Enabling Your Campus to Simplify Research Data Management with Globus Online

Steve Tuecke and Raj Kettimuthu
Computation Institute
University of Chicago and Argonne National Laboratory
The goal of this session is to show you (hands-on) how to take a resource and turn it into a GO endpoint.

Each of you is provided with an amazon EC2 machine for this tutorial.
Log into your host

Your slip of paper has the host information.

Log in as user “xsede12”:

```
ssh xsede12@ec2-xx-xx-xx-xx-xx.compute-1.amazonaws.com
```

Use the password on the slip of paper.

*xsede12* has passwordless sudo privileges.
Setup SSH keys

Upload xsede12's SSH key to Globus Online (/home/xsede12/.ssh/id_rsa.pub).

Each host also has five users: joe, bob, sue, ann, and sam.

Password for joe, bob, sue, ann: the same password as the “xsede12” user.

Password for sam is “xsede12tutorial” for all hosts.
Globus Online Endpoint Setup with Globus Connect Multi-User

XSEDE 2012
July 16, 2012
Globus Connect Multi-User

• **What is GCMU?**
  - Multi-user version of Globus Connect
  - Packages a GridFTP server and MyProxy CA, pre-configured for use with Globus Online

• **Why GCMU?**
  - Create transfer endpoints in minutes
  - Avoid complex GridFTP install

• **To download:** [https://www.globusonline.org/gcmu/](https://www.globusonline.org/gcmu/)

“We used GCMU to form a campus-wide GSI authentication service spanning multiple servers. Now my users have a fast, easy way to get their data wherever it needs to go, and the setup process was trivial.”

--University of Michigan

“As a resource admin, I've found GCMU an exceedingly useful tool.... With GCMU, setting up a GridFTP server and handling authentication for multiple users is easy.”

--Oak Ridge National Lab
GCMU Deployments
Globus Connect Multi-User

The diagram illustrates the flow of data and authentication processes within the GCMU (Globus Connect Multi-User) system. It consists of several labeled components, including:

- **Client**: Initiates the transfer request.
- **MyProxy Online CA**: Handles certificate and password requests.
- **GridFTP Server**: Processes transfer requests and authorizations.
- **Local authentication & authorization system**: Utilizes LDAP, RADIUS, NIS, etc.

The flow begins with the Client sending a transfer request, which is handled by the MyProxy Online CA. This component requests a certificate and a password, along with a TLS handshake. The GridFTP Server then processes the request, followed by authorization from the local authentication system. The diagram outlines the step-by-step process from the client to the local authentication system, ensuring secure and authorized data transfer.
cd /opt


sudo tar xzf globusconnect-multiuser-latest.tgz

cd gcmu*

sudo ./install
Try doing the following

Create a file called tutorial.txt in /home/joe

Go to the GO Web UI -> Start Transfer

Select endpoint *username#xsede12*

Activate the endpoint as user “joe” (not xsede12). You should see joe's home directory. (Remember: joe's password is the same as xsede12's)

Transfer to/from the endpoint of the person sitting next to you (activate their endpoint as user “sam”).

Does tutorial.txt show up in /home/sam in the host of the person sitting next to you?
Setup SSH keys

Upload xsede12's SSH key to Globus Online (/home/xsede12/.ssh/id_rsa.pub).

Each host also has five users: joe, bob, sue, ann, and sam.

Password for joe, bob, sue, ann: the same password as the “xsede12” user.

