Security and BioSimGrid: A Biomolecular Simulation Database

Bing Wu[a,b,*], Matthew Dovey[a], Kaihsu Tai[b], Muan Hong Ng[c], Stuart Murdock[c,d], Hans Fangohr[c], Steven Johnston[c], Paul Jeffreys[a], Simon Cox[c], Jonathan W. Essex[d] and Mark S.P. Sansom[b]

[a]e-Science Centre, [b]Department of Biochemistry, University of Oxford
[c]e-Science Centre, [d]Department of Chemistry, University of Southampton
[*]to whom correspondence should be addressed: bing@biop.ox.ac.uk

Abstract

The overall aim of the BioSimGrid project (www.biosimgrid.org) is to exploit the Grid infrastructure to enable comparative analysis of the results of biomolecular simulations. In particular this paper discusses the security implementation of the BioSimGrid web portal. To achieve a secured application environment, a dedicated security layer has been built on the layers of SOA (Service Oriented Architecture) framework. The security module integrates PKI (Public Key Infrastructure) and supports two levels of authentication: Grid certificate-based authentication for high security, and user/pass based authentication for maximal flexibility.

1 Introduction

Biomolecular simulations [1, 2] enable us to explore the conformational dynamics of complex molecules such as proteins, membranes and nucleic acids. In particular, molecular dynamics (MD) [3] is widely used to investigate nanosecond to microsecond dynamics for a wide range of biomolecules. Currently, a typical simulation may have a system size of ~100,000 atoms, and a nanosecond timescale simulation may require ~1,000,000 timesteps. Depending upon the efficiency of the simulation code and protocols employed, such a simulation would take a few weeks on between ~8 and ~64 processors and could generate gigabytes of data for subsequent analysis and visualisation. The Grid [4, 5, 6] is a combination of network infrastructure and software framework delivering computing services based on distributed hardware and software resources. One of the main aims of UK biological eScience projects is to exploit the Grid infrastructure to provide secure and reliable data and application services to the biology community across the UK. In particular this paper discusses the security implementation in a pilot eScience project, BioSimGrid [7].

2 Applications

Biomolecular applications are the key of the BioSimGrid project. The project is establishing a formal database for biomolecular simulations within the UK, increasing collaboration via a distributed computing environment. This also involves developing and delivering application services (Figure 1. shows simulation data of two particular proteins generated by individual laboratories can be analysed using BioSimGrid) to the users. The current methodology for developing the distributed system is Service Oriented Architecture (SOA) [8]. The key middleware of BioSimGrid are the services defined within OGSA (Open Grid Services Architecture) [9]. We also investigate compatible web service-based data access middleware, i.e. OGSA-DAI (Open Grid Services Architecture Data Access and Integration) [10], for

distributed data access. Thus the security module has been built into the layers of the open middleware.

