the globus alliance
www.globus.org

Data Movement in Distributed Science Environments

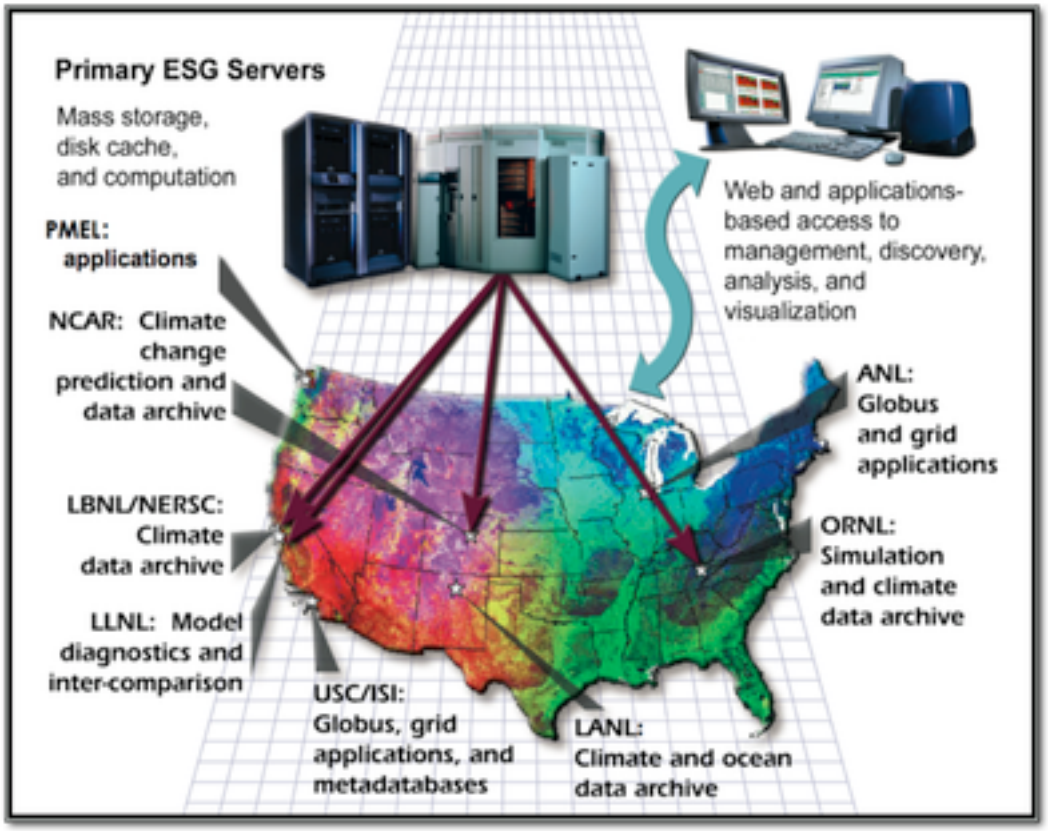
Raj Kettimuthu
Argonne National Laboratory and
The University of Chicago

Outline

- Science Environments
- Data movement problem
- GridFTP
- New features
- Challenges
- Globus.org – Hosted Data Movement Service

