



Globus Toolkit

Data Movement Challenges in the Era of 100 Gigabit and Terabit Networks

Raj Kettimuthu

Argonne National Laboratory and The University of Chicago



ESNet Today

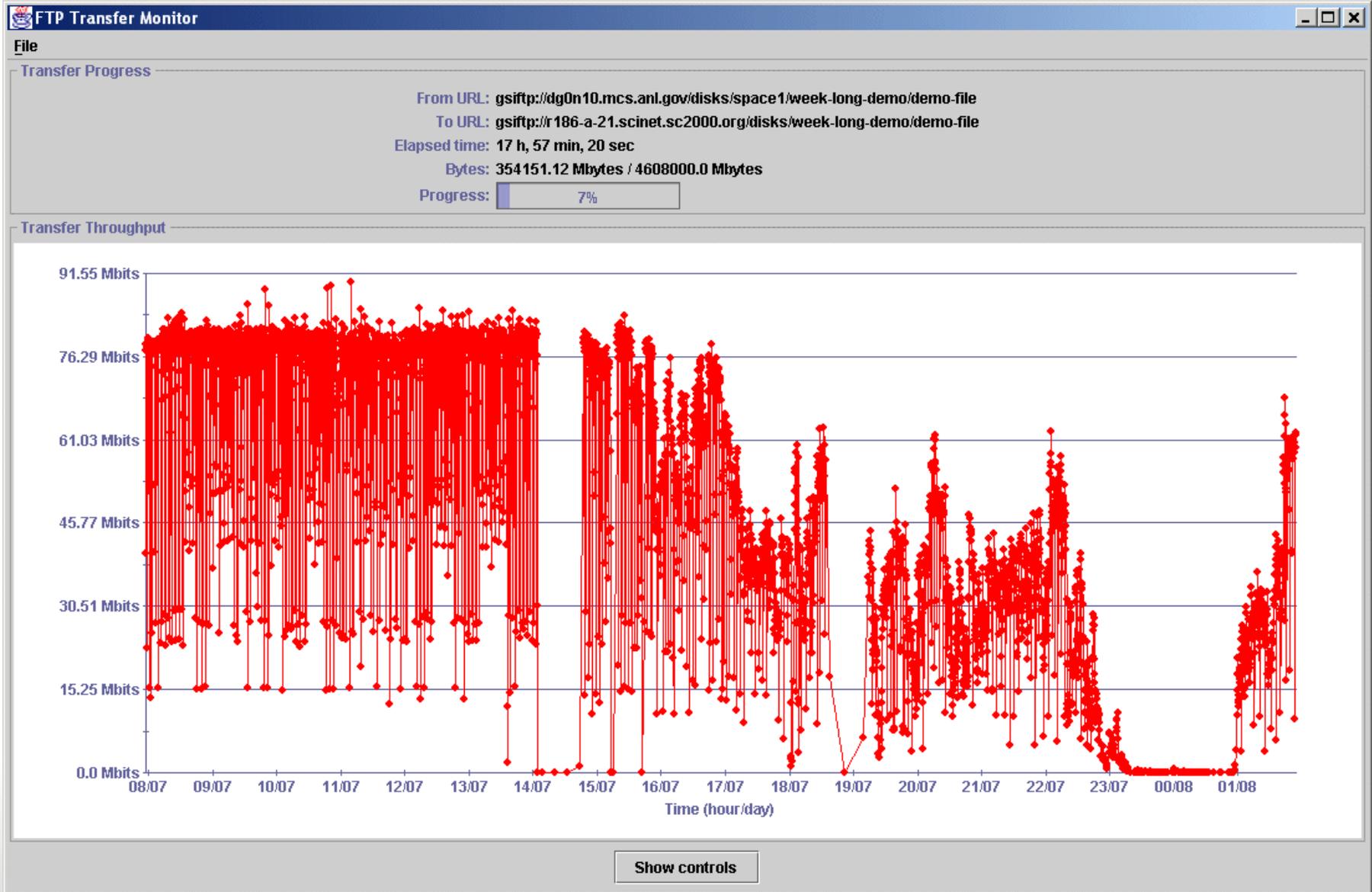
ESnet—Connecting DOE Labs to the World of Science



- Production IP Core (10 Gbps)
- Science Data Network (SDN) Core (20-30-50 Gbps)
- MANs (20-30 Gbps) or Backbone Loops for Site Access
- Major International Connections
- SDN Hubs
- IP Core
- Future Hubs

