Security Impacts on Semantic Technologies in the Coming Decade

Blake Middleton, James Halbert, and Frank P. Coyle

Southern Methodist University, Dallas TX 75205, USA,
[jmiddleton,jhalbert,coyle]@smu.edu,
WWW home page: http://lyle.smu.edu/~jmiddlet

Abstract. As organizations continue to generate value using information assets, foundation technologies such as those that define the Semantic Web will increasingly be leveraged to gain competitive advantages. However following the widespread business and cultural adoption of the Internet, the negative impact of threats and the magnitude of consequences of security failures on the well being of organizations and individuals necessitates effective security services. This paper discusses semantic security topics and forecasts what semantic security impacts will be on organizations adopting Semantic Web technologies.

Keywords: semantic web, security, semantic modeling

1 Introduction

The technologies behind the Semantic Web have been maturing at a rapid pace since their initial inception [31] [36] [25]. The transition from human to machine consumption of information is poised to be as profound as the initial momentum of the commercial and cultural adoption of the Internet. Accelerating this transition is the continued success of multiple technologies; pervasive computing, ubiquitous network access, mobile communications platforms, Big Data technologies and the continued cultural adoption of all things digital. The magnitude of the data available will continue to increase. The rates will continue to rise as a tide, but there will be no ebb and flow. Semantic technologies have been getting significant attention as a way to integrate big data [1]. In parallel with the strides seen over the last decade in information technology, there have been equal strides in the capability of threat agents and malicious users. The awaiting tsunami of data to be unleashed by semantic technology will in necessity remain under the surface until security protection needs are met and risks are acceptable to data owners. This paper extends prior work in semantic web security by outlining basic assumptions from an organizational viewpoint, and provides an assessment of what security solutions will mean to security and semantic technologists. We explore the perceived impact of semantic web technologies on security services, and propose several strategy options for organizations implementing and deploying semantic technologies.

2 Semantic Web Security Foundations

Information is an important asset to organizations and should be secured. Some information may appropriately be shared in the public arena but must be protected from unauthorized modification. Other information may be appropriately shared to authorized entities within an enterprise but must be protected from disclosure or modification by unauthorized entities. Sharing information is necessary and facilitates the efficient operations of organizations. In addition to securing information, organizations must also preserve and secure knowledge such as intellectual property, trade secrets, or digital library resources. The Semantic Web facilitates the sharing of information but most current Semantic Web technologies (for example: RDF, OWL, SPARQL [32] [17]) lack the capability of adequately securing the information being shared. The current state of security with respect to XML and semantic technologies include:

- **eXtensible Markup Language (XML)** is a basic building block for the Semantic Web and must have the capability to be secured in order for layers above it to be secured. XML security has been researched by many including [3], [6], [12], [24] and the W3C has produced standards and recommendations for XML Security and has an XML Security Working Group.
- **Resource Description Framework (RDF)** adds semantic information to data that is represented syntactically in XML. Several researchers have addressed RDF security ([10], [15], [18], [20], [26], [16]) but there are unresolved issues.
- **Web Ontology Language (OWL)** is used to describe ontologies and provides additional reasoning capabilities on top of RDF. Ontologies are semantic models of a domain and are comprised of a vocabulary and rules regarding relationships among objects in the domain. Ontologies need to be secured because they contain descriptive information related to the domain and the relationships between items in the domain. As with RDF, some research related to ontologies and security has been performed ([16]) but open questions remain.
- **Security policies** in the context of the Semantic Web provide rules pertaining to access control, privacy, and trust. Policies are typically written in a natural language but can be specified using XML, RDF, or an ontology language such as OWL. Research in Semantic Web related policies has been conducted by [9], [26], [8], and [11].
- **Query Processing** Recent research ([17]) has proposed using views in SPARQL similarly to how they are used in SQL for access control. Other research ([19]) has extended SPARQL and RDF to create a new query engine that incorporates trust.
- **Secure Semantic Web Services** present security issues for both the service provider and the client. Secure Semantic Web Services have been researched by [7] and [22]. The W3C and the Organization for the Advancement of Structure Information Standards (OASIS) have both proposed standards pertaining to Secure Semantic Web Services.

