

Scientific Report

Period: December 2007 – August 2008

I. Fellow information

Name: Jesús Luna García

Research Institute: University of Cyprus.

II. Research Summary

During this period I did a research stay at the “University of Cyprus” (Cyprus), where the following topics were investigated:

1. Technical and legal security issues for Intensive Care Grids.

The Data Grid is becoming a new paradigm for eHealth systems due to its enormous storage potential using decentralized resources managed by different organizations. The storage capabilities in these novel “Health Grids” are quite suitable for the requirements of systems like the Intensive Care Grid (ICGrid, being developed at the University of Cyprus), which captures, stores and manages data and metadata from Intensive Care Units –ICUs- (Figure 1).

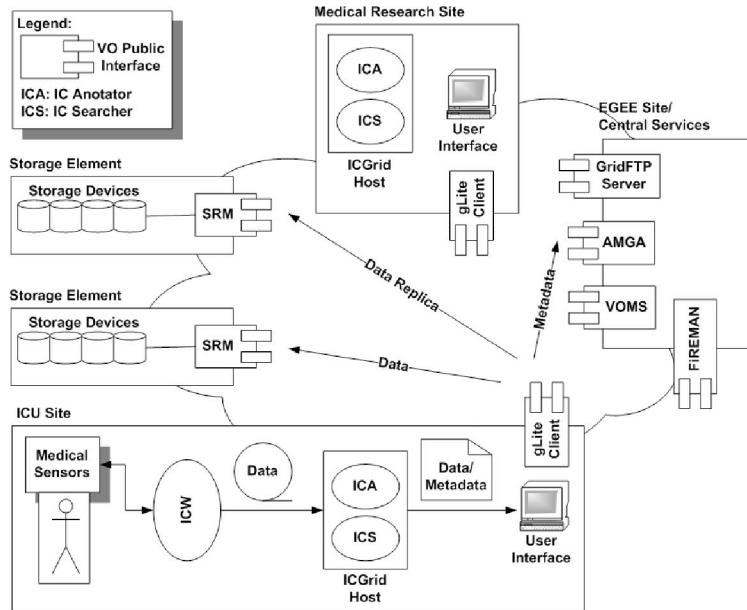


Figure 1. Intensive Care Grid: data and metadata architecture.

However, this paradigm depends on a widely distributed storage sites, therefore requiring new security mechanisms, able to avoid potential leaks to cope with modification and destruction of stored data under the presence of external or internal attacks. Particular emphasis must be put on the patient's personal data, the protection of which is required by legislations in many countries of the European Union and the world in general.

This part of our research has investigated and surveyed underlying data protection legislations (applicable to EU-Member States) and technological data privacy mechanisms (i.e. the Electronic Health Cards and the Grid Security Infrastructure), to obtain the minimum data-centric security requirements for ICGrid. This was the starting point to prototype a privacy protocol, based on the building blocks researched in the first part of this fellowship (FORTH ICS), for the Intensive Care Grid.

2. Security analysis of the Intensive Care Grid.

Taking into consideration underlying data protection legislations and technological data privacy mechanisms (researched in Activity 1 above), in this part of the fellowship we identified the security issues related with ICGrid's data and metadata after applying an analysis framework extended from our previous research on the Data Grid's storage services (done at FORTH ICS).

In a nutshell, the use of this framework consists of determining the basic components related with the system's security (players, attacks, security primitives, granularity of protection and user inconvenience), so that afterwards they can be summarized to clearly represent its security requirements. Figure 2 shows the result of applying the cited framework to ICGrid.

| Damage | Adversary on the wire | | | Revoked user w/Central Service | | | Adversary w/Storage Site | | |
|--------|-----------------------|---|---|--------------------------------|---|---|--------------------------|---|---|
| | L | C | D | L | C | D | L | C | D |
| ICGrid | N | N | Y | Y | Y | Y | Y | Y | Y |

Figure 2. Data-centric security issues related with ICGrid.

