CS 591: Introduction to Computer Security

Lecture 6: Identity and Data Mining

James Hook

(Some material from Bishop, 2004)

Sources

- News stories on Surveillance
  - NY Times article on NSA spying, Dec 2005, http://www.commondreams.org/headlines05/1216-01.htm
  - USA Today article on NSA phone records, May 2006, http://www.usatoday.com/news/washington/2006-05-10-nsa_x.htm
- Readings on Telephone Fraud detection
  - Gary M. Weiss (2005). Data Mining in Telecommunications. http://storm.cis.fordham.edu/~gweiss/papers/kluwer04-telecom.pdf
  - Corinna Cortes, Daryl Pregibon and Chris Volinsky, "Communities of Interest", http://homepage.mac.com/corinnacortes/papers/portugal.ps
- Bishop Chapter 13
  Anderson 17 and 21

Identity

- Mapping from abstract subjects and objects to real people and things

Principal

- A *principal* is a unique entity
- An *identity* specifies a principal
- Authentication binds a principal to a representation of identity internal to a computer system

Uses of Identity

- Access Control
- Accountability

Unix Users

- UNIX uses UID (User identification number) for Access Control
- UNIX uses Username for Accountability
- Users provide a username and password to authenticate
- Password file maps usernames to UIDs
- Common for one principal to have multiple usernames (and UIDs)

Object identity

- Object sharing
- E.g. unix files
  - file names map to inodes
  - inodes map to “real” files

Identity in distributed systems

jghook@pdx.edu	PSU OIT	windows boxes across campus
hook@cs.pdx.edu	PSU CS	unix boxes in CS department
hook@linux.cecs.pdx.edu	PSU MCECS/CAT	linux boxes in Engineering
hook@beethoven.cs.pdx.edu	laptop (owned by PSU)	user administered laptop

Traditional solution

- Within an organization machines trust each other
- Use a central authentication server