the globus alliance
www.globus.org

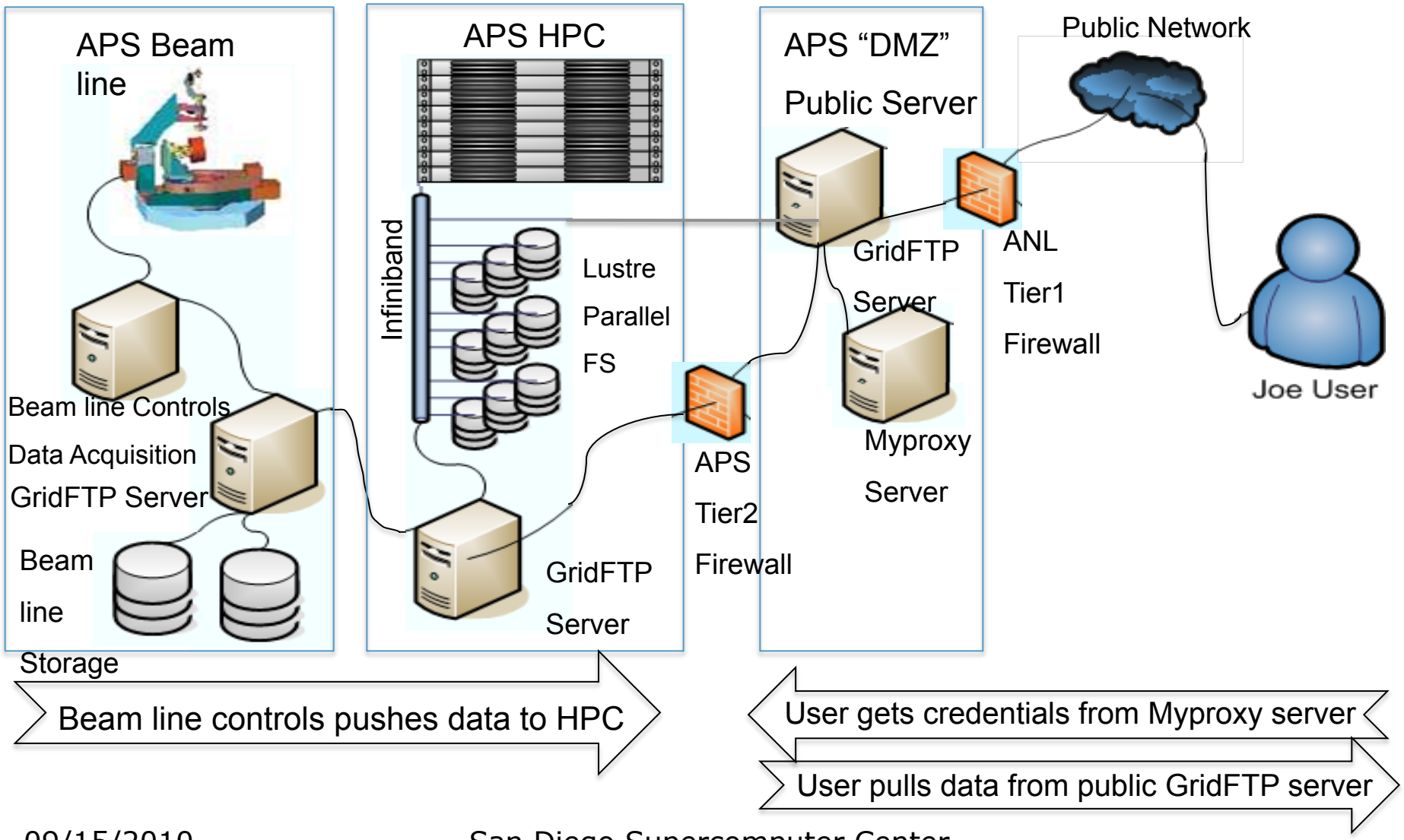
Distributed Science



- Distributed community of users to access and analyze large amounts of data

the globus alliance
www.globus.org

Advanced Photon Source



the globus alliance
www.globus.org

Bandwidth Requirements

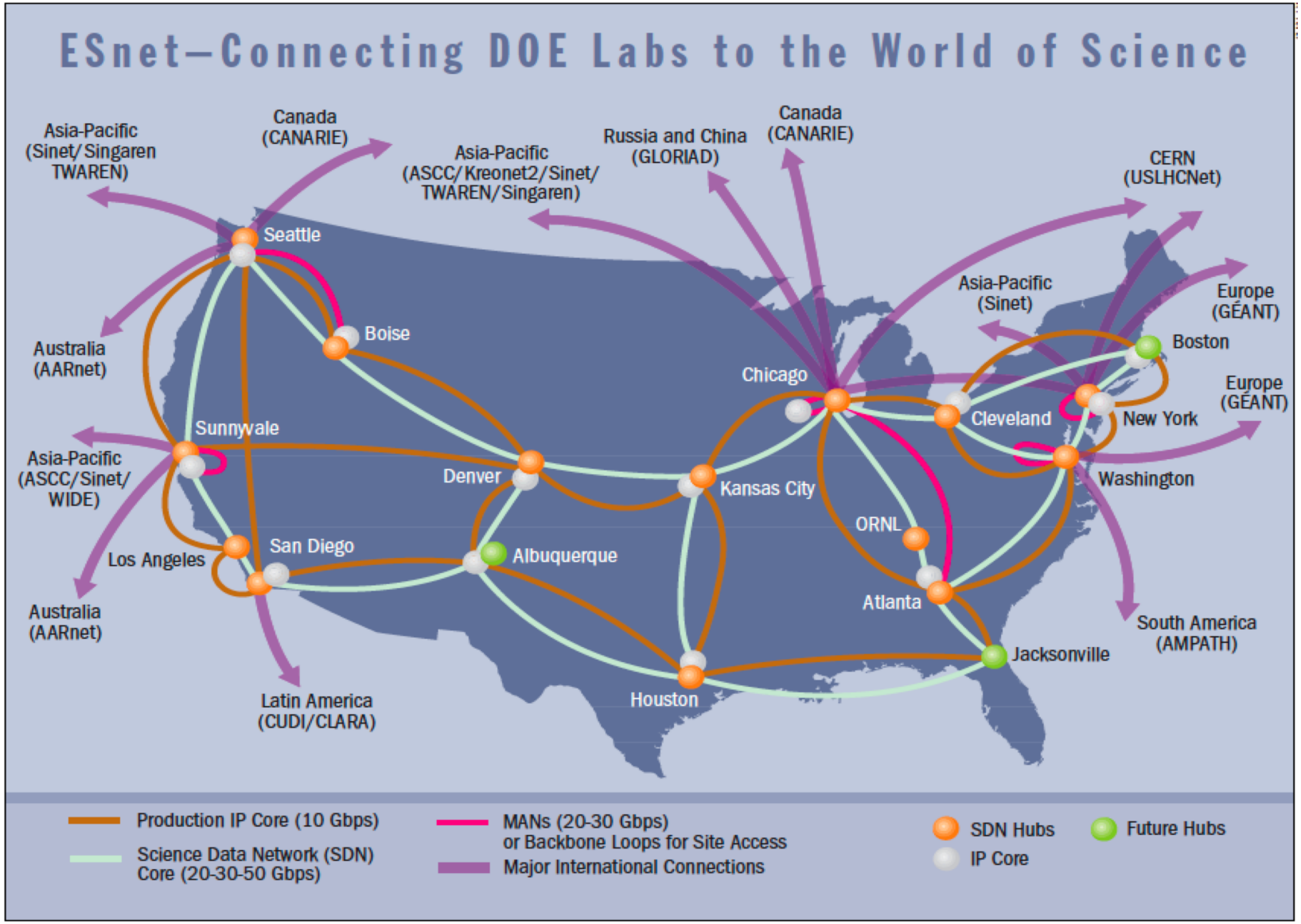
Bandwidth Requirements to move Y Bytes of data in Time X

Bits per Second Requirements

	1H	8H	24H	7Days	30Days
10PB	25,020.0 Gbps	3,127.5 Gbps	1,042.5 Gbps	148.9 Gbps	34.7 Gbps
1PB	2,502.0 Gbps	312.7 Gbps	104.2 Gbps	14.9 Gbps	3.5 Gbps
100TB	244.3 Gbps	30.5 Gbps	10.2 Gbps	1.5 Gbps	339.4 Mbps
10TB	24.4 Gbps	3.1 Gbps	1.0 Gbps	145.4 Mbps	33.9 Mbps
1TB	2.4 Gbps	305.4 Mbps	101.8 Mbps	14.5 Mbps	3.4 Mbps
100GB	238.6 Mbps	29.8 Mbps	9.9 Mbps	1.4 Mbps	331.4 Kbps
10GB	23.9 Mbps	3.0 Mbps	994.2 Kbps	142.0 Kbps	33.1 Kbps
1GB	2.4 Mbps	298.3 Kbps	99.4 Kbps	14.2 Kbps	3.3 Kbps
100MB	233.0 Kbps	29.1 Kbps	9.7 Kbps	1.4 Kbps	0.3 Kbps

