An Implementation of Interactive Jobs Submission for Grid Computing Portals

XIAO Haili **WU Hong** **CHI Xuebin** **DENG Sungen** **ZHANG Honghai**

Supercomputing Center, Computer Network Information Center
Chinese Academy of Sciences, PO Box 100080, Beijing, China
Email: {haili, wh, chi, dsg, zhh}@sccas.cn

Abstract

The Globus Toolkit has been widely used as middleware in Grid computing environments. Java CoG, Web Service and Portlet help to build Grid computing portals easily and productively. Many Grid Portals can provide a customizable interface allowing scientists and researchers to perform Grid operations such as remote submission of their own programs, staging input and ouput files, and querying resources and queues information. However, the Globus Toolkit does not support interactive jobs submission, nor is there an interactive extension for Grid computing portals. In this paper, a Grid Security Infrastructure (GSI) enabled Java SSH client is introduced and implemented as a novel way to submit interactive jobs to the computing resources in ScGrid, which is a scientific computing environment provided by the Supercomputing Center, Computer Network Information Center, Chinese Academy of Sciences (SCCAS).

Keywords: Grid Computing, Grid Portals, Interactive Jobs, GSISSH

1 Introduction

The Grid technology has spurred a revolution in the way science is conducted. More and more scientists and engineers are able to share computing power, databases, scientific instruments, and other tools securely across corporate, institutional, and geographic boundaries. The Globus Alliance focuses on enabling Grid concepts for scientific and engineering applications. The Globus Toolkit has rapidly evolved into the de facto standard for Grid computing (Foster, Kesselman 1998, Foster et al. 2002).

Additionally, interfaces for accessing Grid enabled resources are making great progress. Traditionally, only high-level client side tools were provided that encapsulate a variety of distributed Grid operations such as transferring data, executing jobs, and visualization of its final result across heterogeneous resources. Client tools are capable of providing the most direct and specialized access to Grid resources, but the main limitation of this form is its complication and inconveniency because the end users are required to deploy and configure the specialized client applications, libraries and packages. Therefore many Grid portals were developed to provide application scientists and researchers a customized view of software and hardware resources from a web browser. A Grid portal is defined to be a web based application server enhanced with the necessary software to communicate with Grid services and resources (Novotny 2000).

The Grid Portal Development Kit (GPDK) provides reusable Grid components and a useful framework for the development and deployment of application specific portals. It has been proven to be an extensible and flexible software package. It is based on the Model-View-Controller design paradigm and makes use of commodity technologies including the open source servlet container Tomcat, the web server Apache and the Java Commodity Grid (CoG) toolkit (Novotny 2000).

The Supercompuing Center, Computer Network Information Center, Chinese Academy of Sciences (SCCAS) has several high performance supercomputers (Deepcomp 6800, Dawning 2000, and Hitachi SR2201, etc.), a scientific visualization supercomputer (SGI Onyx 350) and a storage system of 61 TB fiber disk array. Deepcomp 6800, when installed in the SCCAS in 2003, was ranked the 14th (Nov. 16, 2003) most powerful computer in the world at 4.2 Tflops, and also selected as one of the two main nodes of the China National Grid (CNGrid). It is ranked 26th in the latest (June, 2004) top500 list.

As a part of the CNGrid, the Scientific Computing Grid (ScGrid) was developed and deployed to securely integrate and collaboratively use these resources. A web based interface, the ScGrid portal, is provided to access Grid services in the ScGrid. In the current implementation of the ScGrid portal, the Globus Toolkit 2 has been chosen as the middleware to provide Grid services, while the Java CoG and the Grid Portal Development Kit (GPDK) are the rapid prototype tools (with many customized modification) to build the portal. The architecture of the ScGrid is illustrated in Figure. 1.

There are more than eighty ScGrid users, who are scientists, researches or graduates in many different scientific fields, including but not limited to Biophysics, Ecology, Geology, Geophysics, Mechanics, Space Weather and Astronomy. Most of them are from the Chinese Academy of Sciences and universities scattered in China. In ScGrid, they are provided with services such as submitting jobs (including batch jobs and interactive jobs), viewing job outputs, accounting services, querying nodes resources and queues information, plus files and directories operations.

This work was supported by the National High Technology Research and Development (863) Program of China "Creating Grid Environments for Supercomputing" (2002AA104540), and the Informalization Construction of Knowledge Innovation Project of the Chinese Academy of Sciences "Creating Supercomputing Environments and Applications" (INF105-SCE).