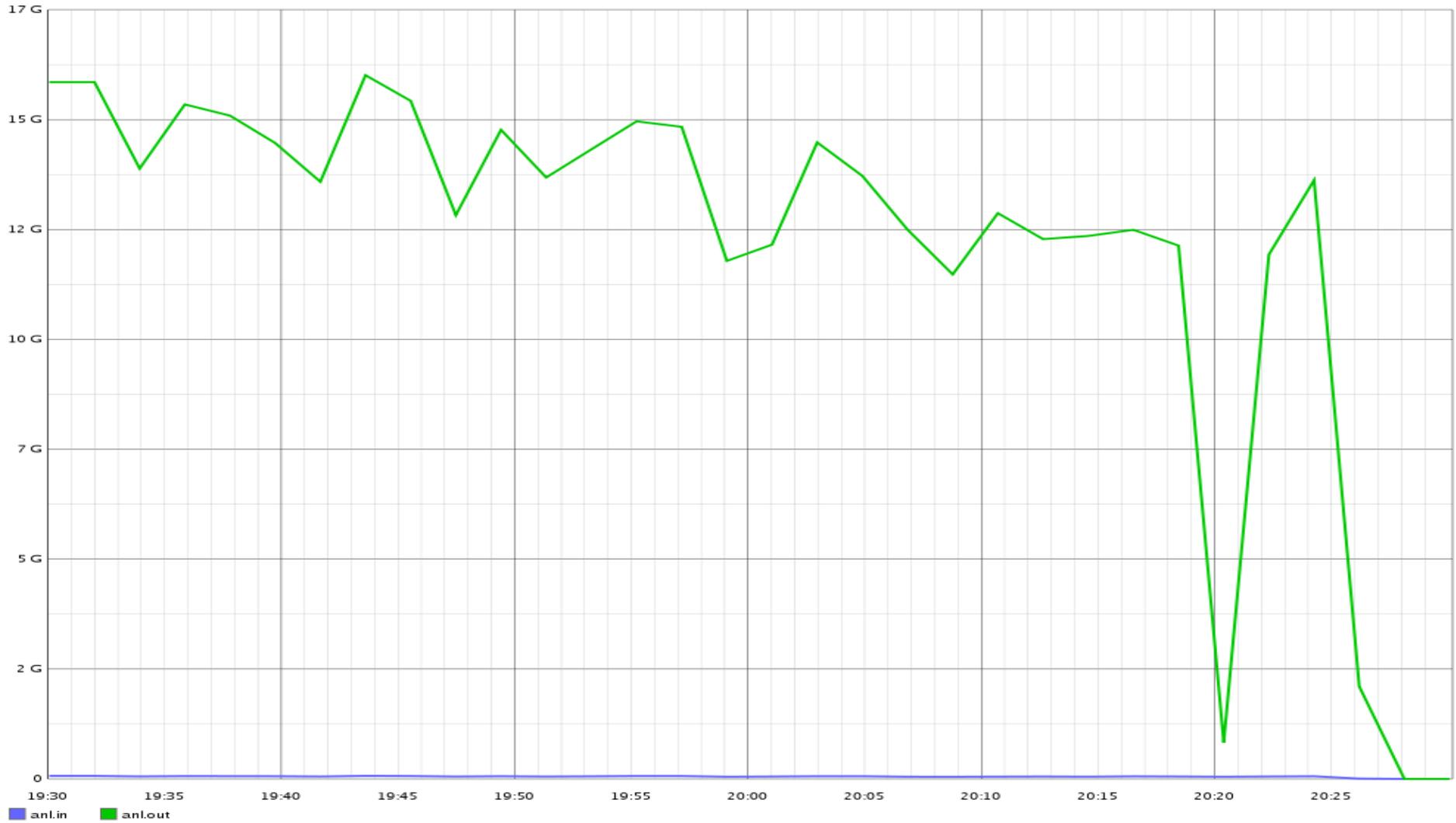


Data Movement in 2000





Data Movement in 2009

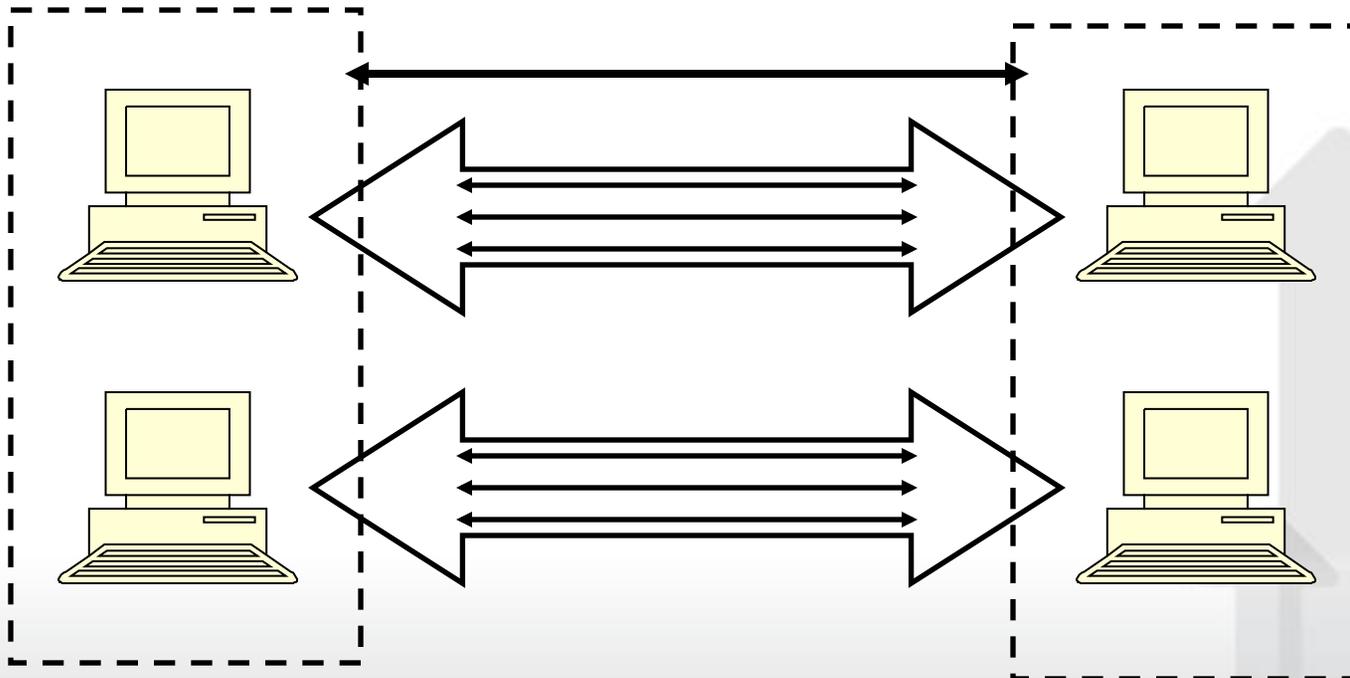
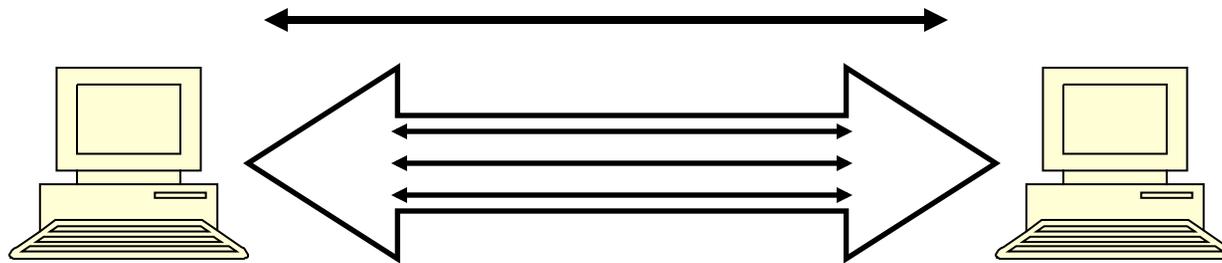




- **High-performance, secure data transfer protocol optimized for high-bandwidth wide-area networks**
- **Backward compatible extension of legacy FTP**
- **Globus GridFTP**
 - Performance
 - Reliability
 - Security

