

# GridFTP to the TeraGrid from non-TeraGrid Hosts

TeraGrid relies on GridFTP for high-throughput file transfers between supercomputer centers. GridFTP is usually installed and configured by TeraGrid system administrators as part of CTSS, and no self-contained GridFTP client is packaged for individual users to download and install for transfers to and from TeraGrid machines.

This document describes how to install the Globus Toolkit on a non-TeraGrid machine and transfer files to and from a TeraGrid host. For example, you can install Globus on your personal workstation, configure it to trust TeraGrid hosts, and then use the `globus-url-copy` command to GridFTP files to a TeraGrid host.

## Software Installation

1. Verify that you need to install the Globus Toolkit

Before you begin, take a moment to verify that the system administrator hasn't already installed the Globus Toolkit. Most official Globus installations create the directory `/etc/grid-security`. Check to see if it exists.

```
$ ls -la /etc/grid-security
```

If this directory exists, there is probably a version of Globus that you can use already installed, and you should ask the administrator for path information or try to find it yourself and then proceed with certificate configuration below.

2. Download and extract the Globus Toolkit

```
$ wget http://www-unix.globus.org/ftppub/gt4/4.0/4.0.5/installers/src/gt4.0.5-all-source-installer.tar.bz2
$ bunzip2 gt4.0.5-all-source-installer.tar.bz2
$ tar xf gt4.0.5-all-source-installer.tar
```

3. Choose an appropriate flavor for installation

In many cases, you can let Globus choose the default flavor and omit the `--with-flavor` argument from Globus configuration commands. However, if you are using an IBM Power 64-bit system, you may wish to use the `gcc32dbg` flavor to avoid known issues on Power platforms.

- For Intel 32-bit platforms, use `gcc32dbg`
- For Intel and AMD 64-bit platforms, use `gcc64dbg`
- For IBM Power systems, both 32-bit and 64-bit, use `gcc32dbg`

4. Install the Globus data management client in your home directory

```
$ export GLOBUS_LOCATION=~/globus-4.0.5
$ cd gt4.0.5-all-source-installer
$ ./configure --prefix=$GLOBUS_LOCATION \
  --disable-wsjava --disable-wsmads --disable-wsdel --disable-wsrf \
  --disable-wsgram --disable-drs --disable-prewsgram --disable-rendezvous \
  --disable-wscas --disable-wsc --disable-tests --disable-wstests \
  --disable-gsiopenssh --disable-webmads --with-flavor=FLAVOR
$ make globus-data-management-client
$ make install
```

5. Install the Globus data management client SDK in your home directory so you can compile other GridFTP tools later

```
$ make globus-data-management-sdk
$ make install
```

6. Remove the downloaded and extracted files

```
$ rm -rf gt4.0.5-all-source-installer
$ rm gt4.0.5-all-source-installer.tar
```

## Certificate Configuration

In order for GSI authentication to work, bidirectional certificate verification must be able to complete successfully. For a GridFTP transfer, the client verifies the authenticity of the target machine, and then the remote site verifies the authenticity of the user, by examining X509 certificates placed in specific locations. In order for the new client to trust the TeraGrid, a collection of certificates must be installed.

## Certificate Authority Certificates

To make your client trust TeraGrid machines, you need to install the TeraGrid certificate authority files. TeraGrid sites use a process known as `gx-map` to do this automatically. For an independent client, you can simply copy the required files to your system.

The easiest way to do this is to retrieve a copy of all the trusted certificate authorities on the target machine. You should already have an account there, so by retrieving its trusted certificate authorities, you guarantee that you will be able to connect using your personal certificate.

For example, to retrieve the certificates from NCAR's Frost system:

```
$ mkdir ~/.globus/certificates
$ scp username@fr0103ge.ncar.teragrid.org:/etc/grid-security/certificates/* ~/.globus/certificates/
```

Files that should be retrieved include:

|                  |                                    |
|------------------|------------------------------------|
| *.0              | Certificate Authority certificates |
| *.r0             | Certificate revocation lists       |
| *.signing_policy | Textual policy summaries           |



#### Important Note

Certificate revocation lists (CRLs, the \*.r0 files) expire after about one month. Once a CRL expires, the certificate can no longer be used. You will need to re-retrieve the \*.r0 files about every two weeks for uninterrupted operation. You can also just delete the \*.r0 files, but that is not recommended for security considerations.

## User Certificate

Next, you need to install your user certificate. You can generate a certificate at several TeraGrid sites. Simply copy your user certificate files `usercert.pem` and `userkey.pem` into the `~/.globus` directory.

## Using GridFTP

When you start a new shell, you first need to configure your environment.

1. Set the environment variable to constrain file transfers to the open port range:

```
$ export GLOBUS_TCP_PORT_RANGE=50000,51000
```

2. Set the pointer to the Globus location and source the environment:

```
$ export GLOBUS_LOCATION=~/.globus-4.0.5
$ source $GLOBUS_LOCATION/etc/globus-user-env.sh
```

Of course, these may be added to your shell startup scripts. Note that there is also a `globus-user-env.csh` script for `tcsh` users.

Next, initialize your proxy certificate:

```
$ grid-proxy-init
```

## Transferring Files using globus-url-copy

The easiest way to transfer files is using the `globus-url-copy` command.

Test the connection's throughput by transferring nothing to nowhere using a 4MB TCP buffer and 4 parallel streams. For example, the following transfer sends data to Frost:

```
$ globus-url-copy -vb -tcp-bs 4096KB -p 4 file:///dev/zero gsiftp://fr0103ge.ncar.teragrid.org/dev/null
```

Once this test completes, you can transfer files by substituting names as appropriate. Remember that you can reverse the order of the file and `gsiftp` URLs to download files from the remote site as well.

Make sure to test file transfers in both directions – depending on the firewall configuration, you may be able to upload files to the server, but not download files. If that is the case, consider using `UberFTP` instead.

## Transferring Files using UberFTP

If the firewall on the client doesn't permit access on ports 50000:51000, then you might try UberFTP instead. Follow the [Simple UberFTP Installation Instructions](#). UberFTP provides a standard FTP command-line interface (don't forget to initialize the environment first):

```
$ ./uberftp fr0103ge.ncar.teragrid.org
220 fr0103ge.ncar.teragrid.org GridFTP Server 2.5 (gcc32dbg, 1182369948-63) ready.
230 User mattheww logged in.
uberftp>
uberftp> get <remote file> <local file>
uberftp> put <local file> <remote file>
```

Troubleshooting

grid-proxy-init Stalls

If grid-proxy-init stalls while generating pages of dots (periods), but never completes, chances are you're running on a Power system compiled with the gcc64dbg flavor. Recompile the Globus Toolkit using gcc32dbg.

Certificate Problems

- Remember, for this to work, bidirectional certificate checking must occur. Some errors that are really certificate problems don't look like them. For example:

```
error: globus_ftp_client: the server responded with an error
500 500-Command failed. : globus_xio: An end of file occurred
500 End.
```

This can be solved by adding the -nodcau option to globus-url-copy. However, it's underlying root cause is a missing certificate in the local ~/.globus/certificates directory.

globus-url-copy stalls

If globus-url-copy stalls indefinitely for downloads, it is probably a firewall issue on the client. Consider UberFTP instead, which supports FTP passive (PASV) mode.

Performance

Using the "transfer nothing to nowhere" test, we've received the following transfer rates to Frost:

| Client     | Upload to Frost | Download from Frost |
|------------|-----------------|---------------------|
| Hemisphere | 36-51 MB/s      | 36 MB/s             |