Week 2 - Monday

COMP 4290

Last time

- What did we talk about last time?
- Attackers
- Controls

Questions?

Assignment 1

Project 1

Security tidbit of the day

- Text computer analysis can be done to compare writing styles
  - Such an analysis confirmed that Robert Galbraith was actually J. K. Rowling
  - It's called **forensic linguistics**
- Can your coding style be analyzed to see if you are the author of some source code?
  - Sure! Programmers have quirks that make it possible to guess that two programs were written by the same person

Security tidbit continued

- But what about the binaries of your code?
  - Yes! Even the **compiled version** of the code is distinctive enough
  - Recent research was able to de-anonymize code from 100 programmers with 96% accuracy
  - They de-anonymized 600 programmers with 83% accuracy
  - They were even able to de-anonymize optimized code from 100 programmers with 89% accuracy
  - Don't forget that compilers do crazy stuff to code!
- So, if you write a virus, it might one day be possible for people to know it's you just from the executable
- Read the paper:
  - https://faculty.washington.edu/aylin/papers/caliskan_when.pdf

Authentication

Authentication vs. identification

- **Identification** is asserting who someone is
- Your identity includes your name, your bank account numbers, your e-mail addresses, TikTok handle, and anything else linked to you
- Identification is not as strong as authentication, which requires proof

Definition of authentication

- **Authentication** is proving an identity
  - Example: Bill Gates (external entity) is a registered user whose identity on this system is **`gatesw`** (identity of system subject)
- The external identity must provide information to authenticate based on
  1. What the entity knows (passwords)
  2. What the entity has (security badge)
  3. What the entity is (fingerprints or voice ID)

Passwords

Passwords

- Passwords are one of the most common forms of authentication mechanisms based on what the entity knows
- The password represents **authentication information** that the user must know
- The system keeps **complementation information** that can be used to check the password
- Real systems generally do not store passwords in the clear but store hashes of them
- Unix chooses one of 4,096 different hash functions, hashes the password into an 11-character string, and then prepends 2 characters specifying which hash function was used

Difficulties using passwords

- Use
  - Supplying passwords for each access is tedious
- Disclosure
  - If someone else learns a password, it provides no protection and has to be changed, requiring all users to get a new password
- Revocation
  - Changing the password also requires all users to get a new password
- Loss
  - Secure systems usually cannot tell you what your password was, requiring you to choose a new one

Attacking a password system

- A **dictionary attack** is an attack based on guessing the password from trial and error
  - A dictionary attack can work on the complementary information (hashes of passwords)
  - If this information is unavailable, a dictionary attack can directly attack the authentication functions (literally trying to log in repeatedly)
- Let ***P*** be the probability that an attacker guesses the password over a certain span of time
- Let ***G*** be the number of guesses that can be made per unit time
- Let ***T*** be the number of time units of guessing
- Let ***N*** be the number of possible passwords
- Then,

$$P \geq \frac{TG}{N}$$

Random passwords

- One way of protecting against attacks is by making an attacker search the largest possible number of passwords
- You can maximize this time by making all passwords in the set of possible passwords equally likely
- To do this, you use a strong source of randomness to generate your password
- Advantages and disadvantages?

Pronounceable passwords

- Because it is difficult to memorize truly random passwords, randomly generating **pronounceable** passwords is sometimes used instead
- A pronounceable password is one made up of a string of random syllables that can be pronounced together
  - helgoret
  - juttelon
- It is not difficult to write a computer program to produce a string of pronounceable phonemes
- Advantages and disadvantages?

