Access Control for Ontologies

Willy Chen
SysTec-CAx GmbH, Germany
Knowledge Representation and Knowledge Management Research Group
Mannheimer Zentrum fuer Wirtschaftsinformatik, Germany

Heiner Stuckenschmidt
Knowledge Representation and Knowledge Management Research Group
Mannheimer Zentrum fuer Wirtschaftsinformatik, Germany

1 Motivation

Conjunctive query answering plays an important role in industrial applications of ontologies such as the integration of heterogeneous information sources and the management of information resources and services. If confidential information is included in such settings, the ability to enforce access control policies is a critical security requirement. In presence of logical inferences, however, ensuring confidentiality is not straightforward. We propose an approach to rewrite queries so they meet given access control polices in order to protect the information from unauthorized use. We thereby rely on well-known standard technologies and tools. We particularly show how our solution aligns with the OWL2 QL profile, in which conjunctive query answering can be executed via SQL queries on relational database management systems and analyze complexity issues of such an approach.

1.1 Ontologies and Industrial Applications

Semantic technologies originally developed for knowledge representation and reasoning on the web are gaining attention from industry as an adequate means for solving complex information management tasks. A constantly growing number of companies that offer professional services or tools in this area reflect this trend (Davis, 2008). Typical applications of ontologies include knowledge and skill management (Staab, 2002) as well as web service and business process management (Fensel, 2001). Successful projects (Syldatke et al., 2007) emphasize the relevance of these technologies in the area of product lifecycle management (PLM), in particular in the automotive sector. Their application in large-scaled industrial use cases (e.g., (Chen et al., 2008)) quickly shows that theoretical issues like expressiveness and decidability are less a limiting factor for their uptake in industry than non-functional aspects. In commercial scenarios we are particularly faced with ever changing domains that need to be reflected within a knowledge base, restricted resources for its development and maintenance, and the uncertain perspicuity and acceptance of such an information source by the actual users (Hepp, 2007).

1.2 The Need for Access Control

Another important requirement of many industrial applications is the need to protect knowledge against unauthorized access. This need is particularly evident in many potential application areas of ontologies such as the area of data and application integration. The integrated data sources often hold confidential information that should not be accessible for everybody. In some cases, such as the deregulation of the energy markets, there are even legal constraints on the accessibility of strategic information between different parts of the same enterprise.

Tailoring the integration of data sources for different target groups is too much effort in most cases. Therefore, the only sensible solution is a complete integration of information along with the implementation of a fine-grained access control mechanism that grants access to parts of the integrated model based on a suitable set of security policies. While this problem has been investigated in details for standard technologies such as relational databases, so far there are no convincing solutions for providing fine-grained access control for ontology-based knowledge in the presence of logical inference. This hinders the uptake of semantic technologies in industrial applications dealing with sensible information. Providing a solution to this problem therefore increases the usefulness and the potential impact of semantic technologies in practice.

1.3 Outline and Contributions

Within this paper we present a security architecture for enabling access control of ontologies within standard semantic web infrastructures without the need to modify existing reasoners. In fact, we propose the use of a security proxy, which rewrites incoming SPARQL queries so they meet prior defined access control policies. For creating and maintaining those policies together with the corresponding ontologies in an integrated manner, we present an approach based on the Model-driven Architecture (MDA). To this aim, the paper is organized as follows. In section 2 we briefly introduce MDA. We recall previous work on a model-driven approach for managing lightweight ontologies for industrial applications and provide an example ontology in section 3. In section 4, we extend the approach with a role-based access control model that uses the XACML standard (OASIS, 2005) for defining policies and show how it applies to our example. In section 5, we discuss our approach for enforcing security policies by rewriting SPARQL queries posed to the knowledge model and discuss the alignment to OWL 2 QL query rewriting. We conclude with a brief review of related work in section 6 and a discussion of the strength and weaknesses of our approach.

2 Model-driven Architecture

While the Web Ontology Language (OWL) is considered the representation standard for ontologies, the Extensible Access Control Markup Language (XACML) is the standard for representing access control policies. Instead of extending or changing those established standards we propose a loose integration of them. This allows for the reuse tools and infrastructures from both worlds while ensuring confidentiality for ontologies.

To this aim, we apply the principles of the Model-Driven Architecture (MDA), a software architecture proposed by the Object Management Group (OMG). Its basic idea is the focus on models as a primary result artifact of the development process. The MDA provides concrete means to create, exchange, and store such models and is grounded on a four-layered stack (cf. figure 1):

- Layer M3 - Meta-Metamodel: This layer defines a common, abstract language for the specification of metamodels.

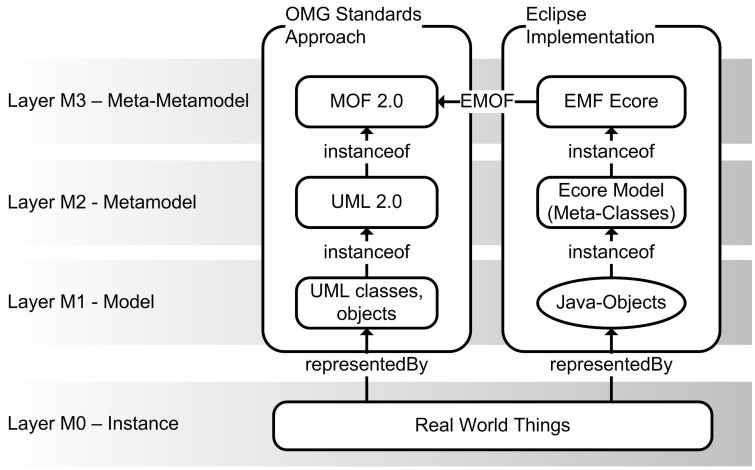

Figure 1: MDA Layers and Eclipse EMF Implementation

- Layer M2 - Metamodel: A metamodel defines the language capabilities for describing a concrete model.
- Layer M1 - Model: A model is a concrete instantiation of a metamodel.
- Layer M0 - Instance: This layer reflects the actual domain – i.e. the real world – that is described by the model.

This architecture is among others enabled by a number of standards hosted by the OMG. The Meta-Object Facility (MOF) serves as the standard language for defining metamodels – i.e. it represents layer M3. The Unified Modeling Language (UML) is a M2 metamodel that comes with a visual notation for its elements. It serves as a versatile visual modeling language for MDA. All models from layer M1 to M3 can be represented in XML using the XML Metadata Interchange (XMI) standard. XMI files are used within MDA to exchange models. Model-to-model transformations are covered by the MOF Query View Transformation (QVT) standard and allow for the transformation of models conforming to a metamodel to into an equivalent model conforming to another metamodel.

