Chapter 8: Authorization

It is easier to exclude harmful passions than to rule them,
and to deny them admittance
than to control them after they have been admitted.
— Seneca

You can always trust the information given to you
by people who are crazy;
they have an access to truth not available through regular channels.
— Sheila Ballantyne

Authentication vs Authorization

- Authentication (身份验证) — Are you who you say you are?
  - Restrictions on who (or what) can access system
- **Authorization** (资质确认) — Are you allowed to do that?
  - Restrictions on actions of authenticated users
- Authorization is a form of **access control**
- But first, we look at system certification…

System Certification

- Government attempt to certify "security level" of products
- Of historical interest
  - Sorta like a history of authorization
- Still important today if you want to sell a product to the government
  - Tempting to argue it's a failure since government is so insecure, but…

Orange Book（橙皮书）

- 可信计算系统评估标准Trusted Computing System Evaluation Criteria (TCSEC), 1983
  - Universally known as the "orange book"
  - Name is due to color of it's cover
  - About 115 pages
  - Developed by U.S. DoD (NSA)
  - Part of the "rainbow series"
- Orange book generated a pseudo-religious fervor among some people
  - Less and less intensity as time goes by

Orange Book Outline

- Goals
  - Provide way to assess security products
  - Provide general guidance/philosophy on how to build more secure products
- Four ***divisions*** labeled D thru A
  - D is lowest, A is highest
- Divisions split into numbered ***classes***

D and C Divisions

- D — minimal protection 最低保护
  - Losers that can't get into higher division
- C — discretionary protection 自主保护, i.e., don't enforce security, just have means to detect breaches (audit)
  - C1 — discretionary security protection
  - C2 — controlled access protection
  - C2 slightly stronger than C1 (both vague)

B Division

- B — mandatory protection 强制保护
- B is a ***huge*** step up from C
  - C: break security, you might get caught
  - B: "mandatory", so you can't break it
- B1 — labeled security protection
  - All data labeled, which restricts what can be done with it
  - This access control cannot be violated

B and A Divisions

- B2 — structured protection
  - Adds covert channel protection onto B1
- B3 — security domains
  - On top of B2 protection, adds that code must be ***tamperproof*** and "small"
- A — verified protection 可验证保护
  - Like B3, but ***proved*** using formal methods
  - Such methods still (mostly) impractical

Orange Book: Last Word

- Also a 2nd part, discusses rationale
- Not very practical or sensible, IMHO
- But some people insist we'd be better off if we'd followed it
- Others think it was a dead end
  - And resulted in lots of wasted effort
  - Aside… people who made the orange book, now set security education standards

Common Criteria

- Successor to the orange book (ca. 1998)
  - Due to inflation, more than 1000 pages
- An international government standard
  - And it reads like it…
  - Won't ever stir same passions as orange book
- CC is relevant in practice, but usually only if you want to sell to the government
- Evaluation Assurance Levels (评估保障等级EALs)
  - 1 thru 7, from lowest to highest security

EAL（Evaluation Assurance Level ）

- Note: product with high EAL may not be more secure than one with lower EAL
  - Why?
- Similarly, product with an EAL may not be any more secure than one without
  - Why?

EAL 1 thru 7

- EAL1 — functionally tested 功能
- EAL2 — structurally tested 结构
- EAL3 — methodically tested, checked 系统地
- EAL4 — ***designed***, tested, reviewed 系统地
- EAL5 — semiformally designed, tested 半形式化
- EAL6 — verified, designed, tested 半形式化
- EAL7 — formally … (blah blah blah) 形式化

Common Criteria

- EAL4 is most commonly sought
  - Minimum needed to sell to government
- EAL7 requires formal proofs
  - Author could only find 2 EAL7 products…
- Who performs evaluations?
  - Government accredited labs, of course (for a hefty fee, like 6 figures)

Authentication vs Authorization

- Authentication — Are you who you say you are?
  - 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**
- Classic view of authorization…
  - Access Control Lists (ACLs 访问控制列表)
  - Capabilities (C-lists 访问能力列表)

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
- Could be 100's of users, 10,000's of resources
  - Then matrix has 1,000,000's of entries
- How to manage such a large matrix?
- Note: We need to check this matrix before access to any resource by any user
- How to make this more efficient/practical?

Access Control Lists (ACLs)

- ACL: store access control matrix by **column**
- Example: ACL for **insurance data** is in **blue**

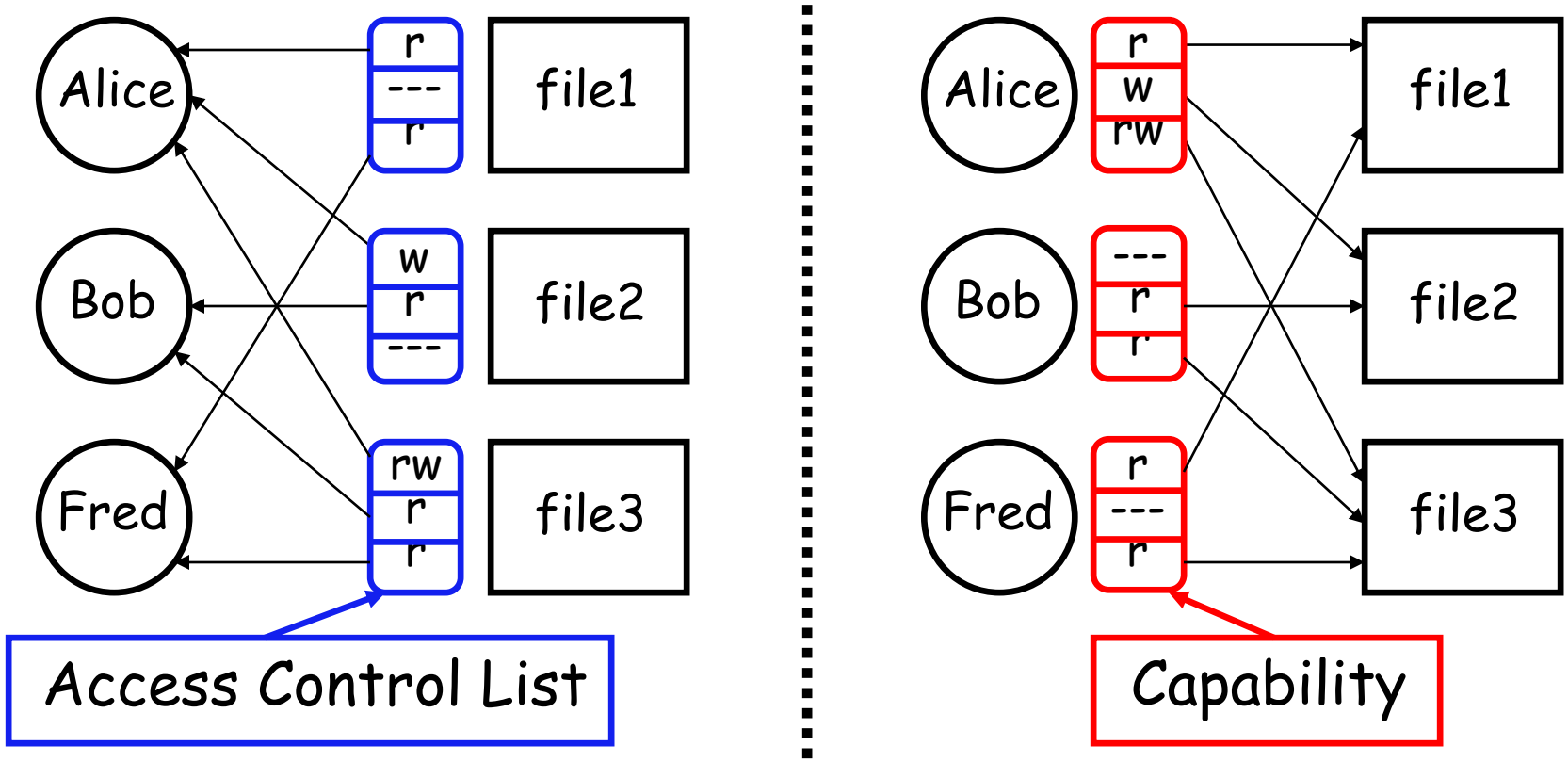
	OS	Accounting program	Accounting data	**Insurance data**	Payroll data
Bob	rx	rx	r	—	—
Alice	rx	rx	r	rw	rw
Sam	rwx	rwx	r	rw	rw
Accounting program	rx	rx	rw	rw	rw