3. Security model to protect ICGrid metadata.

The specific goal of this research was to avoid metadata attacks (leakage, change or destruction just as identified by the previous activity) while at-rest into the untrusted Storage Elements. Because our security analysis found that ICGrid's metadata and data require different security policies, the enforcement mechanisms designed for each one of them also required a differentiated approach.

Our research proposed the implementation of a Mandatory Access Control (MAC) model via the access control lists provided by state of the art metadata servers (i.e. gLite's AMGA server). This "multi-layer" security mechanism was inspired on the Bell-Lapadula's model and the Electronic Health Card, currently being deployed in the European Union (as found by previous research). Despite its simplicity, the researched approach also enforces different levels of authorization for a patient's personal data, in compliance with the eHealth Legislations studied in our previous activities. The proposed MAC model is able to provide a basic level of confidentiality to the patient's private metadata, while at the same time "protecting" him from accidentally disclosing this information to the lower-security levels. As a proof of concept we applied this model to the ICGrid metadata, which uses a basic access control mechanism with POSIX-like permissions. Based on this premise it was possible to provide a basic MAC model like the one seen in Figure 3, where we have defined three different players (Patient -owner-, Paramedics -group- and the Intensive Care Unit Receptionist -others-) and also, three levels of authorization (Public, Semi-Private and Private).

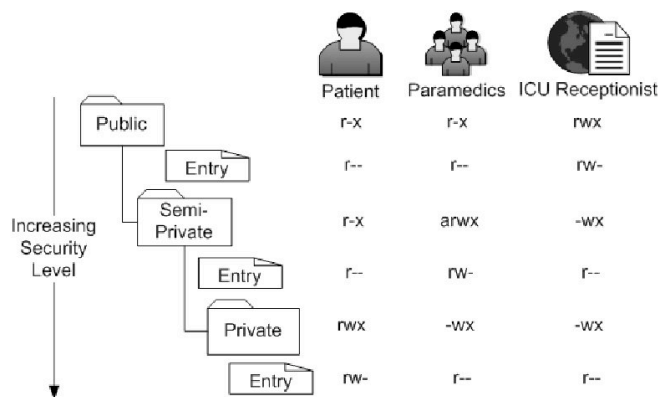


Figure 3. Implementing a Mandatory Access Control for ICGrid metadata. The proposed mechanism is based on the glite's AMGA service.

4. Security model to protect ICGrid data.

In this final part of the fellowship's research (and complementary to the one described above), we designed and prototyped the main components of an architecture proposed to provide security to the data being managed by ICGrid. It is worth noticing that performance issues related with the cryptographic and fragmentation mechanisms being used by the mentioned architecture were carefully analyzed via simulations.

Figure 4 and Figure 5 show how the different *Privacy Services* of the proposed architecture interact with the elements of ICGrid.