User selection of passwords

- Instead of either of the previous methods for randomly generating passwords, most systems allow users to pick their own passwords
- Unfortunately, users are notoriously bad at picking passwords
  - Everyone picks "babygirl" or, worse, "password"
- **Proactive password checkers** allow users to pick passwords but reject them if they violate certain conditions

Easy to guess passwords

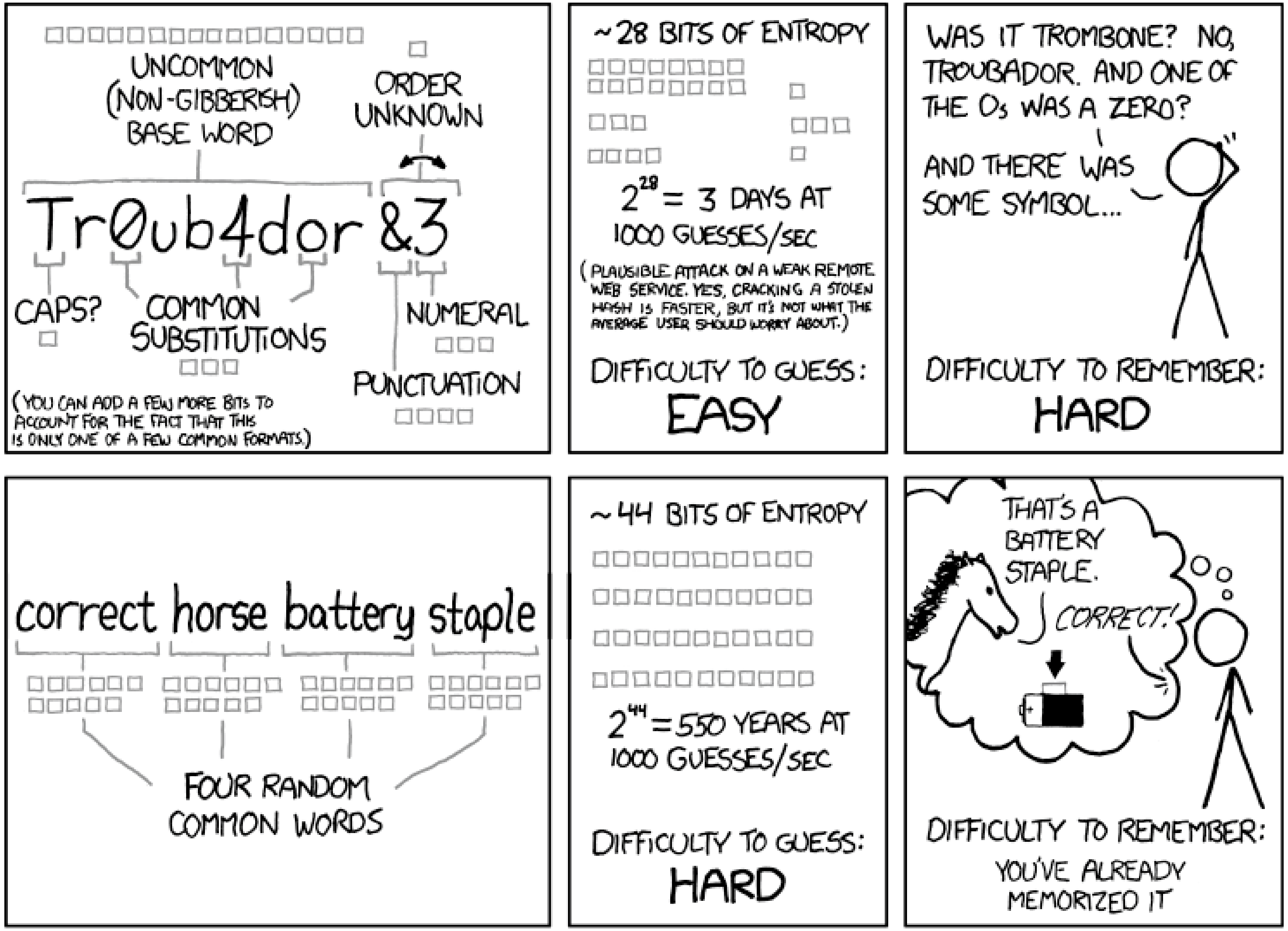
1. Passwords based on account names
2. Passwords based on user names
3. Passwords based on computer names
4. Dictionary words (and reversed versions)
5. Dictionary words with some or all letters capitalized (and reversed versions)
6. Dictionary words with some letters turned into control characters or 1337 substitutions
7. Conjugations of dictionary words
8. Keyboard patterns
9. Passwords shorter than 6 characters
10. Passwords containing only digits
11. Passwords containing just letters, letters and numbers, or letters and punctuation
12. Passwords that look like license plate numbers
13. Acronyms
14. Past passwords
15. Concatenations of dictionary words
16. Dictionary words with digits, punctuation, or spaces preceding or following
17. Dictionary words with all vowels deleted
18. Dictionary words with white spaces deleted
19. Passwords too similar to the previous password

