Privacy

Professor Patrick McDaniel
CSE545 - Advanced Network Security
Spring 2011
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?
• How are online privacy concerns different from offline privacy concerns?
The online privacy landscape

• Web privacy concerns
  ▸ Surveys
  ▸ Fair information practice principles

• How do they get my Data?
  ▸ Browser chatter
  ▸ Cookies 101
  ▸ Online and offline merging
  ▸ Subpoenas
  ▸ Spyware
  ▸ Monitoring devices

• Solutions
  – Privacy policies
  – Voluntary guidelines
  – Seal programs
  – Chief privacy officers
  – Laws and Regulations
  – Software tools
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
Surveys show concerns ...

- 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 web site 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
Few read privacy policies

- 3% review online privacy policies carefully most of the time
  - Most likely to review before providing credit card info
  - Policies too time consuming to read and difficult to understand
- 70% would prefer a 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
OECD fair information principles

• Organization of Economic Cooperation and Development, 1980

• Guidelines:
  ‣ Collection limitation
  ‣ Data quality
  ‣ Purpose specification
  ‣ Use limitation
  ‣ Security safeguards
  ‣ Openness
  ‣ Individual participation
  ‣ Accountability

http://www.privacyrights.org/ar/fairinfo.htm
OECD fair information principles

- **Collection Limitation.** There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.

- **Data quality principle.** Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date.

- **Purpose specification.** The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfilment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose.
OECD fair information principles

- **Use limitation principle**: Personal data should not be disclosed, made available or otherwise used for purposes other than those specified in accordance with Paragraph 9 except:
  - (a) with the consent of the data subject; or
  - (b) by the authority of law.

- **Security safeguards principle**: Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorized access, destruction, use, modification or disclosure of data.

- **Openness principle**: There should be a general policy of openness about developments, practices and policies with respect to personal data. Means should be readily available of establishing the existence and nature of personal data, and the main purposes of their use, as well as the identity about usual residence of the data controller.
OECD fair information principles

- **Individual participation principle.** An individual should have the right:
  - (a) to obtain from a data controller, or otherwise, confirmation of whether or not the data controller has data relating to him;
  - (b) to have communicated to him, data relating to him
    - within a reasonable time;
    - at a charge, if any, that is not excessive;
    - in a reasonable manner; and
    - in a form that is readily intelligible to him;
  - (c) to be given reasons if a request made under subparagraphs (a) and (b) is denied, and to be able to challenge such denial; and
  - (d) to challenge data relating to him and, if the challenge is successful, to have the data erased; rectified, completed or amended.

- **Accountability principle.** A data controller should be accountable for complying with measures which give effect to the principles stated above.
Simplified principles

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

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=66FUQULL0QBT8MMTVSC5MMNKBJFWDVH7; 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, e.g.,

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

First visit to site:
- Site asks browser to store cookie xyzzy

Later visits:
- Site receives cookie xyzzy from browser
How cookies work –

• 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.
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
Ad networks
Ad networks

search for medical information

Search Service

CD Store
Ad networks

- Set cookie
- Search for medical information

- Search Service
- CD Store
Ad networks

- Search for medical information
- Set cookie
- Buy CD

Search Service

CD Store
Ad networks

Search Service

CD Store

search for medical information

set cookie

replay cookie

buy CD
Ad networks

Ad company can get your name and address from CD order and link them to your search.
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…

A family’s 25 most frequent grocery purchases (sorted by nutritional value)!
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
Drive by downloads

• Using a deceptive means to get someone to install something on their own (spyware/adware)

  ‣ Once you have one, then it starts downloading lots of others, their friends, …

  ‣ A personal favorite: **extortion-ware** -- pay us 40$ for our popup blocker, etc ….

    • The real gambit is that they demand 40$ for the uninstall option

• Answer: go get **adaware** and install it (its free)!
Spyware

• Definition: hidden software that uses local host to transmit user secrets
  ‣ e.g., browsing habits, forms data
• Typically found in “free” software
  ‣ Gnutella, game tools, demo software, MP3 tools ...
  ‣ Implemented using spyware “engines” - gator
• Imbeds in local host to
  ‣ Adds shared libraries (.dlls), adds to startup as TSR programs
  ‣ Often difficult or impossible to remove
    • You are never really sure it is gone (advice: reinstall)
• Gets installed by user action or via some of IEs ability to “help” the user via tools such as Active-X
Early devices ...

