Security Architecture for Distributed Medical Information Systems

Luigi LO IACONO, Hariharan RAJASEKARAN

NEC Laboratories Europe, IT Research Division
Rathausallee 10, D-53757 Sankt Augustin (Germany)
{lo_iacono, rajasekaran}@it.neclab.eu

Abstract: This paper presents the security architecture of the @neurIST medical information system. @neurIST aims at a research and decision support system for treating diseases that unites multiple medical institutions and service providers offering technical solutions based on the Service Oriented Architecture (SOA) paradigm. The security architecture provides secure access to federated medical data spread across multiple sites and protects the privacy of the patients by pseudonymisation of the medical data required for the study.

1 Introduction

The Service Oriented Architecture (SOA) paradigm offers an effective framework to develop systems which need to incorporate data, applications, and systems from multiple partners. One such example in the health domain is the @neurIST[1] medical system. It combines Data and Compute Grid technologies with the ones from the SOA sphere to create a medical IT infrastructure that offers clinicians and researchers a tool that brings together medical data that is distributed across multiple sites and offers extensive computational power to run complex simulations to analyze the characteristics of aneurysms. Security is a vital part of such a system since sensitive medical data and cost intensive computational services are involved.

This paper presents the security architecture developed for the @neurIST system and is organised as follows: The @neurIST project is introduced in section 2 followed by a detailed description of the security architecture in section 3. The privacy aspects are discussed in section 4 and section 5 concludes the paper.

2 The @neurIST Project

@neurIST is an Information Society Technologies (IST) Integrated Project funded within the European Commission's Sixth Framework Programme. The @neurIST consortium brings together 30 multi sectorial partners representing hospitals, universities, research institutes and the industry across Europe and incorporates the diverse technologies and services offered by the partners into the @neurIST system. The @neurIST system focuses on supporting the research and treatment of cerebral

[1] http://www.aneurist.org/

aneurysms. The project aims at building a distributed IT infrastructure that consolidates complex data from multiple sources, and enables personalized patient management (i.e. data capture, referral, decision support, treatment planning), as well as clinical research in cerebral aneurysms [1]. Though the system is currently developed to showcase its use in the research and treatment of aneurysms, the service-oriented infrastructure being developed is generic enough to be used for other diseases as well.

The @neurIST system has two modes of operation - a clinical decision support system (CDS) used for treatment planning and a research system used to conduct research studies. When used as a CDS, the clinicians can use the system to view the images of the aneurysm, conduct simulations to ascertain the risk of rupture and make use of knowledge gathered from studying factors such as genes, physiological conditions, etc to arrive at a treatment plan for the patient. In the research mode, the system allows researchers to access pseudonymized data of the patients taking part in the research study and conduct research using simulation tools to ascertain risk factors that influence the risk of an aneurysm bursting.

The entire system is divided into three layers with (1) the application layer providing the tools for end users, (2) the middleware layer providing the communication infrastructure and (3) the resource layer hosting the databases and computational services to store and provide compute power for simulations. The application layer consists of four software suites. **@neuLink** provides researchers with data that links genetic information and disease information. **@neuFuse** provides a visual interface to visualize data from medical imaging and biomedical instrumentation. This application also enables the clinician to order blood flow simulation through the affected blood vessels using the computational services. **@neuEndo** is an application used to assess mechanical and flow performance of stents used for treatment of aneurysms. It is used both by clinicians to test the suitability of a stent for treatment and also by stent manufacturers to optimize their stent designs. **@neuRisk** is the CDS offering treatment planning based on the risk factors identified using @neuLink, @neuFuse and @neuEndo.

The middleware component **@neuInfo** hides the exact physical or logical sources of the data or computing resources for the application suites by providing data access to multiple databases using semantic mediation. The suites use @neuInfo for accessing clinical databases or compute resources spread across different sites without having to worry about the nature of the underlying resource. The computational services in **@neuCompute** also use @neuInfo to acquire data needed for simulations and to store the computed results back into the clinical databases. A more detailed description of the @neurIST architecture can be found in [2, 3].

3 Security Architecture in @neurIST

The security work in @neurIST is focused on the specific security requirements of the distributed and federated service environment for conducting research and supporting treatment in the health domain and tries to provide a security system which ensures that

- the patient data is made available only to authorized personnel, and that
- the privacy of the patient is preserved at all times.

