BUILDING TRUSTWORTHY SEMANTIC WEBS

Bhavani Thuraisingham

Auerbach Publications
Taylor & Francis Group

BUILDING TRUSTWORTHY SEMANTIC WEBS

OTHER AUERBACH PUBLICATIONS

Ad Hoc Mobile Wireless Networks: Principles, Protocols and Applications
Subir Kumar Sarkar, T.G. Basavaraju and C. Puttamadappa
ISBN 1-4200-6221-2

Computer Fraud: An In-depth Framework for Detecting and Defending against Insider IT Attacks
Kenneth C. Brancik
ISBN 1-4200-4659-4

Design Science Research Methods and Patterns: Innovating Information and Communication Technology
Vijay K. Vaishnavi and William Kuechler Jr.
ISBN 1-4200-5932-7

Determining Project Requirements
Hans Jonasson
ISBN 1-4200-4502-4

Digital Privacy: Theory, Technologies, and Practices
Alessandro Acquisti, Stefanos Gritzalis, Costos Lambrinoudakis and Sabrina di Vimercati
ISBN 1-4200-521-79

Effective Communications for Project Management
Ralph L. Kliem
ISBN 1-4200-6246-8

Effective Transition from Design to Production
David F. Ciambrone
ISBN 1-4200-4686-1

Elements of Compiler Design
Alexander Meduna
ISBN 1-4200-6323-5

How to Achieve 27001 Certification: An Example of Applied Compliance Management
Sigurjon Thor Arnason and Keith D. Willett
ISBN 0-8493-3648-1

Inter- and Intra-Vehicle Communications
Gilbert Held
ISBN 1-4200-5221-7

Manage Software Testing
Peter Farrell-Vinay
ISBN 0-8493-9383-3

Managing Global Development Risk
James M. Hussey and Steven E. Hall
ISBN 1-4200-5520-8

Mobile WiMAX: Toward Broadband Wireless Metropolitan Area Networks
Yan Zhang and Hsiao-Hwa Chen
ISBN 0-8493-2624-9

Operational Excellence: Using Lean Six Sigma to Translate Customer Value through Global Supply Chain
James William Martin
ISBN 1-4200-6250-6

Physical Principles of Wireless Communications
Victor L. Granatstein
ISBN 0-8493-3259-1

Practical Guide to Project Planning
Ricardo Viana Vargas
ISBN 1-4200-4504-0

Principles of Mobile Computing and Communications
Mazliza Othman
ISBN 1-4200-6158-5

Programming Languages for Business Problem Solving Price
Shouhong Wang and Hai Wang
ISBN 1-4200-6264-6

Retail Supply Chain Management
James B. Ayers and Mary Ann Odegaard
ISBN 0-8493-9052-4

Security in Wireless Mesh Networks
Yan Zhang, Jun Zheng and Honglin Hu
ISBN 0-8493-8250-5

Service-Oriented Architecture: SOA Strategy, Methodology, and Technology
James P. Lawler and H. Howell-Barber
ISBN 1-4200-4500-8

The Strategic Project Leader: Mastering Service-Based Project Leadership
Jack Ferraro
ISBN 0-8493-8794-9

Simplified TRIZ: New Problem Solving Applications for Engineers and Manufacturing Professionals, Second Edition
Kalevi Rantanen and Ellen Domb
ISBN 1-4200-6273-5

Value-Added Services for Next Generation Networks
Thierry Van de Velde
ISBN 0-8493-7318-2

WiMAX: A Wireless Technology Revolution
G.S.V. Radha Krishna Rao and G. Radhamani
ISBN 0-8493-7059-0

AUERBACH PUBLICATIONS

www.auerbach-publications.com
To Order Call: 1-800-272-7737 • Fax: 1-800-374-3401
E-mail: orders@crcpress.com

BUILDING TRUSTWORTHY SEMANTIC WEBS

Bhavani Thuraisingham

Auerbach Publications
Taylor & Francis Group
New York London

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

Auerbach Publications
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-13: 978-0-8493-5080-1 (Hardcover)

Library of Congress Cataloging-in-Publication Data

Thuraisingham, Bhavani M.
Building trustworthy semantic webs / Bhavani Thuraisingham.
p. cm.
Includes bibliographical references and index.
ISBN-13: 978-0-8493-5080-1 (alk. paper)
ISBN-10: 0-8493-5080-8 (alk. paper)
1. Semantic Web. 2. Database security. I. Title.

TK5105.88815T59 2008
025.04--dc22 2007027962

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the Auerbach Web site at
http://www.auerbach-publications.com

Dedication

In Memory of My Parents

Nitchingam
3 October 1913 – 30 October 1971
and
Gnanam
25 September 1919 – 28 October 1984

Although your time with me was short,
you gave me the strength to be brave and take on challenges.

Contents

Preface

Background

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 also extracting information from these databases in the form of patterns and trends. 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 a semantic Web was conceived by Tim Berners Lee who heads the World Wide Web Consortium (W3C).

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 semantic Web technologies.

This book 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 for 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.

We have written a series of books for Taylor & Francis on data management, data mining, and data security.

Book 1 (*Data Management Systems Evolution and Interoperation*) focused on general aspects of data management and also addressed interoperability and migration.
Book 2 (*Data Mining: Technologies, Techniques, Tools and Trends*) discussed data mining. It essentially elaborated on Chapter 9 of Book 1.
Book 3 (*Web Data Management and Electronic Commerce*) discussed Web database technologies and E-commerce as an application area. It essentially elaborated on Chapter 10 of Book 1.
Book 4 (*Managing and Mining Multimedia Databases for the Electronic Enterprise*) addressed both multimedia database management and multimedia data mining. It elaborated on both Chapter 6 of Book 1 (for multimedia database management) and Chapter 11 of Book 2 (for multimedia data mining).
Book 5 (*XML, Databases, and the Semantic Web*) described XML technologies related to data management. It elaborated on Chapter 11 of Book 3.
Book 6 (*Web Data Mining Technologies and Their Applications in Business Intelligence and Counterterrorism*) elaborated on Chapter 9 of Book 3.
Book 7 (*Database and Applications Security: Integrating Data Management and Information Security*) examines security for technologies discussed in each of our previous books. It focuses on the technological developments in database and applications security. It is essentially the integration of information security and database technologies.

