European Research Network on Foundations, Software Infrastructures and Applications for large scale distributed, GRID and Peer-to-Peer Technologies

A Network of Excellence funded by the European Commission

A data-centric security analysis of ICGrid

Jesus Luna, Marios D. Dikaiakos, Harald Gjermundrod
{jluna, mdd, harald}@cs.ucy.ac.cy
Department of Computer Science,
University of Cyprus,
PO Box 1678. Nicosia, Cyprus

Michail Flouris, Manolis Marazakis, Angelos Bilas
{flouris, maraz, bilas}@ics.forth.gr
Institute of Computer Science (ICS),
Foundation for Research and Technology - Hellas (FORTH),
PO Box 1385. GR-71110. Heraklion, Greece.

Theodoros Kyprianou
drtheo@cytanet.com.cy
Intensive Care Unit,
Nicosia General Hospital
Nicosia, Cyprus

CoreGRID Technical Report
Number TR-0145
May 19, 2008

Institute on Knowledge and Data Management

CoreGRID - Network of Excellence
URL: http://www.coregrid.net

CoreGRID is a Network of Excellence funded by the European Commission under the Sixth Framework Programme

Project no. FP6-004265

A data-centric security analysis of ICGrid

Jesus Luna, Marios D. Dikaiakos, Harald Gjermundrod
{jluna, mdd, harald}@cs.ucy.ac.cy
Department of Computer Science,
University of Cyprus,
PO Box 1678. Nicosia, Cyprus

Michail Flouris, Manolis Marazakis, Angelos Bilas
{flouris, maraz, bilas}@ics.forth.gr
Institute of Computer Science (ICS),
Foundation for Research and Technology - Hellas (FORTH),
PO Box 1385. GR-71110. Heraklion, Greece.

Theodoros Kyprianou
drtheo@cytanet.com.cy
Intensive Care Unit,
Nicosia General Hospital
Nicosia, Cyprus

CoreGRID TR-0145
May 19, 2008

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 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. 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.

1 Introduction

Modern eHealth systems require advanced computing and storage capabilities, leading to the adoption of technologies like the Grid and giving birth to novel *Health Grid* systems. In particular, Intensive Care Medicine uses this paradigm when facing a high flow of data coming from Intensive Care Unit's (ICU) inpatients. These data needs to be stored, so for example data-mining techniques could be used afterwards to find helpful correlations for the practitioners facing

This research work is carried out under the FP6 Network of Excellence CoreGRID funded by the European Commission (Contract IST-2002-004265).

similar problems. Unfortunately, moving an ICU patient's data from the *traditionally isolated* hospital's computing facilities to Data Grids via public networks (i.e. the Internet) makes it imperative to establish an integral and standardized security solution to avoid common attacks on the data and metadata being managed.

As mandated by current Data Protection Legislations [24], a patient's personal data must be kept private because *data privacy means eHealth trust*, therefore comprehensive privacy mechanism are being developed for the Health Grid, harmonizing legal and technological approaches. To provide solutions it is necessary to consider privacy from a *layered* point of view: legal issues are the common base above which state-of-the-art security technologies are deployed. In our previous research related with the security analysis of Grid Storage Systems [20] we concluded that current technological mechanisms are not providing comprehensive privacy solutions and worst of all, several security gaps at the storage level are still open.

There is a clear need not only to identify the vulnerabilities associated with Health Grids, but also for designing new mechanisms able to provide confidentiality, availability, and integrity to the Data Grid in general. Towards this end, the first part of the research presented in this paper shows the result of applying a security analysis framework (extended at the Foundation for Research and Technology - Hellas) over an *Intensive Care Grid* scenario (the ICGrid system developed by the University of Cyprus [17]); this has proven that the greatest threat to patient's privacy comes in fact from the Data Grid's Storage Elements, which are untrusted and may easily leak personal data. In an effort to cover these privacy gaps, the second part of this paper contributes with a *low-level* protocol for providing privacy to current Intensive Care Grid systems from a data-centric point of view, but taking into account the legal framework and keeping compliance with *high-level* mechanisms. The contributed protocol proposes the use of two basic mechanisms to enhance a patient's data assurance: cryptography and fragmentation.

The rest of this paper is organized as follows: Section 2 reviews the basic terminology related with Intensive Care Medicine and the ICGrid system. The basic underlying technological and legal security approaches for Health Grids are presented in Section 3. Section 4 briefly presents and then applies the security analysis framework to ICGrid's data and metadata. Section 5 describes a gLite-based middleware architecture required to implement the proposed privacy protocol for ICGrid. Our first experimental results on the cryptographic performance achieved by our proposal are shown in Section 6. Section 7 briefly presents the State of the Art related with our research. Finally, Section 8 presents our conclusions and future work.

2 The ICGrid system

In this Section we introduce the required background and the respective terminology for Intensive Care Medicine, which is the basis of the ICGrid system analyzed in this paper.

2.1 Intensive Care Medicine

An Intensive Care Unit (ICU) is the only environment in clinical medicine where all patients are monitored closely and in detail for extended periods of time, using different types of *Medical Monitoring Devices (MMD)*. An MMD may be defined as a collection of sensors that acquire the patients' physiological parameters and transform them into comprehensible numbers, figures, waveforms, images or sounds. Taking clinical decisions for the ICU patients based on monitoring can be a very demanding and complex task requiring thorough analysis of the clinical data provided: *even the most skilled physicians are often overwhelmed by huge volumes of data, a case that may lead to errors, or may cause some form of life threatening situation* [12]. Providing systems that actively learn from previously stored data and suggest diagnosis and prognosis is a problem that, to our knowledge, has been overlooked in previous Intensive Care Medicine research.