Capabilities (or C-Lists)

- Store access control matrix by **row**
- Example: Capability for **Alice** is in **red**

	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

ACLs vs Capabilities



- Note that arrows point in opposite directions…
- With ACLs, still need to associate users to files

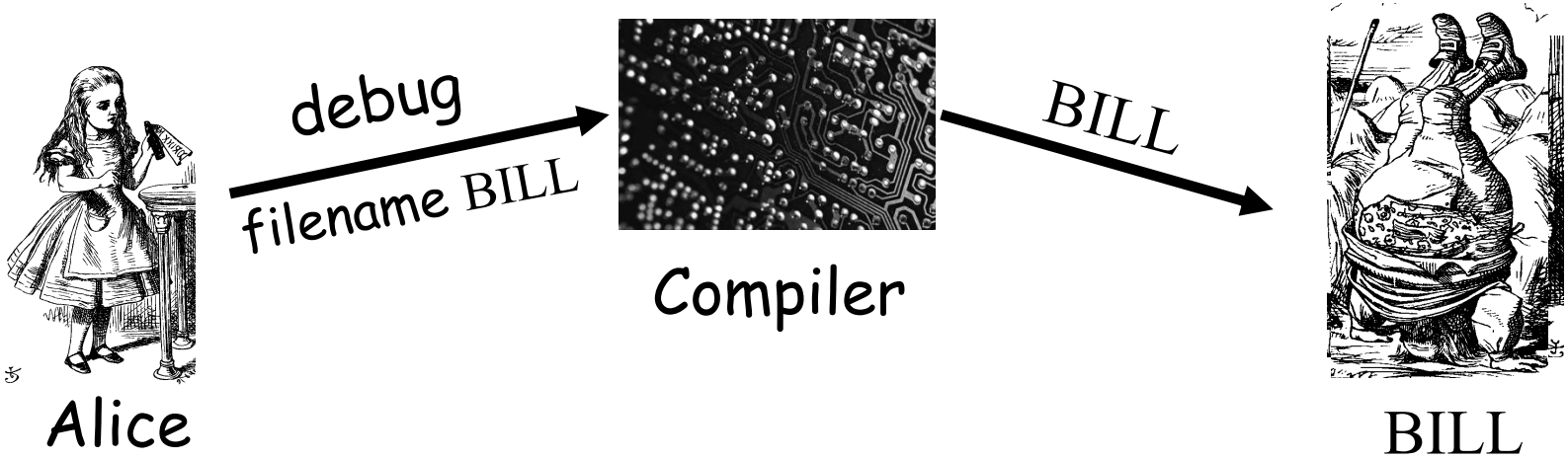
Confused Deputy（混淆代理人）

- Two resources
  - Compiler and BILL file (billing info)
- Compiler can write file BILL
- Alice can invoke compiler with a debug filename
- Alice not allowed to write to BILL

- Access control matrix

	Compiler	BILL
Alice	x	—
Compiler	rx	rw

ACL's and Confused Deputy



- Compiler is **deputy** acting on behalf of Alice
- Compiler is **confused**
  - Alice is not allowed to write BILL
- Compiler has confused its rights with Alice's

Confused Deputy

- Compiler acting for Alice is confused
- There has been a separation of **authority** from the **purpose** for which it is used
- With ACLs, more difficult to prevent this
- With Capabilities, easier to prevent problem
  - Must maintain association between authority and intended purpose
- Capabilities — easy to **delegate** authority

ACLs vs Capabilities

- ACLs
  - Good when users manage their own files
  - Protection is data-oriented
  - Easy to change rights to a resource
- Capabilities
  - Easy to delegate — avoid the confused deputy
  - Easy to add/delete users
  - More difficult to implement
  - The "Zen of information security"
- Capabilities loved by academics
  - Capability Myths Demolished

Multilevel Security (MLS) Models

Classifications and Clearances

- **Classifications** apply to **objects**
- **Clearances** apply to **subjects**
- US Department of Defense (DoD) uses 4 levels:

  TOP SECRET

  SECRET

  CONFIDENTIAL

  UNCLASSIFIED

Clearances and Classification

- To obtain a **SECRET** clearance requires a routine background check
- A **TOP SECRET** clearance requires extensive background check
- Practical classification problems
  - Proper classification not always clear
  - Level of granularity to apply classifications
  - Aggregation — flipside 反面 of granularity

Subjects and Objects

- Let O be an **object**, S a **subject**
  - O has a classification
  - S has a clearance
  - Security **level** denoted L(O) and L(S)
- For DoD levels, we have

  TOP SECRET > **SECRET** >

  CONFIDENTIAL > **UNCLASSIFIED**

Multilevel Security (MLS)

- MLS needed when subjects/objects at different levels access **same system**
- MLS is a form of **Access Control**
- Military and government interest in MLS for many decades
  - Lots of research into MLS
  - Strengths and weaknesses of MLS well understood (almost entirely theoretical)
  - Many possible uses of MLS outside military

MLS Applications

- Classified government/military systems
- **Business example:** info restricted to
  - Senior management only, all management, everyone in company, or general public
- Network firewall
- Confidential medical info, databases, etc.
- Usually, MLS not really a technical system
  - More like part of a legal structure

MLS Security Models

- MLS models explain **what** needs to be done
- Models **do not** tell you **how** to implement
- Models are descriptive, not prescriptive
  - That is, high-level description, not an algorithm
- There are many MLS models
- We'll discuss simplest MLS model
  - Other models are more realistic
  - Other models also more complex, more difficult to enforce, harder to verify, etc.

Bell-LaPadula

- BLP security model designed to express essential requirements for MLS
- BLP deals with **confidentiality**
  - To prevent unauthorized reading
- Recall that O is an object, S a subject
  - Object O has a classification
  - Subject S has a clearance
  - Security level denoted L(O) and L(S)

Bell-LaPadula

- BLP consists of
  - **Simple Security Condition**: S can read O if and only if $L(O) \leq L(S)$
  - ***-Property** (**Star Property**): S can write O if and only if $L(S) \leq L(O)$
- **No read up, no write down**

McLean's Criticisms of BLP

- McLean: BLP is "so trivial that it is hard to imagine a realistic security model for which it does not hold"
- McLean's "system Z" allowed administrator to reclassify object, then "write down"
- Is this fair?
- Violates spirit of BLP, but **not** expressly forbidden in statement of BLP
- Raises fundamental questions about the nature of (and limits of) modeling

B and LP's Response

- BLP enhanced with **tranquility property**
  - Strong tranquility: security labels never change
  - Weak tranquility: security label can only change if it does not violate "established security policy"
- Strong tranquility impractical in real world
  - Often want to enforce "least privilege"
  - Give users lowest privilege for current work
  - Then upgrade as needed (and allowed by policy)
  - This is known as the **high water mark** 低水位线 principle
- Weak tranquility allows for **least privilege** (high water mark), but the property is vague