One can regard our Book 7 to be the start of a new series in data security. Our current book (Book 8) is an elaboration of Chapter 25 of Book 7. It also integrates security with the contents of Book 5. The relationships between our texts will be illustrated in Appendix A.

Developments and Directions for Trustworthy Semantic Webs

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 eXtensible Markup Language (XML), semantics-based languages such as Resource Description Framework (RDF), and ontology

languages such as Web Ontology Language (OWL). Agents use these technologies, negotiate contracts with each other, and carry out activities. 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. To carry out negotiations, various inferencing systems have been developed. Although 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, whereas 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 levels attached to them. Certain parts of the ontologies could be secret, whereas 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. It becomes 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 some research 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 100 percent security at all times. What is needed is a flexible security policy. During some situations we may need 100 percent security, whereas during some other situations some security (e.g., 60 percent) may be sufficient.

Closely related to security is privacy. The challenge here is protecting sensitive information about individuals. Other challenges include trust management and negotiation. How do we determine the trust that agents place in 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 will be developed for securing the semantic Web. Note that although security is essentially about confidentiality, we use the term trustworthiness to include not only confidentiality, but also privacy, trust, and integrity.

Organization of this Book

This book is divided into five parts, each describing some aspect of trustworthy semantic Webs. Part I, consisting of three chapters, discusses concepts in trustworthy-information systems. Note that the supporting technologies for trustworthy semantic Webs are trustworthy-information systems and semantic Webs. Trustworthy-information systems consist of many aspects. We will focus on three aspects. Chapter 2 discusses concepts in trustworthy systems including secure systems as well as features such as integrity, trust, and privacy. Chapter 3 discusses secure data, information, and knowledge management. We discuss topics such as secure database systems, secure information systems such as secure multimedia systems, and secure knowledge management. Chapter 4 discusses concepts in semantic Webs.

Part II, consisting of five chapters, discusses secure semantic Webs. Note that this part is the heart of the book. In Chapter 5 we provide an overview of secure semantic Webs. In Chapter 6 we discuss XML security. In Chapter 7 we discuss RDF security. Security and ontologies including security for OWL are discussed in Chapter 8. Integrating security into Web rules is the subject of Chapter 9.

Part III, which consists of six chapters, discusses dependability of the semantic Web. Note that whereas security (i.e., confidentiality) has been our main focus, we also address other features such as trust management and privacy. Chapter 10 discusses trust management for the semantic Web. We discuss trust policies and describe how automatic trust management may be included in the operation of the semantic Web. Note that trust is already discussed in the definition of the semantic Web by Tim Berners Lee. For example, how can we trust the statements? Logicians are working on proof systems to determine trust. However, the security community has also investigated trust extensively. For example, if A trusts B and B trusts C, then should A trust C? Chapter 11 discusses privacy for the semantic Web. We examine the Platform for Privacy Preferences Project (P3P) and discuss ways to extend it. We also examine the privacy problems that arise through semantic Web mining and discuss approaches for privacy-preserving semantic Web mining. Chapter 12 discusses integrity and data quality for the semantic Web. How do we ensure that the information that is exchanged is of high quality? Multilevel security for the semantic Web is the subject of Chapter 13. Note that multilevel security is an aspect of confidentiality. However, we decided not to include it in Part II to give

Part II more focus. Managing the policies is an important aspect. Therefore, policy engineering is discussed in Chapter 14. Finally, in Chapter 15 we elaborate on the developments about the semantic Web from research, standards, products, and applications points of view. We decided to include Chapter 15 in Part III mainly for completion.

Part IV discusses applications that utilize trustworthy semantic Webs. Chapter 16 discusses secure Web services that utilize semantic Web technologies. Semantic Web technologies for managing secure databases is the subject of Chapter 17. Secure semantic interoperability for heterogeneous information sources is discussed in Chapter 18. Chapter 19 discusses a semantic Web for secure E-business applications. Chapter 20 discusses a semantic Web for secure digital libraries. Chapter 21 discusses semantic Web technologies for an important applications area called assured information sharing.

Part V consists of three chapters and describes specialized and domain-specific semantic Webs. In Chapter 22 we discuss domain-specific semantic Webs for defense, financial, and medical domains, among others. Trustworthy semantic Webs for geospatial data as well as for sensor data are discussed in Chapter 23. In particular, the work carried out on Geography Markup Language (GML) as well as the interoperability work of the Open Geospatial Consortium (OGC) is discussed. Chapter 24 discusses pervasive computing applications including secure mobile sensor semantic Webs as well as pervasive semantic Webs.

Chapter 25 summarizes the book and discusses future directions. We have included three appendices. Appendix A provides an overview of data management and discusses the relationship between the texts we have written. A summary of data and applications security (which is our Book 7) is given in Appendix B to give the reader a better understanding as to where we are coming from. Various standards efforts related to the semantic Web are detailed in Appendix C. This book ends with a bibliography and an index. Each of Chapters 1 through 25 includes references for that chapter as well as exercises that will be useful for students.

Data, Information, and Knowledge

In general, data management includes managing the databases, interoperability, migration, warehousing, and mining. For example, the data on the Web has to be managed and mined to extract information, patterns, and trends. Data could be in files, relational databases, or other types of databases such as multimedia databases. Data may be structured or unstructured. We repeatedly use the terms data, data management, and database systems and database-management systems in this book. We elaborate on these terms in Appendix A. We define data-management systems to be systems that manage the data, extract meaningful information from

the data, and make use of the information extracted. Therefore, data-management systems include database systems, data warehouses, and data-mining systems. Data could be structured data such as that found in relational databases, or it could be unstructured such as text, voice, imagery, and video.

There have been numerous discussions in the past to distinguish between data, information, and knowledge. In some of our previous books on data management and mining, we did not attempt to clarify these terms. We simply stated that data could be just bits and bytes, or it could convey some meaningful information to the user. However, with the Web and also with increasing interest in data, information, and knowledge management as separate areas, in this book we take a different approach to data, information, and knowledge by differentiating between these terms as much as possible. For us data is usually some value like numbers, integers, and strings. Information is obtained when some meaning or semantics is associated with the data such as John's salary is 20K. Knowledge is something that you acquire through reading and learning and, as a result, understand the data and information and take actions. Data and information can be transferred into knowledge when uncertainty about the data and information is removed from someone's mind. It should be noted that it is rather difficult to give strict definitions of data, information, and knowledge. Sometimes we will use these terms interchangeably. Our framework for data management discussed in Appendix A helps clarify some of the differences. To be consistent with the terminology in our previous books, we distinguish between database systems and database-management systems. A database-management system is that component which manages the database containing persistent data. A database system consists of both the database and the database-management system.

