3. GAINING ACCESS ... POSSIBLE WAYS

ACCESS CONTROL

Two parts to access control

Authentication: Who goes there?

- Determine whether access is allowed
- Authenticate human to machine
- Authenticate machine to machine

Authorization: Are you allowed to do that?

- Once you have access, what can you do?
- Enforces limits on actions

Note: Access control often used as synonym for authorization

AUTHENTICATION

WHO GOES THERE?

How to authenticate a human to a machine?

Can be based on…

- Something you **know**
  - For example, a password
- Something you **have**
  - For example, a smartcard
- Something you **are**
  - For example, your fingerprint

SOMETHING YOU KNOW

Passwords

Lots of things act as passwords!

- PIN
- Social security number
- Mother’s maiden name
- Date of birth
- Name of your pet, etc.

TROUBLE WITH PASSWORDS

“Passwords are one of the biggest practical problems facing security engineers today.”

“Passwords are the first line of defense against cyber criminals”

WHY PASSWORDS?

Why is “something you know” more popular than “something you have” and “something you are”?

Cost: passwords are free

Convenience: easier for admin to reset pwd than to issue user a new thumb

KEYS VS PASSWORDS

Crypto keys

Suppose a key is 64 bits long

Then 2^{64} keys

Choose key at random

Then attacker must try about 2^{63} keys

Passwords

Suppose passwords are 8 characters long, and 256 different characters

Then $256^8 = 2^{64}$ pwds

Users do **not** select passwords at random

Attacker has far less than 2^{63} passwords to try (**dictionary attack**)

GOOD AND BAD PASSWORDS

Bad passwords

- frank
- Fido
- password
- 4444
- Pikachu
- 102560
- AustinStamp

Good Passwords?

- jflej,43j-EmmL+y
- 09864376537263
- P0kem0N
- FSa7Yago
- 0nceuP0nAt1m8
- PokeGCTall150

PASSWORD EXPERIMENT

Three groups of users — each group advised to select passwords as follows

- **Group A:** At least 6 chars, 1 non-letter
- winner → **Group B:** Password based on passphrase
- **Group C:** 8 random characters

Results

- **Group A:** About 30% of pwds easy to crack
- **Group B:** About 10% cracked
  - Passwords easy to remember
- **Group C:** About 10% cracked
  - Passwords hard to remember

PASSWORD EXPERIMENT

User compliance hard to achieve

In each case, 1/3rd did not comply (and about 1/3rd of those easy to crack!)

Assigned passwords sometimes best

If passwords not assigned, best advice is

- Choose passwords based on passphrase
- Use pwd cracking tool to test for weak pwds
- Require periodic password changes?

ATTACKS ON PASSWORDS

Attacker could…

- Target one particular account
- Target any account on system
- Target any account on any system
- Attempt denial of service (DoS) attack

Common attack path

- Outsider → normal user → administrator
- May only require **one** weak password!

PASSWORD FILE

Bad idea to store passwords in a file

But need a way to verify passwords

Cryptographic solution: **hash** the passwords

- Store y = h(password)
- Can verify entered password by hashing
- If attacker obtains password file, he does not obtain passwords
- But attacker with password file can guess x and check whether y = h(x)
- If so, attacker has found password!

DICTIONARY ATTACK

Attacker pre-computes $h(x)$ for all x in a **dictionary** of common passwords

Suppose attacker gets access to password file containing hashed passwords

- Attacker only needs to compare hashes to his pre-computed dictionary
- Same attack will work each time

Can we prevent this attack? Or at least make attacker's job more difficult?

PASSWORD FILE

Store hashed passwords

Better to hash with **salt**

Given password, choose random s, compute

$$y = h(\text{password}, s)$$

and store the pair (s,y) in the password file

Note: The salt s is **not secret**

Easy to verify password

Attacker must recompute dictionary hashes for each user — lots more work!

PASSWORD CRACKING: DO THE MATH

Assumptions

Pwds are 8 chars, 128 choices per character

- Then $128^8 = 2^{56}$ possible passwords

There is a **password file** with 2^{10} pwds

Attacker has **dictionary** of 2^{20} common pwds

Probability of 1/4 that a pwd is in dictionary

Work is measured by number of hashes

OTHER PASSWORD ISSUES

Too many passwords to remember

- Results in password reuse
- Why is this a problem?

Who suffers from bad password?

- Login password vs ATM PIN

Failure to change default passwords

Social engineering

Error logs may contain “almost” passwords

Bugs, keystroke logging, spyware, etc.

PASSWORDS

The bottom line

Password cracking is too easy!