BLP: The Bottom Line

- BLP is simple, probably too simple
- BLP is one of the few security models that can be used to prove things about systems
- BLP has inspired other security models
  - Most other models try to be more realistic
  - Other security models are more complex
  - Models difficult to analyze, apply in practice

Biba's Model 完整性等级

- BLP for confidentiality, Biba for **integrity**
  - Biba is to prevent unauthorized writing
- Biba is (in a sense) the dual of BLP
- Integrity model
  - Suppse you trust the integrity of **O** but not **O**
  - If object **O** includes **O** and **O** then you cannot trust the integrity of **O**
- Integrity level of O is minimum of the integrity of any object in O
- **Low water mark** principle for integrity

Biba

- Let $I(O)$ denote the integrity of object O and $I(S)$ denote the integrity of subject S
- Biba can be stated as

  Write Access Rule: S can write O if and only if $I(O) \leq I(S)$

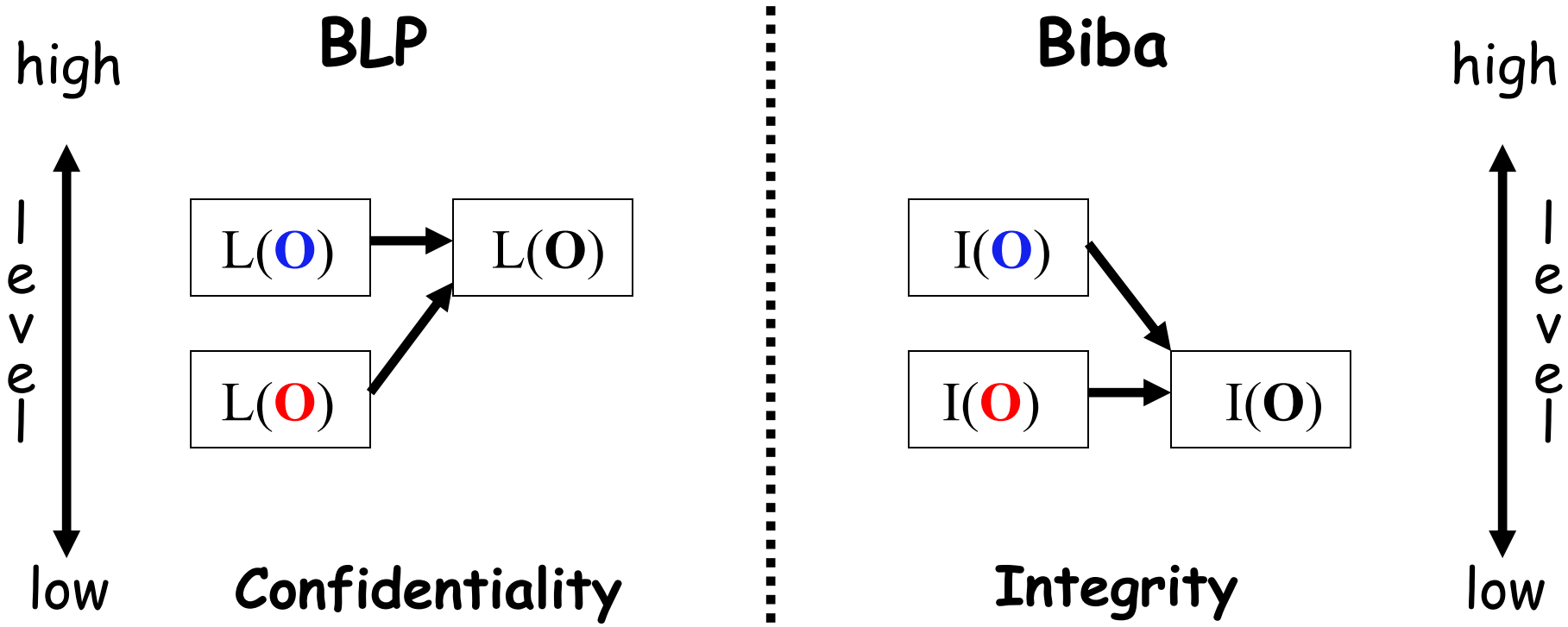
  (if S writes O, the integrity of $O \leq$ that of S)

  Biba's Model: S can read O if and only if $I(S) \leq I(O)$

  (if S reads O, the integrity of $S \leq$ that of O)
- Often, replace Biba's Model with

  Low Water Mark Policy 低水位线策略: If S reads O, then $I(S) = \min(I(S), I(O))$

BLP vs Biba


BLP
high
level
low
L(O)
L(O)
L(O)
Confidentiality
Biba
high
level
low
I(O)
I(O)
I(O)
Integrity

Compartments（分隔项）

Compartments

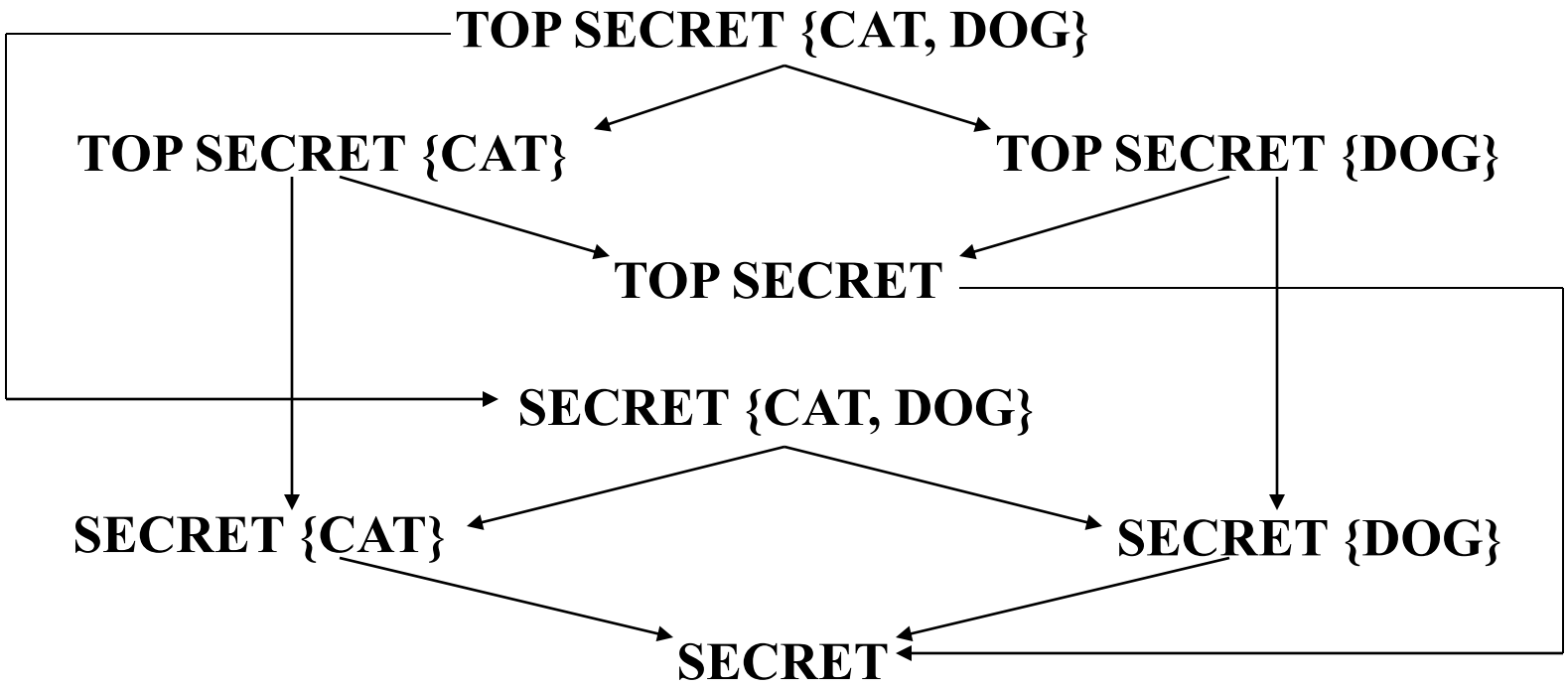
- Multilevel Security (MLS) enforces access control **up and down**
- Simple hierarchy of security labels is generally *not* flexible enough
- Compartments enforces restrictions **across**
- Suppose **TOP SECRET** divided into **TOP SECRET {CAT}** and **TOP SECRET {DOG}**
- Both are **TOP SECRET** but information flow restricted across the **TOP SECRET** level

