

Security Issues for the Use of Semantic Web in e-Commerce

Andreas Ekelhart¹, Stefan Fenz¹, A Min Tjoa², and Edgar R. Weippl²

¹ Secure Business Austria, A-1040 Vienna, Austria
{aekelhart, sfenz}@securityresearch.at

WWW home page: <http://www.securityresearch.at>

² Institute for Software Technology and Interactive Systems - Vienna University of Technology, A-1040 Vienna, Austria
{atjoa, eweippl}@ifs.tuwien.ac.at

WWW home page: <http://www.ifs.tuwien.ac.at>

Abstract. As the ontologies are the pivotal element of the Semantic Web in E-Commerce, it is necessary to protect the ontology's integrity and availability. In addition, both suppliers and buyers will use an ontology to store confidential knowledge pertaining to their preferences or possible substitutions for certain products. Thus, parts of an ontology will need to be kept confidential. We propose to use well established standards of XML access control. E-commerce processes require the confidentiality of customer information, the integrity of product offers and the availability of the vendors' servers. Our main contribution-the introduction of a Security Ontology-helps to structure and simulate IT security risks of e-commerce players that depend on their IT infrastructure.

1 Introduction

We emphasize on the large potential of applying the semantic web technology to electronic commerce. Autonomous or semi autonomous agents can use the semantic information to search for and compare products or suppliers and negotiate with other agents [GTM99] [TBP02] [Sch03]. Generalizing previous work we propose the following short definition for semantic e-commerce:

Semantic e-commerce is the processing of buying and selling via the semantic web.

Even though concepts of solutions already exist for years, they were not successful on the market. Thus till today information asymmetries still exist [Gup02] and one of the resulting shortcomings is the fact that the better informed buyer increasingly gets a better value for his money. Unfortunately searching is still a costly task and due to current data structures often an inefficient, economic activity. Research projects such as [ebS06] attempt to address these issues. The aim of this innovative project is to offer suppliers the option to publish their

products and services in a machine-readable language based on open-source, domain specific structures i.e. an ontology. Such semantically enriched descriptions enable intelligent software agents to query and read product information autonomously and prepare it for human customers in an appropriate way.

2 Introducing Semantic e-Commerce

2.1 Architecture

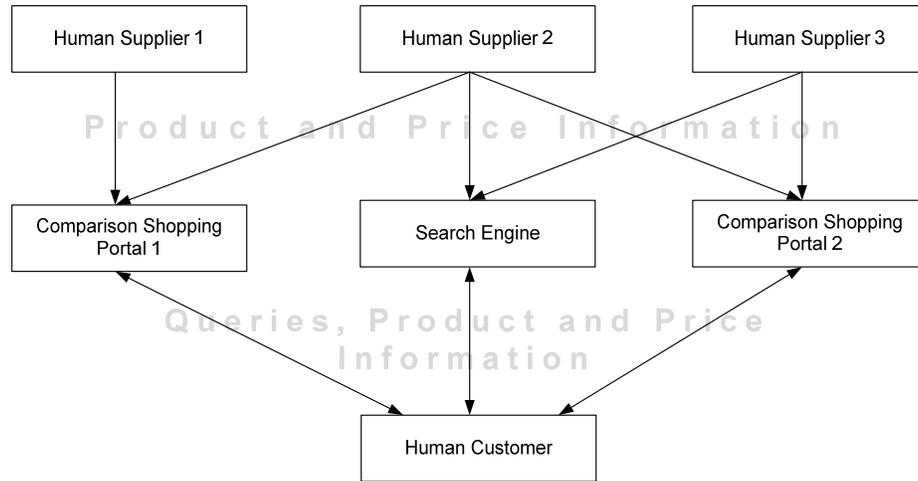


Fig. 1. e-Commerce - state of the art

Customers and suppliers are confronted with a very diversified market environment. Figure 1 shows the typical situation of a customer/supplier who intends to buy/sell a certain article over the world wide web. Compared to the conventional real life market environment, tools such as comparison shopping portals (e.g. www.geizhals.at) and search engines ease the search for the favored product and give suppliers the possibility to offer their products on a central marketplace. Despite these tools the customer is usually still overwhelmed with a big amount of offers and different product descriptions. Even though comparative-shopping-portals offer the possibility to search within specific product groups the customer still has to compare the different product descriptions to figure out which article matches his requirements most.

