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

Natural Languages for Policy

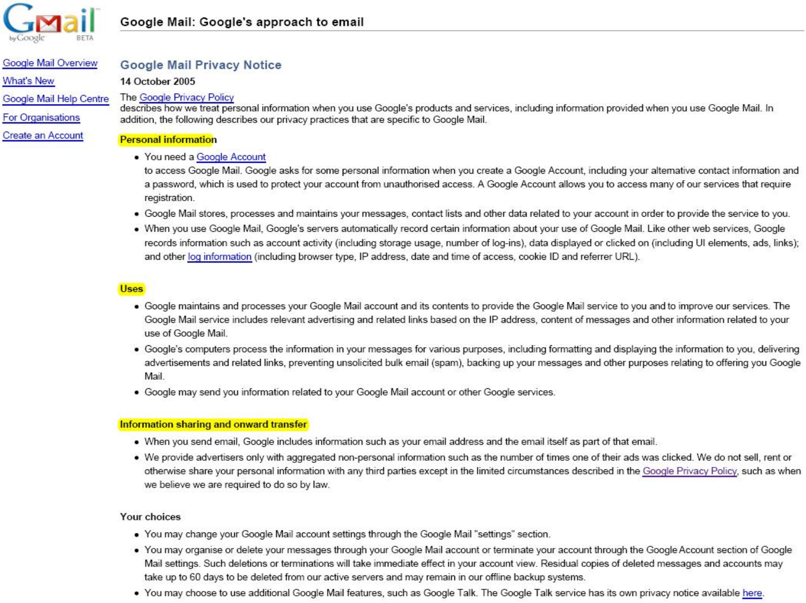
- *Lack formal and unambiguous semantics*
- *Please show me the path for:*
  - *Natural Languages ⇒ Controlled Languages*
  - *Controlled Languages ⇒ Semantic Web Languages*
- *Semantic Web Languages = Ontology Languages+Rule Languages*
- *Ontology Languages: RDF(S), OWL*
- *Rule Languages: RuleML, RIF, 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 Privacy Notice



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 Mail 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. Residual copies of deleted messages and accounts may take up to 60 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 ($\subset$ 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*

▸ Layer Cake

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* $(\Rightarrow \wedge \nLeftarrow)$ *rules*
    - *Bi-directional for ontologies* $(\rightleftharpoons)$ *rules*
- *Tight or Loose Integration*

Ontologies+Rules (O+R) Combination [Ros06b]

Tight Integration

- *CARIN [LR96]* ⇐ *limited expressive power*
- *DLP [G+03]* ⇐ *too restricted expression*
- *SWRL [H+04]* ⇐ *undecidable computation*

Ontologies+Rules (O+R) Combination [Ros06b]

Loose Integration

Positive Datalog Rules

- *(Disjunctive) AL-log [D$^+$98] ⇐ decidability of ALC plus positive, recursive DL-safe rules*
- *DL-Safe Rules [MSS04] ⇐ decidability of SHOIN plus positive, recursive DL-safe rules*

Ontologies+Rules (O+R) Combination [Ros06b]

Loose Integration

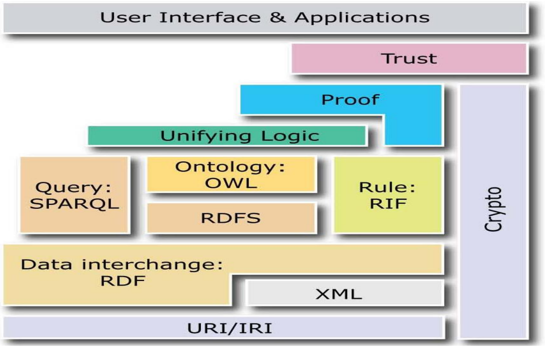
Positive Datalog Rules

- *(Disjunctive) AL-log [D$^+$98] ⇐ decidability of ALC plus positive, recursive DL-safe rules*
- *DL-Safe Rules [MSS04] ⇐ decidability of SHOIN plus positive, recursive DL-safe rules*

Non-Monotonic Datalog Rules

- *DL-log [Ros05] ⇐ decidability of DLs/FOL plus non-monotonic, recursive DL-safe rules*
- *DL + log [Ros06a] ⇐ decidability of arbitrary DLs plus non-monotonic, recursive weakly DL-safe rules*
- *MKNF [M$^+$06] ⇐ mixes OWA and CWA reasoning in DL-safe rules*

Semantic Web Well-Known Layer Cake



Return

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.)