the globus alliance
www.globus.org

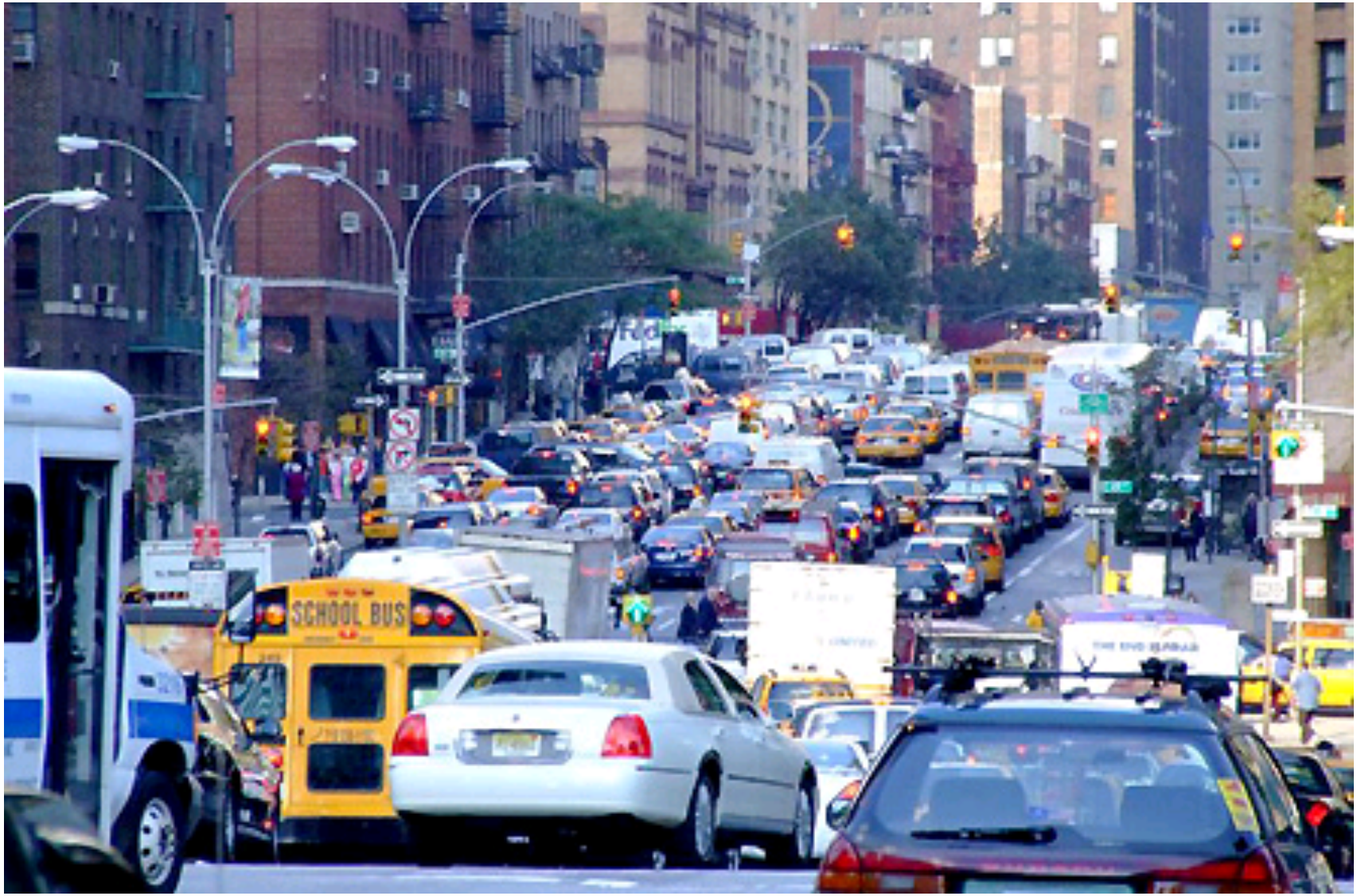
ESNET



TeraGrid

the globus alliance
www.globus.org

End-to-End Problem



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

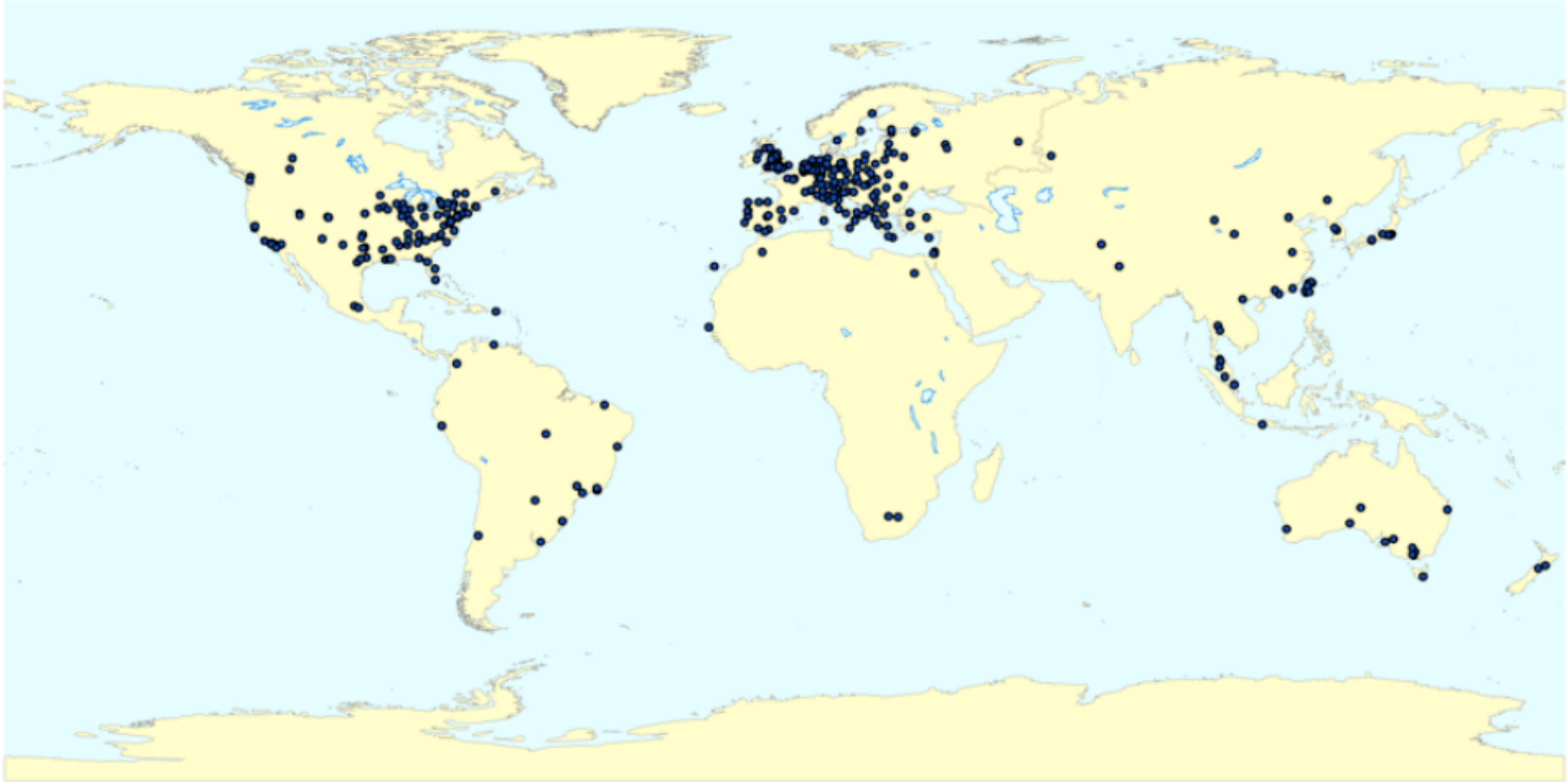
Standards

- Interoperability
  - Big selling point for adoption
- Multiple independent implementations
  - Globus provides a reference implementation
    - Server
    - Client tools
    - Development libraries
  - Fermi Lab and U. Virginia have home grown servers that work with ours

Robustness

- It has to work ALL the time
  - Hard to get a solid stable code base
    - Harder to extend it
  - Race conditions
  - Recover from errors
  - Ability to check point transfers
  - A session crash can't be a service crash
    - Fork()/setuid()/exec()

