Is this a Privacy Incident?

Using News Exemplars to Study End User Perceptions of Privacy Incidents
(Work in Progress)

Pradeep K. Murukannaiah\textsuperscript{1}, Jessica Staddon\textsuperscript{2}, Heather Richter Lipford\textsuperscript{3}, and Bart P. Knijnenburg\textsuperscript{4}

\textsuperscript{1}Rochester Institute of Technology, \textsuperscript{2}NC State University, \textsuperscript{3}UNC Charlotte, and \textsuperscript{4}Clemson University
Broader Context: Privacy Incidents Database (PrIncipedia)

http://go.ncsu.edu/privacyincidents

### Privacy Incidents Database

<table>
<thead>
<tr>
<th>Date</th>
<th>Keywords</th>
<th>Description</th>
<th>Source</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015-09</td>
<td>#whsmith</td>
<td>WHSmith accidentally emailed customer personal details of those completing contact form to all customers in their mailing list.</td>
<td><a href="http://www.business">www.business</a> Insider.com</td>
<td>09-2015- PersonalInfoLeak-WHSmith</td>
</tr>
<tr>
<td>2015-09</td>
<td>#countrygovernment</td>
<td>Lebanon, New Hampshire library stops participating in the anonymous communication network, Tor, after concerns are raised by police.</td>
<td><a href="http://www.concordmonitor.com">www.concordmonitor.com</a></td>
<td></td>
</tr>
<tr>
<td>2015-09</td>
<td>#Google</td>
<td>Google agreed to remove links regarding an old criminal conviction, but the removal inspired lots of news articles that Google does not want to remove. ICO then issued the first public enforcement notice based on the “right to be forgotten”.</td>
<td><a href="http://www.jdispra.com">www.jdispra.com</a></td>
<td></td>
</tr>
<tr>
<td>2015-09</td>
<td>#Facebook</td>
<td>Pictures of under-age girls in Maine and elsewhere repeatedly posted on Facebook. Considered revenge porn.</td>
<td><a href="http://www.centralmaine.com">www.centralmaine.com</a></td>
<td></td>
</tr>
<tr>
<td>2015-09</td>
<td>#Google</td>
<td>Google begins allowing advertisers to target users based on email address. Restrictions are in place to deter individual profiling of users.</td>
<td><a href="http://www.theverge.com">www.theverge.com</a></td>
<td></td>
</tr>
<tr>
<td>2015-09</td>
<td>#PresidentialCandidates</td>
<td>Presidential candidate web sites found to often be poor on privacy (e.g. no privacy policy or inappropriate policies).</td>
<td><a href="http://www.nbcnews.com">www.nbcnews.com</a></td>
<td>09-2015-PrivacyPolicy-PresCandidateSites</td>
</tr>
</tbody>
</table>
Broader Context: Privacy Incidents Database (PrIncipedia)

http://go.ncsu.edu/privacyincidents

Is this a Privacy Incident?
Defining Privacy Incidents: Who should Care and Why?

Organizations

- Identify and respond to privacy incidents

Research community

- Analyze privacy incidents to inform technology improvements

Government

- Inform policy

These require an **actionable** and **accurate** definition of a *privacy incident*
Defining Privacy Incidents

Defining privacy incidents is nontrivial

- Attitudes vary by cultures and context
- Perceptions vary by stakeholder groups (e.g., end users, product developers, policy makers, and legal scholars)

We conduct a first study on end user understanding of privacy incidents involving 482 Amazon Mechanical Turk (AMT) users
Methodology

1. Working definition
Formulate a working definition of privacy incident

2. Research questions (RQs)
Formulate RQs corresponding to different aspects of the definition

3. User study
Provide examples and counterexamples to end users

4. Definition validation
Measure precision of the definition based on users’ perceptions
A privacy incident is

1. An instance of accidental or unauthorized **collection**, **use** or **exposure** of sensitive information, OR,

2. An event that creates the **perception** that unauthorized collection, use or exposure of sensitive information may happen, AND,

3. Involves information that is either being collected, used or shared in **digital** form.
Research Questions

RQ1  Do end users perceive the collection, usage and/or sharing of sensitive data to be closely associated with privacy incidents?

RQ2  Do end users recognize events in which it is perceived or anticipated that sensitive data is or will be collected, used or shared, as privacy incidents?

