CHALLENGES FOR ACCESS CONTROL IN KNOWLEDGE FEDERATIONS

Sergei Evdokimov, Benjamin Fabian and Steffen Kunz
Institute of Information Systems, Humboldt-Universität zu Berlin, Spandauer Straße 1, 10178 Berlin, Germany

Keywords: Security, Access Control, Knowledge Federation, Semantic Technologies.

Abstract: Based on ongoing work in the Aletheia project on knowledge federation for the product lifecycle, we present the most urgent challenges for designing access control solutions for semantic-based knowledge federations across multiple companies.

1 INTRODUCTION

The expansion of the Internet of Things (IoT) and the Internet of Services (IoS) will result in rapidly growing volumes of distributed and heterogeneous information and knowledge. Providing the right knowledge, at the right time, in the right place, and to the right people will be one of the major challenges of the future Internet (Benjamins et al., 2008; Ameri and Duta, 2005). Today's information systems are not able to deal with this challenge because most of them work in isolated domains instead of supporting a federation of knowledge, and due to a lack of semantic integration and reasoning capabilities.

This is also true for knowledge created throughout the lifecycle of a product. Corporate information systems are virtually not capable of federating the necessary product information, though a huge amount of information is already provided by information sources such as business applications, databases, data warehouses, but also from the Web and Web 2.0 (wikis, blogs, social networks etc.), web services, and smart item infrastructures using Radio Frequency Identification (RFID), sensor networks, and the emerging EPCglobal Network. Product knowledge created from these information sources is crucial for many business processes and would offer improvement potentials for various stakeholders in the value chain of a product.

The Aletheia project (Aletheia, 2009) – funded by the German Federal Ministry of Education and Research and organized as a consortium of industry and academic partners – tries to face those challenges by developing an information system that is able to federate all information created throughout the lifecycle of a product. Through semantic technologies and logical reasoning, the federated information is transformed into product specific knowledge that can be accessed by various stakeholders of the lifecycle.

Orthogonal to the issue of federation and orchestration, the problem of assuring security in process and information flows, especially controlling which parties are allowed to have access to what knowledge, is a big challenge.

In this paper, we especially address the security issues of the Aletheia system. In particular, the questions of access control (AC) are investigated. The structure of the paper is as follows: First, Aletheia's application scenarios along the lifecycle of a product are described (Section 2). In Section 3, a draft of the Aletheia architecture is presented. Section 4 focuses on the security requirements of the Aletheia system. The challenges for AC are presented in Section 5. Finally, in Section 6 current and future work are presented.

2 ALETHEIA PROJECT

The aim of the Aletheia project is to build a prototype for federating and creating business-relevant knowledge across all phases of a product's lifecycle. We adopted the definition by Ameri and Duta (2005) in the following, stating "knowledge is evaluated and organized information that can be used purposefully in a problem solving process".

Evdokimov S., Fabian B. and Kunz S. (2009).
CHALLENGES FOR ACCESS CONTROL IN KNOWLEDGE FEDERATIONS.
In *Proceedings of the International Conference on Knowledge Management and Information Sharing*, pages 224-229
DOI: 10.5220/0002331602240229

In order to assess the requirements for Aletheia, the different phases of a product's lifecycle (cf. Ameri and Duta, 2005) have been analyzed, including product requirements analysis, development & design, manufacturing, sales, operation, maintenance, recycling, as well as the connections between those phases, represented by the logistics processes (see Figure 1).

Aletheia's application scenarios for knowledge federation in the product lifecycle as well as the corresponding requirements have been identified with the help of several industrial project partners, who conducted interviews in their operational business units, complemented by personal face-to-face interviews. With the help of these interviews, we identified different use cases in each of these eight stages of the product lifecycle. This allowed us to develop a holistic description, interpretation, and understanding of relationships and the processes in the product lifecycle phases. From these uses cases, we derived requirements, which were collected according to the specification of the Volere Template (Robertson and Robertson, 2006). This approach helped us to reduce possible misunderstandings during the process of data collection resulting in broader understanding of the requirements and challenges.

In the following, we will present the current state of the requirements analysis and the desired functionality of the Aletheia applications.

