How Private Is Your Privacy
&
What Can You Do About It?

Indrajit Ray
Colorado State University
indrajit@cs.colostate.edu
http://www.cs.colostate.edu/~indrajit
Privacy of Personal Information

• Personal information is more than just one’s name, address or social security number
  – Medical history, Work history, Shopping habits, Driving record, Political views, Sexual orientation, Credit score

• Privacy is an interest that the user has in maintaining such information securely under control without that control being compromised by others for their personal gain

• Privacy violations occur everybody without us being aware of it
Outline of Presentation

• Case study of real life incidents
  – Identify fraud
  – Demographic re-identification
• How privacy violations occur
• Privacy enhancing technologies
• Research at Colorado State University
• Conclusions
Privacy Violation Case Study #1

Identity Fraud from Identify Theft
Identity Fraud – The Michelle Brown Story

- Started with mishandling of rental application of Michelle Brown at property management office and subsequent theft
  - Perpetrator obtains duplicate driver’s license with assumed identity
  - Sets up cellular service, residential telephone service, utility services
  - Obtains department store credit card and other loans
  - $32,000 truck
  - Rents property under assumed identity
  - Gets $5,000 worth of liposuction on her body
Identify Fraud – The Michelle Brown Story (2)

• Perpetrator runs drug trafficking business under assumed identity
  – Presents forged identification papers when arrested for carrying 1,300 Kg of marijuana
    • Results in erroneous criminal record under Michelle Brown’s name

• Perpetrator becomes fugitive
  – Warrant for Michelle Brown’s arrest

• Perpetrator arrested
  – Prison record established in Michelle Brown’s name
Identify Fraud – The Michelle Brown Story (3)

- Real Michelle Brown eventually cleared after months of anguish and enormous financial drain
- However, months later was wrongly stopped by an international airport’s customs agents and held for several hours
Analysis of Data Breaches in 2006

<table>
<thead>
<tr>
<th>Cause of Incident</th>
<th>Private Sector (incidents n = 126)</th>
<th>Public Sector (incidents n = 114)</th>
<th>Higher Education (incidents n = 52)</th>
<th>Medical Centers (incidents n = 30)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outside Hackers</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Insider</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td>10</td>
</tr>
<tr>
<td>Malfeasance</td>
<td>15</td>
<td>30</td>
<td>40</td>
<td>15</td>
</tr>
<tr>
<td>Human / Software</td>
<td>20</td>
<td>40</td>
<td>50</td>
<td>20</td>
</tr>
<tr>
<td>Incompetence</td>
<td>25</td>
<td>50</td>
<td>60</td>
<td>25</td>
</tr>
<tr>
<td>Theft (non laptop)</td>
<td>30</td>
<td>60</td>
<td>70</td>
<td>30</td>
</tr>
<tr>
<td>Laptop Theft</td>
<td>40</td>
<td>80</td>
<td>90</td>
<td>40</td>
</tr>
</tbody>
</table>

Source - Beth Rosenberg (Sandstorm.net). Available from Privacy Rights Clearinghouse [www.privacyrights.org](http://www.privacyrights.org)
## Analysis of Privacy Breaches in 2006

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Number of Major Data Breach Incidents reported</td>
<td>327</td>
</tr>
<tr>
<td>Approximate Minimum Total # of Personal Records Potentially Compromised in 2006</td>
<td>100,453,730</td>
</tr>
<tr>
<td># Data-Breach Identity Thieves Sentenced in 2006</td>
<td>5</td>
</tr>
<tr>
<td># Individual Victims of Sentenced Identity Thieves</td>
<td>238</td>
</tr>
</tbody>
</table>
## Average Cost of Data Breach

**Total # of affected records = 250,000**

<table>
<thead>
<tr>
<th>Internal Investigation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cybercrime Consulting</td>
<td>533,508.0</td>
</tr>
<tr>
<td>Attorney Fees</td>
<td>540,930.5</td>
</tr>
<tr>
<td></td>
<td><strong>1,074,439.0</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Notification / Crisis Management</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer notification (Certified mail)</td>
<td>983,510.5</td>
</tr>
<tr>
<td>Call center support</td>
<td>695,880.0</td>
</tr>
<tr>
<td>Crisis management consulting</td>
<td>389,693.0</td>
</tr>
<tr>
<td>Media management</td>
<td>77,010.5</td>
</tr>
<tr>
<td></td>
<td><strong>2,146,095.0</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Regulatory / Compliance</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Credit monitoring for affected customers</td>
<td>4,472,189.0</td>
</tr>
<tr>
<td>Regulatory investigation defense</td>
<td>1,654,338.5</td>
</tr>
<tr>
<td>State / Federal fines or fees</td>
<td>3,509,091.0</td>
</tr>
<tr>
<td></td>
<td><strong>9,635,619.0</strong></td>
</tr>
<tr>
<td></td>
<td><strong>12,856,153.0</strong></td>
</tr>
</tbody>
</table>

Privacy Violation Case Study #2

Demographic Re-identification
&
Its Consequences
Latanya Sweeney’s Work (1)

• In Massachusetts, the Group Insurance Commission (GIC) is responsible for purchasing health insurance for state employees

• GIC has to publish the data for research purposes

GIC (zip, dob, gender, diagnosis, procedure, ...)

Dr. Indrajit Ray, Associate Professor, Computer Science Department
Latanya Sweeney’s Work (2)

• Sweeney paid $20 and bought the voter registration list for Cambridge, MA

GIC (zip, dob, gender, diagnosis, procedure, …)

VOTER (name, party, …, zip, dob, gender)
Latanya Sweeney’s Work (3)

- William Weld (former governor) lives in Cambridge, hence is in VOTER
- 6 people in VOTER share his date of birth
- Only 3 of them were men (same gender)
- Weld was the only one in that zip
- Sweeney learned Weld’s medical records
Medical Records Misuse

• Burlington Northern allegedly obtained genetic tests results on employees who had filed worker’s compensation claims for carpal tunnel syndrome, without their knowledge.
  – The company’s intention was presumably to be able to reject some claims because of genetic predisposition to the condition.

So, How Do They Get My Data?

Beyond Phishing and Other Social Engineering Means
Lack of Knowledge About Privacy Threats

- Users misplace valuable data items
- Users post many personal details on social network sites like Facebook, MySpace, Bebo etc
  - Can be combined with information from other sources to create profile
Carnegie Melon University Heinz School Study of Facebook

Figure 2: Percentages of CMU profiles revealing various types of personal information.
Workplace Monitoring

• 75% of employers monitor their employees website visit
  – Most computer monitoring equipment allows monitoring remotely without user’s knowledge

• Almost all employers review employee email
  – Deleted emails are not really deleted

• 33% track keystrokes and time spent at the keyboard

• Currently there are very few laws regulating employee monitoring
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
Monitoring on the Internet – What Your Browsing Reveals

Privacy.Net Browsing Analysis Results
The Privacy.net Analyzer

This site analyzes the privacy of your Internet connection and shows some of the information web sites can know about you when you visit. The information can be used to display web content based on things such as country of origin and web browser.

Click here to read a description of the tests (opens new window)

You Are Visiting From:
128.93.62.86 is from France (FR) in Western Europe
Host name: dhcp-rocq-86.inria.fr

Cookie Test
No Cookie from this site is on your system from prior visits.

You linked from here
http://network-tools.com/analyze/

Your Browser Type and Operating System:
Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10.5; en-US; rv:1.9.1.7) Gecko/20091221 Firefox/3.5.7

All information sent by your web browser when requesting this web page:

Connection: keep-alive Keep-Alive: 300 Content-Length: 607 Content-Type: application/x-www-form-urlencoded Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8 Accept-Charset: UTF-8,* Accept-Encoding: gzip,deflate Accept-Language: en-us,en;q=0.5 Cookie: bhCookieSaveSess=1; bhCookieSess=1; Privacy.net_Last_Visit=1/18/2010; Privacy.net=Privacy+Analysis; bhawkplt=plt_state=tested&plt_stm=1263815352253&plt_url=null Host: analyze.privacy.net Referer: http://analyze.privacy.net/test.asp?RequestCookies=&Requestdate=&refer=http://network-tools.com/analyze/ User-Agent: Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10.5; en-US; rv:1.9.1.7) Gecko/20091221 Firefox/3.5.7

Firewall Test
The following ports were checked: 16771, 80
Out of the above ports, the following are open and permitting outbound traffic: 16771, 80

Firewall status: PRESENT

Browser Type and Version
Browser: Firefox
Fullversion: 3.5.7
Gecko: True
GeckoBuildDate: 20091221
Crawler: False

Display and Layout
Width: 1440
WidthAvail: 1100
Height: 900
StyleSheet: True

System Details
Platform: MacOSX
Win16: False
WinInstallerMinVer: 0
Browser Security
Session Cookies Not Accepted
Persistent Cookies Accepted
JavaScriptEnabled: True
VBScriptEnabled: False
JavaEnabled: True
ActiveXEnabled: False
SSL: True
SSLActive: False
SSLKeySize: 128
SSLEnabled: True
Firewall: True
OpenPorts: 16771,80
PopupsBlocked: True
ImagesEnabled: True
HighSecurity: False

Connection Details
Broadband: True
Connection Type:
Firewall: True
Proxy: False
CompressGZip: True
AOL: False
MSN: False

Scripting Capabilities
ActiveXControls: False
ActiveXEnabled: False
JavaScript: True
JavaScriptEnabled: True
JavaScriptVer: 1.8
JavaScriptBuild:
VBScript: False
VBScriptEnabled: False
VBScriptBuild:
XML: True
MSXML: 0
XMLHttpRequest: True
DHTML: True
FileUpload: Yes