Traditionally, medical research is guided by either the concept of patients' similarities (clinical syndromes, groups of patients) or dissimilarities (genetic predisposition and case studies). Clinical practice also involves the application of commonly (globally) accepted diagnostic/therapeutic rules (*evidence-based medicine* [14]) as well as *case-tailored approaches* which can vary from country to country, from hospital to hospital, or even from physician to physician within the same hospital. These different approaches in treating similar incidents produce knowledge which, most of the times, remains a personal/local expertise, not documented in detail and not tested against other similar data. Global sharing of this cumulative national/international experience would be an important contribution to clinical medicine in the sense that one would be able to examine and follow up implementation of and adherence to guidelines as well as to get the benefit of sharing outstanding experience from physicians.

2.2 ICGrid: data and metadata architecture

Although a number of dedicated and commercially available information systems have been proposed for use in Intensive Care Units (ICUs) [13], which support real-time data acquisition, data validation and storage, analysis of data, reporting and charting of the findings, none of these systems was appropriate in our application context. Another important issue with ICU is the need for data storage: an estimate of the amount of data that would be generated daily is given in the following scenario. Suppose that each sensor is acquiring data for storage and processing at a rate of 50 bytes per second (it is stored as text) and that there are 100 hospitals with 10 beds each, where each bed has 100 sensors. Assuming that each bed is used for 2 hours per day, the data collected amounts to 33.5275 GB per day. But this number only represents the data from the sensors. Additional information includes metadata, images, etc. Because Grids represented a promising venue for addressing the challenges described above, the Intensive Care Grid (ICGrid) system [17] has been prototyped over the EGEE infrastructure (Enabling Grids for E-sciencE [1]). ICGrid is based on a hybrid architecture that combines a heterogeneous set of monitors that sense the inpatients and three Grid-enabled software tools that support the storage, processing and information sharing tasks.

The diagram of Figure 1 represents a Virtual Organization of the ICGrid system, which depicts the acquisition and annotation of parameters of an inpatient at an ICU Site (bottom left) and the transfer of data replicas to two *Storage Elements (SEs)*. The transfer comprises the actual sensor data, denoted as *Data*, and the information which is provided by physicians during the annotation phase, denoted as *Metadata*. We utilize the notion of a *Clinically Interesting Episode (CIE)* to refer to the captured sensor data along with the metadata that is added by the physician to annotate all the events of interest.

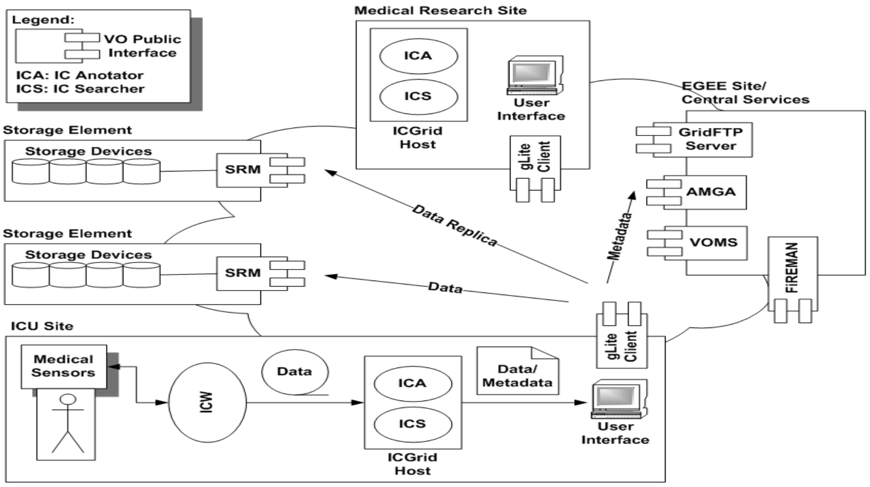


Figure 1: Architecture of an ICGrid's Virtual Organization.

When ICGrid's Data and Metadata are transferred to Storage Elements and Metadata servers (currently a gLite Metadata Catalogue -AMGA- service [28]) respectively, a set of messages are exchanged among the different entities. In particular we should highlight that file catalog services are being provided by FiReMAN (File Replication MAnager [11]) and, authorization mechanisms rely on the X.509 credentials issued by the Virtual Organization Membership Service (VOMS [8]).

3 Health Grid Privacy: legal and technological aspects

As mentioned in Section 1, comprehensive privacy solutions for Health Grids need the synergy of two different factors, legislation and technology.

3.1 Legal aspects

A major concern in eHealth is adequate confidentiality of the individual records being managed electronically. The core component of many eHealth systems is the Electronic Health Record (EHR), which is basically the patient's health record in digital format. Nowadays EHR protection is the focus of privacy legislations around the globe. In the European Union, several Directives of the European Parliament and of the Council protect the processing and free movement of the EHR. The common factor of all these initiatives is the EU Directive on Data Protection [24], which provides the general framework for the protection of privacy with respect to the processing of personal data in its widest sense. This Directive goes further than the protection of the intimacy of the natural persons since it defines *personal data* as all data related to an individual person's private, public, or professional life. However, the European Working Party on Data Protection, which was established under article 29 of the Directive [24] and comprises all national data protection authority of EU Member States, has recently acknowledged that some special rules may need to be adopted for key eHealth applications.