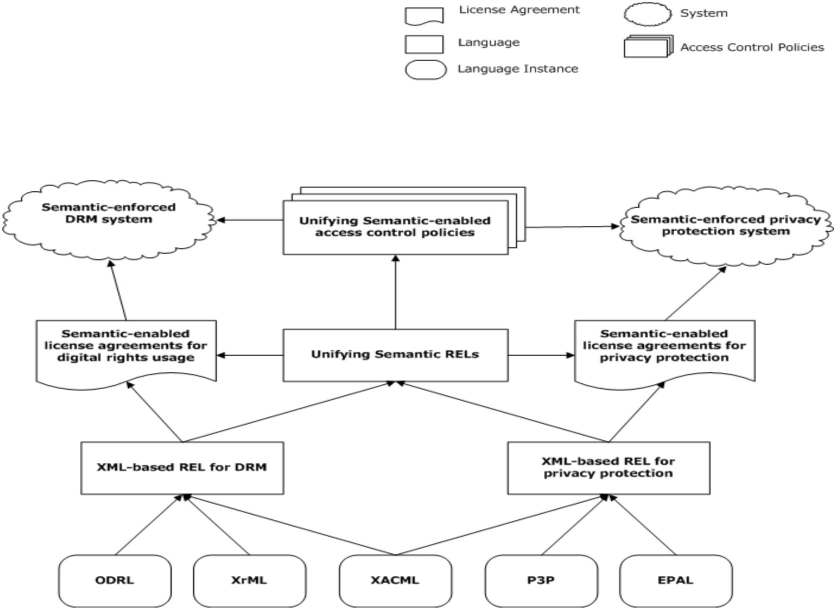
Semantic-Enabled Privacy Protection System

- *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*

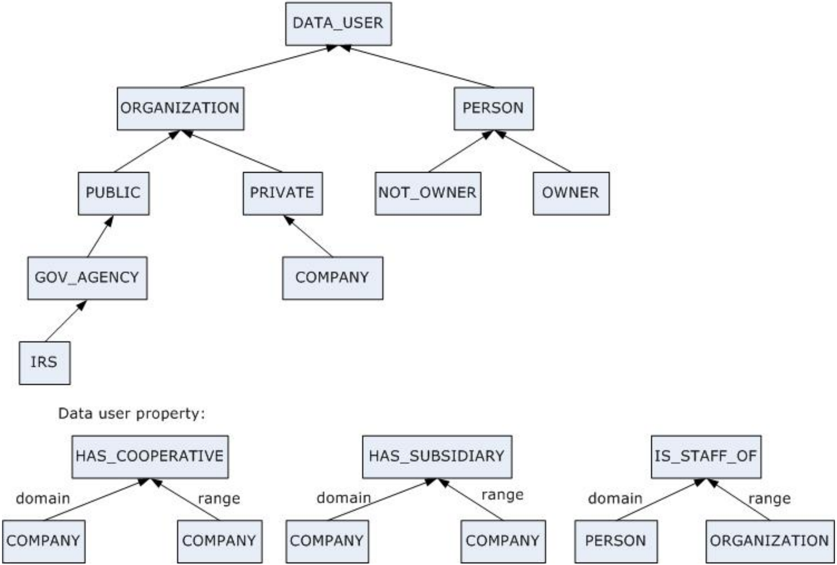
Current Status: Semantic-Enabled Privacy Protection System

- *Exploiting a XML-based P3P/EPAL information model*
- *Designing a semantic right expression Language (REL) for privacy protection*
- *Proposing and implementing an O+R-based privacy protection system*

