Semantic Enforcement of Privacy Protection Policies via the Combination of Ontologies and Rules

Prof.(Dr.) Yuh-Jong Hu

Emerging Network Technology(ENT) Lab.
Department of Computer Science
National Chengchi University, Taipei, Taiwan

June-13-2008

Ambient Semantic Computing (ASC) Workshop 2008
Outline I

1. Background
2. Research Goals
3. Privacy Protection on the Web
4. Privacy Protection for a Mailserver
5. Conclusion
Natural Languages Lack Formal Semantics

- Lack formal and unambiguous semantics
- Please show me the path for:
  - Natural Languages $\Rightarrow$ Controlled Languages
  - Controlled Languages $\Rightarrow$ Semantic Web Languages
- Semantic Web Languages = Ontology Languages+Rule Languages
- Ontology Languages: RDF(S), OWL
- Rule Languages: RuleML, RIF, N3
XML Languages Lack Semantics

XML-based Languages for Policy

- XrML [Con02] ← digital rights expression language
- ODRL [Ian02] ← digital rights expression language
- P3P [C+ 02] ← privacy rights expression language
- EPAL [And06] ← privacy rights expression language
- XACML [And06] ← rights expression language
Google Mail: Google's approach to email

Google Mail Privacy Notice
14 October 2005

The Google Privacy Policy describes how we treat personal information when you use Google's products and services, including information provided when you use Google Mail. In addition, the following describes our privacy practices that are specific to Google Mail.

Personal information
- You need a Google Account to access Google Mail. Google asks for some personal information when you create a Google Account, including your alternative contact information and a password, which is used to protect your account from unauthorised access. A Google Account allows you to access many of our services that require registration.
- Google stores, processes and maintains your messages, contact lists and other data related to your account in order to provide the service to you.
- When you use Google Mail, Google's servers automatically record certain information about your use of Google Mail. Like other web services, Google records information such as account activity (including storage usage, number of log-ins), data displayed or clicked on (including UI elements, ads, links); and other log information (including browser type, IP address, date and time of access, cookie ID and referrer URL).

Uses
- Google maintains and processes your Google Mail account and its contents to provide the Google Mail service to you and to improve our services. The Google Mail service includes relevant advertising and related links based on the IP address, content of messages and other information related to your use of Google Mail.
- Google's computers process the information in your messages for various purposes, including formatting and displaying the information to you, delivering advertisements and related links, preventing unsolicited bulk email (spam), backing up your messages and other purposes relating to offering you Google Mail.
- Google may send you information related to your Google Mail account or other Google services.

Information sharing and onward transfer
- When you send email, Google includes information such as your email address and the email itself as part of that email.
- We provide advertisers only with aggregated non-personal information such as the number of times one of their ads was clicked. We do not sell, rent or otherwise share your personal information with any third parties except in the limited circumstances described in the Google Privacy Policy, such as when we believe we are required to do so by law.

Your choices
- You may change your Google Mail account settings through the Google Mail "settings" section.
- You may organise or delete your messages through your Google Mail account or terminate your account through the Google Account section of Google Mail settings. Such deletions or terminations will take immediate effect in your account view. Resident copies of deleted messages and accounts may take up to 80 days to be deleted from our active servers and may remain in our offline backup systems.
- You may choose to use additional Google Mail features, such as Google Talk. The Google Talk service has its own privacy notice available here.
Generic DL (⊂ FOL) and Pure LP Are Machine Unfriendly

Related Works

- Semantic ODRL [PW06] ⇐ FOL semantics
- Semantic XrML [HW08] ⇐ FOL semantics
- Semantic P3P [YNLA04] ⇐ relational semantics
- FAF [J⁺ 01] ⇐ LP semantics
- Semantic E-P3P (and EPAL) [And06] ⇐ FAF semantics
- Rei, KAoS [T⁺ 03] ⇐ DL-based FOL semantics
Why Use Ontologies+Rules (O+R) Combination?

