Blanquer Espert, I.; Hernández García, V.; Segrelles Quilis, JD.; Torres Serrano, E. (2009). Enhancing Privacy and Authorization Control Scalability in the Grid through Ontologies. IEEE Transactions on Information Technology in Biomedicine. 13(1):16-24. doi:10.1109/TITB.2008.2003369.

Enhancing Privacy and Authorization Control Scalability in the Grid through Ontologies

I. Blanquer, V. Hernández, D. Segrelles, E. Torres

Abstract—**The use of data Grids for sharing relevant data has proven to be successful in many research disciplines. However, the use of these environments when personal data is involved (such in health) is reduced due to its lack of trust. There are many approaches that provide encrypted storages and key shares to prevent the access from unauthorized users. However, these approaches are additional layers which should be managed along with the authorization policies. We present in this paper a privacy enhancing technique that uses encryption and relates to the structure of the data and their organizations, providing a natural way to propagate authorization and also a framework that fits with many use cases. The article describes the architecture and processes and also shows results obtained in a medical imaging platform.**

Index Terms—Grid, Security, Ontologies, OGSA, WSRF

I. Introduction

Data security is a key requirement for biomedical Grid applications. Dealing with the different national legal regulations and procedures accepted by the medical community [1] requires a carefully approach.

One of the challenges for biomedical application is to provide efficient high-level interfaces depending on the applications that enable access to Grids for non-experts, ensuring transparent access to medical resources through services compatible with medical practice. As part of the interfaces, a flexible architecture for the management of the privacy of data is needed, compatible with medical practice and with pre-existing medical information systems.

Besides, the talks which were delivered by the authors of the "Grids: The Top Ten Questions" give us one concluding remark that describes many of the Grid production platforms today: Until security is made easier to use, it will not be used [2]. Grid security systems are complex enough to be considered an obstacle in the successful Grid adoption. The proposed architecture introduces new concepts and methods that need to be expressed in the natural terms of the application community or it will be considered a new barrier.

A Virtual Organization [3] (VO) is formed from different real entities (e.g. Medical Centers, Hospitals, Governmental Centres), and probably also from different communities (e.g. physicians and researchers working in specific projects). Access to data is normally organised around VO membership.

Medical Imaging Grid middlewares using virtual communities for sharing, transferring and processing DICOM medical images in a distributed environment [4] are starting to be adopted by the medical community. The Digital Imaging and Communications in Medicine [5] (DICOM) is the most common standard for medical images. A single DICOM file contains both a header (which stores information about the patient's name, the type of scan, image dimensions, structure report etc), as well as all of the image data (which can contain information in three dimensions) or structured reports in DICOM Structured Reporting Objects [6] (DICOM-SR). TRENCADIS, is a middleware for managing DICOM Structured Reporting Objects [6] that has been used as part of the CVIMO [7] deployment, in which five hospitals in Valencian Region collaborate to share DICOM Studies and DICOM Structured Reports. Three ontologies have been created in CVIMO which define the three oncological target areas implied (i.e. lung, liver and central nervous system). Each area can only access the parts of DICOM studies defined in the ontology that a user belongs to.

The main objective of this paper is to provide Grid middlewares such as TRENCADIS, with efficient and reliable privacy protection for sensitive data. This paper presents a model for long-term storage and management of encrypted data in distributed environments. Furthermore, the paper outlines how this model is implemented to preserve the privacy of patient information in Grid-based collaborative computational infrastructures for biomedical applications.

This paper delineates a dependable security framework in overextended organizations. Throughout the assembly of this framework, organizations will encounter different degrees of data integrity and confidentiality.

The specific objectives of the paper are:

- To propose an on-the-fly cryptographic infrastructure to protect privacy from users with administrative privileges.
- To provide a flexible architecture for organizing key management for long-term storage of encrypted data.
- To propose a model applicable in different environments, compatible with current Grid middlewares.
- To provide an access control mechanism for encryption keys based on ontological groups and roles.

[1] I. Blanquer, V. Hernández, D. Segrelles and E. Torres are with the Institute for the Applications of Advanced Information and Comunication Technologies (ITACA), Polytechnic University of Valencia (UPV). Camino de Vera S/N, 46022 Valencia, Spain (phone: +34 96 387 7007 ext.88254; Fax: +34 96 387 7274)(e-mails: iblanque@dsic.upv.es; vhernand@dsic.upv.es; dquilis@itaca.upv.es; etorres@itaca.upv.es)

The paper is organized as follows. Section II illustrates related works. Section III describes the Security Model and an insight into the security issues presented in previous papers [34][35]. Section IV shows a real deployment of the security model that has been applied in the CVIMO project. After that, results about the model deployed in a controlled environment are described. Finally conclusions are presented.

II. Related Works

Computational Grids offer a number of benefits and opportunities to biomedicine, healthcare and other biomedical domain areas [8]. Several recent systems, focused on new Health-related applications are analysed.

The Medical Data Manager (MDM) [9] is a data management service designed to handle medical images on Grids, strongly based to the gLite middleware. The MDM aims at guaranteeing patient's privacy by keeping private data in acquisition centres. However, this approach comes along with higher complexity in the specification and maintenance of the access policies. Granting full access right to information objects (both image data and header attributes from a DICOM file), requires achieving a number of capabilities kept by different services in the form of access control lists (ACL). This approach has deficiencies in systems where the potential users will not be known beforehand. The higher flexibility of attribute-based approaches enables the model presented in this paper to deal efficiently with these requirements.

The EncFile [10] is an encrypted file management system for biomedical applications in the EGEE [11] project. Although EncFile is not linked to the EGEE Grid components, the system has been implemented over LCG2 [12].