Password for sam is “xsede12tutorial” for all hosts.
GridFTP

- Two channel protocol like FTP
- Control Channel
  - Command/Response
  - Used to establish data channels
  - Basic file system operations eg. mkdir, delete etc
- Data channel
  - Pathway over which file is transferred
  - Many different underlying protocols can be used
    - MODE command determines the protocol
Third Party Transfer

• Client initiates data transfer between 2 servers
• Client forms control channel with 2 servers.
• Information is routed through the client to establish data channel between the two servers.
• Data flows directly between servers
  – Client is notified by each server when the transfer is complete
GridFTP Authentication and Authorization

Client

Authenticate

f

Daemon

setuid

Port 2811

ROOT

USER
Security Configuration

- **Installation**
  - Download, untar, configure, make

- **Security configuration (server admins)**
  - Obtain and install X.509 host certificate from well-known CA
  - Configure trust roots

- **Security configuration (users)**
  - Obtain and install user certificate from well-known CA
  - Configure trust roots

- **Setup authorization (both users and admins)**
  - DN to local username mapping in gridmap file
  - '/DC=org/DC=doegrids/OU=People/CN=Rajkumar Kettimuthu 227852' rajk

- **Too complex for many users and small labs**
• Make GridFTP deployment trivial
  GridFTP transfers can be achieved “instantly” even by non-experts

• **Automate the process of configuring security**
  Avoid the need for any end-user or system administrator involvement in security configuration

• **Reduce burden on both users and administrators**
  Eliminate frequent sources of errors in GridFTP configuration and use.
Globus Connect Multi-User

Diagram:

1. Client
2. MyProxy Online CA
3. GridFTP Server
4. GCMU
5. Local authentication & authorization system (LDAP, RADIUS, NIS etc.)

Steps:
- Step 1: Certificate, Password
- Step 2: Password
- Step 3: Certificate, TLS handshake
- Step 4: Transfer request, Certificate
- Step 5: Authorization
Username is embedded in the certificate, as this cert is used to authenticate with this site only

"/C=US/O=Globus Consortium/OU=Globus Connect Service/CN=84591482-cba0-11e1-b791-1231381b68a7/CN=rajk"

Authorization callout extracts username from DN, if the certificate is issued by local CA
Data Channel Establishment

Client

GridFTP Server A

GridFTP Server B

Connect
IP:PORT

PORT <IP:PORT>

AUTH

PASV

227 <IP:PORT>
Data Channel Authentication

Client

GridFTP Server A

GridFTP Server B

Cred A

Cred B

Delegated Cred A

Delegated Cred B

Control Channel

Data Channel
Data Channel Authentication

Client

Cred A

GridFTP Server A

Delegated Cred A

Has to trust Cred B’s CA

Cred B

GridFTP Server B

Delegated Cred B

Has to trust Cred A’s CA

Control Channel

Data Channel
Data Channel Authentication

Client

GridFTP Server A

GridFTP Server B

Cred A

Cred B

Delegated Cred A

Delegated Cred B

Control Channel

Data Channel

Does not trust Cred B’s CA

Does not trust Cred A’s CA
Data Channel Authentication

Client

GridFTP Server A

GridFTP Server B

Cred A

Shared secret

Cred B

Shared secret

Control Channel

Data Channel
Data Channel Security Context (DCSC)

Client

GridFTP Server A

- Cred A
- Delegated Cred A

GridFTP Server B

- Cred B
- Delegated Cred A
- Supports DCSC

DCSC support not required

Send & accept Cred A for data channel security

Control Channel

Data Channel
• **Site passwords flow through Globus Online**
  • Globus Online does not store passwords
  • Just pass along to MyProxy servers at site
  • Still a security concern for some sites

• **OAuth**
  • Sites run an OAuth server
  • Users enter username and password only on a site’s webpage
  • GO gets an X.509 credential via OAuth protocol
cd /opt

      globusconnect-multiuser-1.2.0rc2.tgz

  . sudo tar xzf globusconnect-multiuser-1.2.0rc1.tgz

cd gcmu*

Sudo apt-get install python-flask

sudo ./install
• GCMU may require firewall configuration
  • Inbound ports: 2811, 7512, and 50000-51000
  • Outbound ports: 50000-51000
• GCMU GridFTP cannot be (easily) used with other GridFTP clients besides GO
  • GCMU uses a GO-issued cert, not a host cert
• GridFTP has many configuration options
  • E.g., Gridmap file can be used to allow other certs
  • Limit access to particular directories
• MyProxy CA default proxy lifetime is 12 hours
Single GridFTP server for different sets of users

- GCMU allows you to serve different sets of users with a single GridFTP server.
- For example, some XSEDE resource providers want to setup a GridFTP server that caters to both XSEDE and non-XSEDE users.
  - Setup a GCMU on top of XSEDE security infrastructure they already have.
  - Separate GO endpoints e.g., xsede#ranger associated with xsede myproxy and tacc#ranger associated with local myproxy.
  - Non-xsede users pick tacc#ranger and enter their tacc credentials to access it.
The Challenge: Moving Big Data Easily

- What should be trivial...

