the globus alliance
www.globus.org

GridFTP GUI: An Easy and Efficient Way to Transfer Data in Grid

Wantao Liu[1,2] Raj Kettimuthu[2,3], Brian Tieman[3],
Ravi Madduri[2,3], Bo Li[1], and Ian Foster[2,3]

[1]Beihang University, Beijing, China

[2]The University of Chicago, Chicago, USA

[3]Argonne National Laboratory, Argonne, USA

Outline

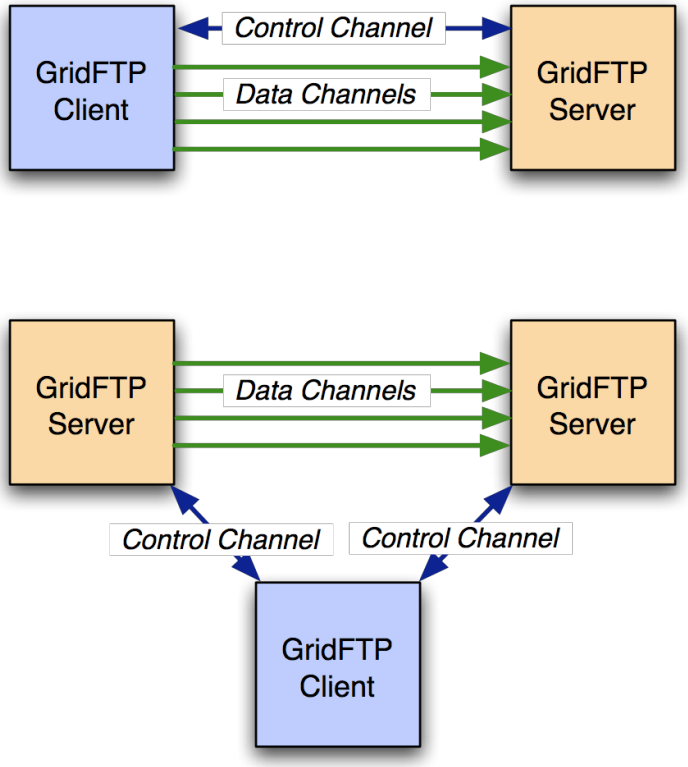
- GridFTP overview
- GridFTP Challenges
- Commonly used GridFTP clients
- Zero configure GUI client
- Experimental results

GridFTP

- A secure, robust, fast, efficient, standards based, widely accepted data transfer protocol
- We also supply a reference implementation:
  - Server
  - Client tools (globus-url-copy)
  - Development Libraries
- Multiple independent implementations can interoperate
  - University of Virginia and Fermi Lab have home grown servers that work with ours.
- Lots of people have developed clients independent of the Globus Project.

