

A Single Sign-On Infrastructure for Science Gateways on a Use Case for Structural Bioinformatics

Sandra Gesing · Richard Grunzke · Jens Krüger · Georg Birkenheuer ·
Martin Wewior · Patrick Schäfer · Bernd Schuller · Johannes Schuster ·
Sonja Herres-Pawlis · Sebastian Breuers · Ákos Balaskó · Miklos Kozlovsky ·
Anna Szikszay Fabri · Lars Packschies · Peter Kacsuk · Dirk Blunk ·
Thomas Steinke · André Brinkmann · Gregor Fels · Ralph Müller-Pfefferkorn ·
René Jäkel · Oliver Kohlbacher

Received: 14 December 2011 / Accepted: 25 July 2012 / Published online: 23 November 2012
© Springer Science+Business Media Dordrecht 2012

Abstract Structural bioinformatics applies computational methods to analyze and model three-dimensional molecular structures. There is a huge number of applications available to work with structural data on large scale. Using these tools on distributed computing infrastructures (DCIs), however, is often complicated due to a lack of

suitable interfaces. The MoSGrid (Molecular Simulation Grid) science gateway provides an intuitive user interface to several widely-used applications for structural bioinformatics, molecular modeling, and quantum chemistry. It ensures the confidentiality, integrity, and availability of data via a granular security concept, which covers all layers of the infrastructure. The security concept applies SAML (Security Assertion Markup Language) and allows trust delegation from the user interface layer across the high-level middleware layer and the Grid middleware layer down to the HPC facilities. SAML assertions had to be integrated into the MoSGrid infrastructure in several places: the workflow-enabled Grid portal WSPGRADE (Web Services Parallel Grid Runtime

This work is supported by the German Ministry of Education and Research under project grant #01IG09006 (MoSGrid) and by the European Commission's 7th Framework Programme under grant agreement #RI-261556 (EDGI), #RI-261323 (EGI-InSPIRE), #261585 (SHIWA), and #RI-283481 (SCI-BUS).

S. Gesing and R. Grunzke contributed equally in this work.

S. Gesing (✉) · J. Krüger · O. Kohlbacher
University of Tübingen,
Tübingen, Germany
e-mail: sandra.gesing@uni-tuebingen.de

R. Grunzke (✉) · R. Müller-Pfefferkorn · R. Jäkel
Technische Universität Dresden,
Dresden, Germany
e-mail: richard.grunzke@tu-dresden.de

G. Birkenheuer · J. Schuster · G. Fels
Universität Paderborn, Paderborn,
Germany

M. Wewior · S. Breuers · L. Packschies · D. Blunk
Universität zu Köln, Köln, Germany

P. Schäfer · T. Steinke
Konrad-Zuse-Zentrum für Informationstechnik
Berlin, Berlin, Germany

B. Schuller
Forschungszentrum Jülich, Jülich, Germany

S. Herres-Pawlis
Ludwig-Maximilians-Universität München,
München, Germany

Á. Balaskó · M. Kozlovsky · A. S. Fabri · P. Kacsuk
MTA SZTAKI, Budapest, Hungary

A. Brinkmann
Johannes Gutenberg-Universität Mainz, Mainz, Germany

and Developer Environment), the gUSE (Grid User Support Environment) DCI services, and the cloud file system XtreamFS. The presented security infrastructure allows a single sign-on process to all involved DCI components and, therefore, lowers the hurdle for users to utilize large HPC infrastructures for structural bioinformatics.

Keywords Single sign-on · Science gateway · Security · DCIs · Structural bioinformatics

1 Introduction

Structural bioinformatics and computational chemistry have become indispensable tools in many fields of biomedical research. Molecular dynamics methods, quantum chemical methods, and protein-ligand docking provide deep insights into the structure of biomolecules and their interactions and are essential tools in such diverse areas as materials science and drug design. While very powerful, most of the tools and applications used for computational chemistry calculations reflect the complexity of the underlying scientific theories. Using these tools thus requires a lot of experience. Their usability is often limited and frequently deters novice users.

The computational complexity of these methods makes them ideal candidates for high-performance computing infrastructures [38]. DCIs, however, have usability issues of their own, which limit their acceptance in the scientific community.

Overcoming these usability issues and thus enabling the use of DCIs for a broader user community was the key goal when the MoSGrid (Molecular Simulation Grid) project [16] was conceived. It is part of the German Grid Initiative (D-Grid) and is designed to address the requirements of both commercial and academic users.

MoSGrid offers a science gateway for the computational chemistry community, providing easy access to tools from the field of quantum chemistry, molecular dynamics, and docking. Currently, the MoSGrid community consists of about 110 users or working groups, respectively. At this stage, the science gateway is opened for beta testing to about 20 users from academia and industry

whose feedback and demands are invaluable for further development. It is planned to offer the science gateway to the whole community in the near future. Novice and advanced users are enabled to run their computations on Grid resources. They are assisted by graphical user interfaces with different levels of sophistication to accommodate both user groups. Additionally, standard methods for specific problem classes are offered. MoSGrid provides a framework for developing, storing and providing simple and complex workflows. Furthermore, users are enabled to archive the results of their calculations in a repository and share them with other users.

Having left the first prototypic state, developments in MoSGrid continue to focus on security requirements of different environments. Distributed computing infrastructures are accessible by a number of users from different locations at the same time. The broad user community has to be provided with an infrastructure that efficiently protects their know how and molecular data.

The MoSGrid science gateway lowers the barrier of utilizing HPC infrastructures by allowing secure access to UNICORE infrastructures [47] via a single sign-on concept based on SAML [45]. This paper describes the recent developments in the MoSGrid security infrastructure. Considering both the demands of academic and commercial users, the paper focuses on the integration and interoperability of the employed components with respect to user authentication and authorization and data security.

The paper is structured as follows: Section 2 introduces to the application domain, the MoSGrid infrastructure, and related work. The developments for the MoSGrid security infrastructure are presented in Sections 3 and 4 demonstrates domain specific workflows utilizing the security infrastructure.

2 Background

Some of the application cases of structural bioinformatics and computational chemistry, in particular applications in pharmaceutical industry, impose strict requirements on data security in order to protect potential intellectual properties. We