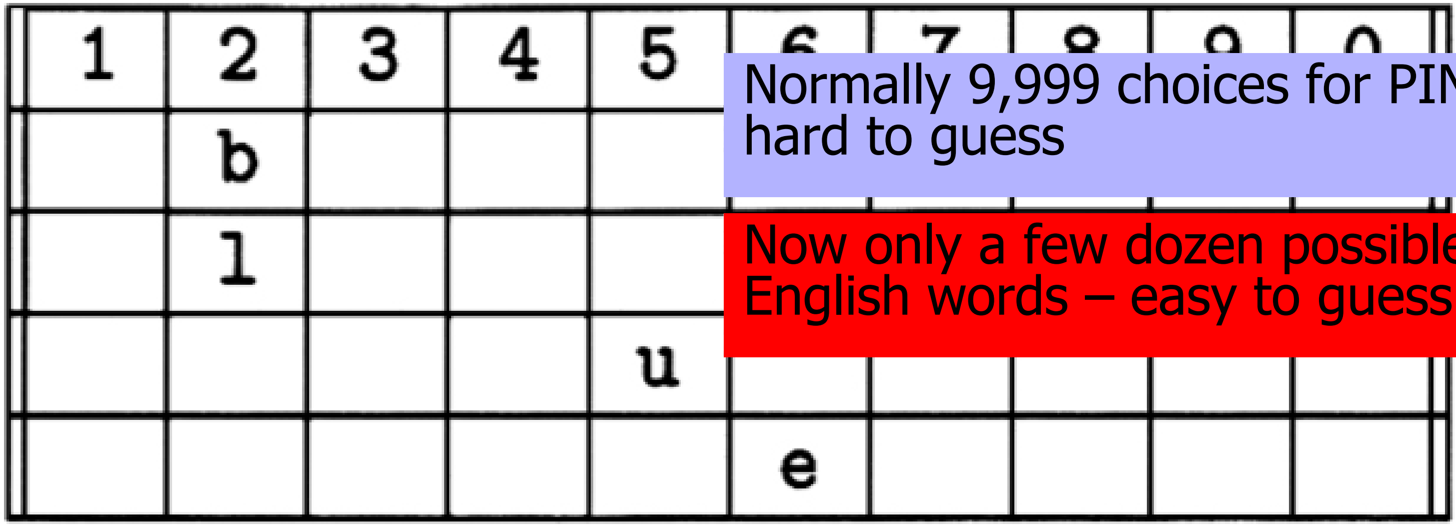

I AM MORDAC, THE PREVENTER OF INFORMATION SERVICES. I BRING NEW GUIDELINES FOR PASSWORDS.
"ALL PASSWORDS MUST BE AT LEAST SIX CHARACTERS LONG... INCLUDE NUMBERS AND LETTERS... INCLUDE A MIX OF UPPER AND LOWER CASE..."
"USE DIFFERENT PASSWORDS FOR EACH SYSTEM. CHANGE ONCE A MONTH. DO NOT WRITE ANYTHING DOWN."
SQUEAL LIKE A PIG!!!
S.Adams E-mail: SCOTTADAMS@AOL.COM
© 1998 United Feature Syndicate, Inc.

Memorability vs. Security

[Ross Anderson]

One bank's idea for making PINs "memorable"

- If PIN is 2256, write your favorite word in the grid



1	2	3	4	5	6	7	8	9	0
	b								
	l								
				u					
					e				

Normally 9,999 choices for PIN hard to guess

Now only a few dozen possible English words – easy to guess!

- Fill the rest with random letters

Password Guessing Techniques

Dictionary with words spelled backwards

First and last names, streets, cities

Same with upper-case initials

All valid license plate numbers in your state

Room numbers, telephone numbers, etc.

Letter substitutions and other tricks

If you can think of it, attacker will, too

Social Engineering

Univ. of Sydney study (1996)

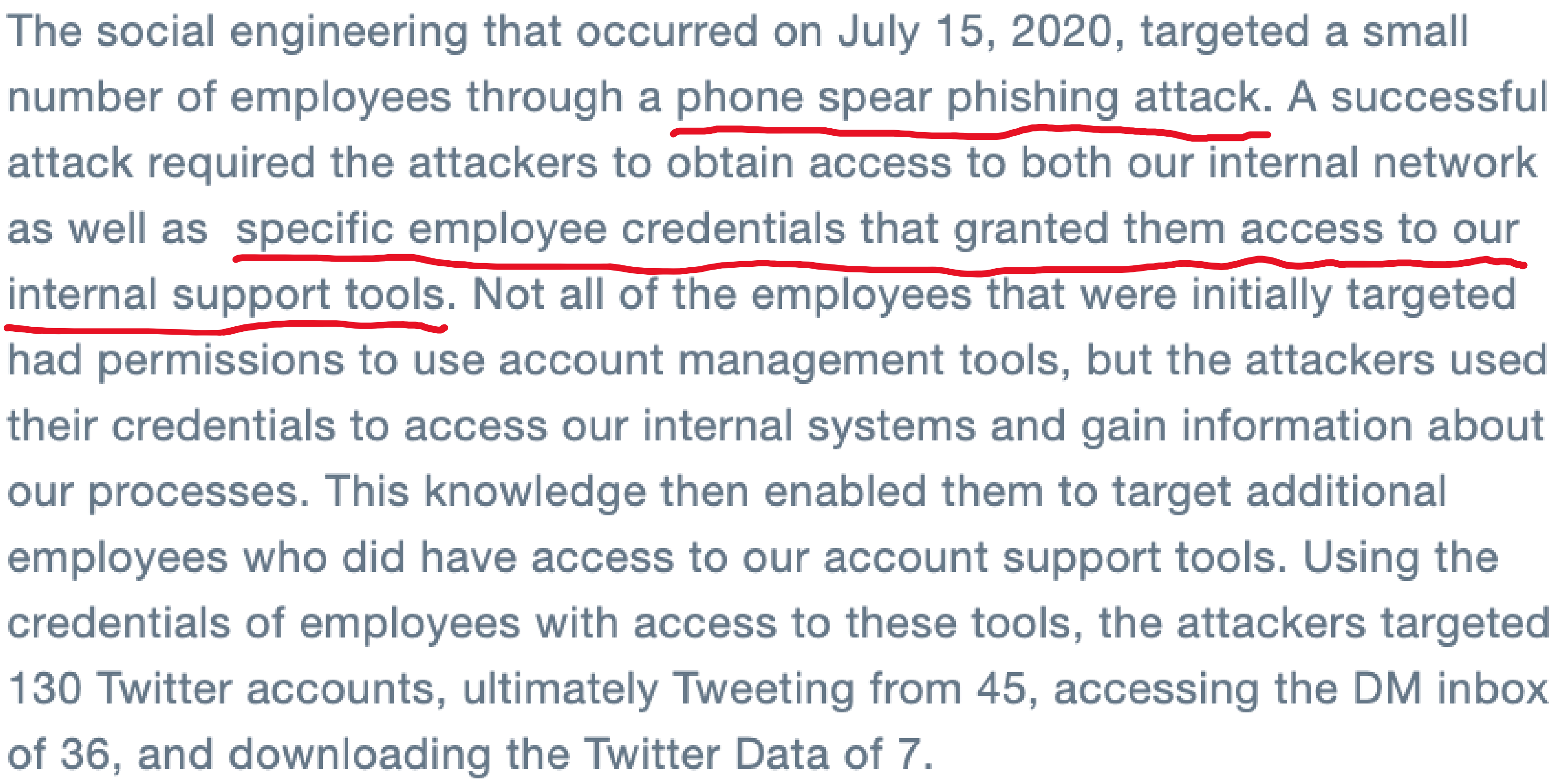
- 336 CS students emailed asking for their passwords
  - Pretext: “validate” password database after suspected break-in
- 138 returned their passwords; 30 returned invalid passwords; 200 reset passwords (not disjoint)

Treasury Dept. report (2005)

- Auditors pose as IT personnel attempting to correct a “network problem”
- 35 of 100 IRS managers and employees provide their usernames and change passwords to a known value

July 2020 Twitter Hack

https://blog.twitter.com/en_us/topics/company/2020/an-update-on-our-security-incident.html



The social engineering that occurred on July 15, 2020, targeted a small number of employees through a phone spear phishing attack. A successful attack required the attackers to obtain access to both our internal network as well as specific employee credentials that granted them access to our internal support tools. Not all of the employees that were initially targeted had permissions to use account management tools, but the attackers used their credentials to access our internal systems and gain information about our processes. This knowledge then enabled them to target additional employees who did have access to our account support tools. Using the credentials of employees with access to these tools, the attackers targeted 130 Twitter accounts, ultimately Tweeting from 45, accessing the DM inbox of 36, and downloading the Twitter Data of 7.

How People Use Passwords

Write them down

Use a single password at multiple sites

- Do you use the same password for Amazon and your bank account? NetID? Do you remember them all?

Forget them… many services use “security questions” to reset passwords

- “What is your favorite pet’s name?”
- Paris Hilton’s T-Mobile cellphone hack

Sara Palin's Email Hack

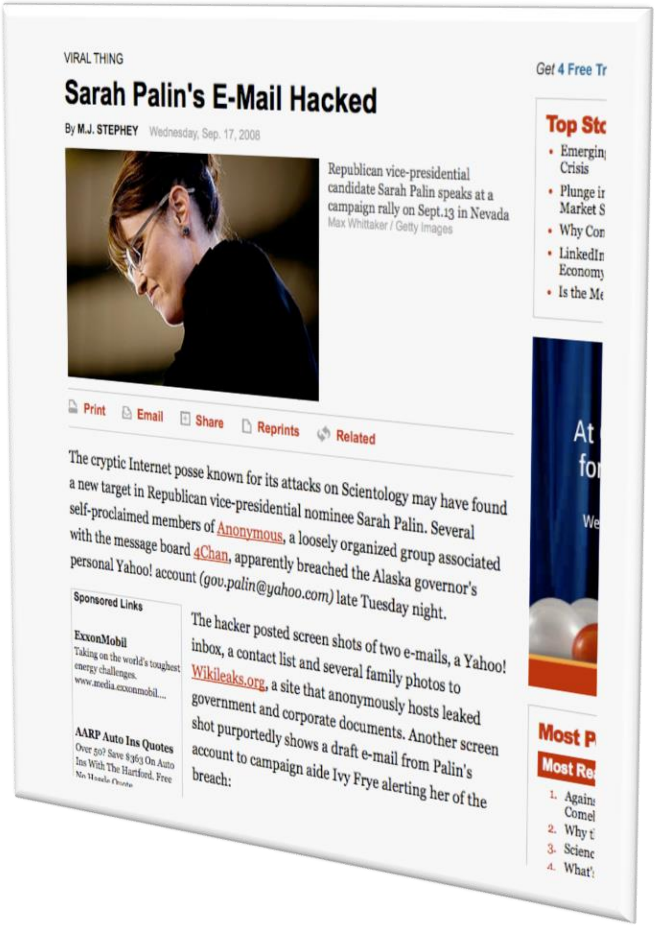
[slide: Gustav Rydstedt]

Reset password for gov.palin@yahoo.com

- No secondary email needed
- Date of birth? Wikipedia
- ZIP code? Wasilla has 2
- Where did you meet your spouse? Wikipedia, Google, …

Changed pwd to "popcorn"

Hacker sentenced to 1 year in prison + 3 yrs of supervised release



VIRAL THING

Get 4 Free Tr

Sarah Palin's E-Mail Hacked

By M.J. STEPHEY Wednesday, Sep. 17, 2008

Republican vice-presidential candidate Sarah Palin speaks at a campaign rally on Sept.13 in Nevada
Max Whittaker / Getty Images

Print Email Share Reprints Related

The cryptic Internet posse known for its attacks on Scientology may have found a new target in Republican vice-presidential nominee Sarah Palin. Several self-proclaimed members of Anonymous, a loosely organized group associated with the message board 4Chan, apparently breached the Alaska governor's personal Yahoo! account *(gov.palin@yahoo.com)* late Tuesday night.