A Grid-based architecture for computer aided diagnosis, was presented in [13]. In order to protect information against unauthorized disclosure, the authors propose an encrypted storage component described in [14]. Although the prototype was validated on a large experimental platform, the architecture has not been tested in real environments.

The Secure Storage Service provides a set of tools to manage confidential information in an encrypted format in a Grid Computing environment [15]. This service has been developed for the gLite [16] middleware. The Secure Storage Service aims to solve the insider abuse problem preventing also the administrators of the storage elements to access the confidential data in a clear format; however it does not specify a means to protect the decryption keys from being accessed by administrators. Moreover, the Secure Storage Service associates an ACL with the decryption key. This ACL contains all users authorized to access the encrypted file. This approach does not scale well as the number of users increases.

Identifying data resources is a fundamental problem within large-scale Grid environments. While traditional solutions enable users from one organization to access data belonging to other organizations by sharing metadata, this may not be acceptable for certain organizations due to privacy concerns.

The MDM client library provides APIs for requesting files based on the metadata attached to the DICOM image. The metadata is internally extracted from the DICOM headers and placed into specialized catalogues.

The role of ontologies [17] in the context of Grid computing for obtaining, comparing and analyzing data is increasing. Ontologies can be used to localize data sets within collaborative environments, and to build on the fly collections of data files based on attributes of the ontology.

Our proposal uses ontologies that define the information which is interesting for a given area or group [4]. In CVIMO, ontology attributes match DICOM fields (headers or DICOM-SR tags) and can be used for filtering, indexing and searching DICOM objects in virtual collections.

There are number of efforts to produce access control languages and standards based on XML (e.g. XACML [18]) and authorization assertion protocols (e.g. SAML [19]). While SAML provides a mechanism for making authentication and authorization assertions and a mechanism for conveying them, XACML provides the language that defines the rules needed to make the necessary authorization decisions.

XACML has been applied with great success [20] for implementations of the Attribute-Based Access Control (ABAC) model. In ABAC, access decisions are based on attributes of the requestor and resource, and users need not be known by the resource before sending a request. ABAC is scalable and flexible and thus is more suitable for distributed, open systems, than identity-based access control models [21].

Finally, there are promising results on applying Semantic Web standards for protecting Grid [22] and web services [23].

III. Security Model

A. Grid Architecture

Most of the current Grid middlewares are based on Web Services protocols. The Open Grid Services Architecture [24] (OGSA) is a specification in progress that aims at defining a standard and open architecture for Grid-based applications.

The Globus Toolkit is a realization of OGSA, which can be used to develop Grid applications. Globus Toolkit Version 4 (GT4) provides services implemented on top of the Web Service Resource Framework [25] (WSRF), a specification that extends Web Services with stateful services and other features. The services of the architecture presented in this paper are all based in OGSA/WSRF.

B. Grid Security Infrastructure

The security services of Grids are not altogether different from those of other distributed system paradigms. Specifically, an effective security model must ensure a set of security primitives: identity verification, authorization, access control, data integrity, data confidentiality and availability.

Modern Grid middlewares provide with the security infrastructure, usually by means of the Globus Security Infrastructure [26] (GSI), which is a set of tools, libraries and protocols used in Globus to securely access resources. Almost all Grid components and Grid middlewares use the GSI for

authentication. GSI also provides mechanisms that deal with secure connections as well as message protection.

GSI lacks from guaranteeing the reliability of the information stored, in terms of authenticity and confidentiality.

On the other hand, in computational Grids, authorization has an importance beyond its common security meaning. Proper Grid authorization eases the administration of the shared resources and provides coherence to the system by consistently preserving the relationships of the participants.

Grid authorization is closely related to the Virtual Organization (VO) concept. The VO administrators define hierarchy relationships (e.g. groups, subgroups) in the VO, different privileges (e.g. roles, capabilities) to resources, and define membership in the groups. They are also in charge of controlling access to the Resources Providers (RPs) (e.g. services in an OGSA approach) on the basis of users' credentials (e.g. groups, roles, capabilities) and the agreements established between the VO participants. In addition, the RPs have their own local security policies that may override the VO policies. Last decisions on the access to resources must be on the side of the owner of the resource, but global policies enable the management of large-scale infrastructures.

In conclusion, the access control to the RPs in the collaborative Grid infrastructure is based on the membership in the group. The actions that users in a given group are allowed to perform (from the point of view of the VO) on a specific resource instance are determined by two policies: the rules that describe the group, and the rules controlling the access to resources. All this is managed in the RPs by a component named GateKeeper, which takes into account resource-specific policies, normally ACLs.

The classic approach of ACLs requires that permissions are explicitly given to individuals or groups. If a piece of information should be made available to different VOs or VO groups (but not all of them), the data owner should explicitly indicate it when sharing the data. This could be complex if many data are created regularly. However, the metadata associated to a piece of information can have enough information to decide which groups should access it.

There are several attribute-based access control systems for Grid environments in the literature (i.e. Akenti [27] , PERMIS [28], Shibboleth [29], and VOMS [30]). Group-based authorisation tools (such as VOMS) enable granting different roles and permissions for a single user. VOMS manages authorization information about the members of virtual organizations, and supplies this information as a X.509 attribute certificate. In the context of the EGEE [11] Grid infrastructure, roles are assigned to users through VOMS.

As VOMS makes use of X.509 attribute certificates to assert user's group memberships, roles and capabilities, users must create a X.509 proxy certificate [31] before accessing the resources. A VOMS server generates the attribute extensions.

C. VO Management and Ontologies

The concept of ontology, as "*the branch of metaphysics that deals with the nature of being*" has been used in many areas of science and literature. In information technologies, an ontology is a vocabulary and a set of terms, rules and relations that define with the needed accuracy a set of entities enabling the definition of classes, hierarchies and other relations among them. The ontologies define the terms to be used to describe and represent a knowledge domain. In this sense, the ontologies organise the knowledge in a reusable way.