(Gasevic et al., 2006) have already discussed the relevance of the Model-driven Architecture for ontology development. Adopting the MDA for integrating ontologies and access control policies requires the development of M2 metamodels based on the MOF standard for OWL and XACML. Having the metamodels for both standards we can establish a loose integration of both formalisms by interconnecting the relevant elements at metamodel level. Thus, there is no need to address any syntactic representation issues in OWL or XACML.

When it comes to defining these metamodels and providing appropriate tool support we can adopt results from the Eclipse Modeling Project (EMP), which provides a number of frameworks to support MDA within the Eclipse platform. Eclipse[1] is an open-source community, which focuses on providing an extensible development platform (e.g., the Eclipse IDE for Java) and application frameworks (e.g., the Eclipse Rich Client Platform) for building software. The success of the Eclipse platform is among others a result of its highly modular and plugin-based architecture following the OSGI Service platform standard[2]. This allows

[1] http://www.eclipse.org
[2] http://www.osgi.org/Specifications/HomePage

for the parallel development of a number of additional frameworks and tools within the Eclipse ecosystem.

Among the EMP project, the Eclipse Modeling Framework (EMF) is a platform specific realization based of the relevant OMG standards to enable a model-driven approach. In order to keep the framework easy to handle, only the Essential MOF (EMOF) subset of the MOF 2.0 specification was considered for defining the meta-metamodel – the Ecore metamodel – of the EMF. Instances of Ecore are concrete metamodels such as a UML 2.0 metamodel. Note that the EMOF is powerful enough to cover the complete UML 2.0 specification of the OMG. Java objects are finally the M1-models. Figure 1 depicts these relationships.

Utilizing the EMF allows us to provide different tools to support our approach to integrate ontologies and access control policies in a loosely way. The model-driven approach allows for integrating both formalisms without the need to extend or modify standards that are defined an implemented by e.g. existing reasoners. We can thus focus on the actual challenge – the enforcement of access control policies in the presence of logical inferences.

3 Indutrial-Strength Ontologies

3.1 A Metamodel for Lightweight Ontologies

The Object Management Group (OMG) has recently finalized the standardization process of the Ontology Definition Metamodel (ODM, (OMG Adopted Specification, 2007)) - a metamodel for OWL ontologies based on the Meta Object Facitiliy (MOF), the standard for describing such models. The participation of big players of the branch such as IBM and AT&T and a first implementation by the Eclipse foundation, the EODM project, shows the high relevance of the ODM particularly to the industry. The coverage of OWL Full however results in a fairly complex metamodel. For OWL DL and Lite the ODM could be significantly simplified, since the use of RDF constructs are restricted in these subsets. Because OWL Full is not relevant to commercial applications due to its undecidability, we focus on more practical subsets of OWL.

To enable high performance reasoning even with large A-Boxes we propose the use of a subset of OWL Lite, which falls into plain Datalog. Such a subset, namely OWL Lite-, has been defined by (de Bruijn et al., 2004). The relevance of such a subset has been discussed by (Volz, 2004) and (Hitzler et al., 2005). In order to enable the direct representation of data from relational databases, we added the primitive data types. The resulting subset, OWL Lite$^{-,p}$, as well as a metamodel have been presented in more details in (Chen & Stuckenschmidt, 2008). Table 1 gives an overview of the included language elements and also presents some mappings to other knowledge representation languages.

To actually utilize the metamodel for implementing an application for ontology development, we build on the Eclipse Modeling Framework (EMF), a project aiming at providing Eclipse-based tools for model-driven development. Among others, we can utilize the framework to automatically generate a Java object representation of metamodels defined in the Ecore - a meta-metamodel language, which is a real subset of MOF. Figure 2 shows the Ecore metamodel for OWL Lite$^{-,p}$.

3.2 An Example

Consider a product ontology, which integrates different information about products and their parts that may exist in different departments of a car manufacturer. Clearly, some information from e.g. sales such as the purchase price of a part should not be visible for developers. Also, external staff that works on specific parts should only see data which they require for their work. For example, somebody who is working on the braking system should only be able to see brake parts and specific electronic control units such as the Anti-Lock Braking System.

