CSE331 Computer Security Fundamentals
Authentication

Nick Nikiforakis
nick@cs.stonybrook.edu
Controlling access

• One of the things that makes security hard is that we want fine-grained access to systems, networks, and information
  – Everyone accessing them is not okay
  – No-one accessing them is also not okay

• We only want specific users (or programs acting on behalf of users) to access specific programs, systems, and data

• So the first step is authentication
  – How do I know you are, who you say you are?
User Authentication

• Authentication based on:
  – What you know (password, answer to security question, personal image etc.)
  – What you have (smartcard, hardware token, etc.)
  – Who you are (biometrics)
  – Where you are (IP address, GPS location)
    • Less popular, but smartphones could change this
Password-Based Authentication

User has a secret password.
System checks it to authenticate the user.

• How is the password communicated?
  – Eavesdropping risk (We will see later how crypto can be used)

• 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
Other Aspects

• Usability
  – Hard-to-remember passwords?
  – Carry a physical object all the time?

• Denial of service
  – Stolen wallet
  – Attacker tries to authenticate as you, account locked after three failures

• Social engineering
Passwords and Computer Security

- In 2012, 76% of network intrusions exploited weak or stolen credentials (username/password)
  – 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
Password Security Risks

• Keystroke loggers
  – Hardware
    • KeyGhost, KeyShark, others
  – Software (spyware)
• Shoulder surfing
• Same password at multiple sites
  – Cascading effects of a single break-in
• Social engineering
How do passwords look like?

• A lot of what we know about passwords is mainly through two ways
  – **Ethical**: Academic studies done in universities where researchers ethically investigated the password practices of their institutions
  – **Less ethical**: Hacking groups hacked and release entire databases of passwords, from their victims.
Figure 1: The percentage of passwords guessed after a given number of guesses (shown in log scale), by college within the university.
Figure 3: The relative likelihoods of passwords with digits, symbols, or uppercase letters in a given location being cracked. For example, a password with all its digits at the end is five times as likely to be cracked as a password with its digits spread throughout, other things being equal.
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?)
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 an SQL injection attack and posted on the Internet
  – More about SQL injection attacks later
## Passwords in RockYou Database

### Password Popularity – Top 20

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

More than 30 million passwords

“#1 Most Trusted Online Dating Site”

SQL injection attack
For sale for $3000
The Dropbox hack is real

31 AUGUST 2016

Earlier today, Motherboard reported on what had been rumoured for some time, namely that Dropbox had been hacked. Not just a little bit hacked and not in that "someone has cobbled together a list of credentials that work on Dropbox" hacked either, but proper hacked to the tune of 68 million records.

Article link: https://www.troyhunt.com/the-dropbox-hack-is-real/
Attackers

• What is the threat model?
  – Online attacker
    • Tries to login to a service by iteratively trying passwords and looking whether he was successful
  – Offline attacker
    • Stole password database and tries to recover the, hopefully protected, passwords
      – Also known as a “dictionary attack”
  – Against one user
  – Against all/any user
Online attacker

• How do we detect an online attacker?
  – Too many wrong tries
    • Distinctly different from a user who first was wrong but then was right
  – Tries multiple accounts instead of just one
    • Tradeoff to allow for NAT usage

• What can we do?
  – Temporarily block the IP address or rate-limit the number of requests
    • Tricky if attack is distributed
  – Temporarily lock the account that is being attacked
    • Rarely a good solution (Harms availability property)
Offline attacker

• Attacker somehow obtains the list of our passwords
  – Break-in to server
    • Credential guessing, SQL injection, Buffer overflow,…
  – Backups
  – Social engineering

• **It’s obvious** that the passwords should not be stored in the clear
  – How do we not store them in the clear, and still check them against users attempting to log in?
Unfortunately it is not obvious for all
Should we use encryption?

• How about encrypting each password with a secret key (e.g. only stored in the memory of the server) which is used to decrypt any single entry, on demand?