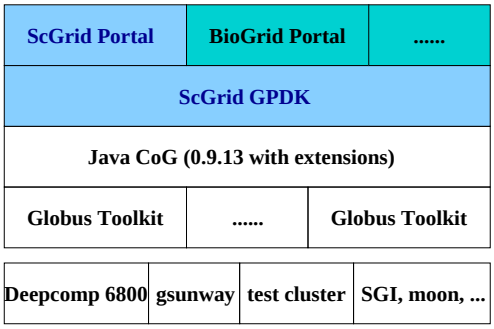

Figure 1: Architecture of the ScGrid

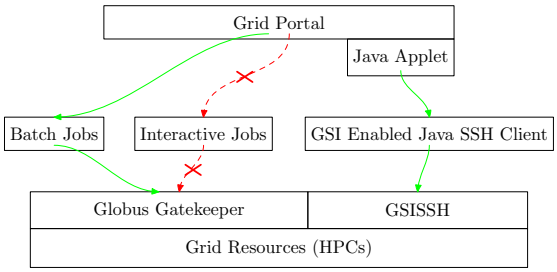

Figure 2: Submit interactive jobs via a Grid Portal. (The dashed curve in the middle is impossible. The solid curve on the right is the new approach.)

2 Overview of Jobs Submission in a Grid Environment

2.1 Batch Jobs and Interactive Jobs

Generally, there are two kinds of job submission methods: batch execution and interactive execution.

When a batch job is submitted, it is submitted to a queue, where it waits until it reaches the top of the queue and the required resources become available.

An interactive job is a job where the nodes are reserved for users to log into and issue commands by hand. An interactive job is run immediately upon submitting if the specified resource are available. Since the interactive jobs never wait, and all the input and output are handled transparently, it is very useful for testing batch scripts and debugging programs .

Although interactive jobs may decrease the utilization of the system, it is vital for a working scientific computing system.

2.2 Jobs Submission Using the Grid Middleware

The Globus Toolkit does not accept any interactive jobs submission. An analysis of the Globus Toolkit source code, especially code snippets related to job submission, reveals that it is impossible to submit interactive jobs to a Globus Gatekeeper, which is the only entrance for submitting jobs. There are two console commands available to submit jobs, `globus-job-submit` and `globus-job-run`. The former is used to submit batch job. It passes the execution request to the gatekeeper, returns immediately with a specified job ID or error messages if failed. The latter will block until the job is finished or cancelled. Neither allows any kind of user input during the lifecycle of a running job.

2.3 Jobs Submission from the ScGrid Portal

When an interactive job submission is requested, this novel approach creates an interactive secure shell process (Figure. 2) in the computing portal rather than directly using the Globus Gatekeeper. With Java Applet, an interactive interface can be easily set up in a normal portal system. Many open source implementations of telnet/ssh for Java are available, but none support the Grid Security Infrastructure (GSI) extensions, which comprises the basic security infrastructure. GSI is used by the Globus Toolkit for enabling secure authentication and communication over an open network. It provides mutual authentication and single sign-on. This project implemented a GSI enabled Java SSH client as an interactive extension to the Globus Gatekeeper.

The usability of the ScGrid portal will be greatly enhanced after this successful implementation of the interactive jobs submission.

3 Implementation of Interactive Jobs Submission

3.1 Latest Progresses

There are some efforts, like EDG 2.0 and GridPort 3, which provide interactive capabilities, but very limited. EDG 2.0 supports interactive grid jobs, however, an X-window server is required running on the user's desktop. It is troublesome for the Windows users since they must install and start X server first. In addition, sometimes this solution would not work due to the firewall restriction. In our design, X server is not required for non-GUI programs. In GridPort 3, interactive job submission provides only immediate output to the user, no standard input. It's not really interactive.

Late 2003, a GSI enabled Java SSH client was being developed by the ScGrid working team, with the aim to provide interactive jobs submission capability via the ScGrid portal.

In February 2004, after the GlobusWORLD, Jean-Claude Cote wrote, "there is not any GSI enabled Java SSH client currently available for Grid computing and that it is on the wish list of many Grid infrastructure developers." Consequently he added a GSI authentication module to the SSHTools, an open source project on sourceforge.net, to authenticate to a GSI-Enabled OpenSSH server.

Unfortunately, this client does not work with any GSISSH server based on the Globus Toolkit 2. These include many Grid projects such as the Scientific Computing Grid from the SCCAS, the Asia Pacific Grid from the Advanced Industrial Science and Technology (AIST), Japan, and the K*Grid from the Korea Institute of Science and Technology Information (KISTI).

Therefore it was crucial to implement a GSI enabled Java SSH client that supports the Globus Toolkit 2.