Final Thoughts

The goal of this book is to explore security issues for the semantic Web and discuss how trustworthy semantic Webs may be applied for Web services, interoperability, and knowledge management, among others. The goal is also to show the breadth of the applications of trustworthy semantic Webs in multiple domains. We have used the material in this book together with the numerous papers listed in the references in each chapter for a graduate level course at the University of Texas at Dallas on *Building Trustworthy Semantic Web*s. In addition to trying out the exercises at the end of each chapter, the students also wrote term papers and carried out a programming project on trustworthy semantic Webs.

One could argue that because the developments in secure semantic Webs are just beginning, this book might be premature. I feel that in many ways it is timely to write such a book so that various viewpoints can be taken into consideration in advancing the field. It is important that appropriate tools and technologies are

developed to secure the semantic Web. Security cannot be an afterthought. Therefore, although the technologies for the semantic Web are being developed, it is important to include security at the onset. Furthermore, a lot of progress has been made on data security, and it is important to take advantage of these developments in securing the semantic Web. I believe strongly in taking as much advantage as possible of the knowledge that is out there rather than reinventing the wheel. It was for these reasons that I decided to write this book now.

Acknowledgments

I would like to thank the Administration at the Erik Jonsson School of Engineering and Computer Science at the University of Texas at Dallas for giving me the opportunity to direct the Cyber Security Research Center and teach courses on data and applications security and building trustworthy semantic Webs. I thank all my Ph.D. and M.S. students for giving me many insights and the students who have taken my classes, especially the students who took my class on Building Trustworthy Semantic Webs during the fall semester of 2006. I am especially grateful to my Ph.D. student Alam Ashraful for the many discussions on trustworthy semantic Webs and giving me examples on query modification, RDF and OWL policy specifications, and geospatial ontology for this book. I also thank my M.S. student Yashaswini Harsha Kumar for example documents on XML, RDF, and OWL for this book. I also thank my M.S. student Ganesh Subbiah for his contributions to geospatial semantic Webs.

I would also like to thank many people who have supported my work including the following:

- My husband Thevendra for his continued support for my work and my son Breman for being such a wonderful person and for motivating me.
- Professor C. V. Ramamoorthy at the University of California at Berkeley for his constant encouragement.
- Henry Bayard at MITRE for his continued mentoring and encouragement.
- Prof. Elisa Bertino (Purdue University), Prof. Tim Finin (University of Maryland [Baltimore County]) for leading the fields of XML security and trust for the semantic Web; Prof. Tim Berners Lee (Massachusetts Institute of Technology) for conceiving the idea of the semantic Web; and Prof. Ravi Sandhu (now UTSA) for RBAC/UCON models that have influenced my research on trustworthy semantic Webs.
- Profs. Elena Ferrari, Barbara Carminati, and Anna Cinzia Squicciarini for collaborating with me on various aspects of XML and RDF security.

- My colleagues at the University of Texas at Dallas, especially Prof. Latifur Khan, Prof. Murat Kantarcioglu, Prof. Kevin Hamlen, and Prof. I-Ling Yen for their collaboration on related topics.
- Prof. Latifur Khan as well as students including Alam Ashraful, Ganesh Subbiah, Nathalie Tsublinik, and Ryan Layfield who have been involved in some aspect of my research on secure semantic Webs.
- My colleagues who have collaborated with me, especially during the past three years since I joined the University of Texas at Dallas.
- The sponsors of my research at the University of Texas at Dallas, especially the Air Force Office of Scientific Research, the National Science Foundation, Raytheon Corporation, the National Geospatial Intelligence Agency, and the Texas Enterprise Funds.
- I also thank the sponsors of my research at MITRE and Honeywell on data and applications security including AFRL, CECOM, SPAWAR, NSA, CIA, IRS, and NASA.
- My former colleagues at the National Science Foundation, the MITRE Corporation, and Honeywell Inc. for their encouragement on my work in secure data management.

I hope that we can continue to make progress in building trustworthy semantic Webs so that we not only have agents that understand Web pages, but also ensure the security of operation in carrying out activities on the Web.

The Author

Bhavani Thuraisingham joined the University of Texas at Dallas (UTD) 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." She was quoted by *Silicon India Magazine* as one of the top seven technology innovators of South Asian origin in the United States in 2002.

Prior to joining UTD, Dr. Thuraisingham was an IPA (Intergovernmental Personnel Act) Program Director at the National Science Foundation (NSF) in Arlington, VA, from the MITRE Corporation. At NSF she established the Data and Applications Security Program and cofounded the Cyber Trust theme and was involved in interagency activities in data mining for counterterrorism. She worked at MITRE in Bedford, MA between January 1989 and September 2001, first in the Information Security Center and was later a department head in Data and Information Management as well as Chief Scientist in Data Management in the Intelligence and Air Force centers. 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. Dr. Thuraisingham's industry experience includes six years of research and development at Control Data Corp. and Honeywell Inc. in Minneapolis, MN. While she was in industry and at MITRE, she was an adjunct professor of computer science and a member of the graduate faculty, first at the University of Minnesota and later at Boston University between 1984 and 2001. She also worked as visiting professor soon after her Ph.D., first at the New Mexico Institute of Technology and later at the University of Minnesota between 1980 and 1983.

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 U.S. patents. She is the author of eight books in data

management, data mining, and data security including one on data mining for counterterrorism and another on database and applications. She has given over 40 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 counterterrorism. She serves (or has served) on editorial boards of leading research and industry journals including several IEEE and ACM Transactions and currently serves as the Editor-in-Chief of *Computer Standards and Interfaces Journal*. She is also an instructor at the Professional Development Center of the Armed Forces Communications and Electronics Association (AFCEA) since 1998 and has served on panels for the Air Force Scientific Advisory Board and the National Academy of Sciences.

