Collaboration and the Knowledge Economy: Issues, Applications, Case Studies
Paul Cunningham and Miriam Cunningham (Eds)
IOS Press, 2008 Amsterdam
ISBN 978–1–58603–924-0

Dynamic Access Control Management for Distributed Biomedical Data Resources

Matthias ASSEL, Onur KALYONCU
High Performance Computer Center of the University Stuttgart,
Intelligent Service Infrastructures, Nobelstr. 19, 70569, Stuttgart, Germany
Tel: +49 711 68562515, Fax: +49 711 68565832, Email: {assel,kalyoncu}@hlrs.de

Abstract: With increasingly dispersed and inhomogeneous resources, sharing knowledge, information or data becomes more and more difficult and manageable for both end-users and providers. To reduce administrative overheads and ease time-consuming management tasks, quite a few solutions for secure data sharing and access have been designed and introduced in several research projects worldwide. In this paper, we focus on the EU-funded research project ViroLab, which tries to build a virtual laboratory that allows health professionals to access distributed data resources in order to perform clinical studies and analyses on the requested data sets. The paper concentrates on the approach and implementation how data sharing in ViroLab is carried out and how these biomedical resources can be easily managed to guarantee the highest level of protection against any abuse. We conclude with an outlook and recommendations on how the system can be enhanced and improved.

1. Introduction

Today, more and more interdisciplinary scientific teams and institutions, which are located at different sites and even in several countries, collaborate together to reach their scientific goals and to facilitate international research activities. Specifically designed collaborative working environments [6] assist scientists during their daily workflows and help them to discuss about their common goals, plan their future tasks, share their knowledge, and evaluate, report and pool their results. These working environments are continuously evolving and becoming more and more complex. A major reason constitutes with the increasing number of inhomogeneous and distributed resources, in particularly data resources and repositories, which could additionally contain sensitive information or private and confidential data sets.

The EU-funded research project ViroLab [1] shall provide researchers and medical doctors a so-called virtual laboratory for infectious diseases. ViroLab develops such a collaborative workspace for different types of end-users like virologists, clinicians, and researchers and enables the interactive sharing of expertise and results while working together on the same data and information sets that are widely dispersed over Europe and currently without cross-national collaboration. The Human Immunodeficiency Virus (HIV) drug resistance problem has become an increasing problem worldwide though it is one medical area where genetic information is widely available and has been collected and used for many years [5]. For these reasons, it was chosen as prototype in the ViroLab project.

In this paper, we describe the approach how access to the distributed and confidential data resources within the ViroLab collaborative working environment is realised and implemented. We explain how existing security concepts like Shibboleth, GSI[i], and XACML[ii] have been combined and linked together in order to build a corresponding environment that ensures the utmost protection for clinical data and provides the users a great flexibility to set up and manage such collaborative working sessions while at the same

time reducing administrative overheads. Finally, we also indicate some limitations and peculiarities and provide advice how future systems can be further improved and enhanced.

2. Motivation and Objectives

Basically, medical experts like doctors or virologists want to use the ViroLab virtual environment for requesting patient information in order to predict any possible drug resistance for their according case. They simply want to retrieve all relevant information without the need of any specific expertise in computer science. The way to gather data must be kept simple and transparent to them but should be as self-explaining as possible

Besides this typical use case, which principally takes place within one organisation or hospital, the virtual laboratory shall also provide possibilities to perform studies on the integrated patient (clinical and genetic) information sets. Those studies usually require lots of data most suitable from several institutions to analyse and observe interesting characteristics of the different virus subtypes and their spread among our society. Since most of the data owners are afraid of sharing all or even single data sets with someone working for ViroLab, the exchange of relevant data shall be handled and controlled by the providers themselves. Hence, the data resource administrators need to be provided with a user-friendly and easy to understand concept, which allows them for dynamically defining and changing access policies in which institutions, departments, or users can be specified to work on their data without browsing and/or accessing their entire repository. In order to reach this difficult endeavour, the ViroLab workspace together with its integrated data resources are developed in close cooperation with clinical partners like hospitals and bioinformatics centres that, in addition to the data and tools, also actively contribute to the overall design of relevant workflows for daily practices being foreseen for integration within the virtual infrastructure. Furthermore, the collaborative and interactive usage of clinical databases beyond the scope of ViroLab shall also be achieved to facilitate future research activities in the field of other infectious diseases.