A common term referenced in current eHealth legislations is the concept of *consent*. Such consent is defined as any unambiguous, freely given, specific and informed indication of the patient's wishes by which he agrees to the processing of his personal data. In other words, *a patient's consent enables the legal processing of his EHR*. However, what happens if, for instance, after an accident the patient is unable to give his consent for accessing his personal data at the Intensive Care Unit? Most of the legal issues and ambiguities related with eHealth regulations are being carefully studied; in the particular case of the European Union, the European Health Management Association (EHMA) along with the Commission established the "Legally eHealth" [9] project to study these. This study gives three basic recommendations regarding the protection of patients' data, using the terminology from RFC 2196 (see [16], section 4). These recommendations can be mapped to the *security services* shown in table 1. The next section presents the security analysis of an eHealth scenario as a way to identify not only its strengths, but in particular its vulnerabilities, according to the requirements stated in table 1. With this information, and towards implementing a comprehensive and harmonized solution, we will introduce in the rest of this paper a novel low-level privacy protocol for Intensive Care Grids.

Table 1: Security requirements for implementing Data Protection Legislations in eHealth environments

Legal Issue	Security Requirement	Example
Patient's Consent	Authentication, Non-repudiation, Integrity	A patient must be confidently identified (authentication) before signing an agreement (non-repudiation) allowing processing his EHR. The signed document should not be modified afterwards (integrity) without notifying the patient.
Specified Purpose	Authorization, Confidentiality, Integrity	If a patient has given his consent to re-use parts of his demographic data for statistical purposes, then a pharmaceutical company should not have access to this information (authorization) and even the personnel authorized to process these statistics should not be able to disclose i.e. the patient's name (confidentiality) or modify the record at all (integrity).

3.2 Technological approach

Enforcing privacy of patient's data in Health Grids have spawned the development of a broad range of mechanisms. Two of these are particularly important for our research because of their wide use: the Grid Security Infrastructure and the Electronic Health Card.

3.2.1 Grid Security Infrastructure

The Grid Security Infrastructure (GSI) [31] is comprised of a set of protocols, libraries, and tools that allow users and applications to securely access Grid resources via well defined Authentication and Authorization mechanisms. In the first case, the Grid client simply uses an X.509 end entity certificate to secure messages and authenticate itself to the Grid service. On the other hand, for Authorization purposes GSI can use XML-based protocols to retrieve security assertions from third-party services to enable features like role-based authorization. One of these third-party Grid authorization services, widely used in EGEE, is the *Virtual Organization Membership Service* (VOMS) [8]: an Attribute Authority that exposes attributes and encodes the position of the holder inside the VO. Despite its functionalities, nowadays Grid Authentication and Authorization systems are unable to enforce access control close to the Storage Elements and the data itself, in other words, an attacker passing over these security mechanisms (i.e. using a local account with administrative privileges or accessing physically the disks) will have full control over the stored data. These vulnerabilities will be analyzed in section 4.

3.2.2 Electronic Health Card

Member States have began testing the Electronic Health Card [23], a new health card that contains basic patient data such as name, age, insurance details, and electronic prescriptions. The card includes also physical features to identify the owner, i.e. a photograph and human-readable information. With time, this card will replace EU's existing health insurance cards. Basically this card is a smartcard that stores information in a microchip supporting authentication, authorization and even digital signature creation. Data protection issues were critical in the design of Electronic Health Cards, so patients must be able to rely on maximum security and confidentiality while operating smoothly in practice. A comprehensive security concept secures the protection of particularly sensitive data, so with few exceptions, the health card can only be used in conjunction with an *Electronic Health Professional Card*, which carries a "qualified" electronic signature (one that meets strict statutory criteria for electronic signatures). In general, Electronic Health Cards represent a big step towards creating a citizen-centered health system, but despite its security advantages, internal storage space is quite limited (just few kilobytes). Thus, the use of the card must rely on external storage services over which the card can not offer protection mechanisms. The next section will analyze in detail these security gaps.

4 Use Case: security analysis of ICGrid

From the point of view of a typical Health Grid system, its subsystems may be attacked in several ways. Nevertheless, for the purposes of our research on data privacy, the framework proposed in [27] and extended in [20] will be used to pinpoint the main concerns linked with the security of its data and metadata. 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. As a proof of concept, the security analysis in this Section will be performed in the content of the *Intensive Care Grid* system (ICGrid) (introduced in Section 2), considering also the underlying security mechanisms presented in Section 3.

4.1 Identifying the Elements for the Security Analysis

As mentioned at the beginning of this Section, the first step in our analysis is to identify the elements that play a security-related role in ICGrid:

1. *Players:* four data readers/writers are involved *(i)* the ICU and Medical Research sites that produce and consume the data; *(ii)* the EGEE Central Services that perform VO authentication and authorization as mentioned in Section 3.2; *(iii)* the EGEE *storage facilities* for data and metadata; and finally *(iv)* the "wire" or WAN links (public and private) conveying information between the other players.

2. *Attacks:* the generic attacks that may be executed over ICGrid are related with *(i)* Adversaries on the wire; *(ii)* Revoked users using valid credentials on the Central Services during a period of time -while the revocation data is propagated through the Grid-; and *(iii)* Adversaries with *full control* of the EGEE storage facilities. Each one of these attacks may result in data being leaked, changed or even destroyed.

3. *User inconvenience:* It is critical for IGGrid operation to have minimum latencies when reading and retrieving the stored data and metadata from the EGEE Site. Since smartcards -like the Electronic Health Card explained in Section 3.2.2- are beginning to be introduced into National Health Systems, it is feasible to consider that

involved entities (i.e. patients and physicians) will require them for performing operations into our Health Grid scenario.

4. *Security Primitives:* Two security operations take place within the ICGrid: *(i) Authentication and Authorization* via GSI-like mechanisms (section 3.2.1) and, *(ii) Consent* just as explained in Section 3.1.

