Contribution ID: **93** Type: **Oral contribution**

# Encrypted Data Storage in EGEE

*Thursday, 2 March 2006 16:45 (20 minutes)*

The medical community is routinely using clinical images and associated medical data for diagnosis, intervention planning and therapy follow-up. Medical imaging is producing an increasing number of digital images for which computerized archiving, processing and analysis are needed.

Grids are promising infrastructures for managing and analyzing the huge medical databases. Given the sensitive nature of medical images, practiotionners are often reluctant to use distributed systems though. Security if often implemented by isolating the imaging network from the outside world inside hospitals. Given the wide scale distribution of grid infrastructures and their multiple administrative entities, the level of security for manipulating medical data should be particularly high.

In this presentation we describe the architecture of a solution, the gLite Encrypted Data Storage (EDS), which was developed in the framework of Enabling Grids for E-sciencE (EGEE), a project of the European Commission (contract number INFSO−508833). The EDS does enforce strict access control to any medical file stored on the grid. It also provides file encryption facilities, that ensure the protection of data sent to remote storage, even from their administrator. Thus, data are not only transferred but also stored encrypted and can only be decrypted in host memory by authorized users.

## Introduction

The basic building blocks of the grid data management architecture are the Storage Elements (SE), which provide transport (e.g. gridftp), direct data access (e.g. direct file access, rfio, dcap) and administrative (Storage Resource Management, SRM) interfaces for a storage system. However the most widely adopted standard today for managing medical data in clinics is DICOM (Digital Image and COmmunication in Medicine).

The simplified goal is to secure the data movement among these blocks, and the client hosts, which actually process the data.

## Challenges

Here we describe the most important challenges and requirements of the medical community and how they are addressed by EDS on the current grid infrastructure.

Access Control

The most basic requirement is to restrict the access to any data, which is on the grid, to permitted users. Although it looks like a simple requirement, the distributed nature of the architecture and the limitations of the building blocks required some work to satisfy the requirements.

The first problem faced is the complex access patterns of the medical community. It is usually not enough to define a single user or group which is allowed to access the file, but instead access is needed by a list of users. The solution is to use Access Control Lists (ACLs), instead of basic POSIX permission bits, however most of the currently deployed Storage Elements do not provide ACLs.

To solve the semantical mismatch, we "wrapped" the existing Storage Elements into a service, which enforced the access control settings, according to the medical community's requirements. This service is called the gLite I/O server, which is installed beside every used storage element.

The gLite I/O server provides a POSIX like file access interface to remote clients, and uses the direct data access methods of the Storage Element to access the data. It authenticates the clients and enforces authorization decisions (i.e. if the client is allowed to read a file or not), so it acts like a Policy Enforcement Point in the middle of the data access.

The authorization decision is not made inside the gLite I/O server. A separate service holds the ACLs (and other file metadata) of every file stored in the Storage Elements. In our deployment it was the gLite File and Replica Management (FiReMan) service, which acts like a Policy Decision Point in the architecture.

The gLite FiReMan service is a central component, which also acts like a file catalog (directory functionality), replica manager (which file has a copy on a given SE) and file authorization server (if a given client is allowed to access a file). The gLite FiReMan service supports rich ACL semantics, which satisfy the access pattern requirements of the medical community.

Encryption

The other important requirement is privacy: the sensitive medical data shall not be stored on any permanent storage or transferred over the network unencrypted, outside the originating hospital.

The solution is to encrypt every file, when it leaves the originating hospital's DICOM server, and decrypt it only inside the authorized client applications.

For the first step we developed a specialized Storage Element, the Medical Data Manager (MDM) service, which "wraps" the hospital's DICOM server and offers interfaces, which are compatible with other grid Storage Elements. In this way the hospital's data storage will look like just another Storage Element, for which we already have grid data managements solutions.

Despite the apparent similarity between the MDM service and an ordinary Storage Element there is an important difference: the MDM service serves only encrypted files. When a file is accessed through the grid interfaces, the service generates a new encryption key, encrypts the file and registers the key in a key

store. Therefore every file which crosses the external network and is stored on stored on an external element stays encrypted during its whole lifetime.

On the client side we provided a transparent solution to decrypt the file: on top of the gLite I/O client libraries, we developed a client library, which can retrieve keys from t he key storage and decrypt files on the fly. The client side library provides a POSIX like interface, which hides the details of the remote data access, key retrieval and decryption.

The key storage had to satisfy several requirements: it has to be reliable, secure and provide fine grained access control for the keys.

To satisfy these requirements we developed the gLite Hydra KeyStore. To satisfy reliability the keys are not only stored at one place, but at least at two locations. To satisfy security, one service cannot store a full key, but only a part of it, thus even when the service is compromised the keys cannot be fully recovered. We implemented Shamir's Secret Sharing Scheme inside the client library to split and distribute the keys among at least three Hydra services, according to the above mentioned requirements.

The key storage also has to provide fine grained access control, similar to the files, on the keys. Our current solution actually applies the same ACLs as the FiReMan service, thus one can be sure that only those who can access the encryption key of a file are allowed to access the file itself.

### Conclusion

The solution for encrypted storage described above has been already released in the gLite software stack and been deployed and demonstrated to work at a number of sites.

As the underlying software stack of the grid evolves we will also adapt our solution to exploit new functionality and to simplify our additional security layer.

## Summary

The medical community is routinely using clinical images and associated medical data for diagnosis, intervention planning and therapy follow-up. Medical imaging is producing an increasing number of digital images for which computerized archiving, processing and analysis are needed.

Grids are promising infrastructures for managing and analyzing the huge medical databases. Given the sensitive nature of medical images, practiotionners are often reluctant to use distributed systems though. Security if often implemented by isolating the imaging network from the outside world inside hospitals. Given the wide scale distribution of grid infrastructures and their multiple administrative entities, the level of security for manipulating medical data should be particularly high.

In this presentation we describe the architecture of a solution, the gLite Encrypted Data Storage (EDS), which was developed in the framework of Enabling Grids for E-sciencE (EGEE), a project of the European Commission (contract number INFSO−508833). The EDS does enforce strict access control to any medical file stored on the grid. It also provides file encryption facilities, that ensure the protection of data sent to remote storage, even from their administrator. Thus, data are not only

transferred but also stored encrypted and can only be decrypted in host memory by authorized users.

**Primary author:** FROHNER, Ákos (CERN)

**Co-authors:** PERA, Christophe (CNRS, CREATIS laboratory); JOUVENOT, Daniel (LAL laboratory); MONTAGNAT, Johan (CNRS, I3S laboratory); GUIO, Patrick (University of Bergen); KUSNZT, Péter (CERN); BRITO DA ROCHA, Ricardo (CERN); FARKAS, Zoltán (MTA-Sztaki LPDS)

**Presenter:** FROHNER, Ákos (CERN)

**Session Classification:** 2b: Data access on the grid

**Track Classification:** Data access on the grid