3. Related Work

Data management, in particular access control to certain resources, in collaborative and Grid environments respectively has been a subject of study and research for quite a few years. Early solutions, such as those employed in common Grid systems basically relied on replicating data, or prior to performing any calculations, data had to be fetched and staged by a specialised middleware component. The identification of users as well as the access to certain data files has been performed in the traditional way of mapping system users who authenticate themselves via corresponding (Grid) certificates onto local user accounts with specific rights allowing them to access, use and transfer appropriate pieces of data. This was a limiting solution as it took no notice of structured data storage technologies such as databases which additionally require proper user authentication and authorisation.

These constraints gave rise to a number of projects aiming at standardisation and increased flexibility of data management in Grids and collaborative environments, one of the most important of them being OGSA-DAI [7]. This project aimed to develop a toolkit for exploiting different types of data resources onto the Grid through a set of standard Web Service interfaces. Although OGSA-DAI provides the possibility to integrate relational databases smoothly, the authorisation for granting access to such resources still based on the certificate provided and hence, the user's distinguished name which reduces the level of dynamicity and limits administrators to perform fine-grain access control down to the level of single data sets. In order to overcome these restrictions of existing authentication and authorisation concepts, which completely rely on GSI, other projects [8] tried to combine both areas, traditional Grid computing together with decentralised management of user

identities to ensure greater flexibilities for both users and resource providers. Unfortunately, these projects do not focus on the dynamic adaption of user rights according to sudden circumstances. However, projects like TrustCoM [9] or BREIN [10] which are dealing with business workflows and collaborations have already applied dynamic policy management but basically for protecting services or applications and without touching the difficult problem of distributed data resources. In ViroLab, we are trying to take and combine results from several previous projects in order to provide the users with the greatest level of flexibility and dynamicity but at the same time guaranteeing the maximum of security and trustworthiness while accessing, sharing and using biomedical data sets.

4. Approach and Methodology

Security issues and policies for collaborative working environments are very different from local ones. The abuse of data not only by third parties but also by so-called trusted organisations must be considered during the design and development of the entire setup. In other words, not all of the integrated resources may be accessible to all of the users known to the system. To protect certain data sources, applications, services, or tools, specific access control policies, which define by whom the shared resources can be accessed and used within the environment, are usually a prerequisite of any virtual collaboration.

In our approach, each partner organisation determines and specifies its own judgment based on the self-defined and self-managed access control policies. According to this judgment the corresponding organisation can make their appropriate authorisation decisions. Thus, in addition to the decentralised control, a fine-grained control of the resources is basically achieved.

To overcome the problems of dynamic and on-demand collaborations, ViroLab makes use of the established security framework Shibboleth [3], which was developed to protect online resources across and within organisational boundaries. Shibboleth provides a federated Web Single Sign-On (SSO), attribute exchange framework and extended privacy functionality, allowing the users and their home site to control the attribute information being released to each service provider. Using Shibboleth-enabled access simplifies the management of identity and access permissions for both identity and service providers.

The access control mechanism that is used within Virolab has some differences from the typical role-based access control. The policy that is created by the resource's owners specifies which minimum set of attributes must be provided in order to gain access to the corresponding resource. In a typical role-based approach within a single organisation a single attribute that defines the users' role could be sufficient to describe a user in a single environment, but for fine-grain access control within multiple organisations supplementary attributes are required which do also reflect the organisation and the organisation type of the user. More attributes are also possible and base on the granularity of entire setup. To make authorisation decisions, the attributes being released by the respective Shibboleth identity provider (IDP) – typically the home organisation of the user (refer to figure 1, step 7) – is evaluated against the attributes specified in the access control policies. Users typically request permissions by being a member of a certain role. Shibboleth eliminates in our attribute-based approach also the need for role assignment policies that define which role can be assigned to which subject like in a role-based approach. In that manner, Shibboleth allows a simple policy management which only concerns about the permission policies. In our approach, permission policies are the single type of policies. The basic security infrastructure components and their interactions are now explained in more detail to clearly depict our approach and its relevance and application within ViroLab .

