Applying the Semantic Web Layers to Access Control*

Mariemma I. Yagüe, Antonio Maña, Javier López, José M. Troya
Computer Science Department, University of Malaga, Spain
{yague, amg, jlm, troya}@lcc.uma.es

Abstract

The Semantic Web, also known as the Web of meaning, is considered the new generation of the Web. Its objective is to enable computers and people to work in cooperation. A requisite for this is encoding data in forms that make web contents (meaning, semantics) more understandable by algorithmic means. In this paper, we present the application of Semantic Web concepts and technologies to the access control area. The Semantic Access Control Model (SAC) uses different layers of metadata to take advantage of the semantics of the different components relevant for the access decision. We have developed a practical application of this access control model based on a specific language, denominated Semantic Policy Language (SPL), for the description of access criteria. This work demonstrates how the semantic web concepts and its layers infrastructure may play an important role in many relevant fields, such as the case of access control and authorization fields.

1 Introduction

The problem of semantic interoperation has been extensively studied. This problem appears when different applications mean different things by similar terms. Semantic heterogeneity is closely tied to the context-dependent interpretations of the concepts represented. Today, interoperability appears as the main challenge to address concerning authorization for open and distributed systems. Access control is a critical component in many environments. Access to systems and resources has to be controlled in a safe way. Nowadays, security issues for semantic web services are becoming more important. This is the reason the Semantic Web has included the access control area among the objectives of its Advanced Development (SAWD) projects. Our work is based on the definition of semantic models upon the components of an access control model to reach interoperability through the semantic integration in heterogeneous and distributed environments.

*Work partially supported by the E.U. through project IST 2001-32446

Specifically, we are concerned with the semantic integration of external authorization entities in the applications and to provide a distributed and scalable framework supporting advanced authorization and access control schemes in an efficient way. In this paper we present an access control model that addresses the aforementioned problems. This model is based on the use of semantic descriptions of the authorization entities; separation of the attribute certification and the authorization management functions, following the layers infrastructure of the Semantic Web. Section 2 states the fundamentals of the access control problem. Section 3 highlights the basis of semantic modelling through XML metadata. The fundamentals of the Semantic Access Control Model are presented in section 4, along with a detailed description of its semantic layers. Finally, section 5 underlines some concluding results and future research lines.

2 The problem of Access Control

When considering the security requirements of many distributed applications, authorization often emerges as a central element in the design of the whole security system [8], because of authorization is the source of the trust chain. Therefore, many security properties are determined by the flexibility, trustworthiness and expressiveness of the authorization scheme. Access control is the mechanism that allows owners of resources to define, manage and enforce access conditions applicable to each resource [7]. Both concepts are related since access control will usually consider authorizations as the basis to produce the access decision.

The shift from centralized to distributed systems and applications poses new requirements in both authorization and access control systems. Moreover, the popularisation of heterogeneous and open systems, such as Digital Libraries, Electronic Commerce, Web Services and Grid Computing, is introducing even more demanding requirements. In the case of centralized systems, the same entity is responsible for the assignment of attributes or privileges to clients (Authorization) and the evaluation of the access requests to determine whether they must be granted or not (Access Control). All the information required to analyse and evaluate

the privileges is stored and managed locally in the same system where the resources reside. The most relevant problem that this scheme presents when applied to open distributed systems is the lack of interoperability. It is not reasonable to expect that heterogeneous systems for different purposes and under control of different parties will be able to define a common homogeneous set of authorization criteria. Let's review some of the characteristics of these new systems and applications that are relevant to the design of the access control model.

Heterogeneity. First of all, it is important to keep in mind that, in open distributed systems it is frequent to have a large number of stakeholders or owners of resources with very different policies and interests. A large number of previously unknown clients that are impossible to classify in advance is also predictable. Moreover, resources found in distributed systems are intrinsically heterogeneous. Heterogeneity affects not only to the type of resource but also to the format, origin, validity, etc. This heterogeneity of resources, clients and owners implies very disparate security requirements and access control criteria.

