



Globus GridFTP: High Performance Data Movement Tool for Grid/HPC Environments

Raj Kettimuthu
Argonne National Laboratory and
The University of Chicago



Outline

- Introduction
- GridFTP Protocol
- Security
- GridFTP Usage
- Failure Handling
- Globus.org – Hosted Data Movement Service

GridFTP

- High-performance, reliable data transfer protocol optimized for high-bandwidth wide-area networks
- Based on FTP protocol - defines extensions for high-performance operation and security
- Standardized through Open Grid Forum (OGF)
- GridFTP is the OGF recommended data movement protocol

GridFTP

- We (Globus Alliance) supply a reference implementation:
 - ◆ Server
 - ◆ Client tools
 - ◆ Development Libraries
- Multiple independent implementations can interoperate
 - ◆ Fermi Lab and U. Virginia have home grown servers that work with ours



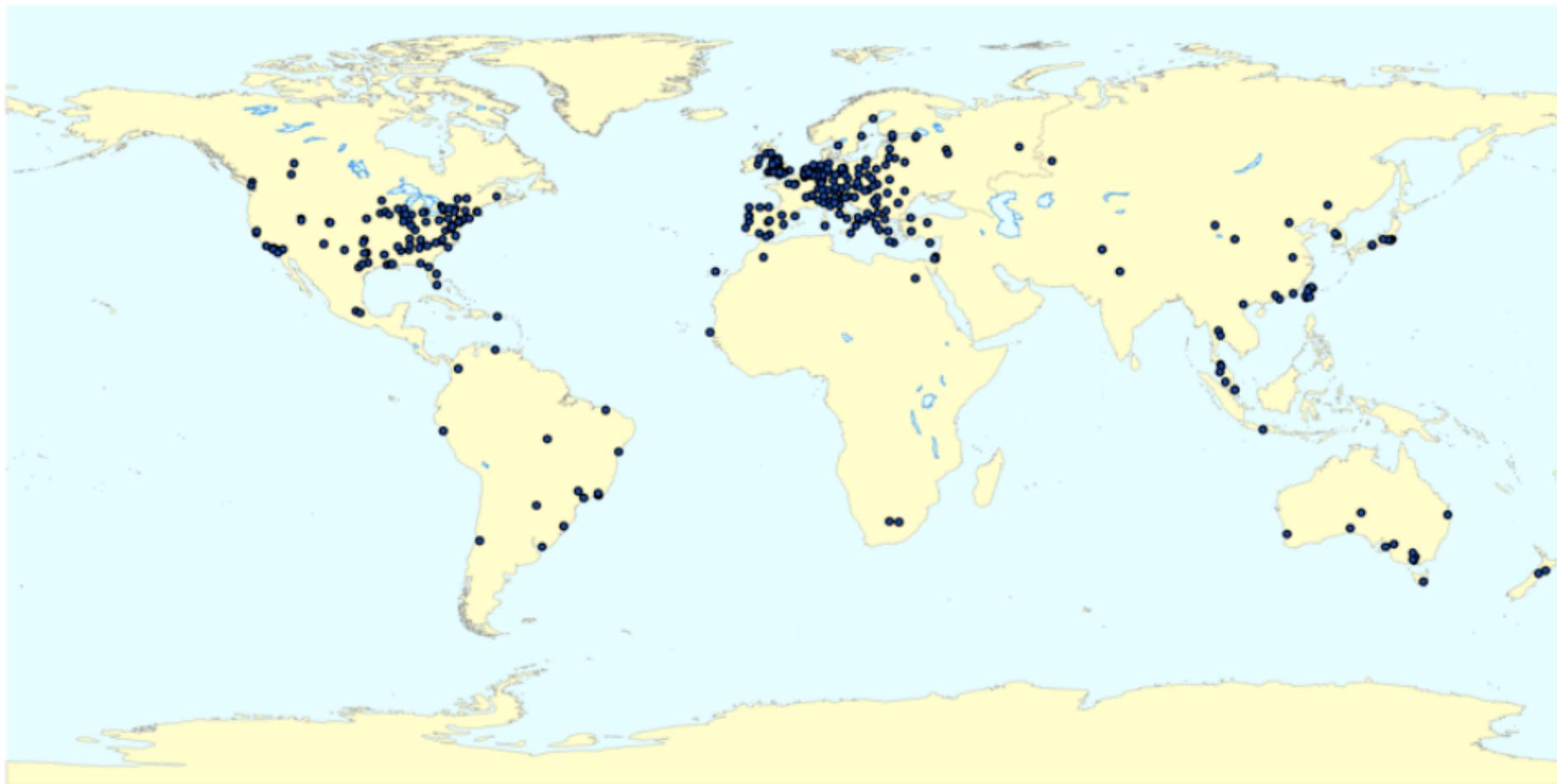
Globus GridFTP

- Performance
 - ◆ Parallel TCP streams, optimal TCP buffer
 - ◆ Non TCP protocol such as UDT
- Cluster-to-cluster data movement
- Multiple security options
 - ◆ Anonymous, password, SSH, GSI
- Support for reliable and restartable transfers