1. A Doctor logs into the ViroLab Portal (a web-based user frontend). He has to provide his credentials, usually username and password (Authentication);

2. The credentials are now transferred to the local identity management system and verified against the locally stored credentials;
3. In case the user is known to the home organisation's users database, a digital identity token (fingerprint) is created and sent back to the portal;

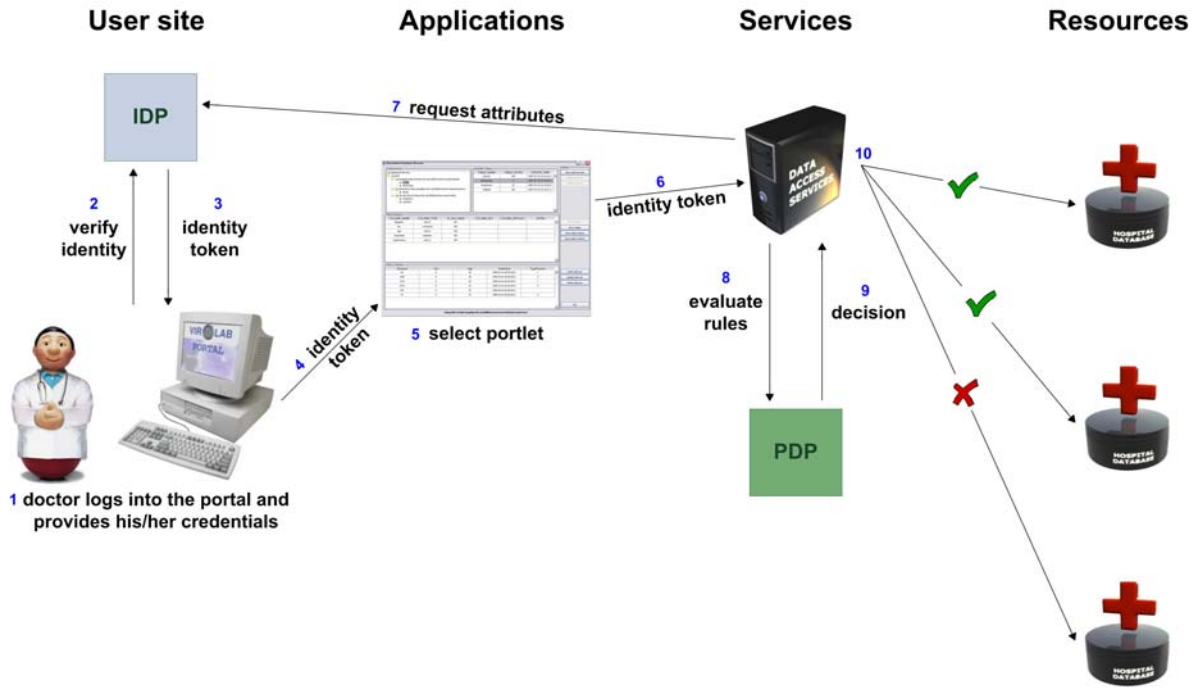

Figure 1: Authentication and Authorisation within ViroLab

4. Once the user is logged in, he can choose between different available applications;
5. He selects the data access application and sends a request to the Data Access Services (DAS) that are responsible for querying distributed databases simultaneously;
6. During that request, the initially created identity token is passed to the DAS that require this token to request the corresponding user attributes;
7. This request is performed by sending the identity token to the user's home organisation (HO) firstly proving whether the user is known and the token is valid. In a second step, the released user attributes like his role, institution or e-mail address are obtained from the local database and returned to the DAS;
8. The final decision whether someone is allowed to access a resource is only known to the PDP (Policy Decision Point). Hence, DAS sends an authorisation request to the PDP;
9. The PDP checks its stored access control policies for the corresponding rules. These rules contain conditions specifying the required set of attributes. If a policy rule matches with the provided attributes, the appropriate resource is cached. Having evaluated each access policy, the PDP returns a list of accessible resources to the DAS;
10. Finally, the DAS take the incoming request and tries to connect to each of the accessible resources and performs the user's query.

5. Technologies and Implementation

In order to meet the specific requirements for exchanging confidential biomedical information within such a virtual environment [2], the solution introduced and developed in ViroLab is built on established Grid technologies – GSI and OGSA-DAI – which provide the core for the own designed services, called Data Access Services (DAS).

These services [4] implement standard user interfaces to support various user groups but they also allow the integration of different data resource types, using the core functionalities of OGSA-DAI, to be smoothly exposed within the entire infrastructure. To guarantee that only persons who are known to the infrastructure can access certain data sets, additional security features, in particular strong access control mechanisms for all integrated data resources, have been defined and applied to protect the sensible information.