<table>
<thead>
<tr>
<th>Hop</th>
<th>(ms)</th>
<th>(ms)</th>
<th>(ms)</th>
<th>IP Address</th>
<th>Host name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td>192.168.255.1</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>6</td>
<td>6</td>
<td>67.222.132.1</td>
<td>router.dfw-datacenter.com</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>14</td>
<td>13</td>
<td>72.249.128.105</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>6</td>
<td>11</td>
<td>206.123.64.82</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>10</td>
<td>15</td>
<td>64.129.174.181</td>
<td>64-129-174-181.static.twtelecom.net</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>45</td>
<td>49</td>
<td>66.192.242.253</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>117</td>
<td>149</td>
<td>122</td>
<td>195.2.25.198</td>
<td>xe-3-1-0-xcr1.lnd.cw.net</td>
</tr>
<tr>
<td>8</td>
<td>186</td>
<td>128</td>
<td>128</td>
<td>195.2.21.210</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>132</td>
<td>127</td>
<td>128</td>
<td>195.2.9.198</td>
<td>xe-0-1-0.xcr1.prp.cw.net</td>
</tr>
<tr>
<td>10</td>
<td>123</td>
<td>137</td>
<td>137</td>
<td>195.2.9.189</td>
<td>xe-0-1-0.xcr1.par.cw.net</td>
</tr>
<tr>
<td>11</td>
<td>142</td>
<td>138</td>
<td>154</td>
<td>195.10.54.66</td>
<td>giprenater-gw.par.cw.net</td>
</tr>
<tr>
<td>12</td>
<td>148</td>
<td>140</td>
<td>146</td>
<td>193.51.189.38</td>
<td>-</td>
</tr>
</tbody>
</table>

Plug-in Information
Plugin Flash Version: 10.0 r22
Plugin Flash Version: 11.5.1
Plugin QuickTime Version: Installed (version 7.6.4)
Plugin RealPlayer Version: 10.1.0.503
Plugin MediaPlayer Version: Installed but version not detected
Plugin Flip4Mac installed

Java Information
JavaApplets: True
JavaEnabled: True

Wireless Device Information
PDA: False
WAP: False
HDML: False

Locale Information
Country: FR
Language: English
User Language: en-us
System Language:
Time Zone Difference: 6
Browser Date and Time: Mon Jan 18 12:49:11 2010
Browser Date and Time ms: 1263815351981

TraceRoute to 128.93.62.86 [dhcp-rocq-86.inria.fr]
Trace aborted.

Checking the domain name records of your connection:

whois query for inria.fr...

Results returned from whois.nic.fr:

%%
%% This is the AFNIC Whois server [clyde.nic.fr].
%%
%% complete date format : DD/MM/YYYY
%% short date format : DD/MM
%% version : FRNIC-2.5
%%
%% Rights restricted by copyright.
%% See http://www.afnic.fr/afnic/web/mentions-legales-whois_en
%%
%% Use `-h' option to obtain more information about this service.
%%
%% [67.222.132.194 REQUEST] >> -n inria.fr
%%
%% RL Net [############] - RL IP [############]
%%

domain: inria.fr
status: ACTIVE
hold: NO
holder-c: INDR18-FRNIC
admin-c: DT5-FRNIC
technical-c: GR1378-FRNIC
technical-c: PB2340-FRNIC
zone-c: NFC1-FRNIC
nsl-id: NSL3778-FRNIC
registrar: RENATER
anniversary: 01/01
created: 01/01/1995
last-update: 11/03/2009
source: FRNIC

ns-list: NSL3778-FRNIC
nservers: dns.inria.fr [193.51.208.13]
nserver: nez-perce.inria.fr [192.93.2.78]
nserver: dns.cs.wisc.edu
nservers: dns.princeton.edu
nservers: imag.imag.fr [129.88.30.1]
nserver: ns2.nic.fr [192.93.0.4 2001:660:3005:1::1:2]
sources: FRNIC
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>registrar</td>
<td>RENATER</td>
</tr>
<tr>
<td>type</td>
<td>Isp Option 1</td>
</tr>
<tr>
<td>address</td>
<td>151 Boulevard de l'hôpital</td>
</tr>
<tr>
<td>address</td>
<td>PARIS</td>
</tr>
<tr>
<td>address</td>
<td>INRIA Sophia-Antipolis</td>
</tr>
<tr>
<td>address</td>
<td>2004, route des Lucioles</td>
</tr>
<tr>
<td>address</td>
<td>B.P. 93</td>
</tr>
<tr>
<td>address</td>
<td>06902 Sophia Antipolis Cedex</td>
</tr>
<tr>
<td>country</td>
<td>FR</td>
</tr>
<tr>
<td>phone</td>
<td>+33 1 53 94 20 30</td>
</tr>
<tr>
<td>phone</td>
<td>+33 4 92 38 77 22</td>
</tr>
<tr>
<td>fax-no</td>
<td>+33 1 53 94 20 31</td>
</tr>
<tr>
<td>fax-no</td>
<td>+33 4 92 38 76 02</td>
</tr>
<tr>
<td>e-mail</td>
<td><a href="mailto:domaine@renater.fr">domaine@renater.fr</a></td>
</tr>
<tr>
<td>e-mail</td>
<td><a href="mailto:daniel.terrer@sophia.inria.fr">daniel.terrer@sophia.inria.fr</a></td>
</tr>
<tr>
<td>e-mail</td>
<td><a href="mailto:rensvp@renater.fr">rensvp@renater.fr</a></td>
</tr>
<tr>
<td>changed</td>
<td>15/10/2000 <a href="mailto:migration-dbm@nic.fr">migration-dbm@nic.fr</a></td>
</tr>
<tr>
<td>anonymous</td>
<td>NO</td>
</tr>
<tr>
<td>registered</td>
<td>01/01/1998</td>
</tr>
<tr>
<td>source</td>
<td>FRNIC</td>
</tr>
<tr>
<td>nic-hdl</td>
<td>DT5-FRNIC</td>
</tr>
<tr>
<td>type</td>
<td>PERSON</td>
</tr>
<tr>
<td>contact</td>
<td>Daniel Terrer</td>
</tr>
<tr>
<td>address</td>
<td>INRIA Sophia-Antipolis</td>
</tr>
<tr>
<td>address</td>
<td>2004, route des Lucioles</td>
</tr>
<tr>
<td>address</td>
<td>B.P. 93</td>
</tr>
<tr>
<td>address</td>
<td>06902 Sophia Antipolis Cedex</td>
</tr>
<tr>
<td>country</td>
<td>FR</td>
</tr>
<tr>
<td>phone</td>
<td>+33 4 92 38 77 22</td>
</tr>
<tr>
<td>phone</td>
<td>+33 4 92 38 76 02</td>
</tr>
<tr>
<td>e-mail</td>
<td><a href="mailto:daniel.terrer@sophia.inria.fr">daniel.terrer@sophia.inria.fr</a></td>
</tr>
<tr>
<td>changed</td>
<td>15/10/2000 <a href="mailto:migration-dbm@nic.fr">migration-dbm@nic.fr</a></td>
</tr>
<tr>
<td>anonymous</td>
<td>NO</td>
</tr>
<tr>
<td>obsoleted</td>
<td>NO</td>
</tr>
<tr>
<td>source</td>
<td>FRNIC</td>
</tr>
<tr>
<td>nic-hdl</td>
<td>GR1378-FRNIC</td>
</tr>
<tr>
<td>type</td>
<td>ROLE</td>
</tr>
<tr>
<td>contact</td>
<td>GIP RENATER</td>
</tr>
<tr>
<td>address</td>
<td>GIP RENATER</td>
</tr>
<tr>
<td>address</td>
<td>Ensam</td>
</tr>
<tr>
<td>address</td>
<td>151, boulevard de l'Hopital</td>
</tr>
<tr>
<td>address</td>
<td>75013 Paris</td>
</tr>
<tr>
<td>country</td>
<td>FR</td>
</tr>
<tr>
<td>phone</td>
<td>+33 1 53 94 20 30</td>
</tr>
<tr>
<td>phone</td>
<td>+33 1 53 94 20 31</td>
</tr>
<tr>
<td>e-mail</td>
<td><a href="mailto:rensvp@renater.fr">rensvp@renater.fr</a></td>
</tr>
<tr>
<td>trouble</td>
<td>Information: <a href="http://www.Renater.fr/">http://www.Renater.fr/</a></td>
</tr>
<tr>
<td>trouble</td>
<td>abuse reports, ... <a href="mailto:CertSVP@Renater.fr">mailto:CertSVP@Renater.fr</a></td>
</tr>
<tr>
<td>trouble</td>
<td>questions: <a href="mailto:RenSVP@Renater.Fr">mailto:RenSVP@Renater.Fr</a></td>
</tr>
<tr>
<td>admin-c</td>
<td>DV321-FRNIC</td>
</tr>
<tr>
<td>tech-c</td>
<td>FS65-FRNIC</td>
</tr>
<tr>
<td>tech-c</td>
<td>BT261-FRNIC</td>
</tr>
<tr>
<td>tech-c</td>
<td>MD5336-FRNIC</td>
</tr>
<tr>
<td>tech-c</td>
<td>SJ94-FRNIC</td>
</tr>
<tr>
<td>tech-c</td>
<td>ED168-FRNIC</td>
</tr>
<tr>
<td>tech-c</td>
<td>CT1053-FRNIC</td>
</tr>
<tr>
<td>tech-c</td>
<td>SEH1-FRNIC</td>
</tr>
<tr>
<td>tech-c</td>
<td>FXA1-FRNIC</td>
</tr>
<tr>
<td>changed</td>
<td>08/11/2005 <a href="mailto:rensvp@renater.fr">rensvp@renater.fr</a></td>
</tr>
<tr>
<td>anonymous</td>
<td>NO</td>
</tr>
<tr>
<td>obsoleted</td>
<td>NO</td>
</tr>
<tr>
<td>source</td>
<td>FRNIC</td>
</tr>
<tr>
<td>nic-hdl</td>
<td>INDR18-FRNIC</td>
</tr>
</tbody>
</table>
Checking domain configuration records. This includes where e-mail is processed for the domain:

127.0.0.1 xxx dhcp-rocq-86.inria.fr

Retrieving DNS records for dhcp-rocq-86.inria.fr...

DNS servers
dns.inria.fr [193.51.208.13]
dns.princeton.edu
dns.cs.wisc.edu
nez-perce.inria.fr [192.93.2.78]
ns2.nic.fr
imag.imag.fr

Answer records
dhcp-rocq-86.inria.fr 1 A  128.93.62.86 3600s

Authority records
inria.fr 1 NS imag.imag.fr 3600s
inria.fr 1 NS dns.princeton.edu 3600s
inria.fr 1 NS ns2.nic.fr 3600s
inria.fr 1 NS dns.cs.wisc.edu 3600s
inria.fr  1 NS  nez-perce.inria.fr  3600s
inria.fr  1 NS  dns.inria.fr  3600s

Additional records
dns.inria.fr  1 A  193.51.208.13  3600s
imag.imag.fr  1 A  129.88.30.1  7200s
nez-perce.inria.fr  1 A  192.93.2.78  3600s

Checking who manages your IP address:

whois query for 128.93.62.86...