An Ontology Server is a service provided by the model that defines the ontologies (in any language: XML, RDF, OWL, etc.) and specifies the relations between VO groups and ontologies. The Ontology Server stores a unique identifier for the ontologies in the context of the VO (namely Ontology Id).

In conclusion, the VOMS server organizes the users into groups, and the Ontology Server organizes the access of groups to ontologically classified resources and data. Each group can manage multiple ontologies, and each ontology can be managed by different groups.

D. Information Object Storage

The Information Object Storage (IOS) is a repository service provided by the model. This repository stores all the encrypted information objects required by the VO, despite of the ontological classifications these object can have. Furthermore, the IOS keeps the relationships between the objects and the ontologies through the Encrypted Object Unique Identifier (EOUID) which uniquely identifies the object in the Grid. In parallel, the ontologies are used for filtering, indexing and searching encrypted objects in virtual collections. These virtual collections are also kept in the IOS.

When a user tries to retrieve an information object, the Gatekeeper at the IOS verifies, first that the user's credentials identify the user as a member of a VO group, secondly that this group is authorized on the object's ontologies (combining the ontology information stored in the Ontology Server), and finally that the local rules allow the user access to the resource.

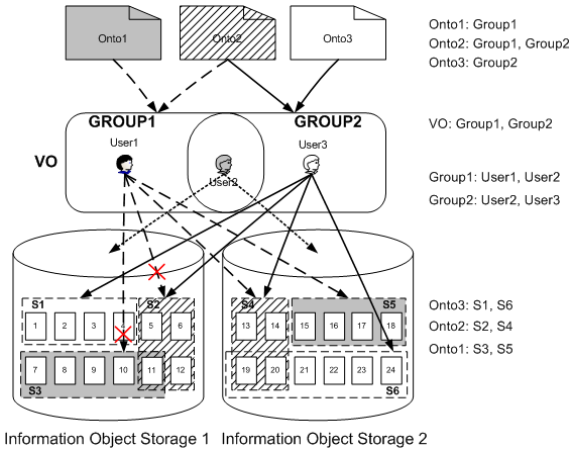

Fig. 1. Access control relation among Ontologies - VO groups – Information Objects Storage.

Figure 1 shows 3 ontologies that classify the objects into three subsets (Onto1, Onto2, and Onto3). Users are organized into two groups (Group1, Group2), and there is one user in

both groups. Group1 is authorized to access data from ontology 1 and 2 (what means accessing objects s2, s3, s4 and s5). Group 2 can access data from ontology 2 and 3 (what means accessing objects s1, s2, s4 and s6). User 2 will be authorized to access data from Onto1 and Onto2, or Onto2 and Onto3, depending on the credentials exposed.

Moreover, an IOS might allow or deny the access from users of specific group (IOS 1 and Group 1 in the figure 1). For this reason, User1, even able to manage objects from ontology 1, cannot access s2 and s3 data in IOS 1.

The key of the authorisation mechanism is that ontologies and VO groups can only be created by the system administrator, which needs the agreement from the deputies of the communities to include an ontology from one group in a different group. Finally, individual users or VO groups can be banned even presenting the right ontologies through a configuration file of the gatekeeper. This is a critical operation and should be performed at the resource administration level.

E. Encryption and Decryption of Data

The model requires a symmetric cryptographic key to encrypt and decrypt the information object. 256-bit keys AES (Advanced Encryption Standard) [32] are used. Submitters of new/updated datasets utilize separate keys for each object. Encryption and decryption operations took place on the client, preventing overload of application servers running the IOS.

Given that the risk of attacks is higher in servers which share multiple services (including public ones) and the impact could be higher since server keep far more data than clients, keeping unencrypted information out of the IOS not only improves performance, but also helps to protect the information from unauthorized disclosure.

F. Data Integrity and Confidentiality

Dependable data storage and sharing among multiple organizations are important features of the proposed model. The security framework guards data integrity and confidentiality, while ensuring that information objects are easily accessible for authorized users.

An integrity code protects both object's integrity as well as its authenticity by allowing users to detect any changes to the object content. We implement this functionality through a 160-bit RIPEMD message digest algorithm. The AES-encrypted blocks of data are used as input for the digest function, joining the encryption/decryption and validation in a single step.

The encrypted objects are stored in the IOS, while redundant copies of the integrity code are kept in secure storages, ensuring that authorized users can compare the integrity code with the digest of the encrypted object.

A message integrity code provides integrity. Additional measures for authenticity are explained in next section, as well as the reason for not encrypting the integrity code.

On the other hand, guaranteeing confidentiality of sensitive data outside the organization's borders additionally requires implementing a decryption key management scheme.

In our model, the management of decryption keys is performed through a secret sharing scheme. The key distribution is achieved by a client that divides the key in N different shares using the Shamir's secret sharing scheme [33]. Key shares are distributed among different administrative domains that contribute with the responsibility of protecting data from unauthorized disclosure. Only k shares $(k < N)$ are needed to reconstruct a key. Key shares are pairs of data that relate to the input and output of a polynomial of degree N. A sharing pair is represented as (*IDKeyPart*, *Key*), where *IDKeyPart* and *Key* are the input and the output (to the polynomial), respectively. Key shares for the same decryption key must be placed at different Key Servers.

The Key Server is a repository service provided by the model. Two Key Servers are different if and only if they are located at different administrative domains. This means that they are managed by different administrators, even if they are sharing the same VO. It also means that any user who has granted access to a share in a given administrative domain cannot reconstruct the decryption key without obtaining permission on other $k-1$ administrative domains.

G. Distribution of the Key Shares and the Integrity Code

Distribution of key shares is one of the novel contributions of this paper. By taking the participants of the secret sharing scheme in different administrative domains, the information is protected from being exposed by users granted with physical or administrator access, ensuring the confidentiality of the encrypted objects in the Grid.