An important non-functional requirement here is the demand for an efficient infrastructure to authenticate and authorize @neurIST stakeholders across multiple domains and borders. The security architecture achieves this by providing the following:

- Virtual Organization (VO) and Trust Model underpinning the security of the system
- The security credentials used for authenticating users (local and remote)
- Policies for authorizing users based on attributes such as roles and location (local and remote)
- Policy enforcement systems (local and remote)
- Filtering and logging schemes to protect user privacy and keep track of data changes
- Fine grained access control at the data level

3.1 VO and Trust Model

The @neurIST system gathers all the participating entities such as clinics, research institutes, device manufacturers and service providers into one large VO. This is important because, a credential issued to an @neurIST user should be recognized (either directly or through the construction of adequate trust paths) by all participating entities in the system to grant access to services and data in the system. To provide the required flexibility and manageability of such a large distributed service environment, the principles of decentralized VO are applied. This means, that no centralized components are required to establish and manage the VO. This is realized by introducing the separation between the local and remote VO management, which e.g. does not require a system-wide harmonisation of local identification and authentication policies and schemes.

The underlying trust in the system is established by the steps shown in Figure 1. The involved medical institutions and service providers run through a contracting phase to setup a research project. An additional step introduced into this common process is the exchange of security tokens. The security tokens here are the certificates of each partner's Security Token Service (STS). The exchanged tokens are henceforth trusted by each participating site to (remotely) authenticate system participants in the scope of the @neurIST study.

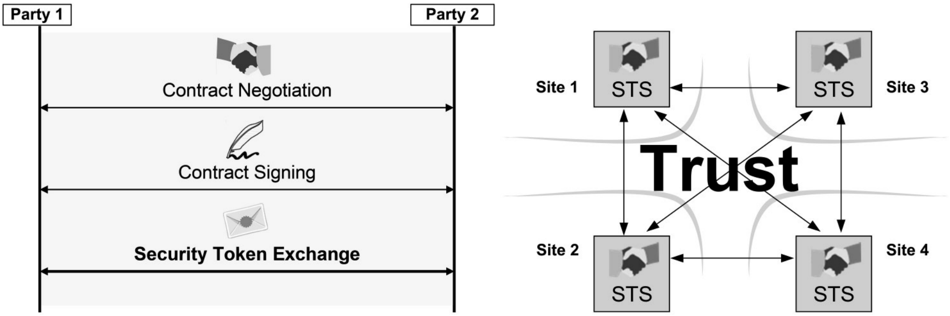

Figure 1 Trust Model

3.2 Credential Provisioning and Access Control in @neurIST

Even though the patient data available for secondary uses does not contain Personally Identifiable Information (PII) and should not allow the linkage to a particular patient without disproportionate effort (see section 4), the usage of this data is bound to the specific purposes of the associated study and must therefore be protected against unauthorized access. Hence, suitable access control policies need to be implemented and enforced.

From a conceptual viewpoint, the access control system for multi-institutional research in life science should follow the common patterns and principles for distributed cross-domain or cross-border information systems, in which heterogeneous environments of multiple distinct institutions need to be interconnected. A widespread approach is to apply hybrid security models to such systems to respect the various security domains and their different security policy implementations. Generally, clinical centres have their own security, access rights management and privacy protection policy according to the role of the user [4], and have the know-how concerning data access and communication using standards such as HL7 (Health Level 7), DICOM (Digital Imaging and Communications in Medicine), IHE (Integrating the Healthcare Enterprise), or business components such as Web services [5]. A hybrid security model underlies the combination of a local model and a distributed model. In other words, within a security domain all the security is concentrated and placed under the responsibility of this domain whereas between different security domains, the chosen approach consists in designating, in each domain, a security entity (which is known as Security Token Service, STS) who will be in charge of issuing and verifying short-term security tokens with the entities of the other security domains.

Figure 2 shows such a distributed architecture consistent with the health application domain. Here, the local security model relies on national e-health infrastructures such as Health Professional Cards (HPC) and local authorization systems to obtain a security claim from the local STS which then can be used to access the distributed services residing in a distinct security domain or even in a distinct country.

The attributes contained in the security claims have a system-wide scope, since the security tokens (also denoted as @neurIST tokens) are accepted as a valid measure for user authentication and the attributes are recognized through out the system and are used in the attribute based access control to resources. The end users' attributes currently consist of their role inside @neurIST (e.g. clinician, researcher, nurse etc) and their affiliation (e.g. name of the hospital, research institute etc) including the location.

Figure 2 shows the access control employed by @neurIST in action. The @neurIST tokens issued by the (local) STS in step 2 are SAML [6] tokens which include each concerning user's attributes. In step 4 the (remote) STS checks the issued token and a policy enforcement point checks if the attributes present in the token entitle the user to access the resource. @neurIST uses XACML [7] for its policy framework.

