What are we talking about?!?

- What is privacy?
- What privacy concerns do you have when you go online?
- What privacy concerns have you heard about (in the press, your text books, etc.)
- How are online privacy concerns different from offline privacy concerns?

A word about CS 643

- Advanced Topics in Computer Security
- Seminar style course
  - students read a paper every week
  - student presentation of the paper on Thursday
    - every student does one or two papers, depending on class size
  - Class discussion on Friday
- Project
  - pick from menu
  - roll your own
- Grading:
  - attendance, class participation, paper presentation, quality of written questions, project
The online privacy landscape

- Web privacy concerns
  - Surveys
  - Fair information practice principles
- How do they get Data?
  - Browser chatter
  - Cookies 101
  - Online and offline merging
  - Subpoenas
  - Spyware
  - Monitoring devices
- Authentication
- Wireless privacy issues
- Solutions
  - Privacy policies
  - Voluntary guidelines
  - Seal programs
  - Chief privacy officers
  - Laws and Regulations
  - Software tools

Web privacy concerns

- Data is often collected silently
  - Web allows large quantities of data to be collected inexpensively and unobtrusively
- Data from multiple sources may be merged
  - Non-identifiable information can become identifiable when merged
- Data collected for business purposes may be used in civil and criminal proceedings
- Users given no meaningful choice
  - Few sites offer alternatives
Privacy surveys find concerns

- Increasingly people say they are concerned about online privacy (80-90% of US Net users)
- Improved privacy protection is factor most likely to persuade non-Net users to go online
- 27% of US Net users have abandoned online shopping carts due to privacy concerns
- 64% of US Net users decided not to use a website or make an online purchase due to privacy concerns
- 34% of US Net users who do not buy online would buy online if they didn’t have privacy concerns

Few read privacy policies

- 3% review online privacy policies carefully most of the time
  - Most likely to review policy before providing credit card info
  - Policies too time consuming to read and difficult to understand
- 70% would prefer standard privacy policy format
- Most interested in knowing about data sharing and how to get off marketing lists
- People are more comfortable at sites that have privacy policies, even if they don’t read them
Simplified principles

- Notice and disclosure
- Choice and consent
- Data security
- Data quality and access
- Recourse and remedies

US Federal Trade Commission, Privacy Online: A Report to Congress
http://www.ftc.gov/reports/privacy3/

Browser Chatter

- Browsers chatter about
  - IP address, domain name, organization,
  - Referring page
  - Platform: O/S, browser
  - What information is requested
    - URLs and search terms
  - Cookies

- To anyone who might be listening
  - End servers
  - System administrators
  - Internet Service Providers
  - Other third parties
    - Advertising networks
  - Anyone who might subpoena log files later
**Typical HTTP request with cookie**

GET /retail/searchresults.asp?qu=beer HTTP/1.0
Referer: http://www.us.buy.com/default.asp
User-Agent: Mozilla/4.75 [en] (X11; U; NetBSD 1.5_ALPHA i386)
Host: www.us.buy.com
Accept: image/gif, image/jpeg, image/pjpeg, */*
Accept-Language: en
Cookie: buycountry=us; dcLocName=Basket;
dcCatID=6773; dcLocID=6773; dcAd=buybasket; loc=;
parentLocName=Basket; parentLoc=6773;
ShopperManager%2F=ShopperManager%2F=66FUQULL0QBT8MMTVSC5MNKBVJFWDV7; Store=107;
Category=0

**Referer log problems**

- GET methods result in values in URL
- These URLs are sent in the referer header to next host
- Example:

  http://www.merchant.com/cgi_bin/order?name=Tom+Jones&address=here+there&credit+card=234876923234&PIN=1234-&index.html
Cookies

- What are cookies?
- What are people concerned about cookies?
- What useful purposes do cookies serve?

