Delivering Research Data Management Services with Globus and the Science DMZ: Endpoint Configuration and Deployment

Steve Tuecke
Vas Vasiliadis
Raj Kettimuthu
Presentations and other useful information available at globus.org/events/sc14/tutorial goo.gl/s2Ew50
Agenda

- Globus Connect Server overview (5 min)
- Demonstration and exercise 5: Installing Globus Connect Server (25 min)
- Exercise 6: Configuring Globus Connect Server (15 min)
- Common Globus Connect Server configurations (15 min)
- Advanced endpoint configuration (10 min)
- Deployment best practice: Science DMZ (10 min)
- Wrap-up and general Q&A (10 min)
Globus Connect Server
Overview
Globus Connect Server
for resource providers

Connect your storage to Globus

Provide an integrated user and admin experience

Reduce your support burden
Globus Connect Server

- Create endpoint in minutes; no complex software install
- Enable all users with local accounts to transfer files
- Native packages: RPMs and DEBs

Local Storage System
(HPC cluster, campus server, …)
What we are going to do:

1. Install Globus Connect Server
   - Access server as user “clusteradmin”
   - Update repo
   - Install package
   - Setup Globus Connect Server

2. Log into Globus (using Globus username)

3. Access the newly created endpoint (as user ‘researcher’)

4. Transfer a file
Globus Connect Server Demonstration
Exercise 5: Set up a Globus Connect Server endpoint and transfer files

• Goal for this session: turn a storage resource into a Globus endpoint
• Each of you is provided with an Amazon EC2 server for this tutorial
• Step 1: Create a Globus account (if you do not have one already)
Step 2: Log into your host

- Your slip of paper has the host information

- Log in as user ‘clusteradmin’:
  ssh clusteradmin@ec2-x-x-x-x-x.compute-1.amazonaws.com

- Use the password on the slip of paper

- ‘clusteradmin’ has passwordless sudo privileges

- NB: Please sudo su before continuing
Step 3: Install Globus Connect Server

‘Cheat sheet’: globus.org/events/sc14/tutorial

```
$ sudo su
$ curl -LOs http://toolkit.globus.org/ftppub/globus-connect-server/globus-connect-server-repo_latest_all.deb
$ dpkg -i globus-connect-server-repo_latest_all.deb
$ apt-get update
$ apt-get -y install globus-connect-server
$ globus-connect-server-setup
```

You have a working Globus endpoint!
Step 4: Access your Globus endpoint

- Go to globus.org; login with your Globus account
- Go to Manage Data → Transfer Files
- Access the endpoint you just created
  - Enter: `<your-Globus-username>#ec2-…` in Endpoint field
  - Log in as user `researcher`; you should see the user’s home directory
- Transfer files
  - Between `go#ep1` and your endpoint (`ec2-nnn-…`)
  - From `esnet#anl-diskpt1/data1` to your endpoint
Exercise 6: Configuring Globus Connect Server

• **Globus Connect Server configuration is stored in:**
  - `/etc/globus-connect-server.conf`

• **To enable configuration changes you must run:**
  - `globus-connect-server-setup`

• “Rinse and repeat”

• **NB: Please `sudo su` before continuing**
Configuration file walkthrough

- **Structure based on .ini format:**
  ```
  [Section]
  Option
  ```

- **Most common options to configure**
  - Hostname
  - Public
  - RestrictedPaths
  - Sharing
  - SharingRestrictedPaths
  - IdentityMethod (CILogon, Oauth)
Basic Configuration

• **Change your endpoint’s name in the Globus Connect Server configuration file:**
  – *Edit* `/etc/globus-connect-server.conf`
  – *Set* `[Endpoint] Name = “dtn”`

• **Run: globus-connect-server-setup**
  – Enter your Globus username and password when prompted

• **Access the endpoint in your browser using the new endpoint name**
Common Globus Connect Server Configurations
Firewall configuration

• **Allow inbound connections to port**
  – 2811 (GridFTP control channel)
  – 7512 (MyProxy CA) or 443 (OAuth)

• **Allow inbound connections to ports 50000-51000 (GridFTP data channel)**
  – If transfers to/from this machine will happen only from/to a known set of endpoints (not common), you can restrict connections to this port range only from those machines

