A DATA-CENTRIC SECURITY ANALYSIS OF ICGRID

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

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

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

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

*This work was carried out for the CoreGRID IST project n^{o}004265, funded by the European Commission.
The author is also with the Dept. of Computer Science, University of Cyprus, PO Box 1678, Nicosia, Cyprus
†Also with the Dept. of Computer Science, University of Toronto, Toronto, Ontario M5S 3G4, Canada.
‡Also with the Dept. of Computer Science, University of Crete, P.O. Box 2208, Heraklion, GR 71409, Greece.

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.

Keywords: Data Grid, eHealth, Intensive Care Grid, privacy, security analysis.

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 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 [1], 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 [2] 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 [3]); 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 uses the analysis' results to introduce a privacy protocol proposed for ICGrid, able to use encryption and fragmentation to protect a patient's personal data at rest. Finally, Section 6 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* [4]. 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* [5]) 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 ICUs [6], 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 [3] has been prototyped over the EGEE infrastructure (Enabling Grids for E-sciencE [7]). 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 depicts the acquisition and annotation of parameters of an inpatient at an ICU Site (bottom left) and the transfer of data replicas to two ICGrid Storage Sites. 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. Data and Metadata are transferred to Storage Elements and Metadata servers (currently a gLite Metadata Catalogue -AMGA- service [8]) respectively, so afterwards they can be accessed by all the authorized and authenticated parties that will be entities of an ICGrid Virtual Organization. About security, the sharing and collaborative processing of medical data collected by different ICUs raises important privacy, anonymity, information integrity challenges that cannot be addressed by existing commercial ICU information systems. The rest of this paper overviews current security solutions, along with our proposal for a comprehensive low-level privacy approach.

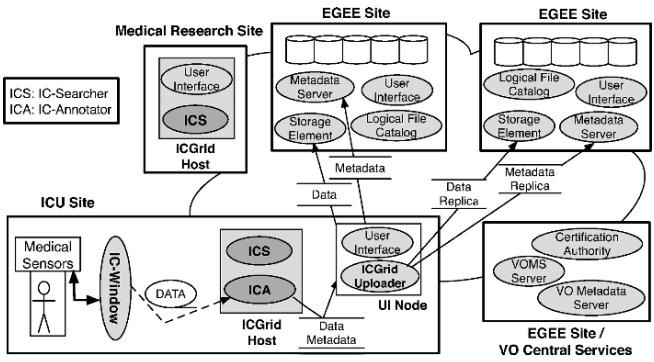

Figure 1. ICGrid System Architecture. White rectangles represent different sites of the infrastructure (each site represents resources of one administrative domain/institution), shaded rectangles represent computer nodes, and shaded ovals depict required Grid services and tools of the ICGrid framework.

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 the confidentiality of the personal data that are stored and 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 the European Council regulate the processing and management of the EHR. The common foundation of all these initiatives is the EU Directive on Data Protection [1], which provides the general framework for the protection of privacy with respect to the processing of personal data in its widest sense. The Directive concerns more than the protection of the privacy of the natural persons, since it defines *personal data* as all data related to an individual's private, public, or professional life. However, the European Working Party on Data Protection, which was established under article 29 of the Directive [1] and comprises all national data protection authorities of EU Member States, has recently acknowledged that some special rules may need to be adopted for key eHealth applications.

A fundamental term referenced in current eHealth legislations is the concept of *consent*, which is defined as any unambiguous, freely given, specific

and informed indication of the patient's wishes, with which she agrees to the processing of her personal data. In other words, *a patient's consent enables the legal processing of her EHR*. However, what happens if, for instance, after an accident the patient is unable to give her consent for accessing her personal data at the Intensive Care Unit? Most of the legal issues and ambiguities related to 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 issues. This document defines the basic recommendations regarding the protection of patients' data, which can be used towards implementing a comprehensive and harmonized technological solution as the one proposed in Section 5.