Dr. Thuraisingham is the president of Bhavani Security Consulting, which provides consulting and training in information technology and security. Dr. Thuraisingham promotes mathematics and science to high school students as well as to women and underrepresented minorities and has given featured addresses at conferences sponsored by Women and Technology International (WITI) and the Society for Women Engineers (SWE). Articles on her efforts as well as her vision have appeared in multiple magazines including the *Dallas Morning News*, *D Magazine*, *MITRE Matters*, and the *DFW Metroplex Technology Magazine*. She is dedicated to advising and motivating her several research students pursuing M.S. and Ph.D. degrees in data mining and data security at UTD and mentors assistant and associate professors related to her field at the university.

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

Chapter 1

Introduction

1.1 Trends

A semantic Web is intelligent and understands and reads Web pages. At present we need the human in the loop to read and understand Web pages and make decisions. The vision of Tim Berners Lee is to develop technologies such as eXtensible Markup Language (XML), Resource Description Framework (RDF), and ontologies so that agents acting on behalf of users can read and understand the Web pages and make decisions.

Although progress has been made on semantic Webs during the past decade, much progress has also been made on trustworthy information systems over the past three to four decades. Such systems include secure systems, high-assurance systems, and high-integrity systems. Until recently much of the focus has been on security (i.e., confidentiality). However, recently there has been work on integrating features such as security, integrity, trust management, fault tolerance, and real-time processing.

One of the major challenges in the development of semantic Webs is to build trustworthy semantic Webs. By trustworthy semantic Webs we mean semantic Webs that are secure, manage trust, have integrity, ensure privacy, and are capable of processing information in a timely manner. In this book we will discuss developments, directions, and challenges for trustworthy semantic Webs.

Trustworthy semantic Webs integrate two major technologies: trustworthy information systems and semantic Webs. In our terminology, trustworthy information includes systems that are secure and dependable. Dependable systems include high-assurance systems that meet timing constraints, recover from faults,

and ensure integrity. We have assumed that features of trustworthy systems include security, integrity, privacy, and trust. It is almost impossible to incorporate all these features in designing a system. Therefore, the challenge is to make tradeoffs between the various features and enforce flexible policies.

Our main focus in this book will be on building secure semantic Web technologies with a focus on confidentiality. However, we will also give consideration to other features such as integrity, privacy, trust, and data quality so that we can build trustworthy semantic Webs. Some books and papers have used the terms *trustworthiness* and *dependability* interchangeably. Note that standard definitions of the terms such as trustworthiness and dependability have yet to be developed. Our goal is to focus on the features of trustworthy semantic Webs based on the definition that we have assumed, that is, trustworthy semantic Webs include secure semantic Webs and dependable semantic Webs. Dependable semantic Webs include semantic Webs that ensure privacy, manage trust, have high integrity, meet timing constraints, and recover from faults.

1.2 Organization of This Chapter

Before we begin discussing the contents of this book, we give an overview of where we are with respect to trustworthy semantic Webs and discuss why we embarked on this book. Although a lot of work has gone on in recent years on trustworthy semantic Webs (especially since my invited talk at the EU-US [European Union–United States] meeting on semantic Webs at Sophia Antipolis, France, in October 2001, and my funding efforts on this topic while I was a program director at the National Science Foundation between October 2001 and September 2004), there was no source where one could go to find out what is going on in building trustworthy semantic Webs. Therefore, I first decided to teach a course on this topic at the University of Texas at Dallas in the fall of 2006 and subsequently started writing this book, although I had planned on this book soon after I finished my previous book on *Database and Applications Security* in 2005.

This book is divided into five parts. Each part is summarized in the ensuing sections of this chapter. First, an overview of research products and standards is given in Section 1.3. Section 1.4, which summarizes Part I, discusses trustworthy systems, secure data and information-management systems, and semantic Webs. We have assumed that supporting technologies for trustworthy semantic Webs are trustworthy systems, secure data systems, and the semantic Web. Trustworthy information systems include numerous types of information systems including secure systems, real-time systems, fault tolerance systems, and high-assurance systems.

The components of trustworthy semantic Webs are secure semantic Webs and dependable semantic Webs. Section 1.5, which summarizes Part II, discusses concepts in secure semantic Webs, in particular, secure XML, secure RDF, secure

ontologies, and secure rules as well as other security issues such as the inference problem for the semantic Web.

Section 1.6, which summarizes Part III, discusses concepts in dependable semantic Webs. Note that we have focused on security, which we assume to be mainly confidentiality, in Part III. We have used dependability to include other features such as trust management, integrity, and data quality in Part III. In addition, we have included privacy as part of dependability. Therefore, in Section 1.6 we discuss semantic Webs that have to manage trust and ensure data integrity as well as timely processing. Privacy issues including the Platform for Privacy Preferences Project (P3P) are also discussed.

Section 1.7, which summarizes Part IV, discusses various applications for secure and trustworthy semantic Webs. In particular, applications such as secure Web services, secure data management, and secure interoperability are discussed.

Section 1.8, which summarizes Part V, discusses special semantic Webs for different user communities including geospatial Webs, sensor Webs, and Webs for medical and financial communities.

Section 1.9 gives further details on the organization of this book. As with our previous books, this book is also based on a framework for trustworthy semantic Webs. The framework consists of the supporting technologies, core concepts, and applications. Future directions are discussed in Section 1.10.

1.3 Research, Products, and Standards

The major research institution conducting research in trustworthy semantic Webs is the University of Maryland (Baltimore County) under the leadership of Prof. Tim Finin (with Profs. Anupam Joshi and Yelena Yesha). Other institutions include the University of Maryland (Prof. James Hendler, now at Rensselaer Polytechnic Institute), the Massachusetts Institute of Technology (Prof. Tim Berners Lee with Dr. Lalana Kagal and Lawyer Daniel Weitzner among others), and the University of Texas at Dallas (Prof. Bhavani Thuraisingham with Prof. Latifur Khan). Work on policy issues is also being carried out at other universities in the United States and Europe. For example, Purdue University is conducting excellent research on XML-based policy management (Prof. Bertino, formerly at the University of Milan). The University of Como in Italy is conducting research on XML and RDF-based policy management (Profs. Elena Ferrari and Barbara Carminati). George Mason University is conducting research on bringing the Usage Control (UCON) Model into the semantic Web framework (Prof. Ravi Sandhu now at the University of Texas at San Antonio).