Interoperability. It has been mentioned that the authorization approach in most of current systems relies on locally-issued credentials related to user identity. This type of credentials presents many drawbacks, but the most important is that they are not interoperable. Taking into account security, scalability and interoperability, the separation of the certification of attributes and access control management responsibilities is widely accepted as a scalable and flexible solution. The external authorization infrastructure is known as PMI. The main entities of a PMI, known as Source of Authorizations (SOAs), issue attribute certificates. Access control systems select which SOAs to trust and which combination of attributes to use as access criteria. To achieve interoperability, a mechanism to convey the semantics of the attributes certified by the SOAs to the access control systems is required.

Flexibility. Due to the heterogeneity and taking into account that our model is designed to be applied in open distributed systems where numerous specific systems will coexist and interoperate, flexibility appears as one of the most important goals to achieve. The model must be flexible enough to be applicable in different scenarios with few or no changes. o Scalability. In the systems that we are considering we deal with very large numbers of resources, access policies, systems, clients and attributes. Therefore, the scalability of the scheme is very important. To achieve this scalability a fully distributed scheme is mandatory.

Dynamism. The access control model must be capable of adapting itself to frequent changes in different parameters such as access criteria, client attributes, environment conditions, resources available, etc. To avoid management overload due to the control of changes, the model must adapt in a transparent and automatic way to these changes.

The basic concepts upon which the access control model is based determine the flexibility of the model to adapt to different environments and systems. Several access control models have been developed based on different schemes. It is important to realize that the existing access control models were developed for closed environments. Consequently, they are built on the basis of modeling the environments that motivated their development. Let's review these models.

Discretionary Access Control (DAC) was designed for multi-user databases and systems with a few, previously known, users. Changes were rare and all resources were under control of a single entity. Access controlled based on the identity of the requestor and on access rules stating what requestors are (or are not) allowed to do [2].

Mandatory Access Control (MAC) had its origins in military environments where the number of users can be quite high, but with a static, linearly hierarchic classification of these users. The model is based on the definition of a series of security levels and the assignment of levels to resources and users. MAC policies control access based on mandated regulations determined by a central authority [6].

Role-based Access Control (RBAC) is inspired in the business world. The development of RBAC coincides with the advent of corporate intranets. Corporations are usually hierarchically structured and access permissions depend on the position of the user in the hierarchy, i.e. the role played by the user. RBAC policies control access depending on the roles that users play within the system and on rules stating what accesses are allowed to users in given roles [1].

Among the previous models RBAC is commonly considered a mature and flexible technology. Consequently, it is the most popular paradigm in use today. The main problem with role based access control is that the mechanisms are built on three predefined concepts: "user", "role" and "group". The definition of roles and the grouping of users can facilitate management, specially in corporate information systems, because roles and groups fit naturally in the organizational structures of the companies. However, when applied to some new and more general access control scenarios, these concepts are somewhat artificial. A more general approach is needed in these new environments.

Groups are a specific use of a more general tool: the attribute. Groups are usually defined based on the values of attributes (position, :.). Other attributes, such as identity, are even built into most of the access control models. The identity is a useful attribute, but it should not be a built-in component of a general model. The static grouping of users of RBAC can suffice in corporate systems, but it is not flexible enough to cope with the requirements of more dynamic environments where the structure of groups can not be foreseen by the administrators of the access control system. In these scenarios each resource may possibly need

a different group structure and access control policy. New resources are incorporated to the system continuously and policies for a given resource may change frequently.

By considering attributes to be the basis of the access control model we can develop a very flexible and open model that is able to be used in most scenarios. In fact, MAC, DAC and RBAC schemes can be specified using the attribute-based approach. In [5] we proposed a modular an dynamic approach based on the separation of the access control criteria from the rules of allocation of policies to resources. We called this scheme Dynamic Access Control. The new model that we present is called Semantic Access Control (SAC) because it complements the use of attributes as the building block of the model with the use of metadata to represent the semantics of the different elements.