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 Electronic Health Card and the Grid Security Infrastructure.

The Electronic Health Card [10] is a new health card that stores basic patient data such as name, age, insurance details, and electronic prescriptions, including also physical features to identify the owner, i.e. a photograph and human-readable information. Basically this is a smartcard that stores information in a microchip supporting authentication, authorization and even digital signature creation, and will soon replace EU's existing health insurance cards. Data protection issues are critical in the design of Electronic Health Cards, since they store sensitive personal data that must be as secure and confidential as possible, while operating smoothly in practice. A comprehensive security concept assures the protection of the 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). Electronic Health Cards being deployed in EU Member States represent a big step towards a citizen-centered health system.

Along with the Electronic Health Card, Health Grids security is strengthened thanks to the Grid Security Infrastructure (GSI) [11] . This is a set of protocols, libraries, and tools that allow users and applications to securely access Grid resources via well defined authentication and authorization mechanisms relying on X.509 entity certificates, and XML-based protocols that retrieve security assertions from third-party services (i.e. the *Virtual Organization Membership Service* VOMS [12] used in EGEE).

Despite their security features, Electronic Health Cards and GSI do not provide adequate confidentiality guarantees for the data at rest, as our security analysis shows in the next Section.

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 [13] and extended in [2] 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- 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 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 1 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 1. 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 1 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. Protecting the patient's personal data at-rest

Up to now we have seen that the most vulnerable and critical part of Health Grid systems are the patient's personal data while at-rest on the storage elements. State of the art distributed storage systems mostly rely on fragmentation[1] ([15] and [16]), encryption ([17]) or even a mix of both ([18], [19] and [20]) for enhancing stored data assurance. Our proposal is a low-level privacy protocol that protects data and metadata from attacks targeting compromised Storage Elements, while *implementing data confidentiality and consent-like mechanisms* (in compliance with current Legislations), by using encryption and fragmentation at the ICGrid Uploader (which uses functionalities of the EGEE Storage Resource Manager -SRM- [21]).

Using the entities from ICGrid architecture (Figure 1), in Figure 2 we show the messages exchanged with the proposed protocol when data and metadata are being stored. Under this scenario the following steps will take place when an IC-Annotator (ICA) is writing a patient's private data file *(D)*:

1 The ICA computes the hash *H(D)* and *signs* this with his private key (using his Electronic Health Professional Card 3.2), that is $E_{KPrivProd}(H(D))$. This enforces non-repudiation, integrity and also provides the basis for an "electronic" consent-like mechanism.

2 Upon reception of $(E_{KPrivProd}(H(D)), D)$, the ICGrid Uploader:

   (a) Generates a nonce *N* and concatenates it to the received hash for generating the symmetric encryption key *H(D)+N*.

   (b) Uses the new key to symmetrically encrypt the data *D*, thus obtaining $E_{H(D)+N}(D)$. This provides patient's data confidentiality, therefore enforcing his right to privacy.

   (c) *Fragments* $E_{H(D)+N}(D)$ into *n-parts* and disperses these to the *Storage Elements at the EGEE Sites*.

   (d) Sends via a secure channel (using GSI) the encryption key *H(D)+N* to a VO Metadata Server hosted at the trusted EGEE Central Services. This service can be seen as a Secure Key Store possibly implemented in cryptographic hardware.

[1] In a fragmentation scheme [14], 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 ($m \leq n$) arbitrarily chosen.