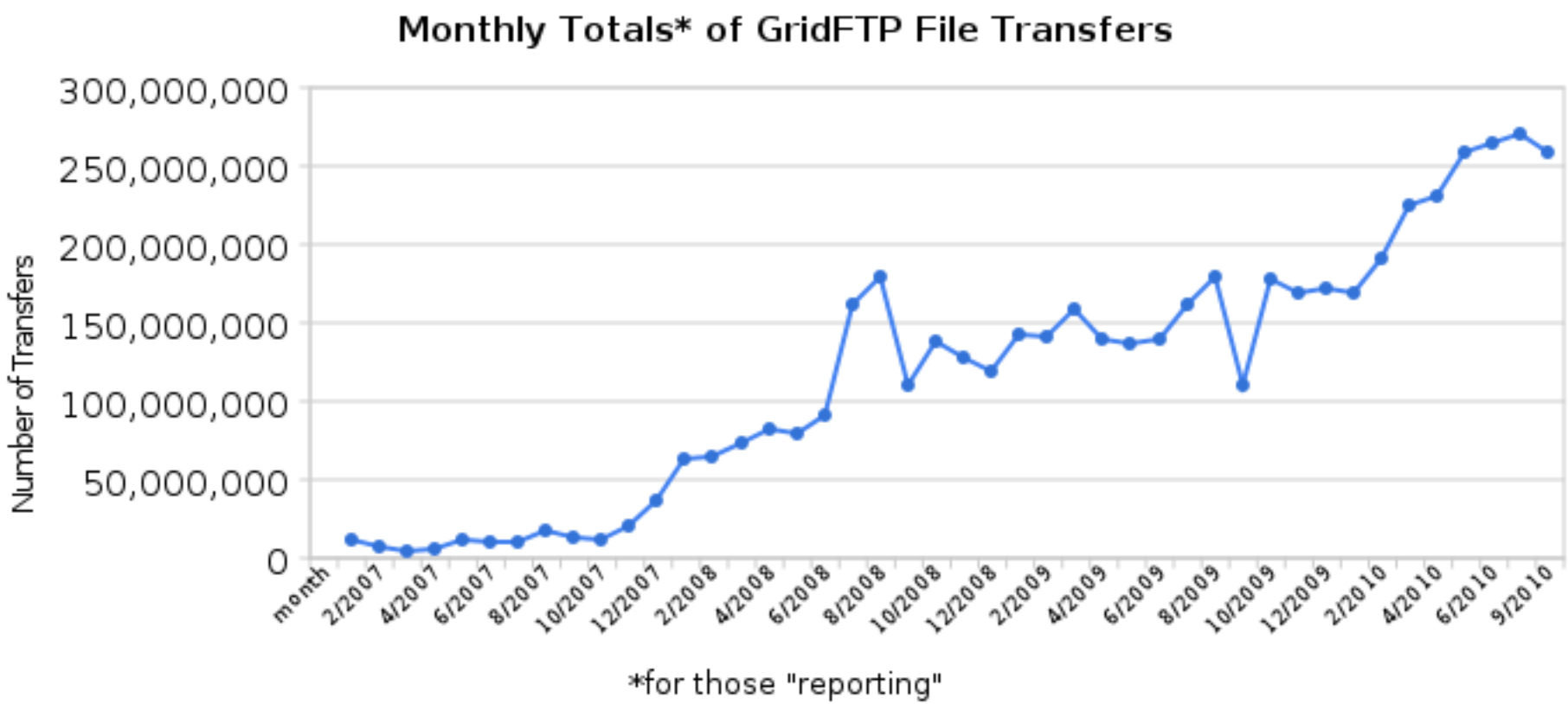
GridFTP Servers Around the World



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

the globus alliance
www.globus.org

GridFTP Usage

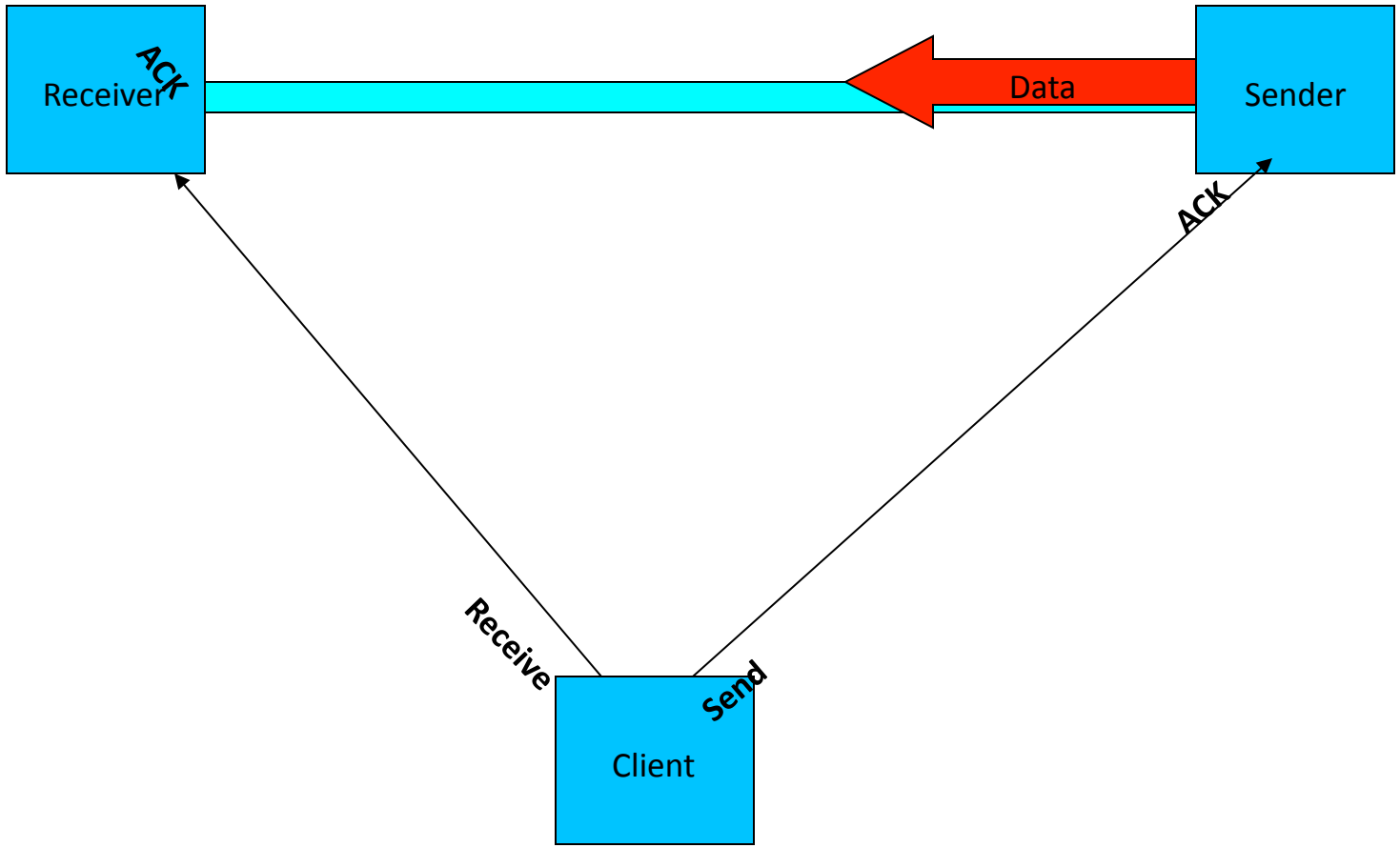


Throughput

- It had to be fast
  - GridFTP was sold on speed
  - Fast varies with the environment
    - LANs, WANs, Long Fat Pipe
  - Parallel TCP streams, optimal TCP buffer
  - Non TCP protocol such as UDT
- Striping
  - Multi-node data movement