**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/](http://www.privacyfoundation.org/)
When good technologies

• 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)
Solutions: Privacy policies

- Policies let consumers know about site’s privacy practices
- Consumers can then decide whether or not practices are acceptable, when to opt-in or opt-out, and who to do business with
- The presence of privacy policies increases consumer trust
- ... but policies are often ...
  - difficult to understand
  - hard to find
  - take a long time to read
  - change without notice
Privacy policy components

- Identification of site, scope, contact info
- Types of information collected
  - Including information about cookies
- How information is used
- Conditions under which information might be shared
- Information about opt-in/opt-out
- Information about access
- Information about data retention policies
- Information about seal programs

- Security assurances
- Children’s privacy

There is lots of information to convey -- but policy should be brief and easy-to-read too!

What is opt-in? What is opt-out?
Seal program problems

• U.S. Chamber of Commerce
  • ‘A "seal" is a readily recognizable emblem, voluntarily displayed on a Web site, which signifies that a site has met recognized industry privacy requirements.’

• Certify only compliance with stated policy
  – Limited ability to detect non-compliance

• Minimal privacy requirements

• Don’t address privacy issues that go beyond the web site

• Nonetheless, reporting requirements are forcing licensees to review policies and think carefully before introducing policy changes
Chief privacy officers

• Companies are increasingly appointing CPOs to have a central point of contact for privacy concerns
• Role of CPO varies in each company
  – Draft privacy policy
  – Respond to customer concerns
  – Educate employees about company privacy policy
  – Review new products and services for compliance with privacy policy
  – Develop new initiatives to keep company out front on privacy issue
  – Monitor pending privacy legislation
Chief privacy officer

- Companies are increasingly appointing CPOs to have a central point of contact for privacy concerns.
- Role of CPO varies in each company:
  - Draft privacy policy
  - Respond to customer concerns
  - Educate employees about company privacy policy
  - Review new products and services for compliance with privacy policy
  - Develop new initiatives to keep company out front on privacy issue
  - Monitor pending privacy legislation

The newest high-level position: Chief privacy officer

Companies are hiring them to ease consumers' concerns in this age of the Internet - and to prevent costly suits.

By D. Ian Hopper
ASSOCIATED PRESS

WASHINGTON - Move over, CEO,
Laws and regulations

• Privacy laws and regulations vary widely throughout the world

• US has mostly sector-specific laws, with relatively minimal protections
  – Federal Trade Commission has jurisdiction over fraud and deceptive practices
  – Federal Communications Commission regulates telecommunications

• European Data Protection Directive requires all European Union countries to adopt similar comprehensive privacy laws
  – Privacy commissions in each country (some countries have national and state commissions)
  – Many European companies non-compliant with privacy laws (2002 study found majority of UK web sites non-compliant)
Some US privacy laws

- Bank Secrecy Act, 1970
- Fair Credit Reporting Act, 1971
- Privacy Act, 1974
- Right to Financial Privacy Act, 1978
- Cable TV Privacy Act, 1984
- Video Privacy Protection Act, 1988
- Family Educational Right to Privacy Act, 1993
- Electronic Communications Privacy Act, 1994
US law – online additions

• HIPAA (Health Insurance Portability and Accountability Act, 1996)
  ‣ When implemented, will protect medical records and other individually identifiable health information

• COPPA (Children’s Online Privacy Protection Act, 1998)
  ‣ Web sites that target children must obtain parental consent before collecting personal information from children under the age of 13

• GLB (Gramm-Leach-Bliley-Act, 1999)
  ‣ Requires privacy policy disclosure and opt-out mechanisms from financial service institutions
Safe harbor

- Membership
  - US companies self-certify adherence to requirements
  - Dept. of Commerce maintains signatory list [http://www.export.gov/safeharbor/](http://www.export.gov/safeharbor/)
  - Signatories must provide
    - notice of data collected, purposes, and recipients
    - choice of opt-out of 3rd-party transfers, opt-in for sensitive data
    - access rights to delete or edit inaccurate information
    - security for storage of collected data
    - enforcement mechanisms for individual complaints

- Approved July 26, 2000 by EU
  - reserves right to renegotiate if remedies for EU citizens prove to be inadequate
Implications of Directive for web sites