Compartments（分隔项）

- Why compartments?
  - Why not create a new classification level?
- May ***not*** want either of
  - **TOP SECRET {CAT} ≥ TOP SECRET {DOG}**
  - **TOP SECRET {DOG} ≥ TOP SECRET {CAT}**
- Compartments designed to enforce the **need to know** principle 限制信息流跨安全等级流动
  - Regardless of clearance, you only have access to info that you need to know to do your job

Compartments

- Arrows indicate "≥" relationship



- Not all classifications are comparable, e.g., **TOP SECRET {CAT}** vs **SECRET {CAT, DOG}**

MLS vs Compartments

- MLS can be used without compartments
  - And vice-versa
- But, MLS almost always uses compartments
- Example
  - MLS mandated for protecting medical records of British Medical Association (BMA)
  - AIDS was **TOP SECRET**, prescriptions **SECRET**
  - What is the classification of an AIDS drug?
  - Everything tends toward **TOP SECRET**
  - Defeats the purpose of the system!
  - Compartments-only approach used instead

Covert Channel（隐藏通道）

Covert Channel

- MLS designed to restrict legitimate channels of communication
- May be other ways for information to flow
- For example, resources shared at different levels could be used to "signal" information
- **Covert channel**: a communication path not intended as such by system's designers

Covert Channel Example

- Alice has **TOP SECRET** clearance, Bob has **CONFIDENTIAL** clearance
- Suppose the file space shared by all users
- Alice creates file FileXYzW to signal "1" to Bob, and removes file to signal "0"
- Once per minute Bob lists the files
  - If file FileXYzW does not exist, Alice sent 0
  - If file FileXYzW exists, Alice sent 1
- Alice can leak **TOP SECRET** info to Bob

Covert Channel Example

Alice:	Create file	Delete file	Create file		Delete file
Bob:	Check file	Check file	Check file	Check file	Check file
Data:	1	0	1	1	0

Time:

Covert Channel

- Other possible covert channels?
  - Print queue
  - ACK messages
  - Network traffic, etc.
- When does covert channel exist?
  1. Sender and receiver have a shared resource
  2. Sender able to vary some property of resource that receiver can observe
  3. "Communication" between sender and receiver can be synchronized

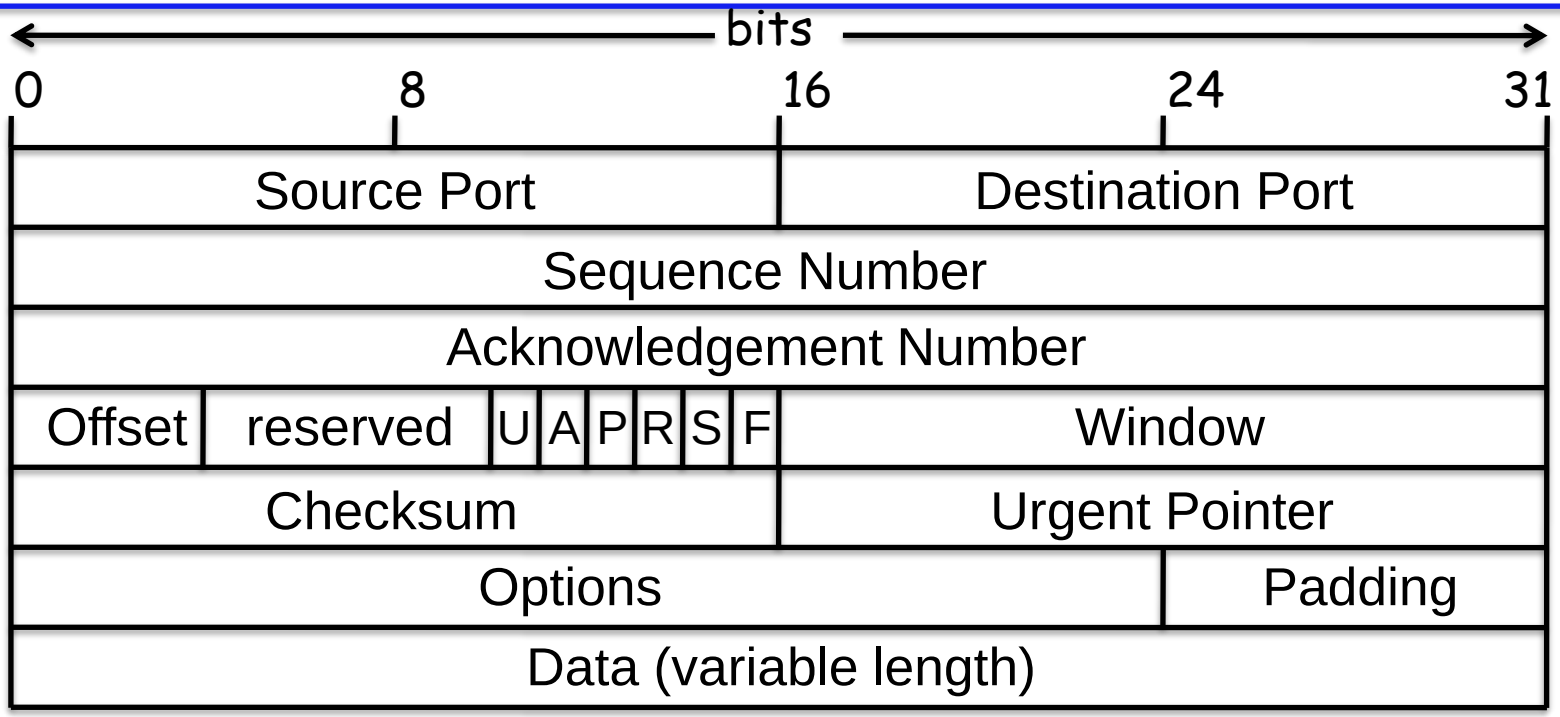
Covert Channel

- Potential covert channels are everywhere
- But, it's easy to eliminate covert channels:
  - "Just" eliminate all shared resources and all communication!
- Virtually impossible to eliminate covert channels in any **useful** information system
  - DoD guidelines: **reduce covert channel capacity** to no more than 1 bit/second
  - Implication? DoD has given up on *eliminating* covert channels

Covert Channel

- Consider 100MB **TOP SECRET** file
  - Plaintext stored in **TOP SECRET** location
  - Ciphertext — encrypted with AES using 256-bit key — stored in **UNCLASSIFIED** location
- Suppose we reduce covert channel capacity to 1 bit per second
- It would take more than 25 years to leak entire document thru a covert channel
- But it would take less than 5 minutes to leak 256-bit AES key thru covert channel!