**Primary Reasons**

- Two Major KRs: Ontologies and Rules
- Semantic Web Research Core
- W3C Web Markup Languages: RDF(S), OWL-DL, RIF, etc
Why Use Ontologies+Rules (O+R) Combination?

**Primary Reasons**
- Two Major KRs: Ontologies and Rules
- Semantic Web Research Core
- W3C Web Markup Languages: RDF(S), OWL-DL, RIF, etc

**Representation and Enforcement for**
- License Agreements
- Access Control Policies
- Protection Systems
Why Use Ontologies+Rules (O+R) Combination?

Why Not Ontologies Alone or Rules Alone?

- Expressive Power Enhancement from Ontologies or Rules
- For Possible Knowledge Representation, Integration, Interchange, and Interoperation
- Options to Use Ontologies Alone or Rules Alone
Which Ontologies+Rules (O+R) Combination?

Criteria for the Selection of O+R

- Computational Decidability
- Expressive Power
- OWA and CWA Semantic Differences
- Knowledge Flow:
  - Uni-directional for ontologies \((\Rightarrow \land \not \Leftrightarrow)\) rules
  - Bi-directional for ontologies \((\Leftrightarrow)\) rules
- Tight or Loose Integration
Ontologies+Rules (O+R) Combination [Ros06b]

**Tight Integration**

- **CARIN [LR96]** $\Leftarrow$ limited expressive power
- **DLP [G⁺ 03]** $\Leftarrow$ too restricted expression
- **SWRL [H⁺ 04]** $\Leftarrow$ undecidable computation
Ontologies+Rules (O+R) Combination [Ros06b]
Loose Integration

Positive Datalog Rules

1. (Disjunctive) AL-log $[D^+ 98] \iff$ decidability of ALC plus positive, recursive DL-safe rules
2. DL-Safe Rules [MSS04] $\iff$ decidability of SHOIN plus positive, recursive DL-safe rules
## Ontologies+Rules (O+R) Combination [Ros06b]

### Loose Integration

<table>
<thead>
<tr>
<th>Positive Datalog Rules</th>
<th>Non-Monotonic Datalog Rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Disjunctive) AL-log [D^+ 98] ⇐ decidability of ALC plus positive, recursive <strong>DL-safe</strong> rules</td>
<td><strong>DL-log</strong> [Ros05] ⇐ decidability of DLs/FOL plus non-monotonic, recursive <strong>DL-safe</strong> rules</td>
</tr>
<tr>
<td><strong>DL-Safe Rules [MSS04]</strong> ⇐ decidability of SHOIN plus positive, recursive <strong>DL-safe</strong> rules</td>
<td><strong>DL + log</strong> [Ros06a] ⇐ decidability of arbitrary DLs plus non-monotonic, recursive <strong>weakly DL-safe</strong> rules</td>
</tr>
<tr>
<td><strong>MKNF</strong> [M^+ 06] ⇐ mixes OWA and CWA reasoning in DL-safe rules</td>
<td></td>
</tr>
</tbody>
</table>
Semantic Web Well-Known Layer Cake
Long Term Research Goals

**Semantic-Enabled Policy Languages**

- To exploit possible ontologies+rules combinations for the semantic-enabled policy languages
- To design and implement the semantic-enabled languages, policies, and systems
- To demonstrate the semantic enforcement of DRM systems on the Web
- To demonstrate the semantic enforcement of privacy protection systems on the Web
Short Term Research Goals

Semantic-Enabled DRM System

- To resolve formal semantic issues of ODRL/XrML
- To construct a formal semantics model for ODRL/XrML
- To exploit semantic enforcement of DRM policies
- To implement a fully semantic-enabled DRM system
Short Term Research Goals

**Semantic-Enabled DRM System**

- To resolve formal semantic issues of ODRL/XrML
- To construct a formal semantics model for ODRL/XrML
- To exploit semantic enforcement of DRM policies
- To implement a fully semantic-enabled DRM system

**Current Status: Semantic-Enabled DRM [Hu07]**

- Exploiting a XML-based ODRL Information Model
- Designing a semantic right expression language (REL) for DRM policies and systems
- Proposing a unifying semantic REL for the DRM and privacy protection systems
- Proposing and implementing an O+R-based DRM system
Short Term Research Goals (conti.)