OWL Abstract Syntax	DL Syntax	F-Logic Syntax	Datalog Syntax
`restriction(`R `allValuesFrom(`C`))`*	$\forall R.C$	see partial class definitions	
`Class(`A `partial` $C_1 \ldots C_n$`)`	$A \sqsubseteq C_i$	$\bigwedge^1 A :: C_i$ $\bigwedge^2 x_1 : A \wedge x_1[R_i \twoheadrightarrow x_2] \rightarrow x_2 : C_i$	$\bigwedge^1 A(x) \rightarrow C_i(x)$ $\bigwedge^2 A(x_1) \wedge R(x_1,x_2) \rightarrow C_i(x_2)$
`Class(`A `complete` $C_1 \ldots C_n$`)`	$A \equiv C_1 \sqcap \ldots \sqcap C_n$	$\bigwedge \left\{ \begin{array}{l} A :: C_i \\ C_i :: A \end{array} \right.$	$\bigwedge \left\{ \begin{array}{l} A(x) \rightarrow C_i(x) \\ C_i(x) \rightarrow A(x) \end{array} \right.$
`EquivalentClasses(`$C_1 \ldots C_n$`)`	$C_1 \equiv \ldots \equiv C_n$	$\bigwedge_{i \neq j} C_i :: C_j$	$\bigwedge_{i \neq j} C_i(x) \rightarrow C_j(x)$
`ObjectProperty(`R			
`super(`R_1`) ... super(`R_n`)`	$R \sqsubseteq R_i$	$\bigwedge x[R \twoheadrightarrow y] \rightarrow x[R_i \twoheadrightarrow y]$	$\bigwedge R(x,y) \rightarrow R_i(x,y)$
`domain(`C_1`) ... domain(`C_n`)`	$\top \sqsubseteq \forall R^-.C_i$	$\bigwedge x[R \twoheadrightarrow y] \rightarrow x : C_i$	$\bigwedge R(x,y) \rightarrow C_i(x)$
`range(`C_1`) ... range(`C_n`)`	$\top \sqsubseteq \forall R.C_i$	$\bigwedge x[R \twoheadrightarrow y] \rightarrow y : C_i$	$\bigwedge R(x,y) \rightarrow C_i(y)$
`[inverseOf(`R_0`)]`	$R \equiv R_0^-$	$\bigwedge \left\{ \begin{array}{l} x[R \twoheadrightarrow y] \rightarrow y[R_0 \twoheadrightarrow x] \\ x[R_0 \twoheadrightarrow y] \rightarrow y[R \twoheadrightarrow x] \end{array} \right.$	$\bigwedge \left\{ \begin{array}{l} R(x,y) \rightarrow R_0(x,y) \\ R_0(x,y) \rightarrow R(x,y) \end{array} \right.$
`[Symmetric]`	$R \equiv R^-$	$x[R \twoheadrightarrow y] \rightarrow y[R \twoheadrightarrow x]$	$R(x,y) \rightarrow R(y,x)$
`[Transitive])`	`Trans(`R`)`	$x[R \twoheadrightarrow y] \wedge y[R \twoheadrightarrow z] \rightarrow x[R \twoheadrightarrow z]$	$R(x,y) \wedge R(y,z) \rightarrow R(x,z)$
`SubPropertyOf(`R_1 R_2`)`	$R_1 \sqsubseteq R_2$	$x[R_1 \twoheadrightarrow y] \rightarrow x[R_2 \twoheadrightarrow y]$	$R_1(x,y) \rightarrow R_2(x,y)$
`EquivalentProperties(`$R_1 \ldots R_n$`)`	$R_1 \equiv \ldots \equiv R_n$	$\bigwedge_{i \neq j} x[R_i \twoheadrightarrow y] \rightarrow x[R_j \twoheadrightarrow y]$	$\bigwedge_{i \neq j} R_i(x,y) \rightarrow R_j(x,y)$
`DatatypeProperty(`U			
`domain(`C_1`) ... domain(`C_n`)`	$\top \sqsubseteq \forall U^-.C_i$	$\bigwedge x[U \twoheadrightarrow y] \rightarrow x : C_i$	(not supported)
`range(`D`))`**	$\top \sqsubseteq U.D$	$C_i[U \Rightarrow D]$.	(not supported)
`Individual(`o `type(`C_1`) ... type(`C_n`)`	$o \in C_i$	$\bigwedge o : C_i$	$\bigwedge C_i(o)$
`value(`R_1 o_1`) ... value(`R_n o_n`)`	$\langle o, o_i \rangle \in R_i$	$\bigwedge o[R_i \twoheadrightarrow o_i]$	$\bigwedge R_i(o, o_i)$
`value(`U_1 v_1`) ... value(`U_n v_n`)`	$\langle o, v_i \rangle \in U_i$	$\bigwedge o[U_i \twoheadrightarrow v_i]$	(not supported)

* only allowed in partial class definitions
** only primitive datatypes allowed
[1]: for named classes
[2]: for universal class restriction

Table 1: Features included in OWL Lite$^{-,p}$

The example ontology is defined using the abstract OWL syntax:

```
Class(Manufacturer partial)
Class(Supplier partial Manufacturer
   restriction(produces allValuesFrom Part))
Class(Product partial)
Class(Car partial Product)
Class(Part partial Product)
Class(BodyPart partial Part)
Class(PrototypeBodyPart partial BodyPart)
Class(Brake partial Part)
Class(ElectronicControlUnit partial Part)
Class(AntiLockBrakingSystem partial ElectroniControlUnit)
Class(TransmissionControlUnit partial ElectronicControlUnit)
Class(TCU partial ElectroniControlUnit)
EquivalentClasses(TCU, TransmissionControlUnit)
ObjectProperty(produces domain(Manufacturer) range(Product))
ObjectProperty(contains domain(Car) range(Part))
DatatypeProperty(hasSalesPrice domain(Product) range(float))
DatatypeProperty(hasPurchasePrice domain(Product) range(float))
```