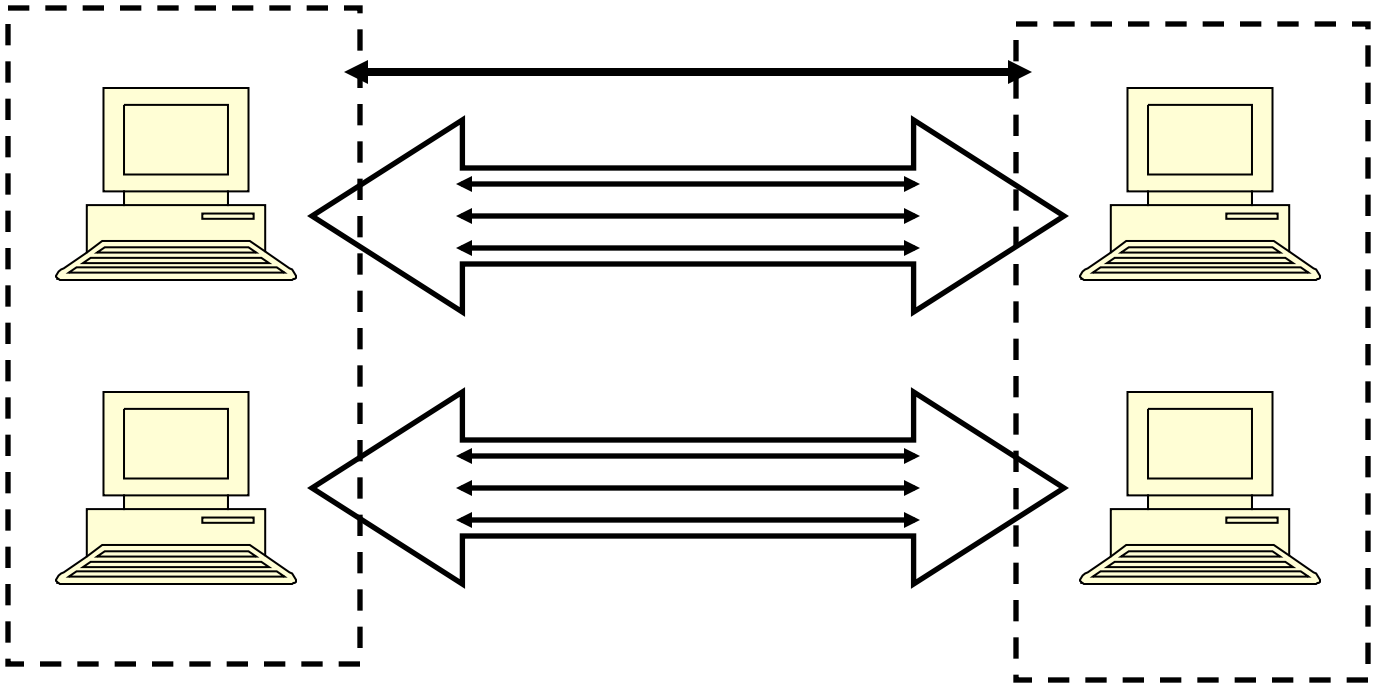
GridFTP

- Two channel protocol like FTP
- Control Channel
  - Communication link (TCP) over which commands and responses flow
  - Low bandwidth; encrypted and integrity protected by default
- Data Channel
  - Communication link(s) over which the actual data of interest flows
  - High Bandwidth; authenticated by default; encryption and integrity protection optional