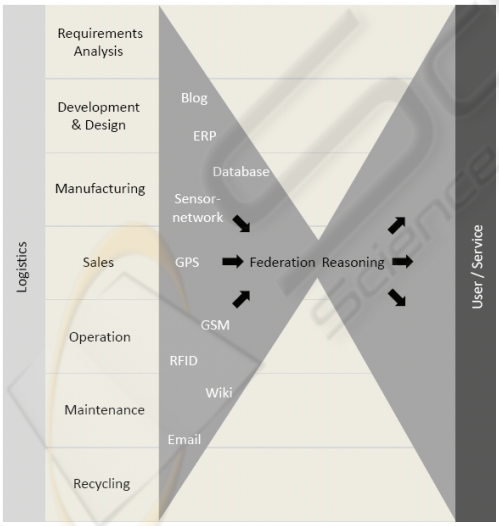

Figure 1: Federated Knowledge for the Product Lifecycle.

2.1 Customer Requirements Analysis

During this phase of the product lifecycle, the basis for the subsequent development & design phase is provided. Knowledge about the individual product lifecycle phases of current, older, or similar products, e.g., about customer satisfaction or problems during operation, can have a major impact on the requirements analysis of a new product. Through Aletheia, important product knowledge about current, older, or similar products could be federated and provided to the product requirements analysts.

2.2 Development & Design

The processes in this phase are rarely standardized, but highly individual. Accordingly, most practices and knowledge of the product developers and designers are implicit and difficult to share. Aletheia could support this phase by documenting the design processes and approaches (e.g. the choice of tools for designing a product) of the individual developers and designers, thus giving their colleagues access to their implicit knowledge and fostering knowledge-sharing among them.

2.3 Manufacturing

Generally speaking, we can distinguish between two major forms of production: mass and individualized production. In the case of mass production, knowledge does not play a major role because processes are highly standardized and automated. In contrast to mass production, individualized production requires frequent changes of the production processes and practices. Accordingly, in case of individualized production the same problems as in the development and design phase arise, i.e., to make implicit individual knowledge explicit and accessible to other employees. Aletheia could provide best manufacturing practices or recommend toolsets for certain individualized products.

2.4 Sales

E-commerce and online shopping have gained increasing importance in the last few years. Nevertheless, many online shops do not provide sufficient information about indirect product attributes that can be easily obtained in regular shops – e.g., how loud the fan of a laptop is, or how the material of a certain dress feels like. Aletheia could deliver more knowledge about indirect product attributes by extracting and federating information from distributed and highly heterogeneous sources (e.g., blogs or wikis).

2.5 Operation

Especially for products with long life spans, the information that is generated in the operation phase of the product lifecycle can be extremely valuable, e.g., software products in the B2C market or turbines in the B2B market. Knowledge created from error messages or log files is extremely valuable. This knowledge could be used for monitoring the operability of a product during the operation phase, but it is also important for other down or upstream phases.

2.6 Maintenance

During the maintenance product lifecycle phase, it is important to support the maintenance processes by providing historical as well as up-to-date knowledge about a product. Examples include historical information from the operation phase, or information about previous maintenance jobs, as well as up-to-date information about problems with suppliers' parts and the corresponding solutions. Aletheia could provide the necessary knowledge base to realize these benefits.

2.7 Recycling

The recycling phase of the product lifecycle is increasingly gaining attention. Especially the recirculation and recycling of old products through the manufacturer is a major issue, due to stricter environmental laws. Furthermore, detailed knowledge about the materials incorporated in the product would help to keep valuable materials in circulation rather than disposing them. Knowledge from all previous phases of the product lifecycle, provided by Aletheia, would considerably improve current practices.

2.8 Logistics

Though logistics cannot be considered as a phase of the product lifecycle, it builds the bridge between the other seven lifecycle phases. In logistics, the major focus is on the tracking and tracing of products with location-aware technology, as well as gathering context-aware information (e.g., with the help of sensors). This knowledge is valuable for many stakeholders in the value chain of a product. Since the amount of the generated data is tremendous and stakeholders who need access to this knowledge are distributed across company borders, smart reasoning and federation technologies could be of great use.

3 ALETHEIA ARCHITECTURE

The application areas of the Aletheia system presented in Section 2 assume that the system is maintained and accessed by a variety of stakeholders participating in different phases of a product's lifecycle. A natural way to achieve such functionality is to implement the system using a decentralized service-oriented architecture (SOA), in which Aletheia's components are implemented as web services (see Figure 2).