Cookies 101

- Cookies can be useful
  - Used like a staple to attach multiple parts of a form together
  - Used to identify you when you return to a web site so you don’t have to remember a password
  - Used to help web sites understand how people use them

- Cookies can do unexpected things
  - Used to profile users and track their activities, especially across web sites
How cookies work - the basics

- A cookie stores a small string of characters
- A web site asks your browser to "set" a cookie
- Whenever you return to that site your browser sends the cookie back automatically

How cookies work - advanced

- Cookies are only sent back to the "site" that set them - but this may be any host in domain
  - Sites setting cookies indicate path, domain, and expiration for cookies
- Cookies can store user info or a database key that is used to look up user info - either way the cookie enables info to be linked to the current browsing session
Cookie terminology

- Cookie Replay - sending a cookie back to a site
- Session cookie - cookie replayed only during current browsing session
- Persistent cookie - cookie replayed until expiration date
- First-party cookie - cookie associated with the site the user requested
- Third-party cookie - cookie associated with an image, ad, frame, or other content from a site with a different domain name that is embedded in the site the user requested
  * Browser interprets third-party cookie based on domain name, even if both domains are owned by the same company

Web bugs

- Invisible “images” (1-by-1 pixels, transparent) embedded in web pages and cause referer info and cookies to be transferred
- Also called web beacons, clear gifs, tracker gifs, etc.
- Work just like banner ads from ad networks, but you can’t see them unless you look at the code behind a web page
- Also embedded in HTML formatted email messages, MS Word documents, etc.
- For more info on web bugs see: [http://www.privacyfoundation.org/resources/webbug.asp](http://www.privacyfoundation.org/resources/webbug.asp)
- For software to detect web bugs see: [http://www.bugnosis.org](http://www.bugnosis.org)
How data can be linked

- Every time the same cookie is replayed to a site, the site may add information to the record associated with that cookie
  - Number of times you visit a link, time, date
  - What page you visit
  - What page you visited last
  - Information you type into a web form

- If multiple cookies are replayed together, they are usually logged together, effectively linking their data
  - Narrow scoped cookie might get logged with broad scoped cookie
What ad networks may know...

- Personal data:
  - Email address
  - Full name
  - Mailing address (street, city, state, and Zip code)
  - Phone number

- Transactional data:
  - Details of plane trips
  - Search phrases used at search engines
  - Health conditions

“It was not necessary for me to click on the banner ads for information to be sent to DoubleClick servers.”

- Richard M. Smith

Online and offline merging

- In November 1999, DoubleClick purchased Abacus Direct, a company possessing detailed consumer profiles on more than 90% of US households.

- In mid-February 2000 DoubleClick announced plans to merge “anonymous” online data with personal information obtained from offline databases

- By the first week in March 2000 the plans were put on hold
  - Stock dropped from $125 (12/99) to $80 (03/00)
Offline data goes online...

Subpoenas

- Data on online activities is increasingly of interest in civil and criminal cases
- The only way to avoid subpoenas is to not have data
- In the US, your files on your computer in your home have much greater legal protection that your files stored on a server on the network
Spyware

- Spyware: Software that employs a user's Internet connection, without their knowledge or explicit permission, to collect information
  - Most products use pseudonymous, but unique ID

- Over 800 known freeware and shareware products contain Spyware, for example:
  - Beeline Search Utility
  - GoZilla Download Manager
  - Comet Cursor

- Often difficult to uninstall!

Anti-Spyware Sites:
- http://grc.com/oo/spyware.htm
- http://www.adcop.org/smallfish
- http://www.spychecker.com
- http://cexx.org/adware.htm

Devices that monitor you

Creative Labs Nomad JukeBox
Music transfer software reports all uploads to Creative Labs.
http://www.nomadworld.com

Sony eMarker
Lets you figure out the artist and title of songs you hear on the radio. And keeps a personal log of all the music you like on the eMarker Web site.
http://www.emarker.com

Sportbrain
Monitors daily workout. Custom phone cradle uploads data to company Web site for analysis.
http://www.sportbrain.com/

:CueCat
Keeps personal log of advertisements you’re interested in.
http://www.crq.com/cuecat.html

See http://www.privacyfoundation.org/
Seisint and Choicepoint

- Say a few words...

Authentication

- Establishes identity
  - Answers the question: To whom am I speaking?