The hacker posted screen shots of two e-mails, a Yahoo! inbox, a contact list and several family photos to Wikileaks.org, a site that anonymously hosts leaked government and corporate documents. Another screen shot purportedly shows a draft e-mail from Palin's account to campaign aide Ivy Frye alerting her of the breach:

Top Stc
- Emergin Crisis
- Plunge ir Market S
- Why Con
- LinkedIr Economy
- Is the Me

At for We

Most P

Most Re

1. Agains Comel
2. Why t
3. Scienc
4. What'

Problems with Security Questions

[redacted] [Rabkin, “Security questions in the era of Facebook”]

Inapplicable

- What high school did your spouse attend?

Not memorable

- Name of kindergarten teacher? Price of your first car?

Ambiguous

- Name of college you applied to but did not attend?

Easily guessable

- Age when you married? Year you met your spouse? Favorite president? Favorite color?

Automatically attackable (using public records!)

Answers Are Easy to Find Out…

Make of your first car?

- Until 1998, Ford had >25% of market

First name of your best friend?

- 10% of males: James/Jim, John, Robert/Bob/Rob

Name of your first / favorite pet?

- Max, Jake, Buddy, Bear…
- Top 500 (covers 65% of names) available online

Information available from Facebook, etc.

- Where you went to school, college athletic rivals, favorite book/movie/pastime, high school mascot

…or Easy to Forget

Name of the street, etc.

- More than one

Name of best friend

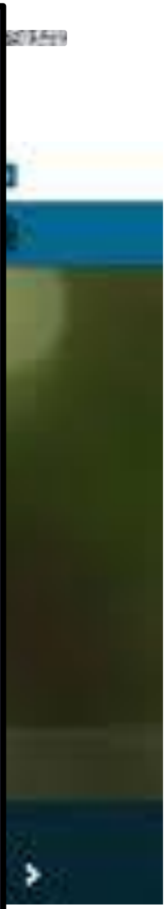
- Friends change

City where you were born?

- NYC? New York? Manhattan? New York City? Big Apple?

People lie to increase security… then forget the answers

HealthCare.gov



Federal:

- What is a relative's telephone number that is not your own?
- Type a significant date in your life?
- What is the name of the manager at your first job?

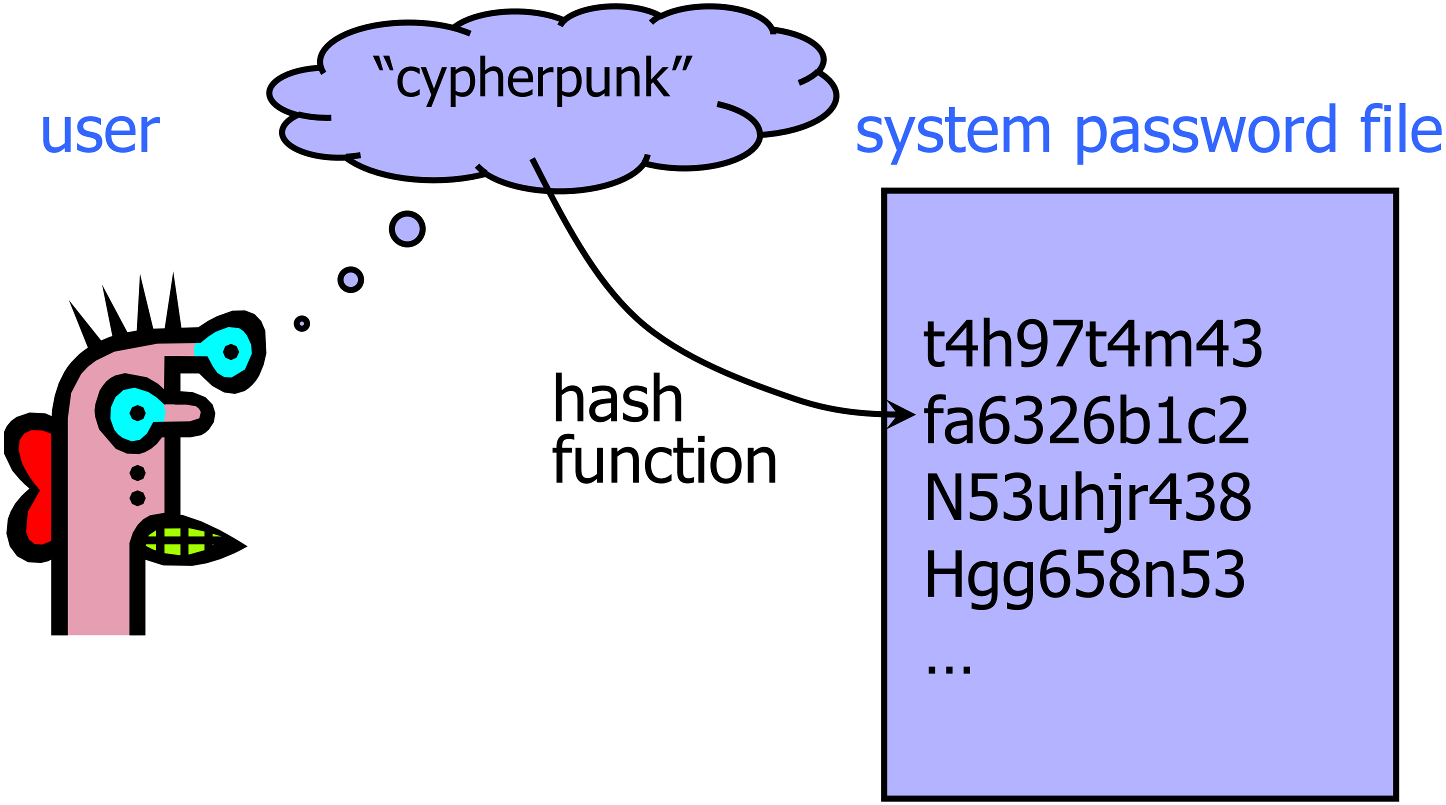
Individual states:

- What is your youngest child's birth weight?
- What color was your first bicycle?
- If you needed a new first name, what would it be?
- What band poster did you have on your wall in high school?
- How many bones have you broken?

Password Management

Countermeasure	Purpose
Password hashing	Database leak doesn’t immediately reveal user passwords; slows **offline guessing attacks**
Strength meters	Nudge / force users to pick stronger passwords to mitigate guessing attacks
Lockout after N failed attempts	Prevent remote guessing attacks (X typically 10, 100, 1000); slows down / prevents **online guessing attacks**
Compromised credential checks	Check if password is in known breaches

Storing Passwords



user

"cypherpunk"

hash function

system password file

t4h97t4m43
fa6326b1c2
N53uhjr438
Hgg658n53
…

Password Hashing

Instead of user password, store Hash(password)

When user enters a password, compute its hash and compare with the entry in the password file

- System does not store actual passwords
- Cannot go from hash to password
  - … except by guessing the password

Hash function H must have some properties

Cryptographic Hash Functions

Cryptographic hash function H maps message to short digest (e.g., 256 bit string)

One-way:

Given y = H(M), hard to compute M

Collision-resistant:

Can’t find M, M’ s.t. H(M) = H(M’)

Cryptographers have designed good hash functions

- SHA256, SHA512, SHA-3
- Deprecated hash functions: MD5, SHA-1

Dictionary Attacks

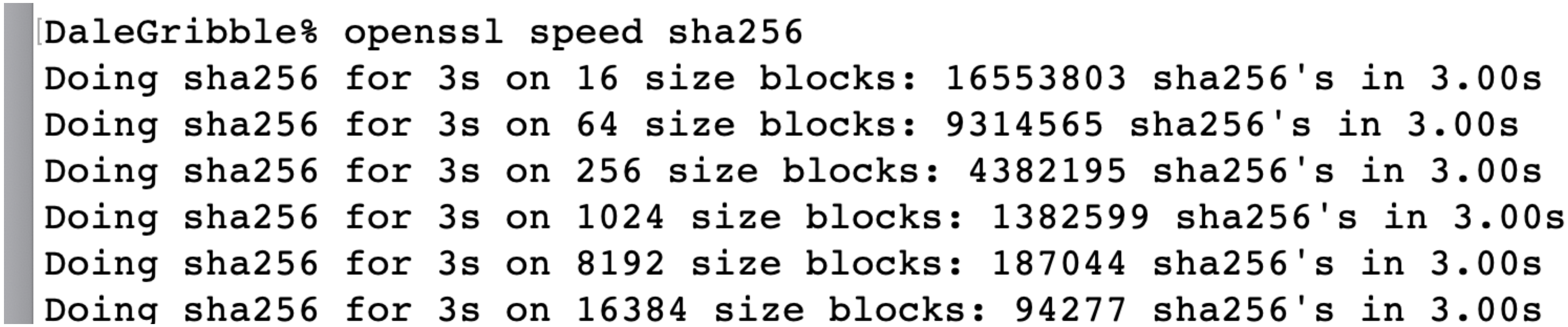
Passwords are not random

- With 52 upper- and lower-case letters, 10 digits and 32 punctuation symbols, there are $94^8 \approx$ 6 quadrillion possible 8-character passwords
- Humans like to use dictionary words, human and pet names $\approx$ 1 million common passwords

Attacker can pre-compute H(word) for every word in the dictionary – only do this once offline

- Once password file is obtained, cracking is instantaneous
- Sophisticated password guessing tools are available
    - Take into account frequency of letters, password patterns, etc.

Brute-Force Password Cracking



```
[DaleGribble% openssl speed sha256
Doing sha256 for 3s on 16 size blocks: 16553803 sha256's in 3.00s
Doing sha256 for 3s on 64 size blocks: 9314565 sha256's in 3.00s
Doing sha256 for 3s on 256 size blocks: 4382195 sha256's in 3.00s
Doing sha256 for 3s on 1024 size blocks: 1382599 sha256's in 3.00s
Doing sha256 for 3s on 8192 size blocks: 187044 sha256's in 3.00s
Doing sha256 for 3s on 16384 size blocks: 94277 sha256's in 3.00s
```

~450,000 hashes per second

How many guesses / hashes needed to crack a password?