Each administrative domain need to be enclosed within the boundaries of one organization. The organization registered as a private data holder or as a private data processor, must carry on with a set of legal responsibilities concerning keeping private data secure from unauthorized access or disclosure.

Key Servers not only store key sharing pairs, but also kept a copy of the integrity code of the encrypted object. The distribution of the integrity code among real administrative domains ensures the integrity of the objects in the Grid, and does not necessarily need to encrypt the integrity code to provide a reliable level of assurance. In this way, it becomes possible to validate the object integrity by comparing the integrity code with its representation in the Key Servers. Unauthorized attempts to modify any encrypted object on the Grid will require compromising the security of a group of services deployed by different administrative domains.

Storing the integrity codes in the Key Servers also serves the purpose of providing the model with a reliable permission revocation mechanism. When a user is revoked from a given VO group, he or she will not be able to access the objects using the VO credentials. The problem arises when the user kept a decryption key after permission revocation and he or she could use local administrative privileges to access the data in the storage elements. In these scenarios, the authenticity and the integrity of the objects is ensured by cross validating the copy of the integrity code within the encrypted object with the copies stored in the Key Servers available at different administrative domains. The complexity of compromising the

integrity of an object is the problem of compromising at least *k* Key Servers located in *k* different administrative domains.

Useful insight into the permission revocation issue was presented in a previous paper [34].

H. Administrative Domains in the VO

The different administrative domains that kept shares of the same decryption key may be part of the same VO. The VO context is the perfect scope for the integration of independent organizations in data protection schemes. VOs are usually associated with a project related to a community where information objects are shared. The VO should agree on the values of *k* and *N*, as well as the number of Key Server replicas that each party should contribute to guarantee operation. Previous works [33] have probed that a very robust key management scheme can be reached by using $N = 2k - 1$.

As the participants of the secret sharing scheme are in different administrative domains, even the minimum value the *k* parameter can take ($k = 2$) enhances the privacy of the data, since any user (even a local administrator of a storage) needs obtaining access on *k* different administrative domains in order to reconstruct a decryption key. On the other hand, with $k = 2$, only $N = 3$ different administrative domains are needed. In the context of the VO, the administrative domains could be defined as the individual organizations which control private information, and contribute with their private data to the VO.

In our model, each administrative domain is revealed by an X.509 organizational unit attribute, along with the common name attribute of the certificate authority.

I. Publication in Monitoring and Information System (MIS)

The Monitoring and Information System (MIS) is a significant piece of Grid technology. This component could be implemented in many different ways (e.g. GMA, MDS2, MDS4) depending on the middleware used (gLite, GT2 and GT4), while the objective is the same: to collect and to deliver information about Grid resources where and when needed.

MIS simplifies the key shares distribution process among parties involved in the secret sharing scheme. Administrative domains integrated in VOs issue information about their key servers to the MIS, and the clients query the MIS for available key servers in trusted and different administrative domains.

The identification of a Key Server and its administrative domain requires The Key Server's URI, the local Key Server identifier (*IDKS*) relative to the administrative domain, and the identifier of the administrative domain. This information is issued by the MIS and queried by the VO clients.

The figure 2 shows a schematic representation of the storage and management of encrypted data in the Grid. The top of the figure shows a user interacting with the MIS of the VO. The user queries the MIS for three Key Servers in three different administrative domains. At the base of the figure 3, independent organizations affiliated with the VO are represented. Each organization contributes with its own Key Servers in the decryption key sharing scheme. At the top of the figure, a CA issues security credentials to the members of the VO. The Key Servers register the organizations in the MIS index. Through the structure of DNs, the administrative domains of the key Servers are revealed. These different administrative domains are used by the encryption mechanism to ensure that different key parts are stored on different domains (as shown at the top of the figure 2). The encrypted object is generated using a new encryption key and the information of the administrative domains.

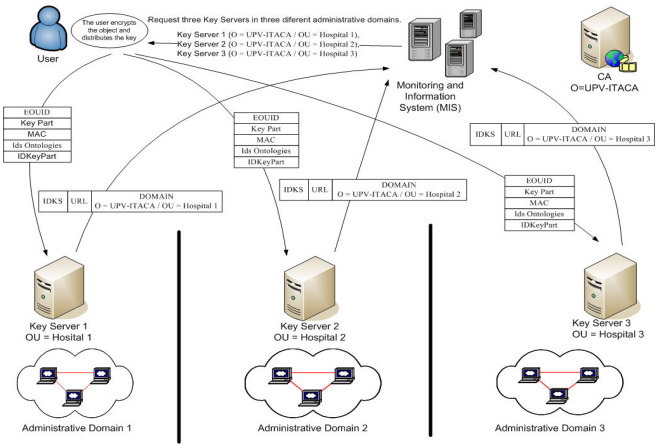

Fig. 2. Schematic representation of the storage and management of encrypted data. It shows the encryption of an information object and the distribution of the key shares among different administrative domains.

Once the decryption key shares are distributed among the parties (additional attributes issued to the Key Servers will be explained in the following sections), the encrypted object is submitted to the IOS. Since the confidentiality and integrity of the information object is protected by the framework through the encryption and the integrity code, the IOS could be deployed across the whole of VO computing environment.

J. Uniqueness of Information Objects

Whether an encrypted information object could be moved to a different IOS (or simply, whether it could be replicated), depends on how complex is to modify the information linking the object with the IOS and with the Key Servers.

The EOUID is a globally unique identifier that guarantees that the encrypted information objects can be unambiguously identified. The EOUID is assigned the first time the object is encrypted, and is based on the Universally Unique Identifier (UUID) standard to guarantee to be unique in time and space.

