Title: Building Trustworthy Semantic Webs

ABSTRACT

Recent developments in information systems technologies have resulted in computerizing many applications in various business areas. Data has become a critical resource in many organizations, and therefore, efficient access to data, sharing the data, extracting information from the data, and making use of the information has become an urgent need. As a result, there have been many efforts on not only integrating the various data sources scattered across several sites, but extracting information from these databases in the form of patterns and trends has also become important. These data sources may be databases managed by database management systems, or they could be data warehoused in a repository from multiple data sources.

The advent of the World Wide Web (WWW) in the mid 1990s has resulted in even greater demand for managing data, information and knowledge effectively. There is now so much data on the web that managing it with conventional tools is becoming almost impossible. New tools and techniques are needed to effectively manage this data. Therefore, to provide interoperability as well as to ensure machine understandable web pages, the concept of semantic web was conceived by. Tim Berners Lee who heads W3C (the World Wide Web Consortium).

As the demand for data and information management increases, there is also a critical need for maintaining the security of the databases, applications and information systems. Data and information have to be protected from unauthorized access as well as from malicious corruption. With the advent of the web it is even more important to protect the data and information as numerous individuals now have access to this data and information. Therefore, we need effective mechanisms to secure the semantic web technologies.

As stated by Tim Berners Lee, the semantic web consists of a collection of technologies that enable machine understandable web pages. The idea is for agents acting on behalf of users to collaborate with one another, invoke web services, understand the web pages and carry out activities such as make airline reservations, plan for a surgery or design a vehicle. The technologies that consist of the semantic web include markup languages such as XML (Extensible Markup Language), semantics based languages such as RDF (Resource Description Framework) and ontology languages such as OWL (web ontology language). Agents use these technologies, negotiate contracts with each other and carry out activities. In order to ensure the security of operation, the semantic web needs to enforce policies for confidentiality, privacy. trust, and integrity among others. That is, policies specify the types of access that agents have to the web resources and also the extent to which the agents trust one another.In order to carry out negotiation various inferencing systems have been developed. While numerous developments have been reported on semantic web technologies, it is only recently that security is getting some attention. Therefore one of the major directions for the semantic web is to ensure the security of operation. We discuss some of the security issues in the next few paragraphs.

Consider the XML layer of the semantic web. One needs secure XML. That is, access must be controlled to various portions of the document for reading, browsing and modifications. There is research on securing XML and XML schemas. The next step is securing RDF. Now with RDF

not only do we need secure XML, we also need security for the interpretations and semantics. For example under certain contexts, portions of the document may be Unclassified while under certain other contexts the document may be classified. As an example one could declassify an RDF document, once the war is over.

Once XML and RDF have been secured the next step is to examine security for ontologies. That is, ontologies may have security levelsattached to them. Certain parts of the ontologies could be Secret while certain other parts may be Unclassified. The challenge is how does one use these ontologies for applications such as secure information integration? Researchers have done some work on the secure interoperability of databases and the use of ontologies is being explored.

We also need to examine the inference problem for the semantic web. Inference is the process of posing queries and deducing new information. Itbecomes a problem when the deduced information is something the user is unauthorized to know. With the semantic web, and especially with data mining tools, one can make all kinds of inferences. Recently there has been someresearch on controlling unauthorized inferences on the semantic web.

Security should not be an afterthought. We have often heard that one needs to insert security into the system right from the beginning. Similarly security cannot be an afterthought for the semantic web. However, we cannot also make the system inefficient if we must guarantee one hundred percent security at all times. What is needed is a flexible security policy. During some situations we may need one hundred percent security while during some other situations some security (e.g., 60%) may be sufficient.

Closely related to security is privacy. The challenge here is protecting sensitive information about the individuals. Other challenges include trust management and negotiation. How do we determine the trust that agents place on one another? Is it based on the reputation of the agents? Another challenge is maintaining integrity. For example, when XML documents are published by third parties, we need to ensure that the documents are authentic and are of high quality. We hope that many of these challenges will be clearer in this book. As more progress is made on investigating these various issues, we hope that appropriate standards would be developed for securing the semantic web. Note that while security is essentially about confidentiality, we use the term trustworthiness to include not only confidentiality, but also privacy, trust and integrity.

This presentation will review the developments in semantic web technologies and describe ways of securing these technologies. The focus will be on confidentiality, privacy, trust, and integrity management fop the semantic web. We will call such a semantic web a trustworthy semantic web. We will also discuss applications of trustworthy semantic webs in secure web services, secure interoperability, secure knowledge management, secure e-business and secure information sharing.

Short Biography:

Dr. Bhavani Thuraisingham joined The University of Texas at Dallas in October 2004 as a Professor of Computer Science and Director of the Cyber Security Research Center in the Erik

Jonsson School of Engineering and Computer Science. She is an elected Fellow of three professional organizations: the IEEE (Institute for Electrical and Electronics Engineers), the AAAS (American Association for the Advancement of Science) and the BCS (British Computer Society) for her work in data security. She received the IEEE Computer Society's prestigious 1997 Technical Achievement Award for "outstanding and innovative contributions to secure data management."

Dr Thuraisingham's work in information security and information management has resulted in over 80 journal articles, over 200 refereed conference papers and workshops, and three US patents. She is the author of nine books in data management, data mining and data security including one on data mining for counter-terrorism and another on Database and Applications Security and is completing her tenth book on Trustworthy Service Oriented Information Systems. She has given over 70 keynote presentations at various technical conferences and has also given invited talks at the White House Office of Science and Technology Policy and at the United Nations on Data Mining for counter-terrorism. She serves (or has served) on editorial boards of leading research and industry journals and currently serves as the Editor in Chief of Computer Standards and Interfaces Journal. She is also an Instructor at AFCEA's (Armed Forces Communications and Electronics Association) Professional Development Center and has served on panels for the Air Force Scientific Advisory Board and the National Academy of Sciences.

Dr Thuraisingham is the Founding President of "Bhavani Security Consulting" - a company providing services in consulting and training in Cyber Security and Information Technology

Prior to joining UTD, Thuraisingham was an IPA (Intergovernmental Personnel Act) at the National Science Foundation from the MITRE Corporation. At NSF she established the Data and Applications Security Program and co-founded the Cyber Trust theme and was involved in inter-agency activities in data mining for counter-terrorism. She has been at MITRE since January 1989 and has worked in MITRE's Information Security Center and was later a department head in Data and Information Management as well as Chief Scientist in Data Management. She has served as an expert consultant in information security and data management to the Department of Defense, the Department of Treasury and the Intelligence Community for over 10 years. Thuraisingham's industry experience includes six years of research and development at Control Data Corporation and Honeywell Inc.

Thuraisingham was educated in the United Kingdom both at the University of Bristol and at the University of Wales.