Designing XML Security Services for Biodiversity Networks

Robert Tolksdorf[1], Lutz Suhrbier[2], Ekaterina Langer[3]

Freie Universität Berlin, Networked Information Systems

Institut für Informatik, Takustraße 9, D-14195 Berlin, http://nbi.inf.fu-berlin.de

[1]tolk@inf.fu-berlin.de, [2]suhrbier@inf.fu-berlin.de, [3]atanasso@inf.fu-berlin.de

Designing XML Security Services for Biodiversity Networks

Summary

We are working on the integration of security services into a biodiversity network system within the scope of the projects SYNTHESYS (EU funded) and GBIF-D (funded by the Federal Ministry of Education and Research BMBF). This paper reports on analysis of user requirements including national specifics and a concept for our rights management and access control.

1 Introduction

In this paper we report on the requirements analysis regarding the integration of security services into the European and German networks for biodiversity. This work is done within the scope of the projects SYNTHESYS [SYNWeb] and GBIF-D [OBWeb] funded by the EU's Sixth Framework Programme and the German Federal Ministry of Education and Research respectively. The main goal of these projects is to create an integrated network infrastructure for making biodiversity data available via Internet. Our work concerns the enhancement of this infrastructure through rights management, access control and other security services.

Technically the biodiversity data are stored in heterogeneous distributed databases. They are integrated by a wrapper component and a XML based protocol developed in the predecessor project BioCASE [BioWeb]. Due to the XML based request response protocol and data exchange schema, we decided to use XML standards for the development of rights management and access control components and to provide security for the communication.

In this document we first give an overview of the technical preconditions. Then we outline our approach to investigate the user demands and analyse the results in a requirements specification. Furthermore, we introduce XML based standards relevant for the design of the security services. Then we describe our concepts for rights management and access control. Finally the outlook gives an overview of future developments regarding communication security aspects like authentication of data sources, integrity and partial confidentiality of transmitted data.

2 Technical Preconditions

The security components to be developed have to be integrated in a pre-existing software architecture developed in the prior project BioCASE [BioWeb]. Thereby, biodiversity data are accessible via a XML based request response protocol – the BioCASE protocol [DöGü03]. This protocol abstracts from the real data to be queried for or transmitted and defines activities or methods understood by a database wrapper or a client application. On the provider domain a CGI/XML database interface written in Python (PyWrapper) enables the connection of an arbitrarily structured database to the BioCASE network. The PyWrapper implements the BioCASE protocol to respond to compatible queries in a standard XML format like ABCD [BABWeb]. On the client side, the Unit Loader is a freely available Java API from the Bio-

CASE developers providing the necessary BioCASE communication protocol functionality to client application developers.

3 Analysis of Users Demand

A requirements specification was performed to get information concerning the needs and constraints of the biodiversity network system. It was based on a questionnaire given to prospective users. The usage of a questionnaire possesses multiple advantages relating to analysis purposes: It saves resources, enables direct comparison of different opinions in single aspects, and provides the chance to establish a lasting dialog for the whole development process. Particularly the last aspect appears considerably important regarding the further development process of security services. Forwarding of questionnaires integrates as many partners as possible in the discussion process, and provides the required input for the development process.

Within the “Framework for European Services in Telemedicine” (FEST) [FESWeb] a question set [FEST95] was designed to broadly cover all relevant aspects regarding the integration of new components into existing environments in the telemedicine domain. This question set was taken as a basis for our questionnaire. The questions were adapted to the biodiversity domain and discussed with a small group of test users. We decided to subdivide the questionnaire into two parts. The first part should contain preliminary and general questions, which will become more specialised and refined in further parts. The final version of the first questionnaire was then sent by email to 93 partners within the projects SYNTHESYS, BioCASE and GBIF-D.

The analysis of the responses shows that the activities of the responding partners are spanning a broad and diversified spectrum of tasks and capabilities. The variety of internal and external data flows does not correspond to a more or less common scheme, hence requiring the implementation of a specialized solution.

From the functional point of view the most wanted features are the allocation of access rights and access control mechanisms. A relatively high demand also exists for the ownership attribution and data integrity during the transmission of documents. Further requested is the confidentiality of sensitive data (e.g. location of endangered species) and topics regarding intellectual property rights of documents. Minor demand exists for features like the installation of an account system and a project wide portal system.

From the organisational point of view the implementation of a hierarchical rights management and access control architecture is necessary, but the final determination of valid access rights policies remains in the hands of every institution. Nevertheless, every institution must be enabled to inherit policies from higher hierarchical levels (e.g. regional, national or project wide nodes) and thus add it to their local access rights decision process.