Referring to an encrypted object by its EOUID, Grid repositories (i.e. IOS and Key Server) guarantees that the information derived from the object is detached from the physical location of the object in the Grid.

K. Encrypted Objects's Data Format

Besides the encrypted bits, the encrypted object carries additional information. Along with the already mentioned integrity code, the encrypted object contains header fields, a body of encrypted bits and a footer field. The prime number used to divide the key is attached to the object in a header field. The rest of the header contains the *N* identifiers of the

administrative domains that keep shares of the decryption key. The footer field is reserved for the integrity code.

L. Access Control with Ontology Attributes

The basic idea of access control with ontology attributes is not to define permissions directly between users and resources, but instead to use the resources' ontology attributes as the basis for authorization. Access control policies grant groups of users with different privileges to ontologically classified resources. All services in the framework must enforce these policies on users, and thereby they must know what services in the Grid store the authorization statements that Policy Decision Points (PDP) will use with the attributes available about the requester and the resource to evaluate authorization.

The previous sections of this paper discussed where policies and other authorization attributes are stored in the framework: the VOMS Server is the repository where VO groups and roles are created and maintained, the Ontology Server stores the different authorization statements that define the relations between VO groups and ontologies, and the IOS defines the ontological classification of the information objects.

As we seen before, the IOS could be deployed anywhere in the Grid. Therefore, an IOS outside the administrative domain is not a trusted source of ontology attributes for Key Servers. On the contrary, when the Key Server itself is a source of ontology attributes for its administrative domain, changing a encrypted object's ontology in the IOS does not affect the security of the object. Keeping a list of ontologies for the object, the Key Server guarantees the security of the key, thus guaranteeing the security of the encrypted object.

Besides the decryption key share, the *IDKeyPart*, the integrity code, and the *EOUID*, the object owner stores a list of ontology identifiers (*Ids. Ontologies*) for the encrypted object in the Key Server. Hereby, authorization to key shares is provided to predefined ontologies that are related to the encrypted object. In this way, ontology identifiers updates must be synchronized among Key Servers. Hence, this model works better for applications where ontological classification of encrypted objects varies little over time.

M. Rebuilding Keys and Decrypting information

When a Grid user wants to retrieve an encrypted object identified by its EOUID, the user is first authenticated, and then the IOS collects the attributes from the user's proxy (figure 3, Step 1). It then consults the Ontology Server to find out if the user belongs to any of the VO groups allowed to access the ontologies related to the object. If authorized by the IOS, the user will retrieve the encrypted object (fig. 3, Step 2).

Once the user retrieves the encrypted object, he or she extracts the administrative domain identifiers from the header of the encrypted object (fig. 3, Step 3). Then the user consults the MIS for the URIs of the Key Servers (see fig. 2), and consults k Key Servers to retrieve the key (fig. 3, Step 4).

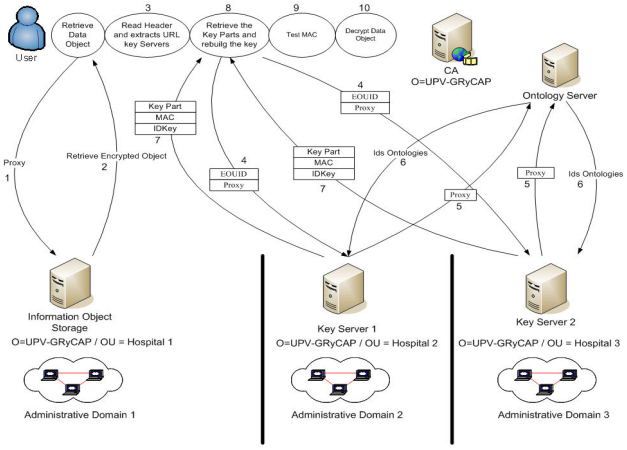

Fig. 3. Schematic representation of the reconstruction of the decryption key and the decryption and validation of the encrypted object.

The role of the different components of the model involved in the security scheme can be explained through an example in the terminology of the XACML standard: when a user requests to the Key Server to retrieve a key share, the code responsible for executing the request contains a Policy Enforcement Point (PEP) creating an access request. The access request contains the attributes that identify the user, and the encrypted object associated with the key share (ontologies), and the action being performed in the resource (retrieving a decryption key). The PEP sends this description of the attempted access to the Gatekeeper. The Gatekeeper implements a Policy Decision Point (PDP) that consults the Ontology Server for policies matching the specified group membership to the ontologies (figure 3, Step 5, 6), and also consulting local security policies (e.g. resource-specific ACLs). The PDP then evaluates the access request and issues an authorization decision, sending this conclusion to the PEP. Finally, the PEP executes the code for retrieving the key share, or throws a denying exception.

If authorized, the user will retrieve *k* different shares and *k* copies of the integrity code (fig. 3, Step 7).

With the *k* sharing pairs, the user reconstructs the decryption key, decrypts the object, and computes the integrity code (figure 3, steps 9 and 10). The user verifies the computed integrity code with the code stored within the encrypted object, and with the codes retrieved from the Key Servers.

IV. Real Implementation and Deployment

Radiological image and report data storage and distribution in clinical practice at intra-corporative level is a well-solved issue with many industrial successful stories. However, sharing data for research and training is an issue that deals with additional problems, such as knowledge organisation, privacy and processing. A representative use case targeted by the present work could be executing a perfusion analysis on all the images from patients suffering a hepatocarcinome and retrieving the flow rate coefficient images. This cannot be done in current image management systems on clinical delivery, even involving only one institution or administrative domain. Integrating multiple sources will increase the

representativeness of the study, and the integration of computing resources will enable complex post-processing.