the globus alliance
www.globus.org

GridFTP Servers Around the World

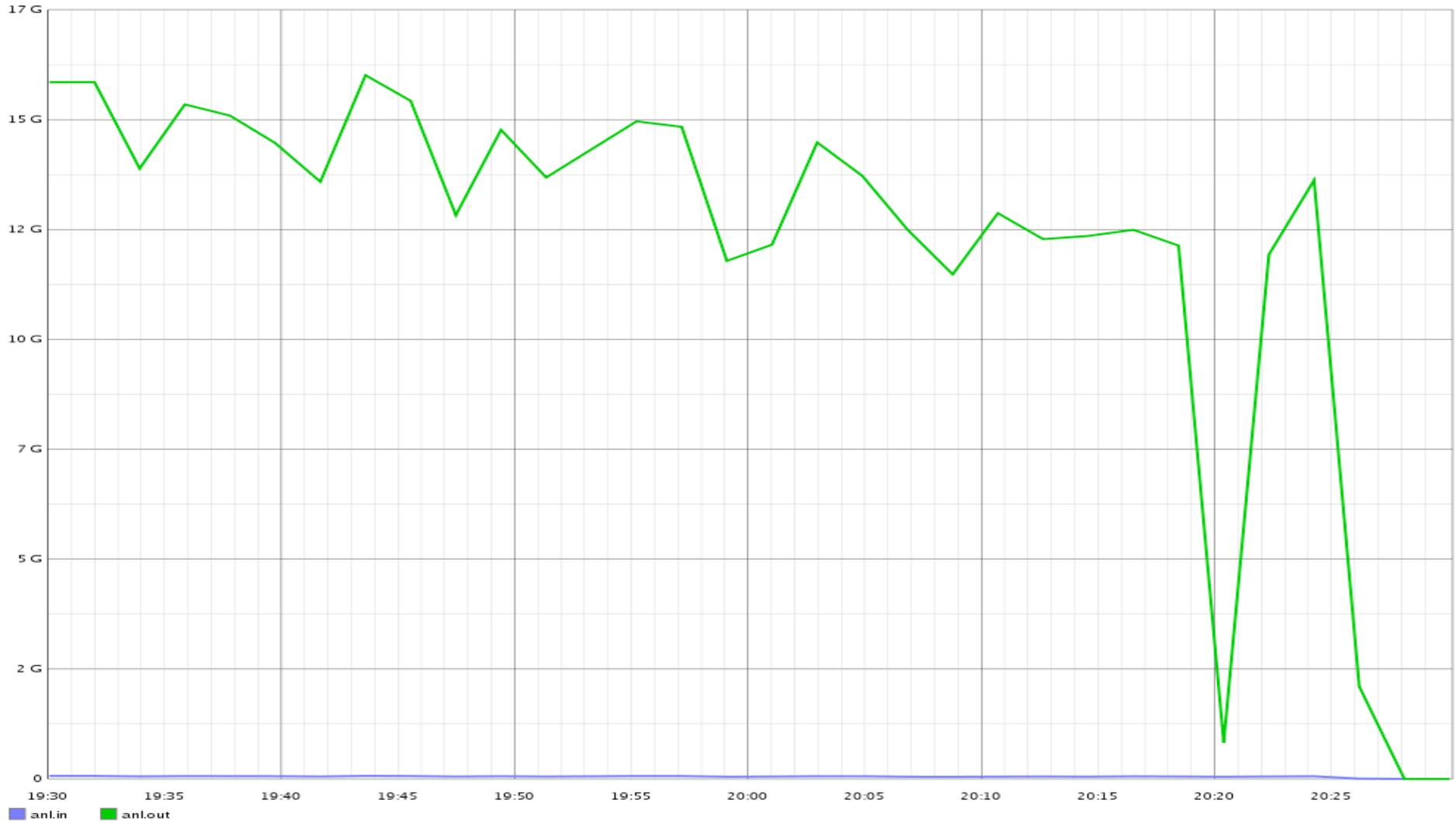


Created by Tim Pinkawa (Northern Illinois University) using MaxMind's GeolP technology (<http://www.maxmind.com/app/ip-locate>).



the globus alliance
www.globus.org

Performance



05/03/2010

Ohio Supercomputer Center



Understanding GridFTP

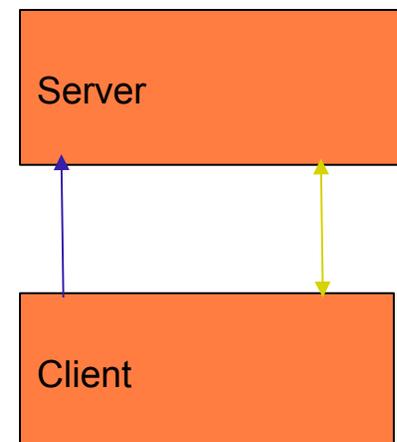
- Two channel protocol like FTP
- Control Channel
 - ◆ Command/Response
 - ◆ Used to establish data channels
 - ◆ Basic file system operations eg. mkdir, delete etc
- Data channel
 - ◆ Pathway over which *file* is transferred
 - ◆ Many different underlying protocols can be used
 - MODE command determines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



Client/Server and 3rd Party Transfers

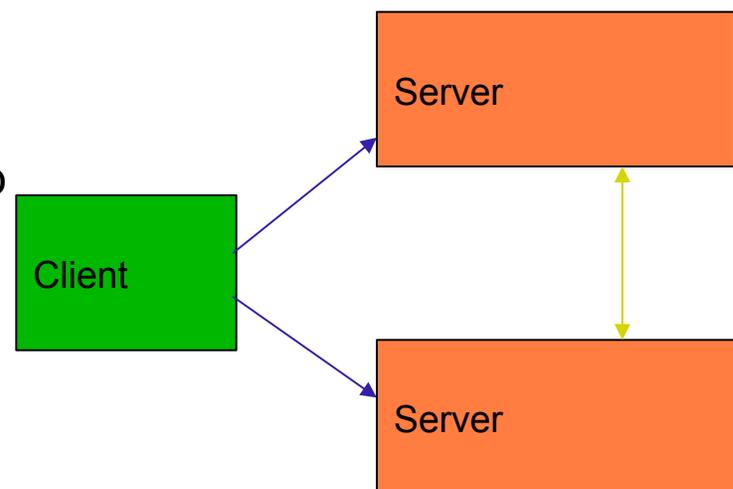
- Two party transfer

- ◆ The client connects and forms a CC with the server
- ◆ Information is exchanged to establish the DC
- ◆ A file is transferred over the DC



- Third party transfer

- ◆ Client initiates data transfer between 2 servers
- ◆ Client forms CC with 2 servers.
- ◆ Information is routed through the client to establish DC between the two servers.
- ◆ Data flows directly between servers
- ◆ Client is notified by each server when the transfer is complete