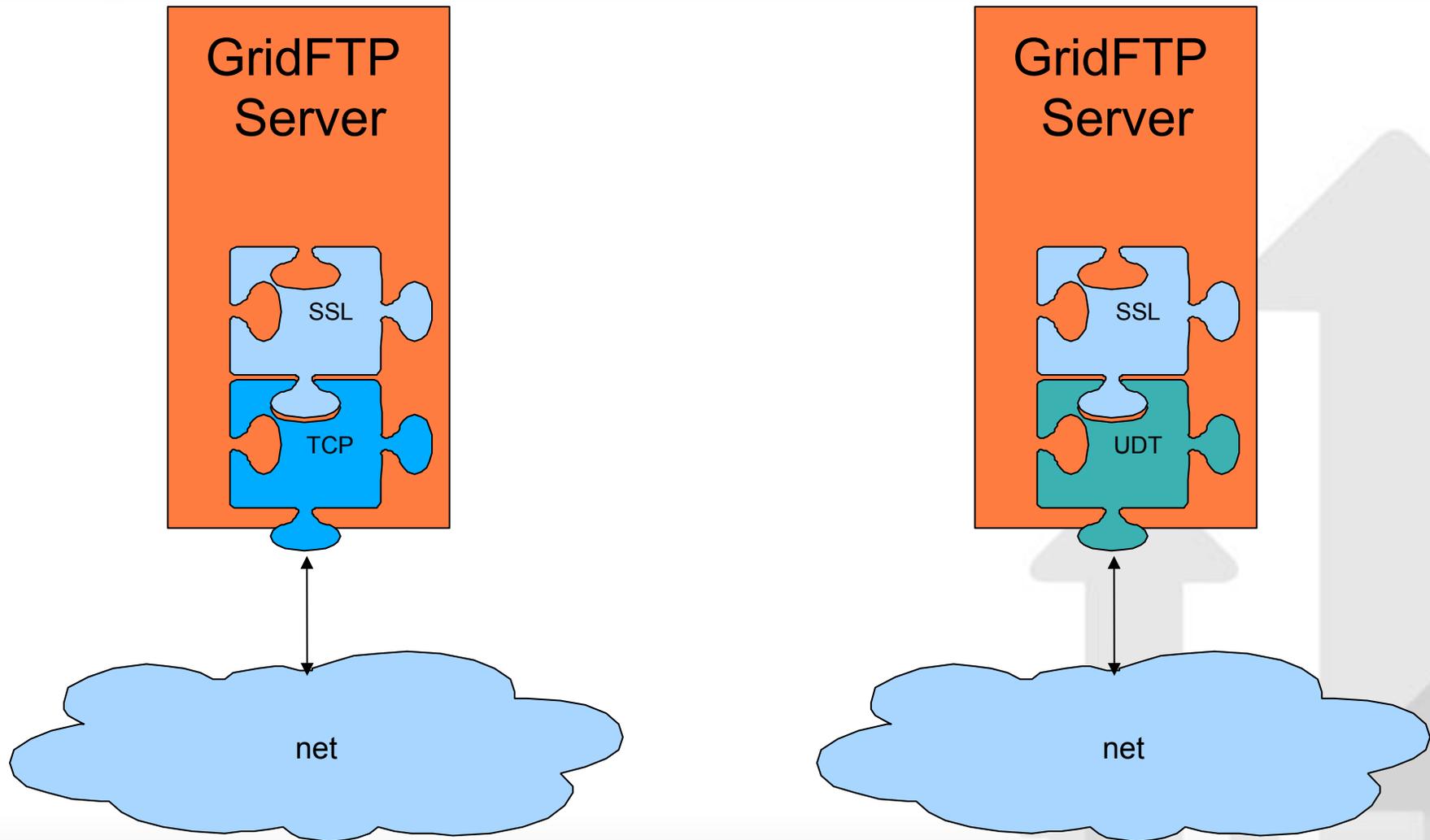


Parallel Streams and Cluster-to-Cluster Data Movement





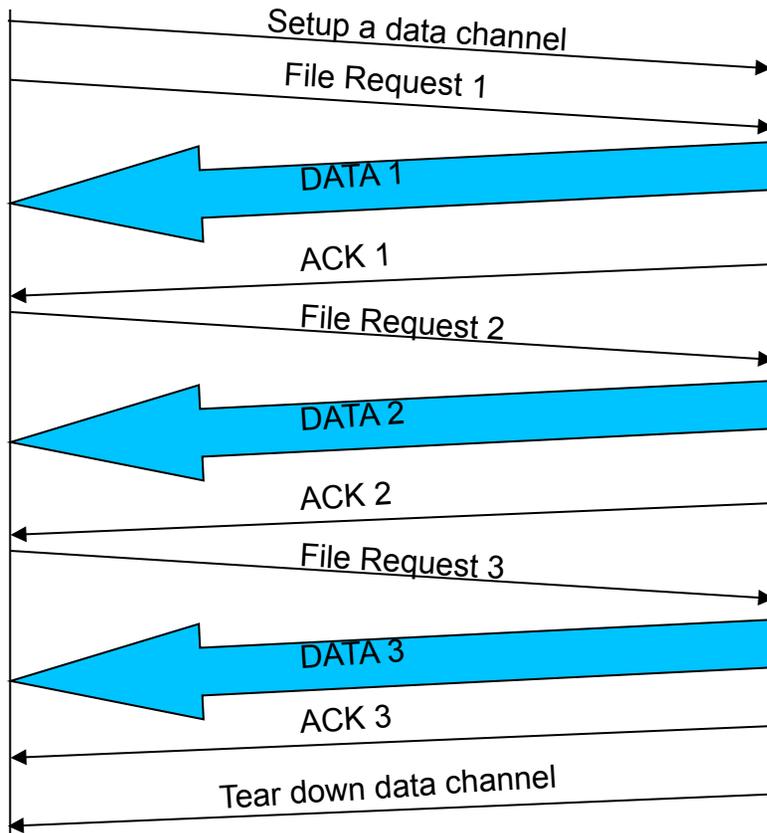
GridFTP over UDT



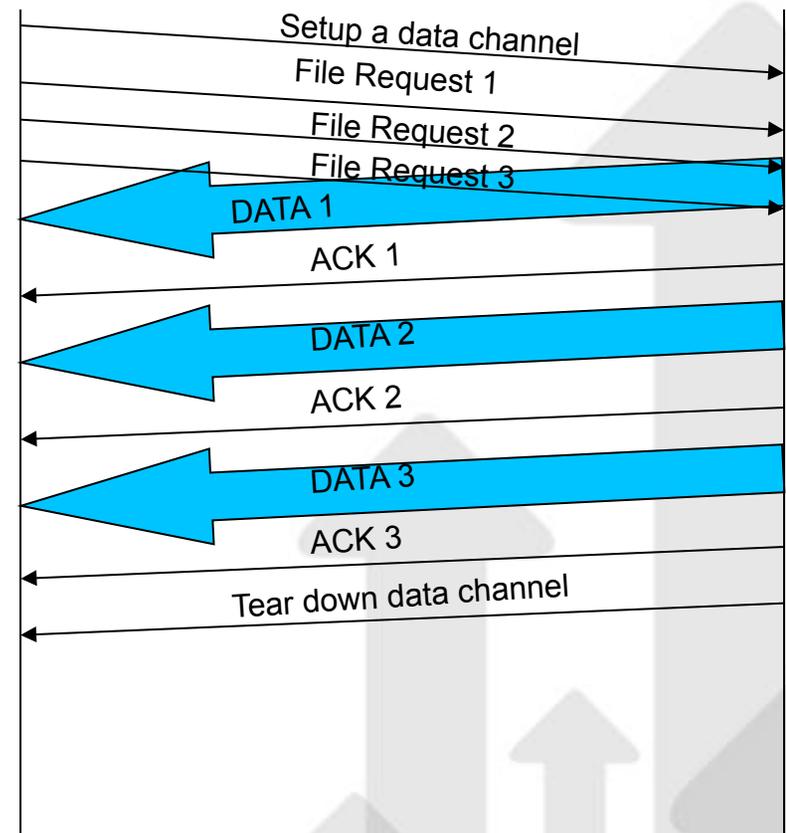


Pipelining

GridFTP

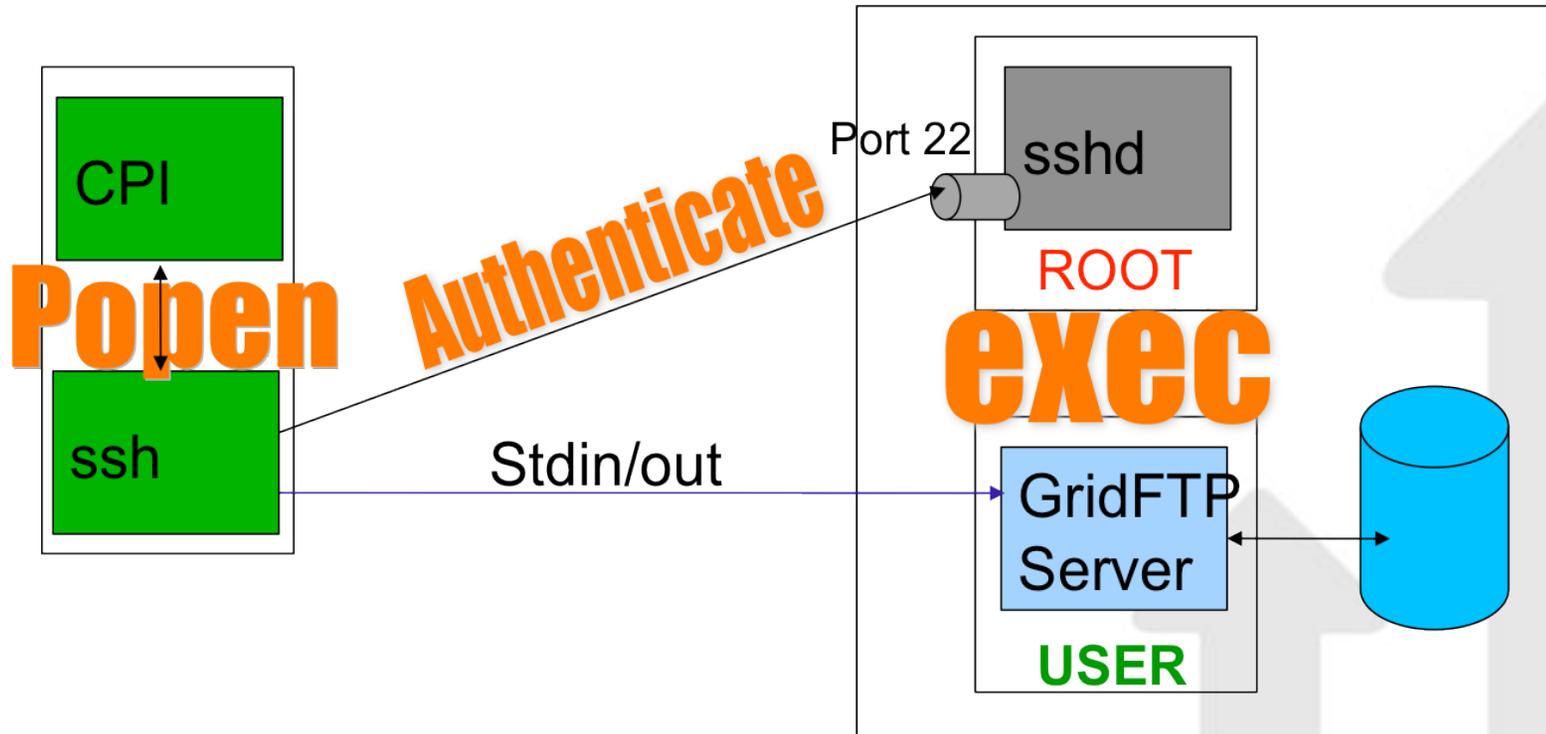