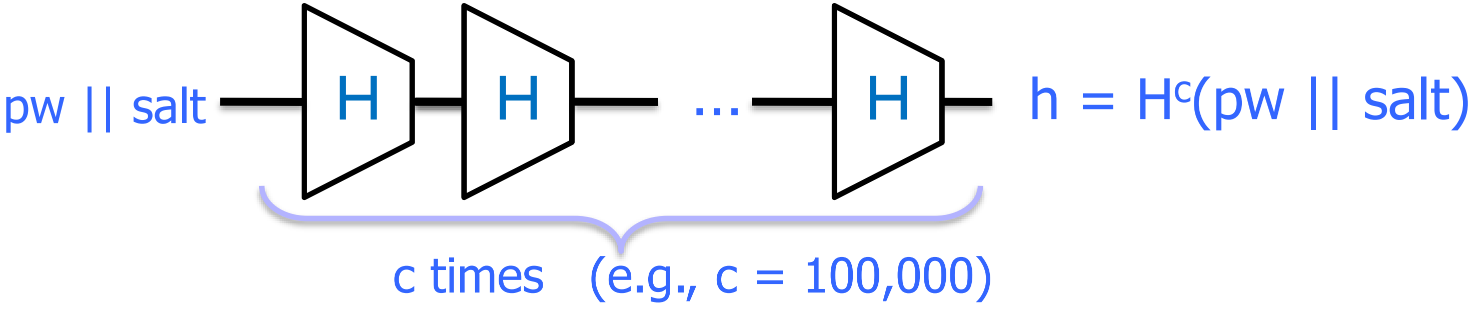
Also rainbow tables:
precompute huge number of hashes
to make a quick-lookup table

Making Cracking Harder

Make hashing slower to slow down cracking attacks

Use random per-user salts to prevent use of rainbow tables

PKCS#5 approach:

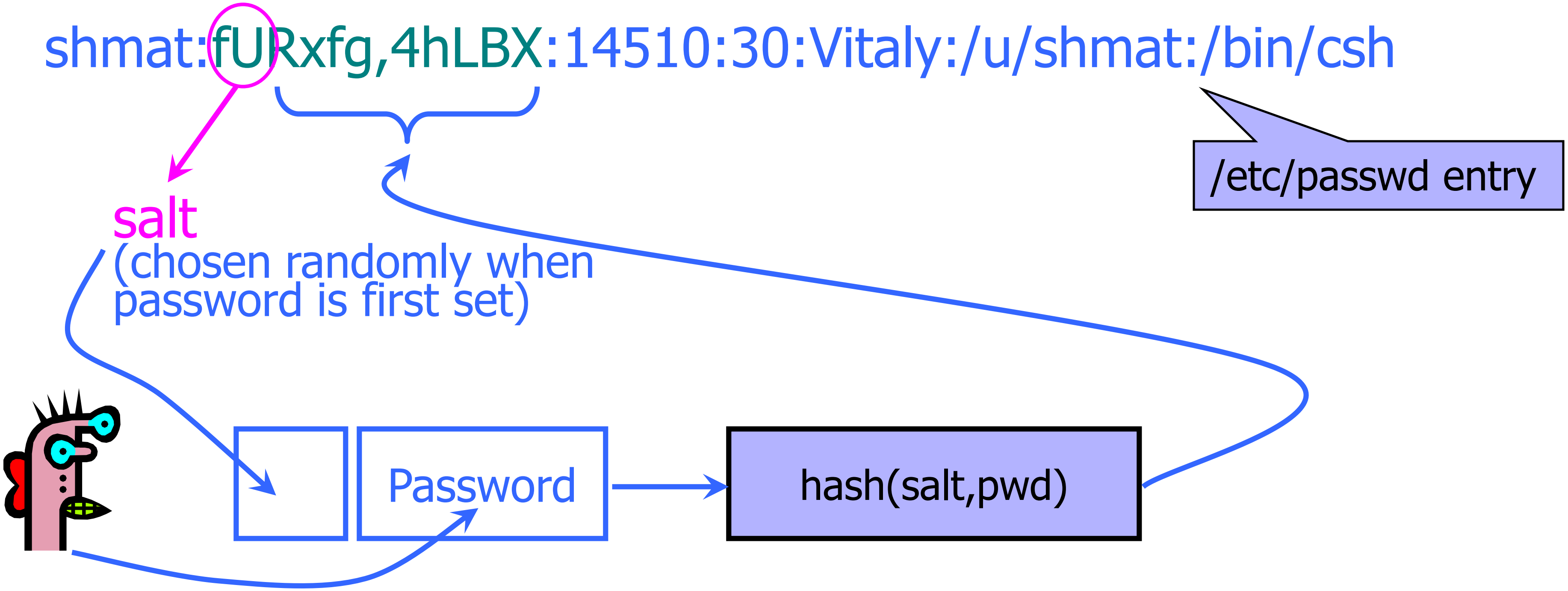


$h = H^c(pw \mid\mid salt)$

c times (e.g., c = 100,000)

Memory-hard hashing: Scrypt and argon2 require lots of memory to compute as well as time

Salt



shmat:fURxfg,4hLBX:14510:30:Vitaly:/u/shmat:/bin/csh

/etc/passwd entry

salt
(chosen randomly when
password is first set)

Password

hash(salt,pwd)

- Users with the same password have <u>different</u> entries in the password file
- Offline dictionary attack becomes much harder

Advantages of Salting

Without salt, attacker can pre-compute hashes of all common passwords once

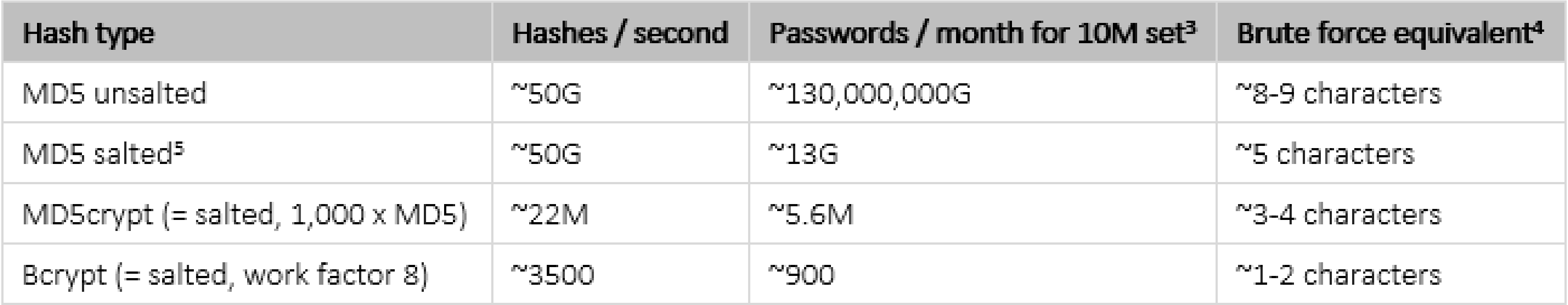
- Same hash function on all UNIX machines; identical passwords hash to identical values
- One table of hash values works for all password files

With salt, attacker must compute hashes of all common passwords for each possible salt value

- With 12-bit random salt, the same password can hash to 4096 different hash values

Modern Hash Cracking

https://medium.com/@ScatteredSecrets/how-to-crack-billions-of-passwords-6773af298172



Hash type	Hashes / second	Passwords / month for 10M set[3]	Brute force equivalent[4]
MD5 unsalted	~50G	~130,000,000G	~8-9 characters
MD5 salted[5]	~50G	~13G	~5 characters
MD5crypt (= salted, 1,000 x MD5)	~22M	~5.6M	~3-4 characters
Bcrypt (= salted, work factor 8)	~3500	~900	~1-2 characters

… with custom GPU and FPGA hardware

IBM X-Force "Cracken"
(circa 2017)

Measuring Password Strength

Hashing slows down but does not prevent guessing attacks

How do we measure password strength?

- Old approaches (deprecated)
  - NIST entropy estimate
  - Shannon entropy
- Today
  - Strength meters based on guess ranks

Shannon Entropy

- Let $\mathcal{X}$ be password distribution.
- Passwords are drawn iid from $\mathcal{X}$
- N is size of support of $\mathcal{X}$
- p_1 , p_2 , …, p_N are probabilities of passwords in decreasing order

Shannon entropy: $$H_1(\mathcal{X}) = \sum_{i=1}^{N} -p_i \log p_i$$

Poor Measure of Guessability

$N = 1{,}000{,}000$

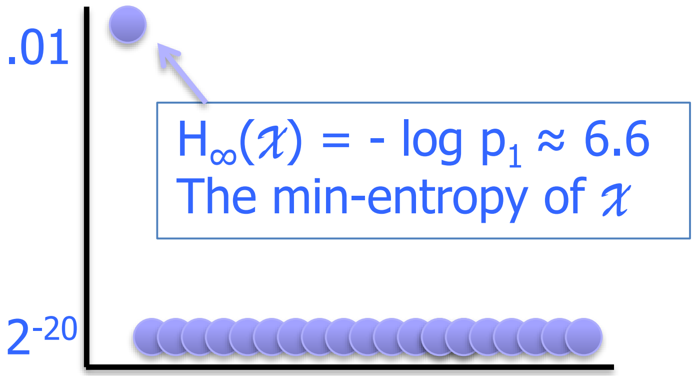
$p_1 = 1 / 100$

$p_2 = (1 - 1/100)/999{,}999 \approx 1 / 2^{20}$

…

$p_N = (1 - 1/100)/999{,}999 \approx 1 / 2^{20}$

$H_1(\mathcal{X}) \approx 19$



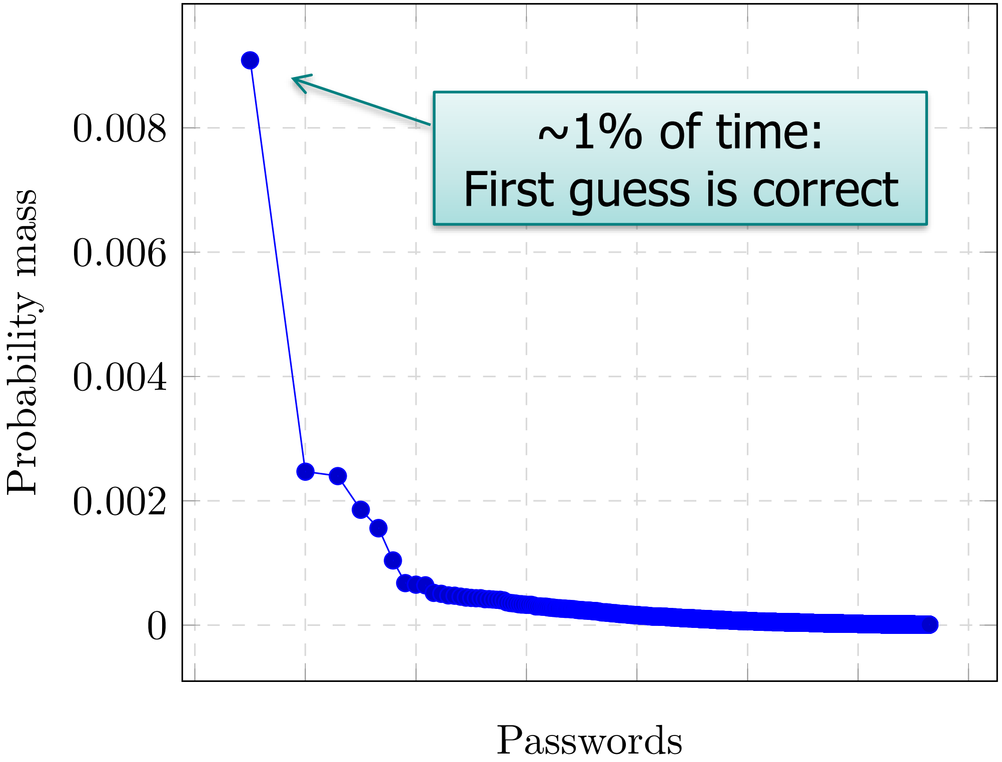
$H_\infty(\mathcal{X}) = -\log p_1 \approx 6.6$
The min-entropy of $\mathcal{X}$

19 bits of “unpredictability”. Probability of success about $1/2^{19}$?