Results returned from whois.arin.net:

OrgName:  RIPE Network Coordination Centre
OrgID:  RIPE
Address:  P.O. Box 10096
City:  Amsterdam
StateProv:  
PostalCode: 1001EB
Country:  NL

ReferralServer: whois://whois.ripe.net:43

inetnum:  128.93.0.0 - 128.93.255.255
netname:  INRIA-NET
descr:  Institut National de Recherche en Informatique et Automatique
The following is a list of all fonts installed on your computer:

Font detection requires IE 5 or higher.
Click to Run The Test Again (Don't use refresh)

Trace other computers on the Internet using Network-Tools.com

Back to Privacy.net
Linking With Cookies

Ad company can get your name and address from CD order and link them to your search.
On-line 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
Should We Be Worried?

Users given no meaningful choice
Few sites offer alternatives
Personal Information = Real Money to Companies

- 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 after complaints from privacy advocates.
  - Stocks dropped from $125 (12/99) to $80 (03/00)
Privacy International’s Privacy Ranking of Internet Service Companies
<table>
<thead>
<tr>
<th>Company</th>
<th>Corporate Leadership</th>
<th>Data Collection and Processing</th>
<th>Data Retention</th>
<th>Openness and Transparency</th>
<th>Responsiveness</th>
<th>Ethical Compass</th>
<th>Customer Control</th>
<th>Fair Gateways</th>
<th>Privacy Enhancing or Invading Innovations</th>
<th>Initial Assessment</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon</td>
<td>Whiteform access to email for those with privacy problems. No postal address given. Must be signed in as an account holder in order to complain.</td>
<td>Privacy notice describes some of processing practices. Does not discuss what is done with ‘clickstream’ and ‘cookie’ data, i.e. whether Amazon tracks usage, popularity, and then profiles.</td>
<td>No information readily available</td>
<td>Policy lacking in information about how information is used to profile customers.</td>
<td>Previously Amazon has been reluctant to introduce privacy measures. Firm seems to have responded to earlier problems.</td>
<td>Customers may close accounts, but only possible through an email sent to Amazon.</td>
<td>Offers the choice to use anonymous or pseudonymous profiles and even informs customers of a variety of RET tools. Amazon Prime accounts offer greater services for an annual fee. Not mandatory and other customers are not penalised.</td>
<td>No privacy enhancing innovations apparent through points to privacy services from other companies.</td>
<td>No privacy enhancing innovations apparent through points to privacy services from other companies.</td>
<td>Amazon has improved much over the years but consumers should be informed on how their clicking, reading, and purchase habits are profiled and used.</td>
<td></td>
</tr>
<tr>
<td>AOL</td>
<td>Contact only available via email at <a href="mailto:privacyqueststore@aol.com">privacyqueststore@aol.com</a> (though with a separate email address for Californians, at <a href="mailto:C4APrivacyInfoAR@aol.com">C4APrivacyInfoAR@aol.com</a>).</td>
<td>Tracks user movements and use of resources. Monitors which e-mails you open and act upon. Monitors searches and how these searches were acted upon. Keeps a track of history of items purchased across AOL services. Supplements data from other firms. Collects IP address and geographic information. Researches use of AOL services, using cookies and web beacons.</td>
<td>No information readily available</td>
<td>Policy is relatively open about the fact that there is personal information processing but is lacking in information about how.</td>
<td>Leakage of search engine data was responded to poorly as though it was not privacy invasive. Investigations showed otherwise.</td>
<td>Poor. Disclosed search data to U.S. Department of Justice for research purposes.</td>
<td>Closing account is possible but nothing is said about how long personal data is kept for afterwards.</td>
<td>Account-only access in many areas of site. Differentiates between different users (e.g. Apple users are prevented from viewing video view content).</td>
<td>No information readily available, though does use web beacons to track users activities.</td>
<td>No privacy enhancing innovations apparent through points to privacy services from other companies.</td>
<td></td>
</tr>
<tr>
<td>Apple</td>
<td>Apple Computer, 1 Infinite Loop, MS65-DR, Cupertino, California, USA, 95014 Privacy policy last updated in 2004. Numerous email addresses given based on geographic region including <a href="mailto:privacy@apple.com">privacy@apple.com</a> and <a href="mailto:privacyar@apple.com">privacyar@apple.com</a>.</td>
<td>Can limit information processing. Shares data with other companies to “manage and enhance customer data”. Collects ‘clickstream’ data. Does not consider IP address as personal information. Also collect ‘ VisitThrough’ data. Ministore collected list of music on home computers.</td>
<td>No specific justification for deletion period. Does not consider itself responsible for data posted in forums, as an result is unlikely to anonymise or delete at any time.</td>
<td>Very little information is available. Though privacy policy states an optimistic tone on data collection, but does not explain whether there is any profiling and marketing activities?</td>
<td>Kept quiet on the potential watermarking of DRM-free iTunes songs. They did respond eventually to the ‘ministore’ controversy. Subject access requests are said to be available according to the policy, by email.</td>
<td>May opt-out of some services. May not access free iTunes services without registering.</td>
<td>Certain features of the Apple website will not be available once cookies are disabled.</td>
<td>Profiles use of music in ‘Ministore’. Mentions privacy enhancing precautions, but no information on technologies. Uses cookies and other technologies to track users. Uses “pixel tags” to identify whether individuals have read emails.</td>
<td>No information readily available</td>
<td>Substantial Threat</td>
<td></td>
</tr>
<tr>
<td>BBC</td>
<td>Data Protection Officer, NC1 01, Media Village, 201 Wood Lane, London, W12 7TQ and email at <a href="mailto:dpsofficer@bbc.co.uk">dpsofficer@bbc.co.uk</a></td>
<td>Use cookies to track movements. Uses Nielsen and SageMetrics cookies to track readership.</td>
<td>Declares in some cases how long personal information is kept.</td>
<td>Privacy policy is relatively explicit about each cookie, describing its detail.</td>
<td>Exports how to opt-out of cookies.</td>
<td>No evidence yet. Change 15 GBP for access to records.</td>
<td>Explains how to opt-out of cookies.</td>
<td>No information readily available</td>
<td>Generally privacy aware</td>
<td>Bare in its openness about processing, what it is, and how to access data and manage cookies.</td>
<td></td>
</tr>
<tr>
<td>Bebo</td>
<td>Customer Support, Bebo, Inc.142 Tenth Street, San Francisco, CA 94103,USA</td>
<td>Co-operates with Child Online Protection Police in UK, after encountering problem cases.</td>
<td>N/A, email address, IP address, age, hobbies, and interests and other content, such as photos. Does not consider IP addresses as personal information.</td>
<td>No information readily available</td>
<td>Inconsistencies in privacy policy. Lacks detail.</td>
<td>Responded to concerns about privacy problems (linked with child safety) but ensuring access is limited to certain age groups.</td>
<td>Can end membership. Can limit information available to people.</td>
<td>Company decides who can contact users based on their age.</td>
<td>No information readily available</td>
<td>Notable lapses.</td>
<td>Poor problems has led to some innovation. Lack of information is problematic. User control increasing.</td>
</tr>
<tr>
<td>Company</td>
<td>Company administrative details</td>
<td>Corporate Leadership</td>
<td>Data Collection and Processing</td>
<td>Data Retention</td>
<td>Openness and Transparency</td>
<td>Responsiveness</td>
<td>Ethical Compass</td>
<td>Customer Control</td>
<td>Fair Gateways</td>
<td>Privacy Enhancing or Invading Innovations</td>
<td>Initial Assessment</td>
</tr>
<tr>
<td>-----------------</td>
<td>---------------------------------</td>
<td>-----------------------</td>
<td>-------------------------------</td>
<td>----------------</td>
<td>--------------------------</td>
<td>---------------</td>
<td>----------------</td>
<td>-----------------</td>
<td>--------------</td>
<td>-----------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>eBay</td>
<td>eBay Inc. Attn: Legal Global Privacy Practices, 2145 Hamilton Avenue, San Jose, California 95129, and via a customer form</td>
<td>Member of Trust-e.</td>
<td>Information collection from other companies included.</td>
<td>No information readily available</td>
<td>Remarkable level of information about how data is shared.</td>
<td>Very responsive to privacy concerns. Changed practice to allow for customer account deletion.</td>
<td>Can opt out of marketing and advertising. Can reject cookies through may have some effects.</td>
<td>Can gain access to much information without authenticating.</td>
<td>Uses web beacons. A lot of the cookies are only session cookies. Anonymised or de-identified information is shared.</td>
<td>Generally privacy aware</td>
<td>Good responsiveness. Web beacons and lack of information on retention detracts from score.</td>
</tr>
<tr>
<td>Facebook</td>
<td>156 University Avenue, Palo Alto, CA 94301; and <a href="mailto:privacy@facebook.com">privacy@facebook.com</a></td>
<td>Member of Trust-e. Signed-up to safe harbor.</td>
<td>Earlier concerns about data matching, data mining and transfers to other companies. Collects data from other sources, including newspapers, blogs, instant messaging services, and other users of the Facebook service through the operation of the service (e.g. photo tags).</td>
<td>No information readily available</td>
<td>Basic privacy policy.</td>
<td>User has responded to some (of many) concerns about security and privacy.</td>
<td>Purports to have two principles: 1. you have control over personal information. 2. you have access to info others want to share. But track history indicates otherwise.</td>
<td>Unable to fully opt out of controversial 'news feed' services. Cookies can be blocked. Many are session cookies. Profiles are only accessible based on privacy settings, though name and profile-photo is available to all.</td>
<td>In 2005 a number of profiles were downloaded to prove weak security. Does not accept liability for security.</td>
<td>No information readily available</td>
<td>Substantial Threat</td>
</tr>
<tr>
<td>Friendster</td>
<td>No specific privacy contact point. General address is given as Friendster, Inc., 568 Howard Street San Francisco, CA 94105 Fax: (415) 638-0074</td>
<td></td>
<td>Personal information types collected through consent and without consent (e.g., IP address). Promises not to share personally identifiable information with third parties. Third party cookies are possible.</td>
<td>No information readily available</td>
<td>Open privacy policy, though vague at times.</td>
<td>User may choose to share with 'friends', Friends of Friends', and anyone, including non-Friendster members. Some profile information is shared with everyone.</td>
<td>Rejecting cookies may prevent access to website.</td>
<td>Access to personal information is said to be limited even to employees.</td>
<td>No information readily available</td>
<td>Variable lapses</td>
<td>Insignificant information to draw compelling conclusions. Lack of main point of contact is problematic.</td>
</tr>
<tr>
<td>Google</td>
<td>Privacy Matters, c/o Google Inc., 1600 Amphitheatre Parkway, Mountain View CA 94043 (USA). Policy not updated since 2005.</td>
<td></td>
<td>Describes data collected. IP addresses are not considered personal information. They do not believe that they collect sensitive information. Do sometimes track links clicked upon. Shares information with consent, or to companies (subordinates, affiliated companies, trusted businesses or persons).</td>
<td>No information readily available</td>
<td>Vague, incomplete and possibly deceptive privacy policy. Document fails to explain detailed data processing elements or information flows.</td>
<td>Generally poor track record of responding to customer complaints. Antiparadox attitude to privacy challenges (for example, complaints to EU privacy regulators over Gmail).</td>
<td>Privacy mandate is not embedded throughout the company. Techniques and technologies frequently rolled out without adequate public consultation (e.g. Street view level view).</td>
<td>Customers have a right to amend personal details held by Google but does not allow search history to be removed. Most services do not permit user access to specific or aggregated disclosure or tracking data.</td>
