Chapter 8
Access Control on Semantic Web Data Using Query Rewriting

Jian Li
Hong Kong Baptist University, China

William K. Cheung
Hong Kong Baptist University, China

ABSTRACT

Semantic Web technologies allow on-line resources to be semantically annotated to support more effective and intelligent online services. However, ontologies sometimes may contain sensitive information. Providing access to them requires proper control to ensure the data protection requirement. Yet, the protection should not be too restrictive to make the access management inflexible. While there has been recent work on policy-based access control, in this paper, the authors present a policy representation specifically for access control on ontology-based data and explain how issues like policy propagation and policy conflict resolution are addressed. The authors present bucket-based query rewriting algorithms for realizing the access control policies to avoid sensitive resources leakage in the context of the Semantic Web. The authors validate the correctness of the proposed mechanisms by going through some illustrative examples in detail.

1 INTRODUCTION

The recent proliferation of the Semantic Web technologies has resulted in an increasing amount of on-line resources being semantically annotated to support more effective and intelligent online services, including customized search, autonomous information integration, etc. We here call such kind of semantically annotations of resources *semantic web data*, where domain ontologies are needed as some background knowledge for support semantic interpretation of various types. In the literature, projects on developing semantic web data management

DOI: 10.4018/978-1-4666-1577-9.ch008

systems are abundant (Broekstra, et al., 2002; Wilkinson, et al., 2003; Heflin, et al., 1999) and related software tools have been available in the market for quite a while. By specifying the user retrieval requirements on the data as some semantic web queries which involve entities defined in some domain ontologies, relevant data can readily be retrieved. However, allowing better search on such distributed on-line resources also implies that more complicated access control is needed, especially for applications where data containing sensitive and private information (e.g., business, healthcare, national defense, etc.) could be found in the query results. In addition, if the queries are for supporting intelligent on-line services, the on-demand requirement will require the access control system to be not only secure but also flexible, or it will not be operationally viable. The conventional role-based access control model does allow some form of access control to be enforced, but rather restricted (e.g., context specific access control mechanisms are hard to be supported). In recent years, enforcing access control based on some declarative policies is gaining attention (Kagal, et al., 2003), and has been considered to be especially suitable for the dynamic characteristic of semantic web data (Bonatti, et al., 2006). One intuitive idea to implement the policy-based approach is to rewrite queries by adding some restrictions on them so that the rewritten queries will result only the data which can be accessed according to the access control policies. In this paper, we argue that proper use of ontology can allow more flexible access control on semantic web data to be supported. Also, we present a particular policy-based approach to demonstrate the viability of controlling query-based access to the semantic web data.

The main contributions of this paper include (1) an in-depth study of an ontological approach to specify access control policies as permission or forbiddance to access resources; and (2) algorithms for automatically rewriting queries which is made possible due to the unambiguous semantics of the proposed access control policies. For the latter, in particular, we propose algorithms that, given an ontology-based conjunctive query for web resources, can a) select the policies which are relevant to the query, b) extract the restrictions expression from the policies, and c) rewrite the queries accordingly.

2 A BRIEF OVERVIEW ON SEMANTIC WEB

The vision of the Semantic Web (Berners-Lee, et al., 2001) is to have on-line resources expressed in a machine understandable format so that they can be interpreted and used by software agents. The ultimate goal is to enable Web users to find, share, integrate and thus reason on distributed information or knowledge more effectively. In the Semantic Web, ontology is one of the key concepts where entities and relationships intended to be modeled within some domains are described or posited as a form of knowledge. "Concept" and "role" are the two common terms used to refer to the two modeling elements in ontologies. Concepts (also called classes) refer to the abstract definitions of entities within the domain, whereas roles express the relationships between entities. On-line resources annotated with the labels of concepts are generally termed as *instances* of these concepts. Figure 1 shows a simple ontology about the domain of weapons, where concepts (e.g., "Weapon" and "Missile") are annotated using ellipses, "subclass" roles and other user defined roles (e.g., "HasRange") are annotated using dashed and solid lines respectively, and individual values are annotated using rectangles (e.g., "Liquid Fuel" and "Solid Fuel"). Resources about missiles can then be described accordingly using instances of concepts defined in this ontology.

Several markup languages, e.g., Resource Description Framework - RDF (Brickley & Guha,