- This does not scale
  - You trust too many things

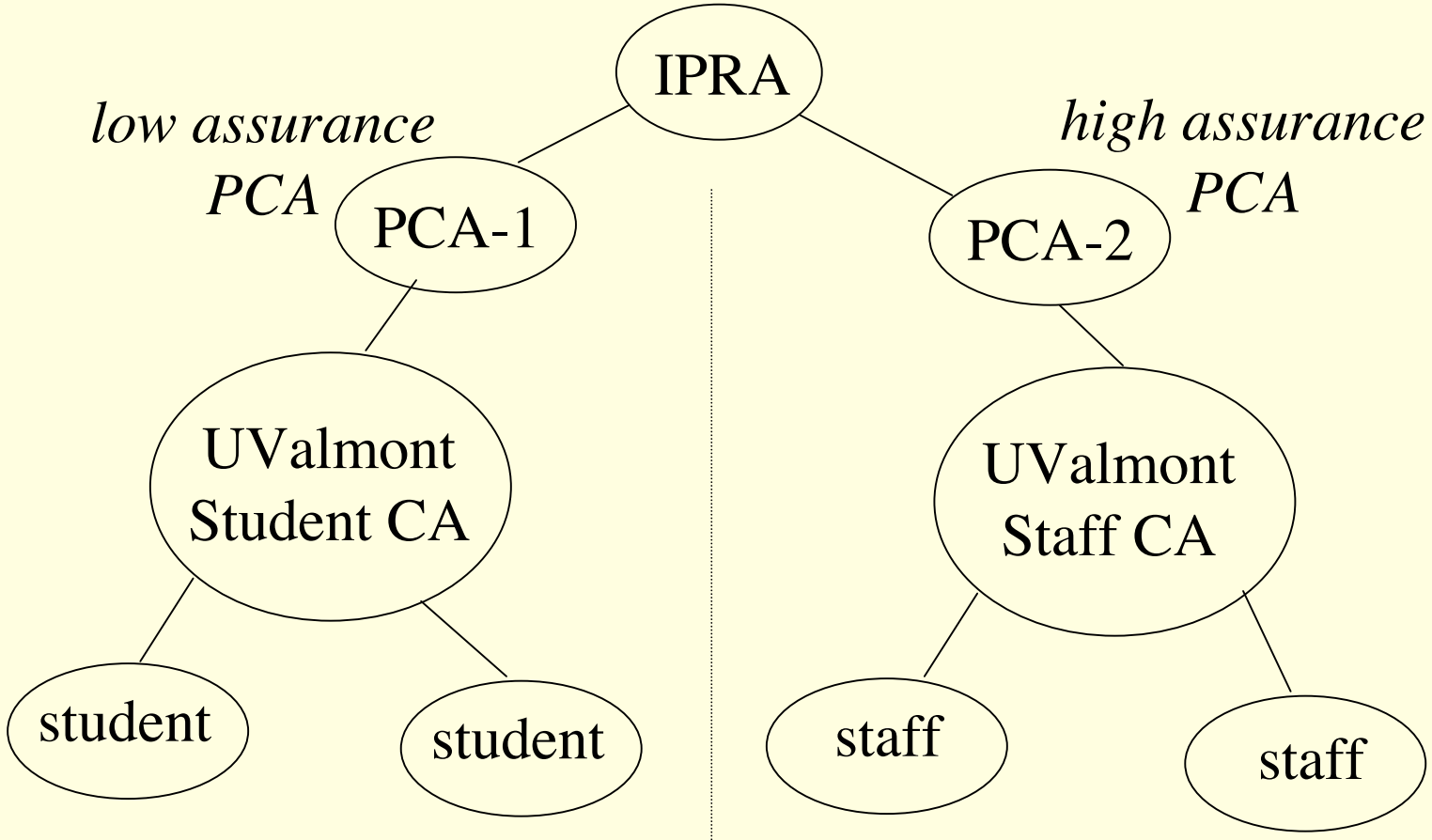
Internet Scale solutions

- Certificates associate "distinguished names" with principals
- Certificates can be managed by Certification Authorities (CA)
- CA has policies:
  - Authentication policy: Level of authentication to identify principal
  - Issuance Policy: Principals to whom the CA will issue certificates

Centralized CA

- CAs can be organized into a hierarchical structure (tree)
- Root CA: Internet Policy Registration Authority (IPRA)
  - Root certifies Policy Certifications authorities (PCAs)
  - PCAs certify individual CAs

UValmont and Certification Hierarchy



Policy and Trust

- In this distributed system policies do not need to be uniform, but they must be public and followed
- Bishop gives example where a university uses a weaker authentication test for students than staff
- Certificates issued by university reflect this
- Clients of certificates can accept or reject based on published policy

Decentralized CA

- Do you need a root CA?
- The PGP ("pretty good privacy") suite of tools uses a decentralized model of certificates
  - A group of principals can decide to trust each other
  - Certificates have a chain of signatures that allow the chain of trust to be evaluated and either accepted or rejected

Other forms of Identity

- Certified identities are a relatively heavyweight mechanism
- Other notions of identity exploit other artifacts that tend to be unique

Host identity

- IP address
- MAC address
- Hostname

- Various vulnerabilities to spoofing
  - Binkley will discuss these issues in more depth

Cookies

- Cookie: a token that contains information about the state of a transaction on a network
  - Name, Value
  - Expires
  - Domain
  - Path
  - Secure (http/https)

Cookies

- Cookies do not authenticate a principal with authority
- They do make a good surrogate to recognize
  - a principal in a session
  - a principal from a previous session

Mechanisms for anonymity

- Bishop discusses anonymous remailers
- Elaborate scenarios using crypto and header stripping to give anonymous email

Example: *anon.penet.fi*

- Offered anonymous email service
  - Sender sends letter to it, naming another destination
  - Anonymizer strips headers, forwards message
    - Assigns an ID (say, 1234) to sender, records real sender and ID in database
    - Letter delivered as if from anon1234@anon.penet.fi
  - Recipient replies to that address
    - Anonymizer strips headers, forwards message as indicated by database entry

Problem

- Anonymizer knows who sender, recipient *really* are
- Called *pseudo-anonymous remailer* or *pseudonymous remailer*
  - Keeps mappings of anonymous identities and associated identities
- If you can get the mappings, you can figure out who sent what

