



A Data Management Framework for Distributed Biomedical Research Environments

Raj Kettimuthu
Argonne National Laboratory and
The University of Chicago

Biomedical Research Environments

- Increasingly depends on access to and analysis of distributed medical and biomedical data
- Datasets are often collected at multiple locations
 - ◆ Difficult to recruit required patient populations at one location
- Strict firewall restrictions
 - ◆ Data repositories as well as researchers and physicians

Datasets

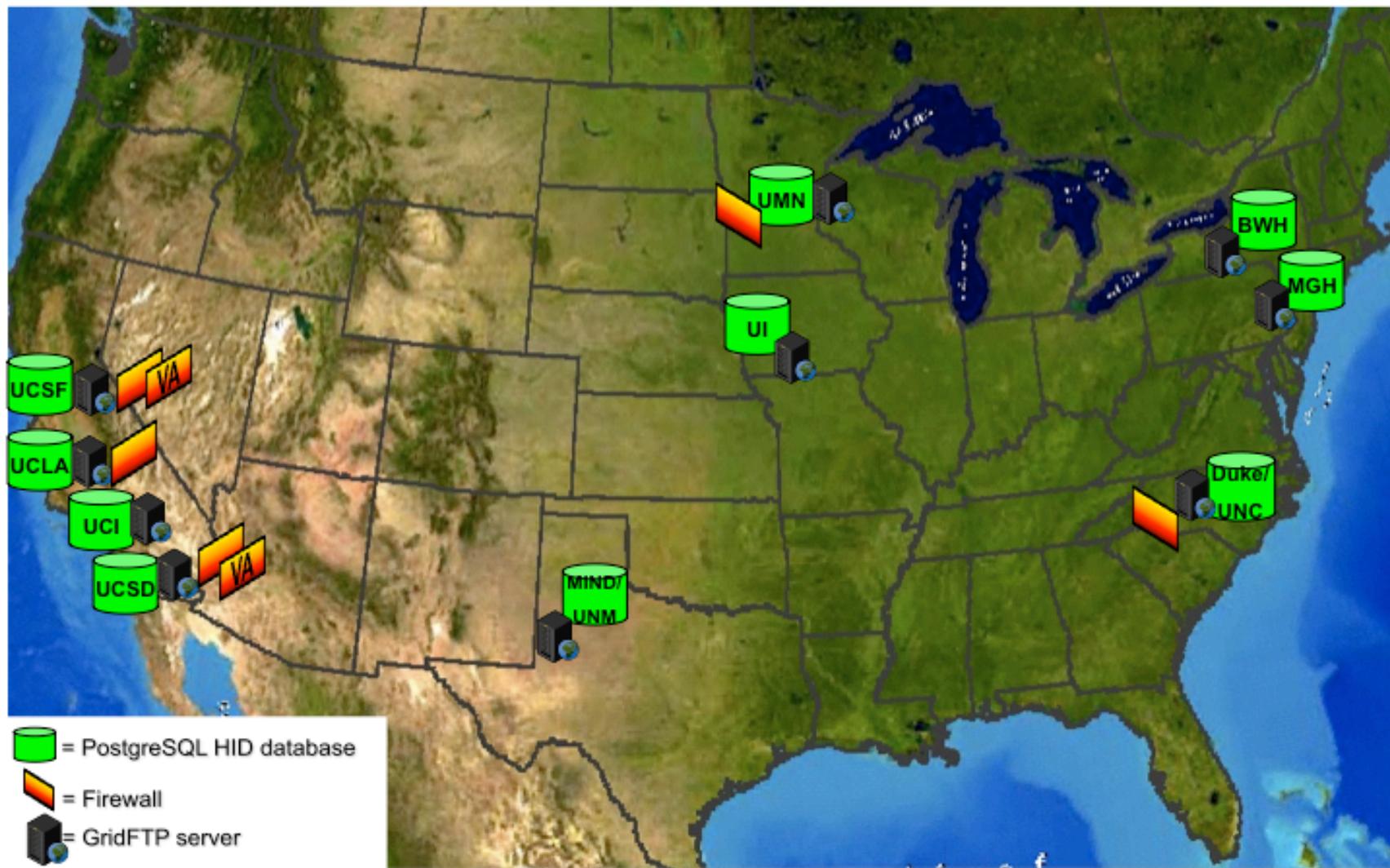
- The datasets are often large (terabytes) and may contain a large numbers (millions) of small files
- These dataset characteristics, combined with strict firewall rules, present significant challenges in transferring these datasets over wide area networks reliably and efficiently



the globus alliance

www.globus.org

FBIRN



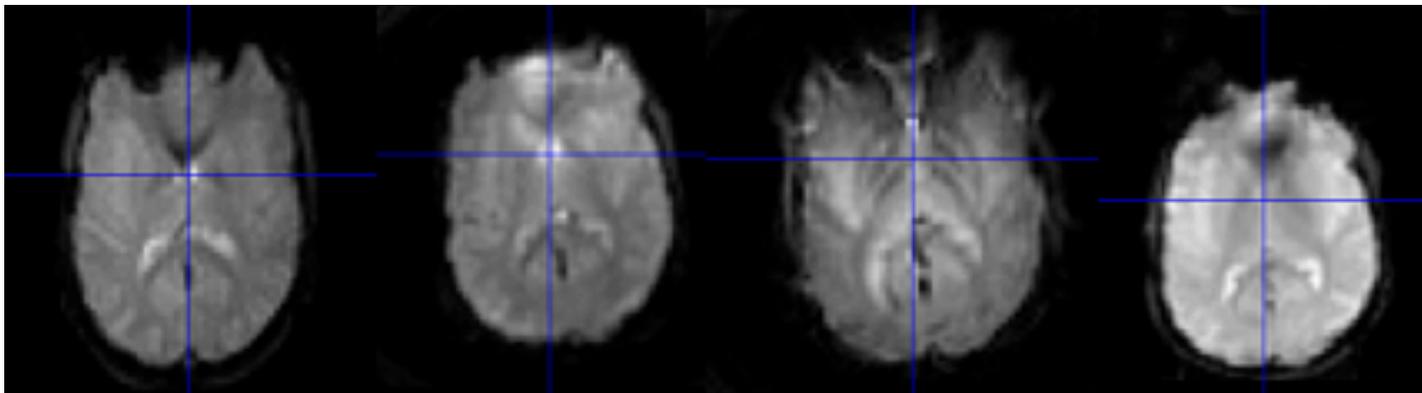
12/07/2010

eScience 2010



FBIRN

- Same person, scanned at MRI centers using different machines and protocols, does not produce the same picture.
- FBIRN is testing improved methods for collecting, sharing, and analyzing multisite fMRI data in clinical populations.





