A brief introduction to GridFTP

Adam Carter, EPCC

EUDAT

What is GridFTP?

- A protocol for efficient file transfer
- An extension of FTP
- Currently the *de facto* standard for moving large files across the Internet
- Multiple implementations exist
  - A popular one is provided by the *Globus Toolkit*
  - *GlobusOnline* is a SaaS GridFTP client
- GridFTP is one of the ways of getting files into and out of EUDAT

EUDAT

Using GridFTP: Some background

- You move files between “endpoints” identified by a URI
- One of these endpoints is often your local machine, but GridFTP can also be used to move data between two remote endpoints
- X.509 (“grid”) certificates are used for authentication to a particular end-point
  - Usually you’ll have one certificate that you use at both end-points, but this need not be the case

EUDAT

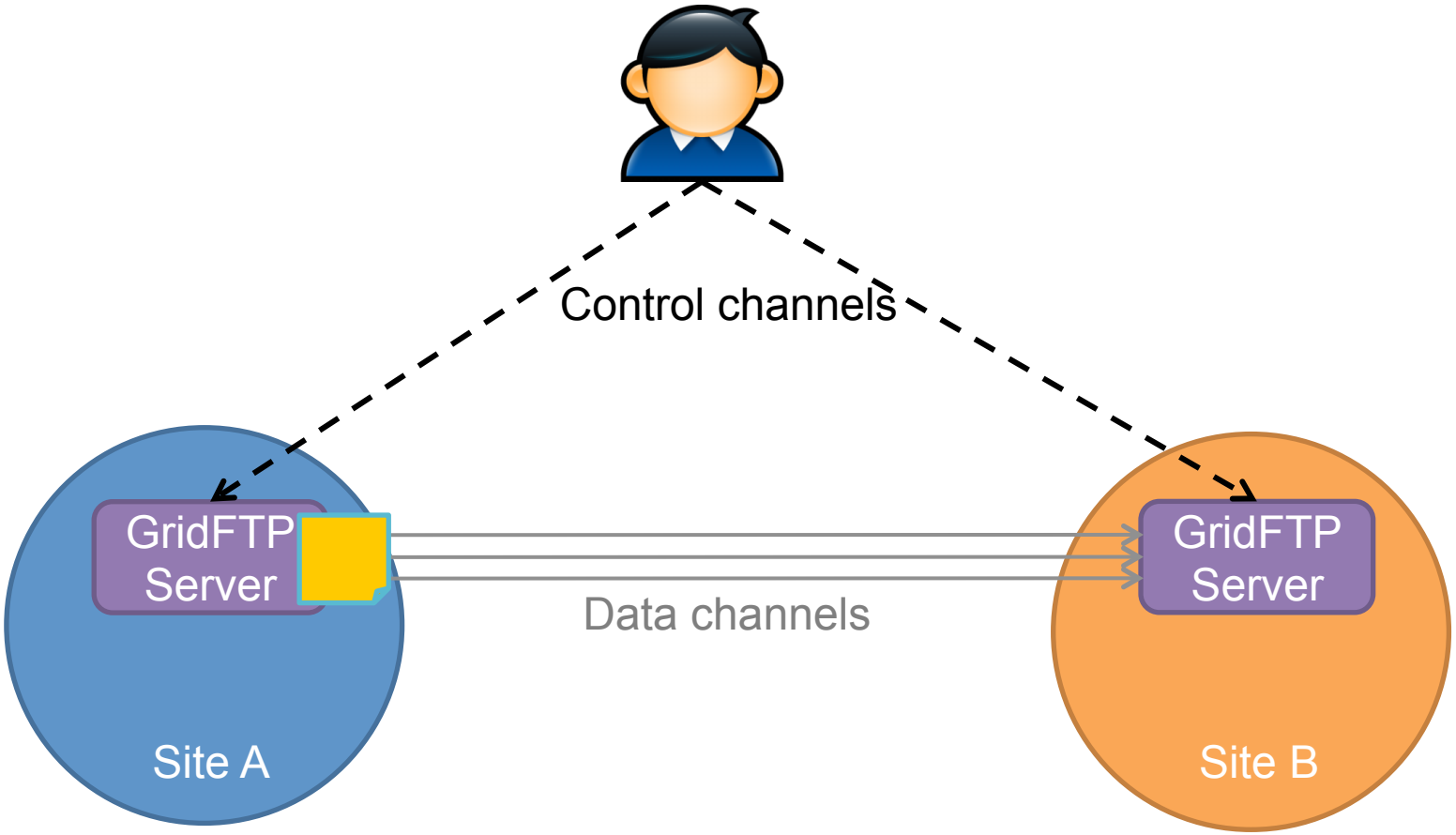
Features of GridFTP

- **Security with GSI**, The Grid Security Infrastructure
- **Third party transfers**
- **Parallel and striped transfer**
- **Partial file transfer**
- **Fault tolerance and restart**
- **Automatic TCP optimisation**

source: http://en.wikipedia.org/wiki/GridFTP

EUDAT

Third Party Transfers



EUDAT

How to use GridFTP
Normal Prerequisites

- Obtain a suitable X.509 certificate
  - depends on the endpoint you're connecting to, but generally for EUDAT sites:
  - any certificate issued by EUGridPMA, e.g. TERENA
- Ensure that the Globus Toolkit (or other GridFTP client) is installed on your local machine (globus.org)
- Install the X.509 certificate on your local machine
- Ensure that you have an account on the machine corresponding to your end point and that your DN is registered with the end point provider

EUDAT

Proxies

- Your (grid) certificate itself is not sent over the network
- A proxy is used instead, which is another certificate, signed by the user (or another proxy)

The commands to execute

- **grid-proxy-init** -cert /path/to/your/cert.pem -key /path/to/your/key.pem

- **globus-url-copy** gsiftp://dtn01.hector.ac.uk:2812/EPCC/home/my_irods_username/myfile_in_irods file:///my/file/on/local/filesystem

EUDAT

Why would you use GridFTP?