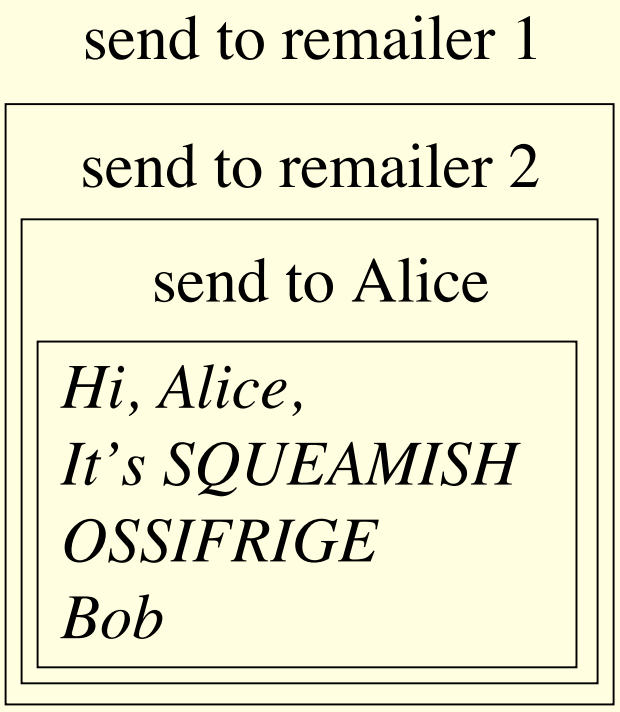
More *anon.penet.fi*

- Material claimed to be copyrighted sent through site
- Finnish court directed owner to reveal mapping so plaintiffs could determine sender
- Owner appealed, subsequently shut down site

Cypherpunk Remailer

- Remailer that deletes header of incoming message, forwards body to destination
- Also called *Type I Remailer*
- No record kept of association between sender address, remailer's user name
  - Prevents tracing, as happened with *anon.penet.fi*
- Usually used in a chain, to obfuscate trail
  - For privacy, body of message may be enciphered

Cypherpunk Remailer Message



- Encipher message
- Add destination header
- Add header for remailer *n*

  ...
- Add header for remailer 2

Weaknesses

- Attacker monitoring entire network
  - Observes in, out flows of remailers
  - Goal is to associate incoming, outgoing messages
- If messages are cleartext, trivial
  - So assume all messages enciphered
- So use traffic analysis!
  - Used to determine information based simply on movement of messages (traffic) around the network

Attacks

- If remailer forwards message before next message arrives, attacker can match them up
  - Hold messages for some period of time, greater than the message interarrival time
  - Randomize order of sending messages, waiting until at least *n* messages are ready to be forwarded
    - Note: attacker can force this by sending *n*–1 messages into queue

