Starlink Software - a Grid and Web enabled future using Globus?

B. K. McIlwrath and D. L. Giaretta

Starlink Project, Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX11 0QX, United Kingdom

Abstract. The rapid development of AstroGrid within the UK community has made it important that as much existing data analysis software as possible is made "Grid aware". With the geographical dispersion of resources on the Grid, web-interfaces (known as "portals") to applications and utilities are likely to became increasingly prevalent. This paper describes preliminary work done by Starlink in these areas and possible future directions.

A prototype Grid-Portal has been created using "MyProxy" technology to demonstrate secure login, access to several Grid utilities and the capability to run several basic Starlink Applications. Two approaches to enable remote data access for existing applications (direct integration with the Globus API and indirect use of Globus utilities) have been shown to be viable.

The Globus Toolkit as well as derived and alternative Grid "middleware" are all in a period of very rapid development. It is, however, expected that this preliminary work by Starlink will place the Project in a good position to take rapid advantage of forthcoming advances.

1. Introduction

Computing Grids (Foster & Kesselman 1998) are a recent and rapidly developing concept of utilising very fast network connections to promote remote use and sharing of computing and data resources. A Grid has been described as *An infrastructure that enables flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions and resources.*

The UK Government, amongst many others, sees great potential in Grid developments and has allocated a total of £98m for a three-year e-Science programme, Out of this the Particle Physics and Astronomy Research Council has allocated £??m to fund an *AstroGrid project.* This will have a profound impact on the way astronomical research is carried out in the UK. AstroGrid is a consortium member of the related European Astronomical Observatory (AVO) project and there will be collaboration with the US National Virtual Observatory (NVO) project.

The Starlink Project has, for many years, supplied applications software to UK-affiliated and other astronomers. These applications have been in the main traditional in style and designed to run on single workstations. However, the fact that the Starlink parameter system is based on observatory control concepts

means that applications can be run via a message passing interface in addition to directly from the shell. This feature has been widely exploited to provide a means of starting and controlling applications from a wide variety of command interfaces based on an ever-expanding number of languages (Tcl/TK, Perl, JAVA JNI, etc.). Data access by applications consistently uses the n-Dimensional Data Format (NDF) (Warren-Smith 19xx) based upon the Hierarchical Data System (HDS) (Warren-Smith & Lawden 19xx)

This paper describes preliminary work to investigate how the traditional suite of *astronomical applications* described above might be developed to work within the somewhat different requirements of a Grid. The *Grid middleware* used in this investigation was GLOBUS (Globus 2001). This software provides a flexible Toolkit to enable software and resources to be *Grid Enabled* to any desired level. Our investigations also used an axiom that, for astronomical data analysis, sharing of data and remote running of applications will, initially, be the most important aspects of Grids.

1.1. The GLOBUS middleware and portals

It is planned that the UK AstroGrid will initially utilise the GLOBUS middleware. This provides a flexible *toolkit* of Grid facilities in the following areas:

- Security
- Information Infrastructure
- Resource Management
- Data management
- Communication
- Fault detection

Not all *Grid enabled* applications are required to use all these facilities.

The area of security is important as the intention of Grids will be to make selected local facilities (CPU power, data etc.) available *ONLY* to authorised remote users. The Globus solution is to use public key encryption, X.509 certificates and the secure sockets layer (SSL). Both services and users must possess certificates signed by a trusted authority in addition to explicit authorisation set by the resource administrator.

Data access for authorised users is available in two forms. The first of these are object library routins which allow executable programs access to the *GLOBUS Access to Secondary Storage* (GASS) mechanism. The second method is to use an adapted form of the standard NCFTP file transfer application – these adaptations allow GLOBUS X.509 security to be used as an alternative to the normal username/password ftp login mechanism. Plans for the future are that GLOBUS will develop its own Grid-ftp application which will add novel features such as partial file transfers and multi-threaded parallel transfers of large datasets.

A related area of Grid applications development is the construction of *Web portals* which allow users access to Grid resources via an easy to use Web page interface – often, but not exclusively, this has been to allow secure access to high performance computers and applications (see summary in, for example, Allan, 2001).

2. Grid data access in action

As stated in the Introduction all Starlink Applications use consistent data access to Hierarchical Data System (HDS) format disk files (frequently accessed on NFS mounted disks) via a set of conventions embodied in the n-Dimensional Data Format (NDF) access layer. Transparent conversion to and from a number of *foreign* data formats (for example, IRAF data) is available at the NDF level via its hidden use of the Starlink *CONVERT* package.