5. *Trust Assumptions:* We assume that *(i)* the security tokens used for authentication and consent (i.e. Electronic Health Cards) are personal, intransferable and tamper-resistant; *(ii)* EGEE Sites and/or ICU premises have full control over the data and metadata stored on them; *(iii)* data are encrypted on the public link thanks to secure functionalities (i.e. via SSL); and *(iv)* the EGEE Central Services are *trusted* because they are managed in a secure manner, therefore providing high assurance to its operations.

4.2 Security Analysis Results

Based on the elements identified in the previous Section, Table 2 summarizes the vulnerabilities identified in the ICGrid system. Results are categorized by possible attacks (main columns) and types of damage – the Leak (L), Change (C), Destroy (D) sub-columns. Cells marked with a "Y" mean that the system (row) is vulnerable to the type of damage caused by this particular attack. Cells marked with a "N" mean that the attacks are not feasible, or cannot cause a critical damage.

Table 2: Summary of security issues related with ICGrid

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

From Table 2 we conclude that current Health Grid Authentication and Authorization systems like the ones presented in Section 3.2 are unable to enforce access control close to the Storage Elements and the data itself. In other words, an attacker that bypasses these security mechanisms (by using a local account with administrative privileges or by physical access to the disks) will have full control over the stored data. Unfortunately, merely using cryptography at the Storage Elements is not a viable solution, and moreover imposes a significant performance penalty. In the following Section, we introduce a protocol designed to address these particular privacy concerns.

5 Secure ICGrid: protecting Metadata and Data

In this section we will present the main components of an architecture proposed to provide security to the ICGrid system introduced in Section 2. The specific goal of our proposal is to avoid data and metadata attacks (leakage, change or destruction) while at-rest into the untrusted Storage Elements. It is worth noticing that performance issues related with the cryptographic mechanism have been carefully considered in our design (more about this in Section 6). Because our previous security analysis [21] found that ICGrid's metadata and data require different security policies, the enforcement mechanisms presented in this section implement a differentiated approach for metadata (Section 5.2) and data (Section 5.3).

5.1 Architecture

Based on ICGrid's current architecture (figure 1), our proposal contributes with the following *Privacy Services*, co-located with the Central Services (scoped at the Virtual Organization level) and interacting directly with the GridFTP Server [15] and AMGA:

- CryptoSRM: This component is a modified Storage Resource Manager that apart from implementing the interface defined in [25], uses a cryptographic engine for encrypting and decrypting staged data stored in its local cache.
- Hydra Key Store: Implements a secure repository for the encryption keys [3]. The repository itself uses a fragmentation algorithm [26] for providing confidentiality and high-availability to the cryptographic material.
- Secure Log: A secure logging service may help to back-trace potential abuses (even those performed by Grid administrators colluded with attackers).

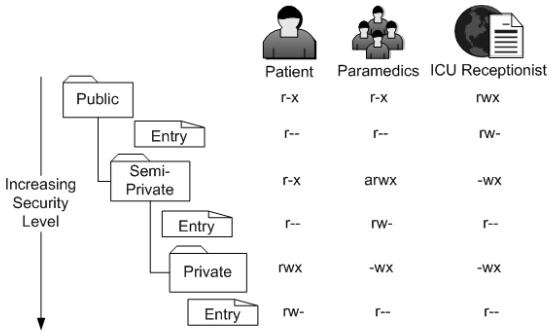
5.2 Metadata Security

AMGA stores metadata in a hierarchical structure that resembles a Unix File System, and also its native authorization model is based on Access Control Lists [5] with POSIX-like permissions per-entry and directory (*r*=read, *w*=write and *x*=change into directory) and, an additional "admin flag" allowing users in a group to administer the ACLs of an entry. Using the latter mechanism, we have defined an authorization model for ICGrid's metadata based on the Bell-LaPadula Model's Mandatory Access Control (MAC) rules [10]:

1. The *Simple Security Property* states that a subject at a given security level may not read an object at a higher security level (no read-up).
2. The **-Property* (read star-property) states that a subject at a given security level must not write to any object at a lower security level (no write-down) and, may only append new data to any object at a higher security level.

Bell-Lapadula's Model applied to ICGrid's metadata (implemented over AMGA) can be seen in figure 2. 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. In this example we have defined three different players (Patient -owner-, Paramedics -group- and the Intentive Care Unit Receptionist -others-) and also, three levels of authorization (Public, Semi-Private and Private). With the proposed AMGA's permissions on directories and entries it is possible to achieve the following Mandatory Access Control:

- Public Metadata: both Patient and Paramedics can read the entries, but only the ICU Receptionist can read and write them (i.e. schedule a new appointment with the physician).
- Semi-Private Metadata: the Paramedics can read and write entries (i.e. emergency information), the ICU Receptionist can only append new ones (the Paramedics group requires the admin flag to set read-only permissions to these newly created entries) and, the Patient is only able to read this metadata.
- Private Metadata: This is the most confidential level of the metadata, therefore only the Patient has full control over it (administrative permissions are implicit since he is the owner of his directories), while Paramedics and ICU Receptionists only can append new entries (the Patient must manage permissions of these newly created entries).



	Patient	Paramedics	ICU Receptionist
Public	r-x	r-x	rwx
Entry	r--	r--	rw-
Semi-Private	r-x	arwx	-wx
Entry	r--	rw-	r--
Private	rwx	-wx	-wx
Entry	rw-	r--	r--

Figure 2: Mandatory Access Control model for ICGrid's Metadata.