Traditional advice on good passwords

- Researchers suggest a password should have at least one digit, one letter, one punctuation symbol, and (ideally) one control character (not possible in many environments)
- Relatively strong passwords can be generated by taking an unusual phrase or line of a poem and taking (say) the third letter out of each word, leaving in punctuation, and capitalizing some letters according to a rule

XKCD fights back

- The author of XKCD makes an argument that longer passwords using only simple dictionary words are better



UNCOMMON (NON-GIBBERISH) BASE WORD
ORDER UNKNOWN
Tr0ub4dor&3
CAPS?
COMMON SUBSTITUTIONS
NUMERAL
PUNCTUATION
(YOU CAN ADD A FEW MORE BITS TO ACCOUNT FOR THE FACT THAT THIS IS ONLY ONE OF A FEW COMMON FORMATS.)
~28 BITS OF ENTROPY
2^28 = 3 DAYS AT 1000 GUESSES/SEC
(PLAUSIBLE ATTACK ON A WEAK REMOTE WEB SERVICE. YES, CRACKING A STOLEN HASH IS FASTER, BUT IT'S NOT WHAT THE AVERAGE USER SHOULD WORRY ABOUT.)
DIFFICULTY TO GUESS: EASY
WAS IT TROMBONE? NO, TROUBADOR. AND ONE OF THE Os WAS A ZERO?
AND THERE WAS SOME SYMBOL...
DIFFICULTY TO REMEMBER: HARD
correct horse battery staple
FOUR RANDOM COMMON WORDS
~44 BITS OF ENTROPY
2^44 = 550 YEARS AT 1000 GUESSES/SEC
DIFFICULTY TO GUESS: HARD
THAT'S A BATTERY STAPLE.
CORRECT!
DIFFICULTY TO REMEMBER: YOU'VE ALREADY MEMORIZED IT

THROUGH 20 YEARS OF EFFORT, WE'VE SUCCESSFULLY TRAINED EVERYONE TO USE PASSWORDS THAT ARE HARD FOR HUMANS TO REMEMBER, BUT EASY FOR COMPUTERS TO GUESS.

XKCD isn't enough

- This comic was popular enough that it affected many user's habits
- Unfortunately, using only four words isn't enough: try six
- The biggest problem is that you should *never* think up the words yourself
  - Humans are pretty terrible at being random
- Instead, they should be pulled randomly from a large dictionary
- Here are some thoughts about passwords by someone who cracked an XKCD-style password of a systems administrator: https://www.unix-ninja.com/p/your_xkcd_passwords_are_pwned

Proactive password checker criteria

- To be a solid proactive password checker, research suggests it must meet certain criteria:
  1. It must always be used
  2. It must be able to reject easily guessed passwords
  3. It must discriminate on a per-user basis (checking family names and birthdays, etc.)
  4. It must discriminate on a per-site basis (no commonly used site acronyms)
  5. It should have a pattern matching facility to catch bad passwords like "aaaaa"
  6. It needs the ability to execute other programs as subroutines
  7. It should be easy to set up

