HAL
archives-ouvertes.fr

A distributed security policy for neuroradiological data sharing

Alban Gaignard, Johan Montagnat

▶ To cite this version:

Alban Gaignard, Johan Montagnat. A distributed security policy for neuroradiological data sharing. HealthGrid 2009, Jun 2009, Berlin, Germany. pp.257-262, 10.3233/978-1-60750-027-8-257 . hal-00677795

HAL Id: hal-00677795
https://hal.archives-ouvertes.fr/hal-00677795

Submitted on 11 Mar 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A distributed security policy for neuroradiology data sharing

Alban GAIGNARD and Johan MONTAGNAT
CNRS / UNS, I3S, MODALIS, 2000 route des Lucioles, Sophia Antipolis, France

Abstract. Grids are key technologies to federate data distributed in multiple neuroscience centers, thus enabling large scale multi-centric studies. However, the take up of these technologies is slow due to the difficulty to manipulate sensitive neuroradiological data in an open environment and the recognized risk of federated sites to loose control over their valuable data. In this paper we propose a distributed data access control policy, enabling the federation of existing data stores, where local security policies prevail, to supports multi-centric neuroscience studies. It achieves a compromise between enabling collaborative work through data sharing and preventing unauthorized access to data in a competitive environment.

Keywords. Distributed security policy. Medical data sharing. Collaborative distributed environment.

1. Distribution of neuroradiological data and access control

Computational neuroscience experiments rely heavily on the ability of brain image databases. Acquisition of this data is a costly process, resulting from the careful selection of target populations (patients and normal controls), the design of rigorous acquisition procedures, and the identification of examinations that show no biases and that can finally be integrated into the study database. Consequently, neuroscience centers are often reluctant to exchange their valuable data sources. However, the challenging population aging problems cause an ever growing need for larger and more specific neuroradiological data sets that can be achieved through multi-centric studies design and selected data sets sharing. Many initiatives are emerging today to propose an HealthGrid-based collaborative environment for supporting multi-centric neuroscience studies [8,**?**,**?**]. The adoption of these new technologies will depend on the confidence the end users will grant to these systems, especially the control they have on the data that is exchanged with other partners.

In the context of the NeuroLOG project [8], we aim at building a federation of neuroscience sites with both collaborative interests and competitive activities. This paper focuses on the distributed security policy designed to accomodate these two partly contradictory aspects. More specifically, the system needs to address four requirements: (R1) medical data protection, (R2) distributed control over data sources with prevailing local policies, (R3) support for multi-centric studies involving data sharing, and (R4) autonomous sites administration. Similar requirements are exhibited in a detailed analysis of security requirements for Neurosciences dedicated e-Infrastructures [12]: *"On the one hand, the local administrator must find a way to translate their local policies to a com-*

mon interface that can be understood by remote users as local. On the other hand, there must be a way for the remote user to gain fine-grained and secure access to individual data fields and parameters". Similar security requirements are also identified in the design of the @neurIST platform, aimed at improving research and clinical care of cerebral aneurisms [9].

Grid technologies have for long set the security infrastructure at the heart of middleware design and grid services interoperability layer. They enable the federation of data distributed over cross-institution storage resources, thus directly addressing R3. The solution proposed in this paper satisfies R1 through a standard X509 certificates-based Public Key Infrastructure (PKI) and the AES encryption algorithm that provide cross-institutional authentication and secured communication. The Shamir secret sharing scheme gained adoption in several medical data access grid services to protect the AES keys against the *insider abuse* threat [11,7,2,10,4]. It is an interesting perspective for our system.

R2 and R4 are more difficult to address. Centralized data access control mechanisms such as the traditional grid VO Management System (VOMS) [3] cause an externalization of data access control that is not accepted by neuroscience centers. VOMS are too heavy-weighted to set up light collaborations and, more importantly, require an agreement on a global authority or system super-users with privileges over multiple sites data access. Conversely, neurosciences request a data federation system which does not compromise sites data beyond the objects of collaboration decided and which does not interfere with the normal autonomous operation of the sites.