Semantic-Enabled Privacy Protection System

- To resolve formal semantic issues of P3P and EPAL
- To construct a formal semantics model for P3P/EPAL
- To exploit semantic enforcement of privacy protection policies
- To implement a semantic-enabled privacy protection system
Short Term Research Goals (conti.)

<table>
<thead>
<tr>
<th>Semantic-Enabled Privacy Protection System</th>
</tr>
</thead>
<tbody>
<tr>
<td>To resolve formal semantic issues of P3P and EPAL</td>
</tr>
<tr>
<td>To construct a formal semantics model for P3P/EPAL</td>
</tr>
<tr>
<td>To exploit semantic enforcement of privacy protection policies</td>
</tr>
<tr>
<td>To implement a semantic-enabled privacy protection system</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Current Status: Semantic-Enabled Privacy Protection System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exploiting a XML-based P3P/EPAL information model</td>
</tr>
<tr>
<td>Designing a semantic right expression Language (REL) for privacy protection</td>
</tr>
<tr>
<td>Proposing and implementing an O+R-based privacy protection system</td>
</tr>
</tbody>
</table>
A Unifying Semantic REL
Data User Ontologies (conti.)

Data user property:

- HAS_COOPERATIVE
  - domain: COMPANY
  - range: COMPANY

- HAS_SUBSIDIARY
  - domain: COMPANY
  - range: COMPANY

- IS_STAFF_OF
  - domain: PERSON
  - range: ORGANIZATION
Data Type Ontologies (conti.)
Purpose Ontology (conti.)
Use Case One Scenario

1. Alice wants to send e-mail to Bob and Charlie

   - e-mail of Bob:
     - from: Alice@gmail.com
     - to: Bob@yahoo.com.tw, Charlie@hotmail.com
     - Subject: Data-Auditing

   Alice

   - e-mail of Charlie:
     - from: Alice@gmail.com
     - to: Charlie@yahoo.com.tw
     - Subject: Data-Auditing

2. Bob doesn’t want to disclose his e-mail address to other recipients not in subsidiary company

3. Charlie will receive the e-mail without displaying the e-mail address of Bob
Use Case Two Scenario

G is a mail server company

- e-mail of Bob:
  from: Alice@government.org
  to: Bob@government.org
  Charlie@hotmail.com
  Subject: Account-Auditing

- e-mail of Charlie:
  from: Alice@government.org
  to: Charlie@hotmail.com
  Subject: Account-Auditing

G

Y
  - Alice
  - Bob

Y is a government agency

M
  - Charlie

M is a company
Ontologies Module

Example (Ontologies Module’s Axiom)

- COMPANY ⊑ PRIVATE
- PRIVATE ⊑ ORGANIZATION
- OWNER ⊑ PERSON
- COMPANY domain ←− HAS_COOPERATIVE range −→ COMPANY
- COMPANY domain ←− HAS_SUBSIDIARY range −→ COMPANY
- HAS_COOPERATIVE ≡ HAS_COOPERATIVE−
- PERSON domain ←− IS_STAFF_OF range −→ ORGANIZATION
- MAIL_TRACE domain ←− HAS_MAIL_TRACE range −→ EMAIL
- EMAIL ⊑ ∃ HAS_MAIL_TRACE_ONLINE−.O_EMAIL_SENDER
- EMAIL ⊑ ∀ HAS_MAIL_TRACE_ONLINE.O_EMAIL_RECEIVER
- DATA_AUDIT_ANNOUN. ⊑ AUDIT_ANNOUN.
Ontologies Module