What is probability of success if attacker makes one guess?

Shannon entropy is almost never useful measure for security

RockYou Empirical Probability



~1% of time:
First guess is correct

(Only first 5,000 points shown)

Password Policies

[redacted] [Inglesant and Sasse, “The True Cost of Unusable Password Policies”]

Overly restrictive password policies…

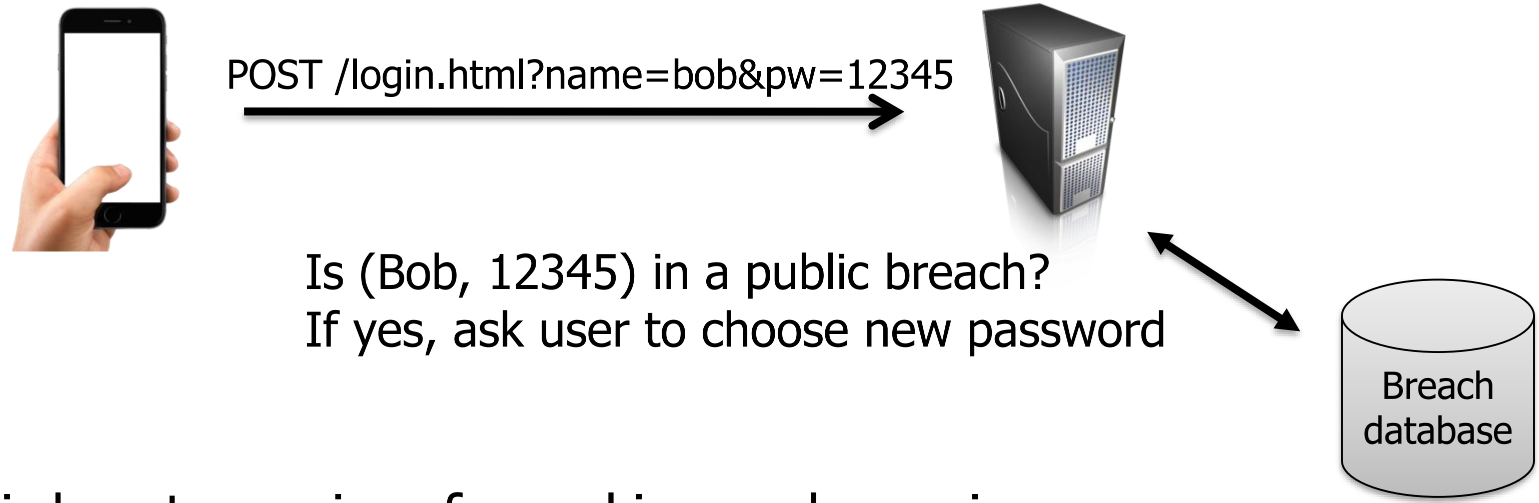
- 7 or 8 characters, at least 3 out of {digits, upper-case, lower-case, non-alphanumeric}, no dictionary words, change every 4 months, password may not be similar to previous 12 passwords…

… result in frustrated users and <u>less</u> security

- Burdens of devising, learning, forgetting passwords
- Users construct passwords insecurely, write them down
  - Can’t use their favorite password construction techniques (small changes to old passwords, etc.)
  - “An item on my desk, then add a number to it”
- Heavy password re-use across systems

Credential Stuffing

Attacker tries multiple credentials from known breaches



POST /login.html?name=bob&pw=12345

Is (Bob, 12345) in a public breach?
If yes, ask user to choose new password

Third-party services for making such queries:

- HaveIBeenPwned
- Google password checker

Credential Tweaking Attacks

Suppose user changes password to 12345**6**

Credential stuffing no longer works, but guessing attacker could try variants of 12345

Deep learning techniques to learn conditional probability distribution

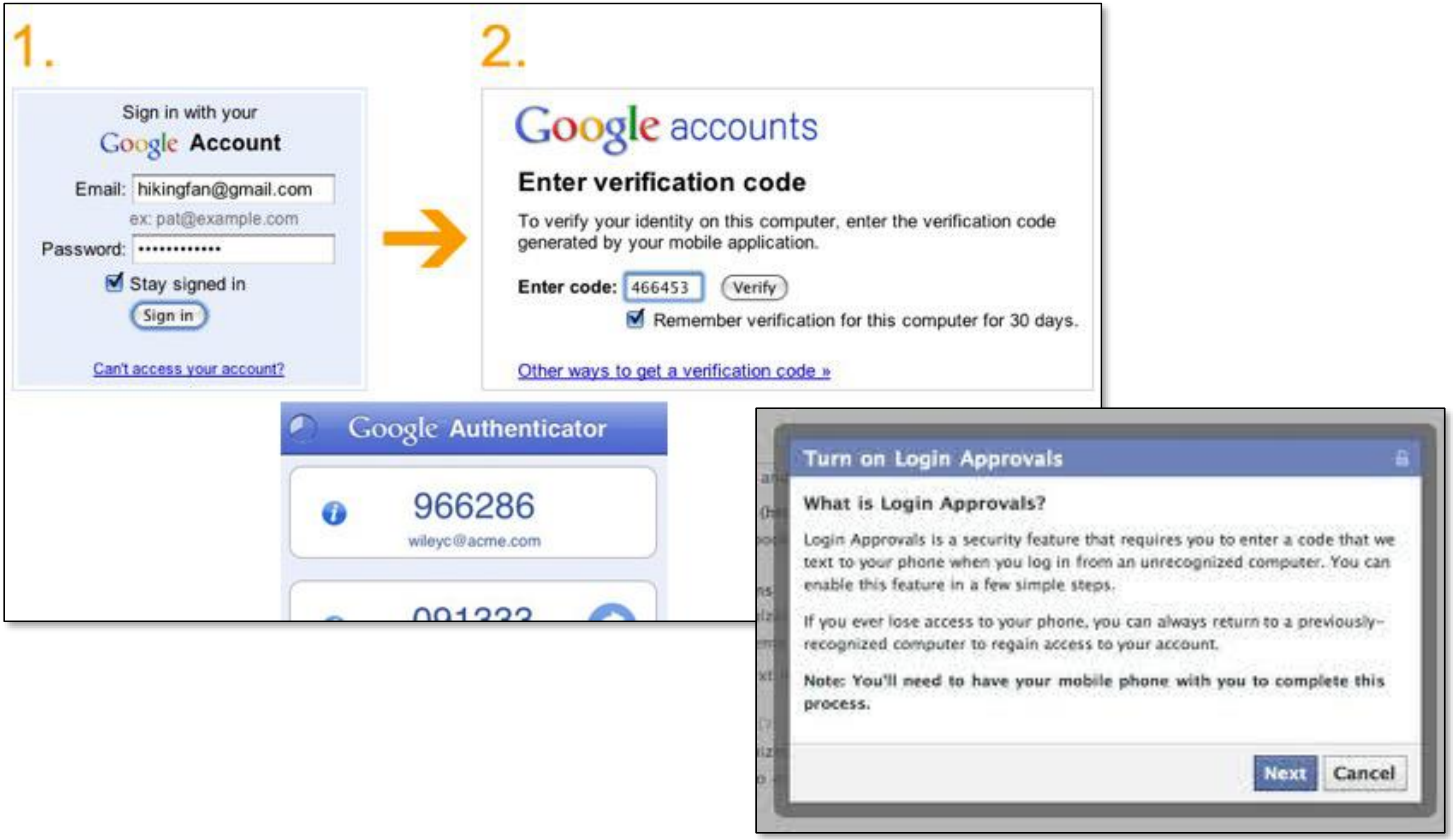
- p(pw’ | pw) where pw is leaked password, pw’ is variant
- Trained from leak data to capture typical password variants [Pal et al. 2019]

Experiments showed that 1,316 Cornell accounts vulnerable

Password Management

5 minutes to brainstorm ideas for how to improve password-based authentication

Multi-Factor Authentication



1.

Sign in with your Google Account

Email: hikingfan@gmail.com

ex: pat@example.com

Password: •••••••••••••

☑ Stay signed in

Can't access your account?

2.

Google accounts

Enter verification code

To verify your identity on this computer, enter the verification code generated by your mobile application.

Enter code: 466453

☑ Remember verification for this computer for 30 days.

Other ways to get a verification code »

Google Authenticator

966286

wileyc@acme.com

Turn on Login Approvals

What is Login Approvals?

Login Approvals is a security feature that requires you to enter a code that we text to your phone when you log in from an unrecognized computer. You can enable this feature in a few simple steps.

If you ever lose access to your phone, you can always return to a previously-recognized computer to regain access to your account.

Note: You'll need to have your mobile phone with you to complete this process.