The model presented in this paper has been implemented in the framework of the Valencian Cyberinfraestructure of Oncological Medical Images [7] (CVIMO) project. All services implemented are based on OGSA/WSRF, which constitutes the Grid architecture and infrastructure of the project. The implementation has been done using the Globus Toolkit 4, which uses MDS4 as MIS [35].

CVIMO is a project funded by the Ministry of Enterprises, University and Science of the Valencia Region. In this case, the ontologies are built from an anonymized set of attributes from DICOM headers or DICOM-SR fields. This set is controlled by the VO at a central level, and under the approval of the management of the system, so no privacy leakages appear. Relevant cases are organised into three communities related with oncology (lung, liver and central nervous system). CVIMO does not compete with intra-hospital System such as PACS or RIS/HIS systems, which are oriented to clinical daily practice, but complements them with a collaborative tool to store and share cases relevant for their research or training.

A VO named CVIMO and three VO groups have been created using a VOMS Server, one for each oncological community implied. The studies relevant for each group can be defined through the ontology using a part of their information. These ontologies are defined in XML. When a user performs a query operation, he or she can only access the information of DICOM studies or DICOM structured reports specified by the ontology that his or her groups have associated.

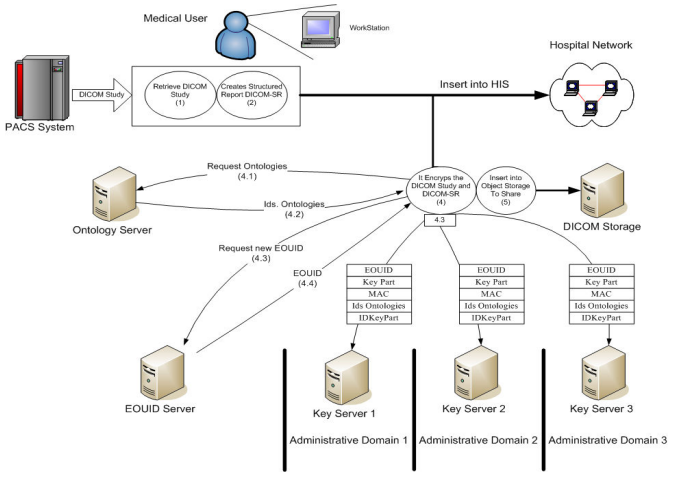

Fig. 4. Path from the creation a relevant DICOM-SR to the share it in the Grid.

When a medical user selects a relevant DICOM Study for sharing (point 1, fig. 4), first creates the structured report DICOM-SR with a given diagnostic (point 2, fig. 4). DICOM Study and DICOM-SR are sent to the IOS to share the information only with the users of the community related to the study. If this happened, the object is encrypted (point 4, fig. 4) and then inserted into the IOS. The encryption operation implies consulting the ontologies' identifiers that the user can manage (points 4.1 and 4.2, fig. 4), to generate an EOUID for the encrypted object (points 4.3 and 4.4, fig. 4) and to create and distribute the encryption key (point 4.3, fig. 4).

The implemented services in this system are the following:

- **Ontology Server.** Keep the ontologies and the relations between VO groups and ontologies.
- **Key Server.** Keep for each key part the associated information (MIC – Message Integrity Code, EOUID and IDs of ontologies, IDKeyPart and the key part).
- **EOUID Server.** This service generates the EOUIDs required to identify the encrypted objects.
- **DICOM Storage.** This service storage the DICOM studies and DICOM-SR encrypted.

V. Results and Discussion

A sample dataset from radiology studies has been created. Each file in the dataset consists of radiology image accompanied by relevant (anonymous) clinical data. Four different studies with different file sizes (see Table I), were used to create the sample set.

TABLE I. SAMPLE DATASET LOCATED IN AN INFORMATION OBJECT STORAGE IN THE VO.

Id.	Study Id.	Image Id.	Image Size (MB)
A	1.2.840.10008.5.1.4.1.1.4	1	0.5
B	1.2.840.10008.5.1.4.1.1.7	1	2.5
C	1.2.840.10008.5.1.4.1.1.1	1	5.8
D	1.2.840.10008.5.1.4.1.1.3	1	7.7

Images were firstly encrypted and stored in an IOS. Unencrypted images are also stored in the IOS, to measure the differences between encrypted and unencrypted objects. The length of registering and retrieving an object in the IOS was measured in a client of the infrastructure. Figure 5 shows the execution time for the set of studies.

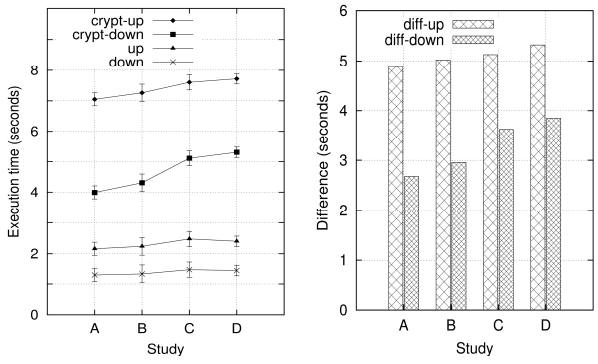

Fig. 5. Graph of the execution time for a set of studies (left) and (right) registering and retrieving objects with and without encryption.

In figure 5 (left), four different series of experimental values are represented. On one hand, "crypt-up" and "crypt-down" show the time used for registering and retrieving objects, including encrypting and decrypting the object and key sharing. On the other hand, "up" and "down" show the time used for registering and retrieving objects without encryption.

Each point in the graph represents the average time measured by the client clock. The error bars shown in the graph show the standard error calculated for each data point.

