P3P

Lorrie Faith Cranor
October 2, 2014

8-533 / 8-733 / 19-608 / 95-818:
Privacy Policy, Law, and Technology
Today’s agenda

• Quiz
• Questions/comments about the readings
• P3P
By the end of class you will be able to:

• Understand the history of P3P and the motivation for its development and adoption

• Understand the major components of P3P

• Understand how web sites are circumventing P3P to avoid IE cookie blocking

• Understand how to read a W3C specification
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
P3P history

• Idea discussed at November 1995 FTC meeting

• Ad Hoc “Internet Privacy Working Group” convened to discuss the idea in Fall 1996

• W3C began working on P3P in Summer 1997
  – Several working groups chartered with dozens of participants from industry, non-profits, academia, government
  – Numerous public working drafts issued, and feedback resulted in many changes
  – Early ideas about negotiation and agreement ultimately removed
  – Automatic data transfer added and then removed
  – Patent issue stalled progress, but ultimately became non-issue

• P3P issued as official W3C Recommendation on April 16, 2002
  – http://www.w3.org/TR/P3P/
P3P1.0 – A first step

• Offers an easy way for web sites to communicate about their privacy policies in a standard machine-readable format
  – Can be deployed using existing web servers

• This will enable the development of tools that:
  – Provide snapshots of sites’ policies
  – Compare policies with user preferences
  – Alert and advise the user
P3P is part of the solution

• P3P1.0 helps users understand privacy policies but is not a complete solution

• Seal programs and regulations
  – help ensure that sites comply with their policies

• Anonymity tools
  – reduce the amount of information revealed while browsing

• Encryption tools
  – secure data in transit and storage

• Laws and codes of practice
  – provide a base line level for acceptable policies
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

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

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

Request web page

Send web page
... 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

Send P3P Policy

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
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
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
Users can click on privacy icon for list of cookies; privacy summaries are available at sites that are P3P-enabled.
Privacy summary report is generated automatically from full P3P policy.
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
Users can view English translation of (part of) compact policy in Cookie Manager.
A policy summary can be generated automatically from full P3P policy.
What’s in a P3P policy?

• Name and contact information for site
• The kind of access provided
• Mechanisms for resolving privacy disputes
• The kinds of data collected
• How collected data is used, and whether individuals can opt-in or opt-out of any of these uses
• Whether/when data may be shared and whether there is opt-in or opt-out
• Data retention policy
Why web sites adopt P3P

• Demonstrate corporate leadership on privacy issues
  – Show customers they respect their privacy
  – Demonstrate to regulators that industry is taking voluntary steps to address consumer privacy concerns

• Distinguish brand as privacy friendly

• Prevent IE6 from blocking their cookies

• Anticipation that consumers will soon come to expect P3P on all web sites

• Individuals who run sites value personal privacy
P3P early adopters

- News and information sites – CNET, About.com, BusinessWeek
- Search engines – Yahoo, Lycos
- Ad networks – DoubleClick, Avenue A
- Telecom companies – AT&T
- Financial institutions – Fidelity
- Computer hardware and software vendors – IBM, Dell, Microsoft, McAfee
- Retail stores – Fortunoff, Ritz Camera
- Government agencies – FTC, Dept. of Commerce, Ontario Information and Privacy Commissioner
- Non-profits - CDT
Web site adoption of P3P

- AT&T study surveyed 5,856 websites in 2003, found 538 P3P policies
  - Adoption highest among popular websites (~30% of top 100 sites)
  - Web site adoption increasing slowly, but steadily
  - Low adoption for government sites – but changed with new regulations

- Large number of P3P policies contain technical errors
  - Most errors due to old version of P3P spec or minor technical issues
  - 7% have severe errors such as missing required components

Legal issues

- P3P specification does not address legal standing of P3P policies or include enforcement mechanisms.

- P3P specification requires P3P policies to be consistent with natural-language privacy policies:
  - P3P policies and natural-language policies are not required to contain the same level of detail.
  - Typically, natural-language policies contain more detailed explanations.

- In some jurisdictions, regulators and courts may treat P3P policies equivalently to natural language privacy policies.