To tackle the centralization issues of VOMS, the Shibboleth's [1] decentralized approach to authentication and authorization is gaining adoption in grid communities. The central idea is that a remote referee authority authenticates and authorizes its users by generating and providing to users a security assertion including user's access rights (security attributes). The main advantage of this approach is the decentralized authentication and attributes assignment. However, with this approach data sharing in the context of multi-centric neurology studies would be under the responsibility of a plethora of site-specific authorities, while none of them owns access control rights on other sites data items. A key contribution of this paper is the decoupling of local data access rights assignment, performed by local site administrators, and distributed studies access rights assignment, delegated to a remote administrator.

In [5], Chadwick et al. propose PERMIS, a modular authorization infrastructure integrated within Globus Toolkit 4 and Shibboleth.It provides a hierarchical access control policies management and an access control decision engine. PERMIS enables the definition of various security policies, involving distribution of attributes in site-wise repositories whereas our approach propose a non-trivial security policy where federation-wide unique roles are recognized and coherently managed in the federation.

2. Security model

2.1. From independent to collaborative trust domains

The collaborating environment should not interfere with the normal autonomous operation of the sites, where multi-sites capabilities is not required (R4). Each center involved

in the federation is responsible for registering its own users, as it is usually the case in autonomous entities. It is administrating a local Certification Authority (Site CA), empowered to deliver and sign site user certificates, thus delineating a local trust domain. To enable multi-centric studies the neuroscience centers involved have to interconnect their trust domains. This is achieved through a standard certification chain: a root CA is delivering all Site CA certificates. The root CA is hosted in a coordination node named *Federation Registry*. This is the only centralized component in the system.

Two data protection policies, corresponding to two levels of privacy, are implemented in our system. Minimally, all communications is protected: data is encrypted during transfers only. Potentially the system can be configured so that data is also encrypted on disk. The first policy is acceptable in a research context where the manipulated data is anonymized prior to importation. It facilitates data manipulation locally while guaranteeing that data is protected when transferred outside. The second policy addresses the stronger data protection need considered in the context of clinical deployment or sites on which all local users should not access the complete data base.

2.2. Decentralized access control policy

As outlined in the introduction, the key problem with most existing grid security environments is the difficulty to define an authorization policy acceptable for the neurosciences community. The main contribution of this work is a decentralized data access control policy where local policies prevail. Our access control mechanism is based on traditional Role-Based Access Control (RBAC) [6]. RBAC assigns permissions to roles (the user possible functions) in such a way that access control policies remain light and easy to understand. Most existing RBAC systems are centralized and they manage two different functions simultaneously: (i) the assignment of roles to users and (ii) the definition of access rights for each role. Our systems decouples these two functions as detailed below.

On each site, at least one administrator has privileges to register the site users and to maintain the site's access control policy. Users may belong to different trust domains (a system administrator usually has no complete view of the potential users of the system) while the access rights to data of a given domain has to be ultimately controlled by this domain's administrator. The compromise to enable collaborative work while ensuring site-wise access control to local data is as follows:

- Site administrators are capable of creating federation-wide roles, as many as needed to describe their access control policies;
- The creator of a role controls the assignment of all system users to that role: the management of a particular role is centralized on one of the sites.
- Each site administrator controls the assignment of federation-wide roles to permissions related to their local resources.

A user is granted access to a data item if she belongs to at least one role that is locally authorized to access this item. This policy framework ensures that sites solely control access to their data: only a site administrator can bind some role to her data. It also ensures that each role is well defined and administrated: only the role creator can bind users to that role. It implies a collaboration between the data owner and the role creator: the data owner agrees to make some data accessible for a particular role (*e.g.* in the context of a particular multi-centric study); the role administrator is trusted and recognized as

the administrator for this particular study. Any user in the federation can collaborate to the study: through the certification chains, the role administrator can validate the identity of any user before assigning the role to her. Finally, the roles are guaranteed to be unique federation-wide through the Federation Registry. Roles can only be created after assignment of a unique name through the unique Registry.

This system is agile and preserves sites autonomy: sites are only tightly coupled to the Registry and they do not depend on it for their normal operation. The authorization scheme is lightweight and can quickly dynamically evolve to adapt to new needs for collaborative studies.