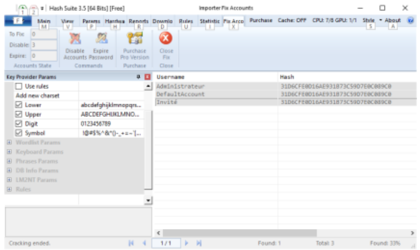
- One weak password may break security
- Users choose bad passwords
- Social engineering attacks, etc.

The bad guy has all of the advantages

All of the math favors bad guys

Passwords are a **big** security problem

PASSWORD CRACKING TOOLS



To Fix: 0
Disable: 3
Expire: 0

☐ Use rules
Add new charset
☑ Lower abcdefghijklmnopqrs...
☑ Upper ABCDEFGHIJKLMNO...
☑ Digit 0123456789
☑ Symbol !@#$%^&*()-_+=~`|...

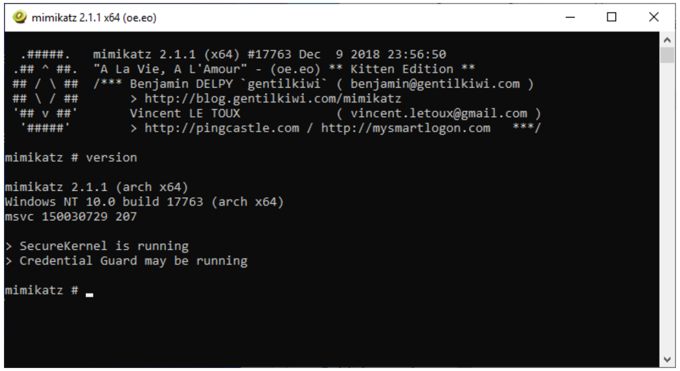
Username	Hash
Administrateur	31D6CFE0D16AE931B73C59D7E0C089C0
DefaultAccount	31D6CFE0D16AE931B73C59D7E0C089C0
Invité	31D6CFE0D16AE931B73C59D7E0C089C0

Cracking ended. 1 / 1 Found: 1 Total: 3 Found: 33%

Popular password cracking tools

- L0phtCrack (Windows) -http://www.l0phtcrack.com/
- John the Ripper (Unix) - https://www.openwall.com/john/ and HashSuite (Windows) - https://hashsuite.openwall.net
- Mimikatz (Windows) - http://blog.gentilkiwi.com/mimikatz

Admins should use these tools to test for weak passwords since attackers will!



```
  .#####.   mimikatz 2.1.1 (x64) #17763 Dec  9 2018 23:56:50
 .## ^ ##.  "A La Vie, A L'Amour" - (oe.eo) ** Kitten Edition **
 ## / \ ##  /*** Benjamin DELPY `gentilkiwi` ( benjamin@gentilkiwi.com )
 ## \ / ##       > http://blog.gentilkiwi.com/mimikatz
 '## v ##'       Vincent LE TOUX             ( vincent.letoux@gmail.com )
  '#####'        > http://pingcastle.com / http://mysmartlogon.com   ***/

mimikatz # version

mimikatz 2.1.1 (arch x64)
Windows NT 10.0 build 17763 (arch x64)
msvc 150030729 207

> SecureKernel is running
> Credential Guard may be running

mimikatz # _
```

BIOMETRICS

Personal Data
Name
Home Address
Business Address
Identity Card No
Passport No
Driving License
Income Tax No
Confidential Data

SOMETHING YOU ARE

Biometric

- **"You are your key"** — Schneier

- Examples
  - Fingerprint
  - Handwritten signature
  - Facial recognition
  - Speech recognition
  - Gait (walking) recognition
  - "Digital doggie" (odor recognition)
  - Many more!

WHY BIOMETRICS?

Biometrics seen as desirable replacement for passwords

Cheap and reliable biometrics needed

Today, a very active area of research

Biometrics are used in security today

- Thumbprint mouse
- Palm print for secure entry
- Fingerprint to unlock car door, etc.

But biometrics not too popular

- Has not lived up to its promise (yet)

IDEAL BIOMETRIC

Universal — applies to (almost) everyone

- In reality, no biometric applies to everyone

Distinguishing — distinguish with certainty

- In reality, cannot hope for 100% certainty

Permanent — physical characteristic being measured never changes

- In reality, want it to remain valid for a long time

Collectable — easy to collect required data

- Depends on whether subjects are cooperative

Safe, easy to use, etc., etc.

BIOMETRIC MODES

Identification — Who goes there?

- Compare one to many
- Example: The FBI fingerprint database

Authentication — Is that really you?