- **Access Control** for the Semantic Web provides protection of information by restricting access to and or modification of information to only those entities authorized to do so. Access control applies to XML documents, RDF stores, ontologies, rules, and queries and is a standard security control used to restrict access to a resource. Access control is typically based on either the accessing entity's identity or role and the security policy in place to protect the resource. Access control has been researched by [26], [17], [13], and [26].
- **Inference** is the ability to deduce information by aggregating other information. Inference is a security issue because one may be able to aggregate information he has access to in order to infer information for which he is not authorized to access. This issue is not new with the Semantic Web as it is also a problem with database systems. Preventing inference has been proven to be impossible ([34]) so reducing the likelihood of inference is the best we can hope to accomplish. Prior research related to the inference problem has been conducted by [14], [15], [16] and [33].
- **Trust** in relation to the Semantic Web can apply to individuals, agents, or data. Research related to trust has been conducted by [21], [4], [5], [27], [28], [35], [30] and [37].
- **Provenance** refers to the history associated with information such as who created the information, who has modified the information, who has accessed the information, etc. Provenance is related to trust because information may become more or less trustworthy based upon who created or modified it. The W3C has recently completed working drafts related to provenance and other provenance research can be found in [29].

3 Semantic Web Security - Where We Are At

Considerable research efforts including the development of W3C standards and recommendations for XML security have produced a reasonable level of security maturity for XML. Progress has been made in securing RDF stores and ontologies but not to the extent that prudent organizations would risk the possible exposure of sensitive data. The security impacts of emerging technologies such as linked data methods (JSON-LD), direct database storage, and HTML encodings (RDFa and HTML microdata) have only been lightly addressed. The basic question of how to control access to RDF stores and ontologies and the proper granularity for that access still remains. Ontologies facilitate interoperation and data exchange but issues still remain here as well. When shared data is governed by disparate security policies, reconciling policy differences is wrought with peril and additional research is needed in security policy engineering and management. Finally, how do we know how much we can trust shared information? Research in provenance is maturing and should help in answering this question.

4 Security Impacts - Where We Are Going

Non-functional requirements for semantic technologies have been acknowledged for some time (see the semantic web layer cake [36]). As previously noted, aca-

demic and industry consortium efforts [25] [36] have continued to investigate potential solutions for many of the needed security services. While there is still considerable research to be completed to define the security services, it is known that there will be changes to the base semantic technology specifications. Additionally, the final implementations of the security services required will result in architectural changes to the coming secure semantic web [25].

4.1 Base Assumptions

In order to forecast the impacts of security on semantic technologies in the coming decade, it is helpful to state the assumptions regarding the fate of "semanity" now. The base semantic technology assumptions made are as follows:

Prevalent Adoption of Semantic Technologies As semantic technologies continue to mature and become secure, businesses and organizations will escalate adoption and exploitation. Domains containing unstructured and behavioral data, for instance social networking, mobility, and pervasive sensor data, will be farmed and processed aggressively. Commercial success will spur competition and technological growth to achieve competitive advantage.

A Tidal Wave of Data is Coming As computing becomes increasingly pervasive and new forms of information analysis such as Big Data become ubiquitous [2], vast quantities of data will be managed. Semantic ontologies and inferencing will aid in understanding and enhance productivity. These technologies will further fuel companies with raw material. New semantic and data technologies will further complicate how security policies are represented and implemented.

Data Drives Revenue and Reputation As knowledge management and analysis of private customer data and related information continue to personalize and increase the value of information-based services, organizational revenues will be directly affected. Likewise, private data and inferred information losses will adversely affect brand and reputation [23].

Semantic Security Solutions Lag Technology Adoption The arms race in new information analysis techniques will result in security services for semantic technologies to lag adoption. Risk of information loss will be accepted by market leaders. Due to application security control improvements, such as advances in application layer firewalls, layered defenses will mitigate some inherent risks.

Security Remains a Priority and Difficult Regardless of risks taken, organizations will continue to tout data security as their number one security priority. Threats will continue to grow in sophistication, and attacks targeted to vulnerabilities in semantic technologies will emerge. This will focus adopters towards acquisition of effective security services in the semantic layer of their information systems.

4.2 Semantic Security Use Cases

Based on the assumptions presented and analyzing both enterprise security and semantic security services, use cases have been identified (see Figure 1) that aid in understanding new operations and impacts to be expected over the next decade. These use cases impact both IT security and semantic technology vendors. The use cases introduce new tasks, new actors/roles, and new responsibilities.

Semantic Data Modeling Current tools and methods to create/model semantic information will require significant modifications. Modeling semantic data with embedded security meta-data will require interactions between a Knowledge Engineer (KE) and a System Security Engineer (SSE). New blended modeling tools will be required to support traditional KE modeling, while assisting an SSE in implementing security policies adding trust service configuration, access control and role assignments. Aggregating data originating from systems with heterogeneous security policies will remain as a challenging problem for the SSE and KE.