3 Semantic Modelling with XML Metadata

XML is a data model designed to provide flexibility and interoperability among different applications and systems. Although usually XML is believed to be a language to represent meaning, this is not completely true. The symbols of XML language does not have any formal semantics for the computer. It is the human user who brings meaning to the tag names, such as `<email>`, `<fax>`, ... For the computer, an XML document describes the structure of the information, that is, its syntax. A language can be understood by a computer if it has associated semantics: the symbols and structures of the language must refer to an underlying model because meaning exists only in relation to something.

However, XML is the basis of new technologies for the formal description of semantics of Web information. Metadata or 'data about data' represents the foundation for achieving a great number of functional requirements in environments such as digital libraries, application integration, or discovery of web resources. The use of XML-related technologies, such as XML Schema, RDF and RDF Schema [3], for the definition of semantic models makes possible that both, humans and machines, take advantage of the potential of the available information.

The possibility of automating the processing of semantic information is a big challenge for the resolution of many relevant problems. This is the case of semantic interoperability. One of our objectives in this work is to reach interoperability through semantic integration in distributed and heterogeneous environments. We think the development of mechanisms for the semantic integration in distributed environments where heterogeneity is common, implies the development of semantic models supported by metadata infrastructures. Therefore, we propose an access control model based on the semantic modelling of its different components, as we will show in the following section.

4 A Semantic Access Control Model

The access control model developed has been called Semantic Access Control (SAC) because semantics are the basis of the access conditions and its design follows a semantic approach. The SAC model is based on the semantic properties of the resources to be controlled, properties of the clients that request access to them, semantics about the context and finally, semantics about the attribute certificates trusted by the access control system.

In the development of the SAC model, we have considered the operation of several independent access control systems and authorization entities. In SAC, the access control to resources is independent of their location. The identification of the user or client is not mandatory. On one hand, the client possess a set of attributes and, on the other hand, the access control to resources is based on the specification of a set of attributes that the client has to present to gain the access to them. For interoperability and security reasons, client attributes must be digitally signed (in the form of an attribute certificate) by a trusted certification entity, external to the access control management system. The independence of the certification of attributes function is the key to the interoperability because it allows attributes to be safely communicated avoiding the necessity of being locally emitted by the system administrator. Additionally, this approach avoids the registration phase of the client, and the emission of a client attribute repeatedly for each access control system. For this approach to be secure, a mechanism to establish the trust between these access control systems and the authorization entities is required. We have addressed this problem using semantic information about the certifications issued by each authorization entity.

As we have yet mentioned, one of the main characteristics of the SAC model is that, opposed to traditional schemes, the attributes required to access a resource may depend on the semantic properties of the resources. The allocation of the policy corresponding to a resource is not based on the storage structure of the resources but on the semantic properties of the resources. Of course, it is also possible to consider the structure of storage.

The approach followed in SAC enables semantic validation of access control criteria. SAC is developed to facilitate the management of the access control system, while guaranteeing simplicity, correction and safety.

SAC has been implemented on the basis of a language to specify the access control criteria and the semantic integration of external authorization entities [10]. This language, called Semantic Policy Language (SPL), is based on the semantic properties about the resources to be accessed and about the context. SPL applies traditional concepts of modularity, parameterisation and abstraction in order to provide simplicity and flexibility to the difficult task

of specifying access control criteria. The modular definition of SPL policies implies the separation of specification in three parts; that is, access control criteria, allocation of policies to resources and semantic information (properties about resources and context). Additionally, SPL makes possible the abstraction of access control components and, as a consequence, the ability to reuse these access control components. All the previous properties help the reduction of the complexity of management. Moreover, the use of semantic information about the context allows the administrator to include relevant contextual considerations in a transparent manner, also helping the semantic validation task.

4.1 Layers of the Semantic Access Control Model

The fundamentals of SAC, shown in Fig. 1, are the definition of several metadata models, described in the following subsections, at different layers of the semantic web. Each component of SAC represents the semantic model of a component of the access control system. Semantic properties contained in that metamodels are used for the specification of access control criteria, dynamic policy allocation, parameter instantiation and policy validation processes.