GridFTP with Pipelining





SSH-based GridFTP

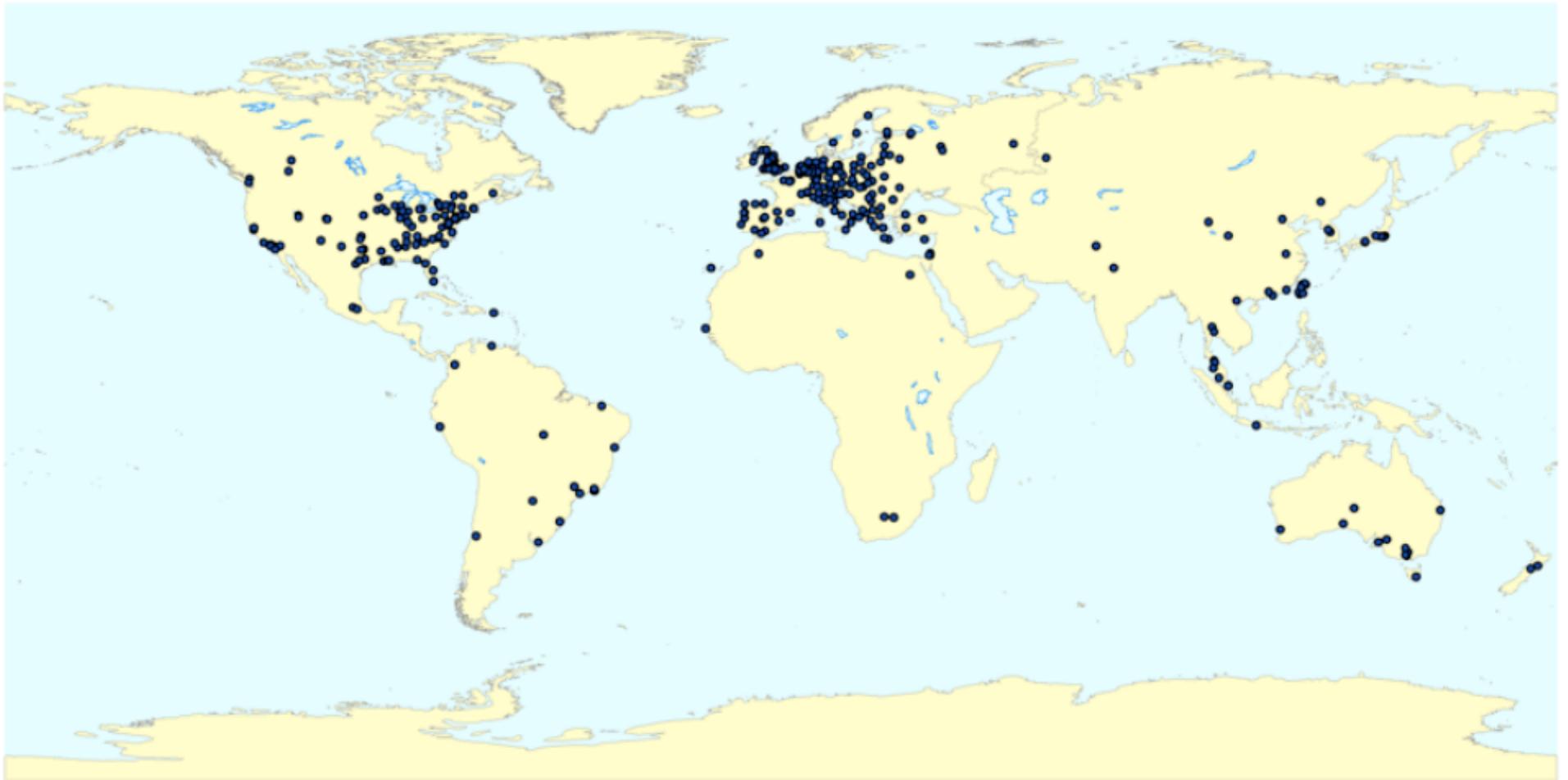




- **Synchronize datasets**
- **Only transfer files where the destination does not exist or differs (size, timestamp, checksum) from the source**
- **‘-sync-level’**
 - 0 - transfer if the destination does not exist
 - 1 - transfer if the size does not match
 - 2 - transfer if timestamp of destination is older
 - 3 - transfer if the checksums do not match.
 - The default sync level is 2.