- Compare one to one
- Example: Thumbprint mouse

Identification problem more difficult

- More “random” matches since more comparisons

We are interested in authentication

ENROLLMENT VS RECOGNITION

Enrollment phase

- Subject's biometric info put into database
- Must carefully measure the required info
- OK if slow and repeated measurement needed
- Must be very precise for good recognition
- A weak point of many biometric schemes

Recognition phase

- Biometric detection when used in practice
- Must be quick and simple
- But must be reasonably accurate

FINGERPRINT HISTORY

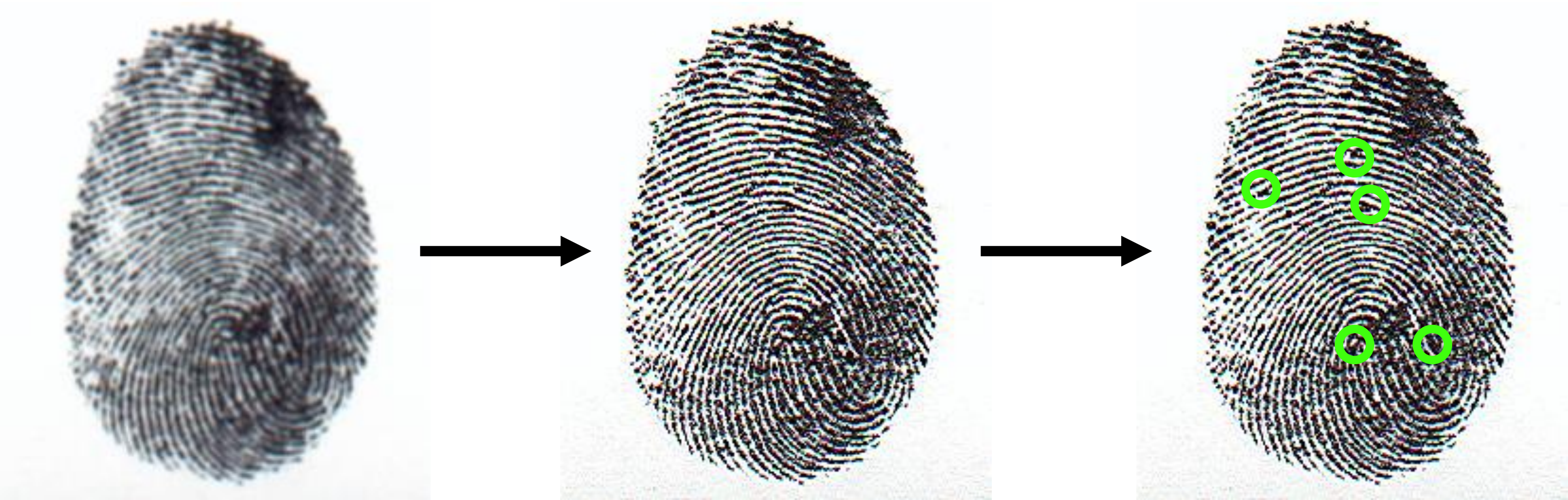
1888 — Sir Francis Galton (cousin of Darwin) developed classification system

- His system of "minutia" is still in use today
- Also verified that fingerprints do not change

Some countries require a number of points (i.e., minutia) to match in criminal cases

- In Britain, 15 points
- In US, no fixed number of points required

FINGERPRINT BIOMETRIC

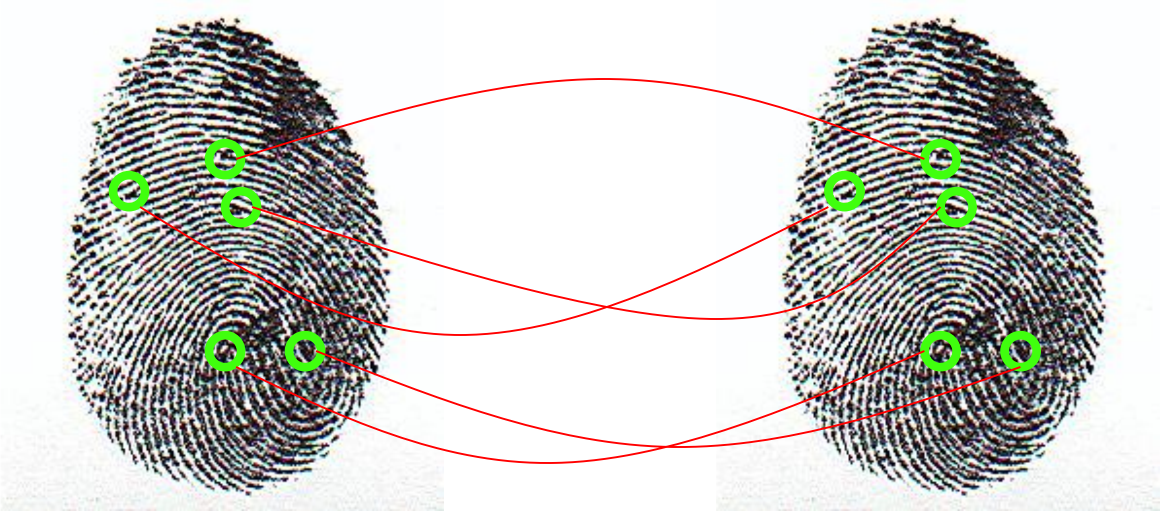


Capture image of fingerprint

Enhance image

Identify minutia

FINGERPRINT BIOMETRIC



Extracted minutia are compared with user’s minutia stored in a database

Is it a statistical match?