Figure 2 shows a possible scenario of a centralized semantic e-commerce environment. The product ontology provides as a central element the knowledge about defined product groups and their specific attributes (e.g. for mobile phones: display size, memory and organizer capabilities). The supplier agent uses the ontology data to dynamically build a user interface for the human supplier

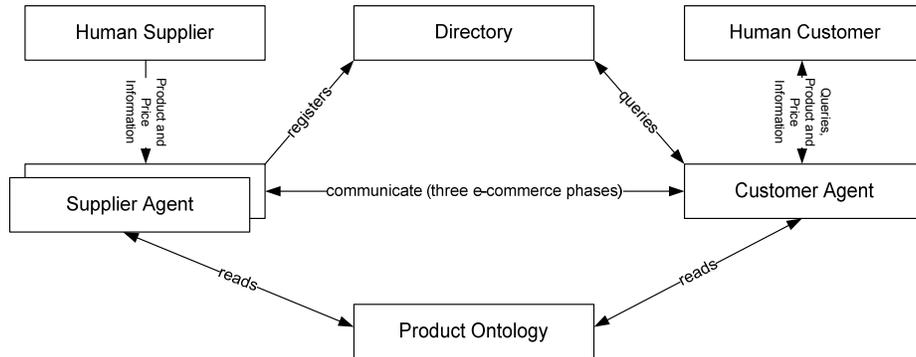


Fig. 2. e-Commerce - the centralized semantic approach

who is then able to feed the supplier agent with relevant product and price information. The last step requires the supplier agent to register itself at a central directory with its virtual location and offered product groups. On the customer side the process is almost identical. Depending on the desired product group the customer agent reads out the proper product ontology and creates a user interface which is capable to find out customer's requirements regarding a specific product. A mixture of questions and checklists could be used to find out what the customer really requires. After the requirement specification the customer agent queries a central directory to find supplier agents which offer the right product group. With a list of all available supplier agents the customer agent is able to start the communication (the three e-commerce phases) with each supplier agent.

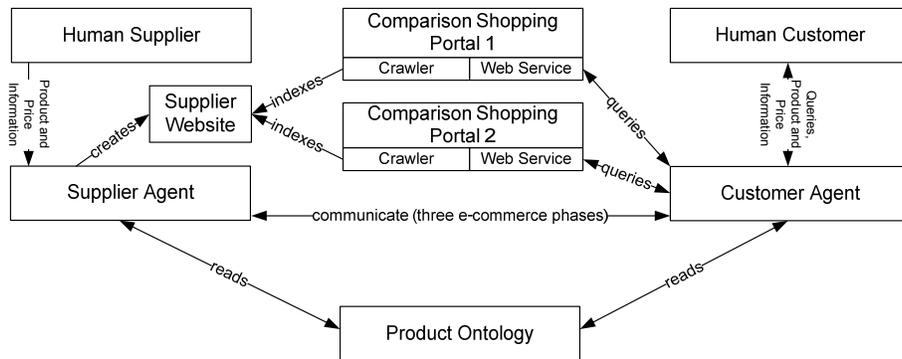


Fig. 3. e-Commerce - the decentralized semantic approach

One shortcoming of the centralized directory approach, is the fact that there has to be a central authority which maintains the directory service. With the

utilization of a central ontology and semantic (in the sense of product and price descriptions) websites a more decentralized architecture which uses web crawlers to identify possible semantic e-commerce websites will be possible (compare Figure 3). Of course these websites have to use the classification of the central product ontology to ensure compatibility with the consumer agents. In realistic terms it will not be possible that every consumer runs its own crawler that processes large parts of the world wide web. Thus some kind of services (e.g. extensions to established comparison shopping portals) which run their own crawlers have to be established and the consumer agent looks for possible supplier agents at these sites to start the three-phase e-commerce communication.