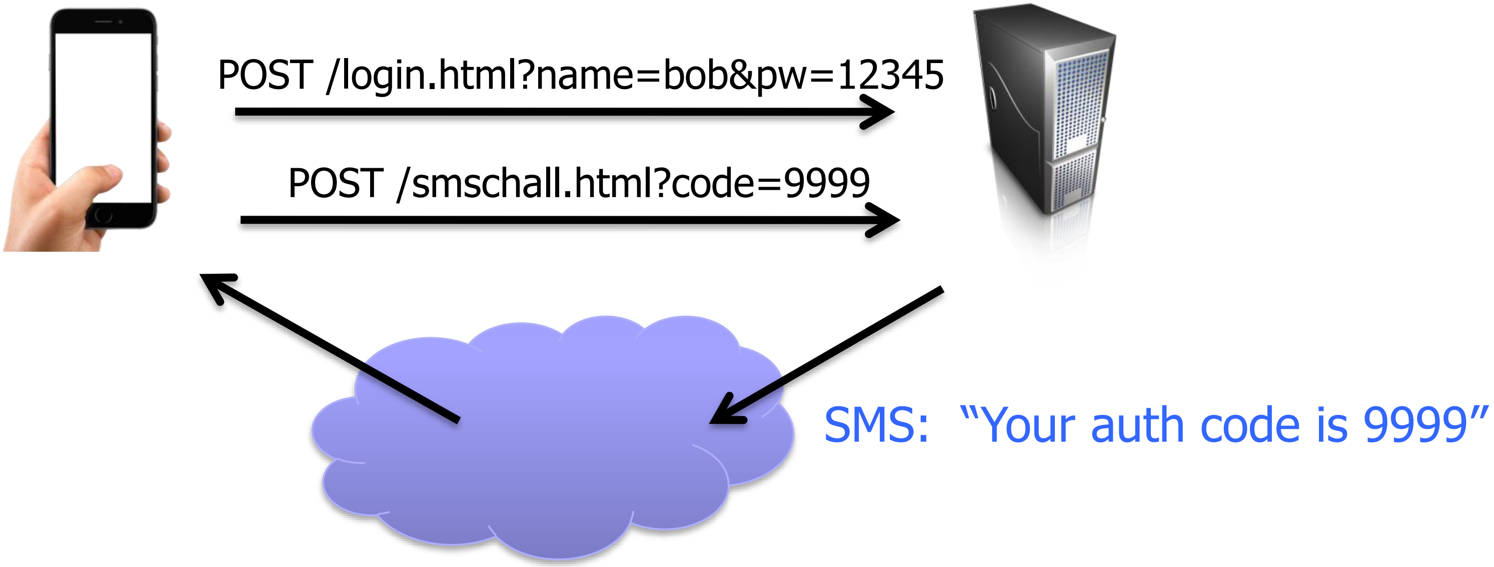
Factors for 2FA

Combine passwords with another way to authenticate user

Second factor is usually proof of ownership of …

- Email address
- Telephone number (via SMS)
- Device (via authenticator app)
- Hardware token (one-time-password token, universal second factor U2F token)

SMS Authentication



POST /login.html?name=bob&pw=12345

POST /smschall.html?code=9999

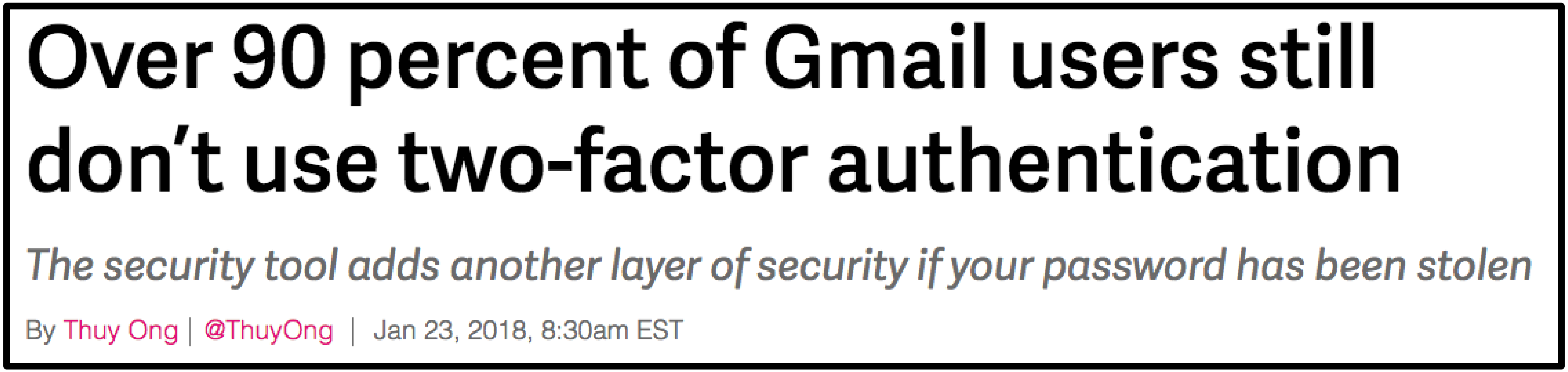
SMS: “Your auth code is 9999”

Suppose you know someone’s password (e.g., due to breach) but their account is protected by SMS-based 2FA. **What can you do as an attacker?**

Circumventing SMS-Based 2FA

- Have physical access to device that receives SMS
- Phishing attacks: confuse user into disclosing SMS to you
- SIM swap: trick phone company into registering victim’s phone # to your device
- SMS hijacking: exploit vulnerabilities in cellular network
    - https://berlin.ccc.de/~tobias/31c3-ss7-locate-track-manipulate.pdf
    - [Doerfler et al. 2019]: SMS 2FA circumvented in ~4% of phishing attacks, ~26% of targeted attacks

- Best practice: authenticator app or hardware token



Over 90 percent of Gmail users still don't use two-factor authentication

The security tool adds another layer of security if your password has been stolen

By Thuy Ong | @ThuyOng | Jan 23, 2018, 8:30am EST

Usability remains a key issue preventing adoption

Other Authentication Signals

Location-based authentication

- IP-based geolocation

Device identification

- Cookies, device fingerprinting

Behaviorial cues

- Typical actions on platform (even after authenticated)

Biometrics

- Fingerprints, etc

User Authentication Is a Huge Pain

Simple typos in passwords cause 3% of Dropbox users to be unable to login in 24-hour period

[Chatterjee et al. 2016]

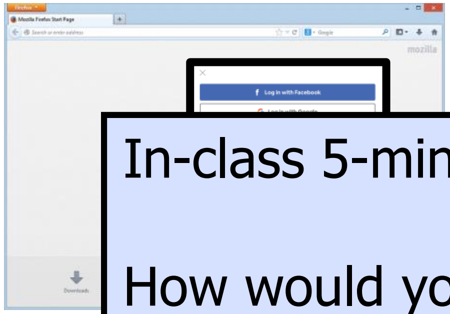
52% of users fail login challenges at Google, 3% don’t get in within short period of time

[Doerfler et al. 2019]

Single Sign-On (SSO)

Identity provider handles authentication

- Google, Facebook, proprietary services, etc.



Identity provider
(e.g., identity.cornell.edu)

In-class 5-minute exercise:

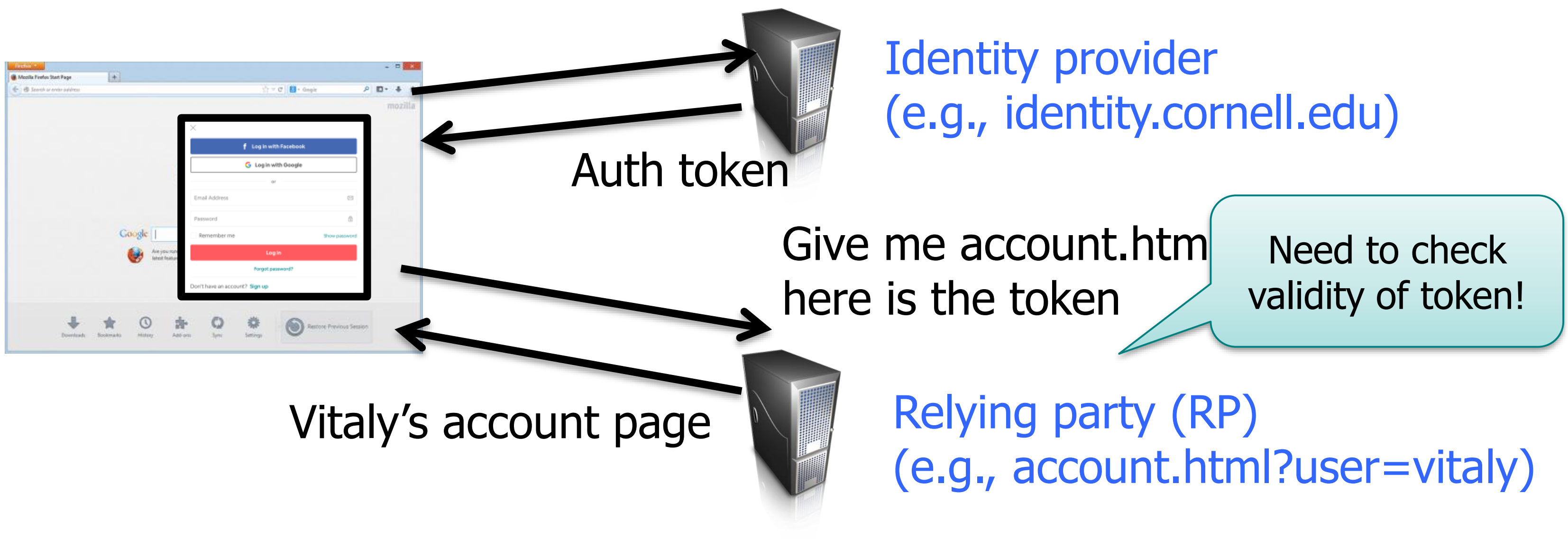
How would you securely authorize browser to access www.cornell.edu/account.html?user=vitaly based on authentication to identity.cornell.edu?

?user=vitaly)

Single Sign-On (SSO)

Many standards and systems

- SAML, OpenID Connect + OAuth 2.0, …



Give me permissions for account.html as vitaly

Auth token

Identity provider (e.g., identity.cornell.edu)

Give me account.html here is the token

Need to check validity of token!

Vitaly's account page

Relying party (RP) (e.g., account.html?user=vitaly)

More about OAuth later…

Strengthening Passwords

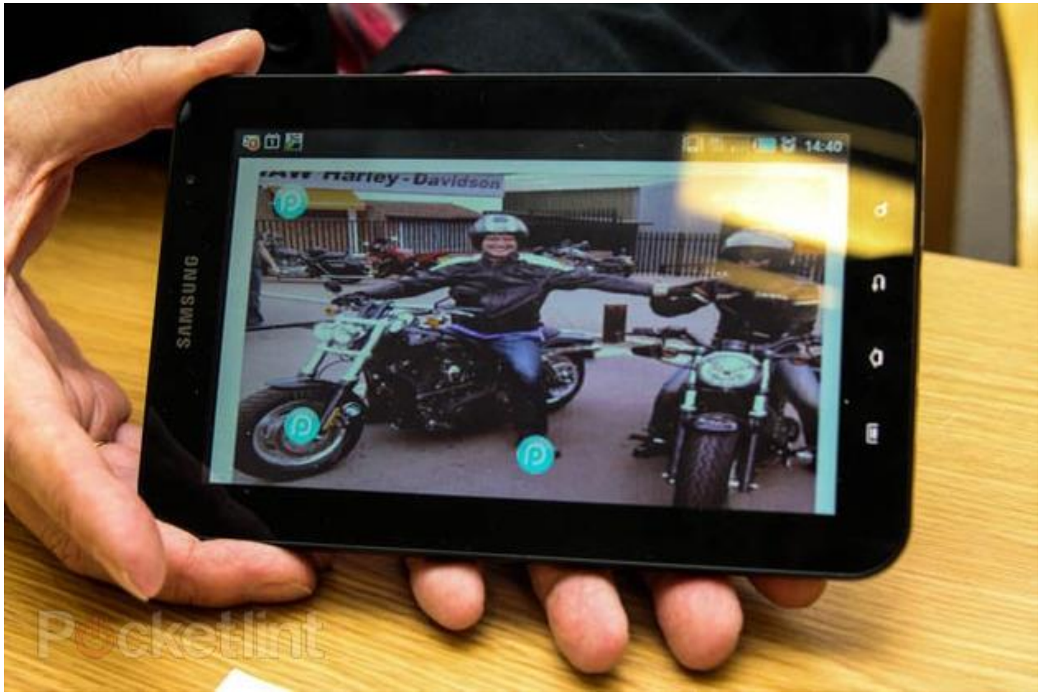
Add biometrics

- For example, keystroke dynamics or voiceprint
- Revocation is often a problem with biometrics

Graphical passwords

- Goal: increase the size of memorable password space
- Dictionary attacks are believed to be difficult because images are very "random" - is this true?