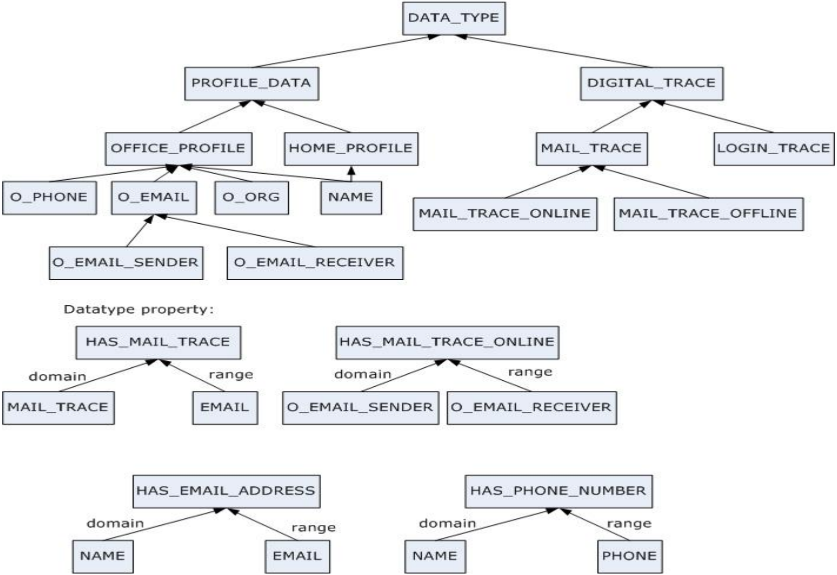
A Unifying Semantic REL


License Agreement
Language
Language Instance
System
Access Control Policies
Semantic-enforced DRM system
Unifying Semantic-enabled access control policies
Semantic-enforced privacy protection system
Semantic-enabled license agreements for digital rights usage
Unifying Semantic RELs
Semantic-enabled license agreements for privacy protection
XML-based REL for DRM
XML-based REL for privacy protection
ODRL
XrML
XACML
P3P
EPAL

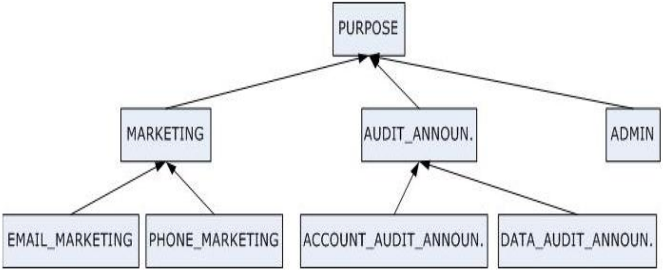
Data User Ontologies (conti.)



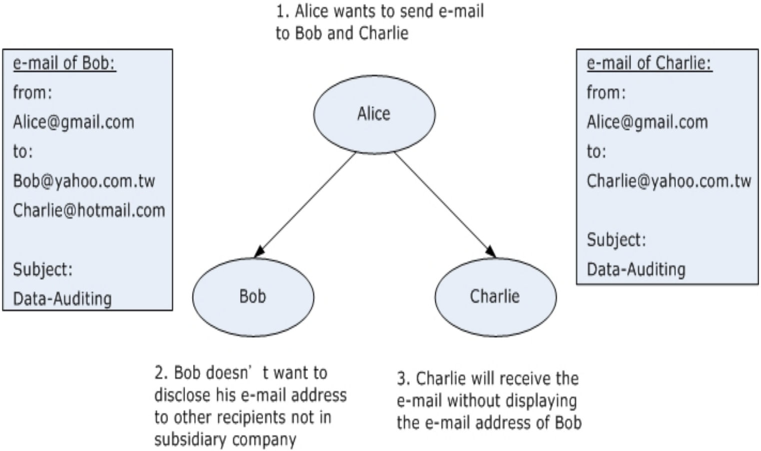
Data Type Ontologies (conti.)