Example (Ontologies Module’s Facts)

- ORGANIZATION(G)
- HAS_SUBSIDIARY(G, J-Corp.)
- HAS_COOPERATIVE(G, Q-Corp.)
- IS_STAFF_OF(Alice, J-Corp.)
- IS_STAFF_OF(Bob, J-Corp.)
- IS_STAFF_OF(Charlie, Q-Corp.)
- HAS_EMAIL_ADDRESS(Charlie, Charlie@hotmail.com)
- HAS_EMAIL_ADDRESS(Alice, Alice@gmail.com)
- HAS_EMAIL_ADDRESS(Bob, Bob@yahoo.com.tw)
- O_EMAIL_RECEIVER(Bob@yahoo.com.tw)
- O_EMAIL_SENDER(Alice@gmail.com),
- O_EMAIL_RECEIVER(Charlie@hotmail.com)
- HAS_MAIL_TRACE_ONLINE
  (Alice@gmail.com, Bob@yahoo.com.tw)
- HAS_MAIL_TRACE_ONLINE
  (Alice@gmail.com, Charlie@hotmail.com)
Example (Rules Module’s Rules)

- **opt-out(?b,?b-email, ?p)**
  $$\iff \text{data-owner(?b), data-user(?c), purpose(?p), data-type(?b-email),}
  \text{IS_STAFF_OF(?b,?c1), IS_STAFF_OF(?c, ?c2), HAS_COOPERATIVE(?c1,?c2),}
  \text{HAS_MAIL_TRACE_ONLINE(?a-email,?c-email),}
  \text{O_EMAIL_SENDER(?a-email), O_EMAIL_RECEIVER(?c-email).} \iff (a4)$$

- **opt-in(?b,?b-email, ?p)**
  $$\iff \text{data-owner(?b), data-user(?c), purpose(?p), data-type(?b-email),}
  \text{IS_STAFF_OF(?b,?c1), IS_STAFF_OF(?c, ?c2), HAS_SUBSIDIARY(?c1,?c2),}
  \text{HAS_MAIL_TRACE_ONLINE(?a-email,?c-email),}
  \text{O_EMAIL_SENDER(?a-email), O_EMAIL_RECEIVER(?c-email).} \iff (a3)$$

- **cando(?c,?b-email, nill)**
  $$\iff \text{opt-out(?b,?b-email,?p)), data-user(?c), data-owner(?b),}
  \text{HAS_EMAIL_ADDRESS(?b, ?b-email).} \iff (a2)$$

- **cando(?c,?b-email, display)**
  $$\iff \text{opt-in(?b,?b-email,?p)), data-user(?c), data-owner(?b),}
  \text{HAS_EMAIL_ADDRESS(?b,?b-email).} \iff (a1)$$
Example (Rules Module’s Facts)

- data-user(Bob), data-owner(Bob),
- data-user(Charlie), data-owner(Charlie),
- purpose(data-auditing),
- data-type(Bob@yahoo.com.tw),
- data-type(Charlie@hotmail.com),
- opt-in(c,Charlie@yahoo.com,data-auditing),
- cando(Bob,Charlie@yahoo.com,display),
- cando(Charlie,Bob@yahoo.com.tw,nill),
- opt-out(b,Bob@yahoo.com.tw,data-auditing)
Discussion
Policy Languages Representation and Enforcement

Natural Language

- **Pros:** human readable and understandable
- **Cons:** machine unfriendly also no formal semantics
## Discussion
### Policy Languages Representation and Enforcement

<table>
<thead>
<tr>
<th>Natural Language</th>
<th><strong>Pros:</strong> human readable and understandable</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Cons:</strong> machine unfriendly also no formal semantics</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pure FOL</th>
<th><strong>Pros:</strong> formal clear syntax and semantics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Cons:</strong> machine unfriendly also possibly undecidable computation complexity and policy writer (reader) needs to be a logician</td>
</tr>
</tbody>
</table>
Discussion (conti.)
Policy Languages Representation and Enforcement