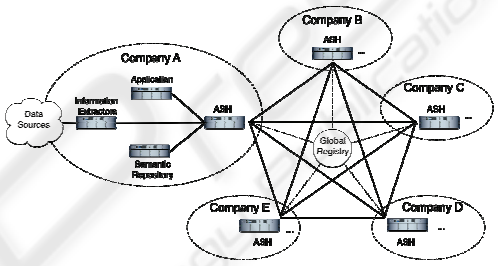

Figure 2: Aletheia Architecture.

The core component of the Aletheia system is the Aletheia Service Hub (ASH) that provides the main functionalities. It constitutes the technical basis for inter-organizational communication using the Aletheia services, which besides semantic-aware query mechanisms also include different types of Information Extractors (IE), adapted to the specific source environments. The IEs, as well as individual stakeholder-specific applications, connect to the ASH in order to pull (or push) information from (respectively to) the Aletheia system.

As displayed in Figure 2, the ASH serves as a gateway from a stakeholder's local domain to other remote domains. In order to leverage the information of other domains, each ASH connects to a global registry that contains information about all ASHs in the Aletheia system, as well as public information about their local information sources.

4 SECURITY REQUIREMENTS

Aletheia will enable knowledge sharing and federation between several stakeholders of multiple product life cycles. However, in real-world settings, not every stakeholder should be granted complete access to every piece of information a company

possesses about a certain product. On the contrary, there will be a high demand on keeping certain information confidential, and share it only on a "need-to-know" basis, which is also referred to as the established least privilege paradigm of information security (Ferraiolo et al., 2007, p.5). For example, production costs or the failure rate of certain products may be kept confidential from shops or consumers, while the latter is of high interest to service technicians using Aletheia as a product source for maintenance and repair.

An important special case is the flow of personal data through the system, where national data protection laws govern special confidentiality requirements for data about individuals (e.g., data drawn from Customer Relationship Management Systems). Further, in some cases users may want to stay anonymous in Aletheia to avoid a profiling of their buying habits by untrusted companies. Depending on access control policies, filtering and perturbation mechanisms in the information flow could provide anonymity. Like with confidentiality, a similar reasoning applies to the integrity of product information: only authorized entities should be able to modify it.

An established method for satisfying confidentiality and integrity requirements is to provide mechanisms for AC: detailed policies can determine who has access to what kind of information, possibly taking also different contexts into account.

The currently established model for access control is based on roles (RBAC, Figure 3): Users are mapped to roles, which in turn are mapped to permission sets (Ferraiolo et al., 1999; Sandhu et al., 2000; Ferraiolo et al., 2007; Neumann and Strembeck, 2002). This enables the creation of role hierarchies and permission inheritance, as well as an easier maintenance, and improved change management of the resulting policies in typical business environments where the user population changes quite frequently.

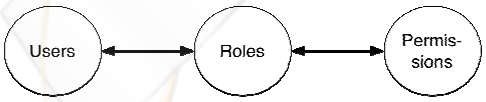

Figure 3: RBAC Relationships (Ferraiolo et al., 2007).

Further, the service-oriented architecture of the Aletheia system implies that the components of the system are communicating with each other by exchanging messages over networks, especially the Internet. This poses the requirement to ensure the confidentiality and integrity of these messages.

In the following, we discuss the important challenges we encountered so far during the process of designing an access control solution for Aletheia.

5 CHALLENGES FOR ACCESS CONTROL

The distributed nature of Aletheia and a sensitive character of data that such a system can contain make it extremely important to consider security requirements when planning the system architecture and proceeding with the developments process.

However, the complexity of the system, the heterogeneity of its data sources, and the fact that the data is stored with semantic annotations make an implementation of these requirements a challenging task. In Figure 4, we list the security topics we identified as being highly relevant to AC in the Aletheia system. We briefly discuss each of them and the challenges they introduce.

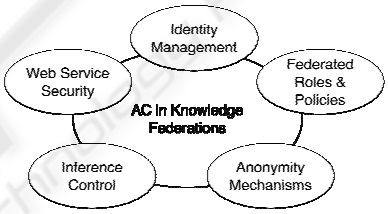

Figure 4: Important Security Areas

5.1 Federated Identity Management

Aletheia's support of AC implies that the system has to distinguish between its users. Usually this is achieved by deploying an identity management system to assign each user an identity. The identity management system is responsible for authenticating and tracking identities of the users during their operations and interactions with the Aletheia system.