Salting

- Some attackers are looking for any password instead of trying to find a specific password
- If they have access to the file with the hashes of passwords, they have much less searching to do if the total number of accounts is large (some hash will match, even if the password doesn't)
- For this case, **salting** is used
- Salting adds data to the password in stored form so that an attacker cannot immediately recognize the password
- In Unix, this is a random choice of 4,096 different hashing functions (the specific choice is recorded with the password)
- Other systems can simply add random bits to the end of the password before hashing (which can all be tried at authentication time)
- The book gives an example where the user name is combined with the password before hashing
- Salting has little or no impact on an attack against a single password

Attacking authentication functions

- In many cases, attackers do not have access to the complementation functions (the raw hash values or the hash functions)
- Instead, they must attack the authentication functions themselves
- In these situations, authentication functions can be protected by one of several common techniques

Defending authentication functions

- **Backoff**
  - Force the user to wait longer and longer between failed authentication techniques
  - Exponential backoff means that the first time waits 1 second before allowing a user to log in, the second waits 2 seconds, the third waits 4 seconds, etc.
- **Disconnection**
  - If the connection is remote and requires significant time to connect (dialing, VPN, etc.), the system can simply break connection after a number of failed attempts
- **Disabling**
  - With ***n*** failed attempts, an account is locked until an administrator resets the account
- **Jailing**
  - In jailing, the user is allowed to enter a fake system that looks like the real one
  - In theory, jailing can be used to learn more about an attacker's goals
  - Attractive data (called honeypots) can be made available, tempting the attacker to spend more time on the system (until he can be caught)

Password aging

- Password aging is the idea that passwords should be changed in approximately the amount of time it would take to guess them
- This concept fuels the requirement that we change our Outlook passwords frequently
- In principle, this is a sound security idea
- In practice, over-frequent (or unwarned) password expirations cause user discontent and unconstructive behavior (changing passwords minimally or writing new passwords on Post-It notes)
- It's more important to have a different password for each site and account
- Unless you can remember hundreds of strong passwords, you need a password manager

Challenge Response

Pass Algorithms

- Some systems have a special function f a user (or user's system) must know
- Thus, the system will give the user a prompt, and the user must respond
- Perhaps the system would issue a random value to the user, who must then encrypt it with his secret key and send it back to the user
- Perhaps it's just some other way of processing the data
- Monkey Island 2: LeChuck's Revenge hand puzzle

Security questions

- Security questions represent a kind of challenge and response
- However, security questions are often easy to defeat since someone with a little knowledge about you might know or be able to guess the answers
- Security experts suggest that you treat security questions as secondary passwords and create strong passwords for them as well

Problems with security questions

- Sarah Palin's personal e-mail account was hacked during the 2008 presidential campaign
- Her e-mail gov.palin@yahoo.com had previously been publicized in the news
- The attacker, 20-year-old college student David Kernell impersonated Palin, claimed to have forgotten the password, and answered her security questions, all widely known facts
  - Birth date
  - Zip code
  - Where she met her husband
- He changed the password to `popcorn`

One-Time Passwords

- A one-time password is invalidated as soon as it is used
- Thus, an attacker stealing the password can do limited damage
  - They can only log in once
  - They have to act quickly before the legitimate user logs in first
- How do you generate all these passwords?
- How do you synchronize the user and the system?

One-time password implementations

- RSA SecurIDs change the password every 30 or 60 seconds
- The user must be synchronized with the system within a few seconds to keep this practical
- Using a secure hash function, we start with a seed value ***k***, then
  - $\boldsymbol{h(k) = k_1, h(k_1) = k_2, \ldots, h(k_{n-1}) = k_n}$
- Then passwords are in reverse order
  - $\boldsymbol{p_1 = k_n, p_2 = k_{n-1}, \ldots\ p_{n-1} = k_2, p_n = k_1}$
- Similar systems are used for two-factor authentication apps

Biometrics

Biometrics

- **Biometrics** means identifying humans by their physical and biological characteristics
- This technology is often seen in spy and science fiction movies
- Although growing more common, it's far from perfect
- Like passwords, the actual biometric scans are usually not stored
  - Instead specific features are stored for later comparison
- Biometrics pose unique privacy concerns because the information collected can reveal health conditions