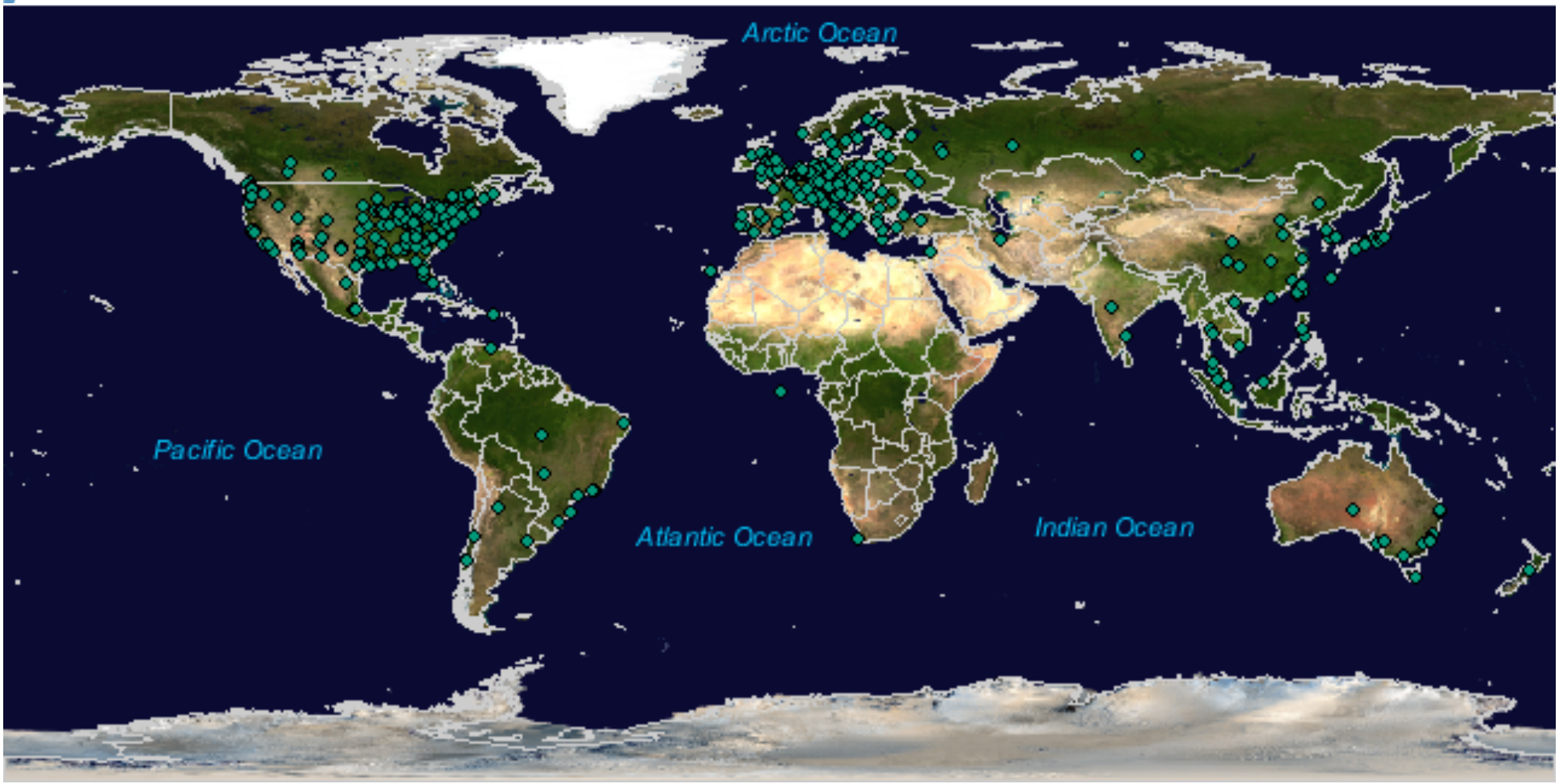
Striping

- GridFTP offers a powerful feature called striped transfers (cluster-to-cluster transfers)



the globus alliance
www.globus.org

GridFTP Servers Around the World



Created by Lydia Prieto ; G. Zarrate; Anda Imanitchi (Florida State University) using MaxMind's GeoIP technology (http://www.maxmind.com/app/ip-locate).

GridFTP in production

- Many Scientific communities rely on GridFTP
  - High Energy Physics – tiered data movement infrastructure for the LHC computing Grid
  - LIGO routinely uses GridFTP to move 1 TB a day
  - 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 5 million data transfers every day

Challenges

- Past success
  - Standard – big selling point for adoption
  - Throughput – GridFTP was sold on speed
  - Robustness – has to work all the time
- Current and future
  - Ease-of-use
    - Zero configuration clients
    - Firewall
  - Scalable
  - Extensible

Globus-url-copy

- Commonly used command line scriptable client
- globus-url-copy [options] srcURL dstURL
- URL format - protocol://[user:pass@] [host]/path
- Users can do client/server and 3rd party transfers using globus-url-copy

Other clients

- UberFTP
- Reliable file transfer service
- Custom clients using globus C and Java client libraries
- All these clients require non-trivial configuration
  - Security setup
- None of these clients provide graphical user interface