Requirements

- High success and with verification for consistency
- Failures at other sites must not prevent users from accessing their own local data
- Experiment will produce over a million files with size of 100s of bytes to several MBs
- System must store files in a manner that is interoperable with conventional systems
- Users part of one or more groups, system must support group-based access control

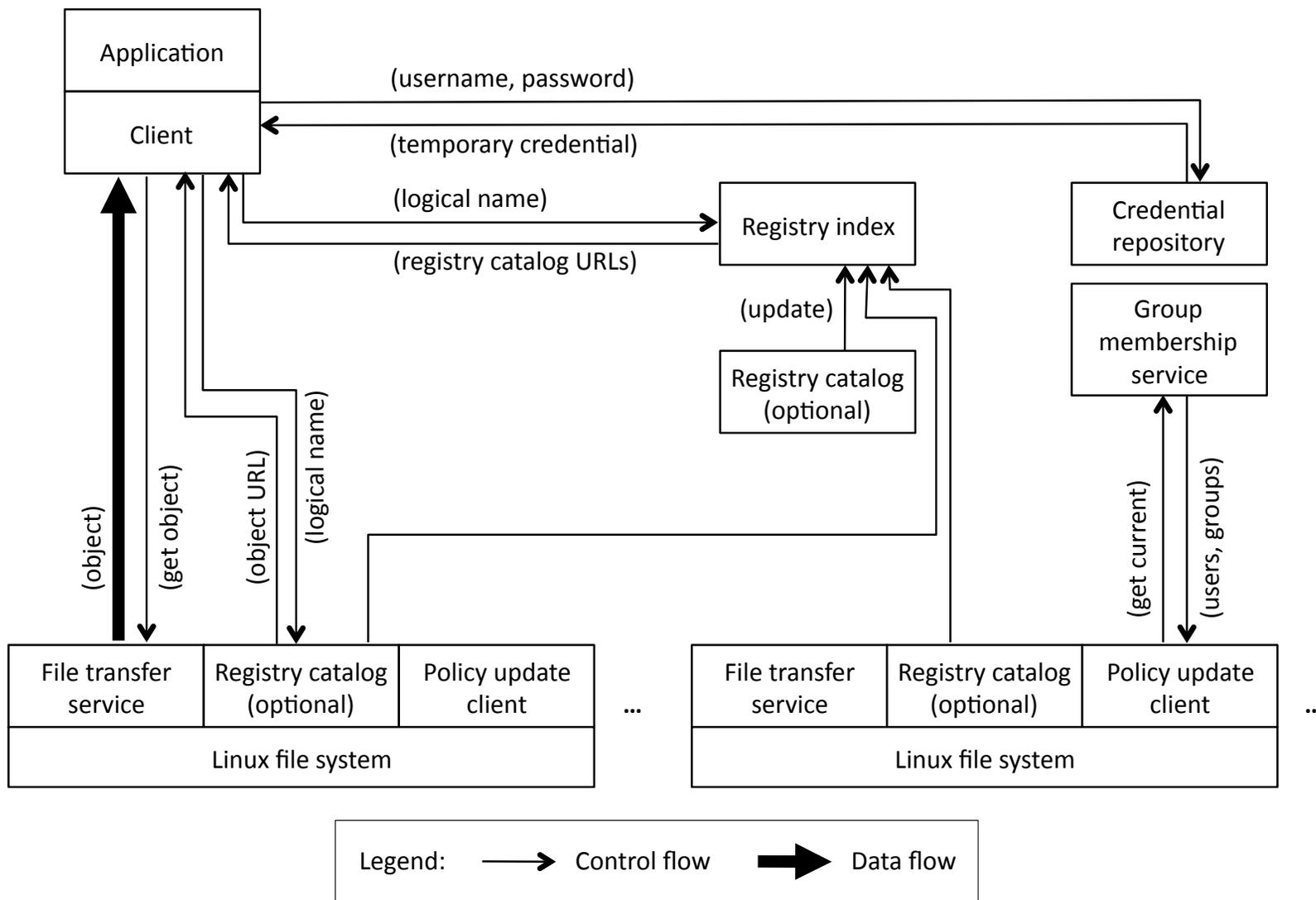


Requirements

- Project lifecycle - data capture phase - quality assurance phase - data access phase
- Support multiple concurrent projects that may be at different phases in lifecycle.
- Data capture - small number of large bulk writes to local storage.
 - ◆ Multiple writes to dataset not common
- Data access - several local and remote bulk data reads
 - ◆ Multiple concurrent reads common