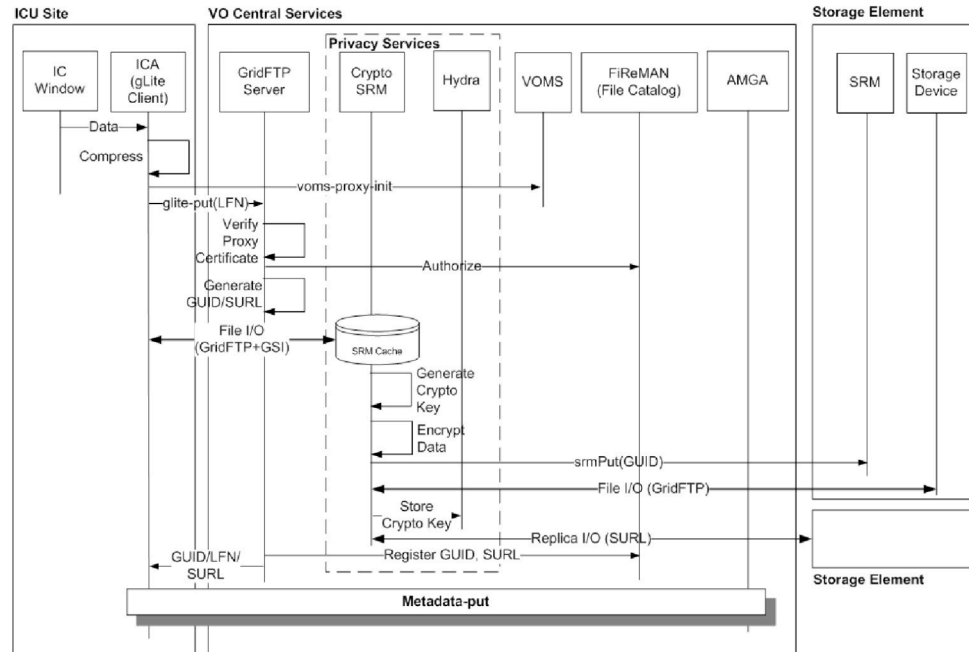


Figure 4. Privacy protocol proposed to transfer data and metadata from ICGrid to EGEE's Storage Elements.

In this protocol we use the following file naming notation when referring to the data being managed by the Grid: (i) Logical File Name -LFN- (a human readable identifier for a file), (ii) Global Unique Identifier -GUID- (a logical identifier which guarantees its uniqueness by construction) and, (iii) Site URL -SURL- (specifies a physical instance of a file replica, which is accepted by the Storage Element's SRM interface).

The core of our proposal is the *CryptoSRM*, which is responsible for symmetrically encrypting the staged data, previously transferred via a secure channel by the ICA's GridFTP client. So far, results obtained with a prototype of our Privacy Protocol have demonstrated an acceptable performance when compared versus more traditional approaches (i.e. implementing the cryptosystem at the Grid Client or Storage Elements). Figure 6 shows an example of our results when using a data sample from a Hospital's ICU.

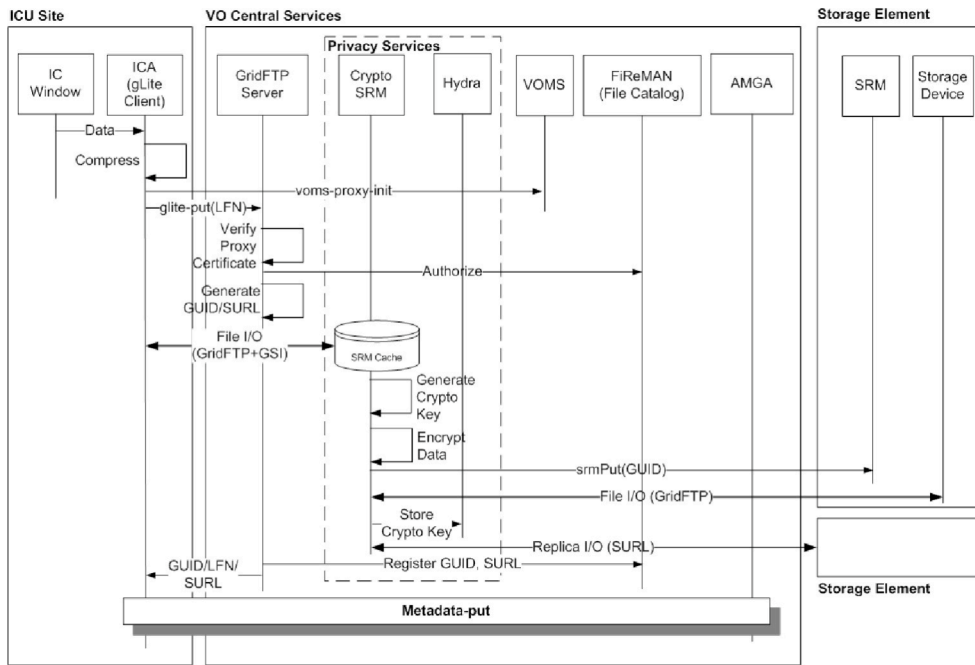


Figure 5. Retrieving data and metadata from ICGrid with the proposed privacy protocol.