PixelPin



Upload a picture,
use 3 or more points as the "password"

random?

Images + Story

Invent a story for an image
or a sequence of images

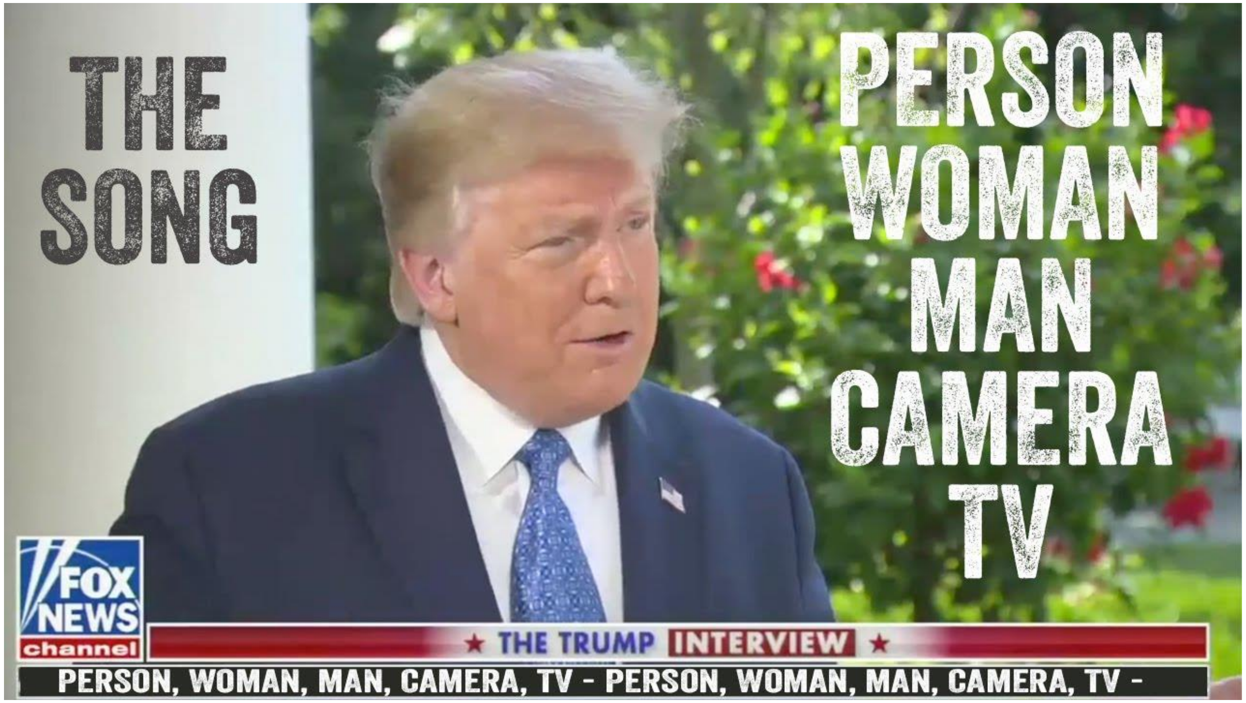
*"We went for a walk
in the park yesterday"*

Fish-woman-girl-corn

Need to remember the order!

'Person, woman, man, camera, TV': Trump insists cognitive test was difficult

US president's pride in his own mental agility on display during interview in which he lists five things repeatedly



User Experiences

50% unable to invent a story, so try to pick four pleasing pictures and memorize their order

- “I had no problem remembering the four pictures, but I could not remember the original order”
- “… on the third try I found a sequence that I could remember, fish-woman-girl-corn. I would screw up the fish and corn order 50% of the time, but I knew they were the pictures”

Picture selection biases

- Males select nature and sports more than females
- Females select food images more often

Alternatives to Passwords

Mobile phones,
USB devices,
special tokens,
etc. etc.

OneID

M-Pin™

LaunchKey

fido™ FORGET PASSWORDS!

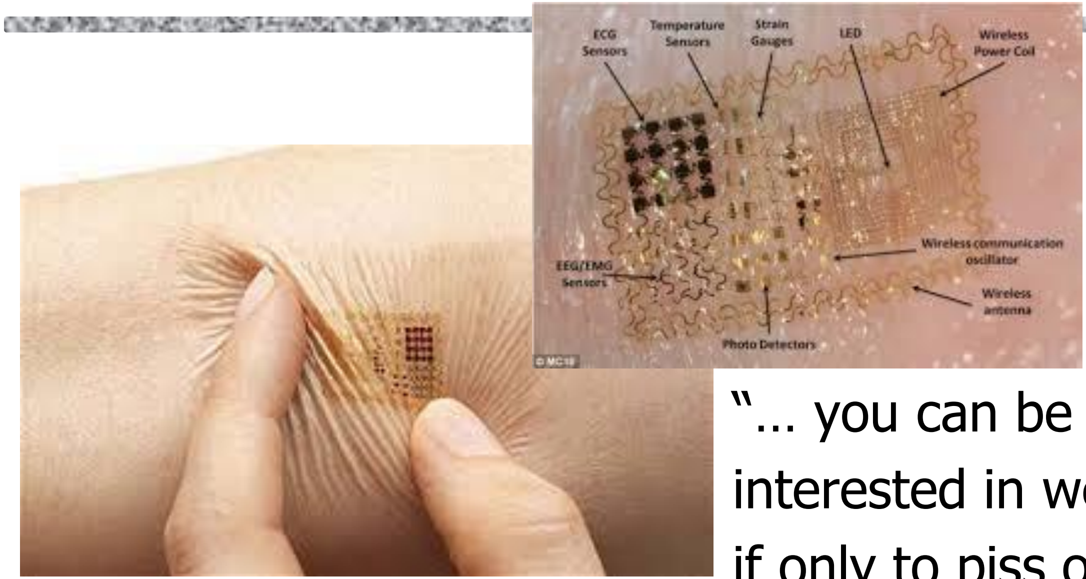
CRYOKEY

yubico
Trust the Net.

Clef

PIN grid

Alternatives from Motorola



"… you can be sure that they'll be far more interested in wearing an electronic tattoo, if only to piss off their parents"

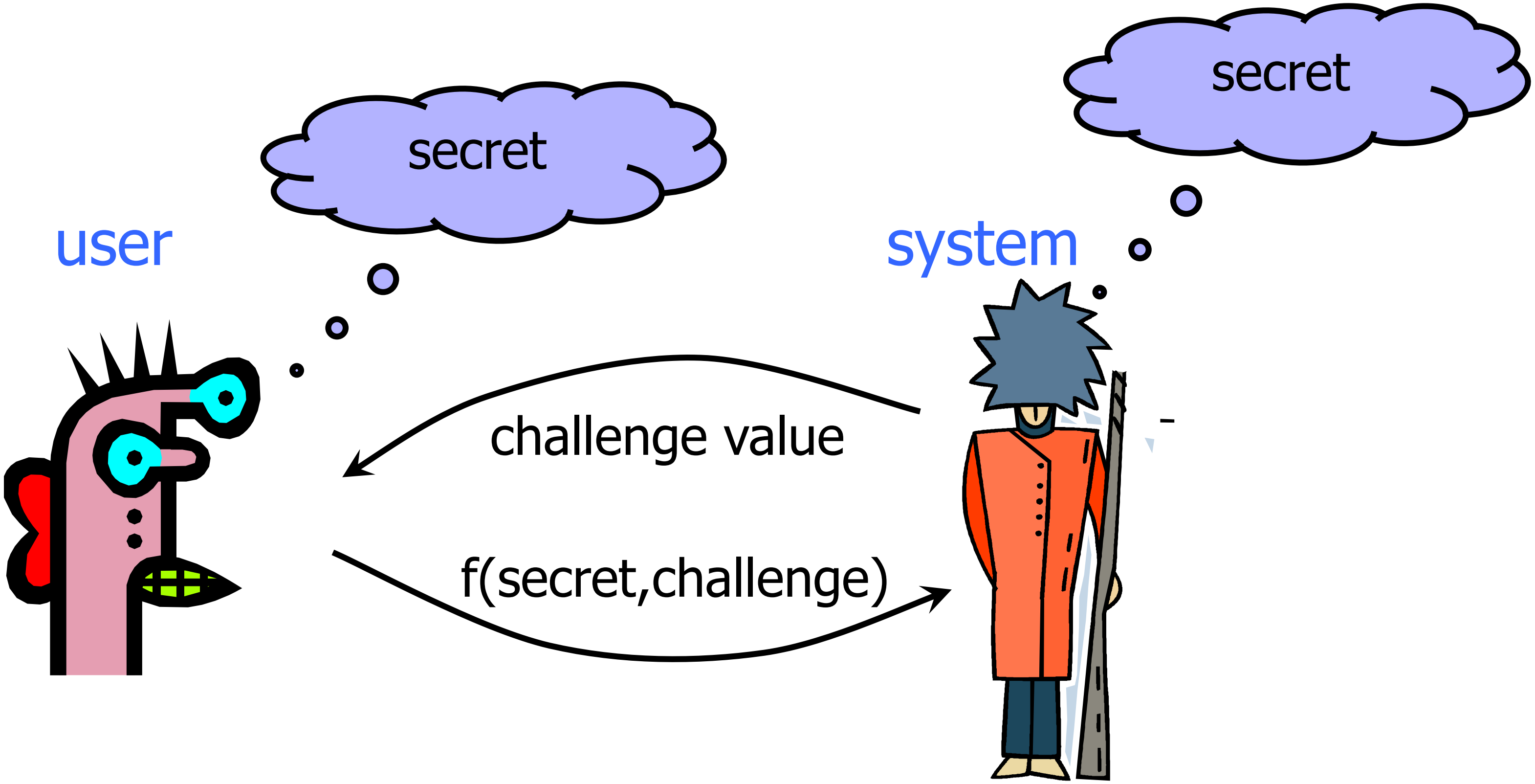
"The pill features a small chip with one switch that uses your stomach acids to activate an 18-bit ECG-like signal inside your body"

One-Time Passwords

Idea: use a shared secret to derive a one-time password

If the attacker eavesdrops on the network, he’ll learn this password but it will be useless for future logins

Challenge-Response



Why is this better than the password over a network?