  "I need my data over there – at my _____" (supercomputing center, campus server, etc.)

- ... can be painfully tedious and time-consuming

  "GAAAH! %&@#&"

  Data Source  ➔  Data Destination
  - Config issues
  - Firewall issues
  - Unexpected failure = manual retry
What is Globus Online?

• **Reliable file transfer.**
  – Easy “fire-and-forget” transfers
  – Automatic fault recovery
  – High performance
  – Across multiple security domains

• **No IT required.**
  – Software as a Service (SaaS)
    • No client software installation
    • New features automatically available
  – Consolidated support & troubleshooting
  – Works with existing GridFTP servers
  – Globus Connect solves “last mile problem”

• **Recommended by XSEDE, Blue Waters, NERSC, ALCF, Advanced Photon Source, many Universities**
Globus Online Demo
Interactive login to command line interface:

```
$ ssh Tuecke@cli.globusonline.org
```

Running commands remotely:

```
$ ssh Tuecke@cli.globusonline.org <command>
```

```
$ ssh Tuecke@cli.globusonline.org scp -r -s 3 -D \ 
  nersc#dtn:~/myfile* mylaptop:~/projects/p1
```

Task ID: 4a3c471e-edef-11df-aa30-1231350018b1

```
$ _
```

Using CLI with gsissh:

```
$ gsissh Tuecke@cli.globusonline.org <command>
```
Globus Storage: For when you want to …

- **Place** your data where you want
- **Access** it from anywhere via different protocols
- **Update** it, **version** it, and take **snapshots**
- **Share** versions with who you want
- **Synchronize** among locations
• Integrate with the Globus research cloud ecosystem
• Write programs that access/manage:
  – user identities, profiles, groups, resources
  – data, compute and collaboration

... via REST APIs and command line programs
• Visit [https://www.globusonline.org/signup](https://www.globusonline.org/signup) to:
  – Get a free account and start moving files

• Visit [www.globusonline.org](http://www.globusonline.org) for:
  – Tutorials, FAQs, Pro Tips, Troubleshooting
  – Papers
  – Case Studies

• Contact [support@globusonline.org](mailto:support@globusonline.org) for:
  – Help getting started
  – Help using the service

• Follow us at [@globusonline](https://twitter.com/globusonline) on Twitter and [Globus Online](https://facebook.com/globusonline) on Facebook
The goal of this session is to show you (hands-on) how to take a resource and turn it into a GO endpoint.

Each of you is provided with an amazon EC2 machine for this tutorial.
Log into your host

Your slip of paper has the host information.

Log in as user “xsede12”:

```
ssh xsede12@ec2-xx-xx-xx-xx-xx.compute-1.amazonaws.com
```

Use the password on the slip of paper.

xsede12 has passwordless sudo privileges.
Setup SSH keys

Upload `xsede12`'s SSH key to Globus Online (`/home/xsede12/.ssh/id_rsa.pub`).

Each host also has five users: Joe, Bob, Sue, Ann, and Sam.

Password for Joe, Bob, Sue, Ann: the same password as the “`xsede12`” user.