• **If your firewall restricts outbound connections**
  – Allow outbound connections if the source port is in the range 50000-51000
Using MyProxy OAuth server

• **Web-based endpoint activation**
  – Sites run a MyProxy OAuth server
    o MyProxy OAuth server in Globus Connect Server
  – Users enter username/password only on site’s webpage to access an endpoint
  – Globus gets short-term X.509 credential via OAuth protocol

• **MyProxy without OAuth (we just did this!)**
  – Site passwords flow through Globus to site MyProxy server
  – Globus does not store passwords
  – Still a security concern for some sites
Making your endpoint public

• Try to access the endpoint created by the person sitting next to you
• You will get the following message:
• ‘Could not find endpoint with name ‘dtn’ owned by user ‘<neighbor’s username>’

Making your endpoint public

• Edit: `/etc/globus-connect-server.conf`
• Uncomment `[Endpoint] Public` option
• Replace ‘False’ with ‘True’
• Run `globus-connect-server-server-setup`
• Try accessing your neighbor’s endpoint: you will be prompted for credentials…
• …but you cannot access it, since you do not have an account on that server
Enable sharing on your endpoint

- Edit: `/etc/globus-connect-server.conf`
- Uncomment `[GridFTP] Sharing = True`
- Run `globus-connect-server-server-setup`
- Go to the Web UI Start Transfer page
- Select the endpoint
- Create shared endpoints and grant access to other Globus users

Note: Creation of shared endpoints requires a **Globus Provider plan** for the managed endpoint
Advanced Endpoint Configuration
Select configuration scenarios

• Customizing filesystem access
• Using host certificates
• Using CILogon certificates
• Enabling sharing on GT GridFTP server
• Configuring multiple GridFTP servers
• Setting up an anonymous endpoint
Path Restriction

• **Default configuration:**
  – All paths allowed, access control handled by the OS

• **Use `RestrictPaths` to customize**
  – Specifies a comma separated list of full paths that clients may access
  – Each path may be prefixed by R (read) and/or W (write), or N (none) to explicitly deny access to a path
  – ‘~’ for authenticated user’s home directory, and * may be used for simple wildcard matching.

• **E.g. Full access to home directory, read access to /data:**
  – `RestrictPaths = RW~,R/data`

• **E.g. Full access to home directory, deny hidden files:**
  – `RestrictPaths = RW~,N~/.*`
Sharing Path Restriction

• Further restrict the paths on which your users are allowed to create shared endpoints

• Use SharingRestrictPaths to customize
  – Same syntax as RestrictPaths

• E.g. Full access to home directory, deny hidden files:
  – SharingRestrictPaths = RW~,N~/.*

• E.g. Full access to public folder under home directory:
  – SharingRestrictPaths = RW~/public

• E.g. Full access to /proj, read access to /scratch:
  – SharingRestrictPaths = RW/proj,R/scratch
Control sharing access to specific accounts

- **SharingStateDir** can be used to control sharing access to individual accounts

- For instance, with
  ```
  SharingStateDir = "'/var/globus/sharing/$USER"
  ```
  user "bob" would be enabled for sharing only if a path exists with the name "'/var/globus/sharing/bob/" and is writable by bob.
Using a host certificate for GridFTP

- You can use your GridFTP server with non-Globus clients
  - Requires a host certificate, e.g. from OSG
- Comment out
  - `FetchCredentialFromRelay = True`
- Set
  - `CertificateFile = <path_to_host_certificate>`
  - `KeyFile = <path_to_private_key_associated_with_host_certificate>`
  - `TrustedCertificateDirectory = <path_to_trust_roots>`
Single Sign-On with InCommon/CILogon

- **Requirements**
  - Your organization’s Shibboleth server must release the ePPN attribute to CILogon
  - Your local resource account names must match your institutional identity (InCommon ID)

- **Set AuthorizationMethod = CILogon** in the Globus Connect Server configuration

- **Set CILogonIdentityProvider = <your_institution_as_listed_in_CILogon_identity_provider_list>**

- **Add CILogon CA to your trustroots**
  - `/var/lib/globus-connect-server/grid-security/certificates/
  - Visit ca.cilogon.org/downloads for certificates
Enabling Sharing on a GT GridFTP Installation

