

**CSE302/ISE302:
Professional Ethics for Computer Science**

Lecture 3: Privacy

Scott Stoller
Computer Science Department
Stony Brook University

Acknowledgement: These slides are based on George Reynolds's slides, modified by prior CSE302 instructors (Scott Smolka, Klaus Mueller, and Jie Gao) and myself.

Privacy Issues

Internet privacy consists of privacy over the medium of the Internet:

- the ability to control what information one reveals about oneself over the Internet, and
- to control who can access that information.

Privacy Issues

Example: You send a message through Yahoo mail, and a third party takes a look at it.

- Someone snooping on your network.
 - The connection is not encrypted (no SSL)!
- Someone (system administrator, customer service representative, etc.) working at Yahoo.

Encryption can help protect your privacy.

Data Encryption

Cryptography

- science of *encoding and decoding* messages
- only sender and intended receiver can understand the messages
- Essential technology for ensuring confidentiality, integrity, and authenticity of electronic messages and online business transactions

Encryption

- process of converting electronic messages into a form understood only by the intended recipients

Data Encryption (continued)

Encryption key

- a (large random) value applied using an algorithm to encrypt or decrypt text
- length of key influences strength of encryption algorithm

Private key (a.k.a. Shared key) cryptosystem

- same key is used to encrypt and decrypt messages
- sender and receiver must both know the key.
 - this is the **key distribution problem**.

Data Encryption (continued)

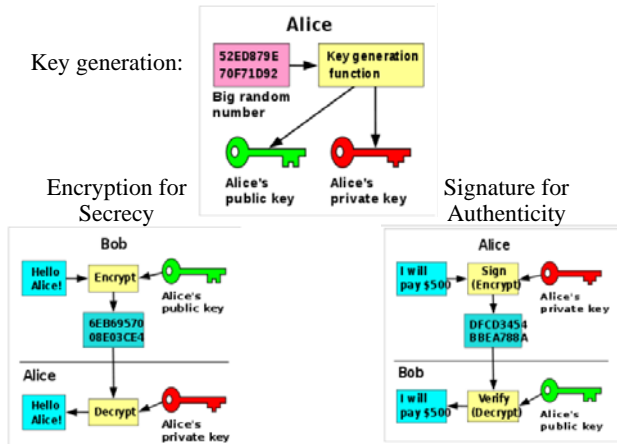
Public key cryptosystem

- Keys are generated in pairs:
 - a **public key** readily available (e.g., posted on the web) and used for encryption and signature verification
 - a **private key** kept secret and used for decryption and signature generation

RSA [Rivest, Shamir, and Adleman, 1978]

- the first, and most famous, public-key cryptosystem
- RSA keys are typically 1024–2048 bits long

RSA



Privacy Issues

Example: You send a message through Yahoo mail, and a third party takes a look at it.

- Someone snooping on your wireless network.
 - The connection is not encrypted (no SSL)!
- Someone (system administrator, customer service representative, etc.) working at Yahoo.

INFORMATION SHARING AND DISCLOSURE

Yahoo! does not rent, sell, or share personal information about you with other people or non-affiliated companies except to provide products or services you've requested, when we have your permission, or under the following circumstances:

- We provide the information to trusted partners who work on behalf of or with Yahoo! under confidentiality agreements. These companies may use your personal information to help Yahoo! communicate with you about offers from Yahoo! and our marketing partners. However, these companies do not have any independent right to share this information.
- We have a parent's permission to share the information if the user is a child under age 13. Parents have the option of allowing Yahoo! to collect and use their child's information without consenting to Yahoo! sharing of this information with people and companies who may use this information for their own purposes.
- We respond to subpoenas, court orders, or legal process, or to establish or exercise our legal rights or defend against legal claims.
- We believe it is necessary to share information in order to investigate, prevent, or take action regarding illegal activities, suspected fraud, situations involving potential threats to the physical safety of any person, violations of Yahoo!'s terms of use, or as otherwise required by law.

Privacy on Google: Gmail

Google may use your messages for "Auditing, research and analysis in order to maintain, protect and improve our services; ...". <http://www.google.com/privacypolicy.html>

From a message to Stony Brook CS faculty:

Student Privacy. The Family Educational Rights and Privacy Act (FERPA) dictates that identifiable information from student education records cannot generally be released to any third party without the consent of the student. ...

Essentially, disclosure of a student course work or grade to other than the student is a violation of FERPA. For example, using your gmail account to send the TA or the student a grade or comments about their course work is considered as released to a third party and thus violating FERPA. Even emailing student grades/comments to his/her own gmail account might be a violation of FERPA, unless the student consents.