Password for Sam is “`xsede12tutorial`” for all hosts.
Globus Connect Multi-User

• **What is GCMU?**
  - Multi-user version of Globus Connect
  - Packages a GridFTP server and MyProxy CA, pre-configured for use with Globus Online

• **Why GCMU?**
  - Create transfer endpoints in minutes
  - Avoid complex GridFTP install

• **To download:** [https://www.globusonline.org/gcmu/](https://www.globusonline.org/gcmu/)

“We used GCMU to form a campus-wide GSI authentication service spanning multiple servers. Now my users have a fast, easy way to get their data wherever it needs to go, and the setup process was trivial.”
--University of Michigan

“As a resource admin, I've found GCMU an exceedingly useful tool.... With GCMU, setting up a GridFTP server and handling authentication for multiple users is easy.”
--Oak Ridge National Lab
Globus Connect Multi-User

**Diagram:**
- **Client** initiates the process.
- **Certificate** and **password** are transferred.
- **GridFTP Server** and **MyProxy Online CA** process the transfer request.
- **PAM** uses the **password** for authentication.
- **Local authentication & authorization system** (LDAP, RADIUS, NIS, etc.) authorizes the transfer.

**Steps:**
1. **Step 1:** Client provides certificate and password.
2. **Step 2:** PAM uses the password for authentication.
3. **Step 3:** MyProxy Online CA processes the request.
4. **Step 4:** Certificate is transferred to the GridFTP Server.
5. **Step 5:** Authorization is granted to the transfer request.
cd /opt


sudo tar xzf globusconnect-multiuser-latest.tgz

cd gcmu*

sudo ./install
Try doing the following

Create a file called tutorial.txt in /home/joe

Go to the GO Web UI -> Start Transfer

Select endpoint username#xsede12

Activate the endpoint as user “joe” (not xsede12). You should see joe's home directory. (Remember: joe's password is the same as xsede12's)

Transfer to/from the endpoint of the person sitting next to you (activate their endpoint as user “sam”).

Does tutorial.txt show up in /home/sam in the host of the person sitting next to you?
Upload `xsede12's` SSH key to Globus Online (`/home/xsede12/.ssh/id_rsa.pub`).

Each host also has five users: joe, bob, sue, ann, and sam.

Password for joe, bob, sue, ann: the same password as the “`xsede12`” user.

Password for sam is “`xsede12tutorial`” for all hosts.
GridFTP

• Two channel protocol like FTP

• Control Channel
  – Command/Response
  – Used to establish data channels
  – Basic file system operations eg. mkdir, delete etc

• Data channel
  – Pathway over which file is transferred
  – Many different underlying protocols can be used
    • MODE command determines the protocol
• Client initiates data transfer between 2 servers
• Client forms control channel with 2 servers.
• Information is routed through the client to establish data channel between the two servers.
• Data flows directly between servers
  – Client is notified by each server when the transfer is complete
GridFTP Authentication and Authorization

Client

Authenticate

Daemon

inetd

Port 2811

ROOT

setuid

USER

fork
Security Configuration

- **Installation**
  - Download, untar, configure, make

- **Security configuration (server admins)**
  - Obtain and install X.509 host certificate from well-known CA
  - Configure trust roots

- **Security configuration (users)**
  - Obtain and install user certificate from well-known CA
  - Configure trust roots

- **Setup authorization (both users and admins)**
  - DN to local username mapping in gridmap file
  - '/DC=org/DC=doegrids/OU=People/CN=Rajkumar Kettimuthu 227852' rajk

- **Too complex for many users and small labs**
GCMU makes it trivial

• Make GridFTP deployment trivial
  • GridFTP transfers can be achieved “instantly” even by non-experts

• Automate the process of configuring security
  • Avoid the need for any end-user or system administrator involvement in security configuration

• Reduce burden on both users and administrators
  • Eliminate frequent sources of errors in GridFTP configuration and use.
Globus Connect Multi-User

Diagram:

- Client
  - Step 1: certificate, password, TLS handshake
  - Step 2: password
  - Step 3: certificate, MyProxy Online CA
  - Step 4: certificate
  - Step 5: Authorization

- GridFTP Server
  - Authz

- Local authentication & authorization system (LDAP, RADIUS, NIS etc.)
• Username is embedded in the certificate, as this cert is used to authenticate with this site only
• “/C=US/O=Globus Consortium/OU=Globus Connect Service/CN=84591482-cba0-11e1-b791-1231381b68a7/CN=rajk”
• Authorization callout extracts username from DN, if the certificate is issued by local CA
Data Channel Establishment

