CyLab Privacy Interest Group
2006 Privacy Policy Trends Report
Final Report

Lorrie Faith Cranor
20 February 2006
Report Outline

- Executive summary
- Introduction
- Privacy policies of Popular and Random web sites
- Focus on financial industry
- Platform for Privacy Preferences
- Discussion
Popular and Random Websites

- Used list of 30,000 most clicked on domains from AOL search engine to create site lists

- Examined only .com domains that were not adult sites or kids sites (to be comparable with previous FTC studies)

- Popular list is top 100 domains on list that met above criteria

- Random list is 75 websites randomly selected from top 12,000 sites on that list, excluding sites not meeting above criteria (12,000 site sample frame selected based on growth of Internet since previous studies)
## Differences Between Random Sample Groups

<table>
<thead>
<tr>
<th></th>
<th>1999</th>
<th>2000</th>
<th>2001</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of sites surveyed</td>
<td>$n = 286$</td>
<td>$n = 281$</td>
<td>$n = 223$</td>
<td>$n = 100$</td>
</tr>
<tr>
<td>Privacy policy</td>
<td>48.3%</td>
<td>65.8%</td>
<td>76.7%</td>
<td>88%</td>
</tr>
<tr>
<td>Provides notice about what personal information is collected</td>
<td>49.7%</td>
<td>71.2%</td>
<td>73.5%</td>
<td>100%</td>
</tr>
<tr>
<td>Provides notice about disclosure to third parties</td>
<td>40.6%</td>
<td>N/A</td>
<td>N/A</td>
<td>83%</td>
</tr>
<tr>
<td>Provides access</td>
<td>27.6%</td>
<td>21.4%</td>
<td>N/A</td>
<td>94%</td>
</tr>
</tbody>
</table>
Differences Between Popular Sample Groups

<table>
<thead>
<tr>
<th></th>
<th>1998</th>
<th>1999</th>
<th>2000</th>
<th>2001</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of sites surveyed</td>
<td>(n = 105)</td>
<td>(n = 91)</td>
<td>(n = 87)</td>
<td>(n = 71)</td>
<td>(n = 75)</td>
</tr>
<tr>
<td>Privacy policy</td>
<td>44.8%</td>
<td>84.6%</td>
<td>96.6%</td>
<td>98.6%</td>
<td>96%</td>
</tr>
<tr>
<td>Provides notice about</td>
<td>N/A</td>
<td>73.6%</td>
<td>90.8%</td>
<td>95.8%</td>
<td>100%</td>
</tr>
<tr>
<td>what personal information is</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>collected</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provides notice about</td>
<td>58.1%</td>
<td>71.4%</td>
<td>N/A</td>
<td>N/A</td>
<td>80%</td>
</tr>
<tr>
<td>disclosure to third parties</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provides access</td>
<td>26.7%</td>
<td>41.8%</td>
<td>49.4%</td>
<td>N/A</td>
<td>95%</td>
</tr>
</tbody>
</table>

Carnegie Mellon • CMU Usable Privacy and Security Laboratory • Lorrie Cranor • http://lorrie.cranor.org/
Privacy Bird evaluation

- **Low:** Sites must not collect health info and share it with other companies or use it for analysis, marketing, or to make decisions that may affect what content or ads the user sees. Sites must not engage in marketing without opt-out. 86% of popular, 70% of random

- **Medium:** Same as low, plus sites must not share PII, financial info, or purchase info with other companies; and sites that collect personally identified data must provide access provisions. 58% of popular, 48% of random

- **High:** Same as medium, plus sites must not share any personal info or use it to determine the user’s habits, interests, or other characteristics; and sites may not contact users for marketing or use financial or purchase info for analysis, marketing, or to make decisions that may affect what content or ads the user sees. 31% of popular, 17% of random

- A site is classified as not sharing data if it shares data only under an opt-in policy or only with agents that use it only to complete the transaction for which it was provided or with delivery companies.