<td>Opt-out possible for some services. Some services may not work well without cookies. May access essential resources without account but when account is created it is sticky.</td>
<td>Good responsiveness. In 2005 a number of profiles were downloaded to prove weak security. Does not accept liability for security.</td>
<td>Substantial Threat</td>
</tr>
<tr>
<td>hi5</td>
<td>General Counsel, hi5 Networks, Inc., 455 Market St., Suite 510, San Francisco, CA 94105,USA.</td>
<td></td>
<td>Collects gender, date of birth, and ZIP. Tracks users with cookies and by IP addresses. Also tracks users movements on site by monitoring clicks-through data.</td>
<td>No information readily available</td>
<td>Relatively blatant about some processing but unnecessarily vague about others.</td>
<td>Poor. Clicking on Privacy Policy opens up a pop-up window advertisement!</td>
<td>User can identify what information is available to members vs. non-members. Can view other users' profiles without their knowledge. Can opt-out of receiving same information. A may delete account.</td>
<td>All visitors can see public content on site (do not need to be registered).</td>
<td>No information readily available</td>
<td>Substantial Threat</td>
<td>Preposterous use of advertising technique (pop-up window) when clicking on privacy policy. Point of contact being a General Counsel leaves little confidence in responsiveness.</td>
</tr>
<tr>
<td>Company</td>
<td>Data Collection and Processing</td>
<td>Data Retention</td>
<td>Openness and Transparency</td>
<td>Responsiveness</td>
<td>Ethical Compass</td>
<td>Customer Control</td>
<td>Fair Gateways</td>
<td>Privacy Enhancing or Invading Innovations</td>
<td>Initial Assessment</td>
<td>Justification</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------</td>
<td>---------------</td>
<td>---------------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>---------------</td>
<td>----------------------------------------</td>
<td>----------------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td>Last.fm</td>
<td>Email address is not required to register. Pseudonymous listening habit data will be available to other users. May sell or licence lists, but not personal data. No personal information collected regarding transactions with third parties. Monitors which songs listened to, whether skipped, etc., recommendations to other users. Does not process PII relating to record collection. Does not collect ZIP post code, city or country unless user explicitly shares. Regards IP addresses as anonymous.</td>
<td>No information readily available</td>
<td>Thorough privacy policy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LinkedIn</td>
<td>Claim that email addresses of friends that user includes are only used for inviting those friends, and sending reminders. Use cookies and web beacons. Permits third-party cookies and beacons. Shares information with other companies “for specific services”.</td>
<td>May close account and then data may be deleted (but not necessarily).</td>
<td>Privacy policy outlines some situations where information is used but could be more explicit.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Last.fm</td>
<td>Email address is not required to register. Pseudonymous listening habit data will be available to other users. May sell or licence lists, but not personal data. No personal information collected regarding transactions with third parties. Monitors which songs listened to, whether skipped, etc., recommendations to other users. Does not process PII relating to record collection. Does not collect ZIP post code, city or country unless user explicitly shares. Regards IP addresses as anonymous.</td>
<td>No information readily available</td>
<td>Thorough privacy policy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LinkedIn</td>
<td>Claim that email addresses of friends that user includes are only used for inviting those friends, and sending reminders. Use cookies and web beacons. Permits third-party cookies and beacons. Shares information with other companies “for specific services”.</td>
<td>May close account and then data may be deleted (but not necessarily).</td>
<td>Privacy policy outlines some situations where information is used but could be more explicit.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Last.fm</td>
<td>Email address is not required to register. Pseudonymous listening habit data will be available to other users. May sell or licence lists, but not personal data. No personal information collected regarding transactions with third parties. Monitors which songs listened to, whether skipped, etc., recommendations to other users. Does not process PII relating to record collection. Does not collect ZIP post code, city or country unless user explicitly shares. Regards IP addresses as anonymous.</td>
<td>No information readily available</td>
<td>Thorough privacy policy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LinkedIn</td>
<td>Claim that email addresses of friends that user includes are only used for inviting those friends, and sending reminders. Use cookies and web beacons. Permits third-party cookies and beacons. Shares information with other companies “for specific services”.</td>
<td>May close account and then data may be deleted (but not necessarily).</td>
<td>Privacy policy outlines some situations where information is used but could be more explicit.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Last.fm</td>
<td>Email address is not required to register. Pseudonymous listening habit data will be available to other users. May sell or licence lists, but not personal data. No personal information collected regarding transactions with third parties. Monitors which songs listened to, whether skipped, etc., recommendations to other users. Does not process PII relating to record collection. Does not collect ZIP post code, city or country unless user explicitly shares. Regards IP addresses as anonymous.</td>
<td>No information readily available</td>
<td>Thorough privacy policy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LinkedIn</td>
<td>Claim that email addresses of friends that user includes are only used for inviting those friends, and sending reminders. Use cookies and web beacons. Permits third-party cookies and beacons. Shares information with other companies “for specific services”.</td>
<td>May close account and then data may be deleted (but not necessarily).</td>
<td>Privacy policy outlines some situations where information is used but could be more explicit.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Last.fm</td>
<td>Email address is not required to register. Pseudonymous listening habit data will be available to other users. May sell or licence lists, but not personal data. No personal information collected regarding transactions with third parties. Monitors which songs listened to, whether skipped, etc., recommendations to other users. Does not process PII relating to record collection. Does not collect ZIP post code, city or country unless user explicitly shares. Regards IP addresses as anonymous.</td>
<td>No information readily available</td>
<td>Thorough privacy policy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LinkedIn</td>
<td>Claim that email addresses of friends that user includes are only used for inviting those friends, and sending reminders. Use cookies and web beacons. Permits third-party cookies and beacons. Shares information with other companies “for specific services”.</td>
<td>May close account and then data may be deleted (but not necessarily).</td>
<td>Privacy policy outlines some situations where information is used but could be more explicit.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Last.fm</td>
<td>Email address is not required to register. Pseudonymous listening habit data will be available to other users. May sell or licence lists, but not personal data. No personal information collected regarding transactions with third parties. Monitors which songs listened to, whether skipped, etc., recommendations to other users. Does not process PII relating to record collection. Does not collect ZIP post code, city or country unless user explicitly shares. Regards IP addresses as anonymous.</td>
<td>No information readily available</td>
<td>Thorough privacy policy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LinkedIn</td>
<td>Claim that email addresses of friends that user includes are only used for inviting those friends, and sending reminders. Use cookies and web beacons. Permits third-party cookies and beacons. Shares information with other companies “for specific services”.</td>
<td>May close account and then data may be deleted (but not necessarily).</td>
<td>Privacy policy outlines some situations where information is used but could be more explicit.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Last.fm</td>
<td>Email address is not required to register. Pseudonymous listening habit data will be available to other users. May sell or licence lists, but not personal data. No personal information collected regarding transactions with third parties. Monitors which songs listened to, whether skipped, etc., recommendations to other users. Does not process PII relating to record collection. Does not collect ZIP post code, city or country unless user explicitly shares. Regards IP addresses as anonymous.</td>
<td>No information readily available</td>
<td>Thorough privacy policy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LinkedIn</td>
<td>Claim that email addresses of friends that user includes are only used for inviting those friends, and sending reminders. Use cookies and web beacons. Permits third-party cookies and beacons. Shares information with other companies “for specific services”.</td>
<td>May close account and then data may be deleted (but not necessarily).</td>
<td>Privacy policy outlines some situations where information is used but could be more explicit.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Last.fm</td>
<td>Email address is not required to register. Pseudonymous listening habit data will be available to other users. May sell or licence lists, but not personal data. No personal information collected regarding transactions with third parties. Monitors which songs listened to, whether skipped, etc., recommendations to other users. Does not process PII relating to record collection. Does not collect ZIP post code, city or country unless user explicitly shares. Regards IP addresses as anonymous.</td>
<td>No information readily available</td>
<td>Thorough privacy policy.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LinkedIn</td>
<td>Claim that email addresses of friends that user includes are only used for inviting those friends, and sending reminders. Use cookies and web beacons. Permits third-party cookies and beacons. Shares information with other companies “for specific services”.</td>
<td>May close account and then data may be deleted (but not necessarily).</td>
<td>Privacy policy outlines some situations where information is used but could be more explicit.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Company</td>
<td>Company administrative details</td>
<td>Corporate Leadership</td>
<td>Data Collection and Processing</td>
<td>Data Retention</td>
<td>Openness and Transparency</td>
<td>Responsiveness</td>
<td>Ethical Compass</td>
<td>Customer Control</td>
<td>Fair Gateways</td>
<td>Privacy Enhancing or Invading Innovations</td>
<td>Initial Assessment</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------------------------</td>
<td>-----------------------</td>
<td>-------------------------------</td>
<td>----------------</td>
<td>--------------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>--------------</td>
<td>------------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Myspace</td>
<td>8391 Beverly Blvd, #349, Los Angeles, CA 90048, <a href="mailto:privacy@myspace.com">privacy@myspace.com</a></td>
<td>Explicit in collecting name, email address, and age; other profile data including but not limited to: personal interests, gender, age, education and occupation. Considers IP addresses as non-identifying information, to track usage, and to share with third parties. Data is recorded for security and monitoring purposes. May opt-out of receiving service information. Email addresses are kept, particularly for invitations, though recipients of invitations can contact myspace to have email address removed. Allow cookies and third party cookies.</td>
<td>No information readily available</td>
<td>Public profiles are no longer mandatory.</td>
<td>Tried to require subpoena before handing over information to law enforcement authorities (on suspected sex offenders).</td>