GridFTP Servers Around the World

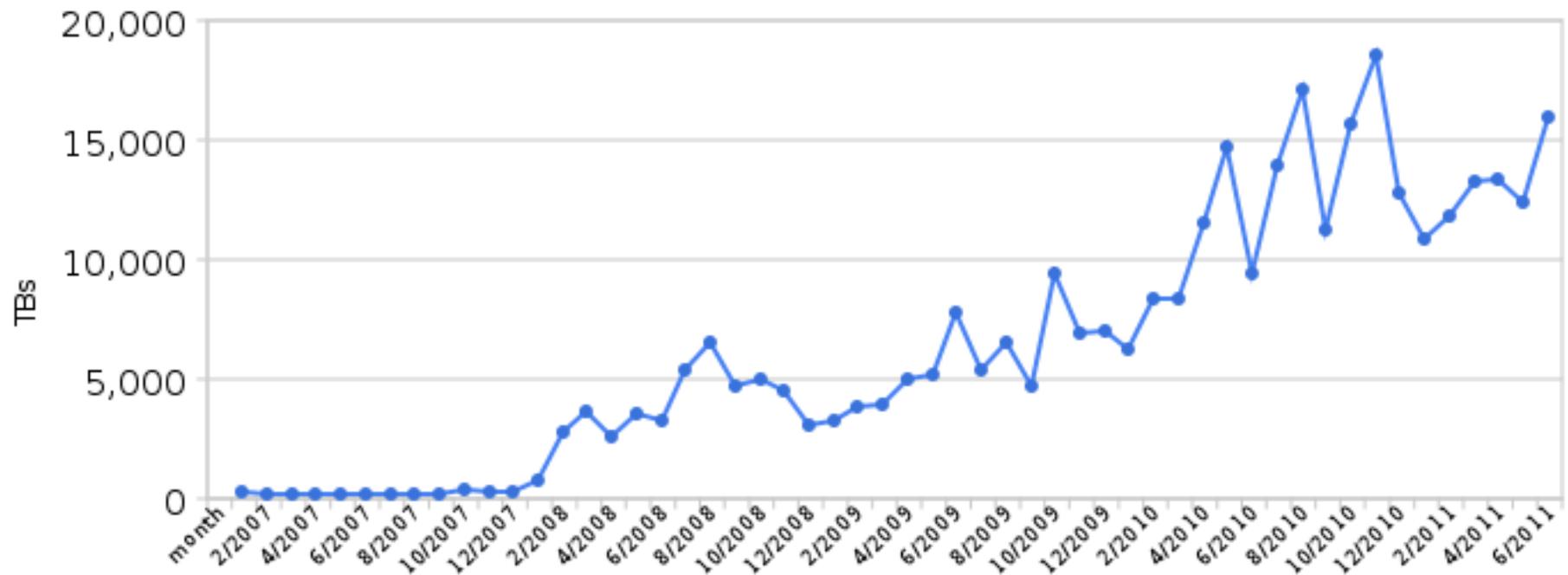


Created by Tim Pinkawa (Northern Illinois University) using MaxMind's GeoIP technology (<http://www.maxmind.com/app/ip-locate>).



GridFTP Transfers

Monthly Totals* of TBs Transferred Via GridFTP



*for those "reporting"



globus online

Reliable File Transfer. No IT Required.

Reliable, high-performance, secure file transfer

Move files fast. No IT required.

+ HOW IT WORKS

Learn more about the service

> GET STARTED

Setup profile in 3 easy steps

Globus Online makes robust file transfer capabilities, traditionally available only on expensive, special-purpose software systems, accessible to everyone.

[Learn more](#)



Why Use Globus Online?

See how easy file transfer can



For HPC Resource Owners

Enable Globus Online



For Developers

Integrate with Globus Online



Globus Connect

Globus Connect Installation

Globus Connect allows you to use Globus Online to transfer files to and from your computer. [Need Help? Click Here](#)

Step One: Choose Your Download

Globus Connect For Mac OS X Globus Connect For Linux Globus Connect For Windows

Step Two: Get Your Globus Connect Setup Key

Endpoint Name:

Description:

[Generate Setup Key](#)

Step Three: Finish Globus Connect Setup

Copy the setup key displayed above. Run Globus Connect and paste the key into the Initial Setup window when prompted. This setup key can only be used once.