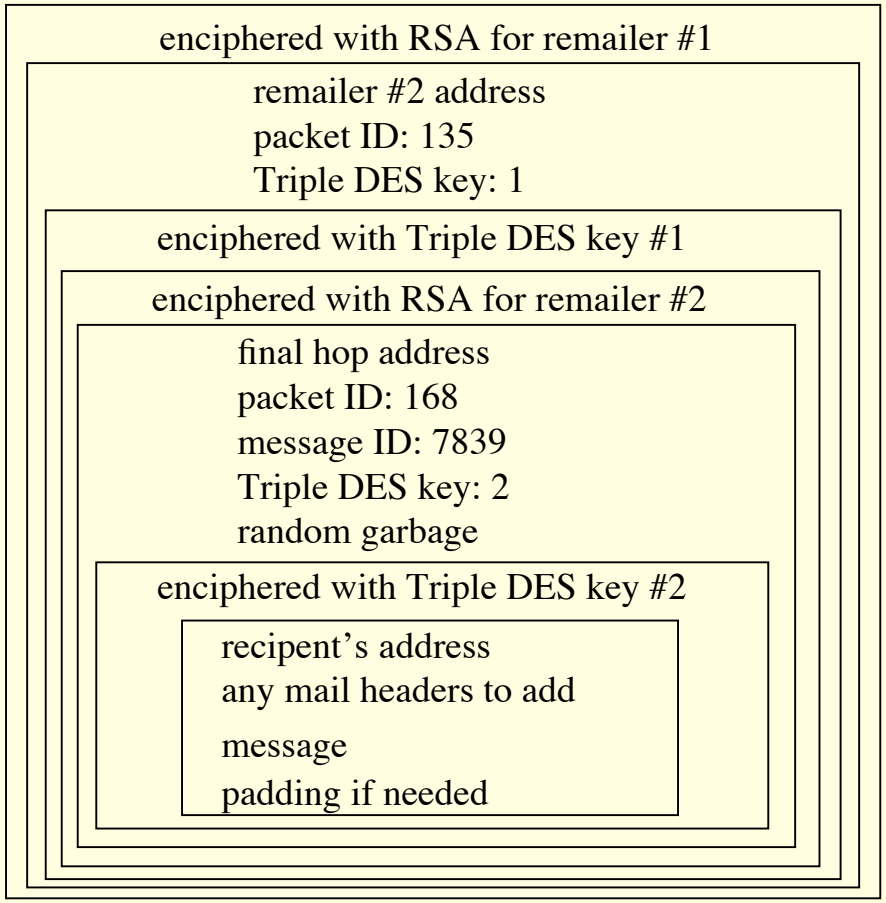
Attacks

- As messages forwarded, headers stripped so message size decreases
  - Pad message with garbage at each step, instructing next remailer to discard it
- Replay message, watch for spikes in outgoing traffic
  - Remailer can't forward same message more than once

Mixmaster Remailer

- Cypherpunk remailer that handles only enciphered mail and pads (or fragments) messages to fixed size before sending them
  - Also called Type II Remailer
  - Designed to hinder attacks on Cypherpunk remailers
    - Messages uniquely numbered
    - Fragments reassembled *only* at last remailer for sending to recipient

Cypherpunk Remailer Message



Anonymity Itself

- Some purposes for anonymity
  - Removes personalities from debate
  - With appropriate choice of pseudonym, shapes course of debate by implication
  - Prevents retaliation
- Are these benefits or drawbacks?
  - Depends on society, and who is involved

Privacy

- Anonymity protects privacy by obstructing amalgamation of individual records
- Important, because amalgamation poses 3 risks:
  - Incorrect conclusions from misinterpreted data
  - Harm from erroneous information
  - Not being let alone
- Also hinders monitoring to deter or prevent crime
- Conclusion: anonymity can be used for good or ill
  - Right to remain anonymous entails responsibility to use that right wisely

Phone Systems

- Phone fraud
  - Attacks on metering
  - Attacks on signaling
  - attacks on switching and configuration
  - insecure end systems
    - dial-through fraud
  - feature interaction

Fraud detection problem

- Subscription fraud
  - customer opens account with the intention of never paying
- Superimposition fraud
  - legitimate account; some legitimate activity
  - illegitimate activity "superimposed" by a person other than the account holder

Fraud detection as identity

- Both Subscription fraud and superimposition fraud are asking if we can identify a principal by their behavior (and without their cooperation)

Communities of Interest

- On the telephone you are who you call
- Coretes, Pregibon and Volinsky paper
  - use "top 9 lists" of ingoing and outgoing calls to characterize a user's Community of Interest (COI)
  - Define Overlap of two COIs to be a distance measure
- Overlap is highly effective at identifying fraudsters
  - "Record Linkage Using COI-based matching"
- NB: Application not limited to phone networks

Phone Fraud

- Where does the data come from?
- Phone switches generate call detail records (Weiss paper)
- These records can be harvested to yield CPV's top 9 lists
  - Hancock is a DSL for writing code to read large volumes of data

Telephone fraud detection

- Historically, COI-based matching is used to detect a deadbeat customer who has assumed a new network identity
- Is this a legitimate business use?
- Is there a potential privacy issue?
- Discuss potential abuses

Credit Card Fraud detection

- Credit Card companies have done nearly real-time analysis of card usage
- Anomalies are flagged; card holder is contacted
- Customers have come to expect this service
  - It is considered a protection and an added value
- Discuss:
  - Abuse potential
  - Does government have a role? Why or why not?

NY Times Story

- Revealed content of international phone calls between "persons of interest" were monitored outside of FISA
  - What not use FISA?
  - What if identity is a surrogate, not a name?
- [Note: I don't know if the COI papers and the news stories reference in this lecture are related.]

USA Today Story

- Several telephone companies providing call detail data to NSA
- "Largest database ever"
- Asserts no content being monitored
- Discussion/Conjecture:
  - What if they are calculating COI? Or COI-like data?
  - Could this serve as the source of the "surrogate identities" used for non-FISA wiretaps
  - If it is reasonable for business to use this technology for fraud detection is it reasonable for the government to exploit it as well?
  - What other personal information could be obtained from this data?

US Constitution
Amendment IV

The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no warrants shall issue, but upon probable cause, supported by oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.

Discussion

- Is a COI a sufficient description to meet the requirement:
  - particularly describing the place to be searched, and the persons or things to be seized