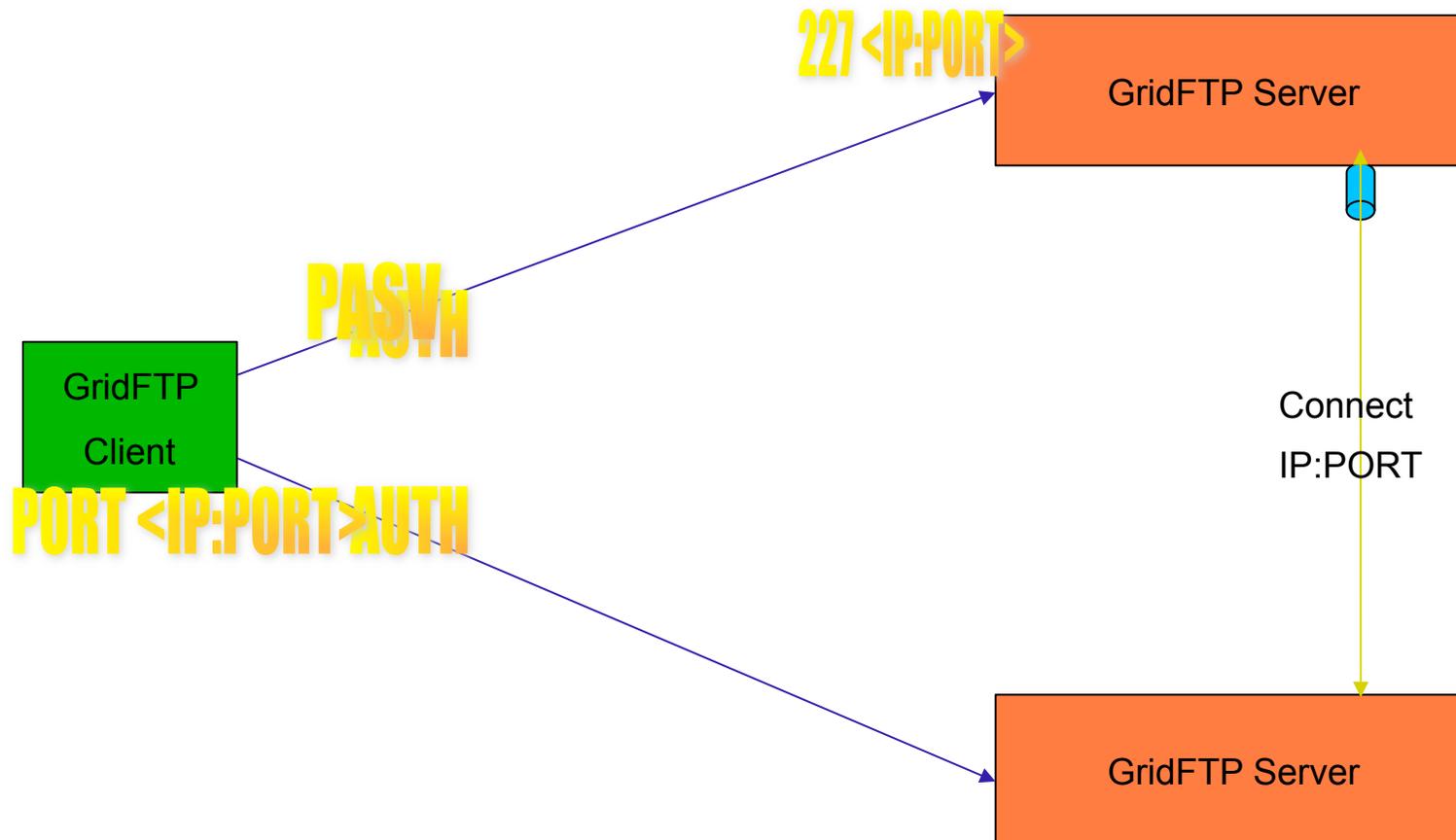




Control Channel Establishment

- Server listens on a well-known port (2811)
- Client form a TCP Connection to server
- 220 banner message
- Authentication
 - ◆ Anonymous
 - ◆ Clear text USER <username>/PASS <pw>
 - ◆ Base 64 encoded GSI handshake
- 230 Accepted/530 Rejected

Data Channel Establishment





Data Channel Protocols

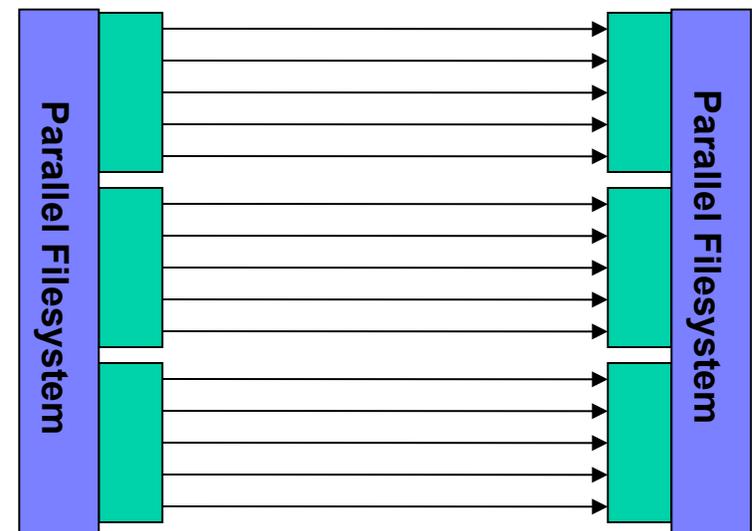
- **MODE Command**
 - ◆ Allows the client to select the data channel protocol
- **MODE S**
 - ◆ Stream mode, no framing
 - ◆ Legacy RFC959
- **MODE E**
 - ◆ GridFTP extension
 - ◆ Parallel TCP streams
 - ◆ Data channel caching

Descriptor (8 bits)	Size (64 bits)	Offset (64 bits)
------------------------	-------------------	---------------------



Cluster-to-Cluster transfers

- Multiple nodes work together as a single logical GridFTP server
- Multiple nodes are used to transfer data into/out of the cluster
 - Each node reads/writes only pieces they're responsible for
 - Head node coordinates transfers
- Multiple levels of parallelism
 - CPU, bus, NIC, disk etc.
 - Maximizes use of Gbit+ WANs



Striped Transfer
Fully utilizes bandwidth of
Gb+ WAN using multiple nodes.