Figure 1 (the DLG slide from the presentation) shows the two methods developed to "Grid enable" Starlink data access in a manner transparent to users.

- Use of CONVERT. A simple change to an NDF subroutine enabled remote files to be accessed via Grid-Enabled-NCFTP from another computer. The only change required by the user is a new file access syntax of the form `<remote machine>/<filename>.sdf.GLOBUS` to automatically invoke the correct CONVERT routine.
- Use of GASS. A special version of the HDS library which included GASS library routines was created. As GASS requires a server process to be running on the remote computer the access syntax is of the form `http://machine name>:<GASS server port>/<filename>` where the server startup gives the *server port number*.

3. Experimental Portal developments

A development Web-Portal was created using the MyProxy software (references and detailed description in Allan 2001). This software allows properly authorised users the ability to create a time-limited, delegated-Proxy certificate analogous to that used for normal Grid access by the Globus command `grid-proxy-init` and to securely transfer this to an agreed location also accessible to a Web server (which may be anywhere on the Internet). Other MyProxy software allows CGI scripts running on this web server to securely acquire this certificate. This certificate is protected by a user-selected password (which is unrelated to that of the high security GLOBUS X.509 authorisation). It is normal to arrange for the CGI script to initially present a *Login form* to a new user requiring him or her to login to using this secondary password and to only allow further actions when this is correct.

This Proxy certificate poses a slight security exposure but, as it is time limited, derived from the users (secure) X.509 certificate and passed to the Web server via secure channels this is, in practice, minor. Also, to acquire access to the certificate the Web server has to present its *OWN* X.509 Globus certificate to the system. Users are confined to operations on the Web server host which have been embedded in the CGI script they are presented with.

Once authenticated to the Web server (in principle from anywhere in the world) users have access to facilities on the server host within the limits imposed by the CGI scripting author. The demonstration Grid Portal described here uses the MyProxy authentication and CGI scripting in two distinct ways:

- Tasks requiring GLOBUS credentials. Although the Web server process will be running under a non-privileged username on its host machine (typ-

ically *nobody* or *webmaster*) it can perform Globus tasks for the remote user by using the X.509 proxy as if it *WAS* that user. In our simple example we have arranged for it to start a GLOBUS GASS file server. More complex Web Portals which give access to supercomputer facilities currently use this facility more extensively and, undoubtably, it will provide a fruitful development area for the future.

- Other tasks. As has been noted above the user *logs in* to the remote Web server in a highly secure manner using features of the Globus GSI security model. This presents the interesting possibility of allowing the CGI script to securely run utility programs (for example a process list or directory listing) up to complex applications on the web server host. Unlike when the web server runs Grid applications these processes will run under the UID of the Web server – which may present problems in some cases. As a *proof of concept* of this we have arranged for several Starlink applications to be run in this manner. These applications are invoked by the Apache Web-server from Perl CGI scripts which use a mechanism developed by at JAC, Hawaii (Jenness & Econonou 2000) to incorporate many disparate Starlink applications into a *data analysis pipeline*.

4. Summary

This paper has described preliminary investigations into making a *traditional* astronomical data analysis system *Grid Aware* with the minimum of changes. It has been shown that the existing data access layer could be made *Grid aware* using two distinct approaches. The use of Grid Portals has also been investigated and they have been shown to have great potential – both in remote access to applications specifically developed for the Grid and, using the security components alone, for remote access to some existing applications.

5. Demos available - for poster only?

References

Allan, R. J., UKHEC draft report, 2001, (www.dl.ac.uk//TCSC//UKHEC/GridWG)

Foster, I. & Kesselman, C. 1998, The Grid - Blueprint for a New Computing Infrastructure (Morgan Kaufmann Publishers - ISBN 1-55860-475-8)

The GLOBUS project (www.globus.org)

Warren-Smith, R. F, & Lawden, M. D. 19xx, HDS – Hierarchical Data System, Starlink Project, (www.starlink.rl.ac.uk/star/docs/sun92/htx/sun92.html)

Warren-Smith, R. F. 19xx, NDF – Routines for accessing the extensible N–Dimensional Data Format, Starlink Project, (www.starlink.rl.ac.uk/star/docs/sun3

Jenness T. J & Economou F. 19xx, STARPERL – Starlink perl modules, Starlink Project, (www.starlink.rl.ac.uk/star/docs/htx/sun228.html)