HAND GEOMETRY

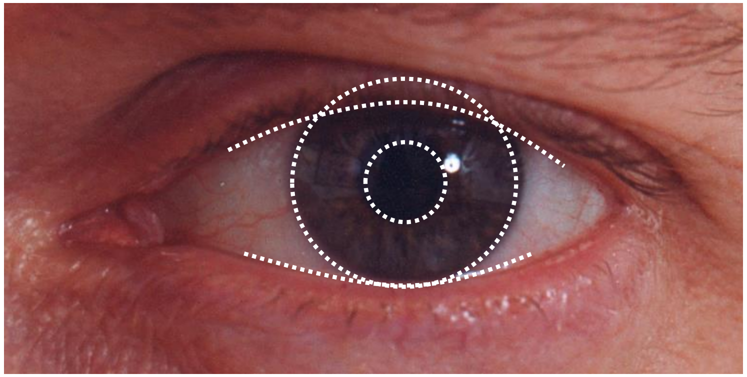
Advantages

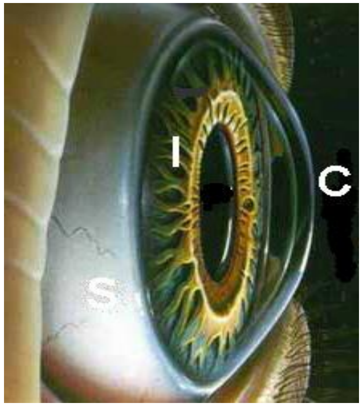
- Quick
- 1 minute for enrollment
- 5 seconds for recognition
- Hands symmetric (use other hand backwards)

Disadvantages

- Cannot use on very young or very old
- Relatively high equal error rate

IRIS PATTERNS





Iris pattern development is “chaotic”

Little or no genetic influence

Different even for identical twins

Pattern is stable through lifetime

MEASURING IRIS SIMILARITY

Based on Hamming distance

Define $d(x,y)$ to be

- # of non match bits/# of bits compared
- $d(0010,0101) = 3/4$ and $d(101111,101001) = 1/3$

Compute $d(x,y)$ on 2048-bit iris code

- Perfect match is $d(x,y) = 0$
- For same iris, expected distance is 0.08
- At random, expect distance of 0.50
- Accept as match if distance less than 0.32

ATTACK ON IRIS SCAN

Good **photo** of eye can be scanned

- Attacker could use photo of eye

To prevent photo attack, scanner could use light to be sure it is a “live” iris

EQUAL ERROR RATE COMPARISON

Fingerprint biometric has ER of about 5%

Hand geometry has ER of about 10^{-3}

In theory, **iris scan** has ER of about 10^{-6}

- But in practice, hard to achieve
- Enrollment phase must be extremely accurate

Most biometrics much worse than fingerprint!

Biometrics useful for authentication…

But ID biometrics are almost useless today

BIOMETRICS: THE BOTTOM LINE

Biometrics are hard to forge

But attacker could

- Steal Alice's thumb
- Photocopy Bob's fingerprint, eye, etc.
- Subvert software, database, "trusted path", …

Also, how to revoke a "broken" biometric?

Biometrics are not foolproof!