At present there is no commercial semantic Web or secure semantic Web product. However, one can develop a semantic Web by putting together technologies such as XML and RDF. Corporations such as International Business Machines Corporation (IBM), Oracle, Microsoft, and SAP among others are developing

semantic Web technologies. For example, Oracle has developed a data-management system to manage XML and, more recently, RDF documents. Oracle also is providing security solutions.

With respect to standards, organizations such as the World Wide Web Consortium (W3C) and the Organization for the Advancement of Structured Information Standards (OASIS) are developing semantic Web standards. Whereas W3C is focusing entirely on standards for semantic Webs, OASIS standards are mainly based on XML technologies. Also organizations such as the Open Geospatial Consortium (OGC) are developing standards such as Geography Markup Language (GML) for geospatial data management.

1.4 Trustworthy Information Systems

Part I of this book will discuss the supporting technologies for trustworthy semantic Webs. These supporting technologies are lumped into what we call trustworthy information systems. These systems consist of many aspects including trustworthy systems, secure data and information systems, and semantic Webs.

Trustworthy systems are systems that are secure and dependable. By dependable systems we mean systems that have high integrity, are fault tolerant, and meet real-time constraints. Trustworthy systems may include information systems including data-management systems, information-management systems, and trustworthy networks. In other words, for a system to be trustworthy it must be secure, fault tolerant, meet timing deadlines, and manage high-quality data. However, integrating these features into a system means that the system has to meet conflicting requirements. For example, if the system makes all the access control checks, then it may miss some of its deadlines. The challenge in designing trustworthy systems is to design systems that are flexible. For example, in some situations it may be important to meet all the timing constraints, whereas in other situations it may be critical to satisfy all the security constraints.

Trustworthy systems have sometimes been referred to as dependable systems, whereas in some other cases dependability is considered to be part of trustworthiness. For example, in some papers dependability includes mainly fault-tolerant systems, and when one integrates fault tolerance with security, then one gets trustworthy systems. Regardless of what the definitions are, for systems to be deployed in operational environments, especially for command and control and other critical applications, we need end-to-end dependability as well as security. For some applications not only do we need security and confidentiality, we also need to ensure that the privacy of the individuals is maintained. Therefore, privacy is also another feature of trustworthiness.

For a system to be dependable and trustworthy, we need end-to-end dependability and trustworthiness. Note that the components that comprise a system include

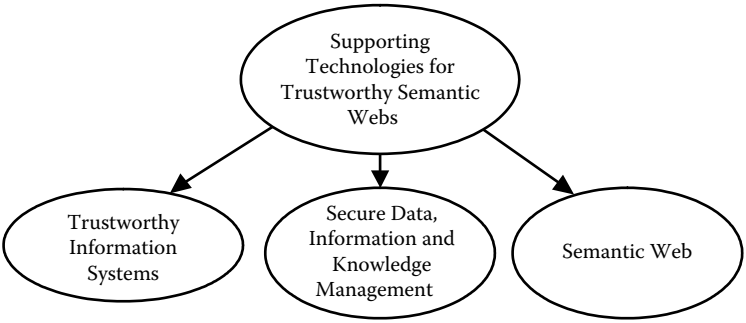

Figure 1.1 Supporting technologies for trustworthy semantic Webs.

the network, operating systems, middleware and infrastructure, data manager, and applications. We need all the components to be dependable and trustworthy.

As stated earlier, other supporting technologies for building trustworthy semantic Webs are secure data, information and knowledge-management systems as well as semantic Webs. Secure data and information systems include secure database systems such as secure relational database systems and secure information systems such as secure multimedia information systems and digital libraries.

The third supporting technology is the semantic Web. As previously stated, the goal of the semantic Web is to ensure that the Web pages are machine understandable. This is the vision of Tim Berners Lee. The idea is to develop common ontologies and specification languages so that agents that act on behalf of the users can read the Web pages and make sense of the data. The ultimate goal is for the system to take actions without the human in the loop. Figure 1.1 illustrates the supporting technologies for trustworthy semantic Webs. Figure 1.2 illustrates the technology stack for the semantic Web as defined by Tim Berners Lee.

1.5 Secure Semantic Webs

As discussed earlier, trustworthy semantic Webs include secure semantic Webs and dependable semantic Webs. Part II discusses technologies for secure semantic Webs. By security we mean confidentiality. For many applications, especially for Command, Control Communications, Computers, Intelligence, Surveillance, and Reconnaissance (C4ISR), the semantic Web has to operate securely. Note that security cannot be considered in isolation. Security cuts across all layers, and this is a challenge; that is, we need security for each of the layers illustrated in Figure 1.3.

For example, consider the lowest layer. One needs secure Transmission Control Protocol/Internet Protocol (TCP/IP), secure sockets, and secure Hypertext Transfer Protocol (HTTP). There are now security protocols for these various lower-layer

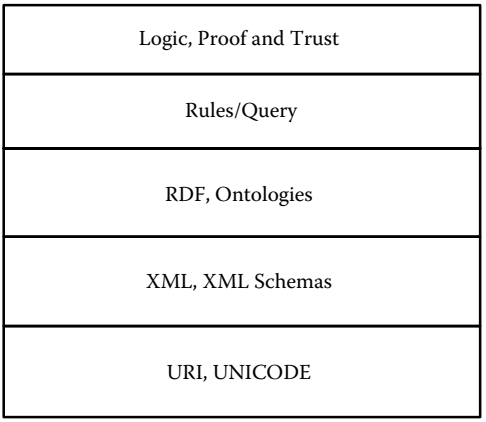

Figure 1.2 Technology stack for the semantic Web.