The users can belong to different security domains (e.g., to different companies). To manage identities of such users, a new dedicated identity management system could be deployed, thus creating a new security domain dedicated exclusively to Aletheia. Though straightforward and simple, this solution, however, first, does not follow Aletheia's distributed paradigm, and second, results in users having extra identities (e.g., a user name and a password) that are used exclusively for accessing Aletheia.

An alternative approach would be to use a federated identity management system, relying, for example, on the OASIS SAML standard (OASIS,

2009). A federated identity system enables the portability of the existing identities between several security domains, thus reducing the number of identities a user has to keep track of. On the other hand, building a federation embracing a number of different identity management systems can entail significant effort.

5.2 Federated Roles and Policies

Different stakeholders of Aletheia can have different AC policies and have different roles structures. Influential standards for defining the policies and the roles have been developed by the OASIS consortium. An important example is the access control policy language XACML (OASIS, 2009).

In case of distributed environments, where the stakeholders use XACML for defining their AC polices, it can be convenient to generate a general policy that would be consistent with the policies of the stakeholders. For that purpose, an approach similar to the one proposed in Mazzoleni et al. (2008) could be considered.

5.3 Web Service Security

To satisfy the requirement for ensuring the confidentiality and integrity of messages exchanged between Aletheia components, Web Services-Security (WSS) protocols can be used. WSS is a generally accepted approach for ensuring end-to-end security in a web service architecture (OASIS, 2009).

Alternatively, a more lightweight approach based on the Transport Layer Security (TLS) protocol can be implemented (see RFC 5246, Dierks and Rescorla, 2008). Compared to WSS, TLS can significantly reduce the performance overhead, but, on the other hand, security options provided by the protocol are limited and ensure only point-to-point security. A decision about which approach will be implemented in Aletheia will be made once the final architecture of the system is released.

5.4 Ontology Access and Inference Control

A key functionality of Aletheia is the integration of very heterogeneous information, also gathered from RFID and the IoT. Further, Aletheia aims to combine semantic reasoning with information extracted from the Web (2.0). Information stored in the Aletheia system will be provided with machine-understandable annotations. These annotations describe semantics and allow to perform reasoning on the stored information. The domain, on which the reasoning is performed, is defined by an ontology, which defines objects, their properties and relations, thus providing a vocabulary and semantics describing the domain. In the Aletheia system, the domain will describe a product lifecycle and the repositories of the stakeholders will contain statements about entities of this domain. Since, in theory, parts of the ontology itself may be subject to confidentiality requirements, it must be investigated if AC for the ontology has to be developed.

Further, the requirements for the confidentiality and integrity of the knowledge stored in the repositories demand mechanisms that could control access to the stored statements, prevent non-authorized inferences, and define access rules for newly generated statements. Existing work in this area appears to be scant. An introduction and survey on many aspects of semantic web security is presented in (Thuraisingham, 2008). The challenging topic of inference control has been presented in Farkas and Jajodia (2002), but they focus mainly on statistical inference.

Today, to our knowledge, no existing semantic data store provides an appropriate AC covering all of those aspects. Typically, the AC support is restricted to the data store as a whole without a possibility to define fine-grained rules covering also inferences. Fortunately, in the recent years a number of works discussing these issues have been published, including Reddivari et al. (2005), Jain and Farkas (2006), Abel et al. (2007), Knechtel and Hladik (2008) – some of which support the AC policies and decision procedure itself by semantic technologies.

However, although these works describe approaches that could enable a comprehensive access control for semantic repositories, currently there is no off-the-shelf solution that could be adopted for the needs of the Aletheia system.

5.5 Anonymity

As was mentioned earlier, Aletheia can process personal and personally identifiable information, which due to national regulations or for the sake of protecting the identities of the customer should be stored or presented in an anonymized form. K-anonymity (Sweeney, 2002) and l-diversity (Machanavajjhala et al., 2007) are established principles in database theory to describe the degree of anonymity a person has in a given data set, though they do not offer perfect protection. Their

application to the field of semantic data integration can pose new challenges and require additional research.

6 CURRENT AND FUTURE WORK

Aletheia has started in 2008 as part of a research alliance on federating and integrating knowledge in a future IoT and IoS, together with the partner projects SemProM and ADiWa. At the time of this writing, the Aletheia system architecture is not finalized, yet. Important current discussions include the construction of a single, system-wide ontology vs. a loosely coupled federation of local ontologies, the existence of central entities like the global Aletheia registry, and possible distribution mechanisms. Many of those choices will affect the security mechanisms we can deploy.