- Generally, to make use of the features mentioned earlier
- Why not scp?
  - Encryption slows transfers
- Why not FTP?
  - Insecure. Not supported at many secure sites (such as HPC centres)
- Why not BitTorrent?
  - Only works well with “popular” files

When would you use GridFTP?

- When you have large amounts of data that you need to move over the Internet (or, e.g. PRACE internal network)
  - Especially to or from third-party storage that you might not have a shell account on
- Remember that disks in the post can sometimes offer a higher bandwidth!

EUDAT

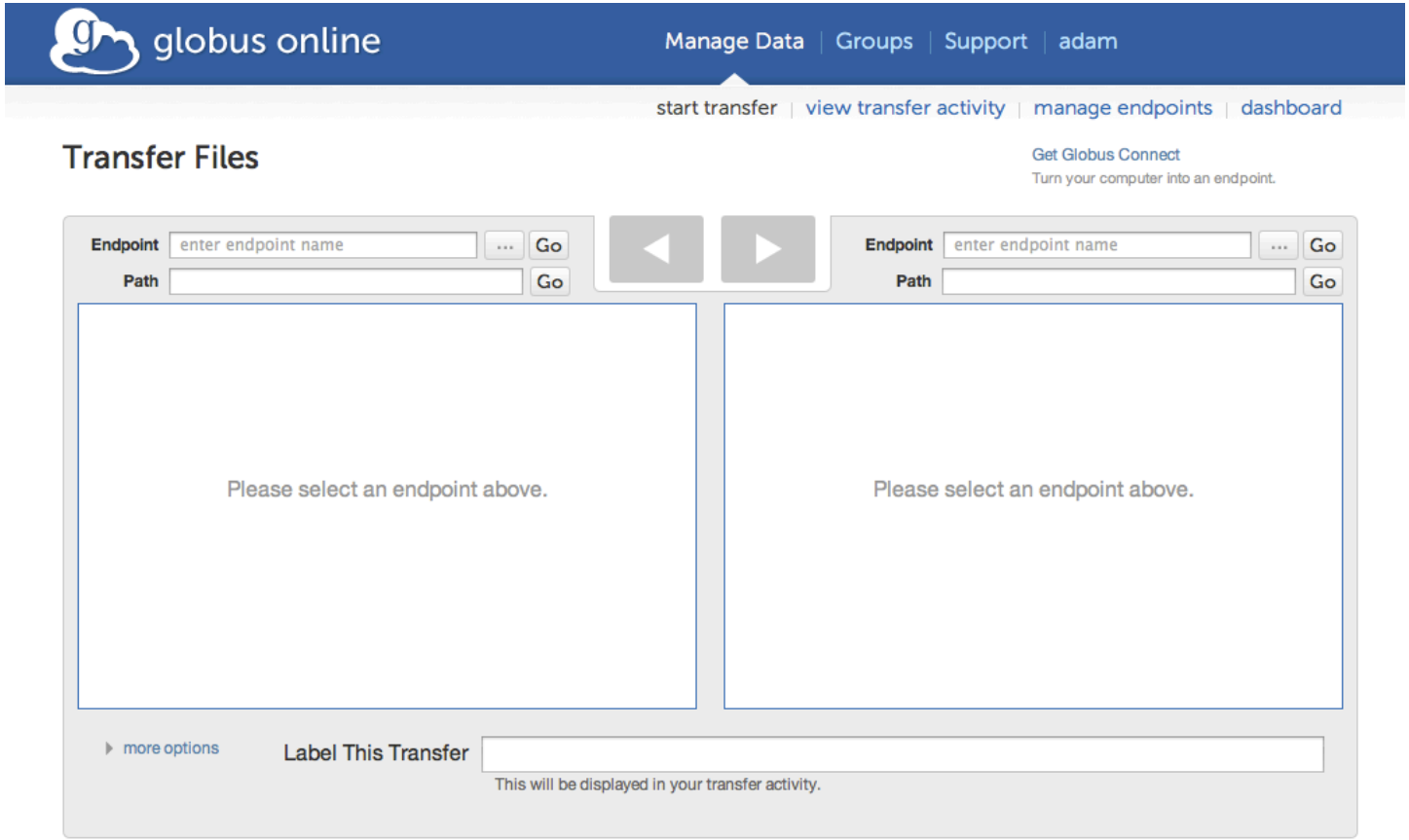
GridFTP endpoints

- CINECA, Bologna, Italy (EUDAT)
- EPCC, Edinburgh, UK (EUDAT)
- PSNC, Posnan, Poland (EUDAT)
- Most PRACE sites
- …

EUDAT

Globus Online

- See https://www.globusonline.org



EUDAT