2.2 e-Commerce Phases

E-commerce transactions, which take place between businesses and customers, consist of three phases: search, negotiate and fulfillment [Pet00] [SKLQ01]. In the following, each of these phases will be discussed in detail, describing the current situation and security relevant issues:

Search Usually an e-commerce transactions starts with a user or business searching for potential trading partners. For this task two general approaches exist: (1) searching for a company with specific characteristics or (2) looking for goods with particular features and subsequently for companies which offer them. Initially all product characteristics are often not specified or not yet known and therefore this phase should result in a list of potential trading partners, each offering products of interest.

We distinguish general-purpose search engines (e.g. Google) and domain-specific portals (e.g. MEDLINEplus) on the Web as proposed by [BCJ⁺03]. In both cases, facing purely syntactic information, only keyword-based search can be conducted, which is known to be inefficient [Sch03]. The obvious need for semantic search approaches has been realized [KB04], and nowadays search portals, taking advantage of proprietary, lightweight semantic definitions, up to companies, offering sound product descriptions based on shared domain specifications in OWL [OWL04], exist. In this paper we concentrate on this last-mentioned newly approach, matchmaking by ontological product descriptions by reason that it is flexible and offers the most accurate search results. Pertaining to the semantic e-Commerce approaches, depicted in Figure 2 and Figure 3, autonomous agents carry out the search instead of the human customer itself. Initially the search parameters are provided to the agent which subsequently queries for supplier agents. Concerning the CIA triad (confidentiality, integrity and availability), ontological product descriptions and offers sometimes have to be confidential (encrypted parts for example), the integrity has to be maintained to counter fraud and availability is necessary for successful matchmaking. Security solutions regarding ontological descriptions, mostly available in XML (RDF or OWL), will be discussed in Section 3.

Negotiate Once potential business partners have been identified in the search phase, the second phase of transaction, namely electronic negotiation, starts. This is performed through an interchange of negotiation proposals describing constraints on an acceptable deal and results in an agreement (which is transformed into a legally binding contract), specifying the terms that both parties consider acceptable. These terms could include the product or service description, the price, delivery date, etc. [TBP02]

Negotiation relies on a shared terminology to guarantee efficient interactions and to avoid misunderstandings and conflicts. Ontologies can provide definitions of concepts and relations, describing the domain of interest as well as negotiation specific concepts. [SBQ⁺02] state that ontology-based negotiation approaches enable efficient, complex and unambiguous exchanges that result in business contracts.

Confidentiality and integrity are of main concern during the negotiation phase pertaining to security. Especially the exchange of private information (including credit card numbers) demands a high level of security and trust and furthermore, non-repudiation must be enforced.

Fulfillment After a contract is agreed upon, the promises set in the negotiation phase and specified in the contract are carried out. Usually automatic workflows are executed to initiate payments or delivery processes which are (automatically) monitored to control and sometimes enforce the correct fulfillment of the contracts. Automatic reasoning on contract obligation fulfillment or non-fulfillment demands formal contract definitions as well as formal transaction information to show the relevant context in which it occurs.

The fulfillment processes and corresponding resources and monitoring installations in place pose as potential targets for attacks, especially pertaining to fraud.

Agent based e-commerce aims to support the whole transaction process by autonomous means. By using sound semantic descriptions it is possible for agents, given a set of initial parameters, to find products and services automatically. Also the negotiation phase can be carried out by agents if the terms are defined and negotiating agents understand each other (using the same vocabulary, which can be achieved by common ontologies). "Intelligent", autonomous agents can unburden users in their daily, time-consuming and complex tasks and even reach better results but legal questions and security issues, including trust between agents, are a crucial point and will be discussed in Section 4.

Another aspect of (semantic) e-commerce security is the business crucial IT-environment, comprising (web-)servers hosting company information and agent services, databases with product and private user-information, ontological file storages for products and domain specific knowledge, etc. Only in a well protected and maintained IT-environment reliable and secure e-commerce can be conducted, which is often overseen, especially by small- and medium-sized enterprises. [Hau00] summarized the problems of SMEs regarding the IT-Security

aspect: (1) Smaller IT budget, relative to total budget as well as in absolute figures (2) Less IT knowledge, information technology is often looked after by employees from other departments (3) IT is not considered as important as within larger enterprises although more and more core processes are processed by IT elements (4) IT environments are not homogeneous. To overcome these problems we introduced a security ontology approach for holistic IT-infrastructure security [EFKW06] and Section 5 refers to the technical details of the security ontology approach.