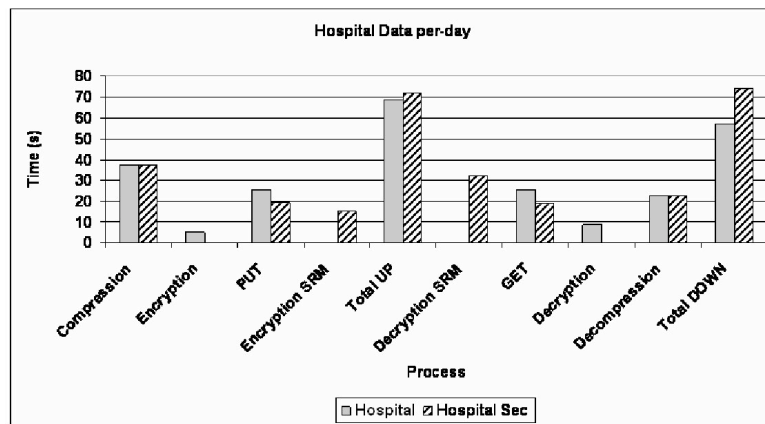


Figure 6. Performance results of the proposed privacy protocol. The test used a data sample from a Hospital's ICU.

III. Publications

- J. Luna, M. Flouris, M. Marazakis, A. Bilas, M. Dikaiakos, H. Gjermundrod, and T. Kyprianou. A data-centric security analysis of ICGrid. In Proceedings of the CoreGRID Integrated Research in Grid Computing, pages 165–176, April 2008. (Also as CoreGRID TR-0145)
Abstract: The Data Grid is becoming a new paradigm for eHealth systems due to its enormous storage potential using decentralized resources managed by different organizations. The storage capabilities in these novel “Health Grids” are quite suitable for the requirements of systems like ICGrid, which captures, stores and manages data and metadata from Intensive Care Units. However, this paradigm depends on widely distributed storage sites, therefore requiring new security mechanisms, able to avoid potential leaks to cope with modification and destruction of stored data under the presence of external or internal attacks. Particular emphasis must be put on the patient's personal data, the protection of which is required by legislations in many countries of the European Union and the world in general. Taking into consideration underlying data protection legislations and technological data privacy mechanisms, in this paper we identify the security issues related with ICGrid's data and metadata after applying an analysis framework extended from our previous research on the Data Grid's storage services.

Then, we present a privacy protocol that demonstrates the use of two basic approaches (encryption and fragmentation) to protect patients' private data stored using the ICGrid system.

- J. Luna, M. Flouris, M. Marazakis and A. Bilas. Providing security to the Desktop Data Grid. CoreGRID Technical Report 0144. May, 2008.

Abstract: Volunteer Computing is becoming a new paradigm not only for the Computational Grid, but also for institutions using production-level Data Grids because of the enormous storage potential that may be achieved at a low cost by using commodity hardware within their own computing premises.

However, this novel "Desktop Data Grid" depends on a set of widely distributed and *untrusted* storage nodes, therefore offering no guarantees about neither availability nor protection to the stored data.

These security challenges must be carefully managed before fully deploying Desktop Data Grids in sensitive environments (such as eHealth) to cope with a broad range of storage needs, including backup and caching.

In this paper we propose a cryptographic protocol able to fulfill the storage security requirements related with a generic Desktop Data Grid scenario, which were identified after applying an analysis framework extended from our previous research on the Data Grid's storage services. The proposed protocol uses three basic mechanisms to accomplish its goal: (a) symmetric cryptography and hashing, (b) an Information Dispersal Algorithm and the novel (c) "Quality of Security" (QoSec) quantitative metric.

Although the focus of this work is the associated protocol, we also present an early evaluation using an analytical model. Our results show a strong relationship between the assurance of the data at rest, the QoSec of the Volunteer Storage Client and the number of fragments required to rebuild the original file.

