Administering the Semantic Web: Confidentiality, Privacy, and Trust Management

Bhavani Thuraisingham, The University of Texas at Dallas, USA
Natasha Tsybulnik, The University of Texas at Dallas, USA
Ashraful Alam, The University of Texas at Dallas, USA

ABSTRACT

The Semantic Web is essentially a collection of technologies to support machine-understandable Web pages as well as Information Interoperability. There has been much progress made on the Semantic Web, including standards for eXtensible Markup Language, Resource Description Framework, and Ontologies. However, administration policies and techniques for enforcing them have received little attention. These policies include policies for security, privacy, data quality, integrity, trust, and timely information processing. This article discusses administration policies for the Semantic Web as well as techniques for enforcing them. In particular, we will discuss an approach for ensuring confidentiality, privacy, and trust for the Semantic Web. We will also discuss the inference and privacy problems within the context of administration policies.

Keywords: algorithms; ontologies; semantic matching; Web architecture

INTRODUCTION

A Semantic Web can be thought of as a Web that is highly intelligent and sophisticated so that one needs little or no human intervention to carry out tasks such as scheduling appointments, coordinating activities, searching for complex documents, as well as integrating disparate databases and information systems (Lee, 2001). Recently there have been many developments on the Semantic Web (see, for example, Thuraisingham, 2002). The World Wide Web consortium (W3C) is specifying standards for the Semantic Web. These standards include specifications for XML (eXtensible Markup Language), RDF (Resource Description Framework), and Ontologies.

While much progress has been made toward developing such an intelligent Web, there is still a lot to be done in terms of security, privacy, data quality, integrity, and trust management. It is critical that the Semantic Web be secure and trustworthy, that is, the components that constitute the Semantic Web have to be secure. The components include XML, RDF, and Ontologies. In addition, we need secure information integration. We also need to examine trust issues for the Semantic Web. Essentially what we need is a set of administration policies as well as techniques for enforcing these policies for the Semantic Web.

This article focuses on administration issues for the Semantic Web with an emphasis...
on confidentiality, privacy, and trust management. In the case of security policies, which we will also call confidentiality policies, we will discuss XML security, RDF security, and secure information integration. We also discuss privacy for the Semantic Web. Trust management issues include the extent to which we can trust the users and the Web sites to enforce security and privacy policies.

The organization of this article is as follows. Our definitions of confidentiality, privacy, and trust, as well as the current status on administering the Semantic Web, will be discussed first. This will be followed by a discussion of our proposed framework for securing the Semantic Web, which we call CPT (Confidentiality, Privacy, and Trust). Next we will take each of the features, Confidentiality, Privacy, and Trust, and discuss various aspects as they relate to the Semantic Web. An integrated architecture for CPT, as well as inference and privacy control, will also be discussed. Finally, the article is summarized and future directions are given.

**TRUST, PRIVACY, AND CONFIDENTIALITY**

**Definitions**

Confidentiality, privacy, trust, integrity, and availability will be briefly defined with an examination of how these issues specifically relate to the trust management and inference problem. Confidentiality is preventing the release of unauthorized information. Privacy is a subset of confidentiality in that it is the prevention of unauthorized information from being released in regards to an individual. Integrity of data is the prevention of any modifications made by an unauthorized entity. Availability is the prevention of unauthorized omission of data. Trust is a measure of confidence in data correctness and legitimacy from a particular source.

Integrity, availability, and trust are all very closely related in the sense that data quality is of particular importance, and all require individuals or entities processing and sending information to not alter the data in an unauthorized manner. If all of these issues, confidentiality, privacy, trust, integrity, and availability, are guaranteed, a system can be considered secure. Thus if the inference problem can be solved such that unauthorized information is not released, the rules of confidentiality, privacy, and trust will not be broken. A technique such as inference can either be used to aid or impair the cause of integrity, availability, and trust. If correctly used, inference can be used to infer trust management policies. Thus inference can be used for good or bad purposes. The intention is to prevent inferred unauthorized conclusions and to use inference to apply trust management.

**Current Successes and Potential Failures**

W3C is proposing encryption techniques for securing XML documents. Furthermore, logic, proof, and trust belong to one of the layers of the Semantic Web. However, by trust in that context is meant whether the Semantic Web can trust the statements such as data and rules. In our definition, by trust we mean to what extent we can believe that the user and the Web site will enforce the confidentiality and privacy policies as specified. Privacy has been discussed by the Semantic Web community. The main contribution of this community is developing the Platform for Privacy Preferences (P3P).

P3P requires the Web developer of the server to create a privacy policy, validate it, and then place it in a specific location on the server as well as write a privacy policy in English. When the user enters the Web site, the browser will discover the privacy policy; if the privacy policy matches the user’s browser security specifications, then the user can simply enter the site. If the policy does not match the user’s specifications, then the user will be informed of the site’s intentions, and the user can then choose to enter or leave.

While this is a great start, it is lacking in certain areas. One concern is the fact that the privacy policy must be placed in a specific location. If a Web site, for example, a student Web site on a school’s server, is to implement...