Biometric use is limited today

That should change in the future…

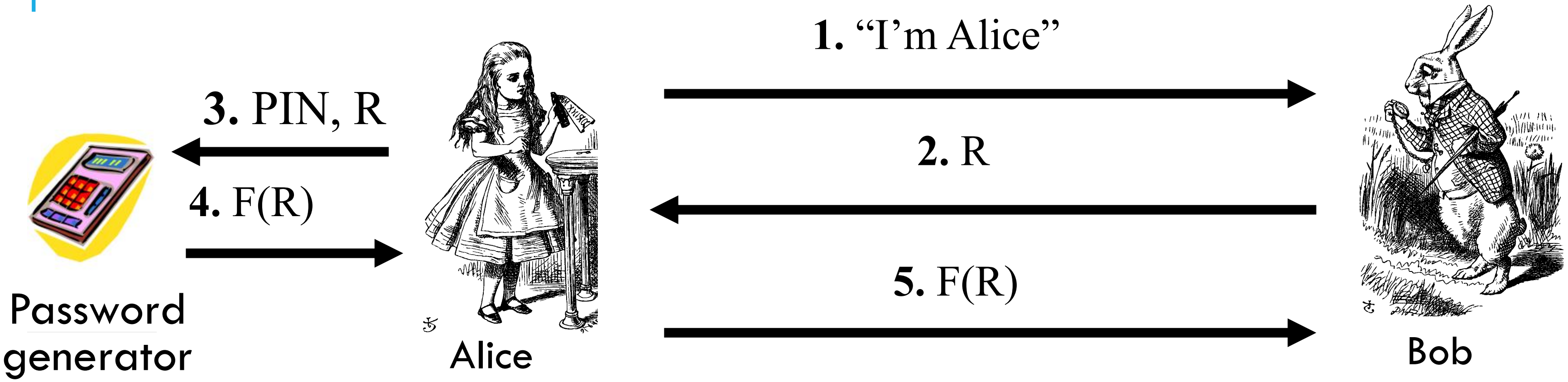
SOMETHING YOU HAVE

Something in your possession

Examples include

- Car key
- Laptop computer
  - Or specific MAC address
- Password generator
  - We'll look at this next
- ATM card, smartcard, etc.

PASSWORD GENERATOR



Alice gets "challenge" R from Bob

Alice enters R into password generator

Alice sends "response" back to Bob

Alice **has** pwd generator and **knows** PINs

2-FACTOR AUTHENTICATION

Requires 2 out of 3 of

1. Something you know
2. Something you have
3. Something you are

Examples

- ATM: Card and PIN
- Credit card: Card and signature
- Password generator: Device and PIN
- Smartcard with password/PIN

SINGLE SIGN-ON

A hassle to enter password(s) repeatedly

- Users want to authenticate only once
- “Credentials” stay with user wherever he goes
- Subsequent authentication is transparent to user

WEB COOKIES

Cookie is provided by a Website and stored on user's machine

Cookie indexes a database at Website

Cookies **maintain state** across sessions

Web uses a stateless protocol: HTTP

Cookies also maintain state within a session

Like a single sign-on for a website

- Though a very weak form of authentication

Cookies and privacy concerns

AUTHORIZATION

AUTHENTICATION VS AUTHORIZATION

Authentication — Who goes there?

- Restrictions on who (or what) can access system

Authorization — Are you allowed to do that?

- Restrictions on actions of authenticated users

Authorization is a form of **access control**

Authorization enforced by

- Access Control Lists
- Capabilities

LAMPSON'S ACCESS CONTROL MATRIX

Subjects (users) index the rows

Objects (resources) index the columns

	OS	Accounting program	Accounting data	Insurance data	Payroll data
Bob	rx	rx	r	---	---
Alice	rx	rx	r	rw	rw
Sam	rwx	rwx	r	rw	rw
Accounting program	rx	rx	rw	rw	rw

ARE YOU ALLOWED TO DO THAT?

Access control matrix has all relevant info

But how to manage a large access control (AC) matrix?

Could be 1000's of users, 1000's of resources

Then AC matrix with 1,000,000's of entries

Need to check this matrix before access to any resource is allowed

Hopelessly inefficient

INFERENCE CONTROL

INFERENCE CONTROL EXAMPLE

Suppose we query a database

- Question: What is average salary of female CS professors at UPS?
- Answer: about 0,000 €
- Question: How many female CS professors at UPS?
- Answer: 50

Specific information has leaked from responses to general questions!

INFERENCE CONTROL AND RESEARCH

For example, medical records are private but valuable for research

How to make info available for research and protect privacy?

How to allow access to such data without leaking specific information?

NAÏVE INFERENCE CONTROL

Remove names from medical records?

Still may be easy to get specific info from such “anonymous” data

Removing names is not enough

- As seen in previous example

What more can be done?

LESS-NAÏVE INFERENCE CONTROL

Query set size control

- Don't return an answer if set size is too small

N-respondent, k% dominance rule

- Do not release statistic if k% or more contributed by N or fewer

Randomization

- Add small amount of random noise to data

Many other methods — none satisfactory

INFERENCE CONTROL: THE BOTTOM LINE

Robust inference control may be impossible

Is weak inference control better than no inference control?

- **Yes**: Reduces amount of information that leaks and thereby limits the damage

Is weak crypto better than no crypto?