An open and probably political question concerns organisational aspects of the assignment of responsibilities for the definition and distribution of corresponding high-level policies and also for the maintenance of a project wide user and/or provider certification authority. The installation of one or more certification authorities, to build up a Public Key Infrastructure, will be mandatory regarding the implementation of access control and authenticity mechanisms.

From the technical point of view the implementation should leave the given software architecture untouched as much as possible. Nobody should be burdened with tasks or responsibilities potentially going beyond organisational capabilities. The installation, configuration and maintenance processes should be kept as simple as possible and be integrated into the installation process of the PyWrapper Software. Thus, the rights management and access control components should provide predefined access rights and rules as well as roles and user groups. Easy to use administration tools should support configuration and maintenance.

From the legal point of view mainly the European data protection and telecommunication acts have to be respected technically. Via the necessary usage of a PKI, also the European Digital Signature Act has to be respected. The responses of the questionnaires include special university guidelines and new laws concerning collections in general (Switzerland, France). This requires the implementation of hierarchical policy structures allowing the definition and inheritance of national specifics into the local decision process. These requirements have been covered and elucidated within the organisational requirements formulated above.

From an economic viewpoint the results show that in general no more than 5-10% of the available funds could be expected for the development, installation and maintenance of security services. Therefore the application of license free and even better open source software components is mandatory to save costs and allow future adaptations of software components.

4 Service Design

Since the existing software architecture is based on XML, we use security related XML standards to build the security components. The following standards have to be considered. The *eXtensible Access Control Markup Language (XACML)* [OAXWeb] is an OASIS standard of XML based languages for the definition of policies and access control decision requests and responses. The *Security Assertion Markup Language SAML* [OASWeb] is an OASIS standard defining a framework for the exchange of security information between online communication partners.

The World Wide Web Consortium (W3C) [W3CWeb] has defined *XML Signature* [BBF+02] and *XML Encryption* [ImDS02] to provide Digital Signature and encryption within the XML context. The *XML Key Management Specification XKMS* [Hall04] defines protocols for public key management services and helps sharing public keys needed for signature verification or decryption of messages. Due to the diversity of the clients and the need to keep the client site applications as simple as possible, the use of a trust service according to the XKMS as a front-end to a PKI is a suitable choice. XKMS is a “young” specification from the W3C and currently no freely available implementation of this standard exists.

The *eXtensible rights Markup Language XrML* [XRMWeb] is an XML-based specification language for expressing rights and conditions associated with digital content, service or any digital resource. In the SYNTHESYS context XrML can be used to enforce the compliance with the Intellectual Property Rights (IRP) of the published biodiversity data.

Our proposed system design for rights management and access control is shown in figure 1 and discussed in the following 2 chapters.