• Still a bad idea....
  – The attacker can steal your key and decrypt everything
  – The administrators can know users’ passwords (no reason that they should)
Password Hashing

• Instead of user password, store $\text{Hash(password)}$
• 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
  – Given $H(\text{password})$, hard to find any string $X$ such that $H(X)=H(\text{password})$ - why?
## Sample Cryptographic hash functions

<table>
<thead>
<tr>
<th>Name</th>
<th>Year of release</th>
<th>Digest size (output size)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD5 (Media Digest 5)</td>
<td>1992</td>
<td>128-bit</td>
</tr>
<tr>
<td>SHA-1 (Secure Hash Algorithm 1)</td>
<td>1995</td>
<td>160-bit</td>
</tr>
<tr>
<td>SHA-256 (Part of the SHA-2 family)</td>
<td>2001</td>
<td>256-bit</td>
</tr>
</tbody>
</table>

MD5(“helloworld”) = d73b04b0e696b0945283defa3eee4538  
SHA-1(“helloworld”) = e7509a8c032f3bc2a8df1df476f8ef03436185fa  
SHA-256(“helloworld”) = 8cd07f3a5ff98f2a78cfc366c13fb123eb8d29c1ca37c79df190425d5b9e424d
Examples

SHA1(“mysecretpassword”) = 08cd923367890009657eab812753379bdb321eeb
SHA1(“mysecretpassword”) = 0c894b9cd0fef7d1ccfe0729d5ff7af9509731ed
SHA1(“mysecretpassword”) = 27c2d31b648cf7773032d1a06c8ee610c3f5b32c

Small changes in input

Large differences in output

This is called the “avalanche” effect
Hashing vs. Encryption

• Hashing is one-way. There is no “uh-hashing”!
  – A ciphertext can be decrypted with a decryption key... hashes have no equivalent of “decryption”
• Hash(x) looks “random”, but can be compared for equality with Hash(x’)
  – Hash the same input twice → same hash value
• Cryptographic hashes are also known as “cryptographic checksums” or “message digests”
Steps for authenticating when hashing is utilized

• When an existing user is trying to login:
  1. Hash the provided password
  2. Compare it to the stored hash for that user
  3. If the hashes match then the provided password is the same as the original one

• Better than before, but still has issues
  – Same passwords of different users will have the same hash
  – Attacker can precompute hashes of popular words and try them against all accounts (rainbow tables)
Salting

• Instead of just hashing the user’s password, hash the user’s password when concatenated with a per-user random value

<table>
<thead>
<tr>
<th>Username</th>
<th>Salt</th>
<th>Password</th>
</tr>
</thead>
<tbody>
<tr>
<td>nick</td>
<td>199654</td>
<td>a4de0a7f8905a0e3832cca84eff4f76a465c85a3</td>
</tr>
</tbody>
</table>

SHA1(“mysecretpassword”)

SHA1(“199654mysecretpassword”)

Username | Password
----------|-------------------
nick      | 08cd923367890009657eab812753379bdb321eeb
Salting steps

• When an existing user is trying to login:
  1. Concatenate the database-stored salt to the user-provided password
  2. Hash the entire thing
  3. Compare it to the stored hash for that user
  4. If the hashes match then the provided password is the same as the original one

• Precomputed tables don’t work
• Same passwords for different users have different hashes
• Attacker must perform hashing “live” in order to crack passwords
HMAC

• Concatenation of keys with data can lead to some exploitable cryptographic scenarios
  – Outside the scope of this course

• HMAC (Keyed Hashed Message Authentication Code) allows us to combine the salt with the hash of the password in a more secure way
HMAC with SHA

HMAC performs two iterations of a chosen cryptographic hash to create a "keyed hash"
One more thing

• Our steps so far allow us the following guarantees:
  – User passwords should not be recoverable from a database
  – Identical/Similar passwords will have different hashes
  – The database does not “leak” the length of a user’s password

• The only problem remaining is that offline attackers, if they are dedicated enough, they can still brute-force their way into users with weak passwords
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
Password Hash Cracking


• Custom GPU-based hardware
  – A 5-server rig with 25 Radeon GPUs
  – 348 billion NTLM passwords per second
    • NTLM = Microsoft’s suite of security protocols
    • 6 seconds to crack a 14-character Windows XP password
  – 77 million md5crypt-hashed passwords per second
    • md5crypt() is used by FreeBSD and Linux

• Cloud-based cracking tools
  – Project Mars, Crackq, etc.
  – Password-cracking as a service
Hash stretching

• Why restrict ourselves to only one hash operation?