Challenge-Response Authentication

User and system share a secret (key or password)

Challenge: system presents user with some string

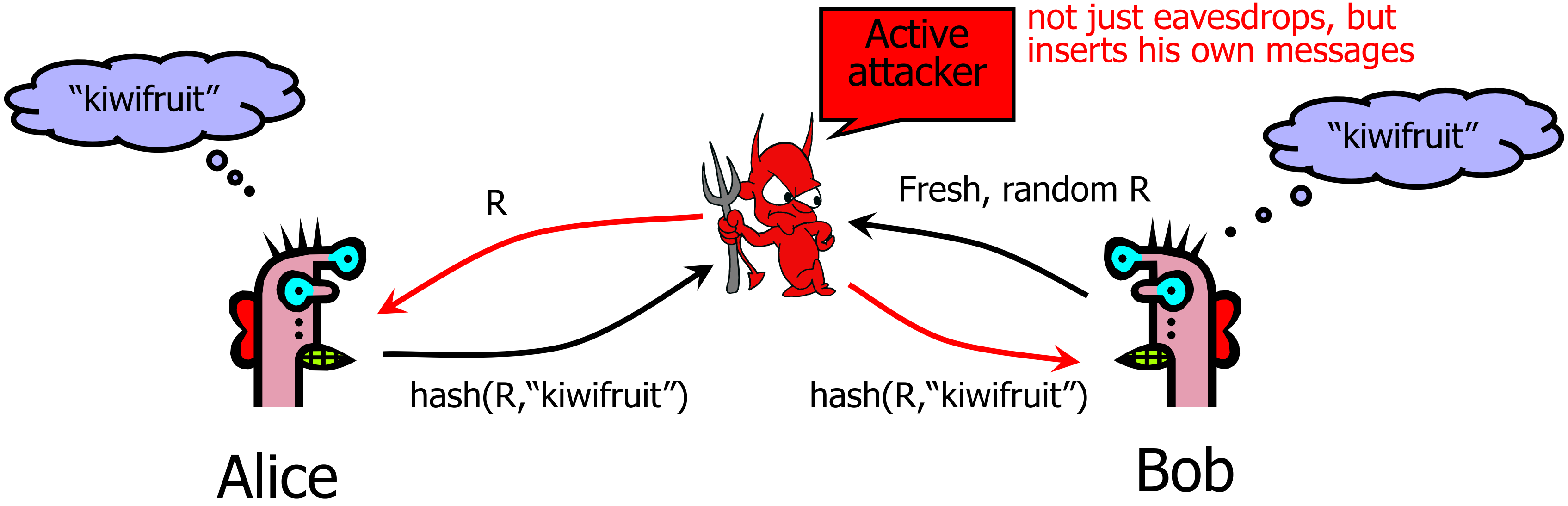
Response: user computes the response based on the secret and the challenge

- Secrecy: difficult to recover secret from response
  - Cryptographic hashing or symmetric encryption work well
- Freshness: if the challenge is fresh, attacker on the network cannot replay an old response
  - Fresh random number, counter, timestamp….

Good for systems with pre-installed secret keys

- Car keys; military friend-or-foe identification

Man-in-the-Middle Attack



Man-in-the-middle attack on challenge-response

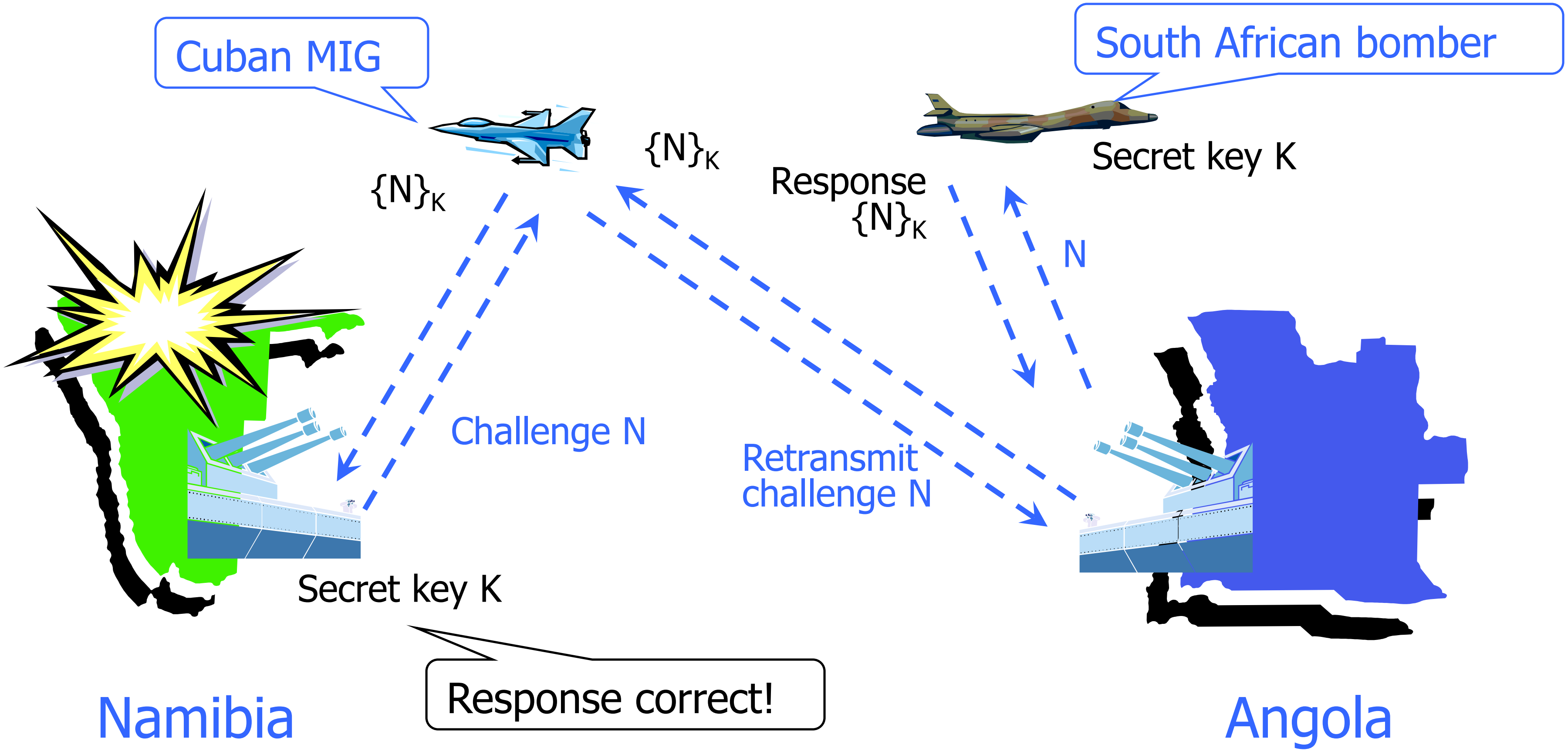
- Attacker successfully "authenticates" as Alice by simple replay

This is an online attack

- Attacker does not learn the shared secret
- Attacker cannot "authenticate" as Alice when she is offline

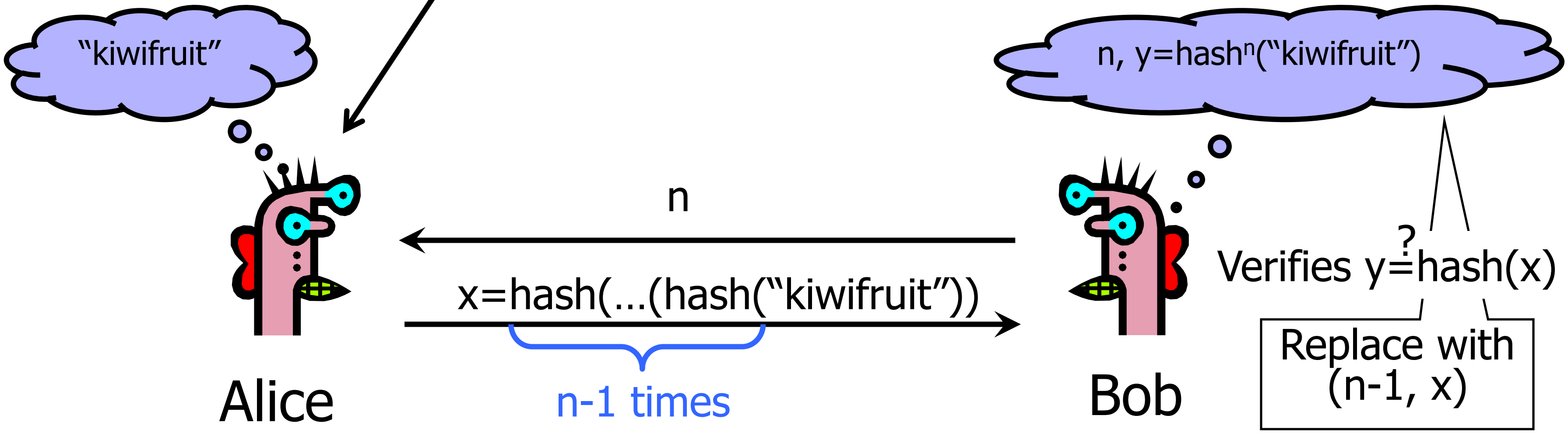
MIG-in-the-Middle

[Ross Anderson]



Lamport's Hash / S-Key

A sheet of paper with N "passwords", cross out a password after using it, move to next one



"kiwifruit"

n, $y=\text{hash}^n$("kiwifruit")

n

$x=\underbrace{\text{hash}(\ldots(\text{hash}}_{\text{n-1 times}}(\text{"kiwifruit"}))$

Alice

Bob

Verifies $y \overset{?}{=} \text{hash}(x)$

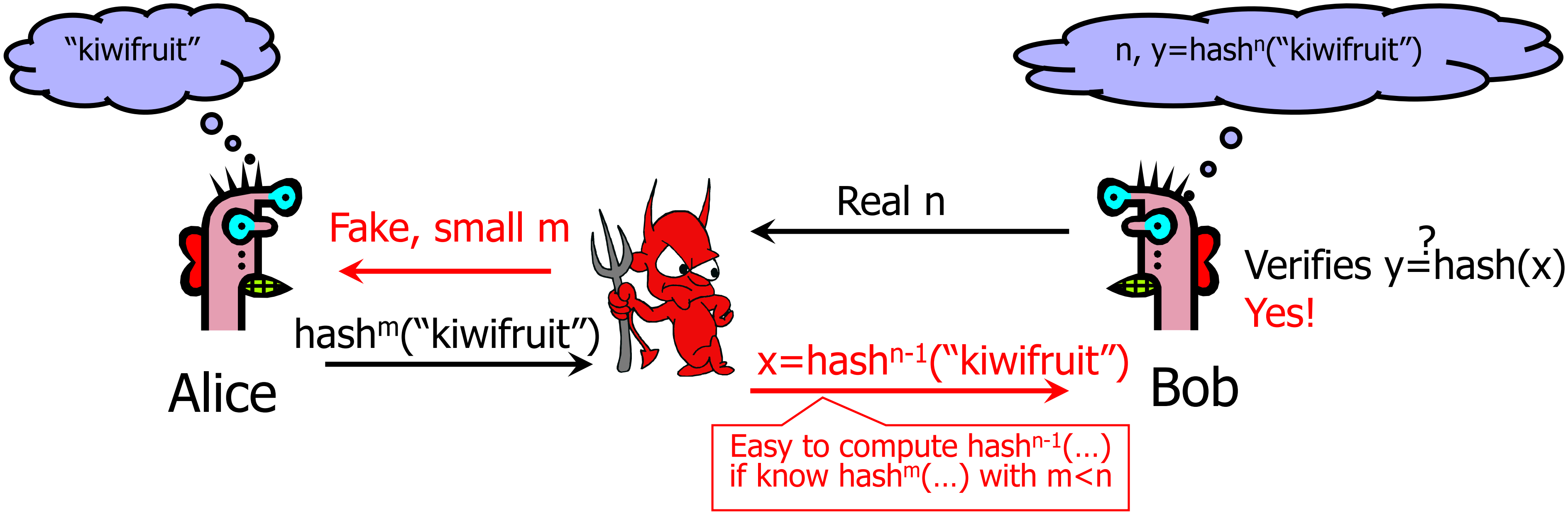
Replace with (n-1, x)