The Policy Decision Point (PDP) introduced within ViroLab has been developed during the TrustCoM project [12]. To fulfil certain new requirements related to data resources, it has been slightly modified and extended. The PDP is implemented as a web service and is responsible for controlling any access to a certain resource. The PDP makes its decision based on the access control policies stored in its repository to decide whether the user can access that particular resource and perform queries. It provides quasi dynamic authorisation and user-defined policies which are created and managed by data providers themselves. For the simplicity of the policy management, each data provider writes its own access control policies and a data provider has one access control policy for each of its resources.

The Extensible Access Control Markup Language (XACML) that provides a general policy language as well as an access decision language is used to define these attribute-based access control policies in our approach. XACML is implemented in XML and standardised by the Organisation for the Advancement of Structured Information Standards (OASIS). A XACML policy typically consists of multiple policy rules, and within each policy rule the data providers can describe specific conditions specifying the required attributes a user needs to provide in order to become authorised. In the following figure, this general interpretation is being presented.

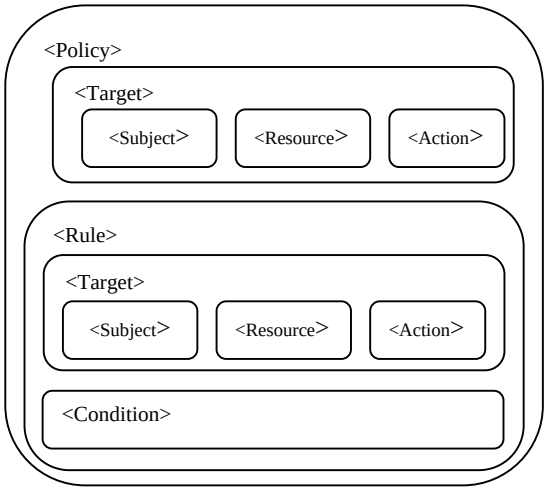

Figure 2: Structure of XACML Policies

In a typical scenario, the PDP initially receives from the DAS a list of available resources and the set of the attributes of the user for whom the access permission is requested. For every available resource, the PDP creates the relevant authorisation requests (see figure 3). The subject field of this request contains the received set of attributes and the resource field corresponds to the one of the available resources (refer to figure 4). In the evaluation process, the PDP matches the subject, resource, and action fields of the request with the target fields of the policies in its repository. In case of a successful match, it evaluates the policy and makes its decision. After the evaluation of the corresponding policies, PDP combines the individual decisions and creates a final authorisation decision being shared with the DAS that provide the users with a list of all accessible resources. The modification and the deletion of corresponding access control policies are also directly provided by the PDP.