3.2 Implementation

3.2.1 GSS-API and extensions

The Generic Security Service Application Program Interface (GSS-API) is defined in RFC 2743 and RFC 2744. It provides security services such as confidentiality, integrity and delegating rights. The services and primitives are independent of underlying security mechanism and programming language (Linn 2000, Wray 2000).

The GSI Working Group of the Global Grid Forum defines the GSS-API extensions. The extensions include credential export and import, delegation at

any time, and credential extensions handling (Meder et al. 2003).

GSS-API adds authentication, delegation and cryptography to distributed applications. In 1997, the Globus Project (subsequently it is the Globus Alliance) introduced the Grid Security Infrastructure (GSI), which was an implementation of GSS-API. Based on this implementation, a lot of Grid applications were developed and deployed in distributed computing environments. Great success has been achieved, however, the deficiencies in the existing GSS-API have emerged.

3.2.2 Secure Shell

The Secure Shell (SSH) is implemented at the application layer. It is designed to run over reliable but insecure transport layer protocols, like TCP/IP. SSH provides strong security against cryptanalysis and protocol attacks. It supports remote login, file transfer, TCP/IP and X11 forwarding. SSH is maintained by an IETF working group, which has produced 14 Internet Drafts about SSH (Ylonen Oct. 2003, Lehtinen et al. 2003, Ylonen Sep. 2002).

The SSH protocol consists of three layers: transport protocol, user authentication protocol and connection protocol. The SSH transport layer provide integrity and confidentiality protection. It is designed to be flexible. Many parameters, such as the key exchange method, are negotiated. The SSH authentication protocol is a general user authentication protocol and runs atop the transport layer. Authentication method is also negotiated. The SSH connection protocol multiplexes the encrypted tunnel into several logical channels (Ylonen Oct. 2003, Lehtinen et al. 2003, Ylonen Sep. 2002).

3.2.3 SSH Authentication and Key Exchange with GSS-API

There is an SSH user authentication method defined for using the GSS-API for authentication and key exchange in SSH. A specified GSS-API mechanism and a family of SSH key exchange methods are used to authenticate a user. The new user authentication method allows an authorization name to be used in conjunction with any prior authentication via key exchange. These definitions allow any GSS-API mechanism to be used with the SSH protocol (Hutzelman et al. Jan. 2002, Hutzelman et al. Nov. 2002, Hutzelman et al. 2003).

There are several similar key exchange and user authentication methods used during the evolution of the Globus Toolkit. The Globus Toolkit 2 uses "external key exchange user authentication". This user authentication method relies on the key exchange to authenticate both the client and the server.

3.2.4 A GSI Enabled Java SSH Client

The open source project SSHTermApplet was modified by adding a GSI authentication module which used the external key exchange user authentication described above. This applet enhances the ScGrid portal by providing interactive jobs submission capability. It is an interactive extension to the Globus middleware that provides a novel approach to submitting interactive jobs via a Grid portal.

3.2.5 An Example

There is a simple interactive script in the EDG Users' Guide. It reads a user's name from the standard input then writes it to the standard output. Figure. 3 shows the execution process of this interactive script.

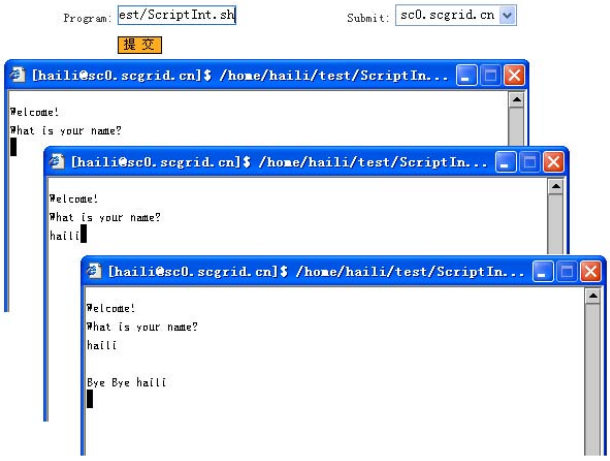

Figure 3: Submit an Interactive Shell Script

4 Conclusions

The Globus Toolkit is widely used as middleware to successfully provide Grid services. However, previously it was impossible to submit interactive jobs to a Globus Gatekeeper. A novel way is introduced to submit interactive jobs to Grid resources via a web-based Grid computing portal. It is based on a GSI enabled Java SSH client which uses the "external-keyx" user authentication extension. Interactive jobs are submitted to an interactive secure shell process rather than directly to a Globus Gatekeeper.