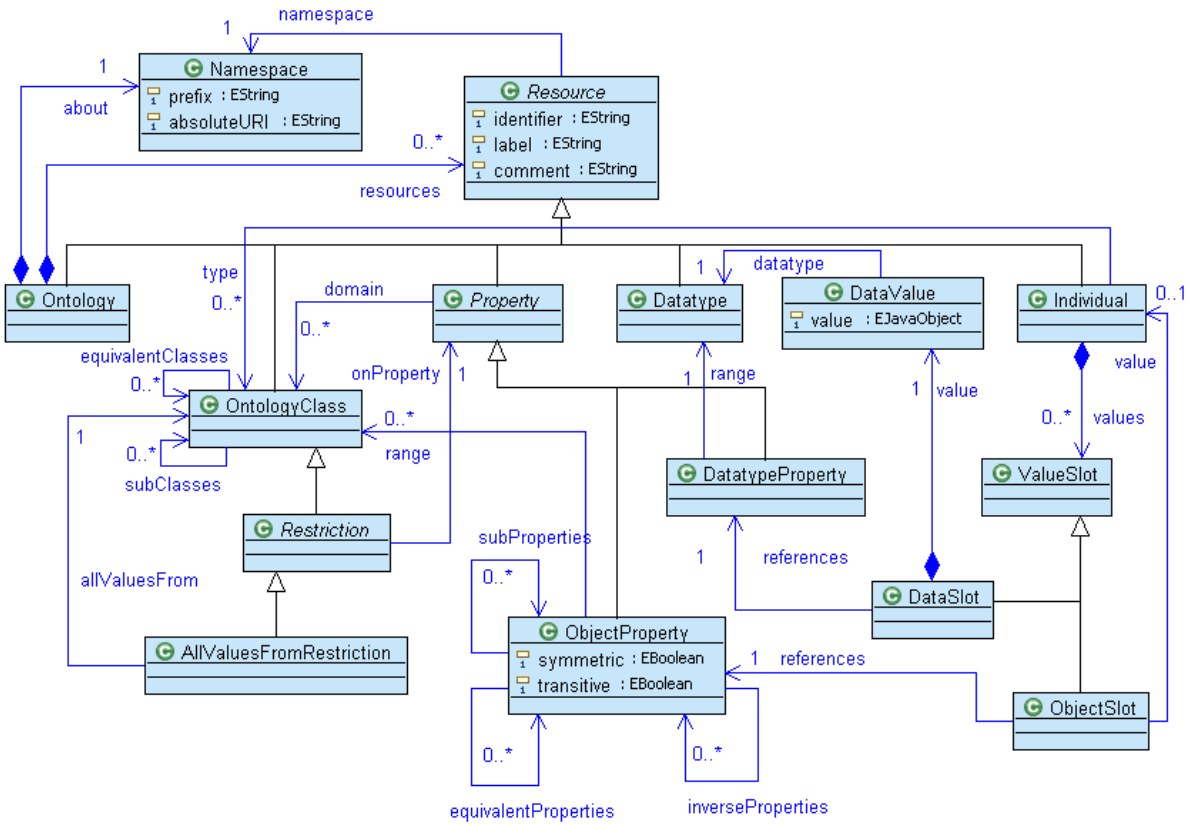

Figure 2: Lightweight Ontology Metamodel

4 The Security Framework

4.1 Access Control and Policy Specification

Access control systems enable the regulation of access to protected resources (i.e. objects) in distributed systems by subjects such as users or system processes (cf. (Eckert, 2003)). They can be categorized in discretionary access control (DAC), mandatory access control (MAC), and role-based access control (RBAC) models. In DAC-based systems, the permissions to access an object are defined by its owner. In MAC models, the system determines the access to objects either by utilizing access rules or lattices for assigning permissions to subjects. It thus removes the ability of the users to control access to their resources. RBAC systems finally remove the explicit use of subjects within access rules or lattices and replace them with roles, which form a logical group of a number of subjects. In fact, permissions are assigned to roles and the subjects are assigned members of a number of roles. Thus changes of single subjects do not necessarily have consequences in the actual access control policies.

The Extensible Access Control Markup Language (XACML) is an OASIS standard that describes both a policy language and an access control decision request/response language (both written in XML) (OASIS, 2003). The policy language consists of following basic constructs:

- PolicySet: A PolicySet is a container for a number of Policies or other PolicySets.

- Policy: A policy represents a single access control policy, expressed by a set of Rules. It applies for a defined Target.
- Rule: A Rule consists of a Target element and has an attribute Effect, which can be either permit or deny.
- Target: The Target element declares a set of Subjects, Resources, and Actions for which a Rule or Policy applies.

Note that the actual XACML specification defines a number of additional constructs to further refine the access control policies. Altogether XACML is an extensible standard that allows enterprises to build fine-grained access control systems.

4.2 Protecting Resources in Ontologies

In order to apply access control mechanisms to ontologies, we need to identify the resources, which should be protected. Since the T-Box of ontologies in generally represents a shared vocabulary of a specific domain, it is not reasonable to restrict access at this level. In fact, we aim at controlling the access to instances of specific classes and properties i.e. the A-Box, while the T-Box remains unrestricted. We thereby assume that the access to individuals of all classes and all property values are denied by default for avoiding unintended security leaks. Because of this, access to specific classes and thus corresponding individuals as well as property values have to be explicitly granted to subjects. For this purpose, we consider following basic access rule types:

1. permit access to all individuals and values respectively of a class, a datatype property, or an object property
2. deny access to a specific individual
3. deny the access to individuals and values of subclasses and subproperties of prior permitted resources

As common in modern access control systems, we envision a RBAC model for simplifying the maintenance of the access control policies.

4.3 Extending the Metamodel

In order to enable a uniform and interdependent maintenance of the protected ontologies together with the security policies, we extend the earlier presented metamodel of lightweight ontologies by means to model access control policies. Thus, we do not need to commit to a single representation formalism for either ontologies or security policies. We initially cover a subset of the XACML specification, which provides basic features for realizing the access control mechanism for ontologies as presented above (cf. Figure 3). We thereby leave out the PolicySet element and only consider simple restrictions on ontology resources. The use of the Target construct allows grouping the Policies by Subjects, Resources, or any combination of both. If no Action, Subject, or Resource is defined for a Target, we consider that the Policy or Rule applies for all Actions, Subjects, and Resources respectively.

4.4 Security Architecture

We can utilize existing reasoners for OWL Lite to infer implicit knowledge and query an ontology. This is typically done by sending a SPARQL select query to the appropriate endpoint of the used reasoner. In order

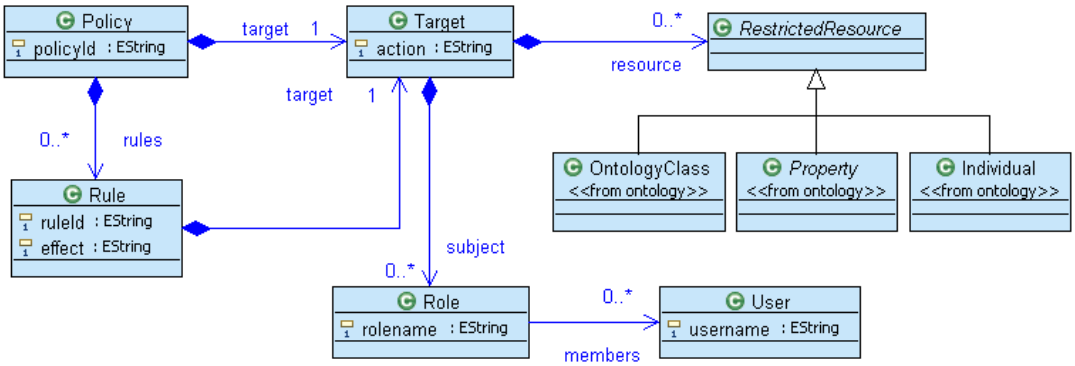

Figure 3: Access Control Policies Metamodell

to ensure confidentiality in such an infrastructure, we enforce the use of a security proxy, which intercepts all communication to the reasoner. Similar to web proxy systems, we require a user authentication prior to its first use. The proxy acts as the Policy Enforcement Point (PEP) within the XACML architecture and constructs an authorization decision query to the Policy Decision Point (PDP). The decision of the PDP is based on the access control policies provided by the Policy Retrieval Point (PRP) and the information provided by the Policy Information Point (PIP) about for example the role of a user. The PEP finally enforces the decision of the PDP either by completely denying the query, by rewriting the query so it meets the security policies, or by directly redirecting the query to the reasoner. Details of this part of the process are discussed in section 4 below. The resulting security architecture (see figure 4) ensures the enforcement of access control policies in a loosely coupled way by utilizing existing frameworks and tools without the need to modify existing reasoners, so they can directly support such a task.