3. Implementation

The NeuroLOG platform [8] involves a single Federation Registry root server dedicated to the coordination of the platform, multiple intercommunicating Site Servers which are implementing most of the middleware functionality and distributed Clients from which users authenticate and connect to the system. Identification and data protection are operationalized through standard OpenSSL, Java Secure Socket Extension, Java Public Key Infrastructure and Java Cryptography Architecture. Authorization concerns different medical services (data, metadata and semantic data management, data analysis workflows management) that are exposed using either Java Remote Method Invocation (RMI) or Web Service (WS) interfaces. We validated that the proposed security model can be integrated in this heterogeneous services framework, noting that authentication of users as needed for implementing the authorization policy is neither straight forward with RMIs nor with WSs.

The authentication at the application level of services communicating through high-level protocols (such as RMI's JRMP or WS's SOAP protocols) over pre-established SSL connections requires specific treatment. Once an SSL communication is established, those protocols do not provide control, at server side, over the identity of the caller emitting high-level protocol messages. For instance, while opening an SSL channel with a Site Server, the identity of a user is validated through SSL handshaking. However, several clients may similarly connect to the same server which receives multiple requests without knowing who effectively performed each of them. This issue brings a major security requirement at application level that is authentication (re-identification) of the caller of a specific operation in the service. This problem is dealt with differently in the cases of RMI and WS servers.

In the case of RMIs, each remote method invocation is guarded by the control of the identity of the client at application level. Authentication tokens are added to RMI calls as additional, non-functional parameters to ensure re-identification of the client. The caller generates and transmits a ciphered and time-stamped invocation token for each call of a sensitive remote method. The server retrieves the client certificate from its SSL context. Re-identification of the client is guaranteed by the success of token deciphering, and time validity checking.

In the case of WSs, the security context depends on the service container. Similarly to the RMIs case, a Web Server transporting messages over HTTPS will benefit from SSL handshaking to identify the caller. However, simple stand alone web services encounter the same limitation: re-identifying the user connected is not possible and an au-

thentication token is needed in the WS operation call to perform access control at the service level. Alternatively, the web services can be hosted in a container. In this case, the container provides the functionality needed to retrieve the SSL identity of any caller, thus enabling re-identification without any extra token.

4. Illustration through a generic use case

Let us consider three sites A, B and C participating in the federation as shown in figure 1. Given a collaborative study initiated by site A, the goal of site B is to share a set of owned data with partners involved in that study The site administrators have to agree on the unique name, say *StudyA*, for the collaboration role. They share the study in several steps: (i) the administrator of *Site A* declares the *StudyA* role on the Registry, thus associating it with her site; (ii) she potentially grant permissions onto hosted data for this shared role through the local access control policy (study read permission, for instance); (iii) she registers any user (local or foreign) participating in the study by assigning her the *StudyA* role; and (iv) after importation of dataset *D* within *Site B*, the administrator of *Site B* shares *D* with participants in *StudyA* through the creation of a local authorization rule.

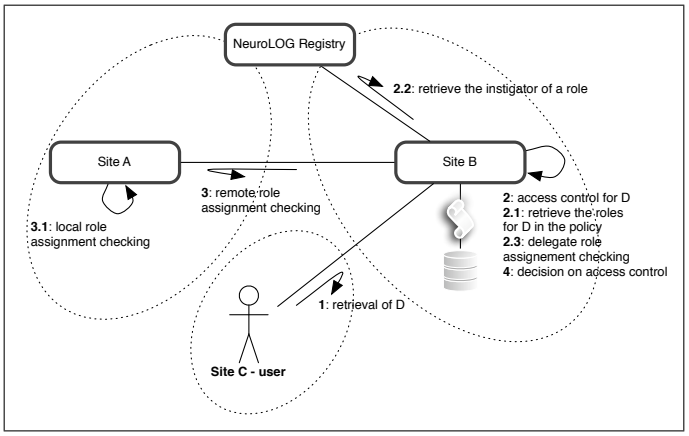

Figure 1. Decentralized access control

Let us now consider a user of *Site C*, involved in *StudyA* who needs to retrieve a shared data hosted by *Site B*. Figure 1 details the distributed access control: (activity 1) the user from *Site C* requests from *Site B* the retrieval of dataset *D*; (activity 2) *Site B* controls the permissions of the user over *D* by (2.1) retrieving in the local access control policy, the collection of roles attached to *D*, (2.2) retrieving the initiator sites for the resulting collection of roles; and (activity 3) delegating to each initiator site, the checking of role assignment to the user. If one of the collaborating sites validates the role assignment checking, then the access control is performed regarding the permission associated to the data and the role.