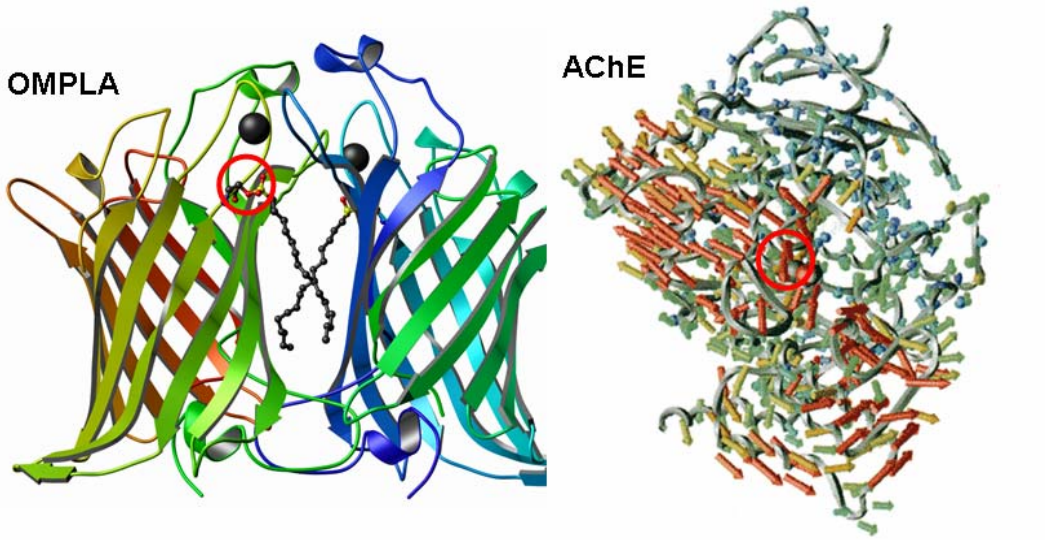

Figure 1. A BioSimGrid example application: comparison of active site dynamics

3 Security

Given the current distributed Grid implementation, security is a critical element of the project. The system needs a secure and robust environment to guarantee the smooth delivery of various BioSimGrid services. In order to achieve this, various security mechanisms have been integrated into the system. We have implemented PKI (Public Key Infrastructure) and X.509 Digital Certificate [11, 12] based authentication. Two kinds of Grid certificates (X.509 certificates) have been used in the system: user certificates for user identities and host certificates for servers. A user certificate is used to access the BioSimGrid web portal [13]. As we use the Apache server in the portal environment, host certificates are converted and configured to support SSL transactions [14, 15]. The converted key pairs can also be used as host certificates in the Globus Toolkit [4].

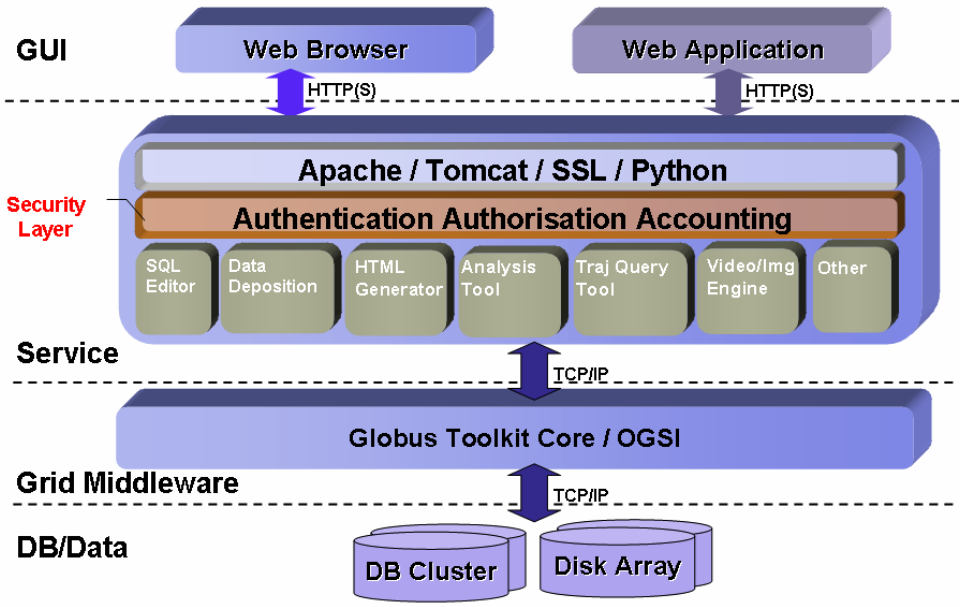

Figure 2. The BioSimGrid security layer in the application framework

The BioSimGrid network has dual firewall protection: the university campus firewall and a

local firewall. All the web transactions of BioSimGrid are based on HTTPS secure channels. A dedicated security layer has been implemented in the BioSimGrid application framework (Figure 2.).

4 Authentication

Two levels of authentication infrastructures have been implemented in the security. A secure and robust authentication mechanism is placed based on digital certificates using Grid certificates issued by UK eScience Certification Authority (CA) [16]. In Figure 3, when a user connects to the BioSimGrid portal, the local web browser creates a digital signature for the access request using the local private key. Then the request and the digital signature will be sent over to the web server. Once the web server receives the request and the user's digital signature, the AAA application will validate the signature against the request using the user's public key. The other level of authentication is a user/password based authentication, which is designed for those who have no digital certificates installed in their client machines. The username and password are collected via text fields from the browser and sent over to the server using HTTPS. While user/password based authentication enables wider access of the portal via a public PC from anywhere in the world, users of certificate based access can have more enhanced level of security for access and authentication. With a digital certificate, the portal can have instant authentication of a user's identity, instead of requiring individual username and password. The certificate is also tied to user privileges which give that user restricted access to specified portal content. As more users in the UK eScience community obtain digital certificates, we will gradually drop the legacy user/password based authentication.

