

# Scientific Report

**Period: March – November 2007**

## ***I. Fellow information***

---

**Name:** Jesús Luna García

**Research Institute:** Foundation for Research and Technology – Hellas.

## ***II. Research Summary***

---

During this period I have been doing a research stay at FORTH ICS (Greece), where the following topics have been investigated:

1. Extended framework for the security analysis of Grid Storage Services.

Storage systems have become an essential piece for the data Grid, thus making it imperative to establish an integral and standardized security solution able to avoid common attacks on the data and metadata being managed. Grid security research has focused in specific high-level services (for example GSI and VOMS) instead of providing a systemic view able to encompass even block-level security features. Work-groups like the Open Grid Forum's OGSA Data Architecture and CoreGRID's Trust and Security have begun to investigate the challenges related with providing security at the Grid Storage Level.

In this research we applied to a typical use case for the Data Grid (designed by encompassing data creation, replication, discovery and retrieval) an extended framework to analyze from a systemic point of view the security guarantees provided by both, underlying infrastructure technologies and storage technologies commonly used in Data Grid sites, focusing in typical attacks that can be mounted on the data and meta-data.

By using the extended framework we were able to identify the security gaps and even redundant security features that may affect the proposed Data Grid scenario. The identified elements were used as requirements for the security mechanism being proposed by our research group in the context of this fellowship (see point 2 below).

Also it is important to mention that the extended framework is so flexible that it can be applied to other scenarios related with the Data Grid, in particular we have utilized it for the Desktop Data Grid (see point 3).

2. Use of fragmentation and encryption to enhance the security of data at-rest in untrusted storage sites.

From our security analysis (see point 1) we have concluded that a mechanism offering data protection for the Grid, should provide and enforce at least the following security policies:

- Confidentiality: the Storage Elements should not have access neither to clear-text data or encryption material, even if an attacker is able to compromise some of them. This assumption is based on the vulnerabilities identified with *untrusted storage sites*.
- Integrity: the proposed mechanisms should enable relying parties (Grid Users) to detect unauthorized modifications to the data being retrieved from the Data Grid. Measures versus attackers destroying data on the wire should be also considered.
- Availability: untrusted storage sites give no guarantees about their availability. Destruction of data at rest is also feasible to occur if we consider that adversaries may be in full control of the storage element.
- Performance: despite it is not directly related with security, there must be a clear balance between both aspects so all the involved parties do not notice degradation on their available resources beyond the storage space they are providing to the Data Grid.

The above requirements have been used to design a cryptographic protocol that achieves its security goals based in two widely used mechanisms. The first is symmetric cryptography, which not only provides high-performance encryption (contrary to using public key cryptosystems), but also a self-contained integrity checking (we use the file's hash as the encryption key) and protection versus replay attacks (use of a nonce).

A second mechanism, the Information Dispersal Algorithm, provides high availability and assurance for the Data Grid by means of data fragmentation. In a fragmentation scheme, a file  $f$

is split into  $n$  fragments, all of these are signed and distributed to  $n$  remote servers, one fragment per server. The user then can reconstruct  $f$  by accessing  $m$  fragments ( $m < n$ ) arbitrarily chosen. Our analytical model have shown the data assurance provided by both mechanisms, even though we are still designing the final version of the protocol (see point 6 below).

3. Basic notions of “Quality of Security” for Grid Storage Services.

In our research we proposed a cryptographic protocol that uses encryption and data fragmentation (see point 2) to secure a typical Data Grid. Despite its advantages, state of the art security protocols still consider that all of the Storage Elements are *homogeneous* in the sense that they have the same probability of being compromised (which means they have the same vulnerabilities, hardware and even software!). Obviously this is not true in real Data Grid, because the storage nodes will most likely have different hardware and/or software features, management policies, etc.

Our hypothesis then is that if subsets of Storage Elements with analogous features are clearly identified and their security level quantified, then any user of the Data Grid may be able to request storage space only into those nodes fulfilling some minimum guarantees or *Quality of Security (QoSec)* level. Moreover, we have to notice that the evaluation and further use of this QoSec into the proposed protocol would enhance the data assurance already provided by the encryption and fragmentation mechanisms, but without affecting the overall performance.

The technique being proposed for the QoSec evaluation is the so-called Reference Evaluation Methodology, originally proposed by the “University of Naples”. Let us mention that from previous work we are quite familiar with this technique.

4. Security protocol for the Desktop Data Grid.

The Desktop Grid is a specific type of distributed system, where shared resources (processor or storage) are provided in a volunteer fashion by the participants. These environments potentially provide commodity resources not only for CPU-intensive tasks, but also for applications that require significant amounts of memory, disk space and network throughput.

However the potential computing power of volunteer systems is much bigger: the number of Internet connected PCs is projected to reach 1 billion by 2015, which means many PetaFLOPS of computing power and a storage capacity (around one Exabyte!) able to exceed the one provided by any centralized system.

Therefore it is not surprising that nowadays a lot of efforts are being done in the direction of using Volunteer Computing as a new paradigm for both, the Computational and the Data Grid. Take for example the Lattice Project, which is expanding the reach of Grid computing by creating a system that combines the Globus Toolkit and BOINC (a widely used Desktop Grid platform).

Along with the ongoing deployment of Volunteer Computing into production environments typically suited for the Data Grid (this particular application field will be referred to as the *Desktop Data Grid*), the security aspect arises as a critical parameter.

Storage systems have become an essential piece for the Desktop Data Grid, thus making it imperative to establish an integral and standardized security solution able to avoid common attacks on the data and metadata being managed.