3 Security within ontologies

Ontologies are at the focus of our approach. We thus need to protect their confidentiality and integrity.

3.1 Access Control

While the proposed product ontology remains public to ensure a shared vocabulary among the market participants, each supplier derives its own ontology, filled with concrete values such as price and delivery information, which has to be secured against unauthorized reading or writing attempts.

Due to the fact that each OWL- or RDF-based ontology uses XML as surface syntax [OWL04], access control models for XML documents can be also applied to OWL- or RDF-based ontologies.

Research in the field of XML access control models is already mature and several approaches for securing ontologies already exist: [FCG04] propose the concept of security views which provide for each user group an XML view consisting of that information that the users are authorized to access. The approach requires a XML query-execution engine that implements the DTD-based access control model. [DdVPS02] present a language for the specification of access restrictions on XML-based files and the corresponding system architecture for access control which should enforce its usage. The proposed XML Access Control Processor (ACP) takes as input a valid XML document requested by the user and the XML Access Sheet listing the associated access authorizations at the instance level. The ACP generates a valid XML document, including only the information the user is allowed to access [DdVPS02]. [BF02] extend the approach by fine-grained XML document encryption and proper key distribution mechanisms to ensure confidentiality within shared XML documents. The Semantic Access Control Language (SACL) proposed by [QA03] is suitable to express concept-level access authorizations within OWL-based ontologies.

Such mechanisms are suitable for enhanced and implemented ontology access control approaches and especially in the semantic e-commerce field with its various actors and different relationships we have to enforce a strong access control technology.

3.2 Integrity

Since the very central product-ontology and the derived supplier ontologies with their price information play an important role in a possible semantic e-commerce scenario there have to be proper mechanisms which ensure the integrity of the ontology structure and its content. Especially the derived supplier ontologies act as a storage for price and delivery information which are used by the agents at the negotiation phase. Therefore the integrity of these data elements is crucial for the long-term establishment of semantic e-commerce systems and due to the XML-based syntax of OWL- and RDF-ontologies we are able to use established standards such as XML Digital Signature [xml02] and XML Key Management Systems (XKMS) [xkm01] to ensure data integrity.

4 Trust issues

Trust is one of the main issues pertaining to e-Commerce, based on the following reasons: 1) a potential buyer has no physical access to the product of interest, 2) seller or buyer may not abide by the agreement reached at the electronic marketplace [Zac99]. Agent based systems add another layer of indirection between a buyer and a seller, resulting in a more complex framework and new trust issues.

[Gam00] defines trust as *a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action, both before he can monitor such action (or independently of his capacity ever to be able to monitor it) and in a context in which it affects his own action.*

We distinguish between two fundamental trust models which are (1) either built on an agent's direct experience of an interaction partner (interaction trust) (2) or reports provided by third parties about their experiences with a partner (witness reputation) [HJS06]. Nowadays, taking eBay [EBA07] as an example, traders receive a feedback (such as +1, 0 or -1) for their reliability in each auction. Furthermore textual comments can be submitted to describe the customer's experience pertaining to the seller. Besides trust based on previous transactions (if they exist), customer feedback (feedback scores and comments) is a crucial element of trust in a seller. According to companies, independent third party evaluation and certification is another possibility to convince customers of their trustworthiness. Concerning to the centralized semantic e-Commerce approach in Figure 2, we identified the following trust issues and possible methods of resolution:

In the first place the human interaction partner has to trust his agent, viz the software system - the underlying lines of code created by the system developer. The agent has to fulfill the promised functionality and should not have any vulnerabilities. Certified providers as well as certified agent systems help to establish the trust needed.

Each agent has to "know" its communication partner before reputation can be considered, thus authentication mechanisms have to be implemented. As a

principle an agent has to provide his identity, usually in form of a public key certificate, issued by a certification authority (CA).

If agents have the ability to purchase products (on the behalf of the agent's principal), the risks can be minimized by only granting a limited payment capability [CPV03]. Furthermore, if digital signatures are required, the use of the private key should be limited to the agent. [RS99] for example propose proxy certificates: in this approach only a new, lifetime limited key pair is handed to the agent. This makes it difficult for malicious hosts to discover the private key before the certificate expires. Additionally, arbitrary transactions can be constrained. To avoid contract repudiation—especially users denying that an agent acted on their behalf—the user instruction parameters should be collected and digitally signed.