Within this work we assume that the security proxy unifies the different components (PxP) defined by the XACML architecture so there is no need to exchange messages between them. In future work, we consider splitting the single parts into separated components.

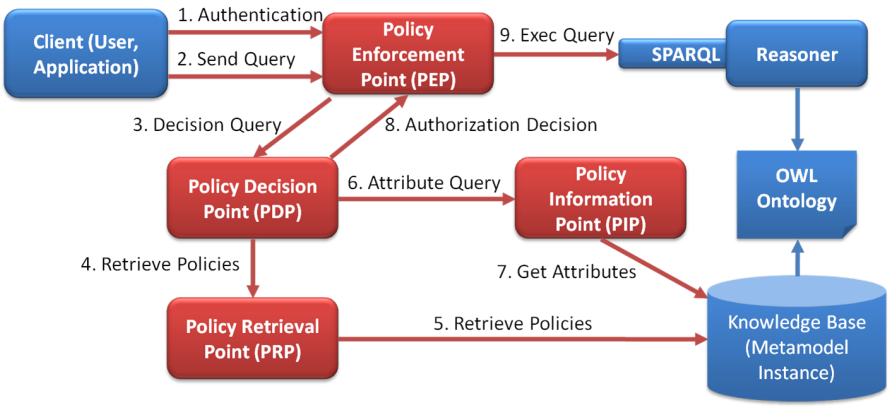

Figure 4: Security Architecture

4.5 Example

An external developer works on the braking system of a car and thus should be able to access the required product information, i.e. information about Brakes and AntiLockBrakingSystems (ABS). Nevertheless, a specific prototype ABS-system should not be visible for him. We define following policy for a role ExternalBrakeDeveloper:

```
Policy P1 = subject="ExternalBrakeDeveloper", rules=R1, R2
Rule R1 = "read", "permit", Class(Brake), Class(AntiLockBrakingSystem)
Rule R2 = "read", "deny", Individual(ABS1)
```

For employees working in the sales department, we define similar rules, but explicitly permit the access to the datatype property ”hasPurchasePrice”:

```
Policy P2 = subject="Sales", rules=R3, R4, R5
Rule R3 = "read", "permit", Class(Product)
Rule R4 = "read", "deny", Individual(ABS2)
Rule R5 = "read", "permit", DatatypeProperty(hasPurchasePrice)
```

5 Enforcing Access Policies

As outlined above, security policies are enforced at a security proxy that receives queries to the ontology and rewrites them in such a way that no policies are violated considering the presence of logical inferences. In this section, we provide details about this rewriting step. The goal of the rewriting step is to create a query that returns the same results as the original query except for those answers that contain restricted information. For this purpose, the proxy first determines the access restrictions that apply for the specific query by determining the role of the issuing user and looking up the restrictions that apply to this role in the policy rules. As defined in the security framework, such rules can apply to classes, properties and instances. In this paper, we further focus on read restrictions on these elements.

5.1 Preliminaries

Select queries in SPARQL typically contain a set of triple patterns, where subject, object and predicate may all be variables. However, for rewriting queries to meet the defined access restrictions it is essential to be able to determine, if a variable will hold an individual or a data value. This can be achieved only if we assume that the predicate position within a triple is not a variable. Knowing the predicate, we can easily determine what kind of resource occurs for the subject and the object position. We refer to $?v_k$ as variables for individuals and $?d_k$ for data values. Temporary variables $?t_k$ are introduced to expand the query with respect to enforce the given policies. When combining several filters we ensure that the variable names remain unique.

5.2 Filtering Query Results

A read restriction on individuals $i_1, \ldots.., i_n$ is interpreted in the way that the user is not allowed to see answers that contain this specific individual. This condition can easily be fulfilled by extending the query with the following filter condition on each of the return variables $?v_k$:

$$\text{FILTER } (?v_k \; != ex:i_1 \; \&\& \; \ldots \; \&\& \; ?v_k \; != i_n) \tag{1}$$

This condition removes all results, where restricted individuals are bound to any of the return variables.

We interpret read restriction on a class C in such a way that the user is not allowed to see answers that contain instances of the restricted class. Therefore not allowing the access to a class is equivalent to restricting the access to any individual of the class. Matters are complicated by the fact that class membership can be inferred using the ontology. Therefore, it is not possible to simply add filter restrictions for all instances that are defined to be member of restricted classes. We can however, delegate this problem to the underlying inference engine by using a filter expression on the schema level:

$$\begin{aligned} &\text{OPTIONAL } \{?v_k \text{ rdf:type } ?t_1.\ \text{FILTER}(?t_1 = C)\ \} \\ &\qquad\qquad \text{FILTER}(!\text{bound } (?t_1)) \end{aligned} \tag{2}$$

This expression checks for each return variable, whether the instance bound to the variable is of type C. If this is true, the corresponding answer is filtered out based on the criterion that variable $?t_1$ is bound to the restricted class.

Read permission on classes can be enforced in a similar manner by adding a filter expression, which ensures that the queried instances are at least member of one of the permitted classes C_i:

$$\begin{aligned} &?v_k \text{ rdf:type } ?t_1. \\ &\text{FILTER}(?t_1 = C_1 || \ldots || ?t_1 = C_n) \end{aligned} \tag{3}$$

We further interpret restrictions on a relation R in the way that the user is not allowed to see answers in which two of the return variables are bound to instances that are in relation R to each other, because queries might be formulated in such a way, that they model the restricted relation without actually mentioning it. We again use a filter expression to filter out the corresponding answers from the result set.

$$\begin{aligned} &\text{OPTIONAL } \{?v_k\ ?t_1\ ?v_j.\ \text{FILTER}(?t_1 = R)\} \\ &\qquad\qquad \text{FILTER}(!\text{bound } (?t_1)) \end{aligned} \tag{4}$$

Here $?v_k \neq ?v_j$ are return variables of the query. The expression checks whether the existence of relation R between $?v_k$ and $?v_j$ can be derived from the ontology. If this is the case, the corresponding answer is filtered out based on the criterion that the variable $?t_1$ is bound to R. This solution again delegates the problem of dealing with derivable information to the underlying inference engine which is used to check whether $(?v_k\ ?t_1\ ?v_j)$ can be derived from other information such as inverse relations, etc. Clearly, $?v_k$ or $?v_j$ may also be concrete resources. Similar to explicitly granting read access to individuals of a specific class, we can add a filter for permitting access to property values R_i:

$$\begin{aligned} &?v_k\ ?t_1\ ?v_j. \\ &\text{FILTER}(?t_1 = R_1 || \ldots || ?t_1 = R_n) \end{aligned} \tag{5}$$