KnowledgeBase Hardening New capabilities/tools will be required to correctly configure and protect semantic data from inferencing and other exposures and provide secure semantic web configuration to external trust services, etc. SSE tasks will include semantic threat, vulnerability, and risk analysis during risk management processes.

KnowledgeBase Situational Awareness In order to detect attacks and potential unauthorized information changes, semantic stores will require query, inference, and audit support. Real-time graph visualization and security dashboard support will support advanced situational awareness.

KnowledgeBase Auditing Semantic data stores will require security reporting capabilities to provide proof of compliance to additional regulation that will address the personal privacy legislation affecting advanced information analysis and knowledge generation.

Web Service/KB Penetration Test In addition to semantic web service testing, specifically designed semantic store penetration test techniques will be required to validate advanced semantic issues and mitigate information loss risks.

4.3 Recommendations

While the use cases offer insight into the changes expected as security is supported in semantic technologies, both security professionals and semantic technology implementors should take heed of the following recommendations.

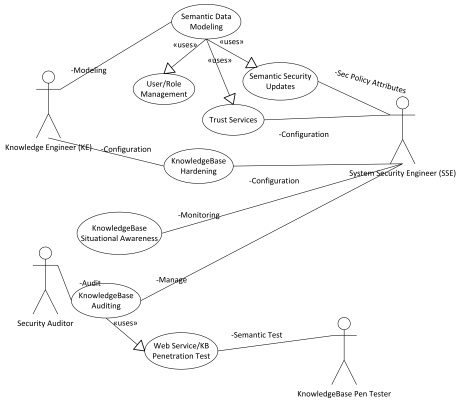

Fig. 1. Semantic Security Use Cases

- **Retooling Required** Security professionals need to become familiar with semantic technologies and semantic security services. Risk management needs to address additional capabilities and issues including information provenance, trust, inferencing, and potential information loss.
- **Certification Effort Complications** Semantic technology products will require additional levels of assurance and resiliency validation. Threats and vulnerability analysis and risk management processes will need to be performed by semantic-aware security professionals.
- **Implement Effective Security Services** Enterprise and government IT departments will continue to have stringent security requirements for operational systems. Vendors that provide effective and interoperable security services that support lower Total Cost of Ownership will be rewarded with larger market shares. Vendors should articulate their security value propositions.
- **Build in Operations Management Support** As operations visualization and performance dashboards mature, semantic technology vendors will need to develop supporting metrics to support advanced operations measurement and management.

5 Summary

This paper provides an overview of semantic security issues as they exist today. Based on current research in semantic web security and a set of educated assumptions on the information technology environment, use cases were provided illustrating potential impacts of security in the semantic web. Conclusions and recommendations were provided for both security professionals and semantic technologists providing insight for the coming decade of the semantic web.

References

1. Anderson, J., Rainie, L.: The Fate of the Semantic Web. Pew Internet and American Life Project, (2010)
2. Anderson, J., Rainie, L.: Big Data. Pew Internet and American Life Project, (2012)
3. Bertino, E., et al.: Access Control for XML Documents. Data Knowledge Engineering, Volume 43 Number 3, 2002.
4. Bertino, E., Ferrari, E., Squicciarini, A.C.: Trust-Chi: An XML Framework for Trust Negotiations. Communications of Mulitmedia and Security, 2003, pp146-157.
5. Bertino, E., Ferrari, E., Squicciarini, A.C.: Trust-X: A Peer-to-Peer Framework for Trust Establishment. IEEE Transactions on Knowledge Data Engineering, volume 16 number5, 2004, pp827-842.
6. Bertino, E., et al.: Secure Third Party Publication of XML Documents. IEEE Transactions on Knowledge and Data Engineering, 2004.
7. Bhatty R., Bertino, E., Ghafoor, A.: Trust-based Context Aware Access Control Models in Web Services. Proceedings Web Services Conference, San Diego, July 2004.
8. Bonatti, P., Olmedilla, D.: Rule-Based Policy Representation and Reasoning for the Semantic Web. Reasoning Web, LNCS, pp 240-268, 2007.
9. Bonatti, P.A., Duma, C., Fuchs, N.E., Nejdl, W., Olmedilla, D., Peer, J., Shahmehri, N.: Semantic Web Policies - a Discussion of Requirements and Research Issues. European Semantic Web Conference, Budva, Montenegro, 2006, pp712-724.
10. Carminati, B., et al.: Security for RDF. Proceedings of the DEXA Conference Workshop on Web Semantics, Zaragoza, Spain, 2004.
11. Carminati, B., Ferrari, E., Thuraisingham, B.: Using RDF for Policy Specification and Enforcement. DEXA '04 Proceedings of the Database and Expert Systems Applications, 15th International Workshop, pp 163-167, 2004.
12. Damiani, E., diVimercate, S., Paraboschi, S., Samarati, P.: Securing XML Documents. Conference on Extending Database Technology, Prague, March 2002.
13. Fabian, A., DeCoi, J., Henze, N., Koesling, A., Krause, D., Olmedilla, D.: Enabling Advanced and Context-Dependent Access Control in RDF Stores. Proceedings of the 6th International Semantic Web Conference, 2007.
14. Farkas, C., et al.: Inference Problem for the Semantic Web. Proceedings of the IFIP Conference on Data Applications Security, Colorado, August 2003.
15. Farkas, C.: Inference problem in RDF. Proceedings ACM SACMAT, 2006.
16. Farkas, C., Gowadia, V., Jain, A., Roy, D.: From XML to RDF: Syntax, Semantics, Security, and Integrity. IFIP International Federation for Information Processing, 2006, Volume 193, pp 41-55.
17. Gabillon, A., Letouzey, L.: A View Based Access Control Model for SPARQL. 2010 Fourth International Conference on Network and System Security, pp 105-112, 2010.
18. Giereth, M.: On Partial Encryption of RDF-graphs. Proceedings of the International Semantic Web Conference LNCS 3729, pp 308-322, 2005.
19. Hartig, O.: Querying Trust in RDF Data with tSPARQL. The Semantic Web: Research and Applications, LNCS, pp 5-20, 2009.
20. Jain, A., Farkas, C.: Secure Resource Description Framework:An Access Control Model. Proceedings of the 11th ACM Symposium on Access Control Models and Technologies pp 121-129. 2006.
21. Kagal, L., Finin, T.W., Joshi, A.: A Policy Based Approach to Security for the Semantic Web. International Semantic Web Conference, Sanibel Island, FL, 2003.