RQ3  Are there privacy events beyond the scope of the working definition (e.g., the release of privacy-enhancing products or privacy laws) that end users recognized as incidents?
Read the news article in the URL below and answer the following questions.

1. Is this article primarily about a privacy incident?
   - Yes
   - No
   - Not sure

2. Is this article primarily concerned with the collection, use or exposure of sensitive information?
   - Yes
   - No
   - Not sure

3. Is this article primarily concerned with information that is either being collected, used or shared in digital form?
   - Yes
   - No
   - Not sure

4. Consider now that for an article to be primarily about privacy incident, the answers to the previous two questions must be yes. Given this, do you think the article is primarily about a privacy incident?
   - Yes
   - No
   - Not sure
User Study: Positive Examples (Set $P$)

**Data Breach**  Blippy allows some credit card numbers to be indexed by search engines

**Emerging Technology**  Police departments building DNA databases of potential suspects

**Surveillance**  PA SD remotely activates cameras to locate school laptops

**Targeting**  Facebook uses visits to sites with like button to target ads

**Revenge Porn**  Woman posts revenge porn pictures; convicted under UK law
User Study: Negative Examples (Set $N$)

Security Breach  Security of an electronic road sign is compromised

Physical Security Breach  Security breach at a Donald Trump rally

Celebrity Privacy Request  Professional athlete asks for privacy as he enters drug rehab

Wildlife Privacy  Park visitors disturb privacy of animals during mating season

Privacy Standards  EFF announces a stronger DNT standard

Security/Privacy Law  Companies warn about privacy implications of cybersecurity bill
Participants and Measurements

Study Units

- \(|P| = 204\) and \(|N| = 63\)

Participants

- 482 unique Amazon Mechanical Turk users from the US
- At least three unique responses per HIT
- Maximum of three HITs per participant
- Payment: USD 0.7 per HIT

Measurements

- Measure Precision\((P)\) and Precision\((N)\)
- Measure initial and final Precision
Findings: Collection/Usage/Sharing (RQ1)

- Initial Precision($P$) = 0.799, and Final Precision($P$) = 0.794
- Initial Precision($N$) = 0.549, and Final Precision($N$) = 0.600

Participants view sensitive data collection, usage and/or sharing as core to their understanding of the term “privacy incident”

Participants more likely to see the collection, usage and sharing aspects as sufficient for a privacy incident, than as necessary
Findings: Anticipated/Perceived Privacy Harm (RQ2)

13 articles about emerging technologies (e.g., DNA testing) in $P$

- Initial Precision($P$) = 0.769, and Final Precision($P$) = 0.769

5 articles about privacy policy changes in $P$

- Initial Precision($P$) = 0.714, and Final Precision($P$) = 0.643

Participants are more inclined than privacy law to view perceived or anticipated privacy issues as grounds for an incident.
Findings: Definition Scope (RQ3)

- Only 3 of the 13 articles about legal privacy events were viewed as non-incidents.
- 2 of the 3 articles about digital security incidents that do not involve a privacy breach were reported to be privacy incidents.

Participants do not consistently view privacy law events and security breaches that do not involve privacy breaches as non-incidents.

- None of the 4 such articles about cyberbullying, involving sexual assault, were identified as privacy incidents.

Cyberbullying events are not consistently viewed as privacy incidents.
Findings: Definition Scope (RQ3)

- 14 articles from $P$ that contain the term “privacy” were reported to not be privacy incidents by a majority of participants.
- 4 of the 25 articles in $N$ that were reported to be privacy incidents, do not contain the term “privacy”.

Participants do not seem to consider the presence of the term “privacy” as a necessary or sufficient condition to treat an event as privacy incident.
Summary and Directions

- High and stable precision for $P$

Our working definition is compatible with end users’ perceptions about privacy incidents

- Increase in precision for $N$ after definition
- Substantial disagreement remains in $N$

Determine if disagreements suggest a way to improve the definition
Read the news article in the URL below and answer the following questions.

1. In a sentence, please summarize the article above.
   ____________________________________________________________

2. Briefly describe why it is or is not a privacy incident.
   ____________________________________________________________

3. Consider now that for an article to be primarily about privacy incident, the answers to the previous two questions must be yes. Given this, do you think the article is primarily about a privacy incident?
   ○ Yes  ○ No  ○ Not sure

4. If you changed your mind, please describe why.
   ____________________________________________________________