<td>Users may block the receiving of myspace invitations by emailing myspace with a subject block.</td>
<td>Email addresses and user names are limited in their disclosure.</td>
<td>No information readily available</td>
<td>Notable lapses</td>
<td>A mixed bag, with some strong protections and a lot of ambiguities. Problematic interpretation of IP addressing data. Invitation recipients can opt-out. Account deletion is unclear.</td>
<td></td>
</tr>
<tr>
<td>Orkut</td>
<td>Privacy Matters, c/o Google Inc., 1600 Amphitheatre Parkway, Mountain View CA 94043 (USA)</td>
<td>Must have a Google Account, including email address. Possible profile information: gender, age, occupation, hobbies, and interests, plus other content, such as photos. Can delete account, completed within 48 hours. Retain contents of messages for indeterminate amount of time.</td>
<td>Very limited privacy policy.</td>
<td>Ethical challenges in blocking site from access in Iran.</td>
<td>Inviters can choose to not receive invites.</td>
<td>Must have a Google Account.</td>
<td>No information readily available</td>
<td>Serious lapses</td>
<td>No Orkut-specific privacy contact information. Limited privacy policy. Account deletion good sign. Checkered history in cooperating with governments. Requires registration to view information, but registration applies across Google services.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reunion.com</td>
<td>Reunion.com, Inc., Attn: Privacy Policy Officer, 2118 Wilshire Blvd, Box 1008, Santa Monica, CA 90403-5764</td>
<td>Collects at a minimum, name, birth date, gender, email address and zip code Uses real names. Company will contact users. May &quot;engage third parties to perform analysis or data processing of our databases that involves access to this information in order to better provide you with the services for which you joined&quot; Shares information with other sites. Tracks movements on site and with partner sites.</td>
<td>No information readily available</td>
<td>Changes to policy are announced but if user continues to use site they have consented to the changes. May transfer information if firm is purchased.</td>
<td>Pos. Admonished by businesses community for misleading advertising practices to bring in new registrants.</td>
<td>Not accepting cookies will limit access.</td>
<td>Data protect email privacy through a relay system. Live &quot;technical, administrative and physical safeguards&quot; to protect security of personal information.</td>
<td>Substantial threat</td>
<td>Promising for use of email relaying. Data sharing is dangerously vague. Tracking usage is problematic. Historical ethics problems.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Company</td>
<td>Company administrative details</td>
<td>Corporate Leadership</td>
<td>Data Collection and Processing</td>
<td>Data Retention</td>
<td>Openness and Transparency</td>
<td>Responsiveness</td>
<td>Ethical Compass</td>
<td>Customer Control</td>
<td>Fair Gateways</td>
<td>Privacy Enhancing or Enabling Innovations</td>
<td>Initial Assessment</td>
</tr>
<tr>
<td>--------------------</td>
<td>--------------------------------</td>
<td>-----------------------</td>
<td>-------------------------------</td>
<td>----------------</td>
<td>---------------------------</td>
<td>---------------</td>
<td>----------------</td>
<td>-----------------</td>
<td>--------------</td>
<td>------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Skype</td>
<td>15, rue Notre Dame, L-2240 Luxembourg and/or Skype Communications S.A though no explicit address given for privacy concerns.</td>
<td>Registration not required. Invitation email addresses are deleted immediately upon sending invitation. No communications from skype other than messages about faults. Shared with third parties for provision of services. Cookies do not contain identifying information. Third-party cookies exist.</td>
<td>Skype. At least deals with the issue in part in the privacy policy without committing in detail. Though for traffic data, commits to “erase Traffic Data, or make Traffic Data anonymous, so soon as it is no longer needed for the purpose of the communication or billing purposes, unless applicable law permit otherwise.”</td>
<td>No way to know if there are back doors in the software. Right to review data, correct, and delete personal data, via email <a href="mailto:delete@skype.com">delete@skype.com</a>. Thorough privacy policy, but no contact information for accountability.</td>
<td>Responded to concerns about DRM and reading motherboard information.</td>
<td>Poor. Co-operated with Chinese government.</td>
<td>Do not need to register to use Skype Software, but registration may be needed for particular services. Blocking cookies may exhibit personalized services.</td>
<td>User profile data not stored centrally on server. Takes “appropriate organizational and technical measures”; authorised employees only. Will take &quot;appropriate technical measures to protect the confidentiality of the communications Content via its Skype Software and VoIP Services”.</td>
<td>Generally privacy aware.</td>
<td>Notice issues.</td>
<td>Good promises on deleting invasion email addresses. Lack of contact details is problematic. Lack of openness about software capabilities is problematic. Deletion of traffic data is good statement though less ambiguity about role of laws would help.</td>
</tr>
<tr>
<td>Wikipedia</td>
<td>No explicit contact, but policy says it was approved by Board.</td>
<td>Can operate under pseudonym, but if not, then logs IP addresses for public view. recommends using pseudonym. IP addresses are stored and can be seen by server administrators and advanced users. Data is combined to investigate abuse.</td>
<td>Raw logs are normally discarded after two weeks. Unable to remove accounts. Deleted ‘content’ is not in fact deleted.</td>
<td>Clear privacy policy, but no main point of contact.</td>
<td></td>
<td></td>
<td></td>
<td>Fully accessible without authenticating.</td>
<td>Session cookies only; and temporary log-in cookies that expire every 30 days.</td>
<td>Generally privacy aware.</td>
<td>Lacking in some information, such as contact details. Good statement on retention policy, though unless there is a contact, this is unverifiable.</td>
</tr>
<tr>
<td>Windows Live Space</td>
<td>Microsoft Privacy, Microsoft Corporation, One Microsoft Way, Redmond, Washington 98052 - 425-882-8180, and webform is available.</td>
<td>Signed up to Trust-e and Safe Harbor.</td>
<td>IP addresses not treated as personal information. Customised links are used to identify users. Once messenger service requires signing up with Verizon. Tracks all requests for maps. Locations are logged when service is used online.</td>
<td>No information readily available.</td>
<td>Unclear about what information is used for and how long it is used for.</td>
<td>Prior. Co-operated with Chinese government.</td>
<td>Unclear policy statement about future co-operation. Recent research hints at profiling based on search requests. Shuffled search data to U.S. Department of Justice for research purposes.</td>
<td>User can designate who has access to which calendar data.</td>
<td>Anyone may review calendar information that is published for public access.</td>
<td>May use beacons to track messages sent by MS to determine whether opened or read. Beacons also used by third parties to aggregate statistics.</td>
<td>Substantial threat.</td>
</tr>
<tr>
<td>Xanga</td>
<td>Contactable through webform for email interaction.</td>
<td>Username, password, email address, date of birth. Email and birthdate are not necessarily disclosed if user wishes. Profile information is optional. For invitations, Xange may send multiple invitations by email. Email addresses can be blacklisted to receive no further invitations. Logs IP data. Targets advertisements based on profile and past activities. Third-party cookies are possible as well. May transfer data if company is purchased.</td>
<td>If account is shut down, Xanga site no longer accessible. Data may be archived, but offline.</td>
<td>Presumes consent by non-U.S. users.</td>
<td>By default information is shared widely, though can be controlled. Can control comments on your section of the site, and whether someone can be blocked from commenting.</td>
<td>Information available to non-registered users. Blocking cookies may limit access.</td>
<td>Postprints’ service allows users to watch visitors on its or her own site (username or geographic information based on IP address).</td>
<td>Generally privacy aware.</td>
<td></td>
<td>Serious lapses.</td>
<td>Invitation process could be better managed. Treatment of IP data is vague. Profiling is mentioned but more clarity is required. Information should not be shared by default. May limit information collected.</td>
</tr>
<tr>
<td>Company</td>
<td>Company administrative details</td>
<td>Corporate Leadership</td>
<td>Data Collection and Processing</td>
<td>Data Retention</td>
<td>Openness and Transparency</td>
<td>Responsiveness</td>
<td>Ethical Compass</td>
<td>Customer Control</td>
<td>Fair Gateways</td>
<td>Privacy Enhancing or Invading Innovations</td>
<td>Initial Assessment</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------</td>
<td>-----------------------</td>
<td>-------------------------------</td>
<td>----------------</td>
<td>----------------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>-----------------</td>
<td>--------------</td>
<td>----------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Yahoo!</td>
<td>Yahoo! Inc. Customer Care - Privacy Policy issues, 701 First Avenue, Sunnyvale, CA 94089, (408) 349-5070</td>
<td>Trust-e and safe harbor</td>
<td>Information collection process can be combined with data from other sources (business partners and other companies). Information collected: name, email, birthdate, gender, ZIP code, occupation, industry, personal interests. May also ask for social security for financial services. Collects transaction data, including information about use of financial products. Collects and stores information including IP addresses and cookies related data. Data can be shared for marketing purposes. Data will be transferred if acquired. Cookies (and third party cookies) are used, as are web beacons. Opt-out of marketing may delete account but some information retained, for 90 days. Log files are used — after they are used they are stored (but said to be inaccessible). No further information on searches.</td>
<td>Data can be shared for marketing purposes. Data will be transferred if acquired. Cookies (and third party cookies) are used, as are web beacons. Opt-out of marketing may delete account but some information retained, for 90 days. Log files are used — after they are used they are stored (but said to be inaccessible). No further information on searches.</td>
<td>May not go out of its way to respond to ethical concerns.</td>
<td>Poor. Cooperates with governments with disclosure of information, including Chinese government. Disclosed search data to U.S. Department of Justice for research purposes.</td>
<td>Overly broad and vague policy.</td>
<td>Did not go out of its way to respond to ethical concerns.</td>
<td>Poor. Cooperates with governments with disclosure of information, including Chinese government. Disclosed search data to U.S. Department of Justice for research purposes.</td>
<td>Use 'physical, electronic, and procedural safeguards that comply with federal regulations to protect personal information' Also limit access to employees.</td>
<td>Registration not necessary for some services.</td>
</tr>
<tr>
<td>YouTube</td>
<td>Contact only available through a contact form.</td>
<td>Media files, once uploaded, can not be modified. No information on deletion of other data.</td>
<td>Use both session and persistent cookies, as well as web beacons. Monitors and tracks IP logs. IP data not considered personal data. Data used to monitor marketing effectiveness and track actions (e.g. cookies). Share personal information with subsidiaries, affiliated companies, or other businesses and persons.</td>
<td>Use of site is considered consent to U.S. law (no safe harbor). Data can be purchased in event of sale.</td>
<td>Has a policy for data breaches.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Blocked cookies may inhibit service.</td>
<td>Web beacons used to track usage, and uses gifs in emails to track usage. &quot;(U)ses commercially reasonable physical, managerial, and technical safeguards to preserve the integrity and security of your personal information&quot;</td>
</tr>
</tbody>
</table>
Ranking Countries for Their Privacy Protection and Surveillance Levels
Leading surveillance societies in the EU and the World 2006

