Running Applications on SURAgrid

(Version 2.0 - March 2007)

SURAgrid is a multi-institutional initiative to develop a multi-purpose grid infrastructure to support the sharing of resources within and between institutions. SURAgrid is both a development and operational environment, where flexibility in deployed processes and technologies supports a growing set of heterogeneous resources, provides interoperability between those resources and future compatibility with emerging national and international grid standards. The SURAgrid user community is also diverse and expected to increase in diversity as both new and traditional grid applications are supported. To view a current list of SURAgrid participants, see http://www.sura.org/programs/sura_grid.html.

This document outlines a "process-in-progress" for running an application on SURAgrid and assumes understanding of the definition and intent of grid technology, as well as some familiarity with the grid middleware involved (e.g., Globus). Application owners should team with their IT support staff if needed to assist with implementation, or through peer support available via the SURAgrid application listserv as described in Step 2 below. If you are not sure if this is the latest version of this document, please check for the latest version on the SURAgrid Web site (http://www.sura.org/suragrid) before proceeding. If you have comments or suggestions for this document, contact sg-implementation@sura.org.

SURAgrid is still under design and deployment, and real and diverse applications are needed to drive technology and policy development. With this, SURAgrid is currently very open to requests from users for access to computation, storage and grid services on SURAgrid resources. This could change in the future, most likely in response to contention for resources, and only with consensus of current SURAgrid participants.

SURAgrid Resources

- SURAgrid's resources are voluntarily contributed and represent a varied set of computing environments that range from small clusters or loosely connected compute nodes to large HPC production clusters.
- The SURAgrid portal lists currently available resources along with their individual system hardware and software details and technical contacts for further information and support: https://gridportal.sura.org/gridsphere/gridsphere
- Resources are managed by local sites and combined to enable SURAgrid users to match their application needs to at least a subset of SURAgrid capability.
- Resources may be dedicated to SURAgrid use or may be shared such that SURAgrid use contends with other existing priorities (e.g., a cluster that has been deployed for a campus grid may have a small queue of nodes available to run SURAgrid jobs). The variety of resources available on SURAgrid makes it difficult to provide a highly managed service.
- If the possibility of job preemption, resource system failure, or other types of outages is of concern to the SURAgrid user, they should work with the application owner to ensure programmatic precautions such as check-pointing are in place.

SURAgrid's Technical Environment

SURAgrid is based on the Globus Toolkit (http://www.globus.org/toolkit) and minimal hardware, software and policy requirements are specified for contributed resources. SURAgrid has not standardized on a single version of the Globus Toolkit so that it can accommodate a high degree of heterogeneity and create an open environment that accommodates the diversity of the SURAgrid community. This may mean that any given application is unable to run on all SURAgrid resources, which is most applicable to the use of pre-Web vs. Web services versions of various Globus services. Specific application requirements can be accommodated through additional documentation and coordination between application owners and owners of the resources where the applications are intended to run (see Step 3 below).

The continually evolving SURAgrid environment includes:

- Operating systems: Linux, SUN, Mac OSX (most SURAgrid resources at this time are running some version of Linux)
- Architectures: clusters, distributed PCs
- Grid middleware: Globus Toolkit (GT) 2.4+ pre-Web services, GT 3.0 and 4.0 pre-Web and Web services; GRAM, GridFTP, RFT, GSI-OpenSSH
- SURAgrid environment requirement and recommendations (available on the SURAgrid Web site at http://www.sura.org/programs/SURAgrid_EnvVar.htm.
- Local schedulers: Not standardized within SURAgrid at this time and a variety are in use - PBS, LSF, Condor; also some use of Globus MDS for metascheduling

Application Deployment

STEP 1: Agreement with SURAgrid Policies

Application use of SURAgrid should be in alignment with the SURAgrid Acceptable Use Policy (AUP), available at: http://www.sura.org/programs/sura_grid_aup.html. Note that SURAgrid users agree to observe the SURAgrid acceptable use policy, and also the acceptable use policies of the individual resources and connections that they are accessing through SURAgrid.

STEP 2: Completion of the SURAgrid Application Description Form

A SURAgrid Application Description form should be submitted for each application that is requested to run on SURAgrid. The form is available at: http://www.sura.org/programs/docs/suragrid_app_template.doc and requests information such as:

- brief description of the nature and impact of the application
- contact information
- resource requirements
- environment requirements (particularly those that are beyond SURAgrid basics)
- estimated frequency and duration of SURAgrid use.

The completed form should be sent to apps-request@sura.org, and provides initial information for the SURAgrid project team to establish contact with those listed as the application contacts, to approve the intended use of SURAgrid resources and to assist in deployment.

After an application form is received (and typically within 2 to 3 business days), a contact from the SURAgrid project team will follow-up to provide additional detail on current SURAgrid policies and next steps for implementation. Following this discussion, the SURAgrid implementation coordinator will be available to track and assist as needed in the remaining steps below.