Figure 5 (right) shows the difference *D*, calculated as:

D = time with encryption - time without encryption

The importance of this graph is to show that even when the difference with and without encryption tends to be greater for large-sized objects, it is possible to estimate a performance level in a given interval. For example, the results of this study show that for those objects in the interval from 0.5 to 7.7 MB, it is possible to anticipate up to 4 additional seconds for retrieving an object using encryption, if compared with the same process without using encryption. This is consistent with the initial studies demonstrating that the overhead due to the security model can be accepted even in an interactive use.

Grouping several key shares for different objects in a single request is also possible, so the overhead for retrieving the decryption keys could be optimized when several objects are used in the same study. This is a common need in medical research and training, the main two objectives of the system.

VI. Conclusions

Healthgrids require supporting the flow of information across hospital network boundaries. Encrypted storage is needed to ensure data privacy on different administrative domains. Sharing encrypted objects requires an infrastructure to manage, protect and control access to the encryption keys.

However, decryption keys have a lifecycle, whose management is proposed in this paper by ontology-organized key management for long-term storage.

The novelty of the approach is to bind automatically the authorisation of users to the actual data automatically through the use of ontologies that specify the data accessible and the relation of VO groups and those ontologies, instead of using the classical ACL approach. Other novelty is in the definition of a distributed security enforcement scheme that takes advantage of the ontologies for distributing and managing the encryption keys in a secure manner. DICOM fields (headers or DICOM-SR tags) used to build the ontologies are previously anonymized, guaranteeing that almost all fields can be used, and resulting in a comprehensive set of ontologies.

The information-centric approach of securing the data combined with protecting and controlling the access to the decryption keys presented in this paper, have proven to be effective in the prevention of incidents of exposed data due to inconsistent encryption and key management policies, in the prevention of incidents of inaccessible data due to mismanagement of decryption keys, and in helping communities to increase the consistency of encryption and key management policies across organization boundaries.

In addition, this work contributes to increasing the clarity of responsibilities and also contributes to the creation of encryption and key management policies and practices.

Overhead due to encryption and decryption is not significant with respect to data transfer overhead, and those processes are performed on the client-side to improve scalability. The ontologies are connected to objects both through the IOS and the Key Server. Duplicating this layer of access control could penalize performance when propagating changes in the ontologies, but deliver higher scalability when the ontologies association do not change in time often. Ontology updates are performed in a lazy revocation. When the ontologies change, a new object with a new EOUID is created, reducing the need for massive re-encryption. This update management could be inefficient for objects frequently changing their ontological classification, medical imaging Grids normally deal with read-only and persistent data which minimises this issue.

Acknowledgment

The authors wish to thank the financial support received from The Spanish Ministry of Education and Science to develop the project “ngGrid - New Generation Components for the Efficient Exploitation of eScience Infrastructures”, with reference TIN2006-12860. This work has been partially supported by the Structural Funds of the European Regional Development Fund (ERDF).

References

[1] I.E. Magnin, J. Montagnat. "The Grid and the Biomedical Community: Achievements and Open Issues" EGEE User Forum, CERN, Geneva, Switzerland, March 1-3, 2006.

[2] J.M. Schopf. “Grids: The Top Ten Questions”. Scientific Programming, 10(2), 2002.

[3] Ł. Skitał, R. Słota, D. Nikolow, J. Kitowski. “Methodology for Virtual Organisation Design and Management”. EGEE User Forum, CERN, Geneva, Switzerland, March 1-3, 2006.

[4] I. Blanquer, V. Hernandez, J.D. Segrelles, “An OGSA Middleware for Managing Medical Images using Ontologies”, Journal of Clinical Monitoring and Computing, Vol. 19, pp. 295–305, October 2005.

[5] Digital Imaging and Communications in Medicine (DICOM) Part 10: Media Storage and File Format for Media Interchange. National Electrical Manufacturers Association, 1300 N. 17th Street, Rosslyn, Virginia 22209, USA.

[6] I Blanquer, V Hernandez, and D Segrelles. “TRENCADIS – a WSRF Grid MiddleWare for Managing DICOM Structured Reporting Objects”. Studies in Health Technology and Informatics. 2006;120:381-91

[7] “Ciberinfraestructura Valenciana de Imagen Médica Oncológica (CVIMO)”, http://www.grycap.upv.es/cvimo

[8] V. Breton, K. Dean and T. Solomonides (Eds.), “The Healthgrid White Paper”. From Grid to Healthgrid – Proceedings of Healthgrid 2005, IOS Press, Studies in Health Technology and Informatics, Vol. 112, 2005, pp 249-321.

[9] J. Montagnat, Á. Frohner, D. Jouvenot, C. Pera, P. Kunszt, B. Koblitz, N. Santos, C. Loomis, R. Texier, D. Lingrand, P. Guio, R. Brito Da Rocha, A. Sobreira de Almeida and Z. Farkas, “A Secure Grid Medical Data Manager Interfaced to the gLite Middleware”, to appear in Journal of Grid Computing (JGC), Springer Netherlands, 2007.

[10] C. Blanchet, R. Mollon and G. Deléage, “Building an Encrypted File System on the EGEE grid: Application to Protein Sequence Analysis”, in: Proceedings of the First International Conference on Availability, Reliability and Security (ARES’06), IEEE, April 2006.