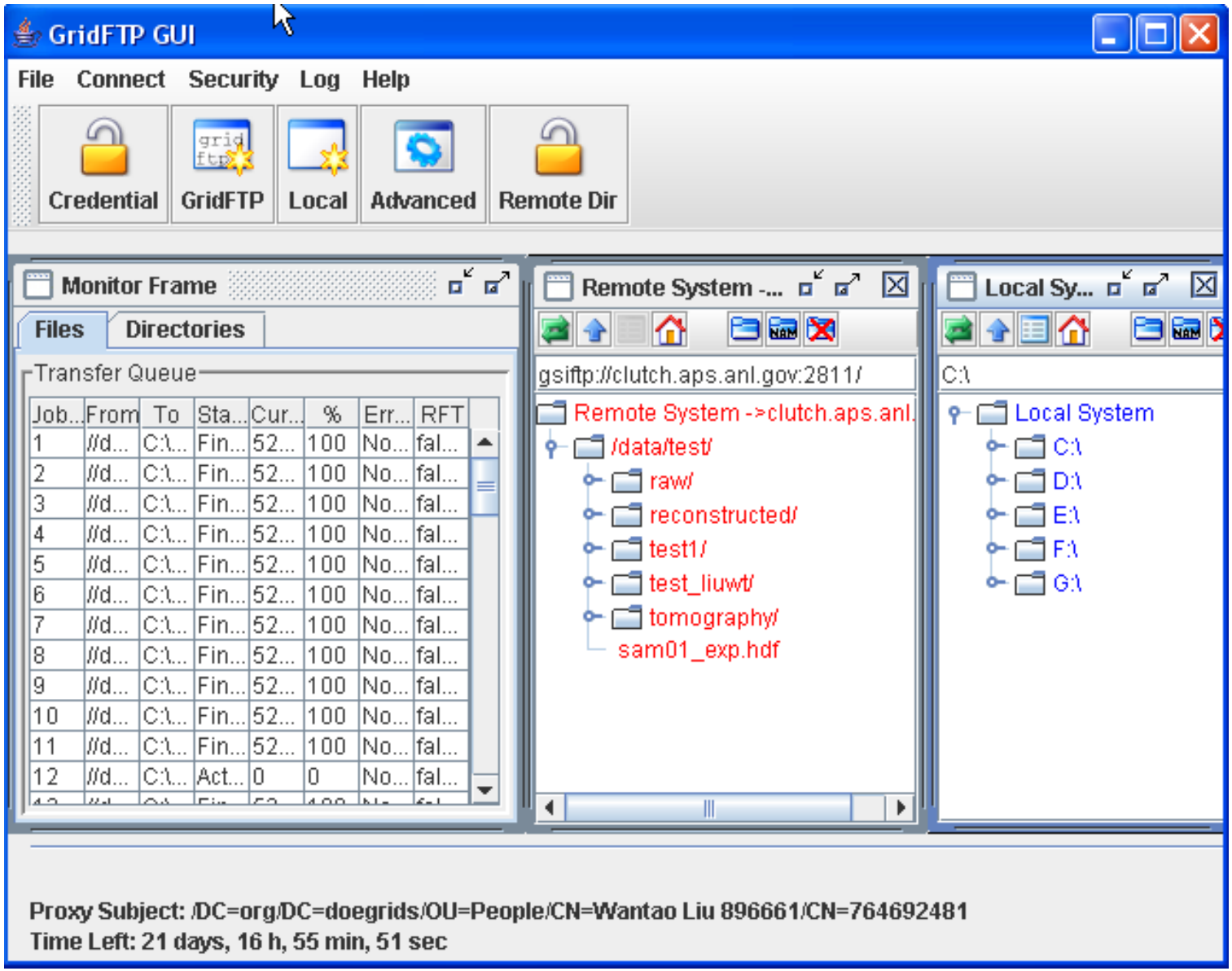
the globus alliance
www.globus.org

GridFTP GUI

- Drag and drop
- Zero configuration
  - Integrated with myproxy
  - Automatically trusts the CAs part of IGTF distribution
- Fault tolerant
- Transfer status monitoring
- Optimized for performance

the globus alliance
www.globus.org

Snapshot of the GUI



Fault tolerant

- Better fault tolerance than other GridFTP clients
  - Like other clients, GUI can recover from transient server and network failures
  - Globus-url-copy can not recover from its own failures
  - GUI can recover from its own failures
  - Unlike RFT, stores information on the local file system

Lots of small files

- Scientific experiments produce huge volume of data
  - the individual file size is modest, on the order of kilobytes or megabytes
  - hundreds of thousands of files to transfer every day
  - the size of the entire dataset is tremendous, from hundreds of gigabytes to hundreds of terabytes

Advanced Photon Source

- Advanced Photon Source at Argonne
  - dozens of samples may be acquired for one experiment every day
  - each sample generates about 2,000 raw data files
  - after processing, each sample produces additional 2,000 reconstructed files
  - each file is 8 to 16 MB in size

