

Chapter 4.6

Molecular Structure Determination on the Grid

Russ Miller

Hauptman-Woodward Medical Research Institute, USA & SUNY-Buffalo, USA

Charles M. Weeks

Hauptman-Woodward Medical Research Institute, USA

ABSTRACT

Grids represent an emerging technology that allows geographically- and organizationally-distributed resources (e.g., computer systems, data repositories, sensors, imaging systems, and so forth) to be linked in a fashion that is transparent to the user. The New York State Grid (NYS Grid) is an integrated computational and data grid that provides access to a wide variety of resources to users from around the world. NYS Grid can be accessed via a Web portal, where the users have access to their data sets and applications, but do not need to be made aware of the details of the data storage or computational devices that are specifically employed in solving their problems. Grid-enabled versions of the SnB and BnP programs, which implement the Shake-and-Bake method of molecular structure (SnB) and substructure (BnP) determination, respectively, have been deployed on NYS Grid. Further, through the Grid Portal, SnB has been run simultaneously on all computational resources on NYS Grid as well as on more than 1100 of the over 3000 processors available through the Open Science Grid.

1. INTRODUCTION

The Grid is a rapidly emerging and expanding technology that allows geographically-distributed resources that extend across administrative boundaries to be linked together in a transparent fashion

(www.gridcomputing.com/; www.globus.org/; Berman *et al.*, 2003; Foster & Kesselmann, 1999). These resources include compute systems, data storage devices, sensors, imaging systems, visualization devices, and a wide variety of Internet-ready instruments. The concept and terminology of the Grid is borrowed from the electrical grid, where utility companies have the ability to share

DOI: 10.4018/978-1-4666-0879-5.ch4.6

and move resources (electricity) in a fashion that is transparent to the consumer. With rare exception, the view taken by a consumer is that they are able to plug a piece of equipment into a power outlet in order to obtain electricity and do not need to know, and in fact, do not want to know, the details pertaining to the manner in which electricity makes its way to the outlet. Similarly, the power of both computational grids (*i.e.*, seamlessly connecting compute systems and their local storage) and data grids (*i.e.*, seamlessly connecting large storage systems) lies not only in the aggregate computing power, data storage, and network bandwidth that can readily be brought to bear on a particular problem, but also on its ease of use.

Numerous government-sponsored reports state that grid computing is a key to 21st century discovery by providing seamless access to the high-end computational infrastructure that is required for revolutionary advances in contemporary science and engineering. In fact, National Science Foundation Director Arden Bement stated that “*leadership in cyberinfrastructure may determine America’s continued ability to innovate – and thus our ability to compete successfully in the global arena.*”

Grids are now a viable solution to certain computationally- and data-intensive computing problems for reasons that include the following.

- Users can access many grids through a Web portal from virtually anywhere in the world. That is, a user only needs an account on a Grid administrative server in order to use a grid. This is similar to how someone uses a search engine or large e-business system. One only needs access to a gateway and not specifically to each individual server that the company has configured to be able to handle the requests/queries/business. For most grids, a user needs access to a Grid Portal, but does not need to be logged in to a site that hosts a particular Grid resource, does not need to be logged in to a computer that is on a Grid, and does

not need to install any additional software on their Web-accessible system (workstation, cellular phone, laptop, etc.) in order to be able to use a Grid.

- The Internet is mature and able to serve as the fundamental infrastructure for network-based computing. In fact, network bandwidth, which has been doubling approximately every 12 months, has increased to the point of being able to provide efficient and reliable services for the vast majority of Grid applications.
- Storage capacity, which has been doubling approximately every 9 months, has now reached commodity levels, where one can purchase a terabyte of disk for roughly the same price as a high-end PC.
- Many instruments are Internet-aware.
- Clusters, supercomputers, storage and visualization devices are becoming more mainstream in terms of their ability to host scientific applications.

As grid computing initiatives move forward, issues of interoperability, security, performance, management, and privacy need to be carefully considered. In fact, security is concerned with various issues relating to authentication in order to insure application and data integrity. Grid initiatives are also generating best practice scheduling and resource management documents, protocols, and API specifications to enable interoperability. Several layers of security, data encryption, and certificate authorities already exist in grid-enabling toolkits such as Globus Toolkit (www.globus.org/toolkit/).

In this chapter, we present the *New York State Grid (NYS Grid)*, which represents a major effort by our research group. The design, implementation, and support of this Grid has provided a wide variety of opportunities for institutions in New York State in terms of science, engineering, scholarship, education, outreach, and training. Indeed, the New York State Grid consists of a

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/molecular-structure-determination-grid/64519

Related Content

The State of the Art and Open Problems in Data Replication in Grid Environments

Mohammad Shorfuzzaman, Rasit Eskicioglu and Peter Graham (2010). *Handbook of Research on Scalable Computing Technologies* (pp. 486-516).

www.irma-international.org/chapter/state-art-open-problems-data/36422

Federated PKI Authentication in Computing Grids: Past, Present, and Future

Massimiliano Pala, Shreyas Cholia, Scott A. Rea and Sean W. Smith (2011). *Cloud, Grid and High Performance Computing: Emerging Applications* (pp. 165-179).

www.irma-international.org/chapter/federated-pki-authentication-computing-grids/54928

A Hybrid Feature Selection Method for Effective Data Classification in Data Mining Applications

Ilangovan Sangaiya and A. Vincent Antony Kumar (2019). *International Journal of Grid and High Performance Computing* (pp. 1-16).

www.irma-international.org/article/a-hybrid-feature-selection-method-for-effective-data-classification-in-data-mining-applications/216478

Cost Efficient Implementation of Multistage Symmetric Repackable Networks

Amitabha Chakrabarty and Martin Collier (2011). *International Journal of Grid and High Performance Computing* (pp. 1-13).

www.irma-international.org/article/cost-efficient-implementation-multistage-symmetric/60251

Web Application Server Clustering with Distributed Java Virtual machine

King Tin Lam and Cho-Li Wang (2010). *Handbook of Research on Scalable Computing Technologies* (pp. 658-681).

www.irma-international.org/chapter/web-application-server-clustering-distributed/36428