Once implementation is underway, information from the completed form will also be added to the SURAgrid Web site as an example of an application that is running on SURAgrid. **If you do not want this information published in this way, note this in the Contact(s) field of the form.**

All contacts listed in the application description form are also automatically added to the SURAgrid application listserv (suragrid-apps@sura.org) for ongoing project coordination and peer support. Additional contacts may be added by sending a request to Mary Fran Yafchak maryfran@sura.org.

STEP 3: Documentation of Application-specific Requirements

The application contact should provide additional detailed documentation of any application requirements that extend beyond or are not addressed within the current SURAgrid technical environment described above. This documentation should be available for ongoing reference by resource owners in advance of running on SURAgrid and should evolve as needed in response to lessons learned through implementation on various resources. See the SURAgrid Web site at: http://www.sura.org/programs/sura_grid_apps.html for examples of application-specific resource set-up instructions.

STEP 4: Establish Access to SURAgrid Resources

In order for an application to run on any given SURAgrid resource, the user(s) that will initiate the application must be authenticated (identity verified) and authorized (accounts and permissions established) to use that resource. For authentication, the Globus Toolkit relies on PKI (public key infrastructure) and its related exchange of certificates, while shared authorization in Globus uses the grid-mapfile and associated system accounts.

The SURAgrid user authentication model leverages authoritative campus identity management for scalable exchange of trust information through a Bridge Certificate Authority (Bridge CA). Within SURAgrid, participating sites typically run their own Certificate Authority (CA) to provide both user and system certificates; these certificates are then cross-certified via the Bridge CA (BCA). Cross-certification between participating SURAgrid sites also supports the use of SURAgrid user accounts for accessing SURAgrid resources.

Obtain a User Certificate Issued by a Cross-certified CA

If your site is already a participating SURAgrid site

If your institution is already a participating SURAgrid site, you should contact your local SURAgrid representative to determine if your site has a cross-certified CA in place (the list of participating SURAgrid representatives is located at http://www.sura.org/programs/sura_grid_contacts.html.)

If your institution has a cross-certified CA, your SURAgrid User Administrator will be able to help you obtain the certificate and SURAgrid account you'll need to access SURAgrid resources. However, not all participating SURAgrid sites have a CA they've cross-certified with the SURAgrid BCA. If that is the case with your site and your SURAgrid representatives don't anticipate that changing, please skip to the step below "If you cannot access a CA from your institution".

If your site is not yet a participating SURAgrid site

Cross-certify your site CA with the SURAgrid BCA

The person that controls your institution's CA can obtain basic instructions for cross-certifying a site CA with the SURAgrid BCA from the University of Virginia, lead for development and maintenance of the SURAgrid Bridge CA, at https://www.pki.virginia.edu/sura-bridge/. The page explains how to download and sign the BCA's certificate request and how to generate a certificate request from a CA and send it to the BCA. Assistance with CA matters can be obtained by sending an email message to <sg-implementation@sura.org>.

If you need to create a new CA

If your institution has no CA in place or has one but it's not available for cross-certification with the SURAgrid BCA it is recommended the SURAgrid representatives at the participating institution create a new CA that can be used to cross-certify their site/resource with the SURAgrid BCA. There is an automated tool available (the simpleCA Bundle) from SURAgrid that simplifies the creation of a new CA. Please see instructions item 3 of "Download Software" at: https://www.pki.virginia.edu/sura-bridge/scl/.

If your site has a CA that cannot be cross-certified

If cross-certifying a new or existing institutional CA with the SURAgrid Bridge CA is simply not practical, ***but your institution does have a CA and can issue user and server certificates for use with your SURAgrid work***, mutual trust can still be established for access to SURAgrid resources. This can be done through the direct exchange of certificates and signing policies between your site and the participating SURAgrid sites whose resources you need to access. You should contact the SURAgrid Resource Admin Contact. Contacts are listed in the SURAgrid portal https://gridportal.sura.org/gridsphere/gridsphere?cid=resource-config&JavaScript=enabled (click on the resource name to view contact information details). Once these items have been exchanged, the SURAgrid Resource Admin Contact will assist with obtaining a local account from their site for you use in accessing their particular resource.

If you cannot access a CA from your institution

Application owners that cannot access an existing or create a new institutional CA can obtain access to SURAgrid resources by contacting Nicole Geiger (istnpg@langate.gsu.edu) at GSU (Georgia State University).

Obtain User Accounts

SURAgrid Accounts

Once you've obtained access to a user certificate issued by a cross-certified CA, you need to obtain a SURAgrid account[1]. If your user certificate was issued by a cross-certified CA from your own institution, your site's designated SURAgrid User Administrator can create a SURAgrid account for you. If you've contacted Nicole Geiger at GSU because you were not able to access a CA from your institution (reference above step "If you cannot access a CA from your institution"), Nicole will guide you through the process of obtaining a SURAgrid account.