- Data from the following P3P categories are considered PII: physical contact info, online contact info, and government issued identifiers.
Data collected
Data use

- System administration
- Current activity
- Research & Development
- Pseudonymous analysis
- Marketing contact
- Pseudonymous decision
- Individual analysis
- Individual decision
- Tailoring
- Other
- Telemarketing

Popular: Collect Data, Unclear
Random: Collect Data, Unclear
Data recipients

- Our practices
- Third parties
- Public forums
- Delivery services
- Other

Collect Data / Unclear

[Bar chart showing data recipients]
Access provisions
Dispute resolution

- Customer service: Popular and Random are both high, indicating a preferred method.
- Independent organization: Popular is lower than Random, suggesting less preference.
- Applicable law: Both Popular and Random are very low, indicating minimal use.
- Court: Both Popular and Random are very low, indicating minimal use.
Data retention policy

Retention Policy

- Unclear
- Indefinitely
- Business practices
- Stated purpose
- Required by Law
- Not retained

Popular
Random
Kincaid readability score

- Popular - 12.4 (standard deviation = 1.8)
- Random - 12.5 (standard deviation = 2.3)
- White house press release - 4.1
- New York Times article - 6.2

  • Readability Info, “Readability Grades,”
Focus on financial industry

- The readability and clarity of policies for financial institutions has improved
- Banks minimally comply with GLB in terms of affiliate sharing
- GLB had little impact on the third party sharing choices available to consumers
- Details in report
Platform for Privacy Preferences Project (P3P)

- Developed by the World Wide Web Consortium (W3C) [http://www.w3.org/p3p/](http://www.w3.org/p3p/)
  - Final P3P1.0 Recommendation issued 16 April 2002

- Standard machine-readable format for web site privacy policies
  - Can be deployed using existing web servers

- Enables the development of tools (built into browsers or separate applications) that
  - Summarize privacy policies
  - Compare policies with user preferences
  - Alert and advise users

- P3P support built into IE6 and Netscape 7
Our P3P policy database

- Our P3P policy database is based on our Privacy Finder search engine cache.

- Whenever someone runs a search, the 10 displayed results get checked for P3P policies and added to our cache.

- We seeded the cache by running 20,000 queries provided by AOL and by checking for P3P policies at 30,000 most clicked on domains.

- We have checked 113,880 Web sites for P3P policies and collected 11,843 policies, including 3,846 unique policies.

- We revisit sites with P3P policies about once a day to see if the policy has changed.

- We revisit other sites about once a month to see if they have added a P3P policy.
# Longitudinal Trends

- Sharpest increase in government sites, probably due to the E-Government Act (1,465%)
- Other large changes in sites targeted to children (106%) and news sites (43.95%)

<table>
<thead>
<tr>
<th></th>
<th># in list</th>
<th>Sites reached in 2003</th>
<th>P3P-enabled in 2003</th>
<th>Sites reached in 2006</th>
<th>P3P-enabled in 2006</th>
<th>% change</th>
</tr>
</thead>
<tbody>
<tr>
<td>PFF Random</td>
<td>302</td>
<td>286</td>
<td>12.23%</td>
<td>282</td>
<td>10.99%</td>
<td>-10.14%</td>
</tr>
<tr>
<td>PFF Most Popular</td>
<td>85</td>
<td>84</td>
<td>30.95%</td>
<td>84</td>
<td>25.00%</td>
<td>-19.22%</td>
</tr>
<tr>
<td>PFF Refined Random</td>
<td>209</td>
<td>195</td>
<td>14.87%</td>
<td>195</td>
<td>12.82%</td>
<td>-13.79%</td>
</tr>
<tr>
<td>Key Measures</td>
<td>500</td>
<td>486</td>
<td>23.46%</td>
<td>474</td>
<td>23.63%</td>
<td>+0.72%</td>
</tr>
<tr>
<td>Netscore Top 500</td>
<td>500</td>
<td>488</td>
<td>22.95%</td>
<td>474</td>
<td>23.84%</td>
<td>+3.88%</td>
</tr>
<tr>
<td>Alexa</td>
<td>500</td>
<td>495</td>
<td>18.59%</td>
<td>470</td>
<td>18.51%</td>
<td>-0.43%</td>
</tr>
<tr>
<td>FirstGov</td>
<td>344</td>
<td>338</td>
<td>2.07%</td>
<td>321</td>
<td>32.40%</td>
<td>+1465.22%</td>
</tr>
<tr>
<td>Froogle</td>
<td>1017</td>
<td>1010</td>
<td>13.17%</td>
<td>964</td>
<td>12.55%</td>
<td>-4.71%</td>
</tr>
<tr>
<td>News</td>
<td>2429</td>
<td>2398</td>
<td>9.42%</td>
<td>2286</td>
<td>13.56%</td>
<td>+43.95%</td>
</tr>
<tr>
<td>Yahooligans!</td>
<td>900</td>
<td>868</td>
<td>3.00%</td>
<td>841</td>
<td>6.18%</td>
<td>+106.00%</td>
</tr>
<tr>
<td>Total</td>
<td>5856</td>
<td>5739</td>
<td>10.25%</td>
<td>5414</td>
<td>13.59%</td>
<td>+32.59%</td>
</tr>
</tbody>
</table>
Geographic distribution