- Evaluates the authenticity of identity proving credentials
  - Credential - is proof of identity
  - Evaluation - process that assesses the correctness of the association between credential and claimed identity
    - for some purpose
    - under some policy
Why authentication?

- We live in a world of rights, permissions, and duties
  - Authentication establishes our identity so that we can obtain the set of rights
  - E.g., we establish our identity with Tiffany’s by providing a valid credit card which gives us rights to purchase goods - physical authentication system

- Q: How does this relate to privacy?

authentication (cont.)

- Same in online world, just different constraints
  - Vendor/customer are not physically co-located, so we must find other ways of providing identity
    - e.g., by providing credit card number - electronic authentication system
  - Risks (for customer and vendor) are different

- Most of computer security is crucially dependent on the proper management and use of authentication system.
What is Identity?

- That which gives you access ... which is largely determined by context
  - We all have lots of identities
  - Pseudo-identities

- Really, determined by who is evaluating credential
  - Driver’s License, Passport, SSN prove ...
  - Credit cards prove ...
  - Signature proves ...
  - Password proves ...
  - Voice proves ...

Credentials ...

- ... are evidence used to prove identity

- Credentials can be
  - Something I am
  - Something I have
  - Something I know
**Something you know ...**

- Passport number, mothers maiden name, last 4 digits of your social security, credit card number

- **Passwords and pass-phrases**
  - Note: passwords are generally pretty weak
    - University of Michigan: 5% of passwords were goblue (followed by: love, beer)
    - Passwords used in more than one place
    - Not just because bad ones selected: If you can remember it, then a computer can guess it
      - Computers can often guess very quickly

---

**Something your have ...**

- Tokens (transponders, ...)
  - Speedpass, EZ-pass

- Smartcards

- Digital Certificates (used by Websites to authenticate themselves to customers)
Something your are ...

- Biometrics
  - Measures some physical characteristic
  - Fingerprint, face recognition, retina scanners, voice, signature, DNA
  - Can be extremely accurate and fast
  - Active biometrics authenticate, passive biometrics recognize

- What is the fundamental problem?
  - Revocation - lost fingerprint?
  - Good for physical security, generally not feasible for on-line systems

Single Sign-On (MS Passport)

- The end-user (customer) is authenticated

User → Website

Gimme’ content
No, you need a ticket

Passport (MS)
Single Sign-On (MS Passport)

- The end-user (customer) is authenticated

User
(password)

Website

Passport (MS)

Gimme' content
... content ...

You are authenticated
Please authenticate me
Single Sign-On Systems (SSO)

- What SSO systems provide
  - Centrally, the idea is that the “provider” (MS) performs authentication on users...
  - Sites subscribing to the service automatically have an “authenticatable” user community
  - Removes barriers to user access
  - From the user's perspective, I don’t have to remember a password for each site.

- What aspect of this model is particularly interesting/troublesome to privacy?
  - Do you trust Microsoft to act responsibly?

Certificate Authentication (SSL/TSL)

- The site is authenticated by the user

User

Gimme’ content (HTTPS)

Setup secure channel

Website
Certificate Authentication (SSL/TSL)

- The site is authenticated by the user

![Diagram](image)

Authentication and key agreement protocol

Certificates

- Technical detail: Certificates provide a (strong) tie between a “private key” and an identity
  - Your browser uses the SSL protocol to prove the site knows the private key
  - If the site knows the key, it is authentic

- Is the site certificate is valid?
  - The browser knows Certification Authority (CA)
  - Certificate is *signed* by a CA
  - The browser uses that signature to infer validity (plus timing information)
Authentication Privacy Principles

Source: Privacy Principles Working Group

1) Control (informed consent at enrollment)
2) Service Diversity (pick one of several soln.)
   - avoid MS controlling all data issue
3) Limit identity authentication
   - Use “proof of valid user” rather than identity
4) Provide notice (of information use)
5) Minimize collection and storage
6) Provide accountability

Lessons

- Ask why you need to authenticate customers, employees, vendors, etc.
- What are you authenticating people to do?
- Match effort and cost of authentication to the value of the things you are protecting.
- Technology imperative: don’t try to invent or build your own authentication service.
Questions

- What risks are present in physical authentication systems that are not present in electronic authentication systems?
- What about vice-versa?