Enforcing the *-Property's append-only mode conveys an administrative overhead for both, Patients and Paramedics, which must manage permissions for entries being created by lower-security subjects. Also it is worth to notice that native AMGA's authorization mechanism can not prevent a malicious System Administrator from accessing the metadata of all the stored patients. To cope with these issues, our future work considers the use of cryptographic techniques to provide greater confidentiality and even a consent-like mechanism (based on electronic signatures) to AMGA's metadata. This research will be briefly introduced in Section 8.

5.3 Data Security

Using the Privacy Services discussed in Section 5.1 it is possible to improve overall security and privacy using cryptography. Figure 3 shows how the different Privacy Services interact with the Central Services when an IC Annotator (ICA) stores data into the ICGrid system. In this figure we use the file naming notation from [18], 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 Identifer -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. Afterwards the encryption key is securely stored in the Hydra service and the encrypted data moved to the untrusted Storage Element. It is obvious that attackers colluded with the latter will be unable to recover the original clear-text. A second scenario (Figure 4) considers an IC Searcher (ICS) retrieving data from the ICGrid: in this case the encrypted data is transferred from the Storage Element, decrypted at the CryptoSRM (the appropriate key is obtained from Hydra) and conveyed through a secure channel to the ICS' GridFTP client. *Notice that the encryption key is never disclosed to the ICS, therefore avoiding its leak by potential attackers (i.e. reading the DRAM like in [6]).* A more comprehensive analysis of the performance issues related with our proposal is presented in the next section.

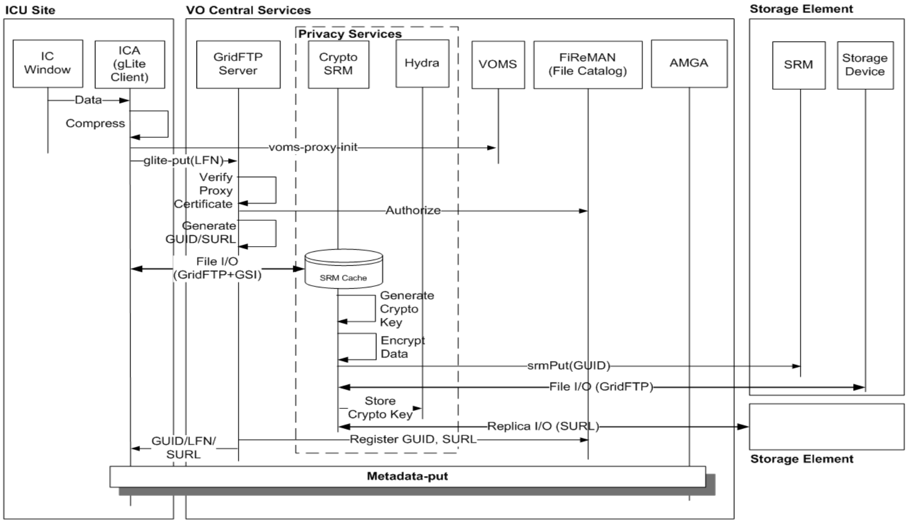


Figure 3: Secure ICGrid: transferring data.

6 Experimental Results

We have setup the following testbed to measure the expected performance to be achieved with the protocol proposed in Section 5.3:

- Grid client (GC): this CentOS4-based node has been configured as a "gLite User Interface". It is an IBM xSeries 335, with two Intel Xeon HT processors @ 2.8GHz and 2GB of RAM.

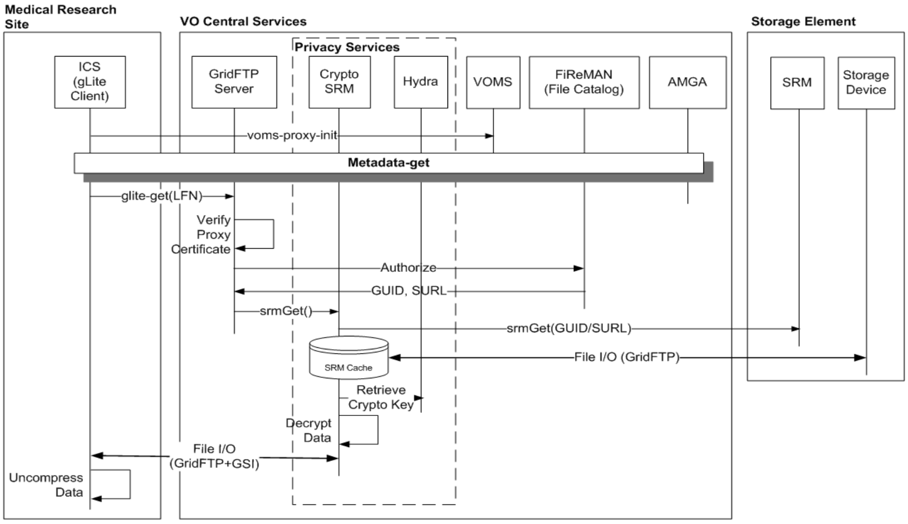


Figure 4: Secure ICGrid: retrieving data.

- Storage Element (SE): To simulate the basic functionalities of the proposed CryptoSRM, we have used for the tests a "DPM_mysql Storage Element" running over Scientific Linux version 3.09. The SE uses a Dell PowerEdge1400, with two Intel Pentium III processors @ 800MHz and 784MB of RAM.