- J. Luna, M. Flouris, M. Marazakis, A. Bilas, M. Dikaiakos, H. Gjermundrod, and T. Kyprianou. Using the gLite middleware to implement a secure Intensive Care Grid System. In CoreGRID Workshop on Grid Middleware 2008. June 2008

Abstract: Storage capabilities in novel "Health Grids" are quite suitable for the requirements of systems like ICGrid, which captures, stores and manages data and metadata from Intensive Care Units. However, this paradigm depends on widely distributed storage sites, therefore requiring new security mechanisms, able to avoid potential leaks to cope with modification and destruction of stored data under the presence of external or internal attacks.

Particular emphasis must be put on the patient's personal data, the protection of which is required by legislations in many countries of the European Union and the world in general. In a previous paper we performed a security analysis of ICGrid, from the point of view of metadata and data, where we found the need to protect the data-at-rest from untrusted Storage Elements (SE). That research also proposed a privacy protocol to protect a patients' private metadata and data.

This paper is the follow-up of our previous research, proposing an architecture based on gLite middleware's components, to deploy the contributed privacy protocol. As a proof of concept we show how to implement a Mandatory Access Control model for the metadata stored into the AMGA service.

To protect the data itself, this paper presents our first experimental results on the performance that can be achieved with a prototyped "cryptographic" Storage Resource Manager -CryptoSRM- service.

Obtained results show that encrypting and decrypting at the CryptoSRM, instead of doing these at the SE or even at the Grid client, not only improve overall security, but also exhibit a higher performance that can be further improved with the aid of specialized hardware accelerators.

- J. Luna, M. Dikaiakos, T. Kyprianou, A. Bilas and M. Marazakis. Data Privacy considerations in Intensive Care Grids. In Global Healthgrid: e-Science meets Biomedical Informatics. Proceedings of HealthGrid Conference 2008, pages 178-187. IOS Press, June 2008. ISBN: 978-1-58603-874-8

Abstract: Novel eHealth systems are being designed to provide a citizen-centered health system, however the even demanding need for computing and data resources has required the adoption of Grid technologies.

In most of the cases, this novel *Health Grid* requires not only conveying patient's personal data through public networks, but also storing it into shared resources out of the hospital premises. These features introduce new security concerns, in particular related with privacy.

In this paper we survey current legal and technological approaches that have been taken to protect a patient's personal data into eHealth systems, with a particular focus in Intensive Care Grids.

However, thanks to a security analysis applied over the Intensive Care Grid system (ICGrid) we show that these security mechanisms are not enough to provide a comprehensive solution, mainly because the data-at-rest is still vulnerable to attacks coming from untrusted Storage Elements where an attacker may directly access them.

To cope with these issues, we propose a new privacy-oriented protocol, which uses a combination of encryption and fragmentation to improve data's assurance while keeping compatibility with current legislations and Health Grid security mechanisms.

- WP2 Working Group. Deliverable KDM.06 –Report on the four years of joint activities, open topics and future research in Knowledge and Data Management, August 31, 2008.

Abstract (Excerpt): This document reports on the four years of joint research activities, open topics and future research in the area of knowledge and data management in Grid and P2P systems of the CoreGRID Institute on Knowledge and Data Management (Workpackage 2, WP2). The report structure reflects the scientific organization of the KDM Institute that is based on a set of research groups that implemented the Institute roadmap and carried out the research activities included in the three tasks of the Institute: Distributed Storage Management, Information and Knowledge Management, and Data Mining and Knowledge Discovery.

IV. Workshops and conferences

- “CoreGRID Workshop on Integrated Research in Grid Computing”, April 2008. Crete, Greece.
- “2nd Workshop on Desktop Grids and Volunteer Computing Systems (co-located with IPDPS 2008)”, May 2008. Florida, U.S.
- “CoreGRID Workshop on Grid Middleware 2008”, June 2008. Barcelona, Spain.
- “HealthGrid Conference 2008”, June 2008. Chicago. U.S.