- **Probably not:** Encryption indicates important data
- May be easier to filter encrypted data

CAPTCHA

TURING TEST

Proposed by Alan Turing in 1950

Human asks questions to one other human and one computer (without seeing either)

If human questioner cannot distinguish the human from the computer responder, the computer passes the test

The gold standard in artificial intelligence

No computer can pass this today

CAPTCHA

CAPTCHA — **C**ompletely **A**utomated **P**ublic **T**uring test to tell **C**omputers and **H**umans **A**part

Automated — test is generated and scored by a computer program

Public — program and data are public

Turing test to tell… — humans can pass the test, but machines cannot pass the test

Like an inverse Turing test (sort of…)

CAPTCHA PARADOX

"…CAPTCHA is a program that can generate and grade tests that it itself cannot pass…"

"…much like some professors…"

Paradox — computer creates and scores test that it cannot pass!

CAPTCHA used to restrict access to resources to humans (no computers)

CAPTCHA useful for **access control**

CAPTCHA: RULES OF THE GAME

Must be easy for most humans to pass

Must be difficult or impossible for machines to pass

- **Even with access to CAPTCHA software**

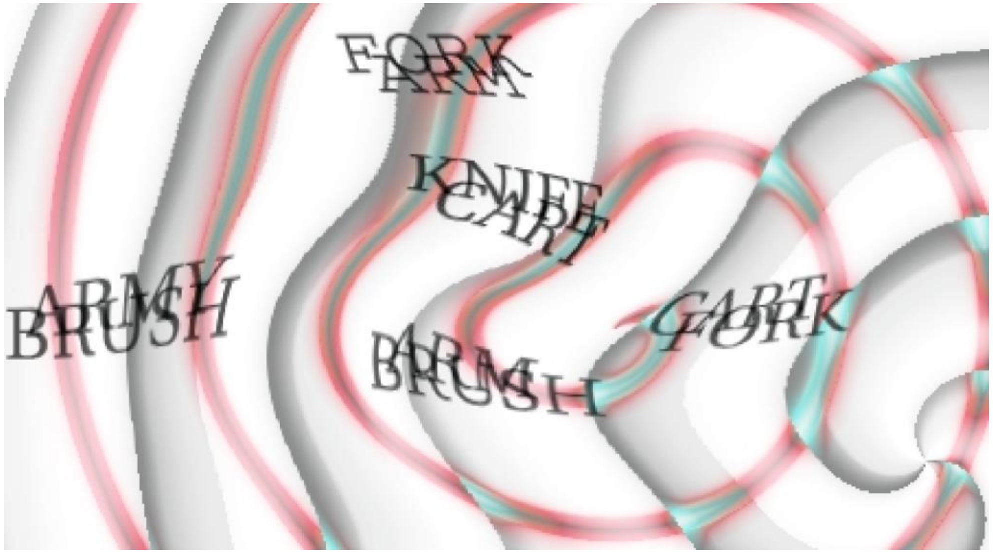
The only unknown is some random number

Desirable to have different CAPTCHAs in case some person cannot pass one type

- Blind person could not pass visual test, etc.

DO CAPTCHAS EXIST?

Test: Find 2 words in the following



- Easy (hum …) for most humans
- Difficult for computers (OCR problem)

CAPTCHAS

Current types of CAPTCHAs

- Visual
  - Like previous example
  - Many others
- Audio
  - Distorted words or music

No text-based CAPTCHAs

- Maybe this is not possible…

CAPTCHA'S AND AI

Computer recognition of distorted text is a challenging AI problem

- But humans can solve this problem

Same is true of distorted sound

- Humans also good at solving this

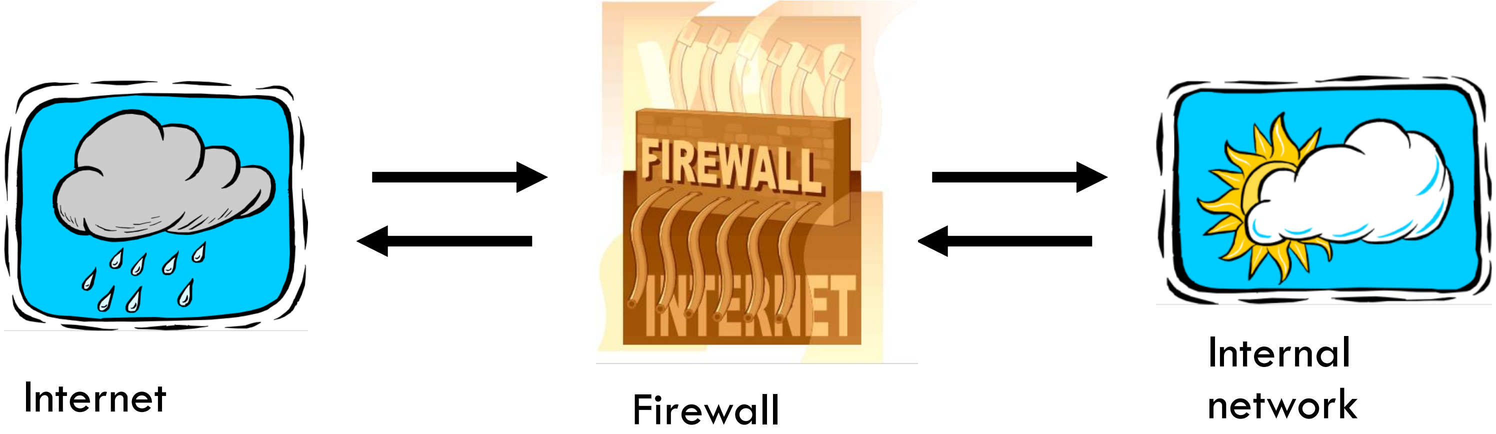
Hackers who break such a CAPTCHA have solved a hard AI problem

