IGI PUBLISHING **ITJ 4914**

701 E. Chocolate Avenue, Hershey PA 17033-1240, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.igi-global.com

This paper appears in the publication, International Journal of Grid and High Performance Computing, Volume 1, Issue 2
edited by **Emmanuel Udoh and Frank Zhigang Wang © 2009, IGI Global**

A Trusted Data Storage Infrastructure for Grid-Based Medical Applications

Guido J. van 't Noordende, University of Amsterdam, The Netherlands*
Silvia D. Olabarriaga, University of Amsterdam, The Netherlands
Matthijs R. Koot, University of Amsterdam, The Netherlands
Cees Th.A.M. de Laat, University of Amsterdam, The Netherlands

ABSTRACT

Most existing Grid technology has been designed with performance and scalability in mind. When using Grid infrastructure for medical applications, privacy and security considerations become paramount. Privacy aspects require a re-thinking of the design and implementation of common Grid middleware components. This article presents a novel security framework for handling privacy sensitive information on the Grid, and describes the privacy and security considerations which impacted its design. [Article copies are available for purchase from InfoSci-on-Demand.com]

Keywords: *Confidentiality; Data Storage; Distributed Systems; Grid Middleware; Medical Information; Privacy; Security*

INTRODUCTION

Most current Grid middleware is designed primarily for high-performance and high-throughput computing and data storage (LHC; Foster, Kesselman, and Tuecke, 2001). Initially, Grid infrastructure aimed mostly at the Physics community, but recently many other domains, such as Biology, Pharmaceutics, and Medical research have shown increasing interest in using Grids for their applications. Grid middleware, including gLite (gLite) and the Globus Toolkit (Globus), hides many aspects such as data distribution and replication from users of the system. As a result, users are often unaware that jobs and data are transferred through multiple Grid com-

ponents in different administrative domains implicitly. This makes it hard for users to understand the security implications of using Grid middleware, in particular when using it for applications that use privacy sensitive information.

Medical applications have very strict requirements on data handling and storage due to privacy concerns and regulations. Therefore, Grid middleware intended for usage in the medical domain should support policies that define where particular data may be stored, in what form, and which jobs from which users may access this data from what hosts or administrative domains.

This article presents a new framework for managing privacy-sensitive data on the Grid, that allows for explicit data-owner control over data access and distribution related aspects. It makes a clear distinction between data storage components, access control, job authentication aspects, and auditing mechanisms for data related operations.

This article is organized as follows: first we describe a use-case for medical research, based on our own experience (Olabarriaga, Nederveen, Snel and Belleman, 2006). Next, we analyze legal requirements with regard to medical data and technical aspects that are relevant when using Grid infrastructure to manage privacy-sensitive data. Finally, we describe a framework that allows data owners to express fine-grained data distribution and access control policies to allow for secure handling of medical data on the Grid.

USAGE SCENARIO

Figure 1 shows a typical Grid infrastructure deployment for medical research. A Grid storage system in one trusted administrative domain is used for storing medical research data. Although data is often replicated across different domains to enhance availability and reliability, we

Figure 1. A use-case for medical imaging research showing grid resources in different administrative domains, with an emphasis on data and job flow