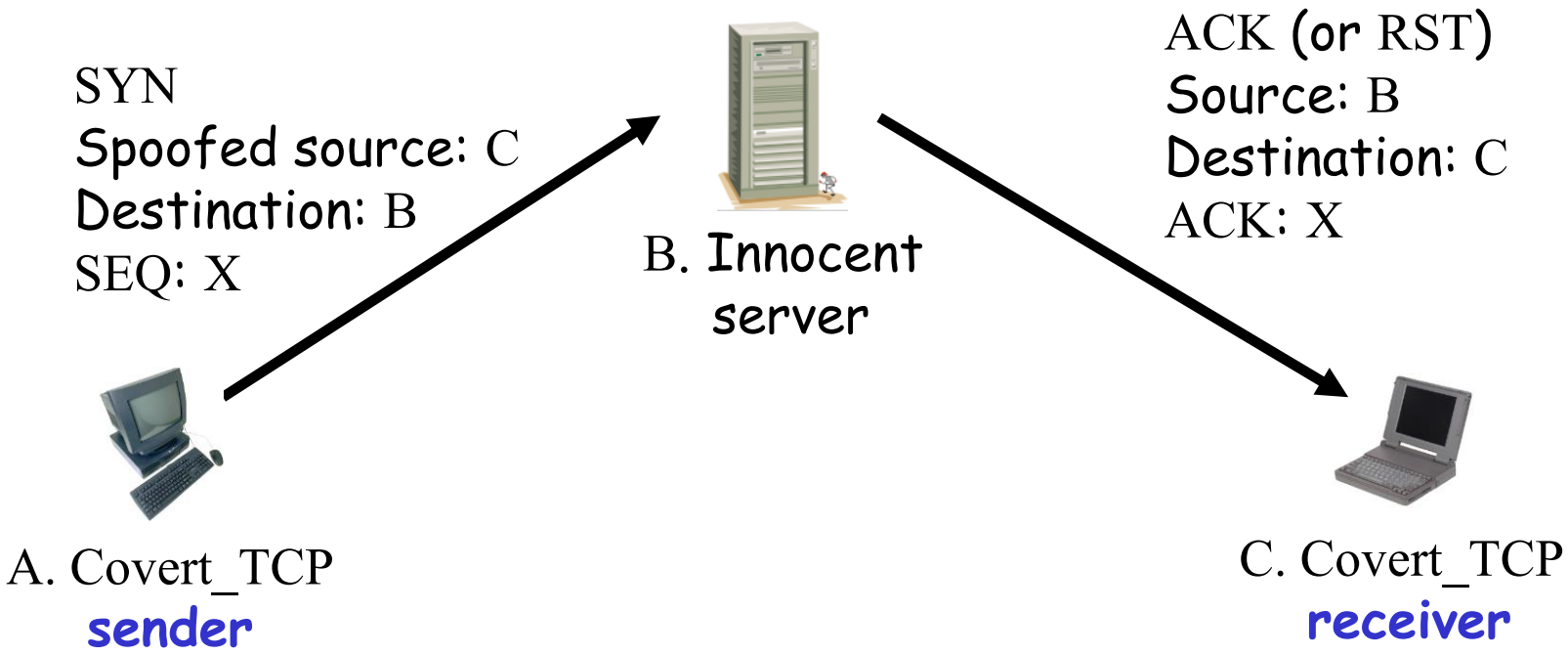
Real-World Covert Channel



- Hide data in TCP header "reserved" field
- Or use covert_TCP, tool to hide data in
  - Sequence number
  - ACK number

Real-World Covert Channel

- Hide data in TCP sequence numbers
- Tool: covert_TCP
- Sequence number X contains covert info



Inference Control(推理控制)

Inference Control Example

- Suppose we query a database
  - Question: What is average salary of female CS professors at SJSU?
  - Answer: $95,000
  - Question: How many female CS professors at SJSU?
  - Answer: 1
- Specific information has leaked from responses to general questions!

Inference Control & 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 支配规则
  - Do not release statistic if k% or more contributed by N or fewer
  - Example: Avg salary in Bill Gates' neighborhood
  - This approach used by US Census Bureau
- Randomization
  - Add small amount of random noise to data
- Many other methods — none satisfactory

Netflix Example

- Netflix prize — $1M to first to improve recommendation system by 10% or more
- Netflix created dataset for contest
  - Movie preferences of real users
  - Usernames removed, some "noise" added
- Insufficient inference control
  - Researchers able to correlate IMDB reviews with those in Netflix dataset

IMDb is the world's most popular and authoritative source for movie, TV and celebrity content.

Something Better Than Nothing?

- Robust inference control may be impossible
- Is weak inference control better than nothing?
  - **Yes**: Reduces amount of information that leaks
- Is weak covert channel protection better than nothing?
  - **Yes**: Reduces amount of information that leaks
- Is weak crypto better than no crypto?
  - **Probably not:** Encryption indicates important data
  - May be easier to filter encrypted data

CAPTCHA

A **CAPTCHA** (an acronym for "**C**ompletely **A**utomated **P**ublic Turing test to tell **C**omputers and **H**umans **A**part") is a type of challenge-response test used in computing to determine whether or not the user is human.

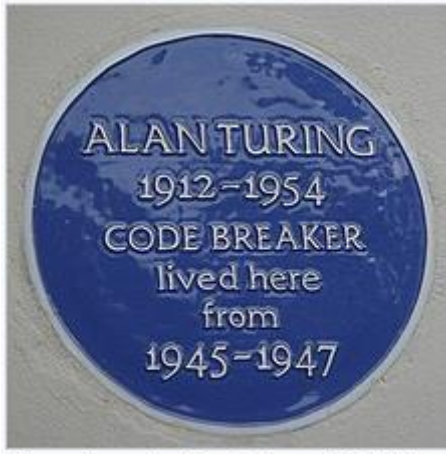
http://en.wikipedia.org/wiki/CAPTCHA

Turing Test

- Proposed by Alan Turing in 1950
- Human asks questions to a human and a computer, without seeing either
- If questioner cannot distinguish human from computer, computer passes
- This is the **gold standard** in AI
- No computer can pass this today
  - But some claim they are close to passing

Alan Mathison Turing

- 《探索发现》阿兰·图灵——破译纳粹密码的人
https://www.bilibili.com/video/BV1Bx411472R/?spm_id_from=333.337.search-card.all.click
- 一位密码破译者的回忆与思考
https://www.global-sci.org/intro/article_detail/mc/11841.html



Blue plaque to Alan Turing at 78 High Street, Hampton

CAPTCHA

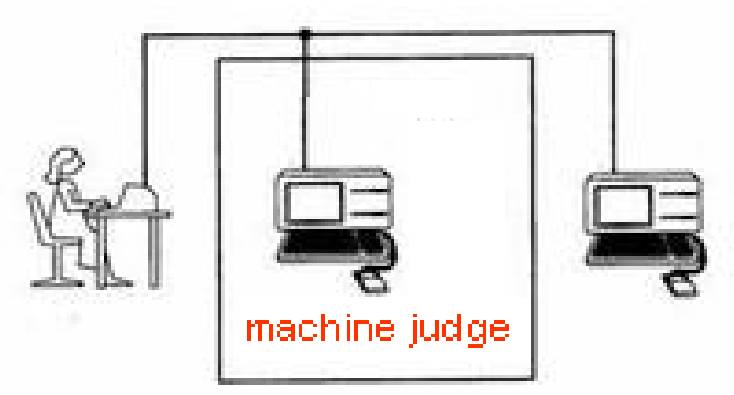
- CAPTCHA
  - Completely Automated Public Turing test to tell Computers and Humans Apart
    验证码(全自动区分计算机和人类的图灵测试)
- Completely Automated — test is generated and scored by a computer
- Public — program and data are public
- Turing test to tell… — humans can pass the test, but machines cannot
  - Also known as HIP == Human Interactive Proof
- Like an inverse Turing test (sort of…)

CAPTCHA == worst acronym in the history of acronyms

CAPTCHA

Reverse Turing Test

Turing Test



CAPTCHA

http://en.wikipedia.org/wiki/CAPTCHA

Turing, Alan (1950). "Computing Machinery and Intelligence", Mind LIX (236): 433–460
Saygin, A. P.; Cicekli, I.; Akman, V. (2000). "Turing Test: 50 years later". Minds and Machines 10 (4): 463–518.