5.3 Controlling the Rewriting Process

In section 5.2 we have presented the concrete filter expressions that can be used to ensure that no restricted information is returned by a query. This section now deals with linking the manipulation of the queries with the policy specifications in order to control the rewriting process. Following the assumption from 4.2, we can formulate a simple rewriting strategy that first extends the query with filters for those individuals explicitly restricted for the role under consideration and then also add filters for concepts and relations access to which is not explicitly permitted. The corresponding simple algorithm is given below:

```
ReWrite(Q:Query, R:Role, P:Policy)
```

```
FORALL i, read, R, deny ∈ P DO AddDenyFilter(Q,i)
FORALL C, read, R, permit ∈ P DO AddPermitFilter(Q,C)
FORALL C, read, R, deny ∈ P DO AddDenyFilter(Q,C)
FORALL P, read, R, permit ∈ P DO AddPermitFilter(Q,P)
FORALL P, read, R, deny ∈ P DO AddDenyFilter(Q,P)
IF C, read, R, permit ∉ P DO DenyAccess(Q)
IF P, read, R, permit ∉ P DO DenyAccess(Q)
```

Basically the rewriting algorithm, if any permit policy exists, will add additional graph patterns to the original SPARQL query depending on the number of used variables used and the number of policies. This can lead to complexity issues when evaluating the rewritten query. As shown by (Perez et al., 2009) the complexity of the evaluation of SPARQL queries is "PSPACE-complete for graph pattern expressions constructed by using only AND, FILTER and OPT operators". Optimizations however can be applied by more intelligent checks, whether a given rule is applicable to a variable in a SPARQL query. Given an query `?s rdf:type ex:Part` and a rule permitting or denying the access to a class C, we only add the appropriate filter if `ex:Part` $\sqsubseteq C$ or $C \sqsubseteq$ `ex:Part`. Similar subsumption checks could be applied for rules restricting the access to properties. The required subsumption hierarchy can clearly be computed and cached before any rewriting is produced. Thus the number of additional graph patterns could be reduced to a minimum.

A proof-of-concept implementation of the algorithm described above can be downloaded at http://www.chenwilly.de/wi.html. It shows how a SPARQL query can be analyzed and rewritten in an automated manner based on the Jena semantic Web framwork[3].

5.4 Example

Consider the example ontology from 3.2 and the policy P_1 from 4.5, consisting of 2 rules R_1 and R_2:

```
Rule R1 = "read", "permit",
    Class(Brake), Class(AntiLockBrakingSystem)
Rule R2 = "read", "deny", Individual(ABS1)
```

If we pose a query for all parts the SPARQL query would be:

```
SELECT ?s WHERE {
   ?s rdf:type ex:Part
   }.
```

For rule R_1 the rewriting rule 3 applies and results in following rewritten query:

```
SELECT ?s WHERE {
   ?s rdf:type ex:Part.
   ?s rdf:type ?t1.
   FILTER(?t1 = ex:Brake || ?t1 = ex:AntiLockBrakingSystem)
   }.
```

By applying the rewriting rule 1 the rule R_2 is enforced:

```
SELECT ?s WHERE {
   ?s rdf:type ex:Part.
   ?s rdf:type ?t1.
```

[3] http://jena.sourceforge.net/

```
FILTER(?t1 = ex:Brake || ?t1 = ex:AntiLockBrakingSystem)
FILTER(?s != ex:ABS1)
}.
```

5.5 Verifying Anomalies in Access Policies

During the creation and maintenance anomalies such as contradictory rules could occur in access policies. Since we use filters to enforce the policies, a deny rule always overrides a permit rule on the same resource. Thus, inconsistent access policies would not result in unintended information leaks. To ensure the consistency of access policies, we additionally provide an initial set of verification rules that can be applied at the metamodel level on rules that apply for the same subject:

```
IF EXISTS R1 = "read", "permit", RestrictedResource(R),
    R2 = "read", "deny", RestrictedResource (R) THEN Anomaly(R1, R2)
IF EXISTS R1 = "read", "permit", OntologyClass(C1),
    R2 = "read", "deny", OntologyClass(C2) AND EquivalentClasses(C1,C2)
    THEN Anomaly(R1, R2)
IF EXISTS R1 = "read", "permit", ObjectProperty(P1),
    R2 = "read", "deny", ObjectProperty(P2) AND
    (EquivalentProperties(P1,P2) OR ObjectProperty(P1 inverseOf(P2)))
    THEN Anomaly(R1, R2)
IF EXISTS R1 = "read", "permit", ObjectProperty(P1),
    R2 = "read", "deny", OntologyClass(C1) AND
    (ObjectProperty(P1 domain(C1)) OR ObjectProperty(P1 range(C1)))
    THEN Anomaly(R1, R2)
IF EXISTS R1 = "read", "permit", DatatypeProperty(P1),
    R2 = "read", "deny", OntologyClass(C1) AND
    DatatypeProperty (P1 domain(C1))
    THEN Anomaly(R1, R2)
```

5.6 Query Rewriting in OWL 2 QL

Within the recent OWL 2 standard proposal (Hitzler et al., 2009) the requirements for query answering with large A-Boxes have been taken into account in a specialized profile, namely OWL 2 QL. It is is based on DL-Lite$_R$ (cf. (Calvanese et al., 2007)) and allows for sound and complete conjunctive query answering in $\mathcal{LOGSPACE}$ with respect to the size of instance data. Technically, the query answering can be implemented by rewriting ontology queries, e.g. in SPARQL, to SQL queries that can be posed to standard relational database systems. Two different rewriting algorithms exist:

- PerfectRef (cf. (Calvanese et al., 2007)): The algorithm transforms a conjunctive query Q and an ontology O into a union of conjunctive queries Q_O that can be answered over the ABox of an ontology only. Table 2 shows the reformulation rules $\mathcal{R}$ that are used for rewriting queries. PerfectRef is defined as follows:

```
PerfectRef(Q, O) {
    // the result contains the original query
    QO = {Q};
    repeat {
        Q'O = QO;
```

```
        forall Q' ∈ Q'_O do {
            forall body atom D in Q' and ontology axiom α ∈ O do {
            // if a reformulation rule is applicable a query is generated
            // the function replace substitutes D with ref(D,α) in Q'
            if (D,α) ∈ R { Q_O = Q_O ∪ replace(Q',D,ref(D,α)); }
        }
        forall D1 and D2 in Q' do {
            // the function reduce applies the most general unifier
            // between D1 and D2 to Q' and replaces unbound
            // variables with the symbol _
            if D1 and D2 unify { Q_O = Q_O ∪ reduce(Q',D1,D2); }
        }
    }
    // no more reformulation could be applied
    } until Q'_O = Q_O;
    return Q_O;
}
```

If the ABox is stored in a relational database management system (RDBMS) the resulting queries Q_O could be executed as a union of SQL queries.