Video about Google's privacy policy: <http://www.youtube.com/watch?v=2IKBke1puFw>

Privacy on Google: Anonymous Blogs

Is It O.K. to Blog About This Woman Anonymously?

- Randy Cohen, *The New York Times*, The Moral of the Story Blog, 24 August 2009.



Last week Judge Joan Madden ordered Google to identify the anonymous blogger whose site, "Skanks in NYC," hosted by a Google subsidiary and now removed, slammed the fashion model Liskula Cohen.

Madden found the blogger's writing, including the assertion that Cohen is a "psychotic, lying, whoring ... skank," to be "reasonably susceptible to a defamatory connotation." That is, Cohen has the basis for a lawsuit and is entitled to know the identity of the blogger in order to seek legal redress.

Google complied, identifying the blogger to Cohen's lawyer. <http://ethicist.blogs.nytimes.com/2009/08/24/is-it-ok-to-blog-about-this-woman-anonymously/>

Identity Theft

Theft of key pieces of personal information to gain access to a person's financial accounts

- using this info, identity thief may apply for new credit or financial accounts, register for college courses, etc.—all in someone else's name

Fastest growing form of fraud in the United States

- over 8 million victims in U.S. in 2007
 - <http://www.privacyrights.org/ar/idtheftsveys.htm>
- victims spend >600 hours over several years recovering from identity theft

Key personal information includes:

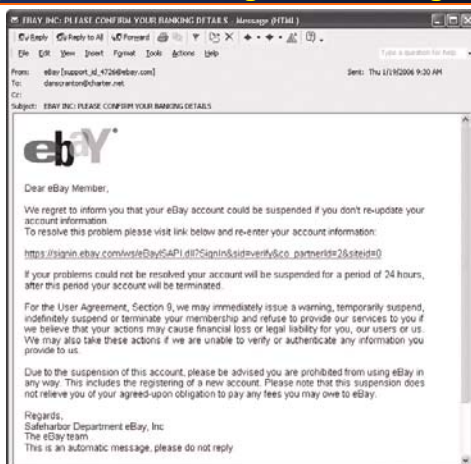
- name, address, date of birth
- Social Security number, driver's license number
- mother's maiden name

Identity Theft: How do they steal the info?

Phishing

- trick users into entering information on a counterfeit Web site
 - link typically sent in a spoofed email message
- spear-phishing - a variation in which employees are sent phony e-mails that look like they came from high-level executives within their organization

Typical E-mail Message for Phishing



Identity Theft: How do they steal the info?

Spyware

- keystroke-logging software downloaded to user's computer without consent
- enables the capture of:
 - account usernames
 - passwords
 - credit card numbers
 - other sensitive information
- operates even if an infected computer is not connected to the Internet
- records keystrokes until users reconnects; data collected then emailed to spy or posted to a web site

Identity Theft: How do they steal the info?

Sometimes it is handed to them. Example:

What happened?

On April 24, 2007, Stony Brook University became aware that files containing personal information were potentially visible on a Health Sciences Center library web site.

What kind of information was visible?

Names, Social Security numbers and University ID numbers of faculty, staff, students, alumni, and other members of the University community. The University is notifying all of the 89,853 people in the database of the incident.

- <http://www.stonybrook.edu/sb/disclosure/>

Identity Theft (continued)

Identity Theft and Assumption Deterrence Act of 1998

- aims to help fight identity fraud by making it a Federal felony, punishable by 3 to 25 years in prison

Consumer Profiling

Companies can collect info about consumers without their explicit permission!

Companies openly collect personal information about Internet users

- when they register at web sites, complete surveys, fill out forms or enter contests online

Cookies

- text files a web site places on user's hard drive so that it can remember info (including user's identity)
- examples: site preferences, contents of electronic shopping cart
- cookie are sent back to server unchanged by browser each time it accesses that server

Consumer Profiling (continued)

Tracking software

- identify visitors to your web site from e.g. pay-per-click accounts

Similar methods used outside the Web environment

- marketing firms warehouse consumer data
- examples: "loyalty" cards at grocery stores, credit card purchases, frequent flier accounts, mail-order catalogue purchases, phone surveys

Databases contain a huge amount of consumer behavioral data

- valuable to marketers, product designers, etc.