Lots of small files

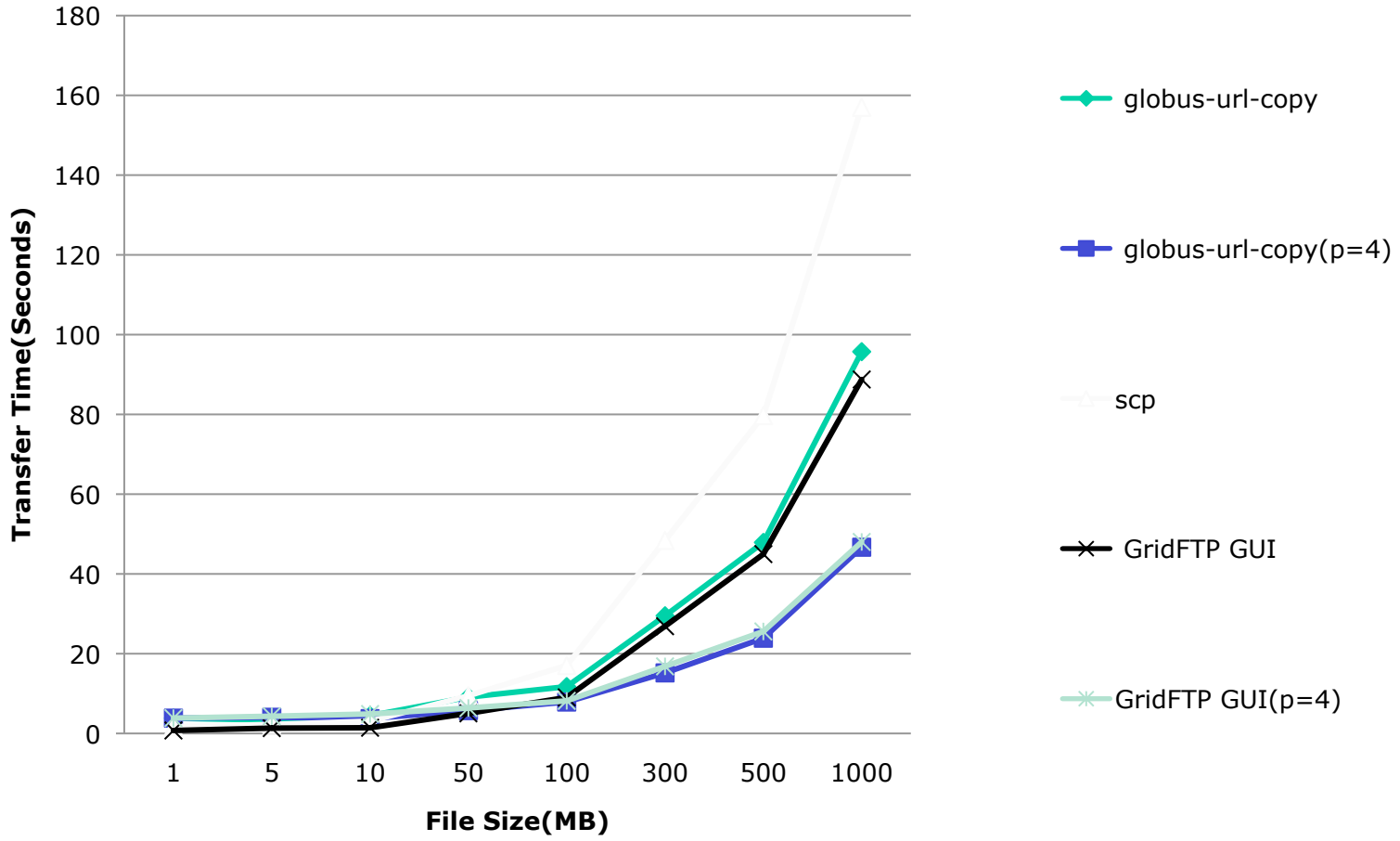
- Transfer threads pool
  - Move multiple files concurrently
  - Maximize the utilization of network bandwidth
  - Improve the transfer performance
- Two windows for status information
  - Directory window lists all directories and their transfer status
  - File window lists all files under the active directory