- Add to your trusted certificates directory (/etc/grid-security/certificates)
- Use '-sharing-dn' option in the server as follows: globus-gridftp-server -sharing-dn "/C=US/O=Globus Consortium/OU=Globus Connect User/CN=__transfer__"
- Use '-sharing-rp' option to restrict the file paths allowed for sharing: globus-gridftp-server -sharing-rp <path>
Deployment Scenarios

• **Globus Connect Server components**
  – globus-connect-server-io, -id, -web

• **Default:** -io and –id (no –web) on single server

• **Common options**
  – Multiple –io servers for load balancing, failover, and performance
  – No -id server, e.g. third-party IdP such as CILogon
  – -id on separate server, e.g. non-DTN nodes
  – -web on either –id server or separate server for OAuth interface
Setting up multiple –io servers

• Guidelines
  – Use the same .conf file on all servers
  – First install on the server running the –id component, then all others

1. Install Globus Connect Server on all servers
2. Edit .conf file on one of the servers and set [MyProxy] Server to the hostname of the server you want the –id component installed on
3. Copy the configuration file to all servers
   – /etc/globus-connect-server.conf
4. Run globus-connect-server-setup on the server running the –id component
5. Run globus-connect-server-setup on all other servers
6. Repeat steps 2-5 as necessary to update configurations
Deployment Best Practice: Science DMZ
Researchers don’t realize full benefits of existing IT infrastructure

- Impedance mismatch between research computing systems and the WAN
- Network “misconfiguration” (10 x 1Gb/s links ≠ 1 x 10Gb/s link)
- Indiscriminate security policies
- TCP: small amount of packet loss = huge difference in performance
Throughput requirements (and expectations!)

<table>
<thead>
<tr>
<th>Data set size</th>
<th>1 Minute</th>
<th>5 Minutes</th>
<th>20 Minutes</th>
<th>1 Hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>10PB</td>
<td>1,333.33 Tbps</td>
<td>266.67 Tbps</td>
<td>66.67 Tbps</td>
<td>22.22 Tbps</td>
</tr>
<tr>
<td>1PB</td>
<td>133.33 Tbps</td>
<td>26.67 Tbps</td>
<td>6.67 Tbps</td>
<td>2.22 Tbps</td>
</tr>
<tr>
<td>100TB</td>
<td>13.33 Tbps</td>
<td>2.67 Tbps</td>
<td>666.67 Gbps</td>
<td>222.22 Gbps</td>
</tr>
<tr>
<td>10TB</td>
<td>1.33 Tbps</td>
<td>266.67 Gbps</td>
<td>66.67 Gbps</td>
<td>22.22 Gbps</td>
</tr>
<tr>
<td>1TB</td>
<td>133.33 Gbps</td>
<td>26.67 Gbps</td>
<td>6.67 Gbps</td>
<td>2.22 Gbps</td>
</tr>
<tr>
<td>100GB</td>
<td>13.33 Gbps</td>
<td>2.67 Gbps</td>
<td>666.67 Mbps</td>
<td>222.22 Mbps</td>
</tr>
<tr>
<td>10GB</td>
<td>1.33 Gbps</td>
<td>266.67 Mbps</td>
<td>66.67 Mbps</td>
<td>22.22 Mbps</td>
</tr>
<tr>
<td>1GB</td>
<td>133.33 Mbps</td>
<td>26.67 Mbps</td>
<td>6.67 Mbps</td>
<td>2.22 Mbps</td>
</tr>
<tr>
<td>100MB</td>
<td>13.33 Mbps</td>
<td>2.67 Mbps</td>
<td>0.67 Mbps</td>
<td>0.22 Mbps</td>
</tr>
</tbody>
</table>

# in green → should be easily achievable on any US campus
# in black → should be achievable on most US campuses
# in blue → should be achievable on a highly tuned system
# in orange → stay tuned!
Science DMZ Components

- “Friction free” network path
- Dedicated, high-performance data transfer nodes (DTNs)
- Performance measurement/test node
- User engagement and education

LOTS of great info available at: fasterdata.es.net/science-dmz
Typical deployment

Science DMZ + Globus

Details at: fasterdata.es.net
Network paths
Enable your resource

- Signup: globus.org/signup
- Enable your resource: globus.org/globus-connect-server
- Need help? support.globus.org
- Follow us: @globusonline