22. Kagal, L., Finin, T., Paolucci, M., Srinivasan, N., Sycara, K., Denker, G.: Authorization and Privacy for Semantic Web Services. IEEE Intelligent Systems Journal, vol. 19, no. 4, July 2004, pp 50-56.
23. Kark, K.: Articulating the Business Value of Information Security. Forrester Research Inc., (2009)
24. Murata, M., Tozawa, A., Kudo, M., Hada, S.: XML Access Control using Static Analysis. Proceedings of the 10th ACM Conference on Computer and Communications Security, pp73-84, 2003.
25. OASIS (Organization for the Advancement of Structured Information Standards). Oasis Web Site: http://www.oasis-open.org/
26. Reddivari, P., Finin, T., Joshi, A.: Policy Based Access Control for a RDF Store. Proceedings of the Policy Management for the Web Workshop, WWW 2005, May 2005, pp78-83.
27. Richardson, M., Agrawal, R., Domingos, P.: Trust Management for the Semantic Web. Proceeding of the Second International Semantic Web Conference, Sanibel Island, Florida, 2003.
28. Shmatikov, V., Talcott, C.: Reputation-Based Trust Management. Journal of Computer Security, 2005.
29. Simmhan, Y.L., Plale, B., Gannon, D.: A Survey of Data Provenance in e-Science. ACM SIGMOD Record, volume 34 issue 3, September 2005, pp31-36.
30. Squicciarini, A.C., Bertino, E., Ferrari, E., Ray, I.: Achieving Privacy in Trust Negotiations with an Ontonlogy-Based Approach. IEEE Transactions on Dependable Secure Computing, volume3, number 1, 2006, pp13-30.
31. Thuraisingham, B. Security Standards for the Semantic Web. Computer Standards and Interfaces 27, 2005, pp 257-268.
32. Thuraisingham, B., Parikh, P. Trustworthy Semantic Web Technologies for Secure Knowledge Management. IEEE/IFIP International Conference on Embedded and Ubiquitous Computing, 2008.
33. Thuraisingham, B., et al.: Administering the Semantic Web, Confidentiality, Privacy and Trust. Journal of Information Security and Privacy, 2006.
34. Thuraisingham, B.: Recursion Theoretic Properties of the Inference Problem. IEEE Computer Security Foundations Workshop Franconia, NH, June 1990.
35. Winsborough, W.H., Li, N.: Safety in Automated Trust Negotiation. IEEE Symposium on Security and Privacy, Oakland, CA, 2004.
36. W3C (World Wide Web Consortium).
W3C Web Site: http://www.w3.org/
37. Yu, T., Winslett, M.: A Unified Scheme for Resource Protection in Automated Trust Negotiation. IEEE Symposium on Security and Privacy, Oakland, CA., May 2003, pp110-122.