Your SURAgrid account provides single sign-on access to SURAgrid resources at cross-certified sites. These sites can be identified by left clicking on a resource listed on the SURAgrid portal's Resource Configuration tab (https://gridportal.sura.org/gridsphere/gridsphere?cid=resource-config&JavaScript=enabled). You'll be taken to the detail page for the resource, which will list "SURAgrid" in the "User Account Access Type" field for resources that can be accessed with a SURAgrid account.

Local Accounts

Not all SURAgrid resources can be accessed via SURAgrid accounts as some sites choose not to cross-certify with the SURAgrid BCA or they may be cross-certified but their institutional security policies require that locally created accounts are used to access their SURAgrid resource. These sites can be identified by left clicking on a resource listed on the SURAgrid portal's Resource Configuration tab[2]. The detail page for the resource will list "Local" in the "User Account Access Type" field for resources that cannot be accessed with a SURAgrid account. To obtain a local account to access such SURAgrid resources, send an email request to the address listed in the "Email to request local account" field (this field is just below the "User Account Access Type".)

STEP 5: Job Submission

In order to be able to submit jobs, you need to first have successfully authenticated to the remote SURAgrid resource. An easy way to check if a job can be submitted to a remote resource is to perform a GRAM Ping test with the command:

globusrun –a –r *remote.resource*

If you receive a message with "GRAM Successful", you are ready to proceed with the steps below. Some SURAgrid sites may also provide GSI-SSH capabilities that you can use to login. Once account access has been verified, any of the below listed approaches can be used to set up your application on a SURAgrid resource:

A. Login to the remote resource using GSI-SSH and compile and run the new application on the resource. This will likely require submit scripts be created for submitting the application as a batch job to the resource manager.

[1] A SURAgrid User Administrator is designated by each participating SURAgrid institution that cross-certifies with the SURAgrid BCA. Typically, the SURAgrid Resource Admin Contact is also the User Administrator, but SURAgrid does not require this arrangement.

[2] https://gridportal.sura.org/gridsphere/gridsphere?cid=resource-config&JavaScript=enabled

B. Submit the job remotely from a local resource using Globus GRAM. Note that the local resource must have Globus installed.
C. Submit the job remotely via the portal. Note that this option does not require Globus be installed on the local resource. The SURAgrid Portal uses MyProxy[3] services to manage SURAgrid user credentials and enable them to be authenticated on SURAgrid resources. First, refer to the instructions on the SURAgrid Portal on how to upload and manage MyProxy credentials, submit and manage jobs, and to transfer files using the SURAgrid Portal (see: https://gridportal.sura.org/gridsphere/gridsphere?cid=documentation&JavaScript=enabled

Below are several application tests that may be run by users to become familiar with using SURAgrid or for use in debugging potential problems. Application-specific resource set-up instructions for other SURAgrid applications can provide further insight into situations or problems that might be encountered; these instructions are available at: http://www.sura.org/programs/sura_grid_apps.html.

A. Test Globus local to SURAgrid resource (confirm cert, MDS, GRAM, GridFTP)
   1) SSH to SURAgrid resource
   2) grid-proxy-init
   3) grid-info-search –anonymous –h $host (This is an optional service that may not be available on all SURAgrid resources.)
   4) globus-job-run $host /bin/date
   5) globus-url-copy file://$FILE gsiftp://$host$FILE

B. Test Globus from local to remote SURAgrid resource (confirm cert, MSD, GRAM, GridFTP)
   1) login to your local grid resource
   2) grid-proxy-init
   3) grid-info-search –anonymous –h <remote SURAgrid resource> (This is an optional service that may not be available on all SURAgrid resources.)
   4) globus-job-run <remote SURAgrid resource> /bin/date
   5) globus-url-copy file://$FILE gsiftp://<remote SURAgrid resource>$FILE

C. Test Globus from SURAgrid portal - reference SURAgrid portal documentation at https://gridportal.sura.org/gridsphere/gridsphere?cid=documentation&JavaScript=enabled
   1) login to SURAgrid Portal and obtain a portal_username if you don't yet have one
   2) login to your local grid resource
   3) myproxy-init -s <myproxy_servername> -l <portal_username>
   4) login to SURAgrid portal
   5) use Get New Proxy function in portal
   6) use SURAgrid portal interface to submit test job such as bin/date (select resource, specify standard.out and standard.err, submit job…)

STEP 6: Ongoing Application Use and Support

Peer support as well as troubleshooting assistance from the SURAgrid team is available via the SURAgrid application listserv: SURAgrid-apps@sura.org. The list includes threaded and

[3] http://grid.ncsa.uiuc.edu/myproxy/ and www.globus.org/alliance/publications/papers/myproxy.pdf

searchable archives for access by subscribers (http://www.sura.org/archives/suragrid-apps) and can support subject line queries.

All contacts listed in the application description form are automatically added to this list once implementation is underway. Additional contacts may be added by sending a request to Mary Fran Yafchak maryfran@sura.org.