02/11/2006

Graphic from the Daily Telegraph, November 2, 2006.

Background information on the table (PDF)
What Can We Do About It?
Privacy Enhancing Technologies

Educating Users to Privacy Threats
Anti-Phishing Phil

- [http://cups.cs.cmu.edu/antiphishing_phil/](http://cups.cs.cmu.edu/antiphishing_phil/)

An interactive game that teaches users how to identify phishing URLs, where to look for cues in web browsers, and how to use search engines to find legitimate sites.
Anti-Phishing Phil

How to Avoid Online Scams

Don’t ignore the URL!

Looking at the address bar can help you figure out if a website is legitimate or a scam!
Privacy Enhancing Technologies

Knowing Privacy Policies
Platform for Privacy Preferences (P3P)

• Allows websites to express their privacy practices in a machine as well human readable way

• Can be retrieved automatically by P3P enabled web browsers and interpreted
  – Users can be made aware of privacy practices
  – Enables automated decision making based on these practices

http://www.w3.org/P3P/
PrivacyFinder – Privacy Enhanced Search Engine

- [http://www.privacyfinder.org](http://www.privacyfinder.org)

**PrivacyFinder** is a privacy-enhanced search engine. Once you state your privacy preferences (low, medium, high, or custom), the search results are ordered based on how their computer-readable privacy policies comply with your preferences. A red bird indicates that the site has conflicts with your preferences while a green bird indicates compliance. The absence of any bird means that a valid computer-readable privacy policy, known as a P3P policy, could not be located.
Privacy Enhancing Technologies

Anonymizing Protocols for Communication
Anonymizing Protocols

• Makes it difficult from someone to trace back a message to its source

• Prevents
  – Linkability
  – Traceability

• Examples
  – Anonymizing Proxy
  – Mix Networks and Onion Routing
    • Tarzan, Tor, I2P
  – Anonymous sharing
    • Freenet
Anonymizing Proxy

Traffic Analysis Breaks Anonymity
Layered encryption: \( E \left[ E \left[ E \left[ M, K_{pub}^1 \right], K_{pub}^2 \right], K_{pub}^3 \right], K_{pub}^4 \right] \)

Prevent edge analysis by introducing cover traffic
Tor – Working Mix Network

- From the Tor project website ([http://www.torproject.org](http://www.torproject.org))

Tor is a toolset for a wide range of organizations and people that want to improve their safety and security on the Internet. Using Tor can help you anonymize web browsing and publishing, instant messaging, IRC, SSH, and other applications that use the TCP protocol. Tor also provides a platform on which software developers can build new applications with built-in anonymity, safety, and privacy features.
Browsing with Tor

Privacy.net Browser Privacy Results with Tor
The Privacy.net Analyzer

This site analyzes the privacy of your Internet connection and shows some of the information web sites can know about you when you visit. The information can be used to display web content based on things such as country of origin and web browser.

Click here to read a description of the tests (opens new window)

You Are Visiting From:
192.251.226.206 is from Germany (DE) in Western Europe
No host name is associated with this IP address or no reverse lookup is configured.
Error: Timed out

Cookie Test
No Cookie from this site is on your system from prior visits.

Your Browser Type and Operating System:
Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.7) Gecko/2009021910 Firefox/3.0.7

All information sent by your web browser when requesting this web page:
Connection: keep-alive Content-Length: 414 Content-Type: application/x-www-form-urlencoded
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8 Accept-Charset: UTF-8,*
Accept-Encoding: gzip, deflate Cookie: bhCookieSaveSess=1; bhCookieSess=1; Privacy.net_Last_Visit=1/18/2010; Privacy.net=Privacy+Analysis; bhawkplt=plt_state=tested&plt_stm=126381585323&plt_url=null Host: analyze.privacy.net Referer: http://analyze.privacy.net/test.asp?RequestCookies=&Requestdate=&referer= User-Agent: Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.0.7) Gecko/2009021910 Firefox/3.0.7

Firewall Test
The following ports were checked: 16771, 80
Out of the above ports, the following are open and permitting outbound traffic: 16771, 80
Firewall status: PRESENT

Browser Type and Version
Browser: Firefox
Fullversion: 3.0.7
Gecko: True
GeckoBuildDate: 2009021910
Crawler: False

Browser Security
Session Cookies Not Accepted

Display and Layout
Width: 1100
WidthAvail: 1100
Height: 700
StyleSheets: True
PNG: True
FontSmoothing: False
FontColor: True
FontSize: True

System Details
Platform: WinXP
Win16: False
WinInstallerMinVer: 2

Plug-in Information
Plugin Flip4Mac installed
### Persistent Cookies Accepted
- JavaScriptEnabled: True
- VBScriptEnabled: False
- JavaEnabled: False
- ActiveXEnabled: False
- SSL: True
- SSLKeySize: 128
- SSLEnabled: True
- Firewall: True
- OpenPorts: 16771,80
- PopupsBlocked: True
- ImagesEnabled: True
- HighSecurity: False

### JavaScript/ActiveX Information
- ActiveXControls: False
- ActiveXEnabled: False
- JavaScript: True
- JavaScriptEnabled: True
- JavaScriptVer: 1.8
- JavaScriptBuild:
- VBScript: False
- VBScriptEnabled: False
- VBScriptBuild:

### Java Information
- JavaApplets: True
- JavaEnabled: False

### Wireless Device Information
- PDA: False
- WAP: False
- HDML: False

### Locale Information
- Country: DE
- Language: English
- User Language: en-us
- System Language:
- Time Zone Difference: 5

### Connection Details
- Broadband: False
- ConnectionType:
- Firewall: True
- Proxy: False
- CompressGZip: True
- AOL: False
- MSN: False

### Scripting Capabilities
- ActiveXControls: False
- ActiveXEnabled: False
- JavaScript: True
- JavaScriptEnabled: True
- JavaScriptVer: 1.8
- JavaScriptBuild:
- VBScript: False
- VBScriptEnabled: False
- VBScriptBuild:
- XML: True
- MSXML: 0
- XMLHttpRequest: True
- DHTML: True
- FileUpload: Yes

### Java Information
- JavaApplets: True
- JavaEnabled: False

### Tables/Colors/Frame Information
- Tables: True
- TableBGColor: True
- TableBGImage: True
- ColorDepth: 24
- Frames: True
- IFrames: True

### TraceRoute to 192.251.226.206

<table>
<thead>
<tr>
<th>Hop</th>
<th>(ms)</th>
<th>(ms)</th>
<th>(ms)</th>
<th>IP Address</th>
<th>Host name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>192.168.255.1</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>67.222.132.1</td>
<td>router.dfw-datacenter.com</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>14</td>
<td>6</td>
<td>72.249.128.105</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>9</td>
<td>11</td>
<td>17</td>
<td>8.9.232.73</td>
<td>xe-5-3-0.edge3.dallas1.level3.net</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>12</td>
<td>17</td>
<td>4.69.145.179</td>
<td>ae-82-80.ebr2.dallas1.level3.net</td>
</tr>
<tr>
<td>6</td>
<td>54</td>
<td>54</td>
<td>56</td>
<td>4.69.137.122</td>
<td>ae-3.ebr4.newyork1.level3.net</td>
</tr>
<tr>
<td>7</td>
<td>73</td>
<td>56</td>
<td>65</td>
<td>4.69.134.114</td>
<td>ae-64-64.csw1.newyork1.level3.net</td>
</tr>
<tr>
<td>8</td>
<td>47</td>
<td>54</td>
<td>43</td>
<td>4.69.134.65</td>
<td>ae-61-61.ebr1.newyork1.level3.net</td>
</tr>
<tr>
<td>9</td>
<td>121</td>
<td>115</td>
<td>135</td>
<td>4.69.137.77</td>
<td>ae-44-44.ebr2.london1.level3.net</td>
</tr>
<tr>
<td>10</td>
<td>132</td>
<td>121</td>
<td>112</td>
<td>4.69.139.106</td>
<td>ae-2-52.edge4.london1.level3.net</td>
</tr>
<tr>
<td>11</td>
<td>133</td>
<td>113</td>
<td>148</td>
<td>195.50.122.2</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>168</td>
<td>129</td>
<td>129</td>
<td>84.16.14.166</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>143</td>
<td>136</td>
<td>140</td>
<td>84.16.9.102</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>144</td>
<td>148</td>
<td>145</td>
<td>195.71.158.21</td>
<td>rmwc-fmk-de01-so-2-1-0-0.nw.mediaways.net</td>
</tr>
<tr>
<td>15</td>
<td>143</td>
<td>151</td>
<td>139</td>
<td>195.71.254.78</td>
<td>rmwc-gtso-de01-ge-0-0-0-0.nw.mediaways.net</td>
</tr>
<tr>
<td>16</td>
<td>149</td>
<td>147</td>
<td>143</td>
<td>195.71.12.59</td>
<td>xmws-gtso-de01-vlan-2.nw.mediaways.net</td>
</tr>
<tr>
<td>17</td>
<td>Timed out</td>
<td>Timed out</td>
<td>Timed out</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Timed out</td>
<td>Timed out</td>
<td>Timed out</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Timed out</td>
<td>Timed out</td>
<td>Timed out</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>Timed out</td>
<td>Timed out</td>
<td>Timed out</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Trace aborted.
Checking who manages your IP address:

whois query for 192.251.226.206...

