Data on Data Breaches: Past, Present and Future

Adam Shostack and Chris Walsh
Emergent Chaos

This presentation represents the official position of the Emergent Chaos blog, not our employers
Welcome to Sevilla
Navigational charts were kept secret during the age of exploration

- Henry the Navigator encouraged exploration
- Wanted the results for competitive advantage
- Columbus ended up in the Caribbean
- Lots of sailors died at sea
- Maps are still secret in some places
  - They don’t like http://maps.google.com
We face navigation hazards, too

We need to:

Know they exist :^)

Know how damaging they can be

Know our weak points if we run into them.

Know how to avoid them.

Case in Point: Security breaches involving personal information

Definitely exist
But how numerous?
How do we know?
Are some more at risk than others?

Can be damaging
But how much so, and to whom?
How do we know?

Weak points driven by economics, not physics

Avoidance techniques must be strategic
Security Breaches: How numerous?

Below the waterline:
1. Undetected incidents
2. Unreported incidents
3. Reported, but unanalyzed
4. Reported, but privileged

Focus here is on 2, 3, and a little bit of 4.
How Do We Know?

Individual reports: News stories, press releases

Collections of same

- For general use - Emergent Chaos breaches category, Attrition.org’s DLDOS, etc.

- Google Alerts are the researcher’s friend

- For specific purposes - data behind a journal article

- Often use commercial news archives such as LexisNexis

Reports are much more numerous now that states have notification laws
Attrition’s DLDOS

http://attrition.org/dataloss/dldos.html

• Provides “date, the company that reported the breach, the type of data impacted, the number of records impacted, third party companies involved, and a few other sortable items”

• 700 records as of June 13, 2007.

• A main data supplier to other well-known sources, academic works, etc.
Attrition.org Incident Archive

Incident Count (Attrition DLDOS)

Breach Sizes
Etiolated.org
The Choicepoint incident certainly spurred legislative action.
U.S. State Breach Notification Laws

It is hard to measure the information security impact of these laws, in part because we only have two years’ worth of data.
Law passage times grow exponentially

This extremely simple model suggests reporting will not be universally required for several years.

December 17, 2010

Take that with a grain of salt, but perhaps we should look closely at what these laws offer us and learn from it.
US Data Breach Laws: Date Passed

Data: National Council of State Legislatures, Perkins Coie
Graphic: IBM Many Eyes
US Data Breach Laws: Entities Covered

Data: National Council of State Legislatures, Perkins Cole
Graphic: IBM Many Eyes
How Do We Know?

Reports required by national regulators

- Oversight committee reports
- FOIA

Reports required by states

- FOIA still needed (except in N.H.) but there are way fewer states than agencies
- Some primary sources available on-line
  http://doj.nh.gov/consumer/breaches.html
  http://www.cwalsh.org/cgi-bin/docview.pl

Question is: Do they add information, or just “more of the same”?

Test: Look at reports obtained by states, and reports obtained through “traditional means”. What, if anything, is added?
Central reporting is uncommon
What is collected by states?
A Quick Test

Look at incidents involving entities based in New York

Should all be reported to the state, since New Yorkers undoubtedly involved
Should appear in “traditional” reports
“Traditional” data set
University of Washington (based on Attrition, Privacyrights.org, news reports)

NY reports
Obtained via FOIA requests

If the picture is markedly different, state reports add value.
Green: University of Washington
Blue: New York reports

This is new information!
Line segments show incident observation rates for multiple sources, over time.
The Bigger Stuff makes the news?
What are the weak points?

<table>
<thead>
<tr>
<th></th>
<th>Exposed Online</th>
<th>External Intrusion</th>
<th>Insider Abuse or Theft</th>
<th>Missing or Stolen Hardware</th>
<th>Mishandled</th>
<th>Other</th>
<th>Unspecified</th>
</tr>
</thead>
<tbody>
<tr>
<td>UWash</td>
<td>3</td>
<td>1</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New York</td>
<td>17</td>
<td>7</td>
<td>3</td>
<td>65</td>
<td>2</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>New York &gt; 99</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>37</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Results for NY, and for NY cases with more than 99 individuals affected, are statistically indistinguishable.