Putting hacker's effort to good use!

May be other ways to defeat CAPTCHAs…

FIREWALLS

FIREWALLS



Internet

Firewall

Internal network

Firewall must determine what to let in to internal network and/or what to let out

Access control for the network

FIREWALL AS SECRETARY

A firewall is like a **secretary**

To meet with an executive

- First contact the secretary
- Secretary decides if meeting is reasonable
- Secretary filters out many requests

You want to meet chair of CS department?

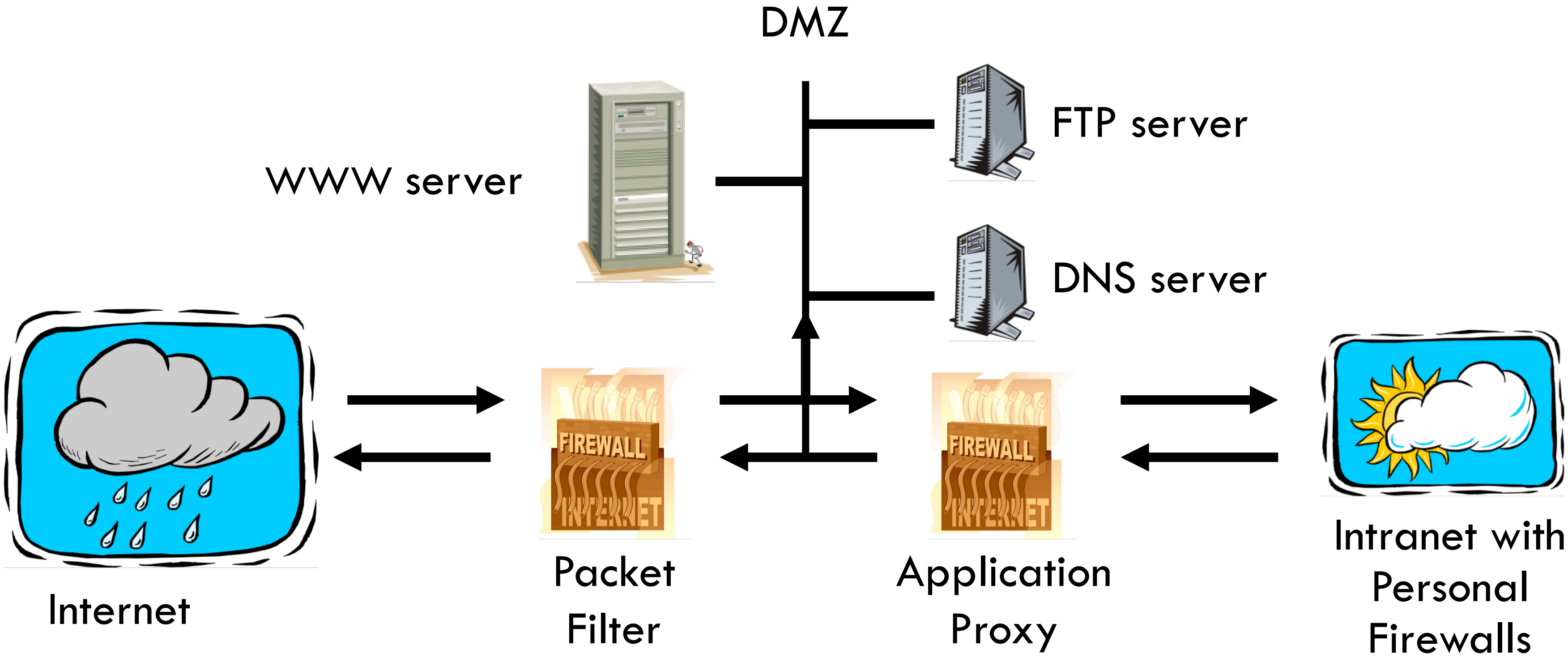
- Secretary does some filtering

You want to meet President Macron?