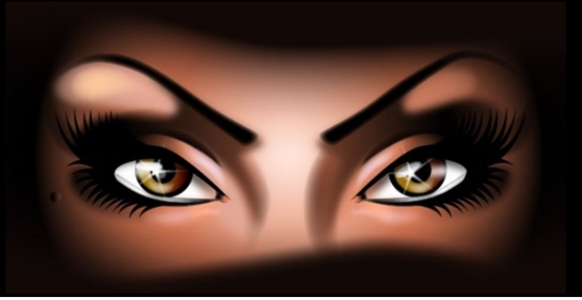
Fingerprints

- Historically, fingerprints are one of the most heavily used forms of biometric identification
  - Especially useful for solving crimes
  - Even identical twins have different fingerprints
  - Fun fact: Koalas have fingerprints so similar to human beings that even experts are fooled
- Optical scanners are available
- Cheap, capacitive scanners are now even available on many laptops and phones
- The image of the fingerprint is usually not stored
- Instead, specific, differentiable features are recorded

Voice recognition

- Voice recognition systems must be trained on your voice
- They can be defeated with recording devices
- If you have a cold, it throws off the characteristics of your voice
- As a consequence, they are particularly susceptible to both false positives and false negatives

Eye recognition



- As the technology matures and hardware becomes cheaper, eye recognition is becoming more common
- Iris recognition looks at the patterns of light and dark areas in your iris (the colored part of your eye)
  - For simplicity, the image is converted to grayscale for comparison
  - Newer iris scanners can make successful identifications at 10 feet away or more, even correcting for glasses!
- Retina scans exist but are unpopular
  - The retina is the tissue lining the inside of your eye and requires pupil dilation to get an accurate picture, blinding you for several minutes
- There are even systems for recognizing the patterns of discolorations on the whites of your eyes

Face recognition

- The shape of your face, the distance between your eyes and nose, and other facial features are relatively distinctive
  - Although they can be nearly the same for identical twins
- Computer vision techniques must be used to locate the face, deal with changes in haircut, glasses, etc.
- Participants must have a neutral facial expression or results can be thrown off
- The US Department of State uses facial recognition and fingerprinting to document foreigners entering the country
  - Although it's for law enforcement, not authentication, the FBI has a database of more than 640 million photos for face recognition
- Phones and computers now use this technology commonly

Other biometrics

- Hand geometry readers measure the shape of your hand
- Keystroke dynamics are the patterns that you use when typing
  - Users are quite distinctive, but distractions and injuries can vary patterns a lot
- Combinations of different biometrics are sometimes used
- DNA sequencing is not (yet) fast enough to be used for authentication
- Researchers are finding new biometrics to use

Problems with biometrics

- People assume that they are more secure than they are
- Attacks:
  - Fingerprints can be lifted off a champagne glass
  - Voices can be recorded
  - Iris recognition can be faked with special contact lenses
- Both false positives and false negatives are possible
- Disabilities can prevent people from using some kinds of biometrics
- It's possible to tamper with transmission from the biometric reader
- Biometric characteristics can change
- Identical twins sometimes pose a problem
- Some people find them intrusive

False positives and false negatives

	Is the Person Claimed	Is Not the Person Claimed
Test is Positive	***a***	***b***
Test is Negative	***c***	***d***

- **Sensitivity** is positive results among correct matches
  - $a / (a + c)$
- **Specificity** is negative results among people who are not sought
  - $d / (b + d)$
- **Accuracy** is how often the test is correct
  - $(a + d) / (a + c + b + d)$
- **Prevalence** is how common a condition is
  - $(a + c) / (a + c + b + d)$

Upcoming

Next time…

- Finish authentication
- Access control
- Adam Garantche presents

Reminders

- Read Section 2.2
- Look at Project 1