The Directory service, shown in Figure 2, should only register and subsequently mediate trustworthy agents. Besides looking for available certificates, customer agents have the possibility to rate their experiences with supplier agents. SPORAS [Zac99] is a possible model for an agent based, centralized rating system.

5 The Security Ontology

Beside the very deep going aspects of securing ontologies and communication between various agents, we also have to consider the IT-Security regarding the company's physical environment. Servers hosting company information and agent services, databases with private user-information or files containing ontological product information have to be secured to ensure a reliable and secure e-commerce service. Especially small- and medium-sized enterprises often oversee the need for a holistic IT-Security approach and thus we developed a Security Ontology [EFKW06] to provide a proper knowledge base about threats and the corresponding countermeasures. In [EFKW07] we extended the threat simulation approach with risk analysis methods to improve quantitative risk analysis methods. The current section summarizes the research results and proposes the implementation of the Security Ontology to enhance the overall IT-Security level.

The most important parts of the Security Ontology are represented by the sub-ontologies *Threat*, *ThreatPrevention* and *Infrastructure*:

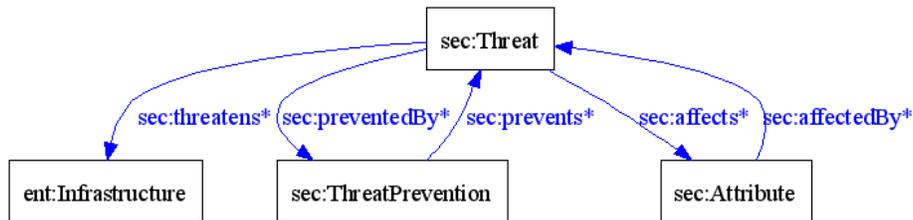


Fig. 4. Sub-ontology: Threat

Figure 4 shows the threat ontology with its corresponding relations: (1) To model the threats which endanger certain infrastructure elements we introduced the *sec:threatens* relation (every threat threatens n infrastructure elements) (2) Of course we want to mitigate the threats and so we created the *sec:preventedBy* and *sec:prevents* relation respectively (3) To enable companies to optimize their IT-Security approach to certain IT-Security attributes such as confidentiality or availability we assigned affected attributes to each threat by the *sec:affects* and its inverse relation.

Figure 5 shows the security ontology's infrastructure area. The building, with its corresponding floors and rooms, can be described using the infrastructure framework. To map the entire building plan exactly on the security ontology, each room is described by its position within the building. The ontology *knows* in which building and on which floor a certain room is located. The attributes *ent:nextToRoomHorizontal* and *ent:nextToRoomVertical* describe the exact location of each room. Each instance of *ent:ITAndTelecommunication* and *sec:TechnicalThreatPrevention* is located in a particular room. A room can, of course, also contain more concepts. The current ontology uses a flexible and easily extendable structure: additional concepts can be included without effort. The concept *ent:TechnicalThreatPrevention* is subdivided into *ent:CounterMeasure* and *ent:Detector*, which are used to model detectors (fire, smoke, noise, etc.) and their corresponding countermeasures (fire extinguisher, alarm system, etc.).

Figure 6 shows the prototype with its four main user interface elements: (1) Selection of a threat: The user is able to choose a certain threat and the SecOnt Manager shows the impact of that threat (2) Threatened infrastructure: The ontology provides an extendable framework for various infrastructure elements to enable the user to create instances of concrete and real infrastructure elements which enables the ontology to show which infrastructure elements are threatened by a certain threat scenario (3) Affected attributes: Works like the threatened infrastructure where the ontology *knows* which threats are affecting certain security attributes (4) Recommendations: Are the most important part for the user, because it gives concrete recommendations to prevent a certain threat. Figure 6 shows an example application for the fire threat and we see that the ontology has to store the whole infrastructure, including the building with its floors and rooms, to make location-based recommendations possible.