Lesson: Keep track of your stuff, and know how to configure your web server.
<table>
<thead>
<tr>
<th>Location</th>
<th>Exposed Online</th>
<th>Insider Abuse or Theft</th>
<th>Missing or Stolen Hardware</th>
</tr>
</thead>
<tbody>
<tr>
<td>UWash</td>
<td>1.6%</td>
<td>0.5%</td>
<td>97.9%</td>
</tr>
<tr>
<td>New York</td>
<td>1.0%</td>
<td>0%</td>
<td>98.7%</td>
</tr>
</tbody>
</table>

Or, maybe ... Just keep track of your stuff!
<table>
<thead>
<tr>
<th>Industry</th>
<th>New York</th>
<th>UWash</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utilities</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Manufacturing</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Retail Trade</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Transportation and Warehousing</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Information</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Finance and Insurance</td>
<td>34</td>
<td>2</td>
</tr>
<tr>
<td>Educational Services</td>
<td>28</td>
<td>0</td>
</tr>
<tr>
<td>Health and Social Assistance</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>Arts, Entertainment, Recreation Accommodation and Food Service</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Public Administration</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>Other Services</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
The California Department of Consumer Affairs reported May 27 that since the state's notification law went into effect in July 2003, it has been aware of 61 significant breach notifications involving an average of 163,500 individuals each. About one-fourth of the breaches occurred at financial institutions and another one-fourth at universities, with 15 percent reported by medical institutions, 8 percent by government and 7 percent by retailers, according to the figures.
June 1, 2005:

The California Department of Consumer Affairs reported May 27 that since the state's notification law went into effect in July 2003, it has been aware of 61 significant breach notifications involving an average of 163,500 individuals each. About one-fourth of the breaches occurred at financial institutions and another one-fourth at universities, with 15 percent reported by medical institutions, 8 percent by government and 7 percent by retailers, according to the figures.
So what now?
Should we only care about lost/stolen media and hardware?

What about low-frequency, huge impact events?
Massive retailer breaches?
Card processor breaches?

Small breaches may also be signs of poor practices.

Additional reporting, and clarification of notification requirements would help us get the information we need to make risk decisions.
More states’ information would help

• Would let us get a better handle on (seemingly) rare events
• Would expose biases (if any) in current, “traditional” reporting
• Would help us to assess whether breaches tend to be local, regional, or national
• Would better inform national and international policy makers
• Would better reveal the role of third parties as “impact magnifiers”
How to obtain this additional information?

• Revise existing laws to add central reporting
• Adopt breach notification requirements beyond U.S.
• Pass US Federal legislation
• Increase voluntary notification
Revise existing laws

• Require reporting to state Attorney General or consumer protection agency
• Standardize reporting to enhance comparability of states’ data
• Close loopholes so that breached entity must report, whether it owns data or not.
Adopt breach notification requirements beyond U.S.

While privacy protections afforded to data subjects are significantly greater in many non-US nations, the extent to which these translate into different rates of data exposure is not known.
Pass US Federal Legislation

Legislation on a national level would eliminate a blind spot: federal agencies not bound by state law.

Central reporting is critical: eliminates need to individually request data from scores of agencies.
Increase Voluntary Reporting

• Higher notification trigger, but mandatory reporting to central entity?
• As means of limiting possible subsequent legal liability
• If you tell people, they can take steps, and thereby limit your risk
• Normative pressure: Customers expect it, law or no law
• Honesty never killed anybody: TJX sales rise after they tell of very large breach!
• Reflexive secrecy could be punished by regulators: why risk it?
• It’s an assurance game: Sharing helps all if sufficient numbers share. We just need to get there.
Things We Might Care About

Breach consequences
Impact on stock price
Impact on customer loyalty/”churn”
Direct notification costs
Impact on identity theft
Repeat offenders? Do they learn?

Aspects of the notifications themselves
Do they show acceptance of responsibility?
Is there a clear “CYA” tone?
What level of detail do they provide?
Do standard forms increase the amount of information provided?
Thanks