FTP transfer pattern

Traditional transfer pattern

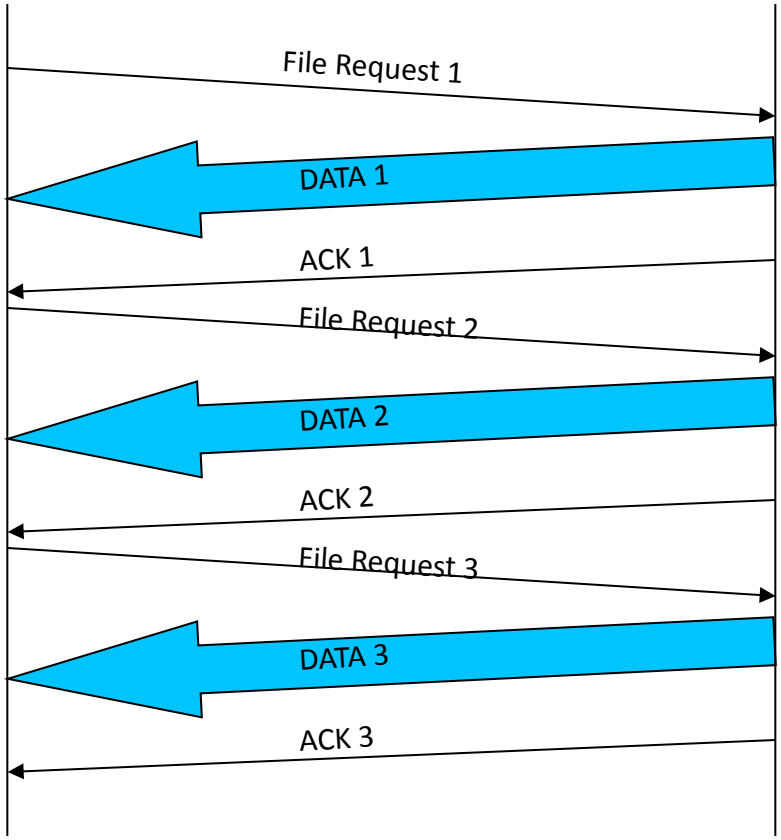


Lots of Small Files

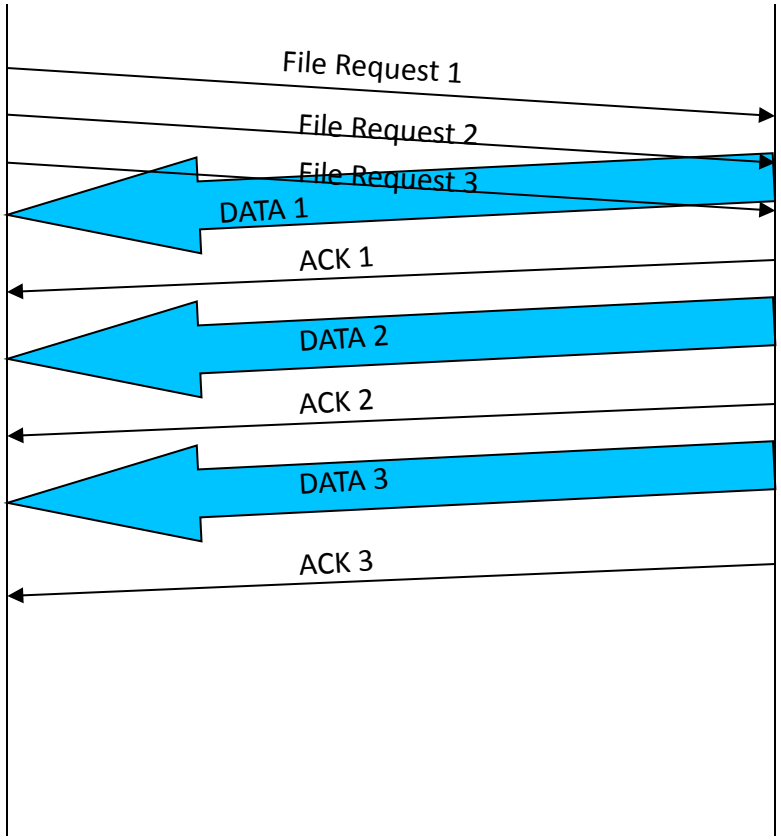
- Large files are easy (but less prevalent)
  - Overhead is low
- Datasets partitioned into many small files
  - Overlap control overhead with data payload
  - Data channel caching
  - Pipelining
  - Concurrent sessions
  - On-the-fly tar