D	α	$ref(D,\alpha)$
$A(x)$	$B \sqsubseteq A$	$B(x)$
	$\exists P \sqsubseteq A$	$P(x,\_)$
	$\exists P^- \sqsubseteq A$	$P(\_,x)$
$P(x,\_)$	$A \sqsubseteq \exists P$	$A(x)$
	$\exists S \sqsubseteq \exists P$	$S(x,\_)$
	$\exists S^- \sqsubseteq \exists P$	$S(\_,x)$
$P(\_,x)$	$A \sqsubseteq P^-$	$A(x)$
	$\exists S \sqsubseteq \exists P^-$	$S(x,\_)$
	$\exists S^- \sqsubseteq \exists P^-$	$S(\_,x)$
$P(x,y)$	$S \sqsubseteq P$	$S(x,y)$
	$S^- \sqsubseteq P^-$	$S(x,y)$
	$S \sqsubseteq P^-$	$S(y,x)$
	$S^- \sqsubseteq P$	$S(y,x)$

Table 2: PerfectRef Query Reformulation Rules $\mathcal{R}$

- RQR (Resolution-based Query Rewriting, cf. (Pérez-Urbina et al., 2009a), (Pérez-Urbina et al., 2009b)): Similar to the PerfectRef algorithm the result of RQR is to produce a rewritten query over the A-Box of an ontology. It uses a resolution based calculus and aims at creating smaller rewritings compared to PerfectRef.

Existing implementations of OWL 2 QL reasoners and rewriting algorithms, respectively, include:

- owlgres[4]: owlgres has been an early implementation of OWL 2 QL, which relies on PostgreSQL as underlying relational database management system. It utilizes the PerfectRef algorithm and is still in alpha phase.
- ROWLKit, QuOnto[5]: The ROWLKit is a simple GUI, which uses QuOnto as reasoner for OWL 2 QL ontologies. It uses an embedded H2 (http://www.h2database.com/html/main.html) database and implements the PerfectRef algorithm.
- REQUIEM[6]: REQUIEM is a prototypical implementation of the RQR rewriting algorithm.
- Quill[7]: Quill is a OWL 2 QL system, which is embedded in the ONTOSEARCH2[8] platform. It implements the PerfectRef algorithm.

During the development of the OWL 2 standard proposal, the need for a metamodel has been raised[9]. An OWL 2 metamodel has been defined, which reconfirms our model-driven approach for enabling access control for ontologies. However, it is currently only available in OWL-format[10].

5.7 Embedding Access Control into PerfectRef

Obviously, the rewriting step for OWL 2 QL and for enforcing access control polices could be integrated. I.e. while expanding the query we can add policy checks to apply defined access control policies. We briefly sketch the possible steps based on an example query $Q = A(x)$ and the reformulation rule $R(D, \alpha, ref(D, \alpha)) = \{A(x), B \sqsubseteq A, B(x))\}$:

- Before applying any reformulation rules, we need to evaluate, whether there exists a rule that permits access to A or C, $A \sqsubseteq C$. If no permit rules exist, we remove the query and PerfectRef terminates.
- During the following reformulation steps, the reformulation rule is only applicable, if there exists no access control rules that denies access to B. Otherwise the rewriting algorithm terminates.
- Finally, the set of individuals, which we want to deny access to, can be passed as a list that must not be included in the query result.

Clearly, the actual idea of transparently add the enforcement of access control policies to a given ontology infrastructure can no longer be pursued in such a setting. In fact, we need to modify the actual rewriting algorithms of OWL 2 QL. Thus, we may also consider including access control mechanisms in RDBMS for a combined enforcement of access control policies in order to reduce the query complexity. The study of the integration of access policy enforcement within OWL 2 QL rewriting algorithms, not only in PerfectRef but also in RQR, is part of future research. In particular, complexity issues and the size of the rewritten queries need to be analyzed and compared in more details.

[4] http://pellet.owldl.com/owlgres/
[5] http://www.dis.uniroma1.it/quonto/
[6] http://www.comlab.ox.ac.uk/projects/requiem/home.html
[7] http://kt.abdn.ac.uk/wiki/Projects/Quill
[8] http://www.ontosearch.org/
[9] http://www.w3.org/2007/OWL/wiki/Metamodel_Proposal
[10] http://code.google.com/p/owl2/

6 Related Work

Recently, there has been quite a bit of work on combining semantic web technologies with security issues. Most of this work, however, is about applying semantic web technologies to the problem of specifying and checking security policies and related aspects (e.g. (Kolovski et al., 2007)). The aspect we are interested in, the protection of knowledge encoded in an ontology has so far received less attention. (Qin & Atluri, 2003) have presented an approach to enable access control at concept level for ontologies. Their approach is based on a propagation mechanism that derives access rights for concepts based on the semantic of the ontology and partial access rights explicitly stated. A similar approach is envisioned by (Knechtel, 2008) who proposes to derive the access rights of derived facts from the access rights based on access restrictions imposed on stated knowledge. In contrast to this our approach relies on the re-writing on queries based on security policies. This allows us to leave the task of optimizing the execution of the query to the query engine.

The inference problem, which is central to our work, has been investigated by researchers in the field of database research (Farkas & Jajodia, 2002). The problem has been investigated in the context of statistical databases, where statistical inference can be used to derive classified information from unclassified ones (Mandujano, 2000). Special attention has been paid to the inference problem in the context of multilevel secure databases. Here it also has to be ensured that access policies are not violated between the different security layers. In our work, we have so far ignored such problems arising in multilevel security policies. Most of the work on secure databases has addressed the relational model and problems that arise from the use of database constraints. The problem of enforcing security constraints in deductive databases, a problem that is closely related to our work has so far only been addressed on a theoretical level with little concerns on implementability (Bonatti et al., 1995).