Globus-url-copy

- Command line scriptable client
- Globus does not provide an interactive client
- Commonly used client for GridFTP
- Syntax overview
 - ◆ `globus-url-copy [options] srcURL dstURL`
 - ◆ `guc gsiftp://localhost/foo file:///bar`
 - Client/server, using FTP stream mode
 - ◆ `guc -vb -dbg -tcp-bs 1048576 -p 8 gsiftp://localhost/foo gsiftp://localhost/bar`
 - 3rd party transfer, MODE E
- URL rules
 - ◆ `protocol://[user:pass@][host]/path`
 - ◆ host can be anything resolvable - IP address, localhost, DNS name



Security Options

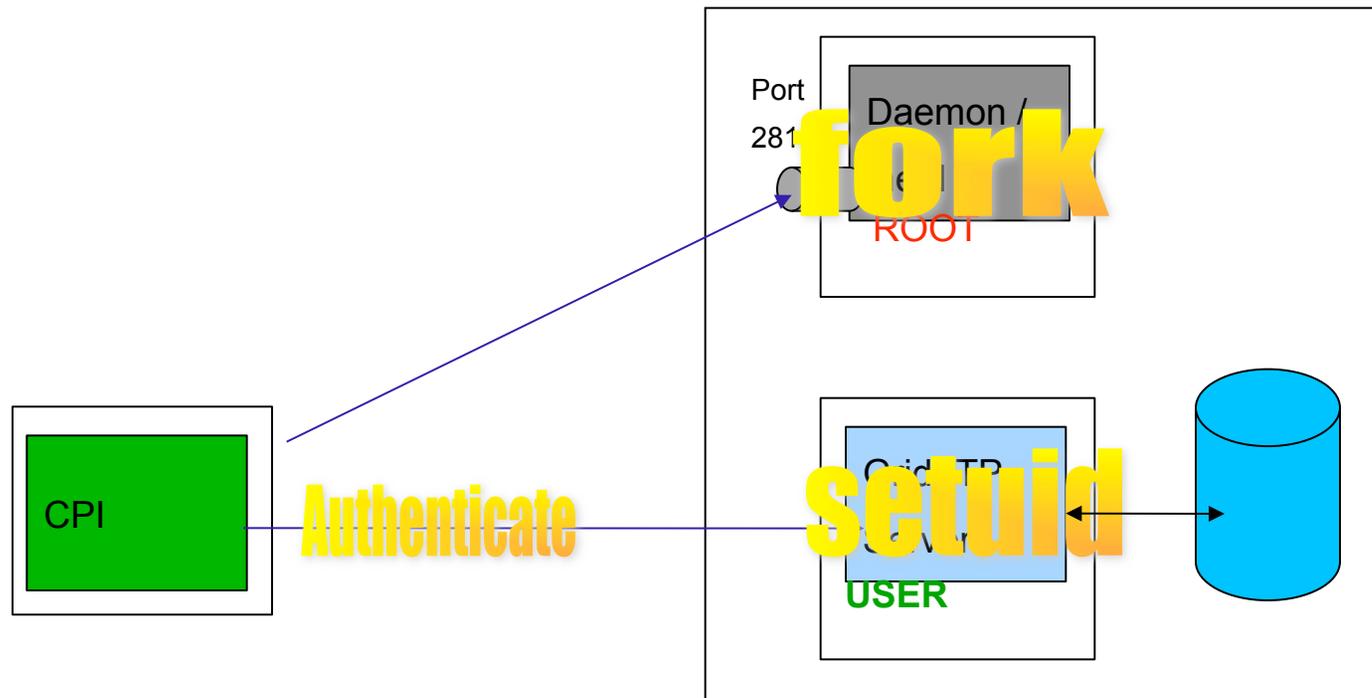
- Clear text (RFC 959)
 - ◆ Username/password
 - ◆ Anonymous mode (anonymous/<email addr>)
- SSHFTP
 - ◆ Use ssh/sshd to form the control connection
- GSIFTP
 - ◆ Authenticate control and data channels with GSI



User Permissions

- User is mapped to a local account and file permissions are handled by the OS
- inetd or daemon mode
 - ◆ Daemon mode - GridFTP server is started by hand and listens for connections on port 2811
 - ◆ Inetd/xinetd - super server daemon that manages internet services
 - ◆ Inetd can be configured to start up a GridFTP server upon receiving a connection on port 2811