• If we perform multiple hashing rounds:
  – An attacker would need significantly more resources per cracking attempt
  – A server can still cope with the increased load because users are not authenticating all at the same time

• Standardized multi-round hashing algorithms
  – PBKDF2, brypt, scrypt
PBKDF2 + HMAC-SHA-256

PBKDF2 using XOR to combine 10,000 successive HMAC-SHA-256 outputs into a final hash

Image source: https://nakedsecurity.sophos.com/2013/11/20/serious-security-how-to-store-your-users-passwords-safely/
Demo: Cracking speed
Back to users – 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 less 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
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. Squeal like a pig!!! Do not write anything down."
Password memorability

• Typically, **strength** of a password and **memorability** are working against each other
  – You can likely remember “jack123” better than “399%(mJjaweeee”

• Various attempts have been made to come up with clever schemes for strong memorable passwords
  – “Abandon hope all ye who enter here” =>
  – aHaYwEh =>
  – aHaYvv3h
How People Use Passwords

• Write them down
  – Password managers attempt to make this okay

• Use a single password at multiple sites
  – Do you use the same password for Amazon and your bank account? Blackboard? 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

- 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
Problems with Security Questions

[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?

• People lie to increase security… then forget the answers
Password Managers

• One place where all your passwords are stored
  – This place is protected with one master password
  – Flavors:
    • Online versus Offline (e.g. LastPass versus KeePass)

• Benefits
  – No need to remember any more passwords (other than the master phrase)
  – Unique password per website (no more password reuse)
  – Most password managers also have their own password generators to automatically create strong passwords

• Disadvantages
  – Single-point of failure
    • This can be easily mitigated by storing multiple copies of the database
  – Lock yourself out
    • If you forget your master password, there is no way to recover passwords
  – Cannot authenticate to services if you don’t have access to the password manager
Replay attacks and possible solutions

• The standard, password-based authentication is vulnerable to replay attacks
  – A network attacker can see the password in traffic, and then later reuse to authenticate as the victim

• We can encrypt the entire channel to protect against this (explore this later in class) but we can also tackle it with one-time passwords (OTP)
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
Making passwords stronger

• Passwords belong to the “what you know” category...

• Using “what you have” to strengthen the overall security of a system

• When more than techniques are used for authentication, then we have multiple-factor authentication
  – E.g. 2 Factor Authentication: password + phone
Something you have

• Things one can have
  – Access to your smartphone
    • Has gained a lot of traction recently due to popular web applications (Gmail, Twitter, etc.) supporting it
  – A bank card
  – A security token
    • A piece of hardware containing crypto that either generates one-time passwords or does a challenge-response protocol
  – A badge

• Problems
  – Stolen / forgotten / lost / duplicated
    • Higher cost to change than passwords
  – Cost of user education and support
Something you have - SMS

- Text messages (SMS) as a 2-factor authentication method is falling out of favor.
  - NIST (National Institute of Standards and Technology) has mentioned that it is deprecated and when possible, services should use hardware tokens or smartphone apps to deliver codes

- Reasons
  - Too many incidents of attackers social engineering phone companies into sending them SIM cards because the real owner “lost their phone”
  - Telcos in authoritarian governments can cooperate with their governments
  - Phone networks and their protocols are not exactly the most secure ones

- Moral of the story
  - Use when possible something other than SMS for 2FA
  - SMS-based 2FA is still *MUCH* better than just password-based authentication
SecurID

Setup: generate random key

Counter: 0 1 ...

v = F(KEY, 0)

v = F(KEY, 1)

Alice

Bob

v = F(KEY, 0)

v = F(KEY, 1)

Verifies v = F(KEY, 0)

Verifies v = F(KEY, 1)

• Advancing the counter
  – Time-based (60 seconds) or every button press
• Allow for skew in the counter value
  – +/- 1-minute clock skew by default
  – Allowed up to +/-10 minute skew but requires extra code after the first one

RSA uses symmetric encryption
Input: 128-bit key, 64-bit ctr
Output: 6-digit value
Something you are