On one hand, SPL access control policies take advantage of the different metadata models (Policy, PAS and SRR) for its creation and syntactic validation (Structure level). Additionally, these models are essential for the semantic and contextual validation of the policies, because they enable us to perform inference processes and formal validation of the SPL specifications (Logical and Inference levels). At the same time, semantics of the resources, represented in the SRR model, are used for the dynamic allocation of policies to resources, and for the instantiation of parameters in policies. Finally, the semantic integration of external certification entities into the access control system, is achieved by means of the SOAD model, which establishes the trust between different authorization entities and access control systems. All SPL documents are defined to conform to XML-Schema templates that facilitate the creation of the specifications, allowing their automatic syntactic validation.

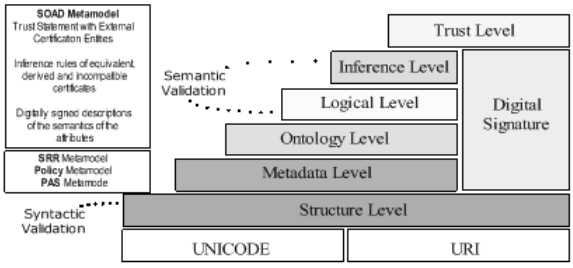

Figure 1. Layers of the SAC Model

4.2 Policy Metamodel

This model of metadata represents the semantics of access control criteria. A policy is described by a set of access rules, each one consisting in sets of attributes. This metamodel is part of the Semantic Policy Language specification, defined to capture the semantics of the different components that participate in the definition of access control policies. SPL policies can include locally defined components as well as imported elements. The ability to import elements enables the modular composition of policies based on the XPath standard. An SPL Policy is composed of a set of access rule elements, each one defining a particular combination of attribute certificates required to gain access, associated with an optional set of actions to be performed before access is granted [5]. The Policy Metamodel is at the Structure Level of the semantic model of SAC.

4.3 PAS Metamodel

This model represents the semantics about the allocation of policies to resources. That is, the Policy Applicability Specification (PAS) model is used to locate the right policy for each resource, based on the relevant properties of the resources. The dynamic policy allocation relies on a rich set of metadata about the resources, which is represented by the Secured Resource Representation model.

Therefore, the PAS documents provides an expressive way to relate policies to resources, either explicitly or based on the metadata about the objects (e.g. type of content, owner, price, etc.). PAS documents include three main elements: policy, objects and instantiation. The policy element indicates which policy is applicable to the specified objects. Optionally, operation elements can be used in PAS to define which operations of the target resource are controlled by the declared policy, allowing a finer grained access control. In case no operation element is included, the policy is applicable to all of the resource operations. The instantiation element describes the mechanism to instantiate parameters in the policies. The PAS Metamodel is located at the Structure Level of the semantic model of SAC.

4.4 SRR Metamodel

The Secured Resource Representation (SRR) model represents semantics about the resources. In this way, resources in PAS specifications are defined by their location and conditions to be fulfilled by the semantics of these resources. The SRR metamodel is a simple and powerful mechanism to describe properties about resources. Properties described in SRR documents are part of the SPL specification and are used for the instantiation of policies and PAS, and to locate the applicable policies. The SRR metamodel is located at the Metadata Level of the SAC model.

4.5 SOAD Metamodel

The Source of Authorization Description model represents the semantics of the different attributes that are certified by an authorization entity. SOADs are XML documents, protected by digital signatures, that state a series of facts about the certification system. Metadata contained in SOADs represent semantics of the different attributes that are certified by the authorization entity, including names, descriptions and relations among attributes. Moreover, the SOADs define rules enabling the derivation of new attributes, specifically incompatible, equivalent or derived ones. The semantic information represented in this model about the attributes certified by each SOA is also used to assist the security administrators in the creation of access control policies. Moreover, this semantic information allows the detection of possible inconsistencies in our SPL policies during the semantic validation process.