Pipelining

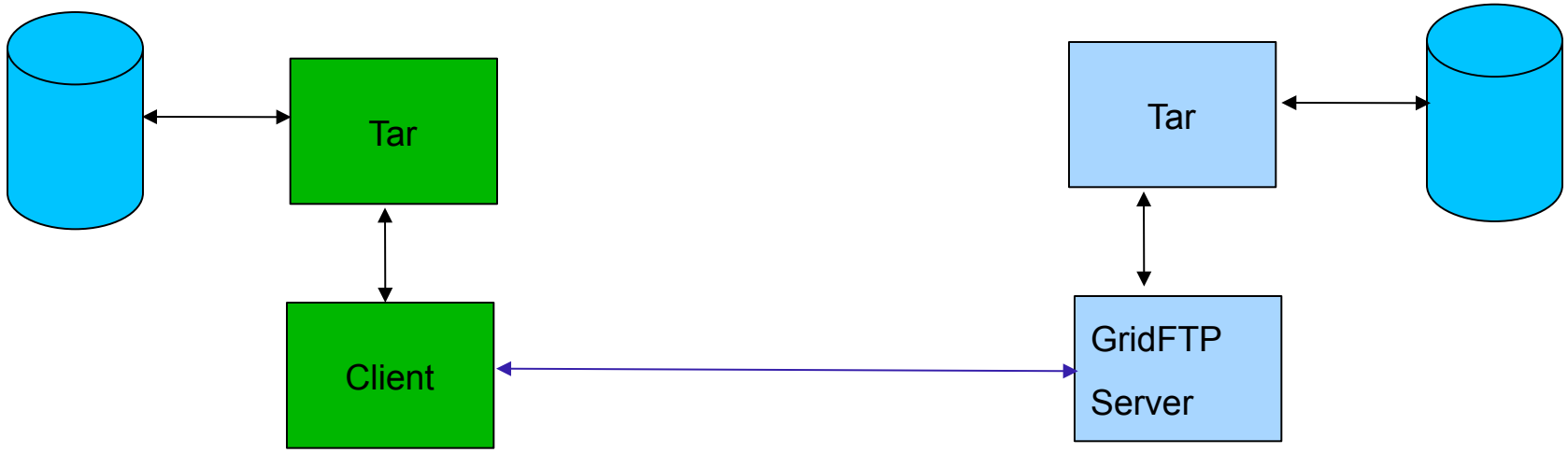
Traditional



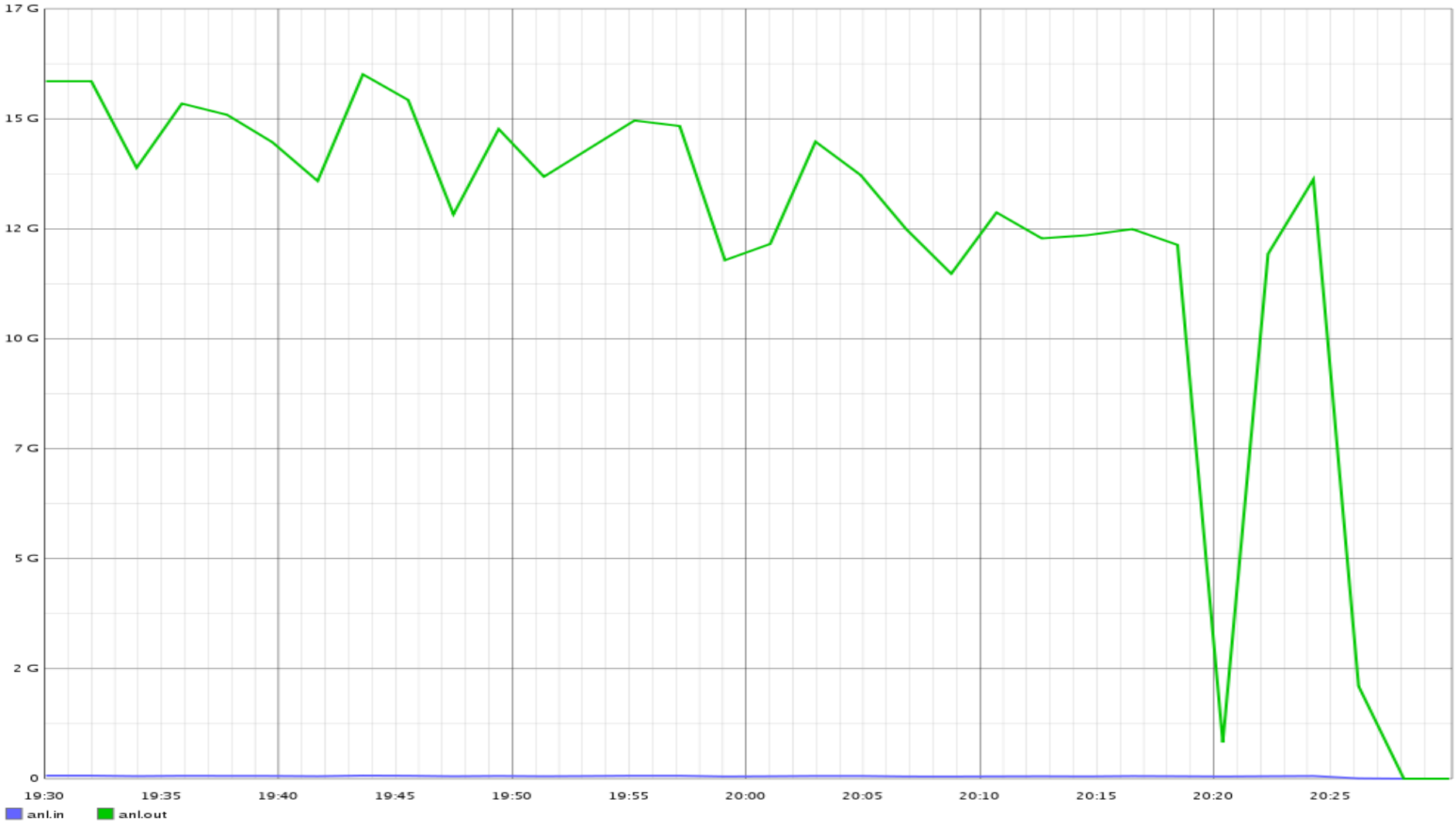
Pipelining



On-the-fly tar



Performance



Security

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

Challenges

- Past success
- Standards
- Robustness
- Throughput
- Future
- Security
- Fire and Forget
- Firewalls

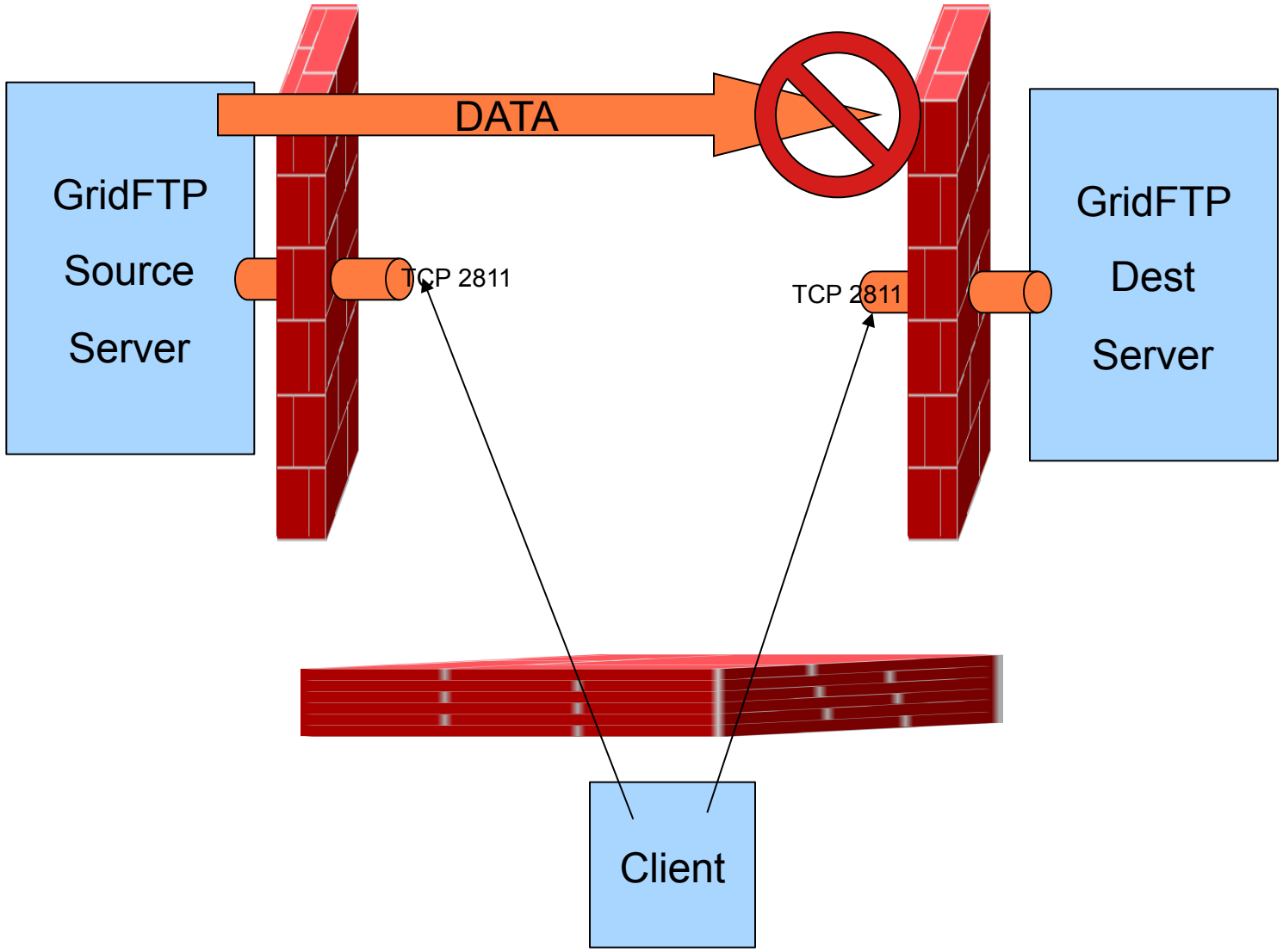
New features

- Sync feature – GT 5.0.2
  - Size, mod time, checksum
- Checkpoint across globus-url-copy sessions – GT 5.0
- Load balancing in globus-url-copy – GT 5.0
- GridFTP for cygwin – GT 5.0
- Chroot GridFTP – GT 5.0.3

Security

- SSH GridFTP limitations
  - Requires SSH access to GridFTP servers
  - No security on the data channel
- Myproxy online CA
  - Use login credentials to obtain certificates
  - These certificates typically not trusted elsewhere
- Incommon?