Standard Rights Expression Languages (RELs)

- **Pros:** XML-based for automatic machine processing
- **Cons:** human unfriendly also no formal semantics
Discussion (conti.)
Policy Languages Representation and Enforcement

Standard Rights Expression Languages (RELs)
- **Pros**: XML-based for automatic machine processing
- **Cons**: human unfriendly also no formal semantics

Ontologies+Rules (O+R)
- **Pros**: formal syntax and unambiguous semantics for automatic machine processing and understanding
- **Cons**: under certain conditions with limited expressing power due to different assumption of ontologies and rules combinations
Policy Languages Classification for Usage Rights Control
We exploit the semantic rights expression languages (RELs) for enforcement of privacy protection policies.
Conclusion

- We exploit the semantic rights expression languages (RELs) for enforcement of privacy protection policies.

- We demonstrate a simple mailserver privacy protection via using one of hybrid loose ontologies+rules combination.
Conclusion

- We exploit the semantic rights expression languages (RELs) for enforcement of privacy protection policies.

- We demonstrate a simple mailserver privacy protection via using one of hybrid loose ontologies+rules combination.

- Semantic-enabled RELs for representation and enforcement of policies and systems on the Web, such as DRM and privacy protection, will be a promising research area.
I. Annie Antón et al.
A roadmap for comprehensive online for privacy policy management.

G. Antoniou et al.
Rule-based policy specification.
References II

A. H. Anderson.
A comparison of two privacy policy languages: EPAL and XACML.

A. P. Bonatti et al.
Semantic web policies - a discussion of requirements and research issues.
In 3rd European Semantic Web Conference (ESWC 2006), Budva, Montenegro, June 2006.
References III


L. Cranor et al. The platform for privacy preferences (p3p) 1.0 (p3p 1.0) specification, 2002. 
http://www.w3.org/P3P/.


S. Fischer-Hübner. 

N. B. Grosof et al.
Description logic programs: Combining logic programs with description logic.

I. Horrocks et al.
http://www.w3.org/Submission/SWRL/.
References VI

Y. J. Hu.
Semantic-driven enforcement of rights delegation policies via the combination of rules and ontologies.
In Workshop on Privacy Enforcement and Accountability with Semantics in conjunction with ISWC+ASWC’07, 2007.

Joseph Y. Halpern and Vicky Weissman.
A formal foundation for XrML.

R. Iannella.
Open digital rights language (ODRL), version 1.1.
References VII

- **S. Jajodia et al.**

- **G. Karjoth and M. Schunter.**
  A privacy policy model for enterprises.

- **G. Karjoth, M. Schunter, and M. Waidner.**
  Platform for enterprise privacy practices: Privacy-enabled management of customer data.
References VIII

Y. Alon Levy and M.-C. Rousset.
CARIN: A representation language combining horn rules and description logics.

N. Li, T. Yu, and A. I. Antón.
A semantics-approach to privacy languages.

B. Motik et al.
Can OWL and logic programming live together happily ever after?
References IX

J. Maluszynski.
Hybrid integration of rules and DL-based ontologies.
REWERS.

B. Motik, U. Sattler, and R. Studer.
Query answering for OWL-DL with rules.

The UCON_{ABC} usage control model.


References XI

R. Rosati.

*DL+log: Tight integration of description logics and disjunctive datalog.*


R. Rosati.

Integrating ontologies and rules: Semantic and computational issues.

References XII

G. Tonti et al.

S. De Capitani di Vimercati et al.
Access control policies and languages in open environments.
References XIII

D. J. Weitzner et al.  
Creating a policy-aware web: Discretionary, rule-based access for the world wide web.  

Y. C. Thomas Woo and S. S. Lam.  

T. Yu, A. N. Li, and I. Antón.  
A formal semantics for P3P.  
http://citeseer.ist.psu.edu/750176.html