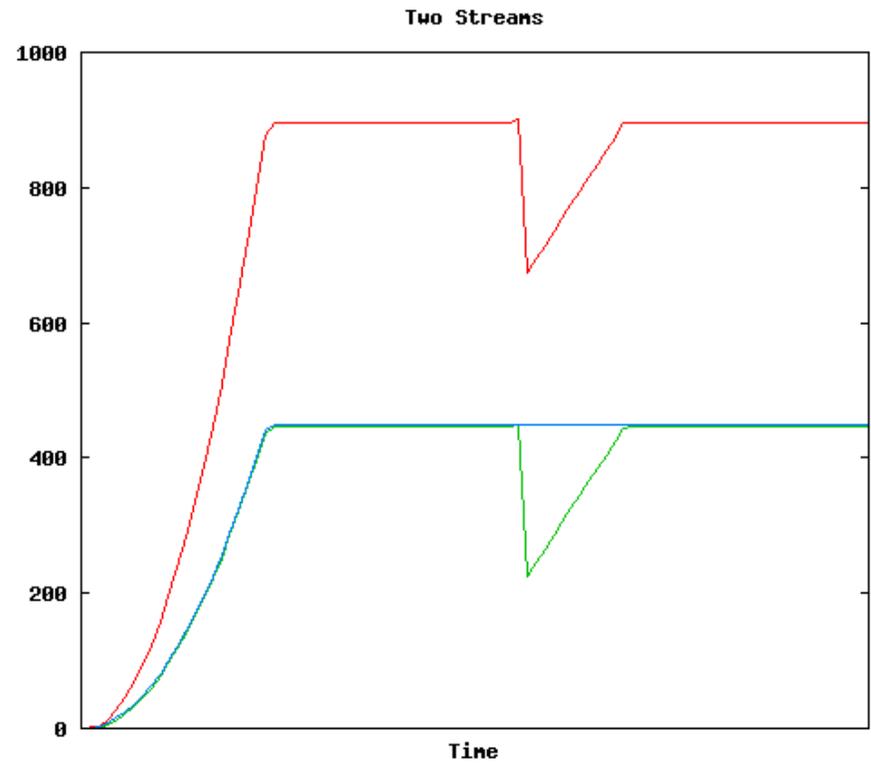
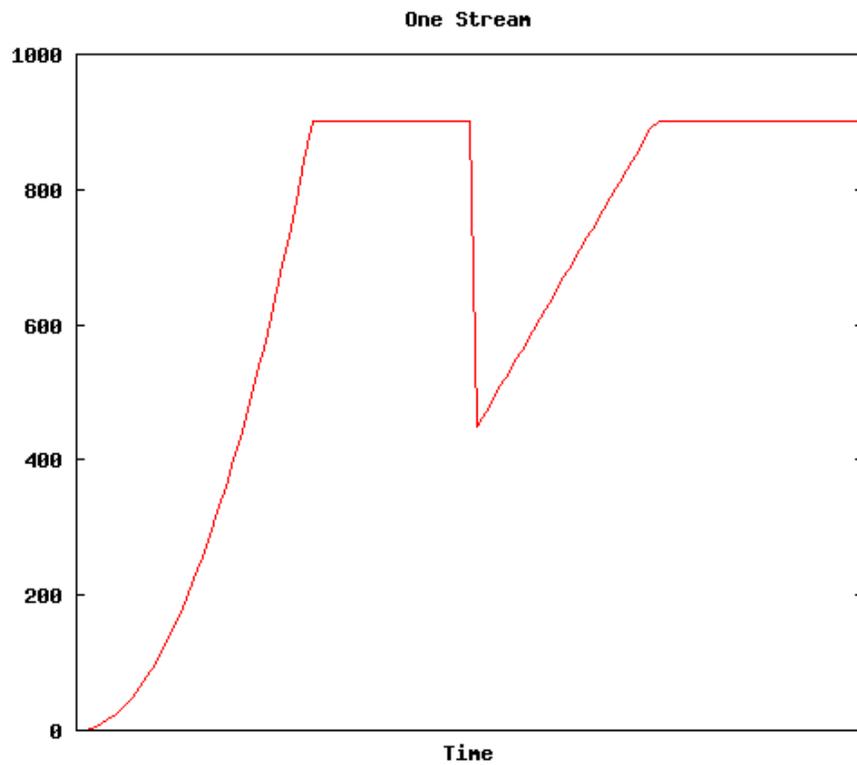
Data Management Architecture



GridFTP

- High-performance, reliable data transfer protocol optimized for high-bandwidth wide-area networks
- Globus GridFTP
 - ◆ Parallel TCP streams, optimal TCP buffer
 - ◆ Non TCP protocol such as UDT
 - ◆ SSH, GSI
 - ◆ Restartable transfers