- The same attorneys and policy makers involved in drafting natural-language policy should help create P3P policy.
<table>
<thead>
<tr>
<th>Privacy policy</th>
<th>P3P policy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Designed to be read by a human</td>
<td>Designed to be read by a computer</td>
</tr>
<tr>
<td>Can contain fuzzy language with “wiggle room”</td>
<td>Mostly multiple choice – sites must place themselves in one “bucket” or another</td>
</tr>
<tr>
<td>Can include as much or as little information as a site wants</td>
<td>Must include disclosures in every required area</td>
</tr>
<tr>
<td>Easy to provide detailed explanations</td>
<td>Limited ability to provide detailed explanations</td>
</tr>
<tr>
<td>Sometimes difficult for users to determine boundaries of what it applies to and when it might change</td>
<td>Precisely scoped</td>
</tr>
<tr>
<td>Web site controls presentation</td>
<td>User agent controls presentation</td>
</tr>
</tbody>
</table>
P3P Interface design challenges

• P3P 1.0 specification focuses on interoperability, says little about user interface
  – P3P 1.1 spec will provide explanations of P3P vocabulary elements suitable for display to end users

• P3P user agents typically need user interfaces for:
  – informing users about web site privacy policies
  – configuring the agent to take actions on the basis of a user’s privacy preferences
Informing users about privacy is difficult

- Privacy policies are complex
  - Over 36K combinations of P3P “multiple choice” elements
- Users are generally unfamiliar with much of the terminology used by privacy experts
- Users generally do not understand the implications of data practices
- Users are not interested in all of the detail of most privacy policies
- Which details and the level of detail each user is interested in varies
Specifying privacy preferences is difficult

- Privacy policies are complex
- User privacy preferences are often complex and nuanced
- Users tend to have little experience articulating their privacy preferences
- Users are generally unfamiliar with much of the terminology used by privacy experts
Iterative design approach

• Four P3P user agent prototypes developed over 4-year period while P3P specification was under development
  – 1997 - W3C prototype
  – 1999 - Privacy Minder
  – 2000 - AT&T/Microsoft browser helper object
  – 2001 - AT&T usability testing prototype

• AT&T Privacy Bird beta released Feb. 2002
  – August 2002 user study
  – Beta 1.2 released Feb. 2003
W3C prototype

- Based on pre-W3C draft of P3P vocabulary with 3 fields, 7x9x2=126 combinations of elements
- Preference interface eliminated the impractical combos, combined 2 dimensions → 7x14=98 combinations
- Matrix represented by tabbed interface
- Feedback: too complicated, too many choices
AT&T Privacy Bird

- Free download of beta from http://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 indicator

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 info

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

Site Statement 1 - All users and customers
 Types of information collected:
Expand/collapse added in beta

1.2

Policy Summary

- Federal Trade Commission Privacy Practices
  - Privacy Policy Check
  - Federal Trade Commission’s privacy policy matches your preference.
  - Privacy Policy Summary
  - Policy Statement 1 - Basic Information
    - Data collected from all Web users: access logs, and search strings (if entered).
    - Types of Information that may be collected:
      - search terms
      - click-stream information
    - How your information may be used:
      - To complete the activity for which the data was provided
      - To do web site and system administration
    - Who may use your information:
      - This web site and the companies that help the site provide services to you.
- Policy Statement 2 - Data Collection
  - Access to your information
    - This site allows you to access your data about you from its records
  - How to reach this site
  - How to resolve privacy-related issues

More Information
Bird checks policies for embedded content
Privacy Bird icons
Evaluating P3P user agents

• Questions
  – Does P3P user agent perform useful function?
  – Can users use it effectively?

• Evaluation techniques
  – User survey
  – Laboratory study
Privacy Bird user survey

• ~20,000 downloads in first six months of beta trial

• Users asked whether they were willing to participate in survey when they downloaded software

• 2000 email addresses randomly selected from those willing to participate

• Sent invitation to fill out online 35-question survey
Privacy settings

• How often did you change your privacy settings?
  – Never: 25%
  – Once or twice: 52%
  – Several times: 21%
  – Ten or more times: 2%

• P3P experts changed their settings more frequently

• A few comments that people did not fully understand what all the choices mean
Example:
Sending flowers
FTD's 'Good as Gold' Guarantee – Fresh, beautiful flowers and plants that will last at least 7 days.