Consumer Profiling (continued)

Types of data collected while surfing the Web



- GET data: affiliated web sites visited and info requested
- POST data: form data
- Click-stream data: monitoring of consumer surfing activity

Ways to limit or stop storage of cookies on hard drive

- set the browser to limit or stop cookies
 - e.g., Private Browsing Mode in Mozilla Firefox
- manually delete the files
- download and install a cookie-management program
- use anonymous browsing services, e.g., Anonymizer.com
 - redirects traffic through proxy to hide your IP address
 - doesn't accept cookies

What Can You Learn From an IP Address?

IP Address Information	
IP Address	130.245.128.99 [Who Is Trace Route]
Hostname	proxy.cs.sunysb.edu
Remote Port	2567
Protocol	HTTP/1.1
Connection	
Proxy Server	1.1 - [DateGate/8.9.5]
IP Behind Proxy	Anonymous Proxy
Geo IP Location Information	
IP Location	US, United States
City	Stony Brook, NY, 11794
Organization	State University of New York at Stony Brook
ISP	State University of New York at Stony Brook
Latitude	40° 8' 23" North
Longitude	72° 6' 37" West
Distance	1724.69 km (1071.67 miles)

Map Location  

<http://cqcounter.com/whois/>

Treating Consumer Data Responsibly

Code of Fair Information Practices and 1980 OECD privacy guidelines: Companies should

- collect only personal info necessary for products/services
- protect this information
- inform customers if it intends to use this info for research or marketing
- provide a means for customers to opt out, correct errors

Typical uses of personal data

- optimize number, frequency, and mixture of ads
- analyze user behavior and improve services

Platform for Privacy Preferences (P3P)

- P3P software in browser downloads privacy policy for each site visited, and notifies user if site's policy does not match their preferences.

Workplace Monitoring

Employers monitor workers

- record email, web browsing, files, even videotaping employees
- goals: evaluate and improve worker productivity; ensure that corporate IT usage policy is followed

Fourth Amendment cannot be used to limit how a private employer treats its employees

Public-sector employees have greater privacy rights

- "reasonable expectation of privacy", *Katz v. U.S.*, 1998 Supreme Court ruling

Privacy advocates want federal legislation

- Requiring employers to inform employees of electronic monitoring devices, and restrict type of info collected

Advanced Surveillance Technology

Camera surveillance

- U.S. cities plan to expand surveillance systems
- London has one of world's largest public surveillance systems
- "Smart surveillance system" identifies people acting suspiciously

Facial recognition software

- identifies criminal suspects and other undesirable characters
- mixed results with current technology
 - Example: in a test at Boston's Logan airport in 2002, two systems made 153 correct identifications, and failed to identify people correctly 96 times.

Advanced Surveillance Technology (continued)

Global Positioning System (GPS) chips

- Placed in many devices to precisely locate users
 - cars, cellphones, etc.
- **Pros:** accurately respond to 911 callers; real-time location-aware marketing
- **Cons:** wireless spamming from local restaurants etc.; your whereabouts always known

Advanced Surveillance Technology

Trade-off between exciting new capabilities and privacy concerns

- **advocates:** people have no legitimate expectation of privacy in public places
- **critics:** creates potential for abuse – intimidation of political dissenters, blackmail of people caught with "wrong" person or in "wrong" place

In-Class Exercise: Instructions

Work in groups of 3 students.

Produce one written statement stating and justifying the group's answer to the question on the next slide.

Everyone in the group must print their name and write their signature next to it, at the top of the paper.

- The names are used for attendance.
- Your signature attests that (1) you are present and (2) everyone else whose name appears on the paper is present. False claims will be penalized. I will count the number of people in the room.

Groups will read their statements aloud.

- As many as time permits.

Submit the written statement at the end of class.

In-Class Exercise

Search engines collect billions of search records, containing search query, IP address, timestamp, cookie, ... What is a reasonable privacy policy for search records? Consider:

- Search data is at the heart of search company's business model: targeted on-line advertising and other forms of marketing
- User may be identified by IP address or by inference from searches (geographical searches, searches for own name, owned cars, favorite sports teams, hobbies, etc.)
- Some search topics are "sensitive", e.g., health and psychological problems, bankruptcy, anti- or pro-abortion info, illegal drugs, erotica, terrorism. User might be writing a report or novel about these topics.
 - 20 million AOL search records, posted against company policy on a website for researchers in 2006, included the query "How to kill your wife".
- What uses should be allowed? Consider uses by search company itself, giving records to law enforcement agencies, selling them to companies. Use/give/sell all records? Only suspicious or relevant records? All except "sensitive" ones? Partially anonymized records (delete or rename cookie; delete part of IP address)? Recent records? Entire history for some or all users?