Experiment Setup

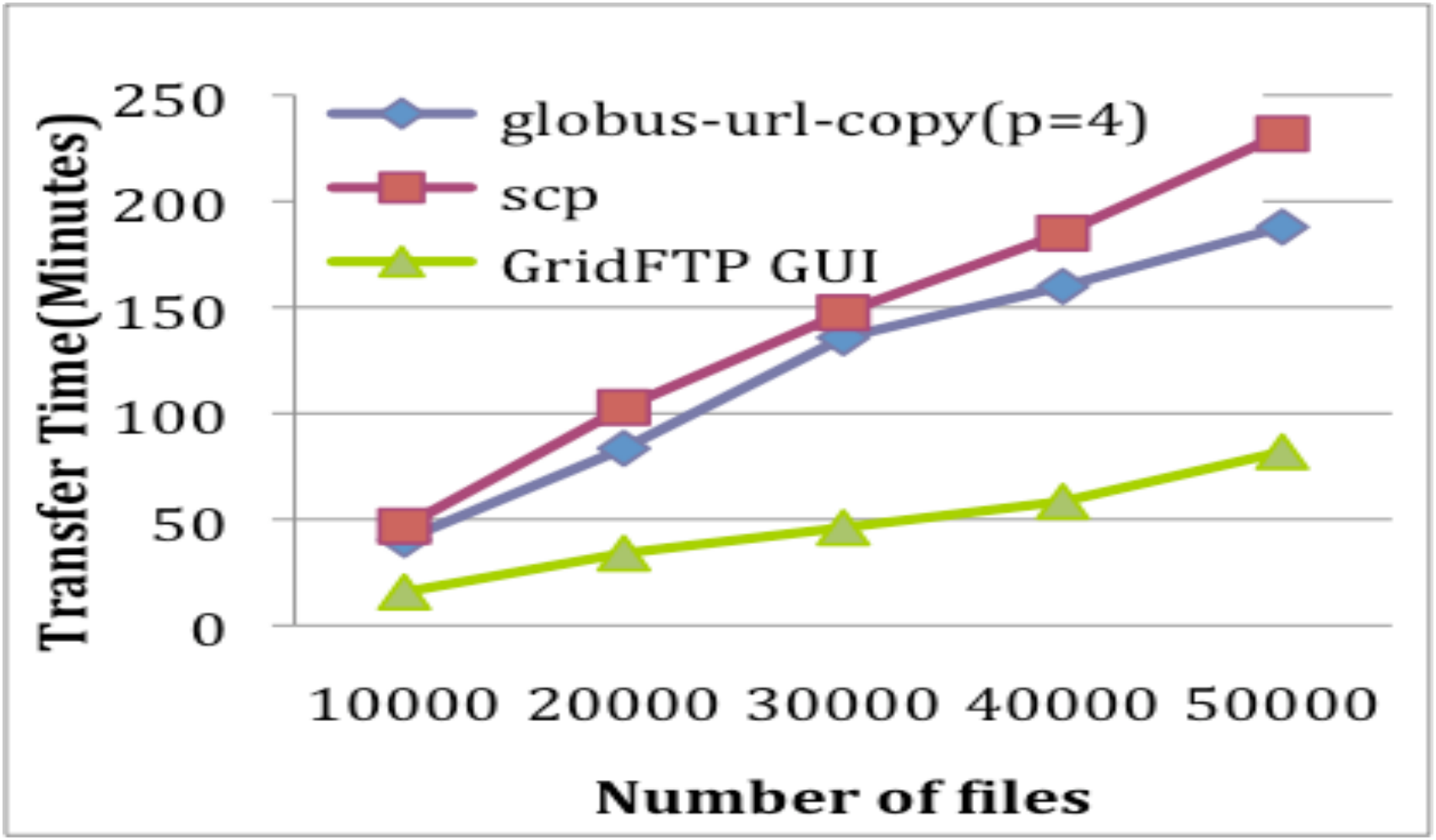
- We conducted all of our experiments using TeraGrid NCSA nodes and the University of Chicago nodes
- GridFTP GUI is compared with scp and globus-url-copy
- TCP is configured as the underlying data transport protocol

the globus alliance
www.globus.org

Experiment Results



Experiment Results(cont.)



Questions