5. Perspectives and future works

Neuroscience centers have both collaborative interests and competitive activities. Their temptation to share neuroradiological resources is often counter-balanced by the fear of loosing control over precious data, which acquisition is a costly process. Building on the foundational grid security layer, the distributed access control framework proposed in this paper addresses neuroscientists needs for enabling multi-centric studies by federating existing, heterogeneous site environments while respecting local data access policies. This infrastructure enables autonomous sites operation and it does not require a centralized administration. This policy is validated through an implementation and practical set up problems related to client-server communications with RMIs and WSs are taken into account.

Acknowledgments

This work is partly funded by French National Agency for Research, NeuroLOG project, under contract number ANR-06-TLOG-024.

References

[1] Shibboleth. http://shibboleth.internet2.edu/.

[2] JRA3: Global Security Architecture, rev. 1. EGEE project deliverable, https://edms.cern.ch/document/602183/, 2005.

[3] R. Alfieri, R. Cecchini, V. Ciaschini, L. dell'Agnel lo, Ákos Frohner, A. Gianoli, K. Lörentey, and F . Spataro. VOMS, an Authorization System for Virtual Organizations. In *European Across Grids Conference (EAGC)*, 2003.

[4] I Blanquer, V Hernandez, D Segrelles, and E Torres. Enhancing privacy and authorization control scalability in the grid through ontologies. *IEEE Trans Inf Technol Biomed*, 13(1):16–24, 2009 Jan.

[5] David W. Chadwick, Gansen Zhao, Sassa Otenko, Romain Laborde, Linying Su, and Tuan-Anh Nguyen. Permis: a modular authorization infrastructure. *Concurrency and Computation: Practice and Experience*, 20(11):1341–1357, 2008.

[6] David F. Ferraiolo, Ravi S. Sandhu, Serban I. Gavrila, D. Richard Kuhn, and Ramaswamy Chandramouli. Proposed NIST standard for role-based access control. *ACM Trans. Inf. Syst. Secur.*, 4(3):224–274, 2001.

[7] Johan Montagnat, Ákos Frohner, Daniel Jouvenot, Christophe Pera, Peter Kunszt, Birger Koblitz, Nuno Santos, Charles Loomis, Romain Texier, Diane Lingrand, Patrick Guio, Ricardo Brito Da Rocha, Antonio Sobreira de Almeida, and Zoltü£¡n Farkas. A Secure Grid Medical Data Manager Interfaced to the gLite Middleware. *Journal of Grid Computing (JGC)*, 6(1):45–59, March 2008.

[8] Johan Montagnat, Alban Gaignard, Diane Lingrand, Javier Rojas Balderrama, Philippe Collet, and Philippe Lahire. NeuroLOG: a community-driven middleware design. In *HealthGrid*, pages 49–58, Chicago, June 2008. IOS Press.

[9] Hariharan Rajasekaran, Luigi Lo Iacono, Peer Hasselmeyer, Jochen Fingberg, Paul E. Summers, Siegfried Benkner, Gerhard Engelbrecht, Antonio Arbona, Alessandro Chiarini, Christoph M. Friedrich, Martin Hofmann-Apitius, Kai Kumpf, Bob Moore, Philippe Bijlenga, Jimison Iavindrasana, Henning Müller, Rod D. Hose, Robert Dunlop, and Alejandro F. Frangi. @neurIST - Towards a System Architecture for Advanced Disease Management through Integration of Heterogeneous Data, Computing, and Complex Processing Services. In *CBMS*, pages 361–366, 2008.

[10] Diego Scardaci and Giordano Scuderi. Managing confidential data in the glite middleware. In *WETICE*, pages 298–299, 2007.

[11] Ludwig Seitz, Jean-Marc Pierson, and Lionel Brunie. Key management for encrypted data storage in distributed systems. In *IEEE Security in Storage Workshop*, pages 20–30, 2003.

[12] Anthony Stell, Richard O. Sinnott, Oluwafemi Ajayi, and Jipu Jiang. Security oriented e-infrastructures supporting neurological research and clinical trials. In *ARES*, pages 629–636, 2007.