CAPTCHA

- Basic requirements of CAPTCHA
  - CAPTCHA should be automatically generated and graded.
  - Test can be taken quickly and easily by human users (<30s).
  - Test will accept virtually all human users (>90%) and reject software agents (<0.01%).
  - Test will resist automatic attack for many years despite the technology advances and prior knowledge of algorithms (similar to cryptography).

Usability

Security

Text-based CAPTCHA

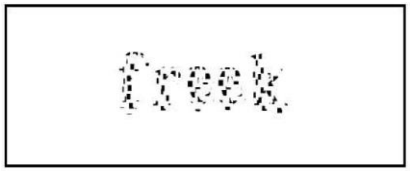
Gimpy

EZ-Gimpy

r-Gimpy

PayPal

BaffleText



ScatterType

Microsoft
2007

Microsoft
2009

Yahoo!
2007

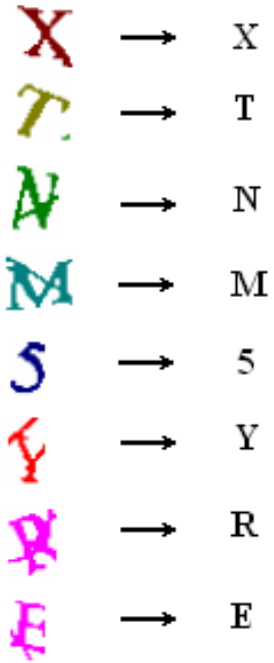
Yahoo!
2009

Google
2005

Google
2009

Breaking Text-based CAPTCHA

+

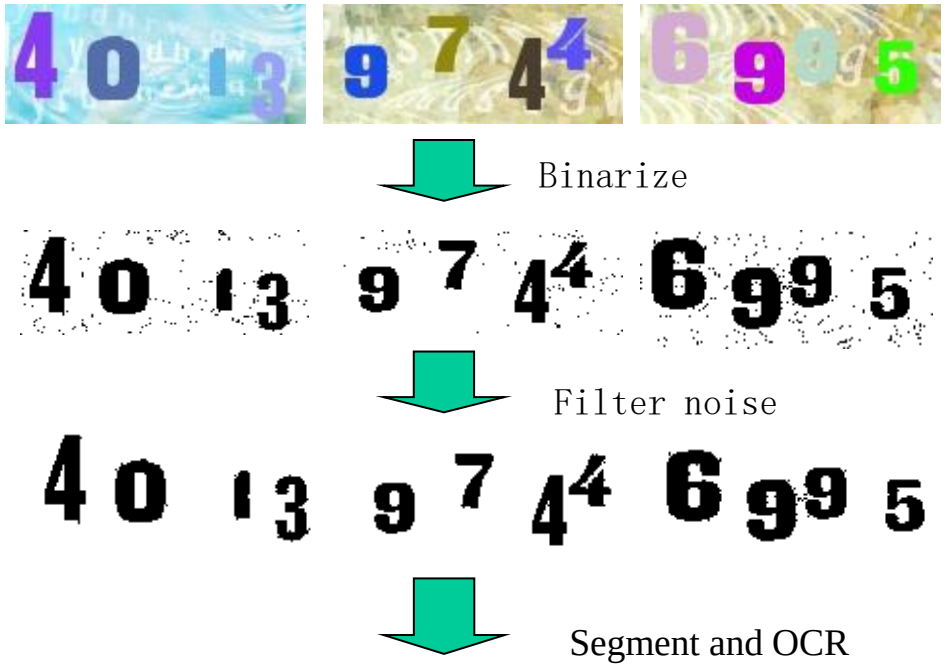


Segment

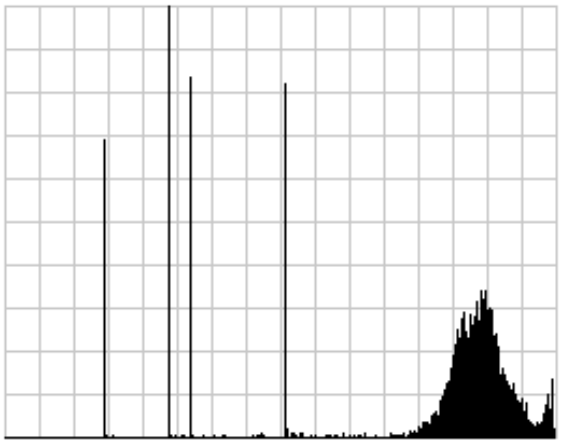
Recognize

K Chellapilla, K Larson, P Simard, M Czerwinski. (2005).Computers beat humans at single character recognition in Reading based Human Interaction Proofs. 2nd Conference on Email and Anti-Spam (CEAS).

Breaking Text-based CAPTCHA



Hist



Gray level

Total Recognition Rate: 88%

Breaking a Microsoft CAPTCHA

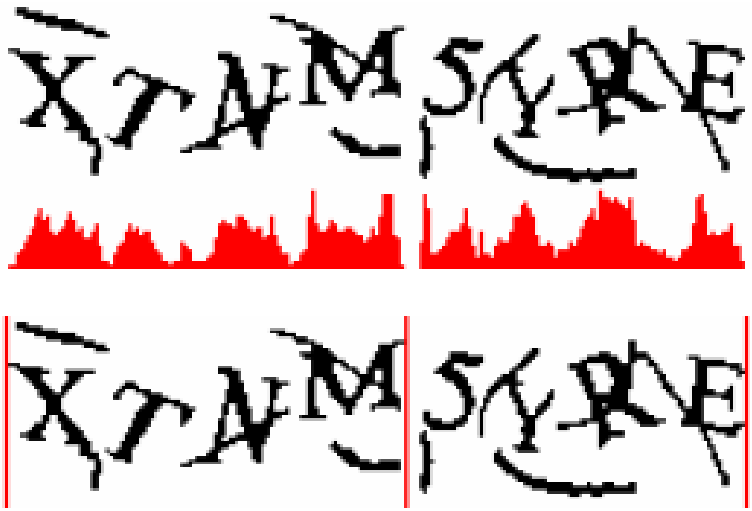
1. Image Binarization

2. Re-connect disconnected strokes

Yan, J., and El Ahmad, A.S. (2008). A low-cost attack on a Microsoft CAPTCHA. In Proc. 15-th ACM Conf. on Computer and Communications Security, 543-554.