[11] EGEE: Enabling Grids for E-sciencE (phase I and II). FP6 European IST Project, Contract Number INFSO-RI-508833: http://www.eu-egee.org

[12] “World Wide Web Computing Grid. Distributed Production Environment of Physics Data Processing “. http://lcg.web.cern.ch/LCG

[13] S. Varrette, J.L. Roch, J. Montagnat, J.M. Pierson, L. Seitz, F. Leprevost, “Safe Distributed Architecture for Image-based Computer Assisted Diagnosis”. ICPS’06, IEEE International Conference on

Pervasive Services, workshop on Health Pervasive Systems, Lyon, France, IEEE, June 2006 http://www.icpsconference.org
[14] L. Seitz, J. M. Pierson, and L. Brunie, "Encrypted Storage of Medical Data on a Grid". Methods of Information in Medicine, vol. 44, no. 2, pp. 198–201, February 2005.
[15] D. Scardaci, G. Scuderi, "A Secure Storage Service for the gLite Middleware", in: IAS 2007 - The Third International Symposium on Information Assurance and Security, Manchester, August 2007 .
[16] gLite middleware: http://www.glite.org
[17] M. Hadzic, E. Chang, "Role of the ontologies in the Context of Grid Computing and Application for the Human Disease Studies". Proceedings of International Conference on Semantics of a Networked World – Semantics for Grid Databases (ICSNW 2004), June 2004.
[18] T. Moses, (Ed.), "eXtensible Access Control Markup Language (XACML) Version 2.0", OASIS Standard, 2005.
[19] Security Assertion Markup Language (SAML), (2005), v2.0, OASIS Security Services TC, http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=security (10.02.2007)
[20] C. Schläger, T. Priebe, M. Liewald, G. Pernul, "Enabling Attribute-based Access Control in Authentication and Authorisation Infrastructures". Proceedings of the 20th Bled eConference "eMergence" (Bled'07), June 2007.
[21] B. Lang, I. Foster, F. Siebenlist, R. Ananthakrishnan, T. Freeman, "Attribute Based Access Control for Grid Computing". Preprint ANL/MCS-P1367-0806 of the Mathematics and Computer Science (MCS) Division of Argonne National Laboratory, August 2006.
[22] W. Xiaopeng, L. Junzhou, S. Aibo, and M. Teng, "Semantic Access Control in Grid Computing". Proceedings of 11th International Conference on Parallel and Distributed Systems (ICPADS 2005), pp. 661 - 667 Vol. 1, 2005.
[23] B. Shields, O. Molloy, G. Lyons and J. Duggan, "Securing Web Services using Semantic Web Technologies". Proceedings of the 1st International IFIP/WG12.5 Working Conference on Industrial Applications of Semantic Web (IASW 2005), pp. 213 – 22, 2005.
[24] "Towards Open Grid services Architecture". http://www.globus.org/ogsa
[25] "The Web Services Resource Framework". http://www.globus.org/wsrf
[26] "Grid Security Infrastructure". http://www.globus.org/security/overview.html
[27] M. Thompson, A. Essiari and S. Mudumbai, "Certificate-based Authorization Policy in a PKI Environment". ACM Transactions on Information and System Security (TISSEC), 6(4):566-588, November 2003.
[28] D. Chadwick and A. Otenko, "The PERMIS X.509 Role Based Privilege Management Infrastructure". Future Generation Computer Systems, 19(2):277-289, February 2003.
[29] V. Welch, T. Barton, K. Keahey and F. Siebenlist, "Attributes, Anonymity and Access: Shibboleth and Globus Integration to Facilitate Grid Collaboration". 4th Annual PKI R&D Workshop, April 2005.
[30] R. Alfieri, R. Cecchini, V. Ciaschini, L. dell'Agnello, Á . Frohner, A. Gianoli, K. Lörentey, and F. Spataro. "VOMS, an Authorization System for Virtual Organizations". Proceedings of the 1st European Across Grids Conference, 2003
[31] S. Tuecke. D. Engert, I. Foster, V. Welch, M. Thompson, L. Pearlman, C. Kesselman. Internet x509 Public key Infraestructure Proxy Certificate Profile, draft-ggf-gsi-proxy-04, 2002.
[32] FIP 197: Announcing the Advanced Encryption Standard, Nov. 26, 2001. http://www.csrc.nist.gov/publications/fips/fips197/fips-197.pdf
[33] A. Shamir. "How to share a secret. In Communications of the ACM". volume 22, pages 612–613, 1979.
[34] E. Torres, C. de Alfonso, I. Blanquer and V. Hernandez, "Privacy Protection in HealthGrid: Distributing Encryption Management over the VO", in Proceedings of HealthGrid 2006, Studies in Health Technology and Informatics, 2006, 120:131–41.
[35] I. Blanquer, V. Hernandez, D. Segrelles, E. Torres, "Long-term storage and management of encrypted biomedical data in real scenarios", In Proceedings of the International Conference on Emerging Security Information, Systems, and Technologies (SECURWARE 2007), pp. 77–82, October 14-20, 2007, Valencia, Spain.

I. Blanquer. Assistant Professor of the Computer System Department (DSIC) of the Polytechnic University of Valencia (UPV) since 1999. He has been involved in Parallel Computation and Medical Image processing since 12 years ago participating in 17 national and European Research Projects. He is a research fellow of the Institute for the Applications of Advanced Information and Comunication Technologies (ITACA) and Network Centre for Biomedical Engineering (CRIB) and member of the board of directors of HealthGrid association.

V. Hernández. Full Professor in Computer Science and Artificial Intelligence and leader of the Grid and High Performance Computing Group (GRyCAP). He has large experience in Parallel and Grid Computing and Numerical Methods. He has managed and participated in more than 25 European projects, from the III to the VI Framework Programme, and national projects. Prof. Hernández has been the Vice-chancellor of Research, Development and Innovation of the UPV during the period 2000-2005. He is currently the Scientific Coordinator of the Spanish e-Science Network (NGI).

D. Segrelles. Researcher in the Institute for the Applications of Advanced Information and Communication Technologies (ITACA) since 2001. He has been involved in Grid Technologies and Medical Image processing since 7 years ago participating in 7 National and European Research Projects. He is a research fellow of the GRyCAP.

E. Torres. Torres is a 3rd year PhD student in the Department of Information Systems and Computation of the UPV. He has been involved in Grid Technologies and Security since 3 years ago participating in 3 National and European Research Projects. He is currently a research fellow of ITACA.