- 437 of 8,661 P3P policies in database have non-US country-specific TLD
  - UK 91
  - Japan 49
  - Australia 48
  - Canada 35
  - Germany 22
  - Netherlands 22
  - Poland 18
  - France 15
  - South Africa 14
  - …

- 45% of non-US TLDs in EU
Data collected
(EU v. Non-EU/Non-US)
Data use (EU v. Non-EU/Non-US)

- System administration
- Marketing contact
- Current activity
- Research & Development
- Historical preservation
- Individual analysis
- Individual decision
- Other
- Pseudonymous analysis
- Pseudonymous decision
- Tailoring
- Telemarketing

European Union vs. Non-EU/Non-US
Data recipients (EU v. Non-EU/Non-US)

- Delivery services
- Other
- Public forums
- Our practices
- Third parties

Data Recipients

<table>
<thead>
<tr>
<th></th>
<th>European Union</th>
<th>Non-EU/Non-US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delivery services</td>
<td>25%</td>
<td>50%</td>
</tr>
<tr>
<td>Other</td>
<td>10%</td>
<td>25%</td>
</tr>
<tr>
<td>Public forums</td>
<td>5%</td>
<td>25%</td>
</tr>
<tr>
<td>Our practices</td>
<td>75%</td>
<td>100%</td>
</tr>
<tr>
<td>Third parties</td>
<td>10%</td>
<td>25%</td>
</tr>
</tbody>
</table>
Access provisions (EU v. Non-EU/Non-US)

Access Policy

- All data
- Contact and other
- Identified contact
- None
- Does not collect
- Other identified

European Union vs. Non-EU/Non-US
Dispute resolution (EU v. Non-EU/Non-US)
Data retention policy (EU v. Non-EU/Non-US)
P3P-enabled sites by category

- 16,919 sites categorized using Yahoo! categories
Rate of change of P3P policies

- Over an eight week span (10/25/06 - 12/20/06) we observed the following changes:
  - 69 policy changes
  - 70 policy “removals”
    - Sites temporarily or permanently unavailable
    - Sites that put up robots files that do not permit retrieval of P3P policy
      - (no sites appear to have explicitly removed P3P policy)
  - 470 new policies, of which 272 are unique

- Net growth rate of 4.16% annually
P3P Errors

- **Syntactic errors**
  - Policies do not follow the P3P specification
  - Some errors are “critical”
    - These policies cannot be parsed
  - Other errors are non-critical
    - We can determine the intent of these policies

- **Semantic errors**
  - P3P policy and natural language policy disagree
## P3P Syntactic Errors