It is important to note that although the credentials to users are issued and administered by the local entity to which they belong; the access to data and compute services is controlled based on policies set by the entities that own them. Therefore, even if a local entity issues a credential to its users that enable them to use an @neurIST service, the authorization decision to allow such a use is finally controlled by the policies set by the entity owning the service. The @neurIST access control system developed provides both the efficient management of the VO and the control to the provided resources for highly distributed service environments.

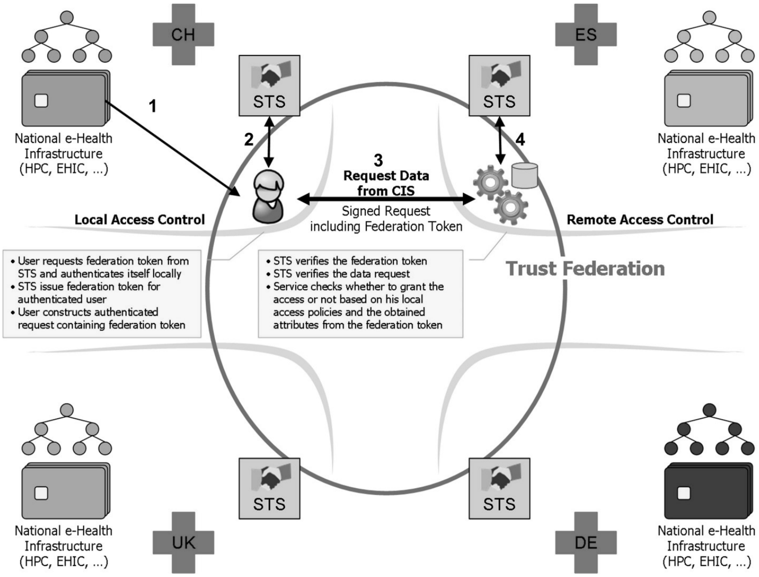
CH
ES
STS
STS
1
National e-Health
Infrastructure
(HPC, EHIC, ...)
2
3
Request Data
from CIS
4
National e-Health
Infrastructure
(HPC, EHIC, ...)
Local Access Control
Signed Request
including Federation Token
Remote Access Control
• User requests federation token from STS and authenticates itself locally
• STS issue federation token for authenticated user
• User constructs authenticated request containing federation token
• STS verifies the federation token
• STS verifies the data request
• Service checks whether to grant the access or not based on his local access policies and the obtained attributes from the federation token
Trust Federation
STS
STS
National e-Health
Infrastructure
(HPC, EHIC, ...)
UK
DE
National e-Health
Infrastructure
(HPC, EHIC, ...)

Figure 2 Access Control in @neurIST

4 Privacy in @neurIST

The privacy of patient data is ensured by enforcing that

- data used for research purposes is pseudonymised to remove PII from the medical data required for the study,
- filtering checks on the data leaving the databases are performed, and
- only authorized personnel are allowed to access the @neurIST system (see section 3).

Personal health data contains in the first instance information relating to the current and historical health, medical conditions and medical tests of its subject. Stored in digital form, it is referred to as an Electronic Health Record (EHR). An EHR is primarily used in the treatment context in which the patient's identity data is needed and protected by medical secrecy, but it also serves as a basis for secondary uses that may vary considerably in purpose and goal including the protection and enhancement of public health and conducting research [8]. Depersonalization must be performed prior to secondary use. Yet, although the identity of the patient is not always important in secondary contexts, it is not always desirable to simply anonymise the EHR. There are many secondary use scenarios for which the ability to form the correct association between a single patient and his EHR from distinct sources or distinct points in time is essential. Examples are the provision of follow-up data at a later point in time, the withdrawal of samples or data after a specific patient's request or the quality control of the data such as checking for double-entries. This usually prevents anonymisation and demands pseudonymisation schemes instead. Furthermore, in some research and ethical frameworks it may be necessary to maintain the possibility to re-contact patients in the event of results relevant to their health being obtained. Here, a reversible pseudonymisation system is required. In general, depending on the kind of research network and its requirements, distinct procedures for anonymisation or pseudonymisation are appropriate. In @neurIST a reversible pseudonymisation scheme is used.

The PII is removed using a two step process. In the first step the patient ID in the health record is reversibly pseudonymised. Since the EHR contains structured (e.g. free text, header in medical images) and unstructured (e.g. medical images) data, the second step consists of removing PII in the structured data (e.g. name, date of birth, etc) and removing physically recognisable information from the images.

The outgoing data set is filtered in addition to ensure privacy is not breached by sophisticated query attacks which tend to identify personal information when the data is restricted to a small set. Since the data access is provided via OGSA-DAI [9] mechanisms, the filtering is applied to the so-called Response Document. Figure 3 illustrates the underlying concept.