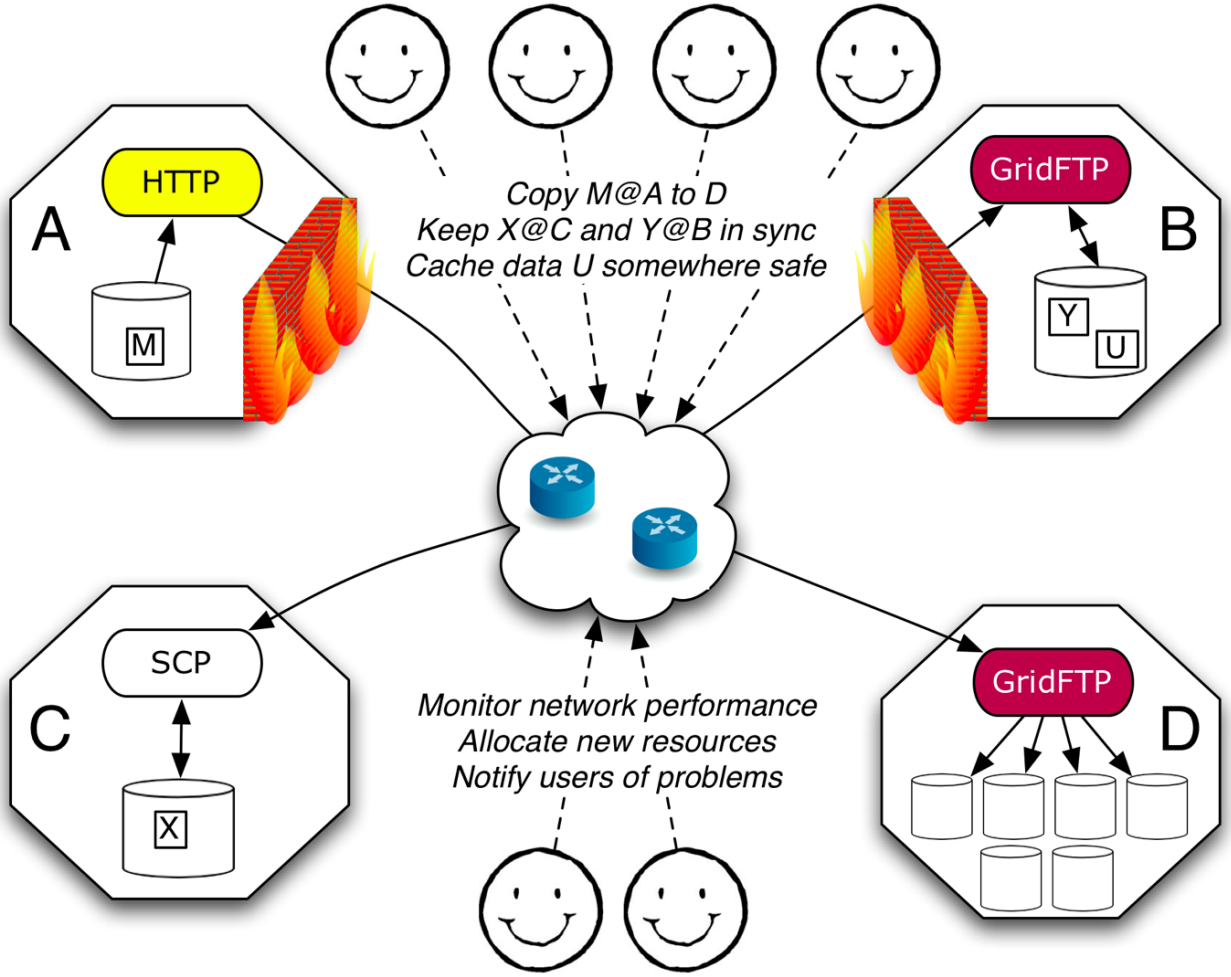
Firewalls


DATA
GridFTP
Source
Server
TCP 2811
TCP 2811
GridFTP
Dest
Server
Client

Fire and Forget transfers

- Data movement is not a fun activity for users
  - Minimize time spent
  - Reduce user intervention as much as possible
- Automatic retries on failures
  - Handle server, network and client failures
  - Server provides checkpoint information
  - Support partial transfer

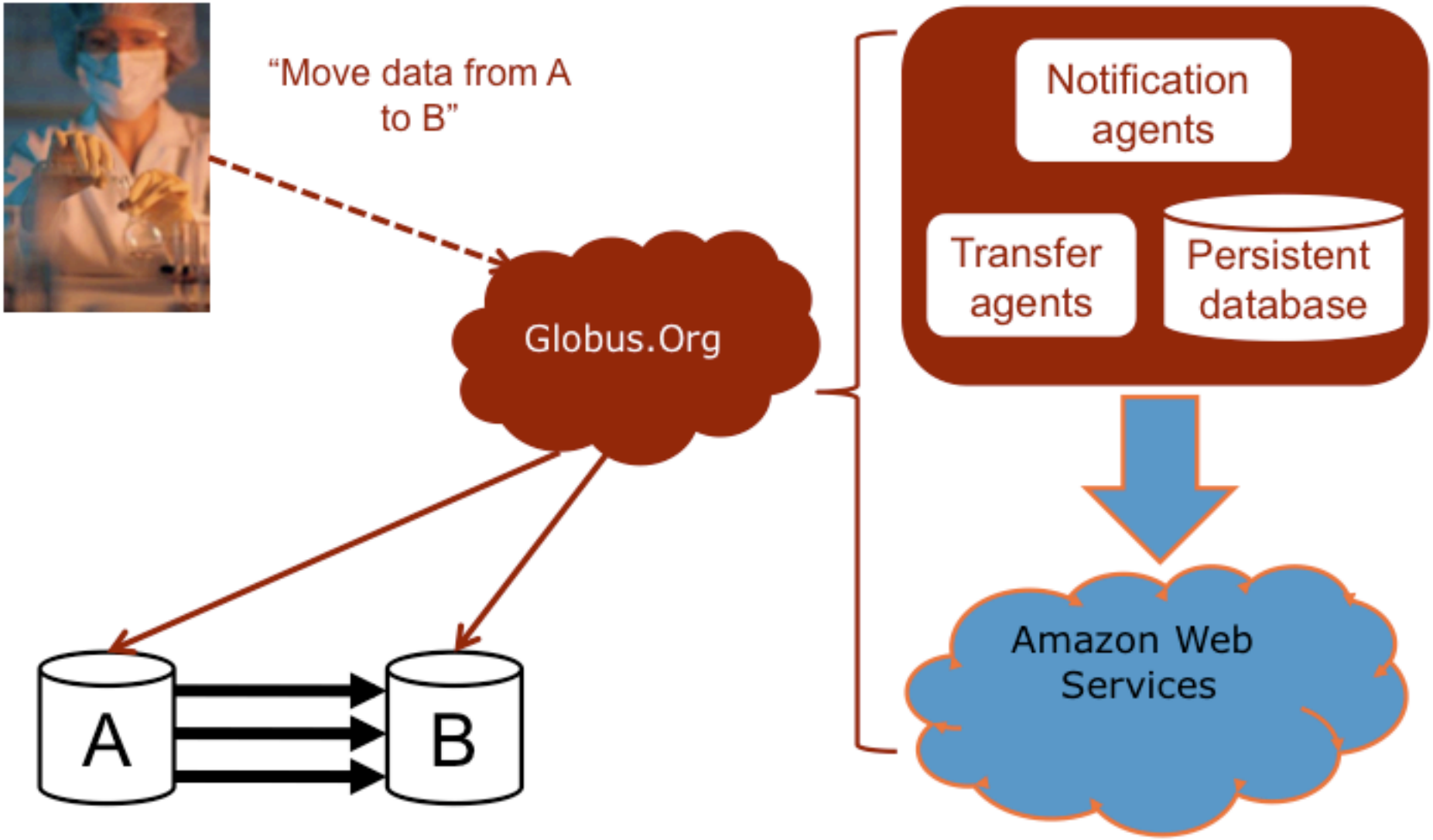
Globus.org – hosted data movement service