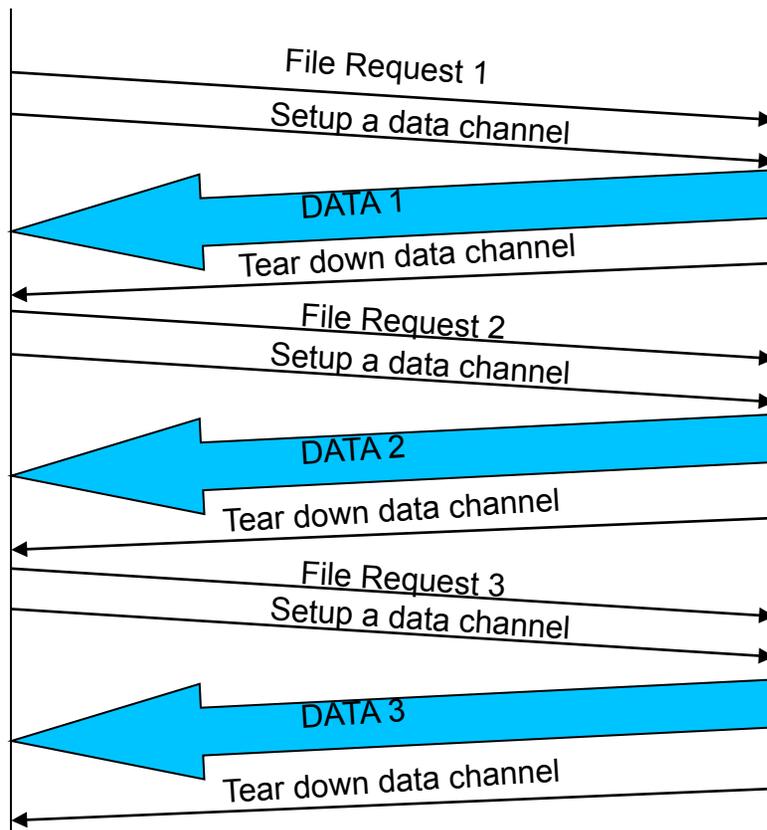
Parallel Streams



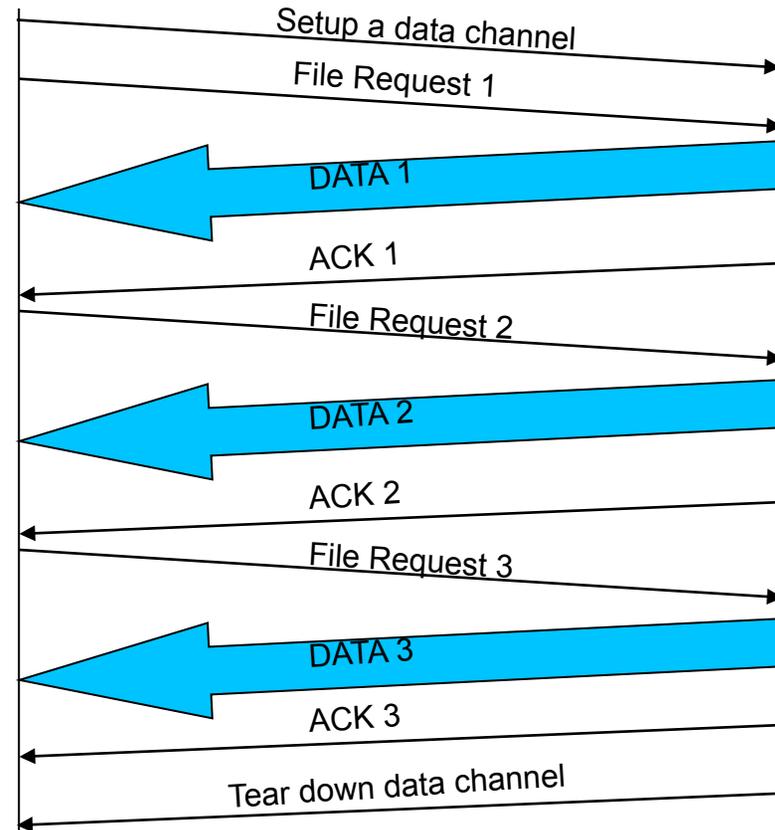


Data Channel Caching

Standard FTP

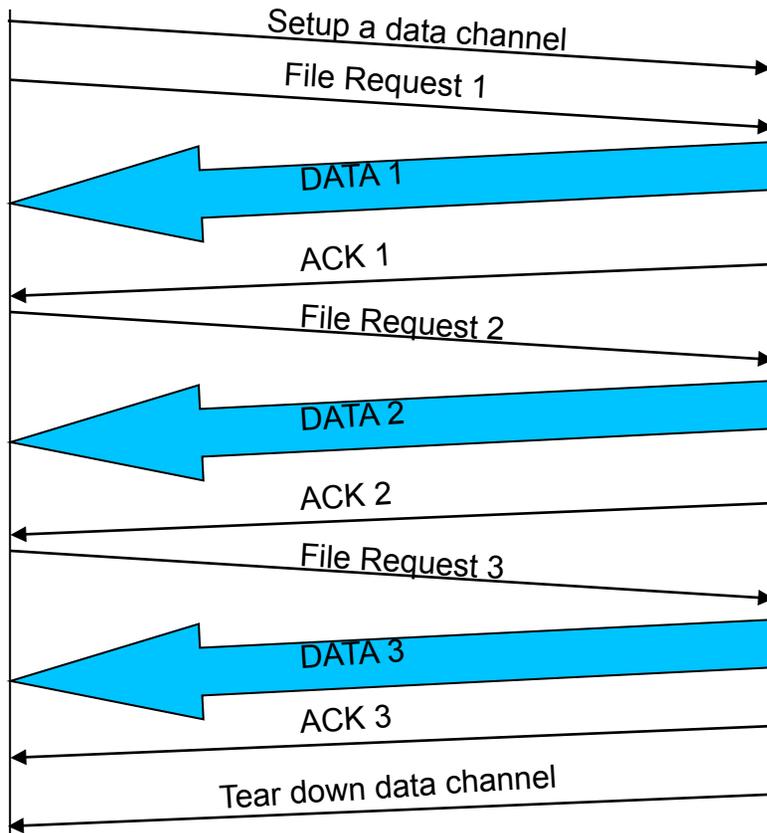


GridFTP

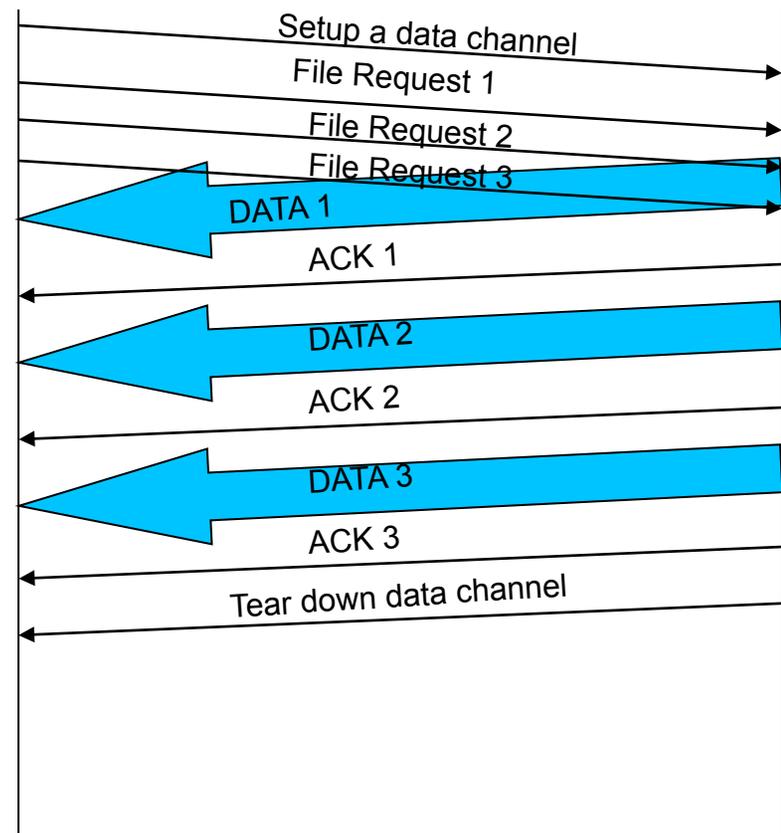


Pipelining

GridFTP



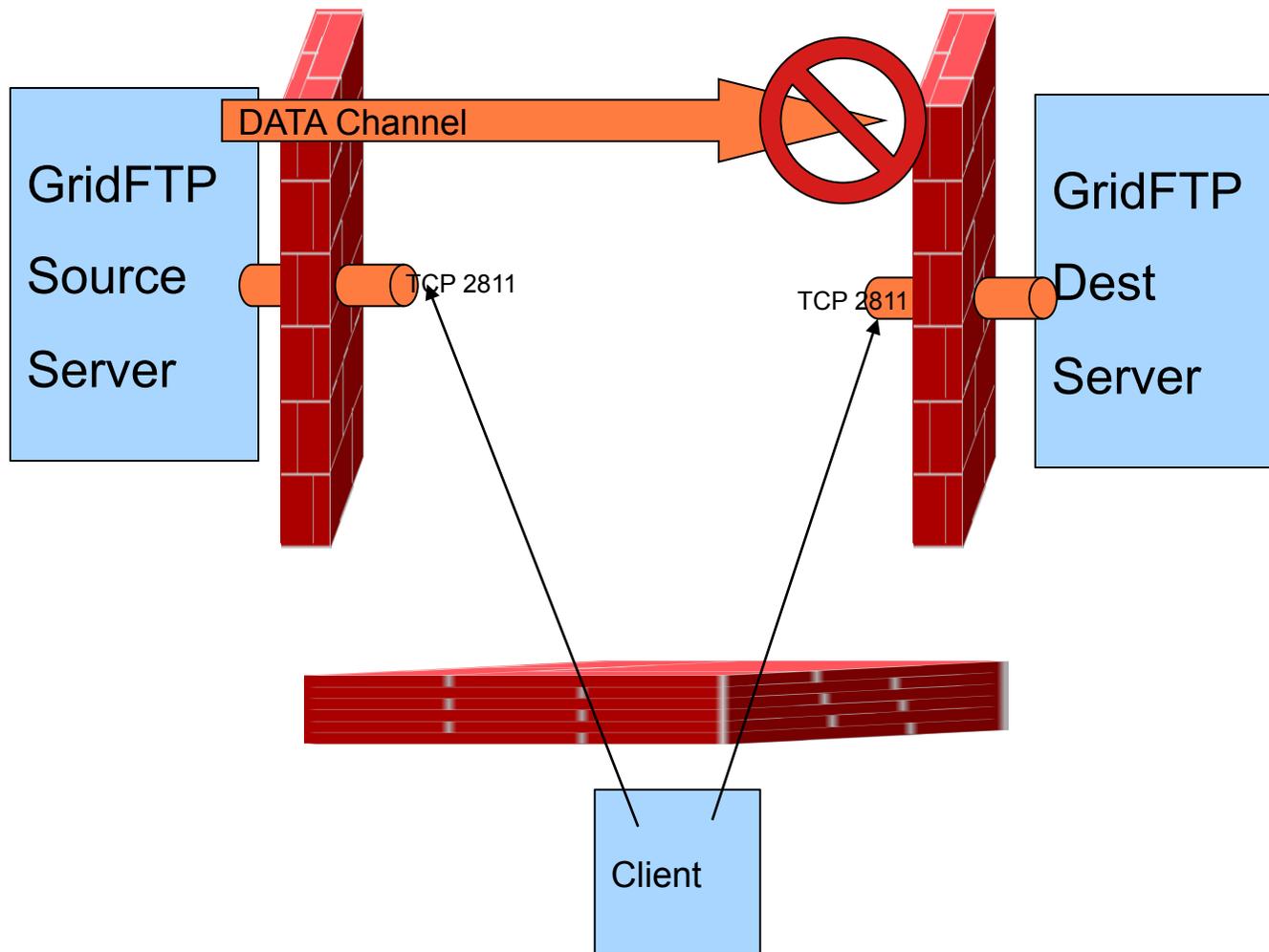
GridFTP with Pipelining





Firewall

- Outgoing allowed at sender, incoming blocked at receiver

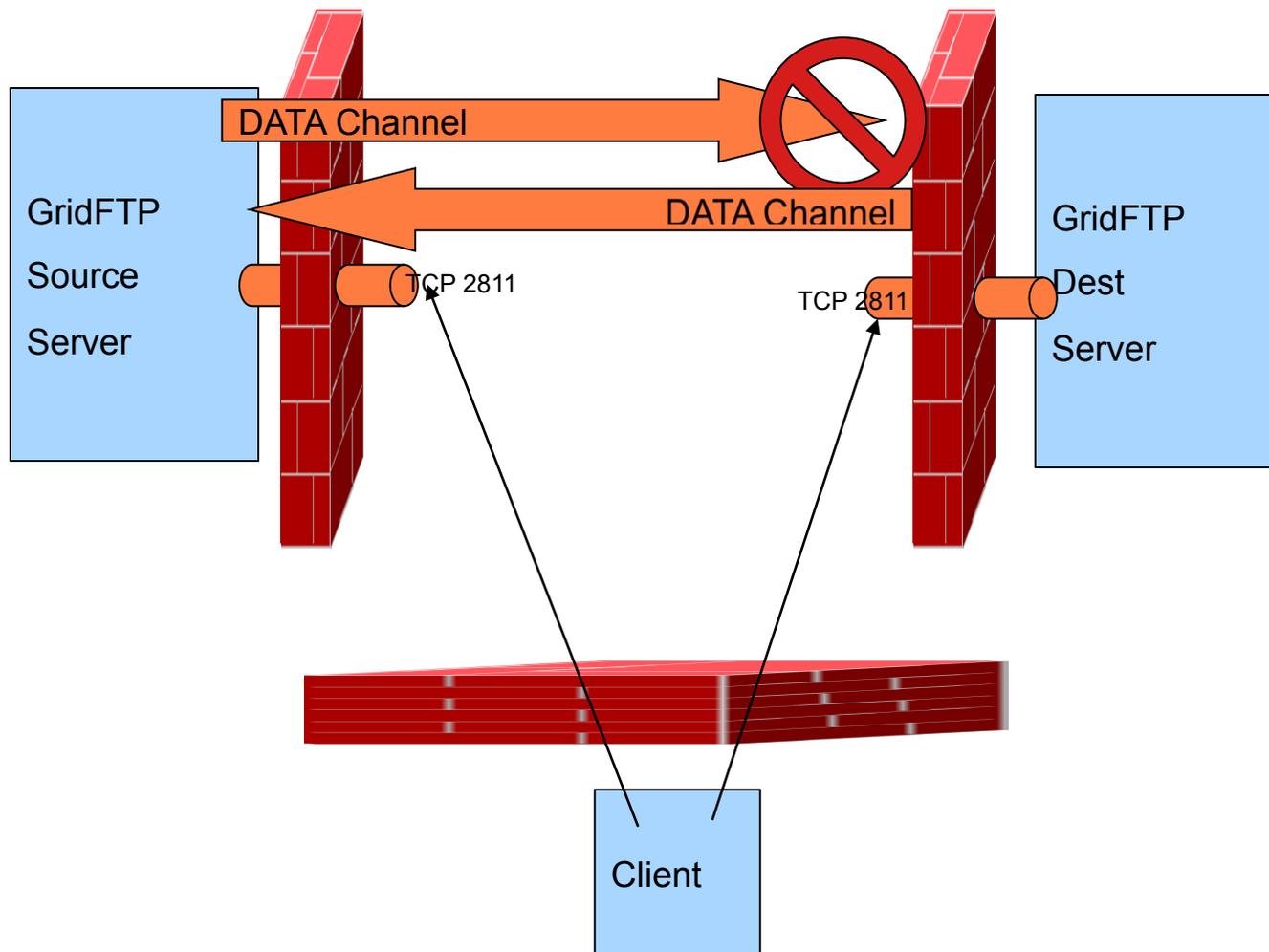