inetd/daemon Interactions





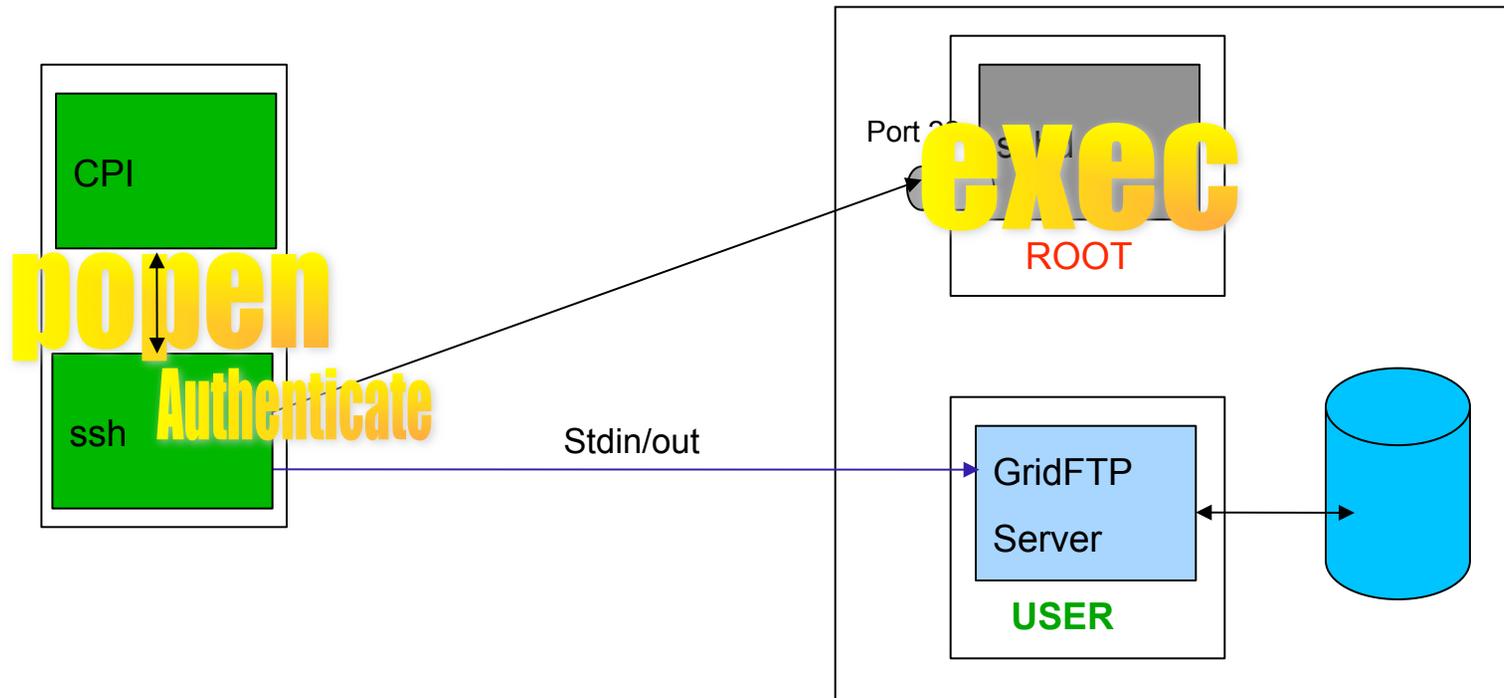
- Based on asymmetric cryptography
 - ◆ Private and Public Key - allows for two entities to authenticate with minimal cross-organizational support
- Certificates - Central concept in GSI
 - ◆ Information vital to identifying and authenticating user/service
 - ◆ Distinguished Name – unique Grid id for user/service
 - ◆ "/DC=org/DC=doegrids/OU=People/CN=Raj Kettimuthu 227852"
- Certificate Authority (CA)
 - ◆ Trusted 3rd party that confirms identity
- Host credential
 - ◆ Long term credential
- User credential
 - ◆ Passphrase protected



Security

- GridFTP provides strong security using GSI
- Protection vs. Ease of use
 - ◆ GSI and CAs were hard for many users
- Speed vs. protection
 - ◆ Users are happy with a minimal amount of data channel protection
- GridFTP over SSH
 - ◆ A big win for many users

sshftp:// Interactions



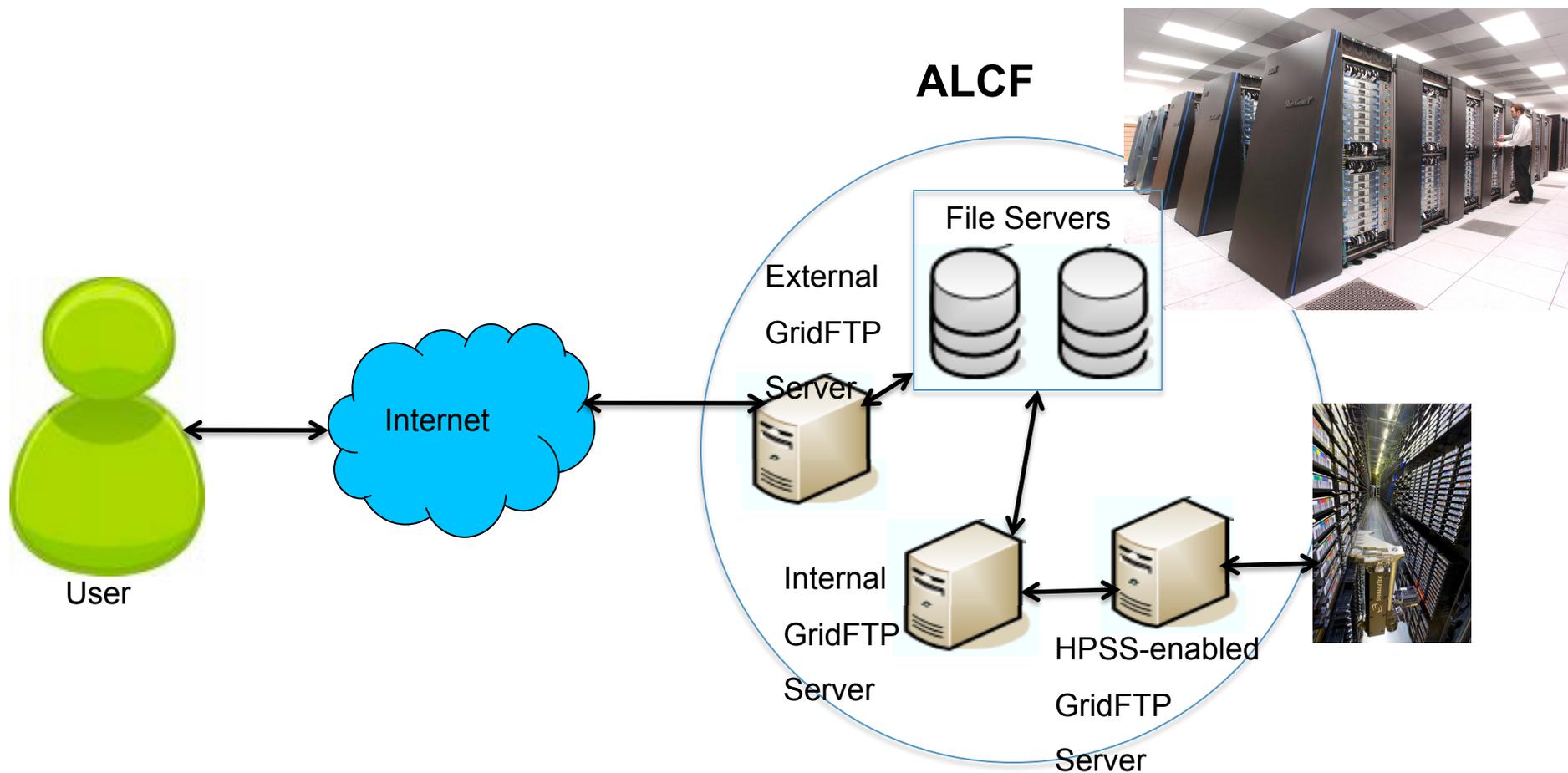


GridFTP in production

- Many Scientific communities rely on GridFTP
 - ◆ High Energy Physics - LHC computing Grid
 - ◆ Southern California Earthquake Center (SCEC), Earth Systems Grid (ESG), Relativistic Heavy Ion Collider (RHIC), European Space Agency, BBC use GridFTP for data movement
- GridFTP facilitates an average of more than 7 million data transfers every day

