Eddy

a privacy specification language

Daniel Smullen, Travis Breaux
Carnegie Mellon University
Privacy is a balancing act.

**Minimize Privacy Risk**
- Collect only what you need.
- Obtain consent for new uses.
- Limit disclosure.
- Destroy what you don’t need.

**Maximize Data Utility**
- Collect everything.
- Open access.
- Retain and disclose as much as you can.
Privacy breach is very costly.

“...average consolidated total cost of a data breach is $3.8 million representing a 23% increase since 2013.” [1]

“the cost incurred for each lost or stolen record ... increased six percent from a consolidated average of $145 to $154…” [1]

Privacy breach is very costly.

<table>
<thead>
<tr>
<th>Breach</th>
<th>Estimated Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Home Depot</td>
<td>$56 000 000</td>
</tr>
<tr>
<td>Sony Pictures Entertainment</td>
<td>$100 000 000</td>
</tr>
<tr>
<td>Anthem</td>
<td>$100 000 000 +</td>
</tr>
<tr>
<td>Heartland Payment Systems</td>
<td>$140 000 000</td>
</tr>
<tr>
<td>TJ Maxx</td>
<td>$162 000 000</td>
</tr>
<tr>
<td>Target</td>
<td>$162 000 000</td>
</tr>
<tr>
<td>Sony Playstation</td>
<td>$171 000 000</td>
</tr>
<tr>
<td>Hannaford Brothers</td>
<td>$252 000 000</td>
</tr>
<tr>
<td>US Veterans Administration</td>
<td>$500 000 000 +</td>
</tr>
<tr>
<td>Epsilon Marketing</td>
<td>$4 000 000 000</td>
</tr>
</tbody>
</table>
What can we do about it?

- Companies protect themselves with privacy policies.
- Limit liability for privacy violations.
- Legal shields are great…
- Are they practical?
- Reusable?
- Can humans understand and use them?
Three Privacy Principles

• **Purpose specification principle:**
  • The purposes for which data is collected should be explicitly stated.

• **Collection limitation principle:**
  • Collection of personal data should be limited (to what will be used).

• **Use limitation principle:**
  • Uses should be limited to the purposes for which the data was originally collected, and nothing else.

• Exceptions for consent and legal compliance.
Three Privacy Principles

• Commonly accepted.
  • U.S. Fair Information Practice Principles (FIPPs)
  • OECD Guidelines on the Protection of Privacy and Trans-border Flows of Personal Data

• If these principles are violated, there are privacy risks.
  • Repurposing
  • Overcollection
Introducing Eddy.

- Language to formally model and analyze privacy policies.
- Leverage the information that’s out there, perform V&V.

Apply requirements, software engineering best practices;
- Make privacy policies PRACTICAL, REUSABLE, TENABLE.
1. Manually annotate policy text with codes.

   Modal phrase “may” indicates a permission

   Either as "Anonymous" user or as registered user, you may also provide
   additional personal information that will be used to personalize the Services,
   for example by providing you the shortest route between work and home,
   and enhance your experience with the Services.

2. Write expression in Eddy.

   SPEC-HEADER
   P personalizing-services > providing-shortest-route, enhancing-service-experience

   SPEC-POLICY
   P COLLECT personal-information FROM waze-user FOR personalizing-services

3. Tool parses Eddy syntax into description logic, and performs validation.

   (A) providing-shortest-route ⊑ personalizing-services
   (B) enhancing-service-experience ⊑ personalizing-services
   (C) p₆ ⊑ COLLECT ∩ ∃hasObject.personal-information ∩  
       ∃hasSource.waze-user ∩ ∃hasPurpose.personalizing-services
   (D) p₆ ⊑ Permission
Notional Scenario

- Hospital shares data with grocery store chain.
- Sell products to hospital patients.
- Suggest ‘healthy’ choices.

- HIPAA governs sharing of medical data.
Notional Scenario

• Hospital’s legal department needs to examine grocery store’s privacy policy.
• Modify policy to express new requirements.
• Establish ontology of terms and shared lexicon.
• Iterate over policy design.
Notional Scenario

- Hospital needs to align their policy with hospital policy.
- Add new statements for intended behaviors.
- Iterate over policy design.
- Hospital wants evidence that the grocery store is obeying privacy principles.
Eddy

a privacy specification language

https://gaius.isri.cmu.edu:8210/eddy/

https://github.com/cmu-relab/eddy
Further Reading

- Detecting Repurposing and Over-collection in Multi-party Privacy Requirements Specifications

- Eddy, A Formal Language for Specifying and Analyzing Data Flow Specifications for Conflicting Privacy Requirements

- Extracting goals from privacy policies
  [Antón et al.,2004; Breaux & Antón, 2005; Young et al., 2011]

- Formal models of privacy-related requirements
Visualizations!

P COLLECT device-id, ip-address, P.COLLECT device-id, location, ... FROM end-user FOR anything
FROM application FOR anything

Legend:
← hasObject
←—— hasSource
←—— hasPurpose
Blue: overflow
Red: underflow
Black: exact flow

P TRANSFER device-id
FROM anyone FOR anything
Visualizations!

Waze

- \( W_{P6} \)
  - P COLLECT personal-information
  - FROM waze-user FOR enhancing-service-experience

- \( W_{P46} \)
  - P TRANSFER unique-device-id
  - FROM anyone FOR anything TO ad-networks

Flurry

- \( F_{P1} \)
  - P COLLECT device-id, device-os, mac-address
  - FROM anyone FOR anything

Legend:
- \( \leftarrow \): hasObject
- \( \leftarrow \): hasSource
- \( \leftarrow \): hasPurpose
- Grey: Waze/Flurry system boundary
- Blue: overflow
- Red: underflow
- Black: exact flow
Eddy

a privacy specification language

https://gaius.isri.cmu.edu:8210/eddy/

https://github.com/cmu-relab/eddy