This implementation is used in the development of the ScGrid portal. Both batch jobs submission and interactive jobs submission from the ScGrid portal are now provided for the ScGrid users. Based on this implementation, more functional modules will be developed to enhance services for the ScGrid users.

5 Acknowledgements

We would like to thank Dr. Coleman Mosley for reviewing drafts of this paper. He offered many constructive suggestions on both the contents and grammar. We would also like to thank Mrs. Suhua Xiao and Mr. Tie Niu for their assistance with working on the supercomputers of the Computer Network Information Center, Chinese Academy of Sciences. Insightful ideas and encouragement from other colleagues helped us to persist in finishing the ScGrid project.

References

European DataGrid Integration Team. (2003), *EDG User's Guide.*

Foster, I., & Kesselman, C. (1998), *The Grid: Blueprint for a New Computing Infrastructure*, Morgan Kaufmann.

Foster, I., Kesselman, C., Nick, J. & Tuecke, S. (2002), *The Physiology of the Grid: An Open Grid Services Architecture for Distributed Systems Integration.*

Friedl, M., Provos, N., & Simpson, W. A. (2003), *Diffie-Hellman Group Exchange for the SSH Transport Layer Protocol*, Internet-Draft, IETF Network Working Group.

Hutzelman, J., Salowey, J., Galbraith, J., & Welch, V. (2002), *GSSAPI Authentication and Key Exchange for the Secure Shell Protocol*, Internet-Draft, IETF Network Working Group.

Hutzelman, J., Salowey, J., Galbraith, J., & Welch, V. (2002), *GSSAPI Authentication and Key Exchange for the Secure Shell Protocol*, Internet-Draft, IETF Network Working Group.

Hutzelman, J., Salowey, J., Galbraith, J., & Welch, V. (2003), *GSSAPI Authentication and Key Exchange for the Secure Shell Protocol*, Internet-Draft, IETF Network Working Group.

Laszewski, G., Foster, I., & Gawor, J. (2000), *CoG Kits: A Bridge Between Commodity Distributed Computing and High-Performance Grids*, Proceedings of the ACM Java Grande Conference.

Laszewski, G., Pieper, G. W., & Wagstrom, P. (2002), *Gestalt of the Grid*.

Laszewski, G., Alunkal, B., & Amin, K. et al. (2003), *The Java CoG Kit User Manual Draft Version 1.1a*.

Lehtinen, S., & Moffat, D. (2003), *SSH Protocol Assigned Numbers*, Internet-Draft, IETF Network Working Group.

Linn, J. (2000), *Generic Security Service Application Program Interface Version 2, Update 1*, RFC 2743, IETF Network Working Group.

Meder, S., Welch, V., Tuecke, S., & Engert, D. (2003), *GSS-API Extensions*, Grid Security Infrastructure (GSI) Working Group of the Global Grid Forum.

Novotny, J. (2000), *The Grid Portal Development Kit*, Concurrency: Pract. Exper. 2000; 00:1-7.

Novotny, J., Tuecke, S., & Welch, V. (2001), *An Online Credential Repository for the Grid: MyProxy*, Proc. 10th IEEE Symp. On High Performance Distributed Computing.

Novotny, J., Russell, M., & Wehrens, O. (2003), *GridSphere: A Portal Framework For Building Collaborations*.

Sterling, T. (2001), *Beowulf Cluster Computing with Linux*, MIT Press.

Wray, J. (2000), *Generic Security Service API Version 2 : C-bindings*, RFC 2744, IETF Network Working Group.

Ylonen, T., & Moffat, D. (2002), *SSH Authentication Protocol*, Internet-Draft, IETF Network Working Group.

Ylonen, T., & Moffat, D. (2003), *SSH Protocol Architecture*, Internet-Draft, IETF Network Working Group.

Ylonen, T., & Moffat, D. (2003), *SSH Connection Protocol*, Internet-Draft, IETF Network Working Group.

Ylonen, T., & Moffat, D. (2003), *SSH Transport Layer Protocol*, Internet-Draft, IETF Network Working Group.

https://3sp.com/tutorials/sshterm/sshtermapplet.htm

http://www.863.org.cn

http://www.ggf.org

http://www.globus.org

http://www.gridforum.org

http://grid.nrc.ca:8080/gsissh

http://gridport.net

http://www.hpclab.niu.edu/mpi

http://ietf.org/html.charters/secsh-charter.html

http://jakarta.apache.org/tomcat

http://javassh.org

http://java.sun.com/products/servlets

http://www.sccas.cn

http://www.scgrid.cn

http://sourceforge.net/projects/sshtools

http://www.top500.org

http://www.unicore.de