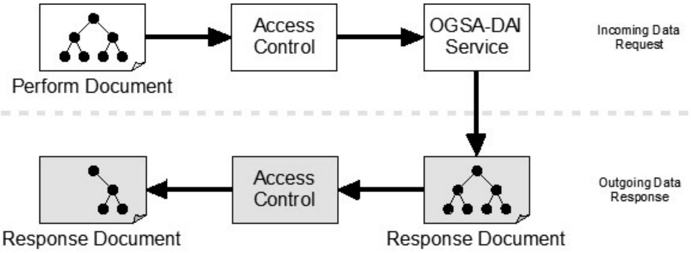

Figure 3: Filtering of outgoing Data Responses

5 Conclusion

Prospective clinical research conducted across multiple institutions is necessary in order to cope with the ever increasing complexity of medical data and to achieve the required accuracy and power in research results. Health-related data platforms relying on SOA principles provide the technological basis to integrate distinct data sources and henceforth to enable multi-institutional research in life science. Important issues to be considered from the very beginning in the system development process are trust, security and privacy. This paper presented a security architecture for such distributed medical information systems building upon a claim-based security model. The introduced security system allows the efficient management of the VO and control to the provided resources. It is itself realized using Web service technologies and SOA principles and integrates therefore seamlessly into the overall service ecosystem.

Acknowledgement

This work was generated in the framework of the @neurIST Integrated Project, which is co-financed by the European Commission through the contract no. IST-027703.

References

[1] A. Arbona, S. Benkner, G. Engelbrecht, J. Fingberg, M. Hofmann, K. Kumpf, G. Lonsdale, A. Woehrer, *A service-oriented grid infrastructure for biomedical data and compute services*. IEEE Transactions NanoBioscience, vol. 6(2), pp. 136-141, June 2007.

[2] H. Rajasekaran, P. Hasselmeyer, L. Lo Iacono, J. Fingberg, P. Summers, S. Benkner, G. Engelbrecht, A. Arbona, A. Chiarini, C.M. Friedrich, M. Hofmann-Apitius, B. Moore, P. Bijlenga, J. Iavindrasana, H. Müller, R.D. Hose, R. Dunlop, A. Frangi, and K. Kumpf, *@neurIST – Towards a System Architecture for Advanced Disease Management through Integration of Heterogeneous Data, Computing, and Complex Processing Services*. In Proceedings of 21 IEEE International Symposium on Computer-Based Medical Systems 2008 (CBMS 2008), Finland, June 2008.

[3] R. Dunlop, A. Arbona, H. Rajasekaran, L. Lo Iacono, J. Fingberg, P. Summers, S. Benkner, G. Engelbrecht, A. Chiarini, C.M. Friedrich, B. Moore, P. Bijlenga, J. Iavindrasana, R.D. Hose, A.F. Frangi, *@neurIST – Chronic Disease Management through Integration of Heterogeneous Data and Computer-interpretable Guideline Services*. In Proceedings of HealthGrid 2008, Chicago, June 2008.

[4] C. Lovis, S. Spahni, N. Cassoni-Schoellhammer, and A. Geissbuhler, *Comprehensive management of the access to a component-based healthcare information system*. Stud Health Technol Inform, vol. 124, pp. 251-256, 2006.

[5] A. Geissbuhler, C. Lovis, A. Lamb, and S. Spahni, *Experience with an XML/http-based federative approach to develop a hospital-wide clinical information system*. Medinfo, vol. 10, pp. 735-739, 2001.

[6] S. Cantor, J. Kemp, R. Philpott, and E. Maler (Editors), *Assertions and Protocols for the OASIS Security Assertion Markup Language (SAML) V2.0*, OASIS Standard, March 2005. Available online at: http://www.oasis-open.org/committees/security/

[7] T. Moses (Editor), *Extensible Access Control Markup Language (XACML) Version 2.0*, OASIS Standard, February 2005. Available online at:
http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml

[8] American Medical Informatics Association (AMIA). *A Taxonomy of Secondary Uses and Re-Uses of Healthcare Data*. Invitational Conference on Secondary Use of Health Data, September 2007. Available online at:
http://www.amia.org/inside/initiatives/healthdata/2007/index.asp

[9] K. Karasavvas, M. Antonioletti, M.P. Atkinson, N.P. Chue Hong, T. Sugden, A.C. Hume, M. Jackson, A. Krause, and C. Palansuriya, *Introduction to OGSA-DAI Services*. Lecture Notes in Computer Science, vol. 3458, pp. 1-12, May 2005.