Firewall

- Outgoing allowed at sender, incoming blocked at receiver

Mode S

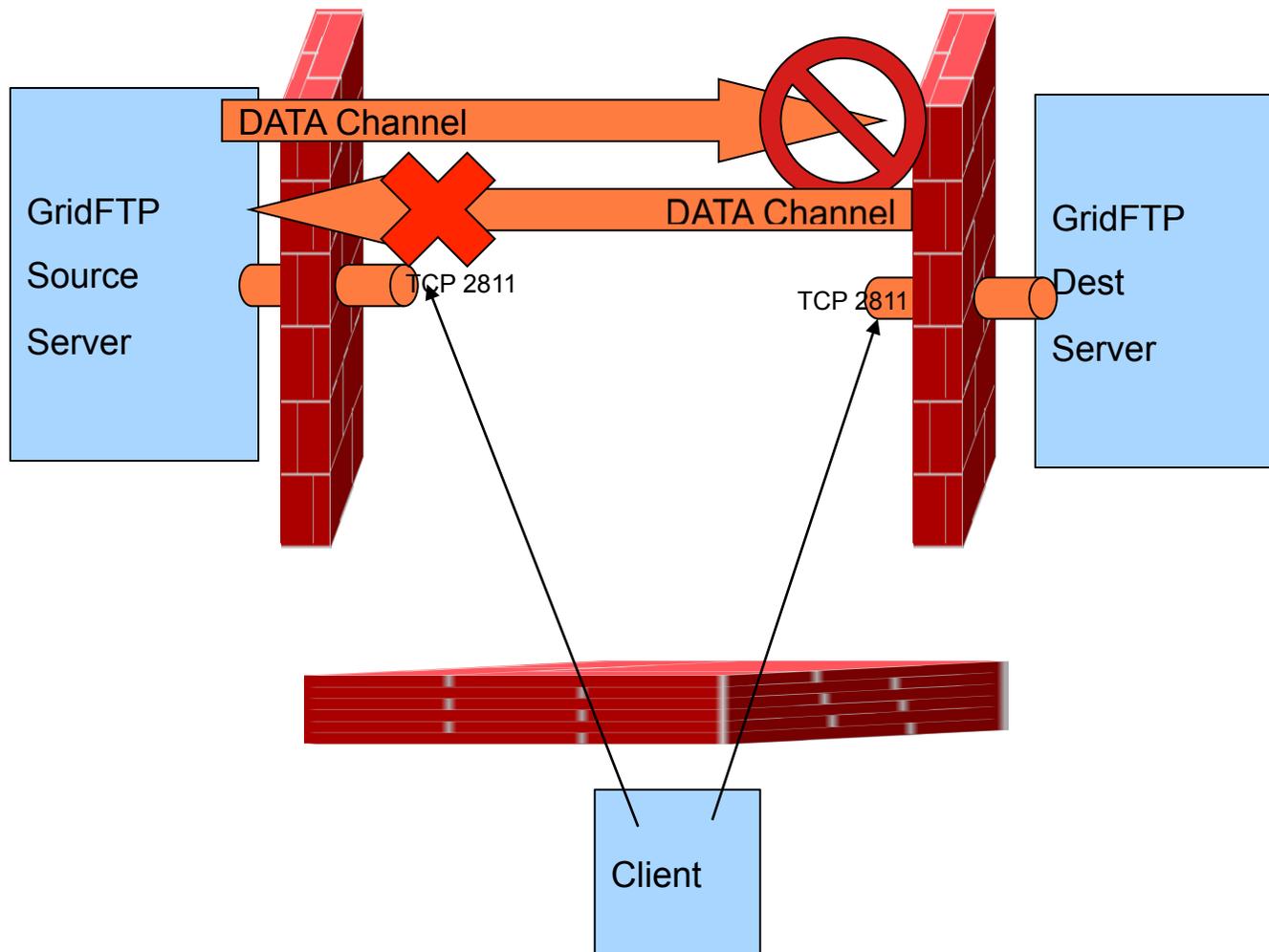




Firewall

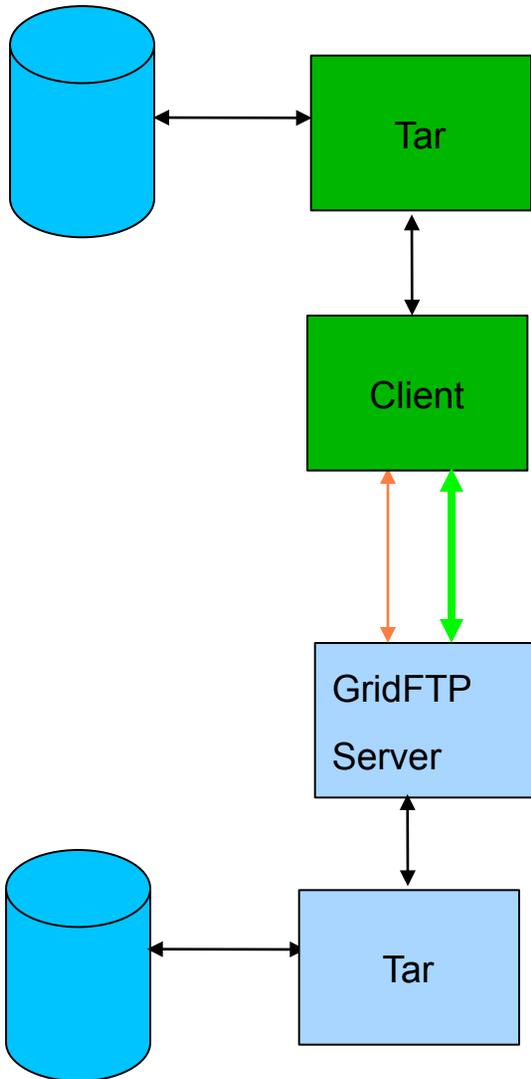
- Outgoing allowed at sender, incoming blocked at receiver

Mode E

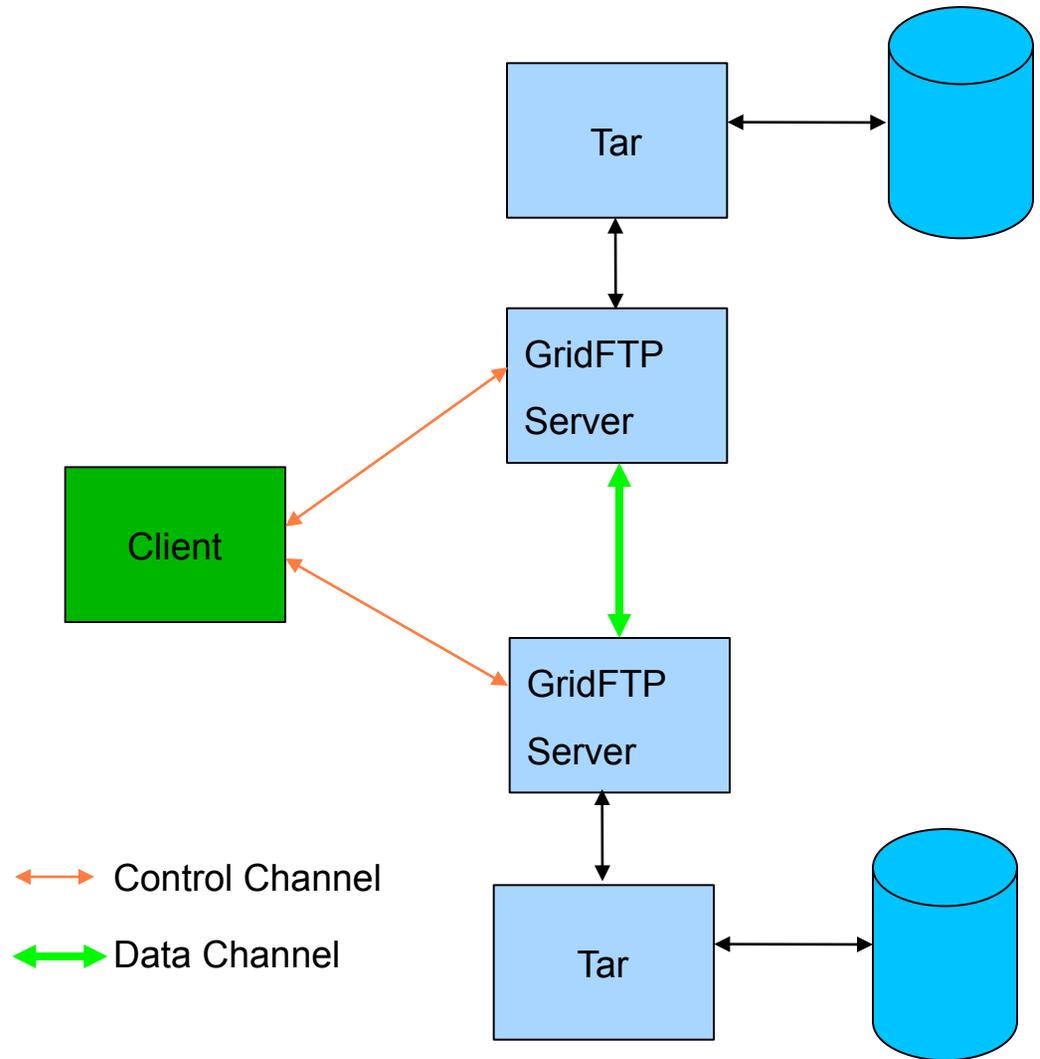


On-the-fly tar

Client – Server Transfer



Server – Server Transfer





the globus alliance

www.globus.org

Transfer characteristics of tar-enabled GridFTP

Concurrency	Size (GB)	Speed (MB/s)	Success Count	Failure Count	Success Rate(%)
1	2	15.44	88	0	100
1	21	10.41	22	0	100
3	2	9.12	174	0	100
3	21	6.77	27	0	100
5	2	6.43	85	0	100
5	21	5.41	45	0	100
7	2	5.48	84	0	100
7	21	4.59	49	0	100
9	2	3.62	90	0	100
9	21	4.40	77	4	95
11	2	2.96	100	10	91
11	21	2.78	60	17	78



Control Channel Disconnect

- No data over the control channel after the connection has been established and while data are being sent
 - ◆ Only after the data is transferred - control channel used to complete the transfer
- Control channel could be idle for long time
 - ◆ When clients or servers behind a NAT proxy or a firewall, connections get disconnected
 - ◆ Caused problems with long-running transfers, control channel connection was dropped silently - transfers hung and failed.