• European Union Data Directive prohibits secondary uses of data without informed consent
  – Creating personally-identifiable online profiles will have to be **opt-in** in most cases
  – Upfront notice must be given when data is collected – no web bugs
  – No transfer of data to non-EU countries unless there is adequate privacy protection
Data protection agencies

- Germany: [http://www.bfd.bund.de/](http://www.bfd.bund.de/)
- Italy: [http://www.privacy.it/](http://www.privacy.it/)
- Switzerland: [http://www.edsb.ch/](http://www.edsb.ch/)

... And many more
Platform for Privacy Preferences Project

- 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
Original Idea behind P3P

• A framework for automated privacy discussions
  – Web sites disclose their privacy practices in standard machine-readable formats
  – Web browsers automatically retrieve P3P privacy policies and compare them to users’ privacy preferences
  – Sites and browsers can then negotiate about privacy terms
The basics

• P3P provides a standard XML format that web sites use to encode their privacy policies
• Sites also provide XML “policy reference files” to indicate which policy applies to which part of the site
• Sites can optionally provide a “compact policy” by configuring their servers to issue a special P3P header when cookies are set
• No special server software required
• User software to read P3P policies called a “P3P user agent”
P3P1.0 Spec Defines

- A standard vocabulary for describing set of uses, recipients, data categories, and other privacy disclosures
- A standard schema for data a Web site may wish to collect (base data schema)
- An XML format for expressing a privacy policy in a machine readable way
- A means of associating privacy policies with Web pages or sites
- A protocol for transporting P3P policies over HTTP
A simple HTTP transaction
A simple HTTP transaction

GET /index.html HTTP/1.1
Host: www.att.com

... Request web page
A simple HTTP transaction

```
GET /index.html HTTP/1.1
Host: www.att.com

. . . Request web page

HTTP/1.1 200 OK
Content-Type: text/html

. . . Send web page
```
... with P3P 1.0 added

GET /w3c/p3p.xml HTTP/1.1
Host: www.att.com

Request Policy Reference File
... with P3P 1.0 added

GET /w3c/p3p.xml HTTP/1.1
Host: www.att.com

Request Policy Reference File

Send Policy Reference File
... with P3P 1.0 added

GET /w3c/p3p.xml HTTP/1.1
Host: www.att.com

Request Policy Reference File

Send Policy Reference File

Request P3P Policy
GET /w3c/p3p.xml HTTP/1.1
Host: www.att.com

Request Policy Reference File

Send Policy Reference File

Request P3P Policy

Send P3P Policy
... with P3P 1.0 added

Web Server

GET /w3c/p3p.xml HTTP/1.1
Host: www.att.com
Request Policy Reference File

Send Policy Reference File

Request P3P Policy

Send P3P Policy

GET /index.html HTTP/1.1
Host: www.att.com
... Request web page
GET /w3c/p3p.xml HTTP/1.1
Host: www.att.com
Request Policy Reference File

Send Policy Reference File

GET /index.html HTTP/1.1
Host: www.att.com
... Request web page

HTTP/1.1 200 OK
Content-Type: text/html
... Send web page

... with P3P 1.0 added
Transparency

- P3P clients can check a privacy policy each time it changes
- P3P clients can check privacy policies on all objects in a web page, including ads and invisible images (web bugs)

http://www.att.com/accessatt/

http://adforce.imgis.com/?adlink|2|68523|1|146|ADFORCE
P3P in IE6

Automatic processing of compact policies only; third-party cookies without compact policies blocked by default.

Privacy icon on status bar indicates that a cookie has been blocked - pop-up appears the first time the privacy icon appears.
P3P in Netscape 7

Preview version similar to IE6, focusing, on cookies; cookies without compact policies (both first-party and third-party) are “flagged” rather than blocked by default.