For the Data, random samples corresponding to one day of ICGrid's operation were generated for *(i)* a sensor (approx. 352 Kb), *(ii)* a bed (approx. 35157 Kb) and, *(iii)* a Hospital (approx. 351563 Kb). The *gzip* utility is used with its default parameters for compression, while for encryption the *aes-128-cbc* algorithm from the OpenSSL library (version 0.9.8g) was used. For comparison purposes we have measured the protocol's performance as the User's time (reported by the Unix'x *time* command) consumed by each phase of the following scenarios:

1. Grid client Encryption: This approach performs encryption/decryption at the Grid client and is commonly used by existing solutions (see Section 7). The steps taking place are: data compression, encryption and transfer to the SE via clear-text FTP. The inverse sequence is used to retrieve it from the SE.

2. CryptoSRM Encryption: This scenario simulates the basic steps proposed by our protocol: data compress, transfer via a GSIFTP encrypted channel to the CryptoSRM and finally, encryption at this entity. The inverse sequence of steps is used to retrieve stored data from the simulated CryptoSRM.

Each test was repeated 50 times to isolate potential overhead being caused by other processes concurrently running at the server. Table 3 shows how the size of the three data samples changed after the compression and encryption processes. It is worth to notice that the compressed data's size is about 60% of the original one, however after encryption the size incremented approximately 35% for all the cases.

Table 3: Reported sizes (in KB) for the three ICGrid's Data Samples after compression and encryption

Data Sample	Original	Compressed	Encrypted
Sensor	352	213	288
Bed	35157	21213	28726
Hospital	351563	212125	287258

Figures 5, 6 and 7 show the performance results using ICGrid's data mentioned in Table 3. The three figures show a side-by-side comparison of the Grid client encryption (the Sensor, Bed and Hospital graphs), versus the CryptoSRM encryption (the Sensor-Sec, Bed-Sec and Hospital-Sec graphs). Aggregated values for the tested scenarios are given

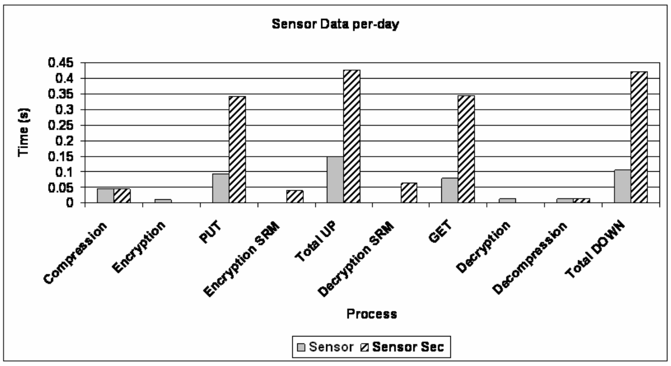


Figure 5: Processing a day of ICGrid's Sensor-data with the proposed privacy protocol.

by the *TOTAL UP* and *TOTAL DOWN* bars. Figure 5 shows the only case in which uploading and downloading Data through a secure GSI channel (the PUT and GET Sensor_Sec graphs), took more time than its equivalent via a clear-text FTP channel. This could be related to the GSI-transfer protocol itself, which for small data sizes requires more processing time (i.e. for encryption or padding). On the other hand for bigger data sizes, the performance achieved when uploading the Bed and Hospital Data (figures 6 and 7) is slightly less with the proposed privacy protocol (between 3%-4%) than with the Grid client encryption. This is because the data's size being uploaded to the SE is *smaller* in clear-text than when encrypted (around 30% according to Table 3), this latter fact helped to masquerade the overhead caused by the SE encryption mechanism (which provided approx. 20% of the TOTAL UP time). When downloading Data the overall performance of the proposed protocol was about 39%-47% less than that of the Grid client encryption, however we have found that most of this overhead is due to the decryption operation taking place at the SE (which spent around 45% of the TOTAL DOWN time). This behavior was predicted, as the used SE is more biased towards storage than processing (this can be easily seen by comparing its hardware configuration with that of the Grid client). Despite this configuration, the experimental results have shown the viability of using the proposed CryptoSRM and it can be foreseen that if both, the SE and the Grid client, would have at-least the same hardware configuration, then for the Hospital's Data our proposal would improve with about 17% for the TOTAL UP time, and approximately with 11% for the TOTAL DOWN time of the Grid client-based encryption approach.

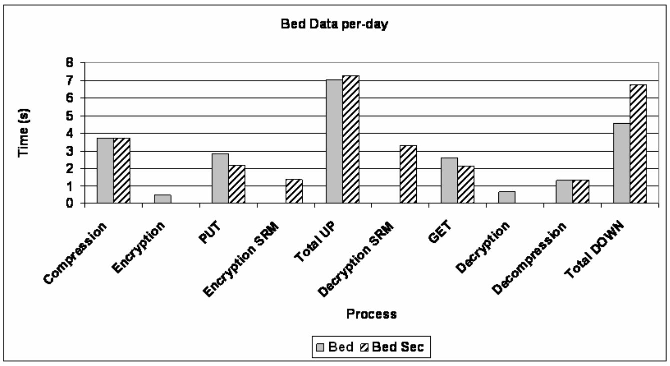


Figure 6: Processing a day of ICGrid's Bed-data with the proposed privacy protocol.