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

e-mail of Charlie:

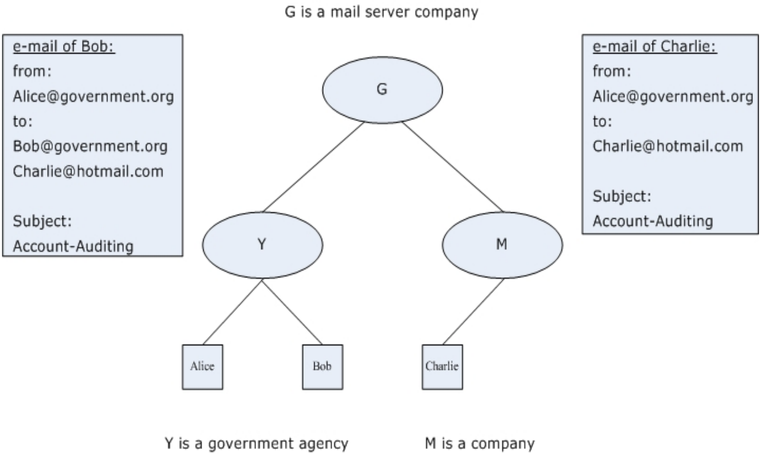
from:
Alice@gmail.com
to:
Charlie@yahoo.com.tw

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

Y is a government agency

M is a company

Ontologies Module

Example (Ontologies Module's Axiom)

- $\mathit{COMPANY} \sqsubseteq \mathit{PRIVATE}$
- $\mathit{PRIVATE} \sqsubseteq \mathit{ORGANIZATION}$
- $\mathit{OWNER} \sqsubseteq \mathit{PERSON}$
- $\mathit{COMPANY} \xleftarrow{domain} \mathit{HAS\_COOPERATIVE} \xrightarrow{range} \mathit{COMPANY}$
- $\mathit{COMPANY} \xleftarrow{domain} \mathit{HAS\_SUBSIDIARY} \xrightarrow{range} \mathit{COMPANY}$
- $\mathit{HAS\_COOPERATIVE} \equiv \mathit{HAS\_COOPERATIVE}^{-}$
- $\mathit{PERSON} \xleftarrow{domain} \mathit{IS\_STAFF\_OF} \xrightarrow{range} \mathit{ORGANIZATION}$
- $\mathit{MAIL\_TRACE} \xleftarrow{domain} \mathit{HAS\_MAIL\_TRACE} \xrightarrow{range} \mathit{EMAIL}$
- $\mathit{EMAIL} \sqsubseteq \exists\, \mathit{HAS\_MAIL\_TRACE\_ONLINE}^{-}.\mathit{O\_EMAIL\_SENDER}$
- $\mathit{EMAIL} \sqsubseteq \forall\, \mathit{HAS\_MAIL\_TRACE\_ONLINE}.\mathit{O\_EMAIL\_RECEIVER}$
- $\mathit{DATA\_AUDIT\_ANNOUN.} \sqsubseteq \mathit{AUDIT\_ANNOUN.}$

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)*

Rules Module

Example (Rules Module's Rules)

- *opt-out(?b,?b-email,?p)*
  ⟸ data-owner(?b), data-user(?c), purpose(?p), data-type(?b-email),
  IS_STAFF_OF(?b,?c1), IS_STAFF_OF(?c, ?c2), HAS_COOPERATIVE(?c1,?c2),
  HAS_MAIL_TRACE_ONLINE(?a-email,?c-email),
  O_EMAIL_SENDER(?a-email), O_EMAIL_RECEIVER(?c-email). ← (a4)

- *opt-in(?b,?b-email,?p)*
  ⟸ data-owner(?b), data-user(?c), purpose(?p), data-type(?b-email),
  IS_STAFF_OF(?b,?c1), IS_STAFF_OF(?c, ?c2), HAS_SUBSIDIARY(?c1,?c2),
  HAS_MAIL_TRACE_ONLINE(?a-email,?c-email),
  O_EMAIL_SENDER(?a-email), O_EMAIL_RECEIVER(?c-email). ← (a3)

- *cando(?c,?b-email, nill)*
  ⟸ opt-out(?b,?b-email,?p)), data-user(?c), data-owner(?b),
  HAS_EMAIL_ADDRESS(?b, ?b-email). ← (a2)

- *cando(?c,?b-email, display)*
  ⟸ opt-in(?b,?b-email,?p)), data-user(?c), data-owner(?b),
  HAS_EMAIL_ADDRESS(?b,?b-email). ← (a1)