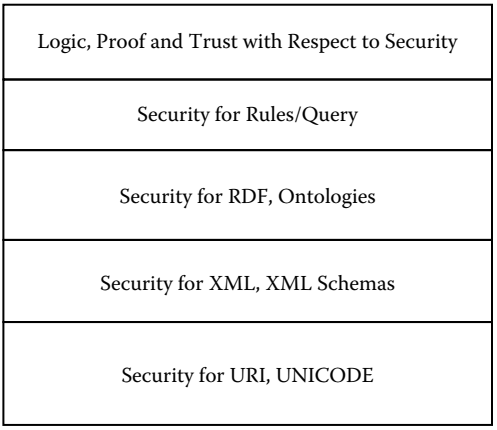

Figure 1.3 Technology stack for a secure semantic Web.

protocols. One needs end-to-end security; that is, one cannot just have secure TCP/IP built on untrusted communication layers. We need network security. The next layer is XML and XML schemas. 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, whereas 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 and interoperation, that is, ontologies may have security levels attached to them. Certain parts of the ontologies could be secret, whereas certain other parts may be unclassified. The challenge is, how does one use these ontologies for secure information integration? Researchers have done some work on the secure interoperability of databases. We need to revisit this research and then determine what else needs to be done so that the information on the Web can be managed, integrated, and exchanged securely.

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 100 percent security at all times. What is needed is a flexible security policy. During some situations we may need 100 percent security, whereas during some other situations 30 percent security may be sufficient. In Part II we discuss secure XML, RDF, and ontologies as well as rules and the inference problem.

1.6 Dependable Semantic Webs

In our definition, trustworthiness consists of security and dependability. By dependable system, we mean systems that have integrity, high assurance, and are fault tolerant and meet real-time constraints. Similarly, a dependable semantic Web is a semantic Web that has integrity; the information is of high quality, is fault tolerant, and meets timing constraints. We have also added privacy, trust management, and rights management as part of dependability. Note that this is not a standard definition; that is, some papers and books have used the terms trustworthiness and dependability interchangeably. Furthermore, some papers have also implied that security includes confidentiality, integrity, and privacy.

Figure 1.4 illustrates aspects of a dependable semantic Web. The challenge is to ensure that the semantic Web has all the features such as privacy, trust, and integrity. Essentially the system has to be flexible. Part III focuses on dependable semantic Webs.

1.7 Applications

A semantic Web is being deployed for many domain applications including medical, financial, and command and control. However, a semantic Web is also a key technology for many other technologies such as Web services, grids, and knowledge management. Therefore, by applications we mean the technical applications such as knowledge management. Domain applications are discussed in Section 1.8.

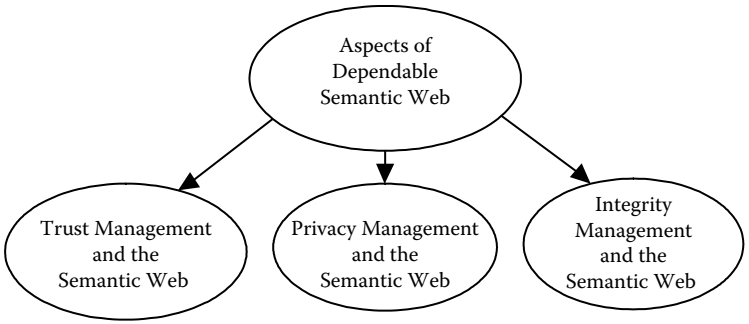

Figure 1.4 Aspects of a dependable semantic Web.

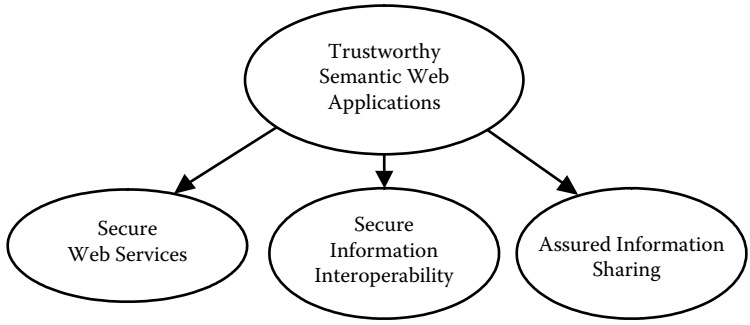

Figure 1.5 Trustworthy semantic Web technology applications.

Web services are the various services that a user can invoke. These services make use of semantic Web technologies such as XML and RDF. Secure semantic Web technologies can be applied to secure Web services. Similarly, secure semantic Web technologies can be applied to secure data, information, and knowledge management. For example, technologies such as RDF and ontologies are useful to capture knowledge, and the reasoning tools could be used to manage the knowledge. Other applications include interoperability and E-business. Figure 1.5 illustrates the technical applications.

1.8 Specialized Trustworthy Semantic Webs

Although much of the research and development about the semantic Web has focused on managing and exchanging text-based and structured data, there is now an urgent need to manage geospatial and sensor data. Languages such as GML and

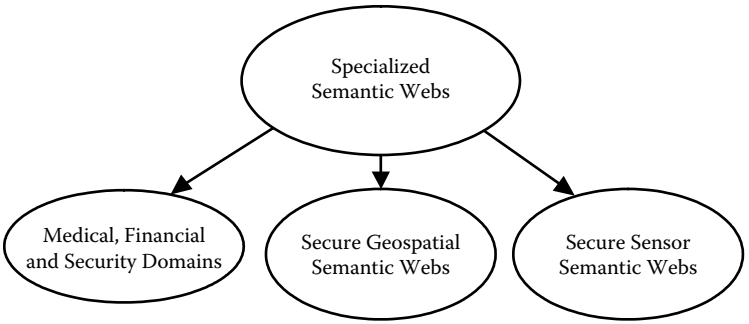

Figure 1.6 Specialized trustworthy semantic Webs.

Sensor Markup Language (SML) are being developed by standards organizations such as OGC. The end result is semantic Webs for geospatial and sensor data.

Very few of the efforts have focused on incorporating security and trust for specialized semantic Webs. Part V of the book discusses these specialized Webs as well as Webs for domains for medical, financial, and defense applications (Figure 1.6).

1.9 Organization of This Book

This book is divided into five parts, each describing some aspect of trustworthy semantic Webs. As mentioned, the major focus of this book will be on security and confidentiality. Other features such as trust management, integrity, data quality, and timely fault-tolerant processing will be addressed for semantic Webs. Applications for trustworthy semantic Webs such as semantic E-business and digital libraries will also be discussed.