7 Related Work

Nowadays most of the work related with Health Grids' security and privacy focuses on "high-level" authentication and authorization mechanisms that rely on Grid-IDs and VOMS-like infrastructures [8], therefore leaving data vulnerable

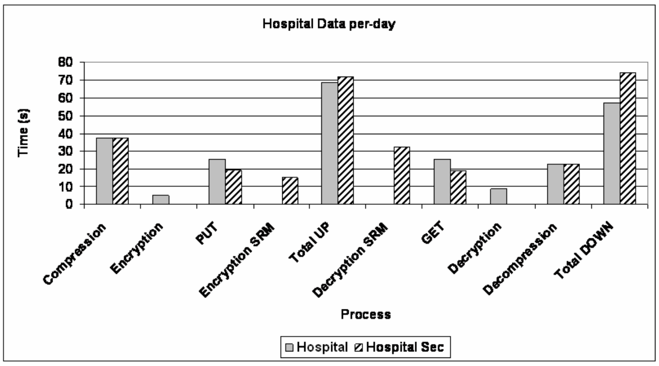


Figure 7: Processing a day of ICGrid's Hospital-data with the proposed privacy protocol.

in the untrusted Storage Elements. An example of these kind of mechanisms can be seen in the BRIDGES [29] and SHARE [4] Health Grids.

The research that is closely related with the work presented in this paper has been presented in [22], where the authors also used the gLite middleware to protect medical images. Their system ensures medical data protection through data access control, anonymization and encryption. A fundamental difference with our approach is the use of encryption at the Grid client, which requires retrieving the encryption key from a Hydra Keystore for decrypting the image. With our research it has been shown that such approach does not only introduce uncertainties about the key's confidentiality (it may be compromised at the Grid client), but also has a performance lower than our "centralized" proposal (using the CryptoSRM).

There are other state of the art distributed storage systems that, even though they have not been specifically designed for the Health Grid, they have focused on low-level data protection by implementing encryption mechanisms at the "Grid's edges" (therefore disclosing the encryption key to the untrusted SEs and Grid Clients). For example in OceanStore [19], stored data are protected with redundancy and cryptographic mechanisms. An interesting feature in OceanStore is the ability to perform server-side operations directly on the encrypted data, this increases system's performance without sacrificing security. On the other hand it is worth to mention the Farsite system [7], which provides security and high availability by storing encrypted replicas of each file on multiple machines.

A second group of related systems do not rely on cryptography, but in a "data fragmentation" scheme for data protection. In the first place let us mention POTSHARDS [30], which implements an storage system for long-time archiving that does not use encryption, but a mechanism called "probably secure secret splitting" that fragments the file to store prior to distributing it across separately-managed archives. A similar approach is given by Cleversafe [2] via an Information Dispersal Algorithm (based on the Reed-Solomon algorithm) for its open-source *Dispersed Storage Project*. In general both, POTSHARDS and Cleversafe, are interesting solutions that solves the management problems posed by cryptosystems and long-living data, however the security achieved only by fragmenting the files could not be strong enough for some highly-sensitive environments.

8 Conclusions

In this paper we have presented a follow-up to our research on data-level security for Health Grids. After analyzing in a previous work the security requirements of the proposed scenario, we found the need to protect Metadata and Data from untrusted Storage Elements and Grid Clients that could compromise sensitive material (i.e. cryptographic keys). The second part of this research proposed a privacy protocol to protect the patient's personal information (metadata) along with his data, using two basic mechanisms: encryption and fragmentation. This paper has proposed building such architecture using components from the gLite middleware, in particular the Hydra Keystore, the AMGA metadata service and a Storage Resource Manager with encryption facilities (the CryptoSRM).

About the Metadata, this paper proposed the implementation of an Mandatory Access Control model via AMGA's access control lists. This model was inspired in the Bell-Lapadula's model and the Electronic Health Card, currently being deployed in the European Union. Despite its simplicity, the proposed approach enforces different levels of authorization for a patient's personal data, in compliance with the eHealth Legislations studied in our previous work. However, we still have a lot of work to do in Metadata confidentiality, because currently AMGA is not able to offer protection from malicious administrators with direct access to its database.

Management of Health Grid's Data has taken a different approach in our proposal, so as a proof of concept to justify –from a performance point of view– the use of a "centralized" encryption mechanism (the CryptoSRM), in this paper we have simulated the former with a SE able to encrypt Data coming from an ICGrid client. Data's transfer operations (upload and download) resulted in most of the protocol's overhead, therefore suggesting us to keep transferred Data as small as possible. Taking into account that the encrypted Data is greater in size than its clear-text counterpart, we highly recommend not performing encryption at the "edges" of the Grid (i.e. Grid client, Storage Element). Notice that this argument is fully compatible with our previous security analysis, which established that Storage Elements are untrusted, thus encryption keys should not be delivered neither to them or even to the Grid Clients. Despite the hardware configuration being used to simulate the CryptoSRM in our experiments, it was possible to conclude its viability for the proposed privacy protocol. We can foresee that an important improvement in overall security and performance can be achieved, if the CryptoSRM uses a hardware-based cryptographic-accelerator, future work should prove this point.

As Future Work we are planning to study, along with AMGA's creators, the repercussions of using encryption at different levels of the Metadata. The next part of our ongoing research will also focus on the fragmentation mechanism originally proposed in [21], which benefits Data's availability and bandwidth use. We are planning to build analytical models able to show the relationship between Data's assurance, Data's fragments and incurred overhead. A prototype using Cleversafe's API (Section 7) will be also developed for our test.

Acknowledgments

The authors would like to thank EGEE-II (contract number INFSO-RI-031688) and Asterios Katsifodimos (University of Cyprus) for his technological support to perform the tests presented in this paper.

References

[1] Enabling Grids for E-SciencE project. http://www.eu-egee.org/.

[2] Cleversafe. http://www.cleversafe.com, 2007.

[3] Encrypted Storage and Hydra. https://twiki.cern.ch/twiki/bin/view/EGEE/DMEDS, September 2007.

[4] SHARE: Technology and Security Roadmap. http://wiki.healthgrid.org/index.php/Share_Roadmap_I, February 2007.