Starting out from a prototype on federated identity and web-service access control that is consisting of just a few partners, we will investigate how existing mechanisms can be extended to cover semantic-aware access control and anonymization. Special emphasis will be placed on the best locations for these mechanisms within the information flow, and their impact on scalability and performance, especially comparing online vs. offline processing for inference protection and anonymity control.

In parallel, we will apply Neumann's and Strembeck's (2002) approach for modelling RBAC policies using their toolset, and investigate legal requirements on data protection.

7 CONCLUSION

In this paper, we presented the main challenges we encountered so far during ongoing work on providing access control solutions for Aletheia, and provided an outlook on current and future work in that area.

REFERENCES

Abel, F., Coi, J. L. D., Henze, N., Koesling, A. W., Krause, D., and Olmedilla, 2007. D. Enabling Advanced and Context-Dependent Access Control in RDF Stores. In *Proc. 6th International Semantic Web Conference*.

Aletheia, 2009. Aletheia Project Web Site URL: http://www.aletheia-projekt.de/.

Ameri, F. and Dutta, D., 2005. Product Lifecycle Management: Closing the Knowledge Loops. In *Computer-Aided Design and Applications* 2(5), pp. 577-590.

Benjamins, V. R., Davies, J., Baeza-Yates, R., Mika, P., Zaragoza, H., Greaves, M.; Gomez-Perez, J. M., Contreras, J., Domingue, J. and Fensel, D, 2008. Near-term prospects for semantic technologies. In *IEEE Intelligent Systems* 23(1), pp. 76-88.

Dierks, T. and Rescorla, E., 2008. The Transport Layer Security (TLS) Protocol Version 1.2, *Request for Comments* 5246. IETF.

Farkas, C. and Jajodia, S., 2002. The Inference Problem: A survey. *SIGKDD Explor. Newsl.* 4, 2, pp. 6-11.

Ferraiolo, D., Barkley, J., and Kuhn, D.R., 1999. A Role-Based Access Control Model and Reference Implementation within a Corporate Intranet. In *ACM Transactions on Information and System Security (TISSEC)*, 2(1), pp. 34-64.

Ferraiolo, D., Kuhn, D.R. and Chandramouli, R., 2007. Role-Based Access Control, Artech House. 2nd ed.

Jain, A. and Farkas, C., 2006. Secure Resource Description Framework: An access control model. In *Proc. of 11th ACM Symposium on Access Control Models and Technologies (SACMAT'06)*. ACM Press.

Knechtel, M. and Hladik, J., 2008. RBAC Authorization Decision with DL Reasoning. In *Proc. of the IADIS International Conference WWW/Internet (ICWI '08)*.

Machanavajjhala, A., Kifer, D., Gehrke, J., and Venkitasubramaniam, M. 2007. L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data* 1(1). ACM Press.

Mazzoleni, P., Crispo, B., Sivasubramanian, S. and Bertino, E., 2008. XACML Policy Integration Algorithms. In *ACM Transactions on Information and System Security (TISSEC)* 11 (1), pp. 1-29.

Neumann, G., Strembeck, M., 2002. A Scenario-driven Role Engineering Process for Functional RBAC Roles, In: *Proc. of 7th ACM Symposium on Access Control Models and Technologies (SACMAT).*

OASIS, 2009. OASIS Standards Web Site, URL: http://www.oasis-open.org/committees/

Reddivari, P., Finin, T. and Joshi A., 2005. Policy Based Access Control for a RDF Store. Proc. *Policy Management for the Web Workshop*. W3C, pp. 78-83.

Robertson, S. and Robertson, J., 2006. *Mastering the Requirements Process*, Addison-Wesley Professional.

Sandhu R., Ferraiolo, D. and Kuhn, D.R., 2000. The NIST Model for Role Based Access Control: Toward a Unified Standard. In *Proc. 5th ACM Workshop on Role Based Access Control*. ACM Press, pp. 47-63.

Sweeney, L. 2002. k-Anonymity: A model for protecting privacy. In *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 10 (5), pp. 557-570.

Thuraisingham, B., 2008. *Building Trustworthy Semantic Webs*, Auerbach Publications.