7 Conclusion

In this paper we have presented an approach to restrict the access to confidential knowledge that is represented using ontologies, which we consider an important feature for real world applications. We thereby fully build on existing technologies and infrastructures. The access control system relies on a security proxy, which intercepts all communication to the SPARQL endpoint of any utilized OWL reasoner. When enforcing fine-grained role-based access policies we need to take possible inferences from the ontology into account. Instead of realizing this complex issue within the security proxy, we delegate this task to the reasoner by adding SPARQL filter statements to incoming queries from users. This allows us to apply our solution with all reasoners supporting OWL Lite and SPARQL. Since the access policies and rules are highly dependent on the ontologies they have been defined for, we utilize a model-driven approach to enable their interdependent development and maintenance. Moreover, we can transform the created models to different representation formalisms. For ontologies, we can directly support OWL, F-Logic (Kifer et al., 1995), and Datalog, while we may rely on the XACML specification for access policies.

Future work includes more detailed analysis on secure query rewriting with the OWL 2 QL profile. We are particularly interested in a combined used of access control mechanisms in relational database management system and policy enforcement via query rewriting for providing a scalable access control mechanism. Also, we intend to research the support of access control for ontologies combined with logical rules. Hierarchical access control models will also be part of future research.

References

Bonatti, P. A., Kraus, S., & Subrahmanian, V. S. (1995). Foundations of secure deductive databases. *IEEE Transactions on Knowledge and Data Engineering*, 7(3), 406–422.

Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., & Rosati, R. (2007). Tractable reasoning and efficient query answering in description logics: The DL-Lite family. *Journal of Automated Reasoning*, 39(3), 385–429.

Chen, W. & Stuckenschmidt, H. (2008). Towards industrial strength knowledge bases for product lifecycle management. In *16th European Conference on Information Systems, Galway, Ireland*.

Chen, W., Syldatke, T., & Hess, C. (2008). Business-oriented CAx-Integration with Semantic Technologies. In *3rd International Applications of Semantic Technologies Workshop*.

Davis, M. (2008). Semantic wave 2008 report - industry roadmap to web 3.0 & multibillion dollar market opportunities. Executive Summary.

de Bruijn, J., Polleres, A., & Fensel, D. (2004). OWL Lite$^-$. WSML Deliverable D20v0.1.

Eckert, C. (2003). *IT-Sicherheit*. Oldenburg.

Farkas, C. & Jajodia, S. (2002). The inference problem: a survey. *SIGKDD Explor. Newsl., ACM*, 4, 6–11.

Fensel, D. (2001). *Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce*. Springer Verlag.

Gasevic, D., Djuric, D., & Devedic, V. (2006). *Model Driven Architecture and Ontology Development*. Springer.

Hepp, M. (2007). Possible ontologies: How reality constrains the development of relevant ontologies. *IEEE Internet Computing*, 11(1), 90–96.

Hitzler, P., Haase, P., Krötzsch, M., Sure, Y., & Studer, R. (2005). DLP is not so bad after all. In *OWL: Experiences and Directions*.

Hitzler, P., Krtzsch, M., Parsia, B., Patel-Schneider, P. F., & Rudolph, S. (2009). OWL 2 web ontology language primer. http://www.w3.org/TR/2009/REC-owl2-primer-20091027/.

Kifer, M., Lausen, G., & Wu, J. (1995). Logical foundations of object-oriented and frame-based languages. *JACM: Journal of the ACM*, 42, 741–843.

Knechtel, M. (2008). Access rights and collaborative ontology integration for reuse across security domains. Poster, ESWC 2008 PhD Symposium.

Kolovski, V., Hendler, J., & Parsia, B. (2007). Analyzing web access control policies. In *Proceedings of the Sixteenth International World Wide Web Conference*.

Mandujano, S. (2000). Inference attacks to statistical databases: Data suppression, concealing controls and other security trends.

OASIS (2003). A Brief Introduction to XACML. http://www.oasis-open.org/committees/download.php/2713/Brief_Introduction_to_XACML.html.

OASIS (2005). eXtensible Access Control Markup Language (XACML) Version 2.0. OASIS Standard.

OMG Adopted Specification (2007). Ontology definition metamodel. http://www.omg.org/docs/ptc/07-09-09.pdf.

Perez, J., Arenas, M., & Gutierrez, C. (2009). Semantics and complexity of sparql. *ACM Trans. Database Syst.*, 34, 45.

Pérez-Urbina, H., Horrocks, I., & Motik, B. (2009a). Efficient Query Answering for OWL 2. In A. Bernstein, D. R. Karger, T. Heath, L. Feigenbaum, D. Maynard, E. Motta, & K. Thirunarayan (Eds.), *Proc. of the 8th Int. Semantic Web Conference (ISWC 2009)*, volume 5823 of *LNCS* (pp. 489–504). Chantilly, VA, USA: Springer.

Pérez-Urbina, H., Horrocks, I., & Motik, B. (2009b). Practical Aspects of Query Rewriting for OWL 2. In R. Hoekstra & P. Patel-Schneider (Eds.), *Proc. of the 5th Int. Workshop on OWL: Experiences and Directions (OWLED 2009)* Chantilly, VA, USA.

Qin, L. & Atluri, V. (2003). Concept-level access control for the semantic web. In *Proceedings of the 2003 ACM workshop on XML security*.

Staab, S. (2002). Wissensmanagement mit ontologien und metadaten. Habilitation Thesis, University of Karlsruhe.

Syldatke, T., Chen, W., Angele, J., Nierlich, A., & Ullrich, M. (2007). How ontologies and rules help to advance automobile development. In A. Paschke & Y. Biletskiy (Eds.), *Advances in Rule Interchange and Applications. International Symposium, RuleML 2007. Orlando, Florida, October 2007. Proceedings*, volume 4824/2007 of *Lecture Notes in Computer Science* (pp. 1–6).: Springer Berlin / Heidelberg.

Volz, R. (2004). *Web Ontology Reasoning with Logic Databases*. PhD thesis, Universität Fridericiana zu Karlsruhe.