[5] Amga: Users, groups and acls. http://project-arda-dev.web.cern.ch/project-arda-dev/metadata/groups_and_acls.html, 2008.

[6] Disk encryption easily cracked. http://www.networkworld.com/news/2008/022108-disk-encryption-cracked.html, 2008.

[7] Atul Adya, William J. Bolosky, Miguel Castro, Gerald Cermak, Ronnie Chaiken, John R. Douceur, Jon Howell, Jacob R. Lorch, Marvin Theimer, and Roger Wattenhofer. Farsite: Federated, available, and reliable storage for an incompletely trusted environment. In *OSDI*, 2002.

[8] R. Alfieri, R. Cecchini, V. Ciaschini, L. dellAgnello and?A. Frohner, A. Gianoli, K. Lorentey, and F. Spataro. VOMS, an Authorization System for Virtual Organizations. In *First European Across Grids Conference*, February 2003.

[9] European Health Management Association. Legally eHealth - Deliverable 2. http://www.ehma.org/_fileupload/Downloads/Legally_eHealth-Del_02-Data_Protection-v08(revised_after_submission).pdf, January 2006. Processing Medical data: data protection, confidentiallity and security.

[10] D. Elliot Bell and Leonard J. LaPadula. Secure computer systems: A mathematical model, volume ii. *Journal of Computer Security*, 4(2/3):229–263, 1996.

[11] JRA1 Data Management Cluster. EGEE: FiReMAN Catalog User Guide. https://edms.cern.ch/document/570780, 2005.

[12] B. Hayes-Roth et al. Guardian: A prototype intelligent agent for intensive care monitoring. *Artificial Intelligence in Medicine*, 4:165–185, 1992.

[13] B.M. Dawant et al. Knowledge-based systems for intelligent patient monitoring and management in critical care environments. In Joseph D. Bronzino, editor, *Biomedical Engineering Handbook*. CRC Press Ltd, 2000.

[14] DL Sackett et al. *Evidence-Based Medicine: How to Practice and Teach EBM*. Churchill Livingstone, 2nd edition, 2000.

[15] Open Grid Forum. GridFTP: Protocol Extensions to FTP for the Grid. http://www.ggf.org/documents/GWD-R/GFD-R.020.pdf, April 2003.

[16] B. Fraser. Site Security Handbook. RFC 2196 (Informational), 1997.

[17] K. Gjermundrod, M. Dikaiakos, D. Zeinalipour-Yazti, G. Panayi, and Th. Kyprianou. Icgrid: Enabling intensive care medical research on the egee grid. In *From Genes to Personalized HealthCare: Grid Solutons for the Life Sciences. Proceedings of HealthGrid 2007*, pages 248–257. IOS Press, 2007.

[18] JRA1. EGEE gLite User's Guide. https://edms.cern.ch/document/570643/, March 2005.

[19] John Kubiatowicz, David Bindel, Yan Chen, Steven E. Czerwinski, Patrick R. Eaton, Dennis Geels, Ramakrishna Gummadi, Sean C. Rhea, Hakim Weatherspoon, Westley Weimer, Chris Wells, and Ben Y. Zhao. Oceanstore: An architecture for global-scale persistent storage. In *ASPLOS*, pages 190–201, 2000.

[20] J. Luna et al. An analysis of security services in grid storage systems. In *CoreGRID Workshop on Grid Middleware 2007*, June 2007.

[21] Jesus Luna, Michail Flouris, Manolis Marazakis, Angelos Bilas, Marios Dikaiakos, Harald Gjermundrod, and Theodoros Kyprianou. A data-centric security analysis of icgrid. In *Proceedings of the CoreGRID Integrated Research in Grid Computing*, pages 165–176, 2008.

[22] 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 Zoltan Farkas. A secure grid medical data manager interfaced to the glite middleware. *J. Grid Comput.*, 6(1):45–59, 2008.

[23] Federal Ministry of Health. The Electronic Health Card. http://www.die-gesundheitskarte.de/download/dokumente/broschuere_elektronische_gesundheitskarte_engl.pdf, Octuber 2006. Public Relations Section. Berlin, Germany.

[24] European Parliament. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal of the European Communities of 23 November 1995 No L. 281 p. 31., Octuber 1995.

[25] T. Perelmutov et al. SRM Interface Specification v2.2. Technical Report, FNAL, USA, 2002.

[26] Michael O. Rabin. Efficient dispersal of information for security, load balancing, and fault tolerance. *J. ACM*, 36(2):335–348, 1989.

[27] Erik Riedel, Mahesh Kallahalla, and Ram Swaminathan. A framework for evaluating storage system security. In Darrell D. E. Long, editor, *FAST*, pages 15–30. USENIX, 2002.

[28] N. Santos and B. Koblitz. Distributed Metadata with the AMGA Metadata Catalog. In *Workshop on Next-Generation Distributed Data Management HPDC-15*, June 2006.

[29] Richard O. Sinnott, Micha Bayer, A. J. Stell, and Jos Koetsier. Grid infrastructures for secure access to and use of bioinformatics data: Experiences from the bridges project. In *ARES*, pages 950–957, 2006.

[30] Mark W. Storer, Kevin M. Greenan, Ethan L. Miller, and Kaladhar Voruganti. Secure, archival storage with pot-shards. In *FAST'07: Proceedings of the 5th conference on USENIX Conference on File and Storage Technologies*, pages 11–11, Berkeley, CA, USA, 2007. USENIX Association.

[31] Von Welch. Globus toolkit version 4 grid security infrastructure: A standards perspective. http://www.globus.org/toolkit/docs/4.0/security/GT4-GSI-Overview.pdf, 2005. The Globus Security Team.