PICKS OF THE WEEK

1. FTD® Star Gazer™ Bouquet #9061X - $109.95
2. Multicolor Roses Bowl #0683T - $59.95
3. Pastel Basket Planter #1112T - $49.95
Privacy Finder

• Prototype developed at AT&T Labs, improved and deployed by CUPS

• Uses Google or Yahoo! API to retrieve search results

• Checks each result for P3P policy

• Evaluates P3P policy against user’s preferences

• Reorders search results

• Composes search result page with privacy annotations next to each P3P-enabled result

• Users can retrieve “Privacy Report” similar to Privacy Bird policy summary
This site may collect the following types of information about you:
- search terms
- HTTP protocol information
- click-stream information
- use of HTTP cookies
  - Information about your tastes or interests
  - Cookies and mechanisms that perform similar functions
  - Which pages you visited on this website and how long you stayed at each page
  - Website login IDs and other identifiers (excluding government IDs and financial account numbers)
  - Information about the computer you are using, such as its hardware, software, or Internet address
  - Email address or other online contact information
  - Name, address, phone number, or other contact information
- third party's name
- home contact information (optional)
- server stores the transaction history
- user's name (optional)

The ways your information may be used:
- To aid in historical preservation as governed by a law or policy described in this privacy policy
- To contact you through means other than telephone (for example, email or postal mail) to market services or products -- unless you opt-out
- To make decisions that directly affect you using information about you, for example to recommend products or services based on your previous purchases -- unless you opt-out
- To customize the site for your current visit only
- To do research and analysis in which your information may be linked to an ID code but not to your personal identity
- To contact you by telephone to market services or products -- unless you opt-out
- For research and development, but without connecting any information to you
- To perform web site and system administration
- To provide the service you requested

With whom this site may share your information:
- Other companies whose privacy policies are unknown to this site -- unless you opt-out
- Companies that have privacy policies similar to this site's -- unless you opt-out
- Delivery companies that help this site fulfill your requests and who may also use your information in other ways

Access to your information

Done
P3P Adoption Studies

- Compiled two lists of search terms:
  - Typical: 20,000 terms randomly sampled from one week of AOL user search queries
  - Ecommerce: 940 terms screen scraped from Froogle front page
- Submitted search terms to Google, Yahoo!, and AOL search engines and collected top 20 results for each term
- Checked each result for P3P policy and evaluated policies against 5 “rulesets” and P3P validator
- Saved 1,232,955 annotated search results in database
- Separately checked for P3P policies on 30,000 domains most clicked on by AOL search engine users

Results: P3P deployment

- 10% of results from typical search terms have P3P
- 21% of results from ecommerce search terms have P3P
- More popular sites are more likely to have P3P
  - 5% of sites in our cache have P3P
  - 9% of 30K most clicked on domains have P3P
  - 17% of clicks to 30K most clicked on domains have P3P
Results: Frequency of P3P-enabled hits

- 83% of searches had at least one P3P-enabled site in top 20 results
- 68% of searches had at least one P3P-enabled site in top 10 results
- For top 20 search results returned by AOL search engine for typical search terms:
  - 29% return at least 1 P3P-enabled hit that matches medium privacy preferences
  - 34% return at least 1 P3P-enabled hit in that does not share data
  - 31% return at least 1 P3P-enabled hit that does not market without opt-in
  - Thus, ~ 1/3 of the time AOL users will find site with “good” privacy policy in first 2 pages of results
Does Privacy Finder influence purchases?

- Yes!

P3P deployment overview

• Create a privacy policy
• Analyze the use of cookies and third-party content on your site
• Determine whether you want to have one P3P policy for your entire site or different P3P policies for different parts of your site
• Create a P3P policy (or policies) for your site
• Create a policy reference file for your site
• Configure your server for P3P
• Test your site to make sure it is properly P3P enabled
One policy or many?

• P3P allows policies to be specified for individual URLs or cookies

• One policy for entire web site (all URLs and cookies) is easiest to manage

• Multiple policies can allow more specific declarations about particular parts of the site

• Multiple policies may be needed if different parts of the site have different owners or responsible parties (universities, CDNs, etc.)
Third-party content

• Third-party content should be P3P-enabled by the third-party

• If third-party content sets cookies, IE6 will block them by default unless they have P3P compact policy
  – But this can be circumvented!