As a preliminary step in designing such a security mechanism for a generic Desktop Data Grid, we applied our analysis framework (see point 1) and identified its particular security issues and challenges. The second part of this research used these requirements to propose a novel cryptographic protocol able to protect the data on the wire and at rest (into the Volunteer nodes), applying the three basic mechanisms reviewed in point 2 and 3: cryptography, fragmentation and QoSec.

As future research on this topic, we are planning to implement the proposed protocol above the BOINC infrastructure.

5. State of the art on security for Desktop Grids.

Quite related with the research from point 4 above, was the state of the art survey performed on Desktop Data Grid’s security. This study allowed us to identify interesting proposals, take for example the *Storage@home* which builds a Desktop Data Grid with a security mechanism analogous to the one proposed by our research.

Also worth to mention is security in widely used volunteer computing systems like BOINC, which cope with problems like the ones identified by our analysis using probabilistic techniques to detect result falsification (data change attack). However contrary to our proposal, no protection is provided against theft of project files (both input and output files are not encrypted) under the assumption that anyway data resides in cleartext in memory, where it is easy to access with a

debugger. For a Desktop Data Grid this is unacceptable, because only the authenticated and authorized Grid user should have access to this information.

Finally we also found several distributed storage systems (even though not directly implementing Desktop Data Grids) that apply security mechanisms somewhat related with our research. From these systems it is possible to highlight POTSHARDS, Farsite, OceanStore and Cleversafe.

6. In coordination with the University of Cyprus, security and privacy issues with ICGrid.

This research represents a joint effort between FORTH and UCY to propose a security mechanism for ICGrid's data and metadata.

As a first step we applied our analysis framework (see point 1) and concluded the need for protecting ICGrid's data and metadata from attacks related with compromised EGEE Storage Sites, but with a particular emphasis on the patient's personal data. But, which are the legal reasons supporting our decision of providing this information with a highly secure mechanism? In the European Union, several Directives of the European Parliament and of the Council protect the processing and free movement of personal data, including for purposes of health care. The EU Directive on Data Protection provides a general framework for the protection of privacy with respect to the processing of personal data in its widest sense.

This (ongoing) research will try not only to solve the privacy issues related with ICGrid, but also to propose a mechanism for protecting a patient's personal data via encryption and fragmentation (applying the protocol described in point 2).

### **III. Publications**

---

- "An analysis of Security Services in Grid Storage Systems". Luna J., et. al. In CoreGRID Workshop on Grid Middleware 2007. June, 2007. Dresden, Germany. (extended version published as CoreGRID TR- 0090. September, 2007. <http://www.coregrid.net>)  
*Abstract:* With the wide-spread deployment of Data Grid installations, and rapidly increasing data volumes, storage services are becoming a critical aspect of the Grid infrastructure. Due to the distributed and shared nature of the Grid, security issues related with state of the art data storage services need to be studied thoroughly to identify potential vulnerabilities and attack vectors. In this paper, motivated by a typical use-case for Data Grid storage, we apply an extended framework for analyzing and evaluating its security from the point of view of the data and metadata, taking into consideration the security capabilities provided by both the underlying Grid infrastructure and commonly deployed Grid storage systems. For a comprehensive analysis of the latter, we identify three important elements: the *players* being involved, the underlying *trust assumptions* and the dependencies on specific *security primitives*.  
This analysis leads to the identification of a set of potential security gaps, risks, and even redundant security features found in a typical Data Grid. These results are now the starting point for our ongoing research on policies and mechanisms able to provide a fair balance between security and performance for Data Grid Storage Services.
- "Providing security to the Desktop Data Grid". Luna J., et. al. Submitted to the 2<sup>nd</sup> workshop on desktop Grids and Volunteer Computing Systems (PCGrid 2008). October, 2007.  
*Abstract:* Volunteer Computing is becoming a whole new paradigm not only for the Computational Grid, but now also for the Data Grid because of the enormous storage potential that is able to achieve at a low cost using commodity hardware. 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 into sensitive environments like eHealth and other production Grids. 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 are applied to provide confidentiality and integrity for the data at rest, while keeping the Volunteer nodes "ignorant" about the cryptographic material being used; (b) an Information Dispersal Algorithm (IDA) is implemented to improve data assurance by fragmenting the encrypted files in a *m-out-of-n* fashion; and finally this paper contributes with a novel concept, (c) a quantitative "Quality of Security" (QoSec) factor which is computed to allow storing file's fragments only in those Volunteer nodes fulfilling the Grid User's minimum requirements about data protection.

Although the focus of this work is the associated protocol, we 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.

- “A data-centric security analysis of ICGrid”. Luna J., et. al. Submitted to the CoreGRID Integration Workshop 2008. November, 2007.

*Abstract:* The Data Grid is becoming a whole new paradigm for eHealth systems because of the enormous storage potential that is able to achieve using decentralized resources managed by different organizations. These storage capabilities are quite suitable for the particular kind of system that manages Intensive Care Units' data and metadata, just like happens with the ICGrid system. However this novel approach depends on widely set of distributed storage sites, therefore requiring new security mechanisms able to avoid the leak, change and destroy of stored data in presence an external or internal attacks. Particular emphasis must be done about the patient's personal data, which protection is required by legislations all over the European Union.

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. A mechanism proposed by our joint research is also introduced to show the advantages and outcomes of their basic approaches (encryption and fragmentation) when protecting ICGrid's patient's personal data.

#### ***IV. Workshops and conferences***

---

- “CoreGRID Workshop on Grid Middleware”. Held in conjunction with ISC'07 conference. June 25-26, 2007, Dresden, Germany.
- “5th meeting of the KDM Institute”. October 1-2, 2007. Cetraro, Italy.