• **Biometrics**
  – Fingerprints
  – Palms
  – Face
  – Iris/Retina scanning
  – Voice
  – How you walk? How you type? How you swipe?
    • Research in continuous authentication

• **Benefits**
  – Nothing to remember
  – Passive (nothing to type, always carrying them around)
  – Can’t share
  – Can be fairly unique
Figure 3.8 A Generic Biometric System. Enrollment creates an association between a user and the user's biometric characteristics. Depending on the application, user authentication either involves verifying that a claimed user is the actual user or identifying an unknown user.
Problems

• Less socially acceptable
  – Put your eye in a large machine that blows air and shoots lasers

• Revocability
  – You can change a password but how do you change your fingerprint?

• Are still spoofable
  – E.g. Pick fingerprints from objects and create molds

• Cost
  – Need special devices to read them
  – Human personnel to support them

• Major difference with something you know/something you have?
  – Probability of you being you, rather than certainty
Biometric Error Rates (Benign)

• “Fraud rate” vs. “insult rate”
  – Fraud = system accepts a forgery (false accept)
  – Insult = system rejects valid user (false reject)

• Increasing acceptance threshold increases fraud rate, decreases insult rate

• For biometrics, U.K. banks set target fraud rate of 1%, insult rate of 0.01%  [Ross Anderson]
Figure 3.9 Profiles of a Biometric Characteristic of an Imposter and an Authorized Users In this depiction, the comparison between presented feature and a reference feature is reduced to a single numeric value. If the input value (s) is greater than a preassigned threshold (t), a match is declared.
'Fake fingerprint' Chinese woman fools Japan controls

A Chinese woman managed to enter Japan illegally by having plastic surgery to alter her fingerprints, thus fooling immigration controls, police claim.

Lin Rong, 27, had previously been deported from Japan for overstaying her visa. She was only discovered when she was arrested on separate charges.

Tokyo police said she had paid $15,000 (£9,000) to have the surgery in China.

It is Japan’s first case of alleged biometric fraud, but police believe the practice may be widespread.

Japanese police suspect Chinese brokers of taking huge sums to modify fingerprints surgically.

Local media reports said Ms Lin had undergone surgery to swap the fingerprints from her right and left hands.

Skin patches on her thumbs and index fingers were removed and then re-grafted on to the matching digits of the opposite hand.
Stealing Biometrics

Malaysia car thieves steal finger

By Jonathan Kent
BBC News, Kuala Lumpur

Police in Malaysia are hunting for members of a violent gang who chopped off a car owner's finger to get round the vehicle's hi-tech security system.

The car, a Mercedes S-class, was protected by a fingerprint scanning system.
Hacker fakes German minister's fingerprints using photos of her hands

Jan Krissler used high resolution photos, including one from a government press office, to successfully recreate the fingerprints of Germany's defence minister.
User-aspects

• Never forget that users are a critical part of securing an infrastructure
  – No matter how good your technology is, users can still ruin everything if someone convinces them that it is “okay”
• Abusing the trust of users: **social engineering**
• You can try to contain it by educating your personnel and setting up standard procedures
  – We will never ask you for your password over email
Path of least resistance

http://xkcd.com/538/
Social engineering

MOVIE HACKING...

IF I CAN JUST OVERCLOCK THE UNIX DJANGO, I CAN BASIC THE DDOS ROOT. DAMN. NO DICE. BUT WAIT... IF I DISENCRYPT THEIR KILOBYTES WITH A BACKDOOR HANDSHAKE THEN... JACKPOT.

REAL HACKING...

HI, THIS IS ROBERT HACKERMAN. I'M THE COUNTY PASSWORD INSPECTOR.

HI BOB! HOW CAN I HELP YOU TODAY?

Recap

• Authentication is the first step to security
  – Before we know whether someone is allowed to do something (authorization) we must first be certain of who they are

• Authentication through
  – What you know (Benefits/Drawbacks)
    • Proper practices for storing secrets at the server side
  – What you have (Benefits/Drawbacks)
  – Who you are (Benefits/Drawbacks)
Questions?
Credits

• Original slides by Vitaly Shmatikov
• Modified and expanded by Nick Nikiforakis