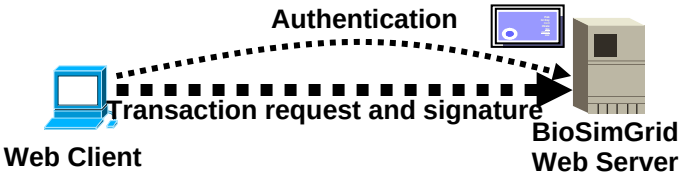

Figure 3. The certificate based user authentication

5 Authorisation

The user authorization is handled by the AAA (Authentication Authorisation Accounting) service module to control the level of user access. The authorisation is based on a distributed database and implements SSO (Single Sign On), which is designed to provide a foundation that gives users role-based access to multiple Web applications from a single, secure point of contact. In BioSimGrid, we have a distributed Grid environment and need to deal with two levels of authentication for high security and easy accessibility. To achieve this, we use SSO based on a distributed database. All the user accounts and corresponding AAA information are stored in the database distributed across the network. This enables a user sign on at any BioSimGrid site to access authorised resources and perform authorised transactions. All transactions and accounting information of the user access are logged in the accounts database locally, and then they will be distributed over the BioSimGrid sites for maximal efficiency and high availability.

6 Conclusion

The security we implemented here provides a foundation for the BioSimGrid prototype system. However, the project is still in its early stage to deliver full services. Our future work on the security involves the integration of user credential delegation using MyProxy [17] and the integration of WS-Security [18] into the system.

7 Acknowledgements

Many thanks to our BioSimGrid partners (C. Laughton, L. Caves, D. Moss, A. Mulholland and O. Smart) for their input to this project. BioSimGrid is funded by BBSRC and DTI. Our thanks to all of our colleagues in the Oxford and Southampton simulation labs, to the Oxford e-Science Center, to the Southampton Regional e-Science Centre, and to the Security Task Force of UK eScience Core Programme for their encouragement, advice and hard work.

References

[1] Oren M. Becker, Alexander D. Mackerell Jr, Benoît Roux, Masakatsu Watanabe, Eds. (2001). Computational Biochemistry and Biophysics. Marcel Dekker.
[2] Bourne, P.E. and Weissig, H. (2003). Structural Bioinformatics, Wiley-Liss, Hoboken.
[3] Karplus, M.J. and McCammon, J.A. (2002). Nature Struct. Biol., 9, 646-652.
[4] http://www.globus.org
[5] Berman, F., Fox, G. and Hey, T., Eds. (2003). Grid Computing: Making the Global Infrastructure a Reality, Wiley.
[6] Foster, I. and Kesselman, C., Eds. (1999). The GRID: Blueprint for a New Computing, Morgan-Kaufmann.
[7] Bing Wu, Kaihsu Tai, Stuart Murdock, Muan Hong Ng, Steven Johnston, Hans Fangohr, Paul Jeffreys, Simon Cox, Jonathan Essex and Mark S.P. Sansom. (2003). BioSimGrid: A Distributed Database for Biomolecular Simulations. Simon J Cox (editor), Proceedings of UK e-Science All Hands Meeting 2003. EPSRC, ISBN 1-904425-11-9.
[8] Hao He (2003). What is Service-Oriented Architecture, http://webservices.xml.com/pub/-a/ws/2003/-09/30/soa.html
[9] Foster, I., Kesselman, C., Nick, J. and Tuecke, S. (2002). The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration, Global Grid Forum.
[10] http://www.ogsa-dai.org
[11] Tuecke, S., Engert, D., Foster, I., Thompson, M., Pearlman, L. and Kesselman, C. (2001). Internet X.509 Public Key Infrastructure ProxyCertificate Profile, IETF.
[12] ftp://ftp.isi.edu/in-notes/rfc2459.txt
[13] Bing Wu, Matthew Dovey, Muan Hong Ng, Kaihsu Tai, Stuart Murdock, Paul Jeffreys, Simon Cox, Jonathan Essex and Mark S.P. Sansom, A Web / Grid Portal Implementation of BioSimGrid: A Biomolecular Simulation Database, Proceedings ITCC 2004: International Conference on Information Technology: Coding and Computing, vol II, pages 50-54. Pradip K. Srimani et al., editors. Los Alamitos, California: IEEE Computer Society. ISBN 0-7695-2108-8
[14] Bing Wu (2003). User Certificate Installation Guide, http://www.biosimgrid.org/docs/-2003/NOTES-/user_certificates.pdf
[15] Bing Wu (2003). Digital Certificate Installation Guide, http://www.biosimgrid.org/docs/-2003/NOTES/certificates.pdf
[16] http://www.grid-support.ac.uk/ca/
[17] http://www.ncsa.uiuc.edu/Divisions/ACES/MyProxy/
[18] http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=wss