Chapter 25
Information Security in Data and Storage Grids through GS³

Vincenzo Daniele Cunsolo
Università di Messina, Italy

Antonio Puliafito
Università di Messina, Italy

Salvatore Distefano
Università di Messina, Italy

Marco Scarpa
Università di Messina, Italy

ABSTRACT

In grid computing infrastructures, the data storage subsystem is physically distributed among several nodes and logically shared among several users. This highlights the necessity of: (i) Availability for authorized users only, (ii) Confidentiality, and (iii) Integrity of information and data: in one term security.

In this work we face the problem of data security in grid, by proposing a lightweight cryptography algorithm combining the strong and highly secure asymmetric cryptography technique (RSA) with the symmetric cryptography (Advanced Encryption Standard, AES). The proposed algorithm, we named Grid Secure Storage System (GS³), has been implemented on top of the Grid File Access Library (GFAL) of the gLite middleware, in order to provide a file system service with cryptography capability and POSIX interface. The choice of implementing GS³ as a file system allows to protect also the file system structure, and moreover to overcome the well-known problem of file rewriting in gLite/GFAL environments. This chapter describes and details both the GS³ algorithm and its implementation, also evaluating the performance of such implementation and discussing the obtained results.

INTRODUCTION

The actual Information Technology (IT) trend definitely brings towards network-distributed paradigms of computing. Among them, the grid is one of the most widely spread. Its success is due to the fact that it manages and makes available large quantities/amounts of computing and storage resources for allocating and elaborating data as required by users' computation workflows. The management of such resources is transparent to the user that only has to specify his/her requirements in terms of resources. Then, the grid system manager automatically determines where the pro-

DOI: 10.4018/978-1-4666-4707-7.ch025

cess is executed and which resources have to be allocated to it (Foster, & Kesselman, 1998). Sharing data in distributed multi-user environments triggers problems of security concerning data confidentiality and integrity. Grid middlewares usually provide resources management's capabilities, ensuring security on accessing services and on communicating data, but they often lacks of data protection from direct malicious accesses, at system level. In other words, the fact that data are disseminated and stored in remote distributed machines, directly accessible from their administrators, constitutes the main risk for data security in grid environment. Security problems, such as insider abuse/attack, identity thefts and/or account hijacking, are often not adequately covered in grid context. It is therefore mandatory to introduce an adequate data protection mechanism, which denies data intelligibility to unauthorized users, also if they are (local) system administrators.

The problem of a secure storage access has been mainly faced in literature as definition of access rights (Junrang et al., 2004), in particular addressing problems of data sharing, whilst the coding of the data is demanded to the user, since no automatic mechanism to access to a secure storage space in a transparent way has been defined.

Scardaci, & Scuderi, (2007) proposed a technique for securing data disseminated over grid gLite (gLite, 2010) environment based on symmetric cryptography (Advanced Encryption Standard, AES). The key security is entrusted to a unique keystore server that stores it, to which all the data access requests must be notified in order to decrypt the data. This algorithm implements a spatial security policy: the security lies in physically hiding and securing the keystore server, and the access to the keystore is physically restricted and monitored in order to protect from malicious users, external attacks and insider abuses. Seitz, Pierson, & Brunie (2003) studied in depth the problem of data access, and propose a solution based on symmetric keys. In order to prevent non-authorized accesses to the symmetric key the authors propose to subdivide it on different servers. A similar technique has been specified by Shamir (1979), used in PERROQUET (Blanchet, Mollon, & Deleage, 2006) to modify the PARROT middleware (Thain, & Livny, 2005) by adding an encrypted file manager. The main contribution of such work is that, by applying the proposed algorithm, the (AES) symmetric key, split in N parts, can be recomposed if and only if all the N parts are available. HYDRA (2010) implements a data sharing service in gLite 3.0 medical environments, securing data by using the symmetric cryptography and splitting the keys among three keystore servers (Montagnat et al., 2006).

All the proposals above mentioned are based on symmetric cryptography. Most of them implement keys splitting algorithms. The underlying idea of the key splitting approach is that at least a subset of the systems (key servers) over which the keys are distributed will be trustworthy. However this approach is weak from three points of views: the security, since the list of servers with key parts must be adequately secured, the system administrators can always access the keys and it is really hard to achieve trustworthy on remote and distributed nodes for users; the reliability/availability, since if one of the server storing a part of the key is unavailable, the data cannot be accessed; the performance, since there is an initial overhead to rebuild a key, depending on the number of parts in which the key is split. A solution for improving reliability/availability is to replicate the key servers, but this contrasts with security challenges.

The goal of our work is to provide a mechanism capable to store data in grid environment in a secure way, the Grid Secure Storage System (GS^3). In order to do that, we propose to combine both the symmetric and the asymmetric cryptography. Therefore, the main contribution of the work is the specification of a lightweight and effective technique for secure data storage in grid environment that conjugates the high security goal with performance issues, as also Tu et al. (2009) have