Rules Module

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

Natural Language

- *Pros: human readable and understandable*
- *Cons: machine unfriendly also no formal semantics*

Pure FOL

- *Pros: formal clear syntax and semantics*
- *Cons: machine unfriendly also possibly undecidable computation complexity and policy writer (reader) needs to be a logician*

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

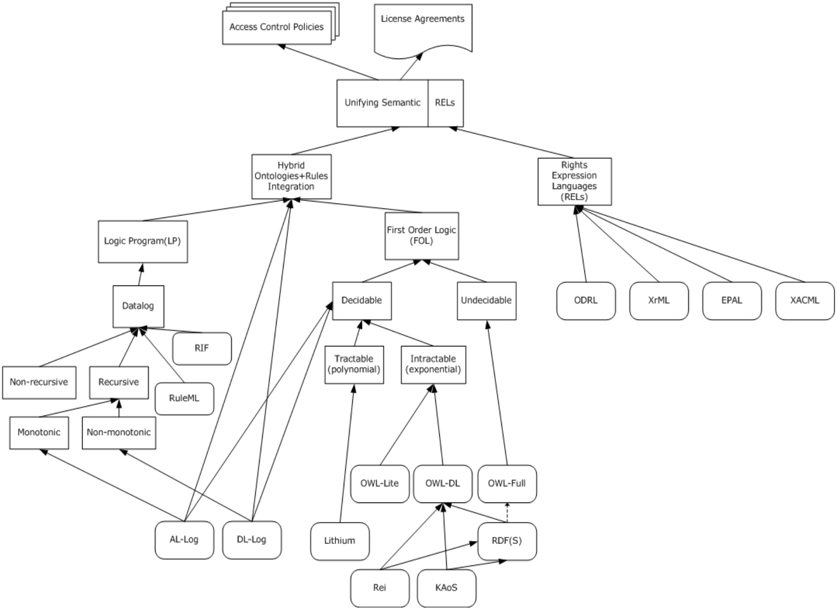
Standard Rights Expression Languages (RELs)

- *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



Conclusion

- *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.*

References I

- I. Annie Antón et al.
  A roadmap for comprehensive online for privacy policy management.
  Comm. of the ACM, 50(7):109–116, July 2007.

- G. Antoniou et al.
  Rule-based policy specification.
  In Ting Yu and S. Jajodia, editors, *Secure Data Management in Decentralized Systems*, pages 169–216. Springer, 2007.

References II

- A. H. Anderson.
  A comparison of two privacy policy languages: EPAL and XACML.
  In *Proceedings of the 3rd ACM Workshop on Secure Web Services (SWS'06)*, pages 53–60. ACM, 2006.

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

References III

A. P. Bonatti, S. De Capitani di Vimercati, and P. Smarati.
An algebra for composing access control policies.
ACM Trans. on Information and Systems Security, 5(1):1–35, February 2002.

M. Blaze, J. Figenebaum, and M. Strauss.
Compliance checking in the policymaker trust management system.
In *Proc. of the Financial Cryptography*, LNCS 1465, pages 254–274. Springer, 1998.

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

References IV

L. Cranor, M. Langheinrich, and M. Marchiori.
A p3p preference exchange language 1.0 (appel 1.0), 2002.
http://www.w3.org/TR/P3P-preferences/.

Inc. ContentGuard.
XrML: The digital rights language for trusted content and services.
Technical report, ContentGuard Inc., 2002.
http://www.xrml.org/index.asp.

M. F. Donini et al.
AL-log: Integrating datalog and description logics.
Journal of Intelligent Information Systems, 10(3):227–252, 1998.

References V

- S. Fischer-Hübner.
  IT-Security and Privacy - Design and Use of Privacy-Enhancing Security Mechanisms.
  LNCS 1958. Springer, 2001.

- N. B. Grosof et al.
  Description logic programs: Combining logic programs with description logic.
  In *World Wide Web 2003*, pages 48–65, Budapest, Hungary, 2003.

- I. Horrocks et al.
  SWRL: A semantic web rule language combing OWL and RuleML, 2004.
  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.
  Journal of the ACM, 55(1):1–42, 2008.