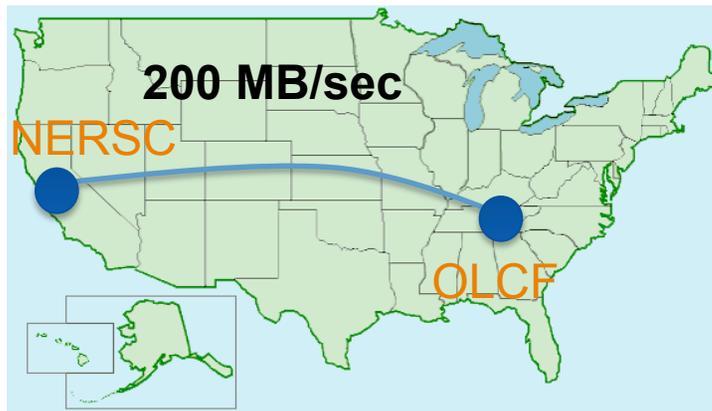


GridFTP in Production

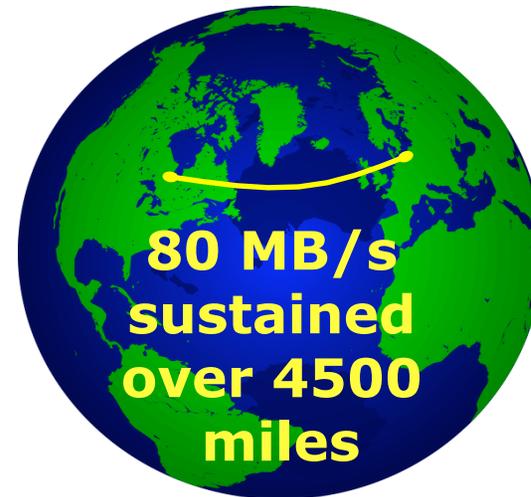




GridFTP in production



Move 40 terabyte (40 trillion bytes) from one DOE center (NERSC) to another (OLCF) in **under 3 days** rather than **several months**

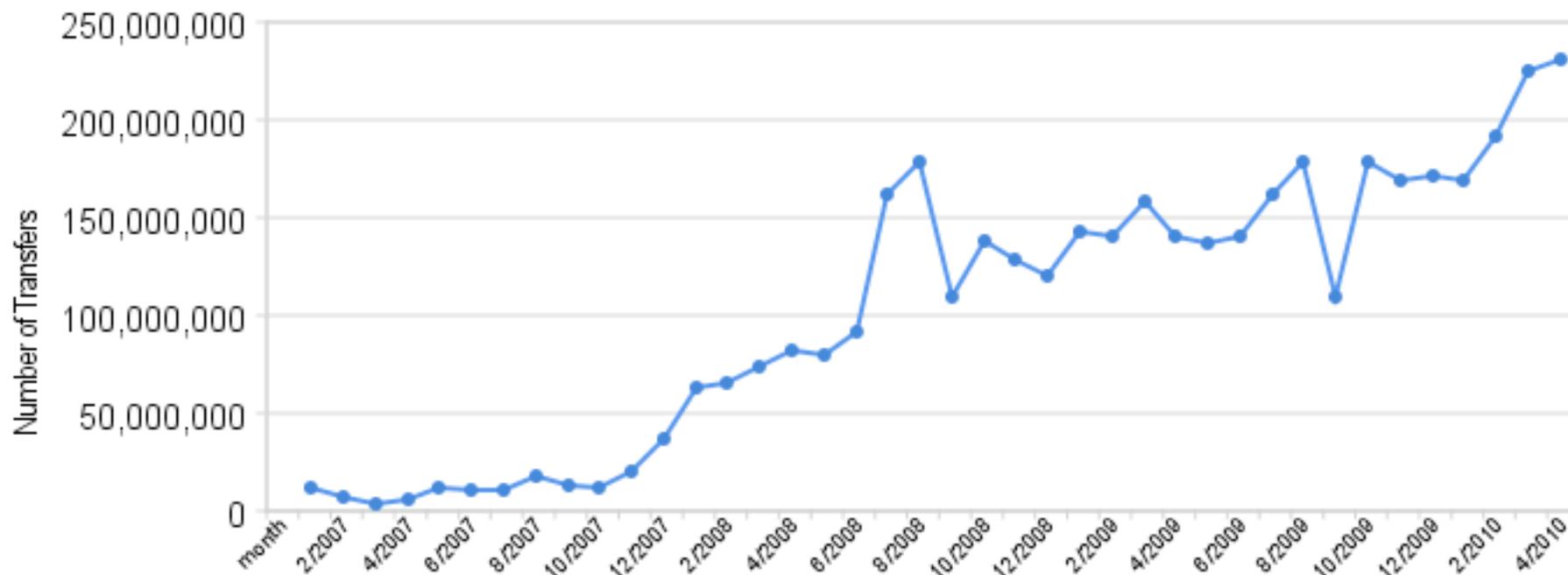


1.5 terabyte moved from University of Wisconsin, Milwaukee to Hannover, Germany at a sustained rate of 80 megabyte/sec



GridFTP Usage

Monthly Totals* of GridFTP File Transfers



*for those "reporting"