• Your first-party cookies may become third-party cookies if your site is framed by another site, a page is sent via email, etc.
Cookies and P3P

- P3P policies must declare all the data stored in a cookie as well as any data linked via the cookie
- P3P policies must declare all uses of stored and linked cookie data
- Sites should not declare cookie-specific policies unless they are sure they know where their cookies are going!
  - Watch out for domain-level cookies
  - Most sites will declare broad policy that covers both URLs and cookies
Generating a P3P policy

• Edit by hand
  – Cut and paste from an example

• Use a P3P policy generator
  – Recommended: IBM P3P policy editor
    http://www.alphaworks.ibm.com/tech/p3peditor

• Generate compact policy and policy reference file the same way (by hand or with policy editor)

• Get a book
  – Web Privacy with P3P
    by Lorrie Faith Cranor
    http://p3pbook.com/
IBM P3P Policy Editor

Sites can list the types of data they collect

And view the corresponding P3P policy
Locating the policy reference file

- Place policy reference file in “well known location” `/w3c/p3p.xml`
  - Most sites will do this

- Use special P3P HTTP header
  - Recommended only for sites with unusual circumstances, such as those with many P3P policies

- Embed link tags in HTML files
  - Recommended only for sites that exist as a directory on somebody else’s server (for example, a personal home page)
Compact policies

- HTTP header with short summary of full P3P policy for cookies (not for URLs)
- Not required
- Must be used in addition to full policy
- Must commit to following policy for lifetime of cookies
- May over simplify site’s policy
- IE6 relies heavily on compact policies for cookie filtering – especially an issue for third-party cookies
Server configuration

- Only needed for compact policies and/or sites that use P3P HTTP header
- Need to configure server to insert extra headers
- Procedure depends on server – see P3P Deployment Guide appendix http://www.w3.org/TR/p3pdeployment or Appendix B of Web Privacy with P3P
Don’t forget to test!

• Make sure you use the P3P validator to check for syntax errors and make sure files are in the right place [http://www.w3.org/P3P/validator/](http://www.w3.org/P3P/validator/)
  or [http://validator.privacyfinder.org/](http://validator.privacyfinder.org/)
  – But validator can’t tell whether your policy is accurate

• Use P3P user agents to view your policy and read their policy summaries carefully

• Test multiple pages on your site
Assertions in a P3P policy

• General assertions
  – Location of human-readable policies and opt-out mechanisms – discuri, opturi attributes of <POLICY>
  – Indication that policy is for testing only – <TEST> (optional)
  – Web site contact information – <ENTITY>
  – Access information – <ACCESS>
  – Information about dispute resolution – <DISPUTES> (optional)

• Data-Specific Assertions
  – Consequence of providing data – <CONSEQUENCE> (optional)
  – Indication that no identifiable data is collected – <NON-IDENTIFIABLE> (optional)
  – How data will be used – <PURPOSE>
  – With whom data may be shared – <RECIPIENT>
  – Whether opt-in and/or opt-out is available – required attribute of <PURPOSE> and <RECIPIENT>
  – Data retention policy – <RETENTION>
  – What kind of data is collected – <DATA>
P3P/XML encoding

```
<POLICIES xmlns="http://www.w3.org/2002/01/P3Pv1">
  <POLICY discuri="http://p3pbook.com/privacy.html" name="policy">
    <ENTITY>
      <DATA-GROUP>
        <DATA ref="#business.contact-info.online.email">privacy@p3pbook.com</DATA>
        <DATA ref="#business.contact-info.online.uri">http://p3pbook.com/</DATA>
        <DATA ref="#business.name">Web Privacy With P3P</DATA>
      </DATA-GROUP>
    </ENTITY>
    <ACCESS><nonident/></ACCESS>
    <STATEMENT>
      <CONSEQUENCE>We keep standard web server logs.</CONSEQUENCE>
      <PURPOSE><admin/><current/><develop/></PURPOSE>
      <RECIPIENT><ours/></RECIPIENT>
      <RETENTION><indefinitely/></RETENTION>
      <DATA-GROUP>
        <DATA ref="#dynamic.clickstream"/>
        <DATA ref="#dynamic.http"/>
      </DATA-GROUP>
    </STATEMENT>
  </POLICY>
</POLICIES>
```
Reading the P3P specification

• http://www.w3.org/TR/P3P11/