Results returned from whois.arin.net:

OrgName:    RIPE Network Coordination Centre
OrgID:      RIPE
Address:    P.O. Box 10096
City:       Amsterdam
StateProv:  
PostalCode: 1001EB
Country:    NL

ReferralServer: whois://whois.ripe.net:43

CIDR:       192.251.226.0/24
NetName:    RIPE-ERX-192-251-226-0
NetHandle:  NET-192-251-226-0-1
Parent:     NET-192-0-0-0-0
NetType:    Early Registrations, Transferred to RIPE NCC
Comment:    These addresses have been further assigned to users in the RIPE NCC region. Contact information can be found in the RIPE database at http://www.ripe.net/whois
RegDate:    2005-02-28
Updated:    2005-02-28

# ARIN WHOIS database, last updated 2010-01-17 20:00
# Enter ? for additional hints on searching ARIN's WHOIS database.
# ARIN WHOIS data and services are subject to the Terms of Use
# available at https://www.arin.net/whois_tou.html

Results returned from whois.ripe.net:

% This is the RIPE Database query service.
% The objects are in RPSL format.
% The RIPE Database is subject to Terms and Conditions.
% See http://www.ripe.net/db/support/db-terms-conditions.pdf

% Information related to '192.251.226.200 - 192.251.226.207'

inetnum: 192.251.226.200 - 192.251.226.207
netname:  BLUTMAGIE
descr:   Olaf Selke
remarks: Inquiries/Abuse: abuse@blutmagie.de
remarks: Inquiries/Abuse: for contact details see OS835-RIPE
country: DE
admin-c: OS835-RIPE
tech-c:   OS835-RIPE
status:  ASSIGNED PA
mnt-by:  MNT-WUSEL
changed: wusel@uu.org 20080129
source:  RIPE

person:  Olaf Selke
address:        Detmolder Strasse 109
address:        33397 Rietberg
phone:          +495246801169
fax-no:         +495246802169
nic-hdl:        OS835-RIPE
changed:        wusel@uu.org 20080123
mnt-by:         MNT-WUSEL
source:         RIPE

% Information related to '192.251.226.0/24AS6805'

route:        192.251.226.0/24
descr:        Kai Siering
origin:       AS6805
mnt-by:       MDA-Z
changed:      ingo.stampe@mediaways.net 20000407
source:       RIPE

The following is a list of all fonts installed on your computer:

Font detection requires IE 5 or higher.

Click to Run The Test Again (Don't use refresh)
Trace other computers on the Internet using Network-Tools.com
Back to Privacy.net

<table>
<thead>
<tr>
<th>Web Spider Software</th>
<th>Stop Bad Bots</th>
<th>GoogleSearchAppliance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extract web content and metadata from websites into your database</td>
<td>Spot your users from the web robots Identify bad crawlers, scrapers</td>
<td>New Kayxo Connector for GSA ` and Microsoft Exchange Server</td>
</tr>
<tr>
<td><a href="http://www.newprosoft.com">www.newprosoft.com</a></td>
<td><a href="http://www.atlbi.com/webcrawlers.html">www.atlbi.com/webcrawlers.html</a></td>
<td>kayxo.com/KayxoConnector</td>
</tr>
</tbody>
</table>
Anonymous Communication

- Not easy to use and administer
- Most rely on a majority of entities being trusted
- Susceptible to collusion among some subsets of entities
- Susceptible to some types of traffic analysis
- Scalability
- Ease of Adaptation
Privacy Enhancing Technology

Personal Data Servers

@ SMIS.INRIA
Paper Based Personal Data

- **Yesterday**
  - Many paper based personal data
    - Invoices, bank, salary form, …
    - Diplomas, official forms, …
    - Health folder, insurance, …

- **Today**
  - More and more electronic versions
    - e-invoice, e-health, e-government, …
  - And new forms of data
    - Bookmarks, navigation history, location tracking …

- ➔ Need to store those data while ensuring important properties !!!
  - Availability: 24/24, 7/7, from everywhere
  - Durability
  - Query facilities
  - Security (confidentiality, controlled sharing)
Baseline = Secure Portable Token

- Embed the traditional chain of software in secure hardware
  - Web server – Application – DBMS – Database

Slide courtesy of the SMIS group @ INRIA
Personal Data Servers (PDS)

• Store data under strong trusted hardware-guaranteed protection on SPTs
• Do not share unless absolutely needed
• Share only what is needed to be shared
• User can specify personal preferences for
  – Data sharing
  – Date usage
  – Date retention
• Ability to audit access to personal data
Data Dissemination can be Useful Too

- Personally identifiable information collected whenever a user
  - creates an account
  - submits an application
  - signs up for newsletter
  - participates in a survey
  - ...

- Data sharing and dissemination is often useful and may be done
  - to study trends or to make useful statistical inference
  - to share knowledge
  - to outsource the management of data
  - ...
Privacy Enhancing Technology

Microdata Disclosure Control
The Anonymity Problem

<table>
<thead>
<tr>
<th>SSN</th>
<th>Name</th>
<th>DOB</th>
<th>Sex</th>
<th>ZIP</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>123456789</td>
<td>W. Carter</td>
<td>04/12/67</td>
<td>F</td>
<td>80542</td>
<td>hypertension</td>
</tr>
<tr>
<td>234567891</td>
<td>P. Jack</td>
<td>04/21/67</td>
<td>F</td>
<td>80541</td>
<td>obesity</td>
</tr>
<tr>
<td>345678912</td>
<td>R. Robert</td>
<td>09/27/63</td>
<td>M</td>
<td>80535</td>
<td>chest pain</td>
</tr>
<tr>
<td>456789123</td>
<td>D. Young</td>
<td>03/13/63</td>
<td>M</td>
<td>80539</td>
<td>obesity</td>
</tr>
<tr>
<td>567891234</td>
<td>J. Lin</td>
<td>09/27/63</td>
<td>M</td>
<td>80539</td>
<td>hypertension</td>
</tr>
<tr>
<td>678912345</td>
<td>K. Martin</td>
<td>03/18/64</td>
<td>F</td>
<td>80538</td>
<td>short breath</td>
</tr>
<tr>
<td>789123456</td>
<td>F. Clark</td>
<td>10/22/64</td>
<td>F</td>
<td>80539</td>
<td>short breath</td>
</tr>
<tr>
<td>891234567</td>
<td>J. Young</td>
<td>04/21/69</td>
<td>F</td>
<td>80549</td>
<td>chest pain</td>
</tr>
<tr>
<td>912345678</td>
<td>R. Smith</td>
<td>04/22/69</td>
<td>F</td>
<td>80541</td>
<td>short breath</td>
</tr>
</tbody>
</table>

Privacy breach: J. Young who lives at 600 Welker Av., Mead, CO suffers from chest pain.
Preserving Privacy: k-Anonymity

- The released data should be indistinguishably related to no less than a certain number, $k$, of respondents
- The respondents must be indistinguishable with respect to a set of attributes (quasi-identifiers)
- k-Anonymity requires that every combination of values of quasi-identifiers in the released table must have at least $k$ occurrences
  - Enforced using generalization and suppression
Generalization and Suppression

- **Generalization**: the values of a given attribute are replaced by more general values
  - ZIP codes 80521 and 80523 can be generalized to 8052*
  - Date of birth 12/04/64 and 12/10/64 can be generalized to 64 or 12/64

- **Suppression**: remove the information altogether
  - Suppression reduces the amount of generalization necessary to satisfy the k-anonymity requirement
Domain Generalization Hierarchy

\[ H^0_{\text{ZIP}} \rightarrow H^1_{\text{ZIP}} \rightarrow H^2_{\text{ZIP}} \]

- \( H^0 \) full specialization
- \( H^1 \) full generalization
- \( H^2 \) generalization level attribute name

Dr. Indrajit Ray, Associate Professor, Computer Science Department
Anonymizing a Tuple

- $H^0_{ZIP}$: {12345} {12346} {12355} {12356}
- $H^1_{ZIP}$: {1234*} {1235*}
- $H^2_{ZIP}$: {123**}

- $H^0_{SEX}$: {M} {F}
- $H^1_{SEX}$: *

- $H^0_{SALARY}$: {< 50K} {≥ 50K}
- $H^1_{SALARY}$: *

(12355, M, <50K)

(1235*, M, *)

Generalize using levels:
- 1 for ZIP
- 0 for SEX
- 1 for SALARY

Dr. Indrajit Ray, Associate Professor, Computer Science Department
Privacy by \(k\)-Anonymity