- Secretary does lots of filtering!

FIREWALLS AND DEFENSE IN DEPTH

Example security architecture



INTRUSION DETECTION SYSTEMS

INTRUSION PREVENTION

Want to keep bad guys out

Intrusion prevention is a traditional focus of computer security

- Authentication is to prevent intrusions
- Firewalls a form of intrusion prevention
- Virus defenses also intrusion prevention

Comparable to locking the door on your car

INTRUSION DETECTION

In spite of intrusion prevention, bad guys will sometime get into system

Intrusion detection systems (IDS)

- Detect attacks
- Look for "unusual" activity

IDS developed out of log file analysis

IDS is currently a very **hot** research topic

How to respond when intrusion detected?

- We don't deal with this topic here

INTRUSION DETECTION SYSTEMS

Who is likely intruder?

- May be outsider who got thru firewall
- May be evil insider

What do intruders do?

- Launch well-known attacks
- Launch variations on well-known attacks
- Launch new or little-known attacks
- Use a system to attack other systems
- Etc.

IDS

Intrusion detection **approaches**

- Signature-based IDS
- Anomaly-based IDS

Intrusion detection **architectures**

- Host-based IDS
- Network-based IDS

Most systems can be classified as above

- In spite of marketing claims to the contrary!

SIGNATURE DETECTION EXAMPLE

Failed login attempts may indicate password cracking attack

IDS could use the rule "N failed login attempts in M seconds" as **signature**

If N or more failed login attempts in M seconds, IDS warns of attack

Note that the warning is specific

- Admin knows what attack is suspected
- Admin can verify attack (or false alarm)

SIGNATURE DETECTION

Advantages of signature detection

- Simple
- Detect known attacks
- Know which attack at time of detection
- Efficient (if reasonable number of signatures)

Disadvantages of signature detection

- Signature files must be kept up to date
- Number of signatures may become large
- Can only detect known attacks
- Variation on known attack may not be detected

ANOMALY DETECTION

Anomaly detection systems look for unusual or abnormal behavior

There are (at least) two challenges

- What is normal for this system?
- How "far" from normal is abnormal?

Statistics is obviously required here!

- The **mean** defines normal
- The **variance** indicates how far abnormal lives from normal

HOW TO MEASURE NORMAL?

How to measure normal?

- Must measure during “representative” behavior
- Must not measure during an attack…
- …or else attack will seem normal!
- Normal is statistical mean
- Must also compute variance to have any reasonable chance of success

HOW TO MEASURE ABNORMAL?

Abnormal is relative to some “normal”

- Abnormal indicates possible attack

Statistical discrimination techniques:

- Bayesian statistics
- Linear discriminant analysis (LDA)
- Quadratic discriminant analysis (QDA)
- Neural nets, hidden Markov models, etc.

Fancy modeling techniques also used

- Artificial intelligence
- Artificial immune system principles
- Many others!

ANOMALY DETECTION (1)

Suppose we monitor use of three commands:

open, read, close

Under normal use we observe that Alice

open,read,close,open,open,read,close,…

Of the six possible ordered pairs, four pairs are "normal" for Alice:

(open,read), (read,close), (close,open), (open,open)

Can we use this to identify unusual activity?

ANOMALY DETECTION (1)

We monitor use of the three commands

open, read, close

If the ratio of abnormal to normal pairs is “too high”, warn of possible attack

Could improve this approach by

- Also using expected frequency of each pair
- Use more than two consecutive commands
- Include more commands/behavior in the model
- More sophisticated statistical discrimination

ANOMALY DETECTION ISSUES

System constantly evolves and so must IDS

- Static system would place huge burden on admin
- But evolving IDS makes it possible for attacker to (slowly) convince IDS that an attack is normal!
- Attacker may win simply by "going slow"

What does "abnormal" really mean?

- Only that there is possibly an attack
- May not say anything specific about attack!
- How to respond to such vague information?

Signature detection tells exactly which attack

ANOMALY DETECTION

Advantages

- Chance of detecting unknown attacks
- May be more efficient (since no signatures)

Disadvantages

- Today, cannot be used alone
- Must be used with a signature detection system
- Reliability is unclear
- May be subject to attack
- Anomaly detection indicates something unusual
- But lack of specific info on possible attack!

ANOMALY DETECTION: THE BOTTOM LINE

Anomaly-based IDS is active research topic

Many security professionals have very high hopes for its ultimate success

Often cited as key future security technology

Anomaly detection is difficult and tricky

Is anomaly detection as hard as AI?