Control Channel Disconnect

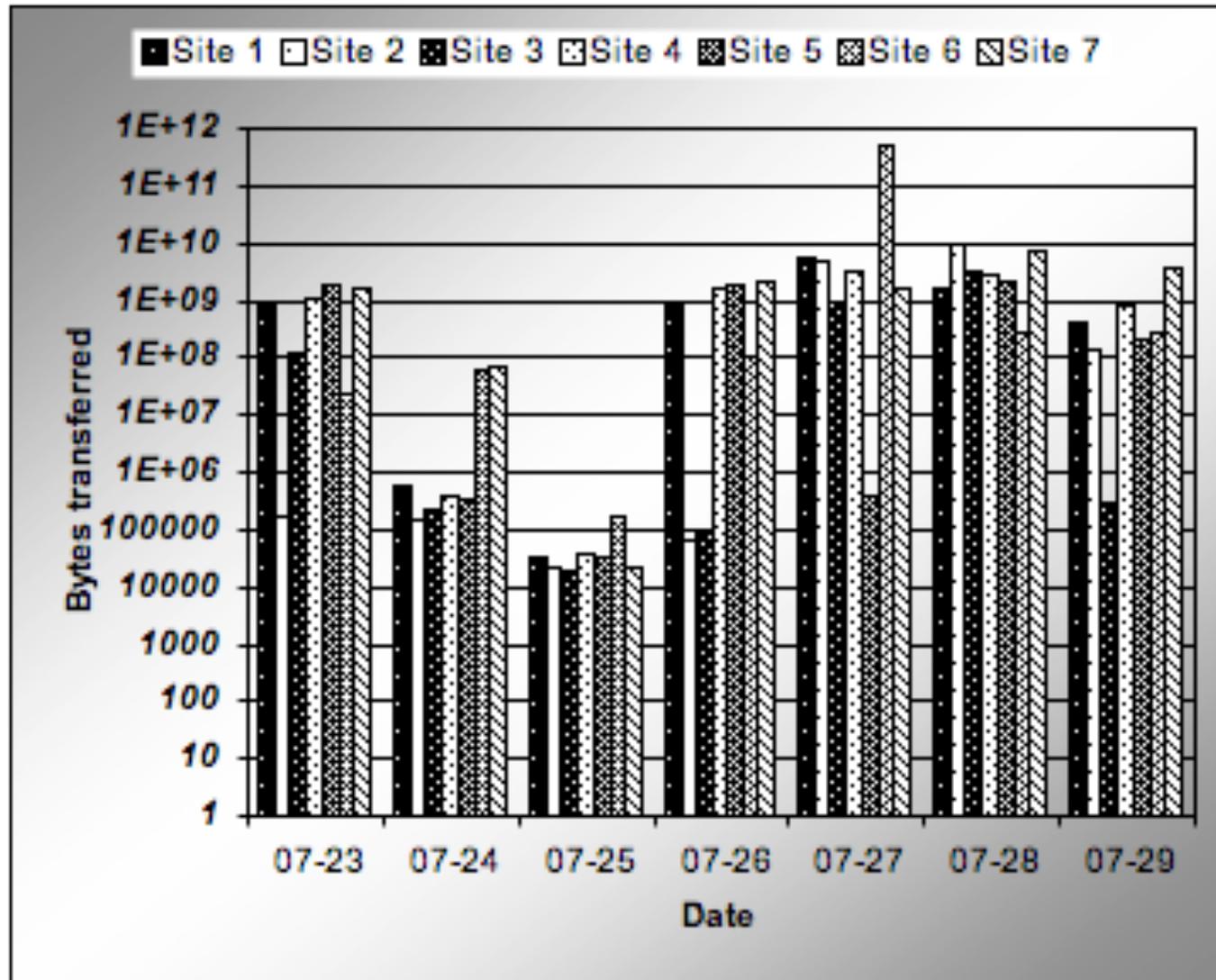
- Solution - we leverage the TCP keepalive mechanism to avoid connection dropping
 - ◆ “probes” are transmitted to network peers at a configurable time interval
- Linux has built-in support for TCP keepalive
 - ◆ Default settings exceed the time limit used by proxies and firewalls
- With TCP keepalive configured at frequent interval, connection dropping was avoided.
 - ◆ All tests listed in Table 1 succeeded



the globus alliance

www.globus.org

FBIRN Data Movement Volume





Continuous Monitoring

- FBIRN scientists need the data transfer services to behave predictably:
 - ◆ Performance (good or bad) should meet expectations based on past behavior.
- This requirement led us to two goals:
 - ◆ Stabilize the system's behavior
 - ◆ Measure the system's behavior and publish results to set reasonable expectations among the users.



Automated Test Mechanism

- Set of tests across all of the FBIRN GridFTP servers on a continuous basis
- Two types of tests are performed
 - ◆ A relatively short data transfer between each pair of data centers is performed frequently (multiple times per day)
 - ◆ A relatively long data transfer is performed infrequently (once per day or every other day) to measure the performance of the system.

Continuous Monitoring

From	To	Throughput (MB/s)									
		5/21	5/22	5/23	5/24	5/25	5/26	5/27	5/28	5/29	
Site A	Site C	*	X	X	X	X	X	X	*	*	
	Site D	**	**	X	X	X	X	X	**	*	
	Site E	O	*	O	*	*	*	*	*	O	
	Site F	**	**	**	X	**	**	**	**	**	
	Site G	**	**	**	X	X	**	**	**	**	
	Site B	*	X	X	X	X	X	X	*	*	
	Site H	**	**	**	**	**	**	**	**	*	
	Site I	**	X	X	X	X	X	X	**	**	
Site J	**	**	**	**	**	**	**	**	**		
Site B	Site A	*	*	*	*	*	*	*	*	*	
	Site C	*	*	*	*	*	*	*	*	*	
	Site D	***	***	***	***	***	***	***	***	***	
	Site E	*	*	*	*	*	*	*	*	*	
	Site F	**	**	**	**	**	**	**	**	*	
	Site G	**	**	**	**	*	**	**	**	**	
	Site H	***	***	***	***	***	***	***	***	**	
	Site I	***	***	***	***	***	***	**	**	***	
Site J	***	***	***	***	***	***	***	***	***		

Legend:

* Complete <25 MB/s

** Complete 25-50 MB/s

*** Complete >50 MB/s

X Failed

O Expired

Questions