- R. Iannella.
  Open digital rights language (ODRL), version 1.1.
  Specifications, The ODRL Initiative, August 2002.
  `http://odrl.net/1.1/ODRL-11.pdf`.

References VII

- S. Jajodia et al.
  Flexible support for multiple access control policies.
  ACM Trans. on Database Systems, 26(2):214–260, June 2001.
- G. Karjoth and M. Schunter.
  A privacy policy model for enterprises.
  In *15th IEEE Computer Security Foundations Workshop (CSFW)*. IEEE, June 2002.
- G. Karjoth, M. Schunter, and M. Waidner.
  Platform for enterprise privacy practices: Privacy-enabled management of customer data.
  In *2nd Workshop on Privacy Enhancing Technologies (PET)*, LNCS. Springer, 2002.

References VIII

Y. Alon Levy and M.-C. Rousset.
CARIN: A representation language combining horn rules and description logics.
In *Proc. of the 12th Eur. Conf. on Artificial Intelligence (ECAI'96)*, pages 323–327, 1996.

N. Li, T. Yu, and A. I. Antón.
A semantics-approach to privacy languages.
Computer Systems and Engineering (CSSE), 21(5), Sep. 2006.

B. Motik et al.
Can OWL and logic programming live together happily ever after?
In *5th International Semantic Web Conference (ISWC) 2006*, LNCS 4273, Athens, GA, USA, Nov. 2006.

References IX

J. Maluszynski.
Hybrid integration of rules and DL-based ontologies.
In J. Maluszynski, editor, *Combining Rules and Ontologies. A survey*, pages 55–72. EU FP6 Network of Excllence (NoE), Feb. 2005.
REWERSE.

B. Motik, U. Sattler, and R. Studer.
Query answering for OWL-DL with rules.
In *3rd International Semantic Web Conference (ISWC) 2004*, LNCS 3298, pages 549–563. Springer, 2004.

J. Park and R. T. Sandhu.
The UCON$_{ABC}$ usage control model.
ACM Trans. on Information and System Security, 7(1):128–174, 2004.

References X

- F. P. Patel-Schneider and J. Siméon.
  Building the semantic web on XML.
  In *ISWC 2002*, LNCS2342, pages 147–161. Springer, 2002.
- R. Pucella and V. Weissman.
  A formal foundation for ODRL.
  arXiv:cs/0601085v1, January 2006.
  http://arxiv.org/abs/cs/0601085.
- R. Rosati.
  On the decidability and complexity of integrating ontologies and rules.
  Web Semantics: Science, Services and Agents on the World Wide Web 3, pages 61–73, 2005.

References XI

- R. Rosati.
  DL+log: Tight integration of description logics and disjunctive datalog.
  In *Proc. of the 10th International Conference on Principles of Knowledge Representation and Reasoning (KR)*, 2006.

- R. Rosati.
  Integrating ontologies and rules: Semantic and computional issues.
  In *Reasoning Web 2006*, LNCS 4126, pages 128–151, 2006.

References XII

G. Tonti et al.
Semantic web languages for policy representation and reasoning: A comparison of KAoS, Rei, and Ponder.
In *2nd International Semantic Web Conference (ISWC) 2003*, LNCS 2870, pages 419–437. Springer, 2003.

S. De Capitani di Vimercati et al.
Access control policies and languages in open environments.
In Ting Yu and S. Jajodia, editors, *Secure Data Management in Decentralized Systems*, pages 21–58. Springer, 2007.

References XIII

D. J. Weitzner et al.
Creating a policy-aware web: Discretionary, rule-based access for the world wide web.
In E. Ferrari and B. Thuraisingham, editors, *Web and Information Security*, pages 1–31. Idea Group Inc., 2006.

Y. C. Thomas Woo and S. S. Lam.
Authorization in distributed systems: a new approach.
Journal of Computer Security, 2(2-3):107–136, 1993.

T. Yu, A. N. Li, and I. Antón.
A formal semantics for P3P.
In *ACM Workshop on Secure Web Services*, Fairfax, VA, USA, Oct. 2004.
`http://citeseer.ist.psu.edu/750176.html`