Handling failures

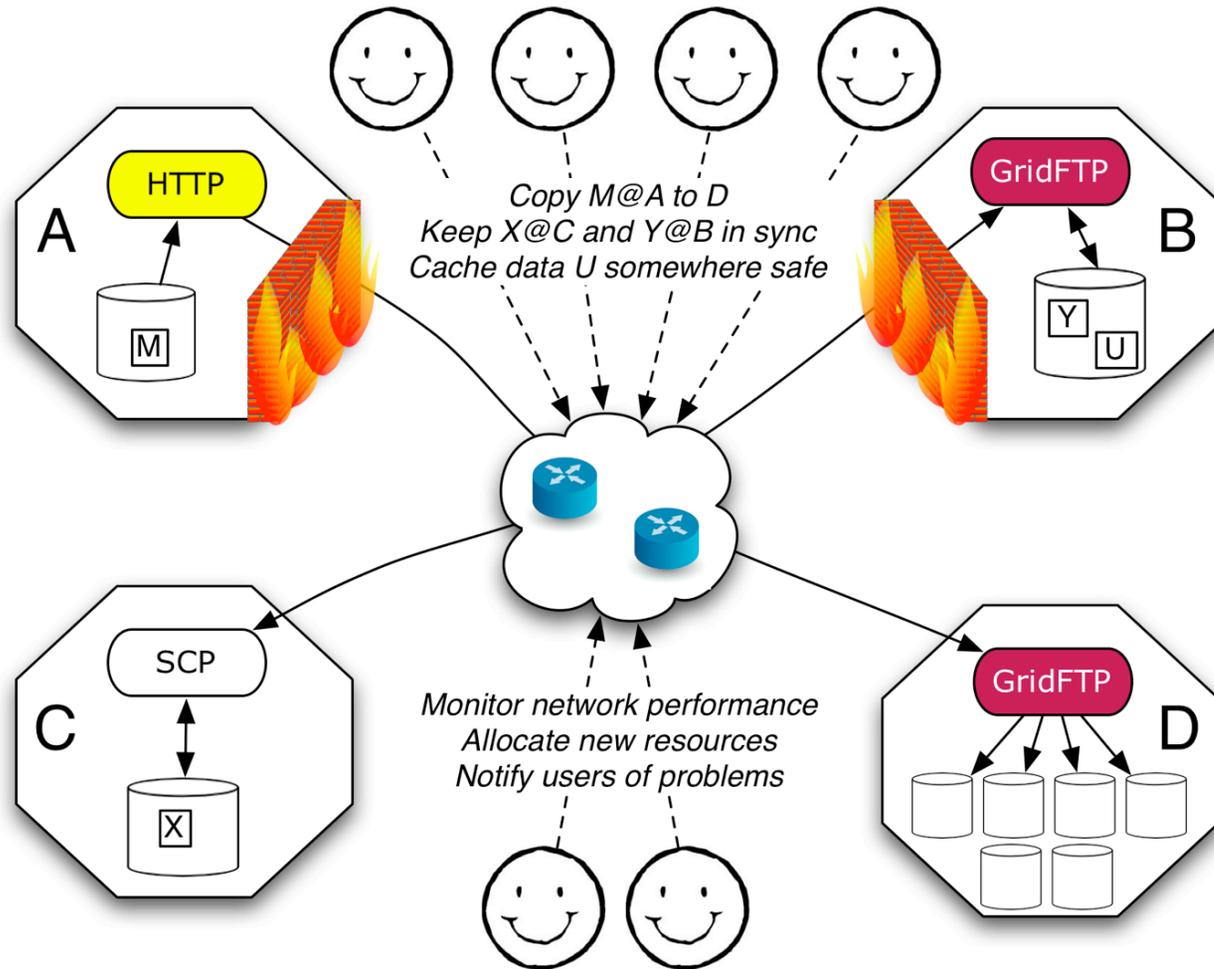
- GridFTP server sends restart and performance markers periodically
 - ◆ Default every 5s - configurable
- Helpful if there is any failure
 - ◆ No need to transfer the entire file again
 - ◆ Use restart markers and transfer only the missing pieces
- GridFTP supports partial file transfers



Server failure

- Command-line client - globus-url-copy - support transfer retries
 - ◆ Use restart markers
- Recover from server and connection failures
- What if the client fails in the middle of a transfer?