<table>
<thead>
<tr>
<th>SSN</th>
<th>Name</th>
<th>DOB</th>
<th>Sex</th>
<th>ZIP</th>
<th>Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>123456789</td>
<td>W. Carter</td>
<td>04/12/67</td>
<td>F</td>
<td>80542</td>
<td>hypertension</td>
</tr>
<tr>
<td>234567891</td>
<td>P. Jack</td>
<td>04/21/67</td>
<td>F</td>
<td>80541</td>
<td>obesity</td>
</tr>
<tr>
<td>345678912</td>
<td>R. Robert</td>
<td>09/27/63</td>
<td>M</td>
<td>80535</td>
<td>chest pain</td>
</tr>
<tr>
<td>456789123</td>
<td>D. Young</td>
<td>03/13/63</td>
<td>M</td>
<td>80539</td>
<td>obesity</td>
</tr>
<tr>
<td>567891234</td>
<td>J. Lin</td>
<td>09/27/63</td>
<td>M</td>
<td>80539</td>
<td>hypertension</td>
</tr>
<tr>
<td>678912345</td>
<td>K. Martin</td>
<td>03/18/64</td>
<td>F</td>
<td>80538</td>
<td>short breath</td>
</tr>
<tr>
<td>789123456</td>
<td>F. Clark</td>
<td>10/22/64</td>
<td>F</td>
<td>80539</td>
<td>short breath</td>
</tr>
<tr>
<td>891234567</td>
<td>J. Young</td>
<td>04/21/69</td>
<td>F</td>
<td>80549</td>
<td>short breath</td>
</tr>
<tr>
<td>912345678</td>
<td>R. Smith</td>
<td>04/22/69</td>
<td>F</td>
<td>80541</td>
<td>short breath</td>
</tr>
</tbody>
</table>

**Generalization:** group attribute (quasi-identifier) values into larger domains

\[ k \text{-Anonymity: size of every equivalence class is at least } k \]
Other Models of Microdata Disclosure Control

- Other models are proposed around the privacy issues identified by k-anonymity
  - l-diversity and t-closeness
  - Mondrian multidimensional k-anonymity
  - k-type anonymizations
  - (α,k)-anonymity, p-sensitive k-anonymity, (k,l)-anonymity
  - Anatomy, personalized privacy, skyline privacy

- All existing anonymity models are minimalistic view models
  - privacy of a table is characterized by the minimum privacy level of all individuals
Privacy Research

@

Colorado State University
Major Research Initiatives

- Optimal microdata disclosure with focus on data publisher’s dilemma and biased privacy
- Identifying trustworthy data recipients
- Location privacy models and techniques
- Network trace data anonymization
- Privacy preserving protocols
  - Trust negotiation
  - E-commerce
  - E-voting
Optimal Microdata Disclosure
Too Much Sanitization

• May reduce the quality of the data to such extent that it may not be useful anymore

• What is too much?
  – Need to assess the degree of data disclosure
  – Need to assess the quality of data resulting from disclosure control
Preserving Data Utility

- microdata

- generalization

- choose the one with lowest information loss (given by some metric)

- k-anonymous tables (for a given k)
Problem – Data Publisher’s Dilemma

• A data publisher must weigh in the the risk of publicly disseminated information against the statistical utility of the content
  – How to decide what a good value of \( k \) is?
  – How to assure that higher \( k \) values or lower information loss is not possible in the neighborhood of a chosen value?
Problem – Biased Privacy

- k-anonymity only specifies a minimum privacy level present for all individuals in the microdata
- Individual privacy levels can be very different for different individuals

<table>
<thead>
<tr>
<th>Race</th>
<th>DOB</th>
<th>Sex</th>
<th>ZIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>asian</td>
<td>64</td>
<td>F</td>
<td>805**</td>
</tr>
<tr>
<td>asian</td>
<td>64</td>
<td>F</td>
<td>805**</td>
</tr>
<tr>
<td>asian</td>
<td>64</td>
<td>F</td>
<td>805**</td>
</tr>
<tr>
<td>asian</td>
<td>63</td>
<td>M</td>
<td>805**</td>
</tr>
<tr>
<td>asian</td>
<td>63</td>
<td>M</td>
<td>805**</td>
</tr>
<tr>
<td>black</td>
<td>64</td>
<td>F</td>
<td>805**</td>
</tr>
<tr>
<td>black</td>
<td>64</td>
<td>F</td>
<td>805**</td>
</tr>
<tr>
<td>white</td>
<td>64</td>
<td>F</td>
<td>805**</td>
</tr>
<tr>
<td>white</td>
<td>64</td>
<td>F</td>
<td>805**</td>
</tr>
</tbody>
</table>

2-anonymized table
Current Research Focus

• Multi-objective analysis
  – to resolve the data publisher’s dilemma

• Quantification of anonymization bias
  – bias may be infused to cater to personal privacy requirements

• Fair comparison of anonymization techniques in the presence of bias

• Alternative characterization of privacy
  – privacy from an individualistic viewpoint

• Optimization framework for alternative privacy models
Most Recent Work

• Identify Pareto-optimal generalizations and report to data publisher for trade-off analysis

Data publisher gets optimal k-value – information loss pairs. Can now decide which one to choose
Domain Hierarchy Lattice

- **full generalization**

- **full specialization**

A node is a vector of integers specifying the **generalization levels** of the attributes.

An edge exists between two nodes that differ in exactly one position by one.
Generalization Graph

- **Full generalization**
  - 2 1 1

- **Root**
  - 1 1 1

- **Increasing generalization**
  - 2 0 1

- **Increasing information loss**

- **Non-decreasing k value**

- **Full specialization**
  - 0 0 0

Dr. Indrajit Ray, Associate Professor, Computer Science Department
Specialization Graph

full generalization

root

full specialization

increasing specialization
non-increasing k value
decreasing information loss

Dr. Indrajit Ray, Associate Professor, Computer Science Department
Pareto-Optimal Node in Generalization Graph

- Node \( N \) is Pareto-optimal iff there is no other node \( M \) such that
  - \( k_M \geq k_N \) and \( \text{loss}_M < \text{loss}_N \), or
  - \( k_M > k_N \) and \( \text{loss}_M \leq \text{loss}_N \)

Identify the Pareto-optimal nodes in a domain hierarchy lattice without evaluating (computing \( k \) and \( \text{loss} \)) every node.
POkA

• **Pareto-Optimal k-Anonymization**
  – start from a known Pareto-optimal node
  – prune nodes that cannot be Pareto-optimal
  – move to the *next* Pareto-optimal node
  – repeat from newly found node
Identifying Trustworthy Data Recipients

• Pareto optimal anonymization produces a set of optimal privacy level, information loss pairs
• No guarantees that data recipients are not malicious and hence misuse data
• Evaluate trustworthiness of data recipients to handle data properly
  – Previous experience with data recipient
  – Properties of data recipient
  – Recommendation for data recipient
• Use specific privacy level (from optimal set) suitable for recipient
Conclusions
So Where Are We?

• Privacy survey indicates concern among Internet users
  – Increasingly people say they are concerned about online privacy (80-90% of Net users)
  – 27% of Net users have abandoned online shopping carts due to privacy concerns
  – 64% of Net users decided not to use a web site or make an online purchase due to privacy concerns
  – 34% of Net users who do not buy online would buy online if they didn’t have privacy concerns

• Improved privacy protection is factor most likely to persuade more users and previous non-Net users to go online
So Where Are We?

- Often users are left with little choice but to accede
  - However, user can benefit from conscious decision
  - Other times there are tools to help user (at least partially)
  - User should know what these tools can or cannot do

- More research needed in privacy enhancing technologies
Thank You
Privacy Research @ CSU

Data Anonymization for Network Traces
## Use of Public Trace*

*Source: Jelena Mirkovic, Privacy-Safe Network Trace Sharing via Secure Queries, NDS ‘08.*

<table>
<thead>
<tr>
<th></th>
<th>SIG 06</th>
<th>IMC 06</th>
<th>SIG 07</th>
<th>IMC 07</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total papers</td>
<td>37</td>
<td>34</td>
<td>35</td>
<td>38</td>
</tr>
<tr>
<td>Used a traffic trace</td>
<td>9</td>
<td>15</td>
<td>10</td>
<td>15</td>
</tr>
<tr>
<td>Used a public trace</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Used a private trace</td>
<td>8</td>
<td>14</td>
<td>6</td>
<td>14</td>
</tr>
</tbody>
</table>
Trace Data Requirements

- Pseudonym Consistency: useful for traffic matrix estimation, connection characterization, etc.
- Header Information: required in analyzing the effects of packet loss and reordering in TCP dynamics
- Transport Protocol: useful for studying round trip times, reassembly and fragmentation
- Port Numbers: protocol classification schemes
Sensitive Information

- Infrastructure: topology, capacities, hardware, etc.
- Participation: no. of customers, volume of traffic, web server hits, etc.
- Identity: IP address
- Data: payloads
Fundamental Difference with Microdata Disclosure Control

- A tuple in a microdata relates to a single individual
- Sensitive information are certain attributes that are part of the data itself
- Anonymity models quantify privacy of an individual
- Preliminary attempts to quantify loss

- A tuple in a trace is just a part of a bigger communication sequence
- Sensitive information is also inferential, derivable by combining information in the attributes
- No privacy model yet
- No privacy quantification yet
- No quantification of information loss
Attacks on Sanitized Traces

- Web page attack: identify web pages based on number and length of objects
- Clock skew attack: identify a host by calculating the skew between the packet sender’s clock and some reference clock
- Link layer attack: identify network topology
- Clustering attack: use address clustering to detect subnets of IPs
- Behavior attack: use behavior models of popular, known servers
Data Anonymization Techniques Used for Traces

- Suppression: outright removal
- Fixed Transformation: constant substitution
- Variable Transformation: replace an IP address with different pseudonym values based upon application layer protocol
- Typed Transformation: prefix-preserving address anonymization
Techniques Used (cont’d)

• Share data with only trusted parties - legal binding (PREDICT model)
  – Who to trust and how much too trust
• Do not publish data; instead publish an access portal to the data
  – Access data using a query language
  – Access is restricted by a privacy policy implemented as part of the language interpreter
  – Publisher’s can modify the policy : balance privacy and utility trade-off according to needs
  – Usefulness of the data is limited by the capabilities of the query language
  – Too much freedom in the language can lead to inference attacks in unknown ways
Our Research in Trace Data Anonymization

• Characterize privacy levels
  – A binary approach is not sufficient
  – Introduce ambiguity in attacker inference

• Is there a way to quantify the information content of a trace?
  – Data usefulness classification
  – Levels of anonymization: the more you do, the more you lose
  – But remember, we need a “one for all” anonymization; multiple versions may be dangerous

• Develop policy model for trace data dissemination
  – Integrate trust into policy model
Grand Vision

• A trace anonymization tool
  – Query the user to determine requirements
  – Determine anonymization possibilities
  – Estimate value of anonymized trace data
  – Maximize utility of anonymized trace data to user