<table>
<thead>
<tr>
<th>Error</th>
<th>Popular</th>
<th>Privacy Finder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Old Version</td>
<td>15 (71.4%)</td>
<td>9,155 (62.2%)</td>
</tr>
<tr>
<td>No Policy Name</td>
<td>3 (14.3%)</td>
<td>6,289 (42.7%)</td>
</tr>
<tr>
<td>No Errors</td>
<td>6 (28.6%)</td>
<td>4,014 (27.3%)</td>
</tr>
<tr>
<td>Policy Validation Error</td>
<td>1 (4.8%)</td>
<td>1,157 (7.9%)</td>
</tr>
<tr>
<td>Bad XML Root</td>
<td>8 (38.1%)</td>
<td>1,125 (7.6%)</td>
</tr>
<tr>
<td>Policy Expired</td>
<td>0</td>
<td>474 (3.2%)</td>
</tr>
<tr>
<td>Policy Vocabulary Error</td>
<td>0</td>
<td>453 (3.1%)</td>
</tr>
<tr>
<td>No Policy Elements</td>
<td>0</td>
<td>252 (1.7%)</td>
</tr>
<tr>
<td>Incorrect XML</td>
<td>0</td>
<td>204 (1.4%)</td>
</tr>
<tr>
<td>Policy Access Error</td>
<td>1 (4.8%)</td>
<td>183 (1.2%)</td>
</tr>
<tr>
<td>No Namespace</td>
<td>0</td>
<td>151 (1.0%)</td>
</tr>
<tr>
<td>Malformed INCLUDE/EXCLUDE</td>
<td>0</td>
<td>56 (0.4%)</td>
</tr>
<tr>
<td>No &lt;META/&gt; Tag</td>
<td>0</td>
<td>21 (0.1%)</td>
</tr>
<tr>
<td>No Policy Found</td>
<td>0</td>
<td>5 (0%)</td>
</tr>
<tr>
<td>Not A Policy</td>
<td>0</td>
<td>2 (0%)</td>
</tr>
<tr>
<td><strong>Total Policies</strong></td>
<td><strong>21</strong></td>
<td><strong>14,720</strong></td>
</tr>
</tbody>
</table>
P3P Syntactic Errors

- Most (over 71%) policies have errors
  - Many are non-critical

- Most (~70%) errors are due to old versions
  - Little change in syntax
  - Many policies missing the name fragment
## Semantic Errors

- We compared P3P policies with our manual coding
- Differences may be due to one policy being more specific
- Some differences are conflicts indicating an error in one or both policies

<table>
<thead>
<tr>
<th>Policy</th>
<th>ACCESS</th>
<th>CATEGORIES</th>
<th>DISPUTES</th>
<th>NON-IDENTIFIABLE</th>
<th>PURPOSE</th>
<th>RECIPIENT</th>
<th>REMEDIES</th>
<th>RETENTION</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>yahoo.com</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>geocities.com</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>hotmail.com</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>superpages.com</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>16</td>
</tr>
<tr>
<td>angelfire.com</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>walmart.com</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>go.com</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>microsoft.com</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>ticketmaster.com</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>usps.com</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>dealtime.com</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>rootsweb.com</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>hgtv.com</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>wachovia.com</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>tripod.com</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>sportsline.com</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>qvc.com</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>download.com</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td>usatoday.com</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>about.com</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>wunderground.com</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>Policies with Error</td>
<td>9</td>
<td>19</td>
<td>5</td>
<td>1</td>
<td>21</td>
<td>18</td>
<td>5</td>
<td>11</td>
<td>217</td>
</tr>
</tbody>
</table>
Semantic Errors

- **<ACCESS>** errors
  - 43% specified different policies for reviewing personal information

- **<CATEGORIES>** errors
  - 80% showed different information collected

- **<DISPUTES>** errors
  - Very few
  - Most due to non-reporting of third parties (e.g. TrustE)

- **<NON-IDENTIFIABLE>** errors
  - Only ticketmaster.com used this incorrectly
Semantic Errors

■ <PURPOSE> errors
  • 8 policies mention marketing not mentioned in P3P policy
  • 5 policies mention telemarketing not mentioned in natural language policy

■ <RECIPIENT> errors
  • Significant differences
  • 15 natural language policies specify recipients not in the P3P policies
  • 3 P3P policies overly report sharing
  • 3 P3P policies accurately report sharing
Semantic Errors

- `<RETENTION>` errors
  - No natural language policies mention retention
    - Eleven sites require it from their P3P policies
Error Summary

 Syndactic errors show P3P policies not being adequately checked and maintained

 Semantic errors show misunderstanding of P3P
  • Wachovia claims not to use P3P
    ▪ P3P policy is available
    ▪ Only references cookies from their site
  • <financial> tag overused
    ▪ Only meant for information beyond purchases
  • Many overly-broad P3P policies
    ▪ Five policies mention concerns not in their natural language policies
      – Worst case scenarios
      – Benefits users

 Errors have little affect on high/medium/low settings in Privacy Finder and Privacy Bird
We want your feedback

- What parts of this report are most useful to you?
- What would you like to see us update next year?
- What new data would you like to see us collect next year?
- Collecting P3P data is easy, coding natural language policies is hard - does data on natural language policies provide value to you?
CMU Usable Privacy and Security Laboratory
http://cups.cs.cmu.edu/
Carnegie Mellon