Wireless communication

- What is a wireless network?
  - The use of radio frequencies (RF) to transmit digital data between computers/devices
  - Collections of often wireless-enabled computers form *ad hoc* networks
  - In more structured environments, computers access the network/Internet via *access-points*

- What is a *mobile* user?
  - User moving around wireless network

- Q: Why wireless now?
### Wireless Technologies

<table>
<thead>
<tr>
<th>Standard</th>
<th>Range</th>
<th>Bandwidth</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>802.11a,g</td>
<td>50 Meters</td>
<td>54Mbps</td>
<td>High-speed LAN (Offices)</td>
</tr>
<tr>
<td>802.11b</td>
<td>100 Meters</td>
<td>11Mbps</td>
<td>Large LAN (Airport)</td>
</tr>
<tr>
<td>Bluetooth</td>
<td>30 Meters</td>
<td>1Mbps</td>
<td>Low-cost devices - e.g.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Keyboard)</td>
</tr>
</tbody>
</table>

---

### Wireless Issues

- **Advantages**
  - Easy to move around and work freely
  - Computers become less physically constraining
  - Affords new models of computing (geography, find/user local service)

- **How does wireless communication change the threats (real or perceived)?**
  - Broadcast technology (eavesdropping)
  - Denial-of-service (DOS, jamming)
  - **Privacy** (monitoring, tracking)
Ubiquitous Computing

- User (devices) communicate with infrastructure that surrounds them
  - Allows geography to become relevant (new)
  - Opportunity to aid development

- E.g., McDonald’s Coupons
  - Walking through a mall
  - ... you have been to McDonald’s
  - ... and you like Big Mac’s
  - Suddenly, Big Mac coupons appear on your PDA

- Anybody have a problem with this?

Tracking (GPS)

- Global Positioning System (GPS) uses 15+ satellites to triangulate (locate) receiver
  - Used to track users, vehicles, E-911
  - Very important for commercial navigation, military applications, and tracking (with transmitter)

- OnStar uses technology
  - Offers roadside assistance
  - Emergency road assistance
  - Navigation, services locator (e.g., where is gas)

- Like many other technologies, can be abused
  - Tracking where people go, when they, who they (potentially) meet, ...
When good technologies go bad ...

- The Nanny Cam
  - X.10 Camera (heavily advertised on web)
  - Allows “small footprint” camera to transmit to local computer/TV (undetectable)
  - Sold to be used to monitor without detection
  - Transmits data via 802.11

- Any problems here?

The Nanny-Cam (NBC)

Wireless Technology ...

- The Motherhood Principle
  - ... wireless makes life easier because we are not longer required to interact with computing environments through fixed media (e.g., Internet) ...
  - ... which allows new ways of working (and new business models) ...
  - ... but, raises user concerns about both privacy and security.
  - So, the realization of technology is only possible where these issues are adequately addressed.
Approaches to privacy protection

- Privacy policies
- Voluntary guidelines and codes of conduct
- Seal programs
- Chief privacy officers
- Laws and regulations
- Software tools

Platform for Privacy Preferences Project (P3P)

- Developed by the World Wide Web Consortium (W3C)
  - Final P3P1.0 Recommendation issued 16 April 2002
- Allows web sites to communicate about their privacy policies in a standard computer-readable format
  - Does not require web sites to change their server software
- Enables the development of tools (built into browsers or separate applications) that
  - Summarize privacy policies
  - Compare privacy policies with user preferences
  - Alert and advise users
- P3P helps users understand privacy policies
  - P3P increases transparency, but it does not set baseline standards or enforce policies
- P3P user agent software available (as of July 2002)
  - Microsoft Internet Explorer 6
  - Netscape Navigator 7
  - AT&T Privacy Bird
    http://privacybird.com/
- For more information
  - http://www.w3.org/P3P/
  - http://p3ptoolbox.org/
  - Web Privacy with P3P
    by Lorrie Faith Cranor
    http://p3pbook.com/