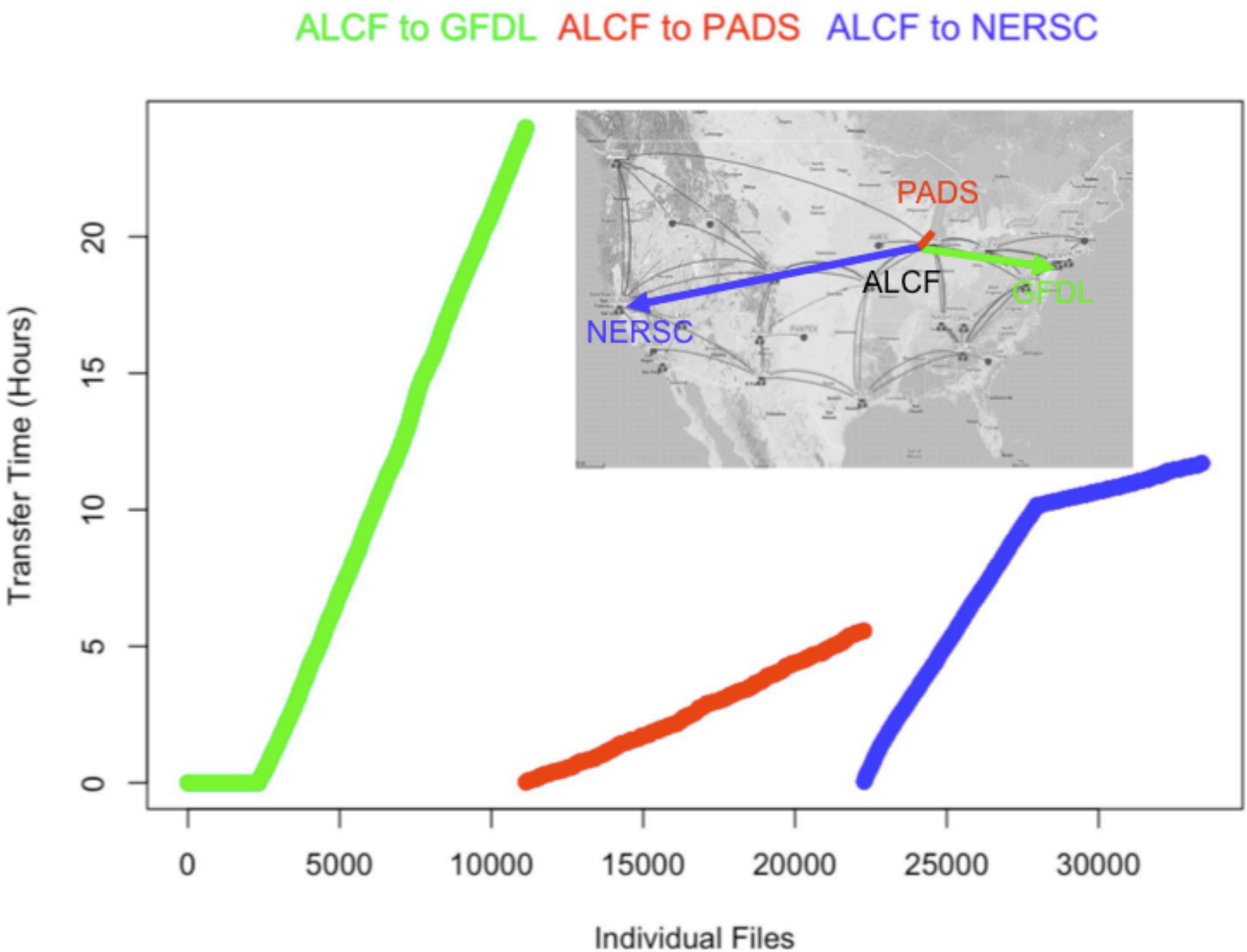


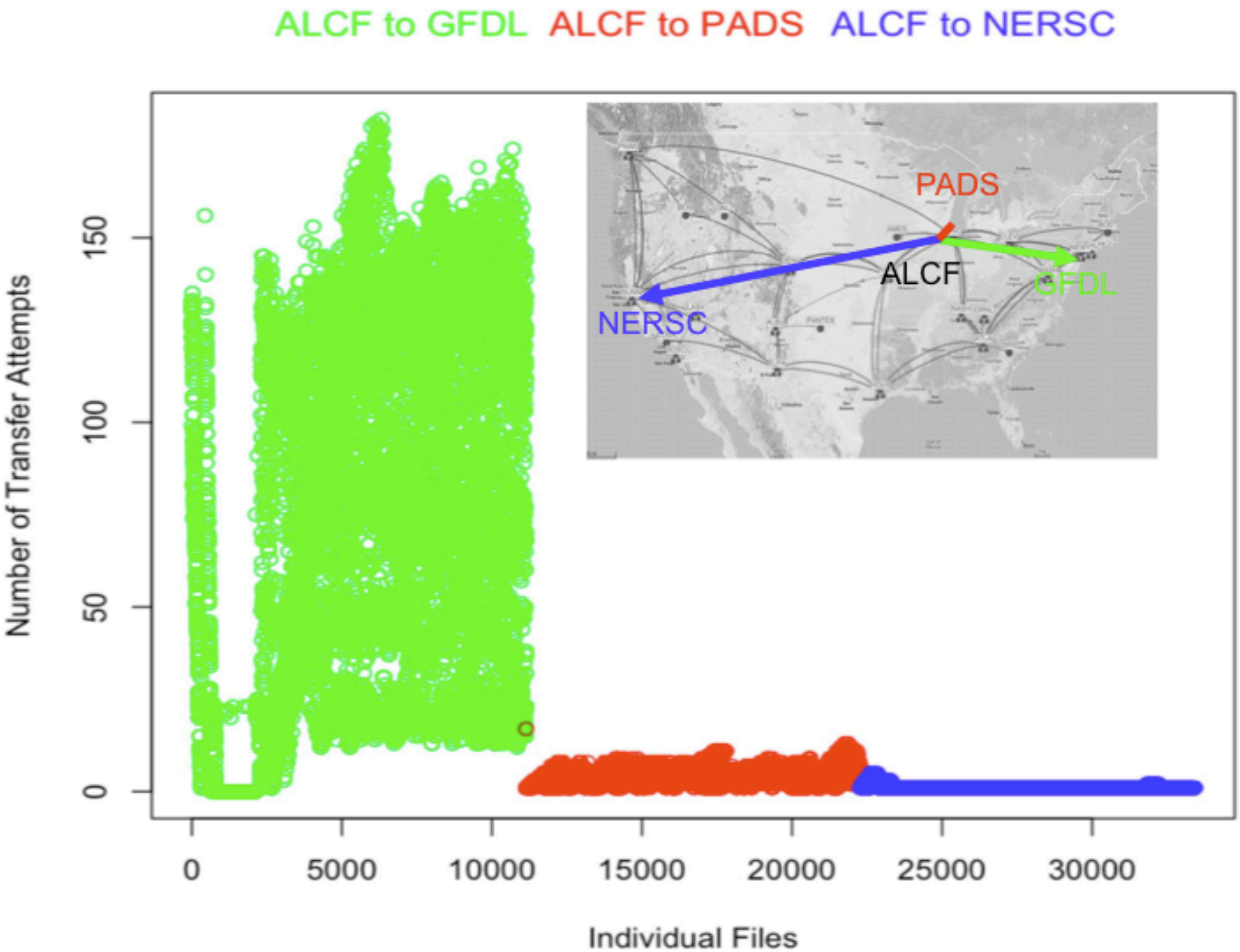
Globus.org

- 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

Globus.org hosted services: Data replication as an example




ALCF to GFDL
ALCF to PADS
ALCF to NERSC
PADS
ALCF
GFDL
NERSC
Transfer Time (Hours)
0
5
10
15
20
0
5000
10000
15000
20000
25000
30000
Individual Files


ALCF to GFDL
ALCF to PADS
ALCF to NERSC
PADS
ALCF
GFDL
NERSC
Number of Transfer Attempts
0
50
100
150
0
5000
10000
15000
20000
25000
30000
Individual Files

the globus alliance
www.globus.org

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

Questions