GridFTP Server A

GridFTP Server B

Client

Connect
IP:PORT
Data Channel Authentication

- **Client**
- **GridFTP Server A**
  - Delegated Cred A
- **GridFTP Server B**
  - Delegated Cred B

Channels:
- **Control Channel**
- **Data Channel**
Data Channel Authentication

- **Client**
  - Cred A
  - Cred B

- **GridFTP Server A**
  - Has to trust Cred B’s CA
  - Delegated Cred A
  - Delegated Cred B

- **GridFTP Server B**
  - Has to trust Cred A’s CA

- Control Channel
- Data Channel
Data Channel Authentication

Client

GridFTP Server A
- Cred A
- Delegated Cred A

GridFTP Server B
- Cred B
- Delegated Cred B

Control Channel

Data Channel

Does not trust Cred B’s CA

Does not trust Cred A’s CA
Data Channel Authentication

Client

GridFTP Server A

GridFTP Server B

Cred A

Shared secret

Cred B

Shared secret

Shared secret

Control Channel

Data Channel
Data Channel Security Context (DCSC)

- **Client**
  - Cred B
  - DCSC: Send & accept Cred A for data channel security

- **GridFTP Server A**
  - Cred A
  - Delegated Cred A
  - DCSC support not required

- **GridFTP Server B**
  - Delegated Cred A
  - Supports DCSC

- **Control Channel**
- **Data Channel**
GO / GCMU Interaction

Step 1: Access Endpoint
Step 2: Username/Password
Step 3: TLS Handshake
Step 4: MyProxy Online CA
Step 5: Certificate
Step 6: Transfer Request
Step 7: GridFTP Server
Step 8: Access Files
Step 9: Authentication & Data Transfer

Campus Cluster
- GCMU
- GridFTP Server
- Remote Cluster/User’s PC
- Local Authentication System (LDAP, RADIUS, Kerberos etc)
- Local Storage
Site passwords flow through Globus Online
- Globus Online does not store passwords
- Just pass along to MyProxy servers at site
- Still a security concern for some sites

OAuth
- Sites run an OAuth server
- Users enter username and password only on a site’s webpage
- GO gets an X.509 credential via OAuth protocol
cd /opt

- sudo wget http://connect.globusonline.org/linux/stable/globusconnect-multiuser-1.2.0rc2.tgz

- sudo tar xzf globusconnect-multiuser-1.2.0rc1.tgz

cd gcmu*

Sudo apt-get install python-flask

sudo ./install
Globus Connect Multi User

1. Access Endpoint
2. Redirect
3. Username, password
4. Oauth Server
5. GCMU
6. Username, password
7. MyProxy Online CA
8. Transfer request
9. GridFTP Server
10. Certificate
11. GridFTP Server

Remote Cluster / User’s PC

Campus Cluster

Local Authentication System (LDAP, RADIUS, Kerberos etc)

Local Storage
• **GCMU may require firewall configuration**
  • Inbound ports: 2811, 7512, and 50000-51000
  • Outbound ports: 50000-51000
• **GCMU GridFTP cannot be (easily) used with other GridFTP clients besides GO**
  • GCMU uses a GO-issued cert, not a host cert
• **GridFTP has many configuration options**
  • E.g., Gridmap file can be used to allow other certs
  • Limit access to particular directories
• **MyProxy CA default proxy lifetime is 12 hours**
• GCMU allows you to serve different sets of users with a single GridFTP server
• For example, some XSEDE resource providers wants to setup a GridFTP server that caters to both XSEDE and non-XSEDE users
  • Setup a GCMU on top of XSEDE security infrastructure they already have
  • Separate GO endpoints e.g., xsede#ranger associated with xsede myproxy and tacc#ranger associated with local myproxy
  • Non-xsede users pick tacc#ranger and enter their tacc credentials to access it