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

Bhavani Thuraisingham
University of Texas at Dallas, USA

Natasha Tsybulnik
University of Texas at Dallas, USA

Ashraful Alam
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 chapter discusses administration policies for the Semantic Web as well as techniques for enforcing them. In particular, the authors 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.
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 paper focuses on administration issues for the semantic web with 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 paper 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 paper is summarized and future directions are given.

TRUST, PRIVACY, AND CONFIDENTIALITY

In this section we will discuss issues on confidentiality, privacy and trust.

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