Indicates flagged cookie
AT&T Privacy Bird

- Free download of beta from http://www.privacybird.com/
- “Browser helper object” for IE 5.01/5.5/6.0
- Reads P3P policies at all P3P-enabled sites automatically
- Puts bird icon at top of browser window that changes to indicate whether site matches user’s privacy preferences
- Clicking on bird icon gives more information
- Current version is information only – no cookie blocking
Chirping bird is privacy

Shane Zachary Cranor

Born May 4, 2001, 7:25 am, 7 pounds, 13 oz., 21 inches

Photo Album | Latest Photos | 2001 Favorite Photos | 2002 Favorite Photos

Shane’s Photo Album

- Shane’s First Year

Shane’s Latest Photos

Shane attended Mom’s Chatham Community Band Concert, but he was so bored he fell asleep. The next day Shane helped Dad change a lightbulb -- climbing a ladder couldn’t be that much
Click on the bird for more

Policy Summary

Shane Cranor's Home Page Privacy Practices

Privacy Policy Check

Shane Cranor's Home Page's privacy policy matches your preferences.

Privacy Policy Summary

This site has the following statements in its policy:

- Site Statement 1

Site Statement 1

Types of Information Collected:

- HTTP protocol information
- Click-stream information

How your information will be used:

- Research and development
- To complete the activity for which the data was provided
- Web site and system administration

Who will use your information:

- This web site and its agents
Privacy policy summary - mismatch

1-800-Flowers.com, Inc. Privacy Practices

Privacy Policy Check

1-800-Flowers.com, Inc.'s privacy policy does not match your preferences:

- Unless you opt-out, site may share financial information or information about your purchases with other companies (other than those helping the site provide services to you)
- Unless you opt-out, site may share information that personally identifies you with other companies (other than those helping the site provide services to you)

Privacy Policy Summary

This site has the following statements in its policy:

- Site Statement 1 - All users and customers
Users select warning

Privacy Preference Settings

These settings control when a warning icon will be displayed at the top of your browser window. You can click on the warning icon for more information.

Select Privacy Level: Low, Medium, High, Custom, Imported

--- HEALTH OR MEDICAL INFORMATION ---
Warn me at web sites that use my health or medical information:
- For analysis, marketing, or to make decisions that may affect what content or ads I see, etc.
- To share with other companies (other than those helping the web site provide services to me)

--- FINANCIAL OR PURCHASE INFORMATION ---
Warn me at web sites that use my financial information or information about my purchases:
- For analysis, marketing, or to make decisions that may affect what content or ads I see, etc.
- To share with other companies (other than those helping the web site provide services to me)

--- PERSONALLY IDENTIFIABLE INFORMATION (name, address, phone number, email address, etc.) ---
Warn me at web sites that may contact me to interest me in other services or products:
- Via telephone
- Via other means (email, postal mail, etc.)
- And do not allow me to remove myself from marketing/mailing lists

Warn me at web sites that use information that personally identifies me:
- To determine my habits, interests, or other characteristics
- To share with other companies (other than those helping the website provide services to me)
- Warn me at web sites that do not allow me to find out what data they have about me

--- NON-PERSONALLY IDENTIFIABLE INFORMATION (demographics, interests, web sites visited, etc.) ---
Warn me at web sites that use my non-personally identifiable information:
- To determine my habits, interests, or other characteristics
- To share with other companies (other than those helping the website provide services to me)
Bird checks policies for embedded content
Types of P3P user agent tools

- On-demand or continuous
  - Some tools only check for P3P policies when the user requests, others check automatically at every site

- Generic or customized
  - Some tools simply describe a site’s policy in some user friendly format – others are customizable and can compare the policy with a user’s preferences

- Information-only or automatic action
  - Some tools simply inform users about site policies, while others may actively block cookies, referrers, etc. or take other actions at sites that don’t match user’s preferences

- Built-in, add-on, or service
  - Some tools may be built into web browsers or other software, others are designed as plug-ins or other add-ons, and others may be provided as part of an ISP or other service
User privacy preferences

• P3P 1.0 agents may (optionally) take action based on user preferences
  – Users should not have to trust privacy defaults set by software vendors
  – User agents that can read APPEL (A P3P Preference Exchange Language) files can
    offer users a number of canned choices developed by trusted organizations
  – Preference editors allow users to adapt existing preferences to suit own tastes, or
    create new preferences from scratch
  – For more info on APPEL see http://www.w3.org/TR/WD-P3P-preferences
Other types of P3P tools

• P3P validators
  – Check a site’s P3P policy for valid syntax

• Policy generators
  – Generate P3P policies and policy reference files for web sites

• Web site management tools
  – Assist sites in deploying P3P across the site, making sure forms are consistent with P3P policy, etc.

• Search and comparison tools
  – Compare privacy policies across multiple web sites – perhaps built into search engines