Chapter 62
Certification and Security Issues in Biomedical Grid Portals:
The GRISSOM Case Study

Charalampos Doukas
University of the Aegean, Greece

Ilias Maglogiannis
University of Central Greece, Greece

Aristotle Chatziioannou
National Hellenic Research Foundation, Greece

ABSTRACT

User authentication and data security are very important aspects for the deployment and proper function of biomedical grid portals, since both sensitive data issues and controlled access to grid resources must be addressed. This chapter discusses certification and security issues in biomedical grid portals and presents the security infrastructure of GRISSOM (Grids for In Silico Systems biology and Medicine) platform. The platform consists of a web-based portal and a Web Service that enables statistical analysis of microarray cDNA data with the use of EGEE Grid infrastructure. The security infrastructure addresses user authentication and access issues, data encryption, Grid secure access and Web Service Security. The appendix of the chapter contains code snapshots on how to implement secure authentication in Web Services and create user SSL certificates on demand.

INTRODUCTION

In the field of bioinformatics, DNA microarray experiments are becoming a standard technique in order to examine patterns of gene expression. As this technology matures and the cost drops significantly, the amount of experimental data produced by laboratories around the world constantly increases, leading to the problem of finding powerful and easy to use analysis tools and platforms. The GRISSOM (Grids for In Silico Systems biology and Medicine) portal is a web-based platform that provides a concrete environment for data normalization, statistical gene selection, clustering and

DOI: 10.4018/978-1-4666-2919-6.ch062

annotation of microarray data exploiting the Grid infrastructure of a project called EGEE (Enabling Grids for E-sciencE, http://public.eu-egee.org/). The EGEE project is funded by the European Commission and aims to develop a service grid infrastructure available to scientists 24 hours-a-day including a Greek portion (HellasGrid), allowing the execution of parallel algorithms. Running the analysis algorithms in a parallel and distributed fashion, decreases the amount of time needed to complete without the occupation of the end user's equipment, offering large scalability. Raw data are uploaded from the user and through an easy to use step-by-step web environment he/she defines the analysis parameters before submitting it to the GRID infrastructure. The analysis is monitored automatically from the GRISSOM platform and the user is properly informed about the status of his experiment. The platform includes also access to external biological repositories and meta-data analysis resources. A web service has been created that provides access to the aforementioned resources and functionality through various application programming interfaces.

The web-based access to the computational resources of the GRID and the handling of biological data introduces many issues concerning authentication, encryption and integrity. This book chapter aims at presenting the certification and security mechanisms developed and deployed specially at the GRISSOM platform for enabling the secure transactions between users and a generic GRID infrastructure. More specifically, the chapter presents an assessment of the risk factors introducing potential vulnerability at all levels (system reliability with respect to result's correctness, system functional robustness, malicious software protection) and data protection. In addition, it discusses the development and deployment of a user registration and authentication mechanism for achieving secure access and data confidentiality against the services and the security infrastructures of a GRID Infrastructure (GSI – Globus Security Infrastructure) and monitoring services. A proper web-based mechanism for the automated provision of user certificates is also presented among with technical details for the secure communication between the platform modules and the integration with the GRID services. Finally, the appendix of the chapter contains code snapshots on how to implement secure authentication in Web Services and create user SSL certificates on demand.

BACKGROUND

Information on Biomedical Grid Portals

Most Biomedical Grid portals are Web applications that provide a front-end interface to accessing various grid computational and storage resources, along with access to special biological repositories that contain information like biological datasets and metadata, tools for meta-analysis (e.g., gene annotation), etc. Although the architecture of Web applications vary, many biomedical grid applications leverage similar sets of software components (see Figure 1). The Web Server (examples: Apache, IIS) processes incoming web requests, often routing them to other pieces of software. This is the outer layer of software for most Web applications, serving up static content when requested or delivering content dynamically generated by an individual application. The Web Application Container (examples: ASP/ASP.NET and J2EE/servlet containers like Tomcat, JBoss, Weblogic, Websphere) is a layer of software/libraries to facilitate writing, deploying and running Web applications. This typically abstracts away details like communicating with the Web server and process lifetime management so Web application authors can focus on exposing useful functionality through dynamically-generated content. Other technologies like CGI and PHP, which run as modules in the web server provide some similar capabilities. The Web Applications themselves (e.g., Grid portals) usually runs in