Main idea: "hash stalk"

- Moving up the stalk (computing the next hash) is easy, moving down the stalk (inverting the hash) is hard
- n should be large - a stalk is only good for n authentications

Verifier only needs the current tip of the stalk

“Small n” Attack



“kiwifruit”

n, y=hash^{n}(“kiwifruit”)

Alice

Fake, small m

hash^{m}(“kiwifruit”)

Real n

x=hash^{n-1}(“kiwifruit”)

Easy to compute hash^{n-1}(…)
if know hash^{m}(…) with m<n

Bob

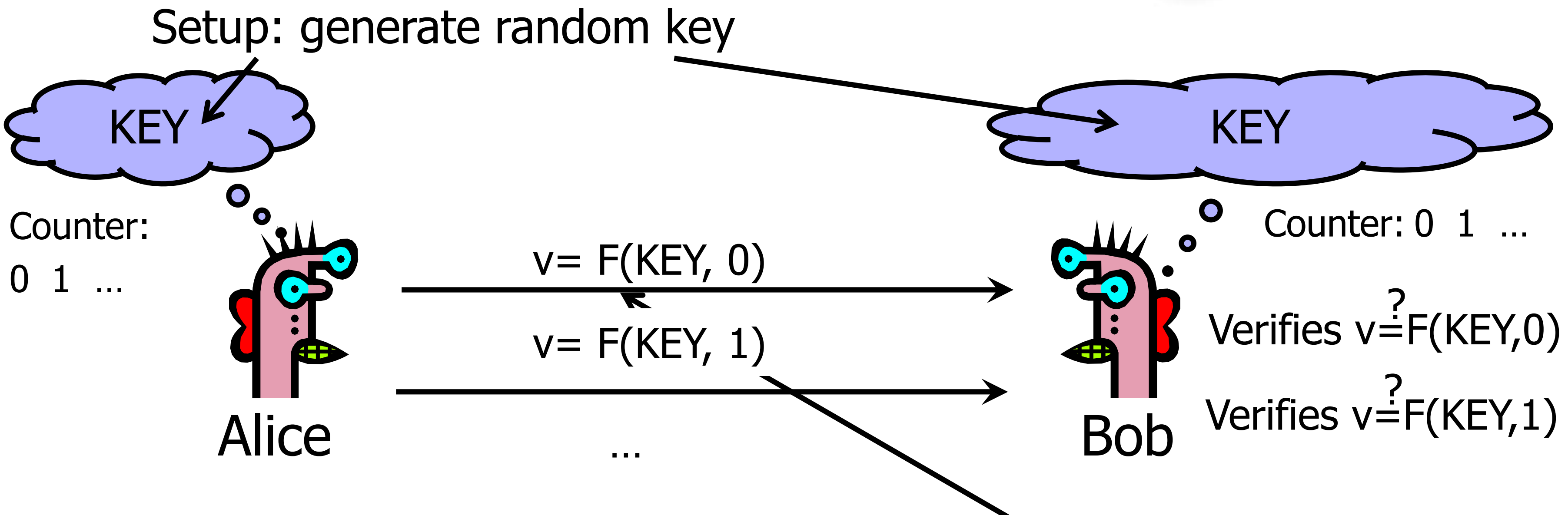
Verifies $y\overset{?}{=}\text{hash}(x)$

Yes!

First message from Bob is not authenticated!

Alice should remember the current value of n

SecurID



Setup: generate random key

KEY

KEY

Counter: 0 1 …

Counter: 0 1 …

Alice

Bob

v= F(KEY, 0)

v= F(KEY, 1)

…

Verifies $v \overset{?}{=} F(KEY,0)$

Verifies $v \overset{?}{=} F(KEY,1)$

RSA uses a custom function
Input: 64-bit key, 24-bit ctr
Output: 6-digit value

Advancing the counter

- Time-based (60 seconds) or every button press

Allow for skew in the counter value

- 5-minute clock skew by default

Guessing Mother’s Maiden Name

[Griffith and Jakobsson]

Griffith and Jakobsson, “Messin' with Texas: Deriving Mother's Maiden Names Using Public Records” (2005)

Insight: MMN is a fact, not a secret

Figure out people’s MMN by creating ancestry trees from records that are public by law

Target: Texas

- Large population
- Close to national averages
- Good online records

Useful Public Records (1)

[Griffith and Jakobsson]

US Census records

- Individual records released with 72-year delay
  - Individual data sheets for the 1940 Census released in 2012
- Can read MMN directly, but difficult

Voter registration records

- 67% of Texans registered to vote (2000)
- Voter information has “Other Name” field, people often put maiden name there
- Also full name, date of birth, address
- Not free!

Useful Public Records (2)

[Griffith and Jakobsson]

Property records

- Match addresses to names (“legally enforced phonebooks”), good in combination with phonebooks
- Include people who have children but haven’t married

Obituaries

- Obituaries of “important” people in local newspapers often mention spouse, children, date of birth, when married, etc.

SSDI (Social Security Death Index)

- Free, comprehensive, but no direct MMN info
- Purpose: prevent mafia from using SSNs of dead people

Useful Public Records (3)

[Griffith and Jakobsson]

Marriage records

- Names and ages of bride and groom, date of marriage, where married

Birth records

- Full name, date of birth, where born

Sources of birth and marriage records

- Mormons
- Rootsweb.com's WorldConnect
  - Family trees for 4499 living Texans
- Rootsweb.com's USGenWeb
  - 11,358,866 birth records, mainly from county records

Texas Bureau of Vital Statistics

[Griffith and Jakobsson]

1966-2002 marriage index online

1968-2002 divorce index online

1926-1995 birth records, taken offline in 2000

- So that adopted children can’t find their natural parents
- Copies still available at archive.org

1965-1999 death records, taken offline in 2002

- Unlinked, but actual files still found at old URLs

Low-Hanging Fruit in Birth Records

[Griffith and Jakobsson]

1923-1949 birth records have MMN in plaintext

- 1,114,680 males auto-compromised

1,069,448 females in records

- Linking females born in 1923-1949 to marriages 1966-2002 gives 288,751 compromises (~27%)
  - Use full name, DoB to connect women to marriages
  - If more than 1 marriage per woman, divorce records help

1950-1995 has 40,697 hyphenated last names

Insights for Guessing MMN

[Griffith and Jakobsson]

Children have same last name as their parents

Suffixed children will have same first and last name as parents

Children often born shortly after parents' marriage

Children born shortly after parents' marriage often born in same county

- Makes guessing much easier than you'd normally think… Especially true for the clustering of names within ethnic groups - don't have to pick the correct parents, just the correct MMN!

Example #1: Unique Last Name

[slide: Virgil Griffith]

Ernest AAKQUANAHANN

Dionne COX

Mother’s maiden name = COX

Example #2: Two Marriages

[slide: Virgil Griffith]

Shawn ZUTTER

Lisa MENDOZA

Chad ZUTTER

Lauren LANDGREBE

Entropy = 1 bit

(need at most 2 guesses)

Example #3: Two Marriages

[slide: Virgil Griffith]

Robert STUGON

Duarte STURNER

Jim STUGON

Luann STURNER

Mother's maiden name = STURNER

Insights for Guessing MMN

[Griffith and Jakobsson]

Last names + birth records

+ 82,272 Texans

- Birth records not very comprehensive

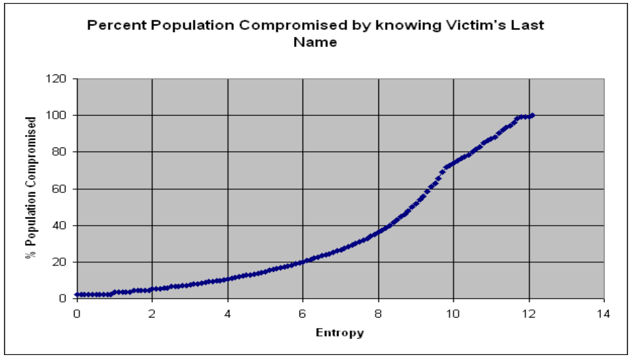
Suffixed last names

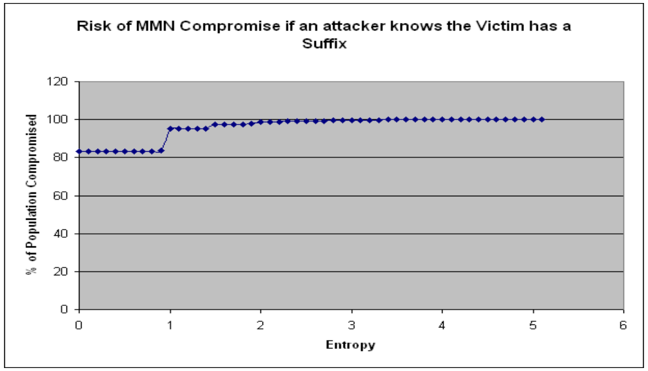
+ 344,463 Texans

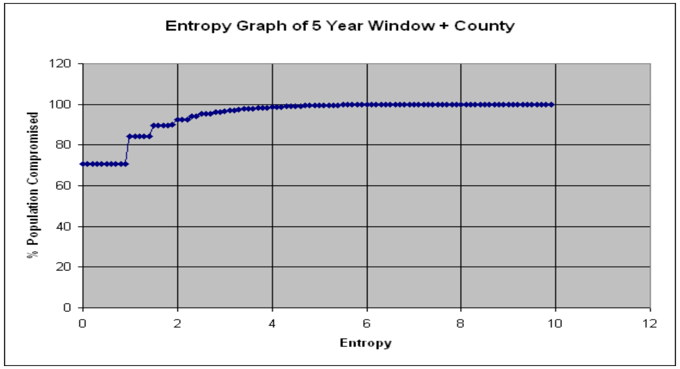
- 60% of suffixed children in birth records

Assume child is born 5 years from marriage, in the same county

+ 2,355,828 Texans


Percent Population Compromised by knowing Victim's Last Name


Risk of MMN Compromise if an attacker knows the Victim has a Suffix


Entropy Graph of 5 Year Window + County

MMN Considered Harmful

Griffith-Jakobsson study figured out mother's maiden name for 4,190,493 Texans using only free, public sources of information

- 1/5 of the state's population

More sources of information available

- More comprehensive birth records available for sale

More sophisticated analyses possible

Conclusion: mother's maiden name is not a secure authentication factor