Globus.org – hosted data movement service





the globus alliance
www.globus.org

Globus.org

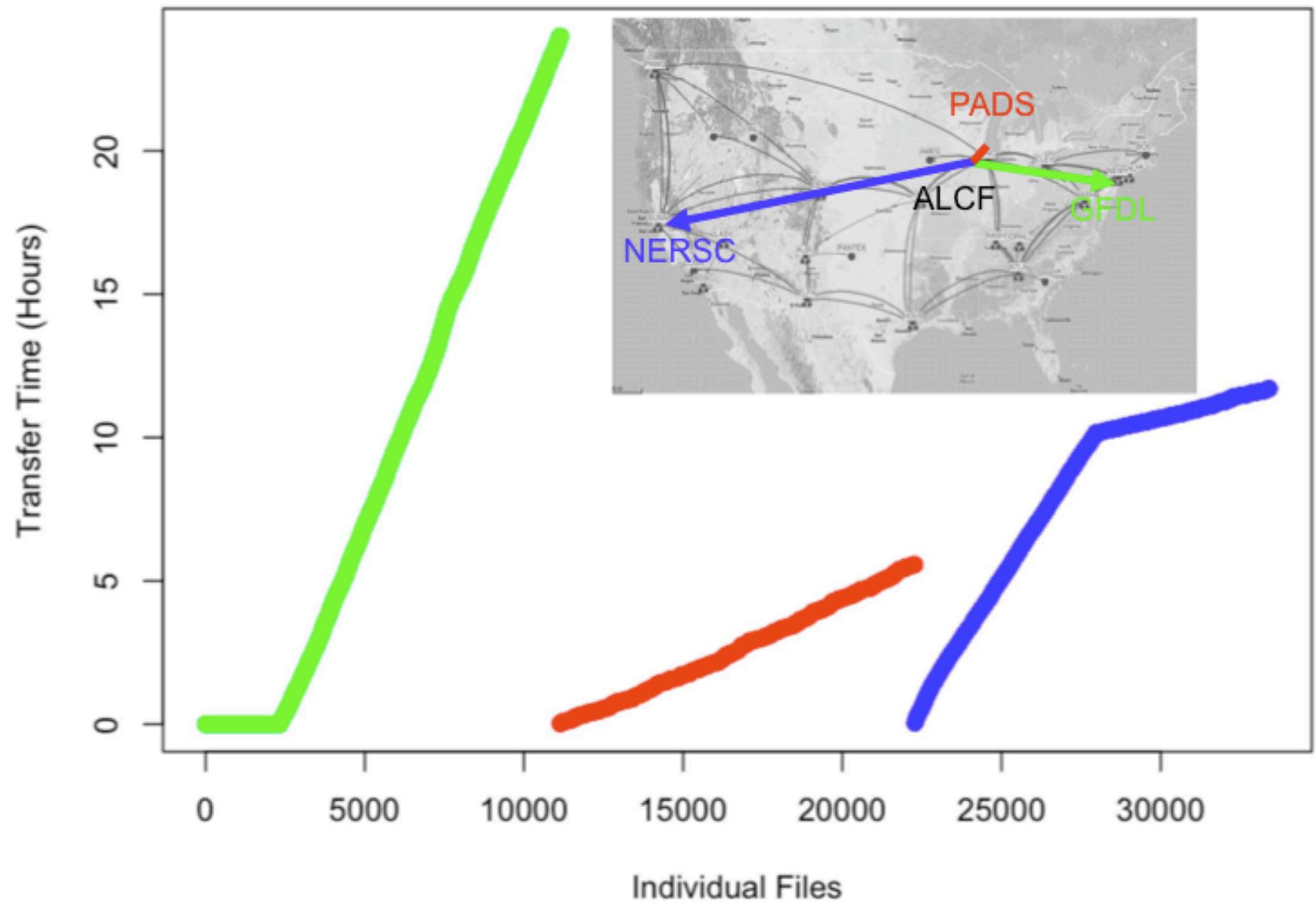
Value Additions for GridFTP

- Fire and forget
 - ◆ Less user interaction
 - ◆ Email notifications
 - ◆ No need to babysit transfers
- Failure handling
 - ◆ Automatic retries
- Familiar user interfaces
- Technology interactions requiring no special expertise
- No client software to install

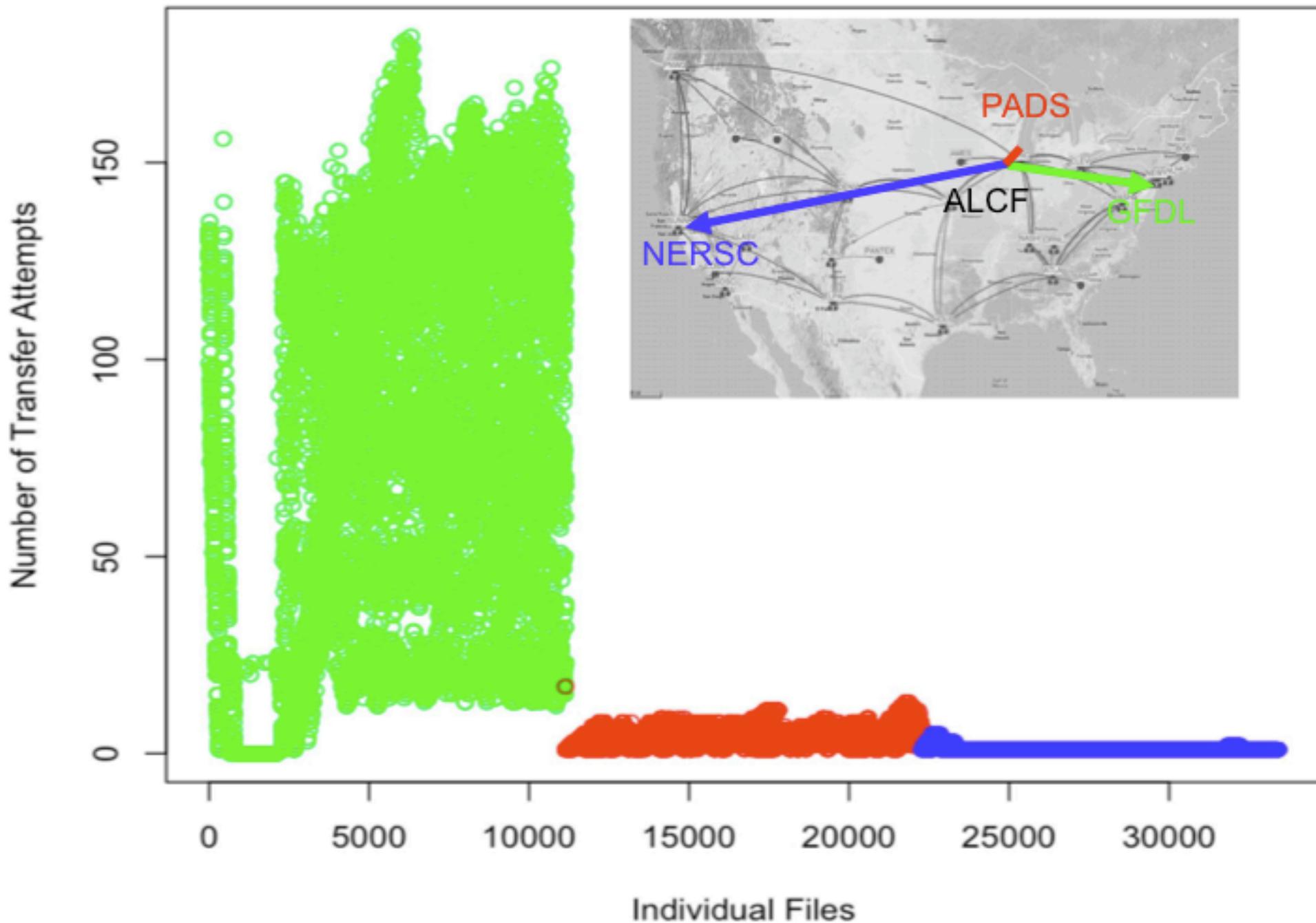
Globus.org

- Enable users to focus on domain-specific work
 - ◆ Manage technology failures
 - ◆ Notifications of interesting events
 - ◆ Provide users with enough information to resolve problems
- Ease the infrastructure providers' support burden
 - ◆ Hosted and supported by Globus team

ALCF to GFDL ALCF to PADS ALCF to NERSC



ALCF to GFDL ALCF to PADS ALCF to NERSC





More Information at
<http://www.gridftp.org>
<http://www.globus.org/service/>

Questions