So why are we using an ontology instead of a database solution which has various advantages over a file-based ontology? The main advantage of an ontology is the possibility of inferring new knowledge by utilizing a reasoning engine which considers existing facts and rules.

$$\begin{aligned}
 & sec : affectsOS(?x, ?z) \wedge ent : hasOS(?y, ?z) \wedge sec : AntiVirusProgram(?c) \\
 & \wedge \neg sec : installedOn(?c, ?y) \wedge sec : prevents(?c, ?x) \\
 & \rightarrow sec : threatens(?x, ?y)
 \end{aligned}
 \tag{1}$$

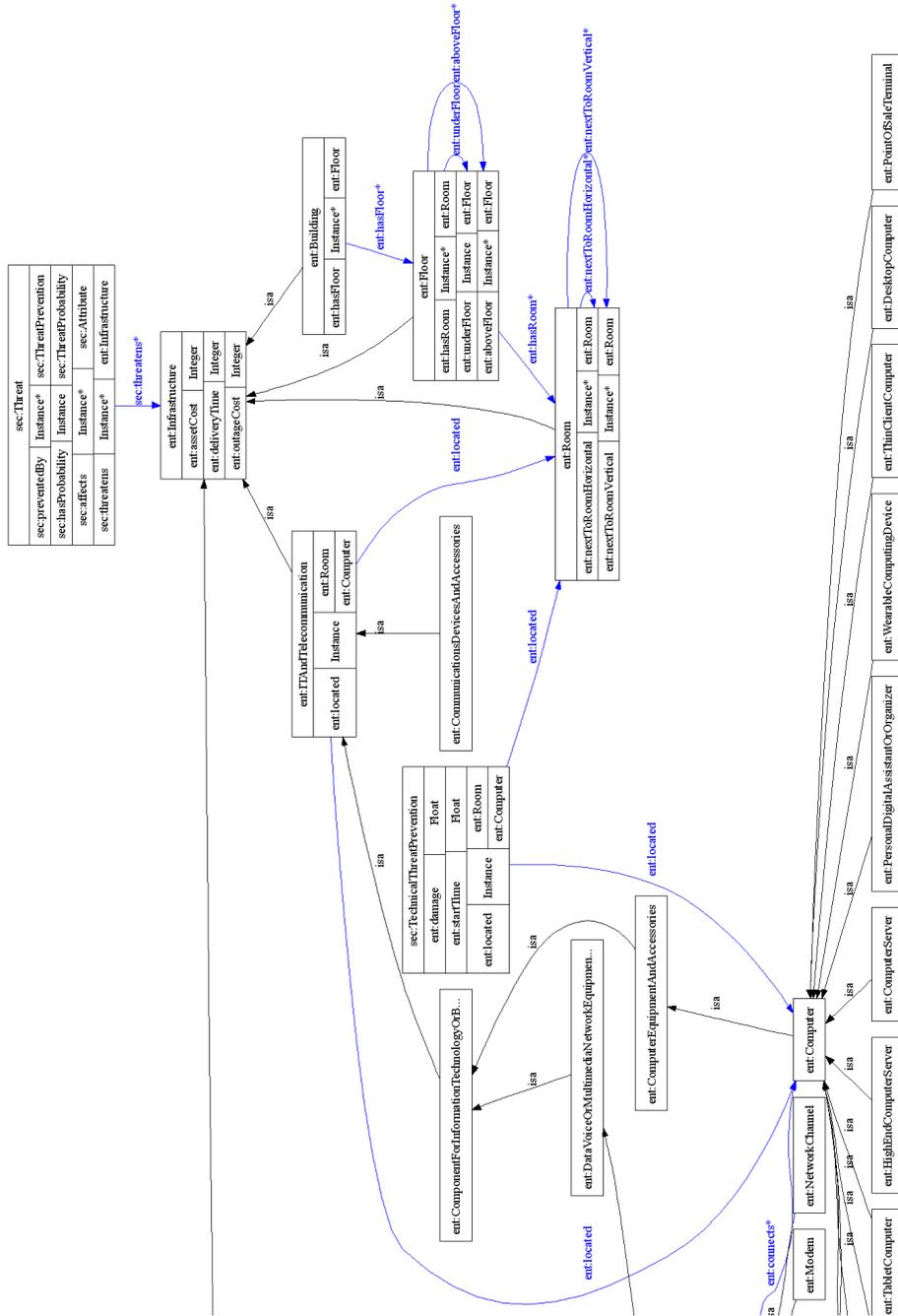


Fig. 5. Sub-ontology: Infrastructure

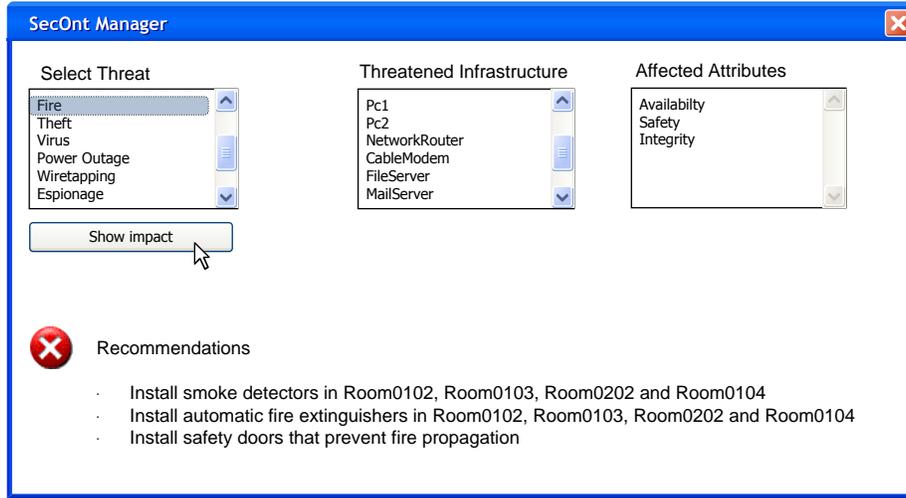


Fig. 6. SecOnt Manager Prototype

Equation 1 illustrates a possible axiom which formalizes the *sec:threatens* relation between a computer virus and a computer device. First *sec:affectsOS* determines which operating systems are endangered by a certain virus and in the second step *ent:hasOS* looks up for all computers and their corresponding operating systems. Variable *?c* stores all available anti virus programs and looks with $\neg sec:installedOn$ for computers that have not installed such a program. With *sec:prevents* it is possible to determine which anti virus protection is useful to a certain virus and so the ontology, equipped with a proper reasoning engine, is able to identify those computers that are directly threatened by a certain virus.

6 Conclusion

In this paper we covered the three phases of e-business (search, negotiation, and fulfillment) and investigated how semantic information and ontologies can support and improve these processes. Moreover, we explored the case for protecting the ontology which is the central element of this approach. Mechanisms of XML access control are used to protect the confidentiality, integrity and availability of ontologies. Finally, we presented how the introduced Security Ontology can be used to secure all assets required by IT-centered companies to ensure CIA (confidentiality, integrity and availability) of information processed in their business processes.

Acknowledgements

This work was performed at the Research Center Secure Business Austria funded by the Federal Ministry of Economics and Labor of the Republic of Austria (BMWA) and the federal province of Vienna.