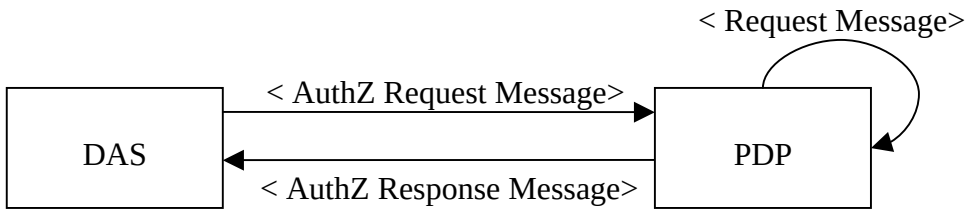

Figure 3: Authorisation Scenario

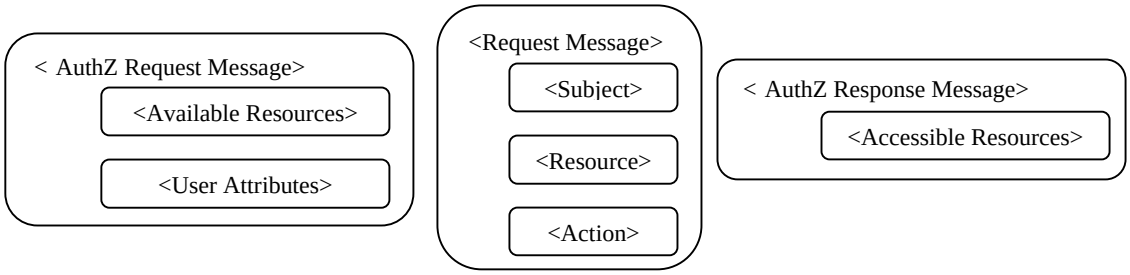

Figure 4: Structure of the Messages

6. Preliminary Results

The ViroLab project is in the middle of its implementation phase developing and realising this approach. However, ViroLab already released a first prototype of the virtual laboratory that supports data access and integration of the heterogeneous resources in a limited way. One can deal with these resources as a federated data space, which can be queried by submitting multiple and concurrent requests for gathering any kind of biomedical data sets that still reside in an inhomogeneous state. Basic security features including the encryption of transferred data messages as well as the support for user authorisation based on Shibboleth's authentication and authorisation infrastructure are also in place. Currently, four user sites have been connected to the overall laboratory environment which provides Shibboleth identity provider capabilities to their users. Access control has also been successfully achieved by using the extended version of the TrustCoM PDP and self-defined XACML access control policies following the abovementioned schema. The dynamicity aspect is supported through the creation of a nice and user-friendly graphical interface that enables fast and easy generation, change, and upload of particular policies.

Both virologists and clinicians have successfully applied these activities requesting data within several pre-defined experiments. A detailed description of corresponding experiments and workflows can be found in [11].

7. Conclusions

XACML provides administrators with a standard access control policy language. Controlling the general access to the resources can be achieved by very simple policies. Highly detailed policies, which are also provided by XACML, can be used in order to support fine-grained access control down to the level of database tables and single data sets. XACML also allows conditional authorisation, policy combination, and conflict resolution. It is a very general extensible language offering lots of flexibility. However, this flexibility and expressiveness can cause complexity and verbosity while creating deeply structured policies. Another limitation of the language is the lack of policy versioning and management in the XACML framework. This must be solved by the administrators themselves.

Using a one standard access control policy language within the entire system has some advantages. The delegated administrators do not have to write their own policies in many

different languages but they can reuse their existing codes. Thus, they can create, change and delete the access control policies in a timely manner. There is also no need for an invention of a new and specific policy language.

The role-based approach allows for modeling a single or several complex organisation/s. Instead of considering every single user in the entire system this can save time and money. Distributed administration also reduces the complexity and provides enhanced flexibility.

The solution presented in this paper shows the interaction of Shibboleth with the DAS and a modified version of the TrustCoM PDP. It still resides in a prototyping phase but already achieves the integration of dispersed and heterogeneous biomedical data resources in an easy to manage and secure way. Future work and further usage scenarios including several hospitals and more complex data queries shall validate the usability and effectiveness of XACML for making fine-grained and attribute-based access control that could be applied and used in real clinical workflows.

