Minining Rule Semantics to Understand Privacy Legislation

Travis D. Breaux and Annie I. Antón
North Carolina State University
{tdbreaux, aianton}@eos.ncsu.edu
WPES 2005, November 7th 2005

Presentation Outline

- Research Motivation
- Semantic Parameterization
- Example Semantic Models
- Parameterized Operators
- Future Work & Summary
Towards Machine-enforceable Policies

- Motivations
  - Privacy laws require companies to enforce their policies.
  - Consumers are increasingly concerned about privacy violations.
  - Companies are increasingly being held accountable for their privacy practices.

- Problem Statement
  ... without machine-readable and machine-enforceable policies, privacy practices will continue to be inconsistently applied and therefore prone to violations.

Need a policy language that can...

- Represent rights and obligations.
  - Rights, like permissions, describe what people and systems are permitted to do.
  - Obligations describe what people and systems are required to do.

- Interface to natural language, policies must...
  - be maintainable by non-technical policy analysts.
  - be implementable by system administrators.
  - be legally enforceable by a court of law.

- Interface to program execution, policies must...
  - exclusively decide policy-governed control flow.
  - associate governance semantics with data.
From Policies to Semantic Models

(a) Policies as Restricted Natural Language Statements (RNLS).
(b) RNLS are parameterized to build semantic models.

Simple Semantic Model

[POLICY’05]

- **RNLS**: The provider may share information with whom?.

  - $\sigma(\text{activity})$
  - $\sigma(\text{activity, actor})$
  - $\sigma(\text{activity, action})$
  - $\sigma(\text{activity, object})$
  - $\sigma(\text{activity, target})$

  - $\delta(\text{actor, provider})$
  - $\delta(\text{action, share})$
  - $\delta(\text{object, information})$
  - $\delta(\text{target, ?whom})$

- The modal “may” indicates a **right**.
  - $\alpha(\text{provider, right})$
  - $\delta(\text{right, activity})$

  **KTL Expression:**
  
  ```
  activity [ right : provider ] { 
    actor = provider 
    action = share 
    object = information 
    target = ?whom 
  }
  ```
Targeted and Open-ended Queries

Two types of queries:
- Boolean queries - pair-wise relational match.
- Wh-queries - pair-wise relational match with variables store corresponding values as query responses.

Example:
- What information may be shared with whom?

<table>
<thead>
<tr>
<th>ID</th>
<th>Object</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>155</td>
<td>transaction information</td>
<td>subsidiary</td>
</tr>
<tr>
<td>156</td>
<td>experience information</td>
<td>affiliate</td>
</tr>
<tr>
<td>954</td>
<td>statistics</td>
<td>third-party</td>
</tr>
</tbody>
</table>

Example from HIPAA Privacy Rule

- Providers will <provide the patient access to their medical records> within <30 days of the patient’s request>.

- Semantic models for two activities as events:
  - M₁: Patient requests access (via right).
  - M₂: Provider provides access (via obligation).

- Unit of time: 30 days.

Rule: if { M₁ } then { M₂ <time { 30 days +_time M₁ } }
Arithmetic, Comparative Operators

[HIPAA Privacy Rule]

<table>
<thead>
<tr>
<th>Keyword</th>
<th>A</th>
<th>C</th>
<th>N</th>
<th>HIPAA Privacy Rule Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>less</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>Not less than 30 days before...</td>
</tr>
<tr>
<td>more</td>
<td>27</td>
<td>10</td>
<td>0</td>
<td>Contains more than 20,000 people...</td>
</tr>
<tr>
<td>before</td>
<td>1</td>
<td>9</td>
<td>9</td>
<td>At least 15 days before the...</td>
</tr>
<tr>
<td>after</td>
<td>20</td>
<td>8</td>
<td>2</td>
<td>180 days after the effective date...</td>
</tr>
<tr>
<td>older</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Age 90 or older...</td>
</tr>
<tr>
<td>smaller</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>Geographic subdivisions smaller than a state...</td>
</tr>
<tr>
<td>longer</td>
<td>2</td>
<td>7</td>
<td>0</td>
<td>No longer than 30 days from the date...</td>
</tr>
</tbody>
</table>

Arithmetic (A), Comparative (C), Neither (N)

Parameterized Operators

Make it possible to ...

- Compare semantic models using nested properties.
  - Evaluate $E_1 < E_2$, comparing times of two events.
  - Evaluate $E_1 < T_1$, comparing an event and a time.
  - Evaluate $E_1 + T_2$, sum of time of an event and time.
- Statically detect ambiguous references.
  - Suppose $E_1$ has a start and end time, then which time is used to evaluate $E_1 < E_2$?
Current and Future Work

- **Case Study:** *The HIPAA Privacy Rule*, enforced by the Dept. of Health and Human Services.
  - Extracting access control rules governing use and disclosure of protected health information.
  - Representing our constraints in RBAC, XACML, Ponder.

- **Case Study:** Organizational Security Policies
  - New theory relating security requirements to business processes.
  - Framework for tracing security goals from managers to implementations by administrators.

Feedback and Questions?

To see more of our work, visit our website:

http://ThePrivacyPlace.org