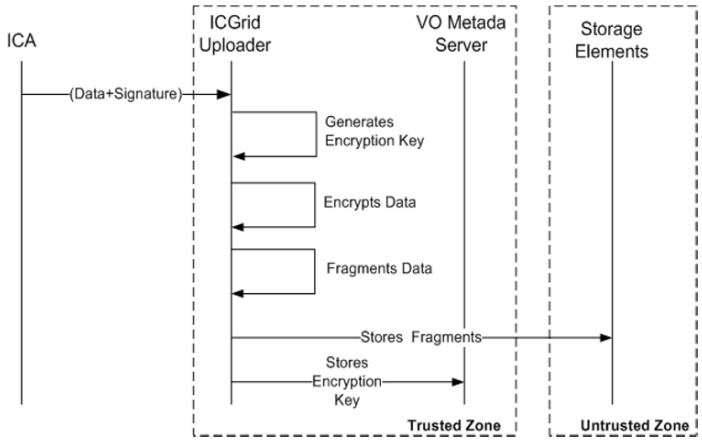

Figure 2. Privacy Protocol proposed to protect patient's data within ICGrid

Correspondingly, when an IC-Researcher (ICS) tries to retrieve a data with this protocol, the inverse sequence takes place, first of all using the ICGrid Uploader to defragment the encrypted file and then by securely retrieving the correspondent encryption key from the Central Service's Key Store. Encryption at the ICGrid Uploader is a promising solution if security issues related with the Key Store (high availability and protection of the symmetric key) and overall performance can be achieved. However, new research lines also have begun to analyze the performance gains that could be achieved if the *untrusted* Storage Elements participate in the whole encryption scheme or, if the whole fragmentation and encryption processes are performed by the *trusted* Key Store.

6. Conclusions

The computing and storage potential of the Grid are projected to play an important role for implementing Health Grid systems, able to store and manage Intensive Care Units' data. However, the deployment of production-level Health Grids, such as the ICGrid system presented in this paper, should provide assurances of the patient's data, in particular when referring to personal information, which is currently the subject of increasing concerns in most countries in the European Union. Unfortunately, when personal data is being transferred from the Hospital to the Grid new vulnerabilities may appear: on the wire, at-rest, with the metadata servers, etc. As a first step on proposing a security mechanism for Health Grids, in this paper we have performed a security analysis of ICGrid's data and metadata by applying a framework previously extended and used in Grid storage services. The results of the analysis show the need to protect the system from *untrusted Data sites*, which have full control

over the stored information, thus allowing its leak, destruction of change due to successful external or even internal attacks. It is also worth highlighting that our analysis takes into consideration the use of commonly deployed security mechanisms. After the security analysis, our research focused on proposing a privacy protocol able to protect the patient's personal data at the Storage Elements with a combination of encryption and fragmentation. The contributed protocol not only provides data confidentiality, but also integrity, high availability and a consent-like mechanism fully compliant with the legal and technological aspects discussed in this paper.

Our next steps include focusing on performance tests that will provide more information about the optimal design of the privacy protocol presented in this paper: encryption/fragmentation at ICGrid Uploader or, fragmentation at ICGrid Uploader with encryption at Storage Elements. A second promising solution refers to using the proposed Central Service's Key Store for encryption and fragmentation; this will greatly improve data assurance (encryption key are never transferred through the network), however communication overhead may become an issue. As future work we plan to base the design of the proposed Key Store into the Hydra system [20], because it resembles our needs in its application field (EGEE's Health Grid) and uses similar security mechanisms.

Acknowledgments

We thankfully acknowledge the support of the European FP6-IST program through the UNISIX project, the CoreGRID Network of Excellence and EGEE-II (contract number INFSO-RI-031688).

References

[1] 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.

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

[3] 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.

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

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

[6] 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.

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

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

[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] 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.

[11] 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.

[12] 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.

[13] 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.

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

[15] Mark W. Storer, Kevin M. Greenan, Ethan L. Miller, and Kaladhar Voruganti. Secure, archival storage with potshards. 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.

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

[17] 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.

[18] Adam L. Beberg and Vijay S. Pande. Storage@home: Petascale distributed storage. In *IPDPS*, pages 1–6. IEEE, 2007.

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

[21] Graeme A. Stewart, David Cameron, Greig A Cowan, and Gavin McCance. Storage and Data Management in EGEE. In *5th Australasian Symposium on Grid Computing and e-Research (AusGrid 2007)*, January 2007.