References

- [BCJ⁺03] Suresh K. Bhavnani, Bichakjian K. Christopher, Timothy M. Johnson, Roderick J. Little, Frederick A. Peck, Jennifer L. Schwartz, and Victor J. Strecher. Strategy hubs: next-generation domain portals with search procedures. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 393–400, New York, NY, USA, 2003. ACM Press.
- [BF02] Elisa Bertino and Elena Ferrari. Secure and selective dissemination of xml documents. *ACM Trans. Inf. Syst. Secur.*, 5(3):290–331, 2002.
- [CPV03] Joris Claessens, Bart Preneel, and Joos Vandewalle. (how) can mobile agents do secure electronic transactions on untrusted hosts? a survey of the security issues and the current solutions. *ACM Trans. Inter. Tech.*, 3(1):28–48, 2003.
- [DdVPS02] Ernesto Damiani, Sabrina De Capitani di Vimercati, Stefano Paraboschi, and Pierangela Samarati. A fine-grained access control system for xml documents. *ACM Trans. Inf. Syst. Secur.*, 5(2):169–202, 2002.
- [EBA07] ebay. <http://www.ebay.com/>, 2007.
- [ebS06] ebsemantics. www.ebsemantics.org, 2006.
- [EFKW06] Andreas Ekelhart, Stefan Fenz, Markus Klemen, and Edgar R. Weippl. Security ontology: Simulating threats to corporate assets. In Aditya Bagchi and Vijayalakshmi Atluri, editors, *Information Systems Security*, volume 4332 of *Lecture Notes in Computer Science*, pages 249–259. Springer, Dec 2006.
- [EFKW07] Andreas Ekelhart, Stefan Fenz, Markus Klemen, and Edgar R. Weippl. Security ontologies: Improving quantitative risk analysis. In *Proceedings of the 40th Hawaii International Conference on System Sciences (HICSS 2007)*, Jan 2007.
- [FCG04] Wenfei Fan, Chee-Yong Chan, and Minos Garofalakis. Secure xml querying with security views. In *SIGMOD '04: Proceedings of the 2004 ACM SIGMOD international conference on Management of data*, pages 587–598, New York, NY, USA, 2004. ACM Press.
- [Gam00] Diego Gambetta. Can we trust trust? In Diego Gambetta, editor, *Trust: Making and Breaking Cooperative Relations*, chapter 13, pages 213–237. Published Online, 2000.
- [GTM99] Robert J. Glushko, Jay M. Tenenbaum, and Bart Meltzer. An xml framework for agent-based e-commerce. *Commun. ACM*, 42(3):106–ff., 1999.
- [Gup02] *Reduction of price dispersion through Semantic E-commerce*, volume 55 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2002.
- [Hau00] Hans Eduard Hauser. Smes in germany, facts and figures 2000. Institut für Mittelstandsforschung, Bonn, 2000.

- [HJS06] Trung Dong Huynh, Nicholas R. Jennings, and Nigel R. Shadbolt. Certified reputation: how an agent can trust a stranger. In *AAMAS '06: Proceedings of the fifth international joint conference on Autonomous agents and multiagent systems*, pages 1217–1224, New York, NY, USA, 2006. ACM Press.
- [KB04] Mark Klein and Abraham Bernstein. Toward high-precision service retrieval. *IEEE Internet Computing*, 8(1):30–36, 2004.
- [OWL04] Owl web ontology language. <http://www.w3.org/TR/owl-features/>, 2004.
- [Pet00] Ralf Peters. Elektronische märkte und automatisierte verhandlungen. *Wirtschaftsinformatik*, 42(5):413–421, 2000.
- [QA03] Li Qin and Vijayalakshmi Atluri. Concept-level access control for the semantic web. In *XMLSEC '03: Proceedings of the 2003 ACM workshop on XML security*, pages 94–103, New York, NY, USA, 2003. ACM Press.
- [RS99] Artur Romao and Miguel Mira Da Silva. Proxy certificates: A mechanism for delegating digital signature power to mobile agents. In *IAT99 Workshop on Agents in Electronic Commerce*, 1999.
- [SBQ⁺02] Mareike Schoop, Andreas Becks, Christoph Quix, Thomas Burwick, Christoph Engels, and Matthias Jarke. Enhancing decision and negotiation support in enterprise networks through semantic web technologies. In *XML Technologien für das Semantic Web - XSW 2002, Proceedings zum Workshop*, pages 161–167. GI, 2002.
- [Sch03] Mareike Schoop. Semantic web technology for electronic commerce. In *Proceedings of the The Tenth Research Symposium on Emerging Electronic Markets*, 2003.
- [SKLQ01] Mareike Schoop, Joerg Koeller, Thomas List, and Christoph Quix. A three-phase model of electronic marketplaces for software components in chemical engineering. In *I3E '01: Proceedings of the IFIP Conference on Towards The E-Society*, pages 507–522, Deventer, The Netherlands, The Netherlands, 2001. Kluwer, B.V.
- [TBP02] D. Trastour, C. Bartolini, and C. Priest. Semantic web support for the business-to-business e-commerce lifecycle, 2002.
- [xkm01] Xml key management specification (xkms). <http://www.w3.org/TR/xkms/>, 2001.
- [xml02] Xml-signature syntax and processing. <http://www.w3.org/TR/xmlsig-core/>, 2002.
- [Zac99] Giorgos Zacharia. Trust management through reputation mechanisms. In *Third International Conference on Autonomous Agents (Agents '99)*, New York, NY, USA, May 1999. ACM Press.