Acknowledgement

The results presented in this paper are partially funded by the European Commission through the support of the ViroLab Project Grant 027446. The authors want to thank all who contributed to this paper, especially the members of the project consortium.

References

[1] ViroLab – EU IST Project (IST-027446), http://www.virolab.org

[2] M. Assel, B. Krammer, and A. Loehden. Management and Access of Biomedical Data in a Grid Environment. In Proceedings of the 6th Cracow Grid Workshop 2006, pp. 263-270. Cracow, Poland (2006)

[3] M. Assel and A. Kipp. A Secure Infrastructure for Dynamic Collaborative Working Environments. In Proceedings of the 2007 International Conference on Grid Computing and Applications, pp. 212-216. CSREA Press, USA (2007)

[4] M. Assel, B. Krammer, and A. Loehden. Data Access and Virtualization within ViroLab. In Proceedings of the 7th Cracow Grid Workshop 2007, pp. 77-84. Cracow, Poland (2007)

[5] P. Sloot, C. Boucher, M. Bubak, A. Hoekstra, P. Plaszczak, A. Posthumus, D. van de Vijver, S. Wesner and A. Tirado-Ramos. VIROLAB - A Virtual Laboratory for Decision Support in Viral Diseases Treatment. In Proceedings of the 5th Cracow Grid Workshop 2005, Cracow, Poland (2005)

[6] A. Kipp, L. Schubert and M. Assel. Supporting Dynamism and Security in Ad-Hoc Collaborative Working Environments. In Proceedings of the 12th World Multi-Conference on Systemics, Cybernetics and Informatics: WMSCI2008, In Press. Orlando, USA (2008)

[7] M. Antonioletti , M.P. Atkinson , R. Baxter , A. Borley, N.P. Chue Hong, B. Collins, N. Hardman, A. Hume, A. Knox, M. Jackson, A. Krause, S. Laws, J. Magowan, N.W. Paton, D. Pearson, T. Sugden, P. Watson, M. Westhead. The Design and Implementation of Grid Database Services in OGSA-DAI. Concurrency and Computation: Practice and Experience, Volume 17, Issue 2-4, pp. 357-376 (2005)

[8] T. Barton, J. Basney, T. Freeman, T. Scavo, F. Siebenlist, V. Welch, R. Ananthakrishnan, B. Baker, M. Goode, and K. Keahey. Identity Federation and Attribute-based Authorization through the Globus Toolkit, Shibboleth, Grid- Shib, and MyProxy. Proceedings of 5th Annual PKI R\&D Workshop, Gaithersburg, USA (2006)

[9] TrustCoM - EU IST Project (IST-2003-01945), http://www.eu-trustcom.com

[10] BREIN - EU IST Project (IST- 034556), http://www.gridsforbusiness.eu

[11] T. Gubala, B. Balis, M. Malawski, M. Kasztelnik, P. Nowakowski, M. Assel, D. Harezlak, T. Bartynski, J. Kocot, E. Ciepiela, D. Krol, J. Wach, M. Pelczar, W. Funika, and M. Bubak. ViroLab Virtual Laboratory. In Proceedings of the 7th Cracow Grid Workshop 2007, pp. 35-40. Cracow, Poland (2007)

[12] M. Wilson, D. Chadwick, T. Dimitrakos, J. Doser, A. Arenas, P. Giambiagi, D. Golby, C. Geuer-Pollmann, J. Haller, S. Ketil, T. Mahler, L. Martino, X. Parent, S. Ristol, J. Sairamesh, and L. Schubert. The TrustCoM Framework V0.5. In Proceedings 6th IFIP Working Conference on Virtual Enterprises (PRO-VE '05), Valencia, Spain (2005)

[i] Grid Security Infrastructure

[ii] eXtensible Access Control Markup Language