Part I discusses concepts in trustworthy information systems. Note that the supporting technologies for trustworthy semantic Webs are trustworthy information systems and semantic Webs. Trustworthy information systems consist of many aspects. We focus on three aspects. Chapter 2 discusses concepts in trustworthy systems including secure systems as well as features such as integrity, trust, and privacy. Chapter 3 discusses secure data, information, and knowledge management. We discuss topics such as secure database systems, secure information systems such as secure multimedia systems, and secure knowledge management. Chapter 4 discusses concepts in semantic Webs.

Part II discusses secure semantic Webs. Note that this part is the heart of the book. In Chapter 5 we provide an overview of secure semantic Web. In Chapter 6 we discuss XML security based on our collaborative research with the University of Milan. In Chapter 7 we discuss RDF security. Security and ontologies including

security for Web Ontology Language (OWL) are discussed in Chapter 8. Integrating security into Web rules is the subject of Chapter 9.

Part III discusses dependability of the semantic Web. Note that whereas security (i.e., confidentiality) has been our main focus, we also address other features such as trust management and privacy. Chapter 10 discusses trust management for the semantic Web. We discuss trust policies and describe how automatic trust management may be included in the operation of the semantic Web. Note that trust is already discussed in the definition of the semantic Web by Tim Berners Lee. For example, how can we trust the statements? Logicians are working on proof systems to determine trust. However, the security community has also investigated trust extensively. For example, if A trusts B and B trusts C, then should A trust C? Finin and others have carried out extensive research on trust for the semantic Web at the University of Maryland (Baltimore County). Chapter 11 discusses privacy for the semantic Web. We examine P3P and discuss ways to extend it. We also examine the privacy problems that arise through semantic Web mining and discuss approaches for privacy-preserving semantic Web mining. Chapter 12 discusses integrity and data quality for the semantic Web. How do we ensure that the information that is exchanged is of high quality? Multilevel security for a semantic Web is the subject of Chapter 13. Note that multilevel security is an aspect of confidentiality. However, we decided not to include it in Part II to give that part more focus. Managing the policies is an important aspect. Therefore, policy engineering is discussed in Chapter 14. Finally, in Chapter 15 we elaborate on the developments discussed in Section 1.2. We decided to include this in Part III mainly for completion.

Part IV discusses applications that utilize trustworthy semantic Webs. Chapter 16 discusses secure Web services that utilize semantic Web technologies. Semantic Web technologies for managing secure databases is the subject of Chapter 17. Secure semantic interoperability for heterogeneous information sources is discussed in Chapter 18. Chapter 19 discusses secure E-business applications. Chapter 20 discusses semantic Web for secure digital libraries. Chapter 21 discusses semantic Web technologies for an important application area called assured information sharing.

Part V describes special semantic Webs. In Chapter 22 we discuss domain-specific semantic Webs for financial and medical domains, among others. Trustworthy semantic Webs for geospatial data as well as sensor data are discussed in Chapter 23; in particular the work carried out on GML as well as OGC's interoperability work are studied. Chapter 24 discusses pervasive computing applications including secure mobile-sensor semantic Webs that we will call pervasive semantic Webs.

Each part begins with an introduction and ends with a conclusion. Furthermore, each of Chapters 2 through 24 starts with an overview and ends with a summary and references. Chapter 25 summarizes the book and discusses future directions. We have included three appendices. Appendix A provides an overview of data management and discusses the relationship between the texts we have writ-

ten. This has been the standard practice with all of our books. Note that although Book 7, *Database and Applications Security*, essentially ends our series of books on data management, it also begins our new series on data security. Our current book is an elaboration of Chapter 25 of Book 7. A summary of *Database and Applications Security* is given in Appendix B to give the reader a better understanding as to where we are coming from. Various standards efforts related to a semantic Web are detailed in Appendix C. This book ends with a bibliography and an index.

We have essentially developed a five-layer framework to explain the concepts better in this book. This framework is illustrated in Figure 1.7. Layer 1 is the supporting technologies layer and consists of trustworthy information systems technologies. Layer 2 is the core technologies layer that consists of the key technologies for secure semantic Webs including secure XML, RDF, ontologies, rules, integrity, privacy, and trust management. Layer 3 is the dependability layer and consists of features for privacy, trust, and integrity. Layer 4 is the applications layer and includes applications such as Web services and semantic interoperability. Layer 5 is the specialized semantic Web layer and consists of trustworthy geospatial semantic Webs and sensor Webs. Each layer uses the technologies of the lower layers. Furthermore, the technologies in Layers 1 through 5 feed into the research, products, and standards that are evolving. Figure 1.8 illustrates how Chapters 2 through 24 in this book are placed in the framework. Essentially the technologies of Parts I through V belong to Layers 1 through 5, respectively.

1.10 Next Steps

This chapter has provided an introduction to the book. We first presented a brief overview of the supporting technologies for a trustworthy semantic Web including trustworthy information systems and semantic Webs. Then we discussed secure semantic Webs and dependable semantic Webs. Applications such as semantic Web services and assured information sharing were discussed next. Finally, we discussed specialized semantic Webs. The organization of this book, detailed in Section 1.9, includes a framework for organization purposes. Our framework is a five-layer framework, and each layer is addressed in one or more parts of this book.

This book provides the information for a reader to become familiar with a secure semantic Web and trustworthy systems. We discuss some topics such as a secure semantic Web in more depth as we have carried out much research on this topic. Some other topics are less concrete such as sensor-based semantic Webs and security. In fact many of the topics we discuss are still in the research stages.