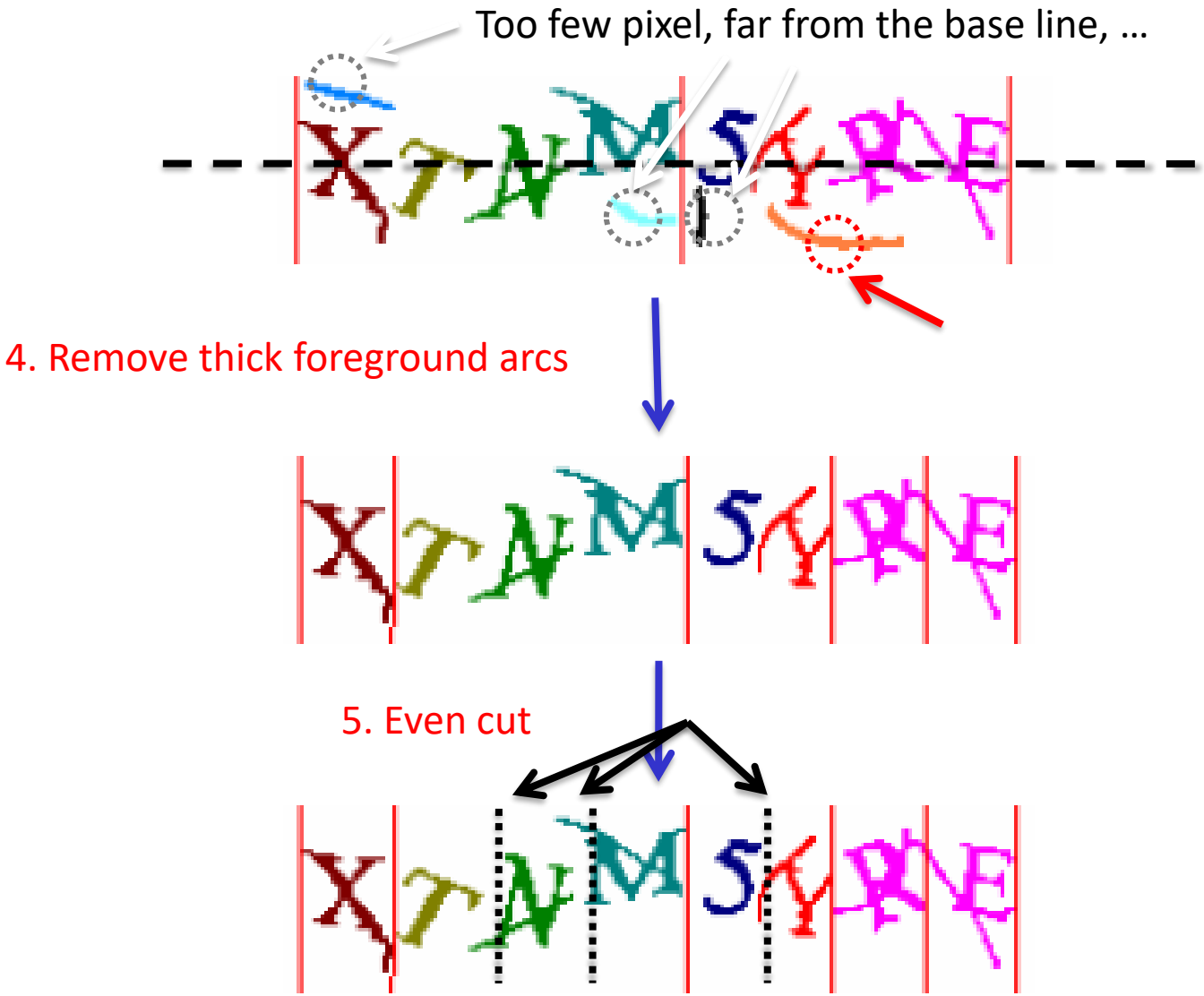
Breaking a Microsoft CAPTCHA

3a. Counting bit – Vertical Segment (VS)



3b. Find connected components – Color Filling Segmentation (CFS)

Breaking a Microsoft CAPTCHA



Usability vs Security

c + l = d ?

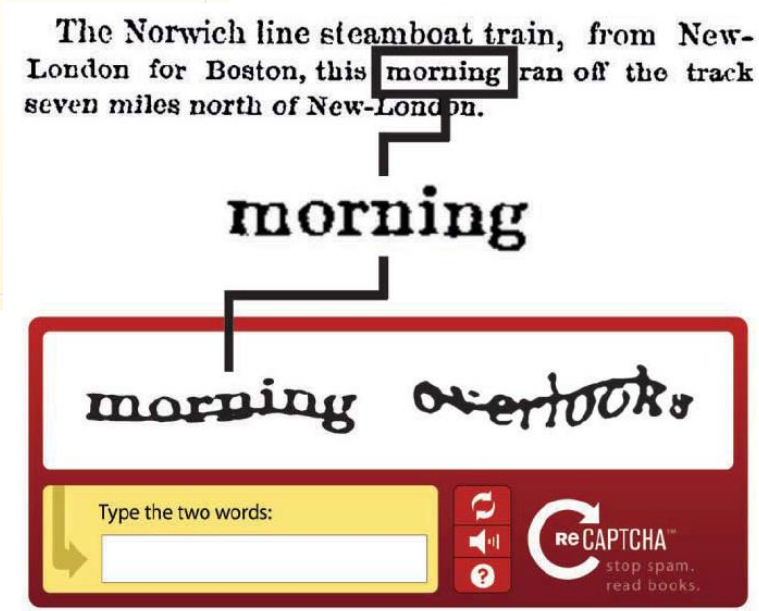
r + n = m ?

v + v = w ?

Any idea about 2nd Character?

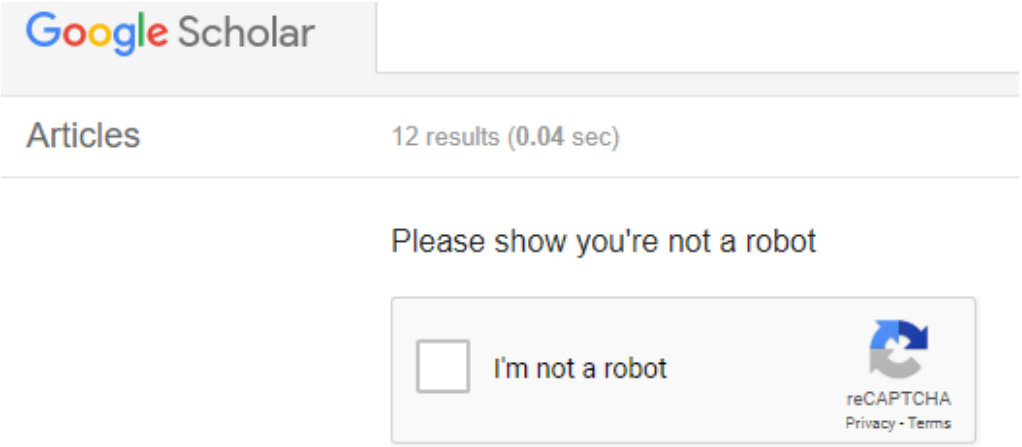
Jeff Yan and Ahmad Salah El Ahmad. 2008. Usability of CAPTCHAs Or usability issues in CAPTCHA design. ACM International Conference Proceeding.