The SOAD model is the key to achieve the necessary interoperability because it represents the mechanism to establish the trust between the client system and the authorization entities. The SOAD model describes the semantics of the different attribute certificates and enables the integration of external authorization entities at the trust level. Each client system selects which authorization entities to trust and which combination of attributes to use. Therefore, the SOAD metamodel reaches the higher levels of the Semantic Access Control Model Infrastructrure (Logical, Inference and Trust levels).

5 Conclusions and Future Work

Following the Semantic Web vision, this paper shows the benefits of the application of the concepts of the Semantic Web to a very relevant field. The access control in open, heterogeneous and distributed systems poses important challenges making it a perfect scenario to demonstrate the potential of metadata infrastructures based on the Semantic Web concepts. The application of Semantic Web technologies has been the origin of the SAC Model.

The SAC model is scalable, applicable to different environments with heterogeneous and complex access criteria and avoids the need of a registration phase. Moreover, it covers other access control models. An infrastructure implementing this access control model, along with autonomous enforcement mechanisms called XSCD (XML-based Secure Content Distribution) has been developed,. This infrastructure has been successfully applied to information commerce [5]. Furthermore, this infrastructure provides distributed access control management and enforcement, as well as secure content distribution in digital libraries [9]. Another interesting application scenario for SAC is Web Services, where SAC achieves the desired semantic interoperability [10]. The SOAD metadata model has been applied to the semantic integration of an infrastructure of authorization entities in the CORBA architecture [4]. The SOAD metadata model along the semantic validation algorithms is extensively studied in [9].

In conclusion, the semantic approach of SAC is the foundation to achieve semantic interoperability among the different components of access control systems. SAC reaches the highest layers of the Semantic Web (Logical, Inference and Trust). The SOAD metadata model describes inference rules for deducting new information (incompatible, equivalent or derived attribute certificates) and supports the semantic validation of policies, providing proofs of the correctness of the access control policies. Additionally, the SOAD metamodel enables trusted interoperation between access control systems and external authorization entities.

We are also working on the development of additional semantic models to enable secure delegation of attribute certificates. The extension of the Semantic Policy Language with additional digital rights specification, along with semantic models for its management is under consideration.

References

[1] D. Ferraiolo and D. Kuhn. Role based access control. In *15th NIST-NSA National Computer Security Conference*, 1992.

[2] B. W. Lampson. Protection. *Computer Networks*, 8(1):18–24, 1974.

[3] O. Lassila and R. Swick. Resource description framework (rdf). Technical Report W3C Recommendation 1999-02-22, W3C, 1999.

[4] J. López, A. Maña, J. Ortega, E. Pimentel, J. Troya, and M. Yagüe. Integrating pmi services in corba applications. *Computer Standards and Interfaces*, 25(4):391–409, 2003.

[5] J. López, A. Maña, and M. Yagüe. Xml-based distributed access control system. In *EC-Web'02*, volume 2455 of *LNCS*. Springer-Verlag, 2002.

[6] X. Qian and T. Lunt. A mac policy framework for multilevel relational databases. *IEEE Transactions on Knowledge and Data Engineering*, 8(1):1–14, 1996.

[7] P. Samarati and S. de Capitani di Vimercati. Access control: Policies, models, and mechanisms. In *FOSAD 2000*, volume 2171 of *LNCS*, pages 137–196. Springer-Verlag, 2001.

[8] T. Woo and S. Lam. Designing a distributed authorization service. In *Proc. of IEEE INFOCOM*, volume 2437 of *LNCS*, pages 227–245. Springer-Verlag, 1998.

[9] M. Yagüe, A. Maña, J. López, E. Pimentel, and J. Troya. A secure solution for commercial digital libraries. *Online Information Review*, 2003(3), 2003.

[10] M. Yagüe and J. Troya. Semantic approach for access control in web services. In *Euroweb'02*, eWiC. British Computer Society, 2002.