Note that one could argue that semantic Webs are not yet commercially available as a whole and therefore a book on secure semantic Webs may be somewhat premature. We feel that such a book is very timely. Even though the concepts are not mature, we have discussed many issues and solutions so that the reader has some understanding of what needs to be done to develop a secure semantic Web. Further-

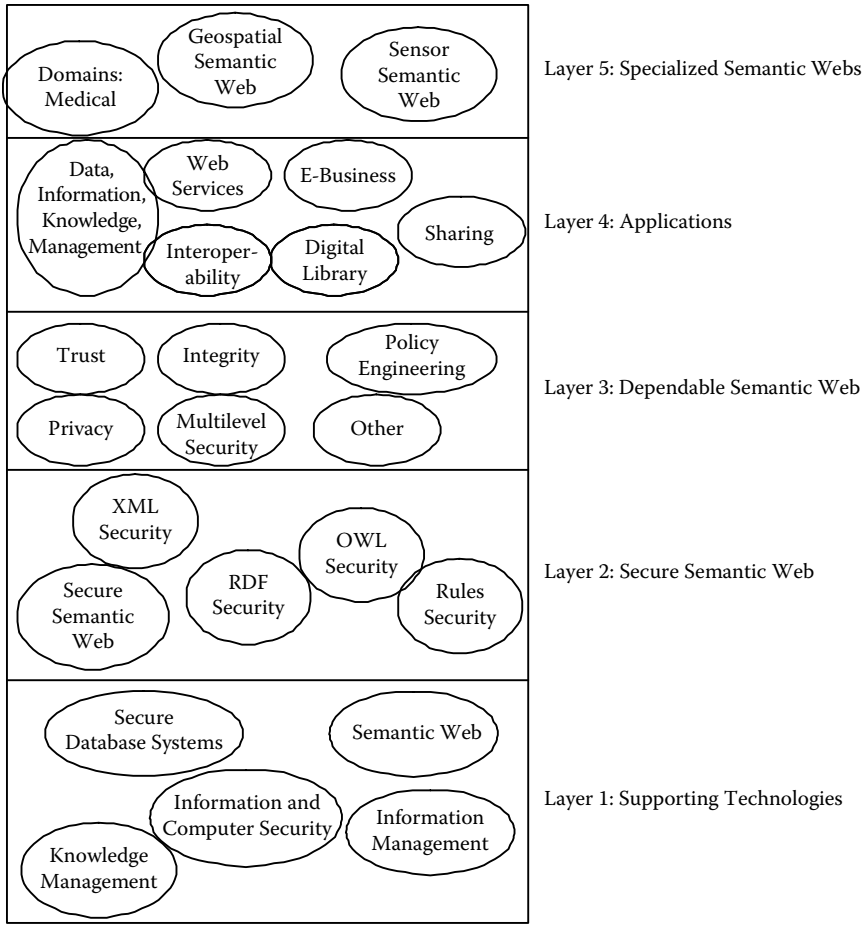

Figure 1.7 Framework for trustworthy semantic Webs.

more, as we have stated, security cannot be an afterthought. It has to be incorporated while the standards for these semantic Webs are being developed by the W3C and others. One of the main contributions of this book is raising the awareness of the importance of security and trustworthiness.

We have also given a set of exercises, intended for those who wish to pursue research in the area, at the ends of Chapters 2 through 24. To be consistent with our previous books, our purpose is to explain, especially to technical managers, what a secure semantic Web is all about. However, because of our fairly extensive research in secure information systems, we have also tried to include technical details that would help the technologists, researchers, and developers.

We provide several references that can help the reader in understanding the details of data security. My advice to the reader is to keep up with the develop-

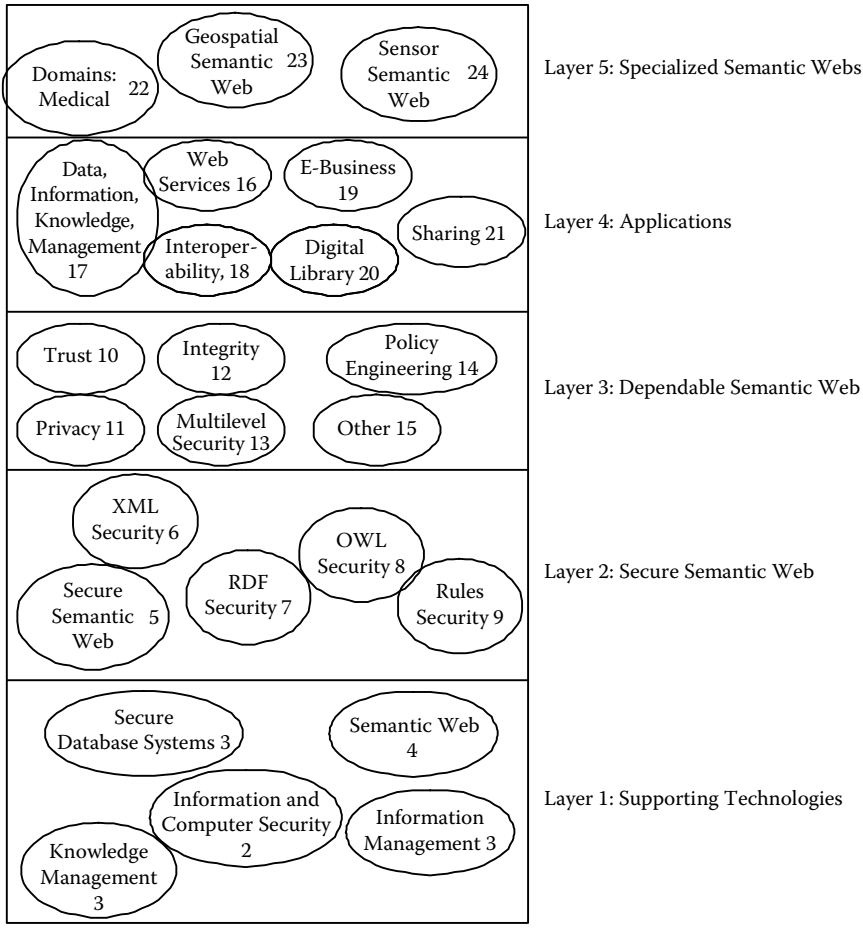

Figure 1.8 Contents of the book with respect to the framework.

ments in semantic Webs as well as in data and applications security, various data and applications security as well as information security-related conferences and workshops that are being held. Most notable is the IFIP11.3 Data and Applications security conference series. Other security conferences include the Institute of Electrical and Electronics Engineers (IEEE) Symposium on Security and Privacy, the Association for Computing Machinery (ACM) Conference on Computers and Communications Security, and the Computer Security Applications Conference. Journals include the *Journal of Computer Security*, *Computers and Security Journal*, *ACM Transactions on Information and Systems Security*, *IEEE Magazine on Security and Privacy*, *IEEE Transactions on Dependable and Secure Computing*, and the *Journal of Privacy Technologies*. Several semantic Web conferences are also being conducted. These include the International Semantic Web Symposium and the WWW