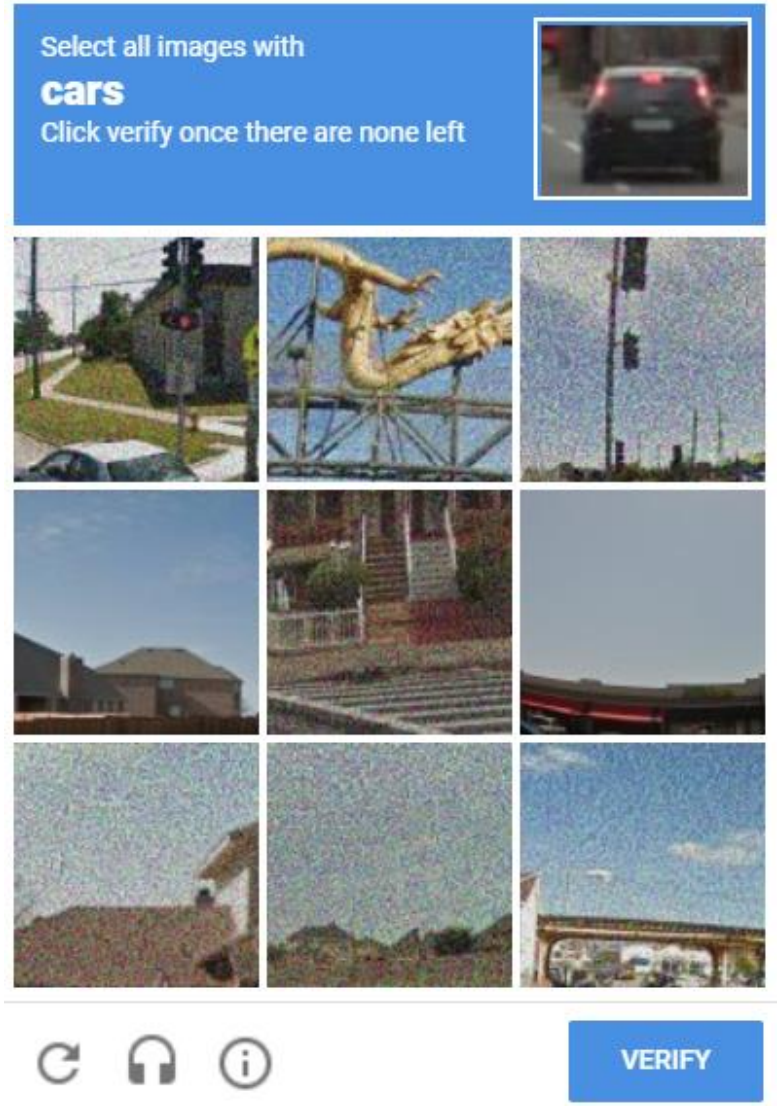
reCAPTCHA – stop spam, read books



Luis von Ahn, Benjamin Maurer, Colin McMillen, David Abraham, Manuel Blum, "reCAPTCHA: Human-Based Character Recognition via Web Security Measures," Science 2008: Vol. 321, Issue 5895, pp. 1465-1468
http://dx.doi.org/10.1126/science.1160379

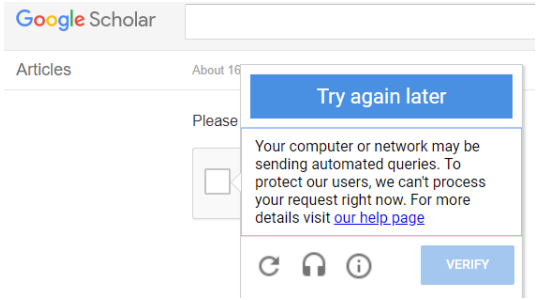
Image-based CAPTCHA





Sound-based CAPTCHA

- Visual CAPTCHAs are not very suitable for visually impaired users.
- Humans are better than computers at understanding spoken language, especially in the presence of distortion and background noise.



Sound-based CAPTCHA is vulnerable to Automatic Sound Recognition attack. Traditional audio CAPTCHAs based on distorted letters or digits can be broken by using machine learning algorithms.

J. Tam, J. Simsa, S. Hyde, L. von Ahn. (2008). Breaking audio CAPTCHAs with machine learning techniques. Neural Information Processing Systems, NIPS 2008.

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 itself cannot pass!
- CAPTCHA purpose?
  - Only humans get access (not bots/computers)
- So, CAPTCHA is for **access control**

CAPTCHA Uses?

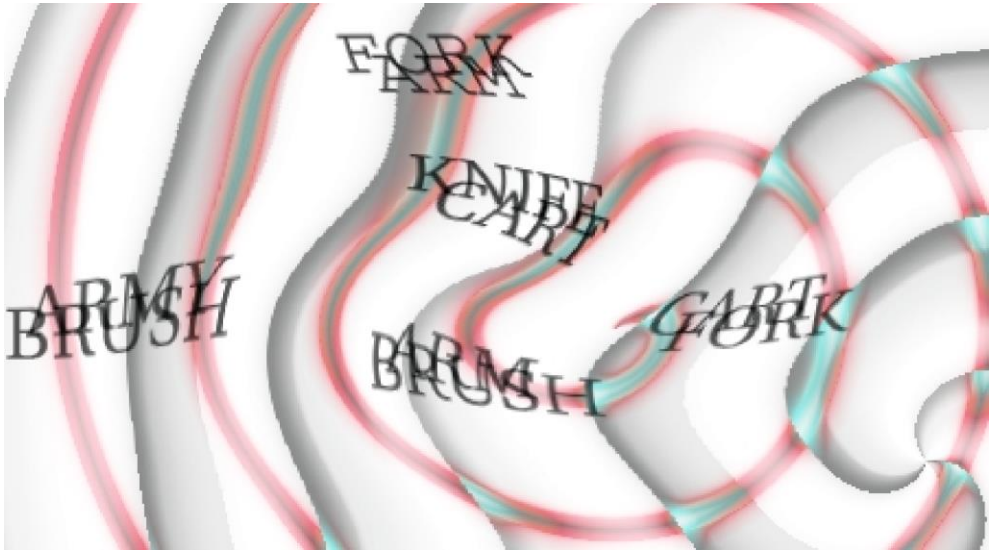
- Original motivation?
  - Automated bots stuffed ballot box in vote for best CS grad school
  - SJSU vs Stanford? No, it was MIT vs CMU
- Free email services — spammers like to use bots to sign up for 1000s of email accounts
  - CAPTCHA employed so only humans get accounts
- Sites that do not want to be automatically indexed by search engines
  - CAPTCHA would force human intervention

CAPTCHA: Rules of the Game

- Easy for most humans to pass
- Difficult or impossible for machines to pass
  - **Even with access to CAPTCHA software**
- From Trudy's perspective, the only unknown is a random number
  - Similar to Kerckhoffs' Principle
- Good to have different CAPTCHAs in case someone cannot pass one type
  - E.g., blind person could not pass visual CAPTCHA

Do CAPTCHAs Exist?

- Test: Find 2 words in the following



- Easy for most humans
- A (difficult?) OCR problem for computer
  - OCR — Optical Character Recognition

CAPTCHAs

- Current types of CAPTCHAs
  - Visual — like previous example
  - Audio — distorted words or music
- No text-based CAPTCHAs
  - Maybe this is impossible…

CAPTCHA's and AI

- OCR is a challenging AI problem
  - Hardest part is the **segmentation problem**
  - Humans good at solving this problem
- Distorted sound makes good CAPTCHA
  - Humans also good at solving this
- Hackers who break CAPTCHA have solved a hard AI problem (such as OCR)
  - So, putting hacker's effort to good use!
- Other ways to defeat CAPTCHAs???

搭建系统模拟攻击过程？

```
# 欢迎回来，李 #

查询任务 长沙 -> 岳阳 正在验证乘客信息
# 乘客验证成功 李 (成人) #

>> 第 2 次查询 长沙 -> 岳阳 2019-12-29 20:29:57
[ 查询到座位可用 出发时间 20200125 车次 K356 座位类型 硬卧 余票数量 有 ]
检查完成 开始提交订单
提交订单成功
检查订单成功
获取排队信息成功，目前排队人数 0，余票还剩余 32 张
# 提交订单成功！#
第 1 次排队，请耐心等待
第 2 次排队，请耐心等待
# 车票购买成功，订单号 EJ78085185 #
```

https://github.com/pjialin/py12306
使用机器学习算法完成对12306验证码的自动识别
https://github.com/zhaipro/easy12306
Another software: 12306智能刷票，订票
https://github.com/testerSunshine/12306