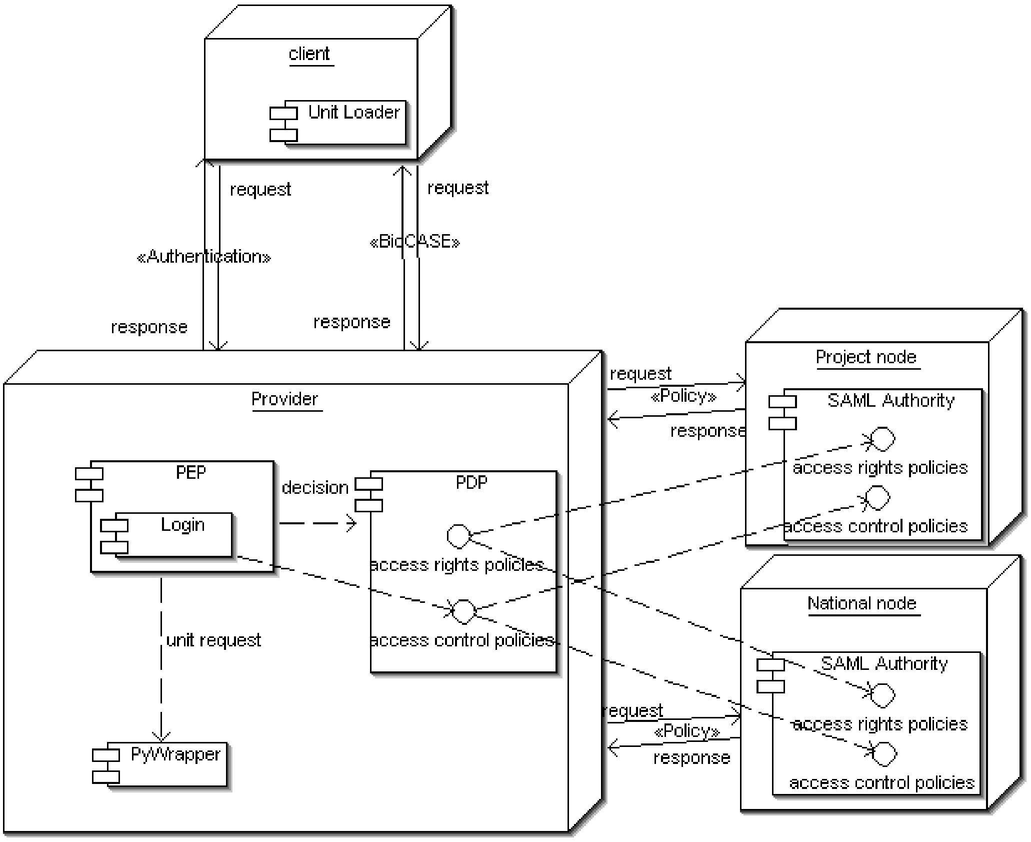

Figure 1 Rights Management and Access Control Concept

4.1 Concept for Rights Management

The concept for rights management builds on XACML. The XACML standard provides a very flexible architecture to define, manage and enforce policies. Thereby the included rules determine the decision process concerning rights based access control. The rights management functionality will be realised from the two main XACML components *Policy Enforcement Point (PEP)* and *Policy Decision Point (PDP)*, which will have to be integrated into the present provider scenario. Then, every request will firstly be processed from the PEP which delegates the final decision process to its related PDP. Dependent on the PDP decision, the PEP rejects, permits or constrains the requested access.

Resources are expressed using XPath statements within the BioCASE requests. By stating corresponding rules, access right policies will be defined describing to whom (e.g. user or role) the system grants what kind of access (e.g. read, write) to the data elements (resources) expressed with the XPath statement of the request.

To meet the specified requirements above regarding the provision of further (higher level) policies, XACML also offers mechanisms to include multiple policies into the decision process of the provider's PDP. Thereby the policies to be included may also reside on locations internal or external to the system boundaries of the current PDP (e.g. on an intranet or Internet server). Thus, the underlying policies of the access right decision process may be updated at every decision request from the corresponding server or regularly e.g. by importing and integrating policy configuration files into the local provider software configuration (or caching of requests for performance reasons).

The linking process between the provider’s PDP and higher level nodes could, in a further step, also be realised using the XACML Profile defined within the SAML standard. Thereby SAML authorities could be introduced to perform policy services accessible via standard SAML requests from the providers. The benefit of this approach is in its combination with access control components restricting policy access e.g. to specific providers. So, provider specific policies could be prepared e.g. from higher-level nodes on behalf of “security or IT aware” providers. On the other side national nodes could take over policy maintenance e.g. to incorporate national (e.g. legal) specifics into national-wide policies. Assuming a related configuration beside the providers, these policies would be automatically included in the policy decision process of their PDPs as shown in Figure 1.

4.2 Concept for Access Control

XACML is also suited for the implementation of a role based access control component into the present scenario. The main objective of an access control component is the surveillance of the compliance with the specified access policies defined within the access rights management component for a given subject. This comprises the authentication of the related subject demanding for access, the assignment of roles to the identified subject and the execution of the access rights decision processes.

For assigning subjects to specific roles and vice versa, version 2.0 of the XACML standard provides a role based access profile for the specification of roles and their mapping to the relating policies. It also covers concepts for mapping between users and roles.

The definition of the subject/role-mapping takes place within the access control policies. By the usage of the same XACML architecture as for rights management, the same mechanisms become usable for the (hierarchical) distribution of correspondent access control policies. So, e.g. a national node may prepare access control policies automatically included in the decision process by mapping roles and thus, access rights to specific subjects or user groups.

The analysis of the user questionnaire showed highly diversified possibilities regarding the interpretation of role understanding. So, in the first development phase only basic roles will be implemented. They will serve as an example for the derivation of similar roles and access control policies in the future: Anonymous, Client, and Expert.

For an incoming request from an unauthenticated subject, the access control policy assigns to him the role Anonymous. This role has minimum access rights allowing only access to publicly available data sets or information. This behaviour would then comply with the present access control situation of the system. Any request from a known subject (guest, normal user) will be assigned the client role permitting access to resources available only for registered users within the network. The expert role permits access to sensitive information also.

The authentication process reflects the identification of a given subject (e.g. user or client system). The authentication has to be done before the subject sends requests to the system. For that, the PEP will be extended by an authentication (commonly called login) component, which is responsible for the correct identification of requesting subjects (see Figure 1). The login process may be based on a Password or Public Key Infrastructure (PKI). We are planning to integrate both methods into a common login component.

To introduce PKI-based login, a public key infrastructure has to be configured. For authentication, the user's certificate will be verified. After a successful verification of the certification chain the public key from the certificate will be used to identify the user by execution of an authentication protocol based on SSL. The integration of SAML authorities into the login process would pave the way for a project wide single-sign-on solution. Once authenticated, the client would be able to request any information without the need for further explicit logins, switching from one connected provider to another.

5 Outlook

Users expect further enhancements from the security components in the medium-term future like authenticity of the data source and data integrity, (partial) confidentiality of documents, copyright protection or a Single-Sign-On expert portal. The following paragraphs will present security related aspects deduced from enhancements demanded within the analysed responses.

Authenticity of the data source enhances together with data integrity the quality of the transmitted information. A provider could digitally sign every document generated from the PyWrapper on request by any client. This will be done using the XML-Signature standard. The usage of Digital Signature assures the authenticity of the sender and the integrity of the received information, which is important for the trust in the submitted data and the image of the data providers. *Confidentiality* would be realised with encryption technology according to the XML Encryption standard. It ensures that sensitive information, e.g. locations of protected species, is treated confidential and is readable only by the intended recipients.

For that, the PEP is extended by an additional component realising the signing and encryption of documents. The corresponding cryptographic processes would be applied on the response documents received on request to the PyWrapper. The required credentials could be obtained from the same policies used for the implementation of access control management functionality. The delivered documents would be secured through the related XML Signature and/or XML Encryption formats. Some updates on the client software are also necessary.

Another future extension could make use of the features of *multiple and partial encryption and signing* of documents. Sensitive and non-sensitive information could be mixed within one document, where the protection of the sensitive elements of a document would be guaranteed through the usage of partial encryption, i.e. only accessible to specific subjects.

Using the features of the XML-Signature standard regarding partial document signing e.g. the marking of copyright information for every data element in a document would become feasible. But it fails, when protection against unlicensed dissemination of this document is required. Even if there is an existing XML standard dealing with this problem (XrML), the available API from ContentGuard supports only the marking of copyright issues. Furthermore, there is not any known working system implementation. However, the implementation of these features requires a very careful analysis of their influence on the current data formats and applications.

The development and integration of these security features to the biodiversity network are objectives of our further work.

Acknowledgements

This work was in part supported by the EU and the German Federal Ministry of Education and Research (BMBF) in the projects SYNTHESYS and GBIF-D. More information can be found at http://nbi.inf.fu-berlin.de/research/SYNTHESYS.

References

[BABWeb] Walter Berendsohn: ABCD Schema - Task Group on Access to Biological Collection Data; http://www.bgbm.org/TDWG/CODATA/Schema/default.htm, last seen 15 October 2004.

[BBF+02] W3C: M. Bartel, J. Boyer, B. Fox, B. LaMacchia, E. Simon: XML-Signature Syntax and Processing (2002), http://www.w3.org/TR/xmldsig-core/ last seen 15 October 2004

[BioWeb] BioCASE Secretariat: A Biological Collection Access Service for Europe; http://www.biocase.org, last seen 15 October 2004.

[DöGü03] Döring, M. & Güntsch, A. 2003: Technical introduction to the BioCASE software modules. 19th annual meeting of the Taxonomic Databases Working Group (TDWG 2003), Oeiras, Lisbon 2003.

[FEST95] FEST Consortium (A 2011); FEST Guide Appendix A: The Question Set; ftp://ftp.cs.tu-berlin.de/pub/local/kit/documents/KIT-Reports/r123/FG-AA.fm.ps, March 1995, last seen 15 October 2004.

[FESWeb] FEST Consortium: The FEST project; http://www.cee.hw.ac.uk/Databases/lachs/fest.html, last seen 15 October 2004.

[Hall04] W3C: Phillip Hallam-Baker XML Key Management Specification XKMS 2.0 (2004), http://www.w3.org/TR/2004/CR-xkms2-20040405/ last seen 15 October 2004

[ImDS02] W3C: T. Imamura, B. Dillaway, E. Simon: XML Encryption Syntax and Processing (2002), http://www.w3.org/TR/xmlenc-core/ last seen 15 October 2004

[OASWeb] OASIS: OASIS Security Services TC: http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=security, last seen 15 October 2004.

[OAXWeb] OASIS: eXtensible Access Control Markup Language TC; http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml, last seen 15 October 2004.

[OBWeb] Susanne Oehlschlaeger, Walter Berendsohn: GBIF-Deutschland; http://www.gbif.de, last seen 15 October 2004.

[SYNWeb] SYNTHESYS project: http://www.synthesys.info, last seen 15 October 2004.

[W3CWeb] W3C: World Wide Web Consortium (W3C); http://www.w3.org, last seen 15 October 2004.

[XRMWeb] Content Guard: XrML 2.0 Specification; http://www.xrml.org/get_XrML.asp; last seen 15 October 2004.