[Close](#)

Globus Connect

1 of 2 selected, 90.4 MB available

globus online

Reliable File Transfer. No IT Required.

[Globus Connect](#) Applications

Setup

Initial Setup

Please type or paste your Globus Connect setup key into the field below and click 'OK' when finished.

Setup Key:

▶ Advanced

[Ok](#)

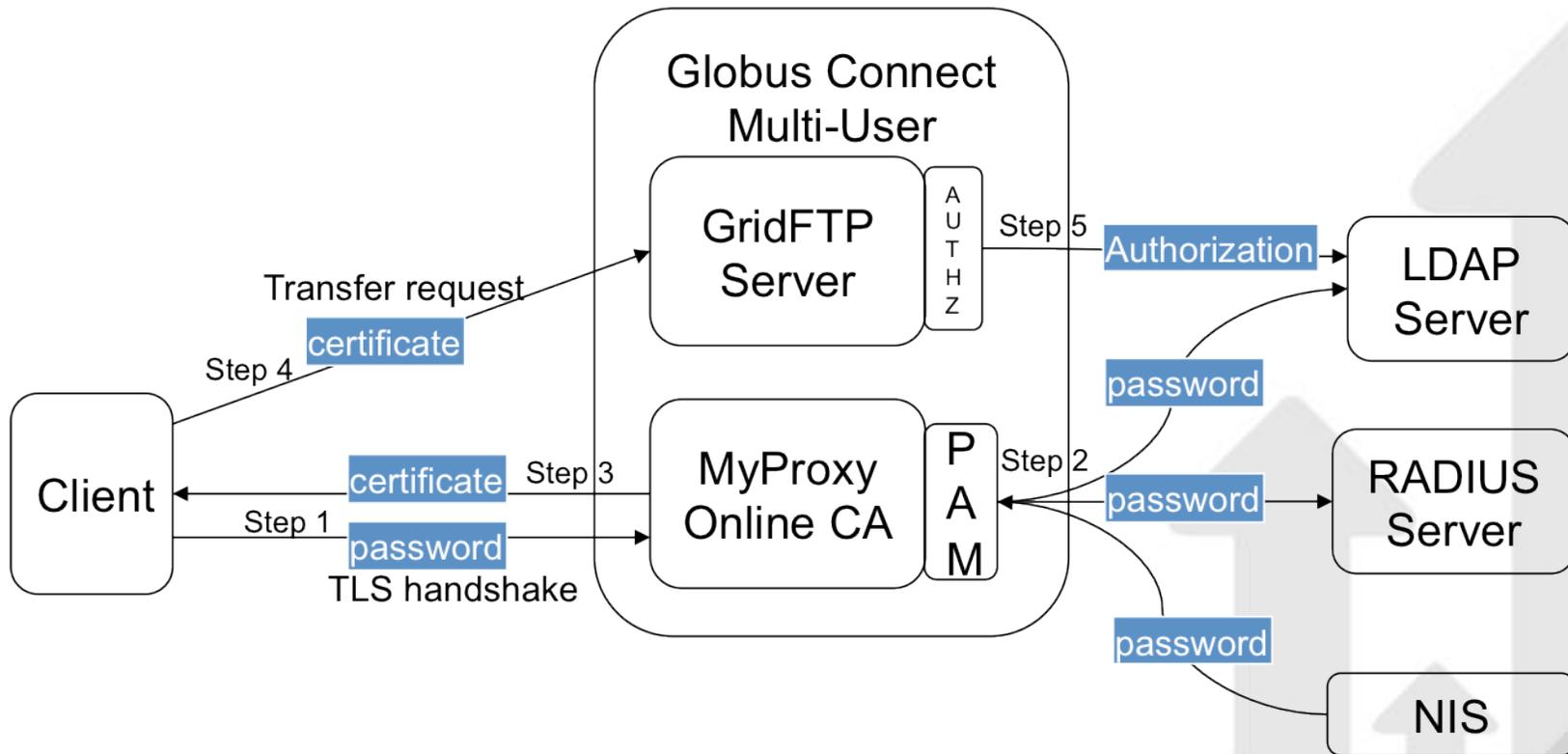


Native Packaging

- GT5.2 to have native packages for RedHat, RHEL, Fedora, CentOS, Scientific Linux, Debian, Ubuntu
- Released two alphas
- Release testing third alpha (GT5.1.1)
 - GridFTP, MyProxy, GSI-OpenSSH
 - CentOS 5, Fedora 13 and 14, RedHat 5, Scientific Linux 5.5 and Debian
- GT5.2 – end of summer

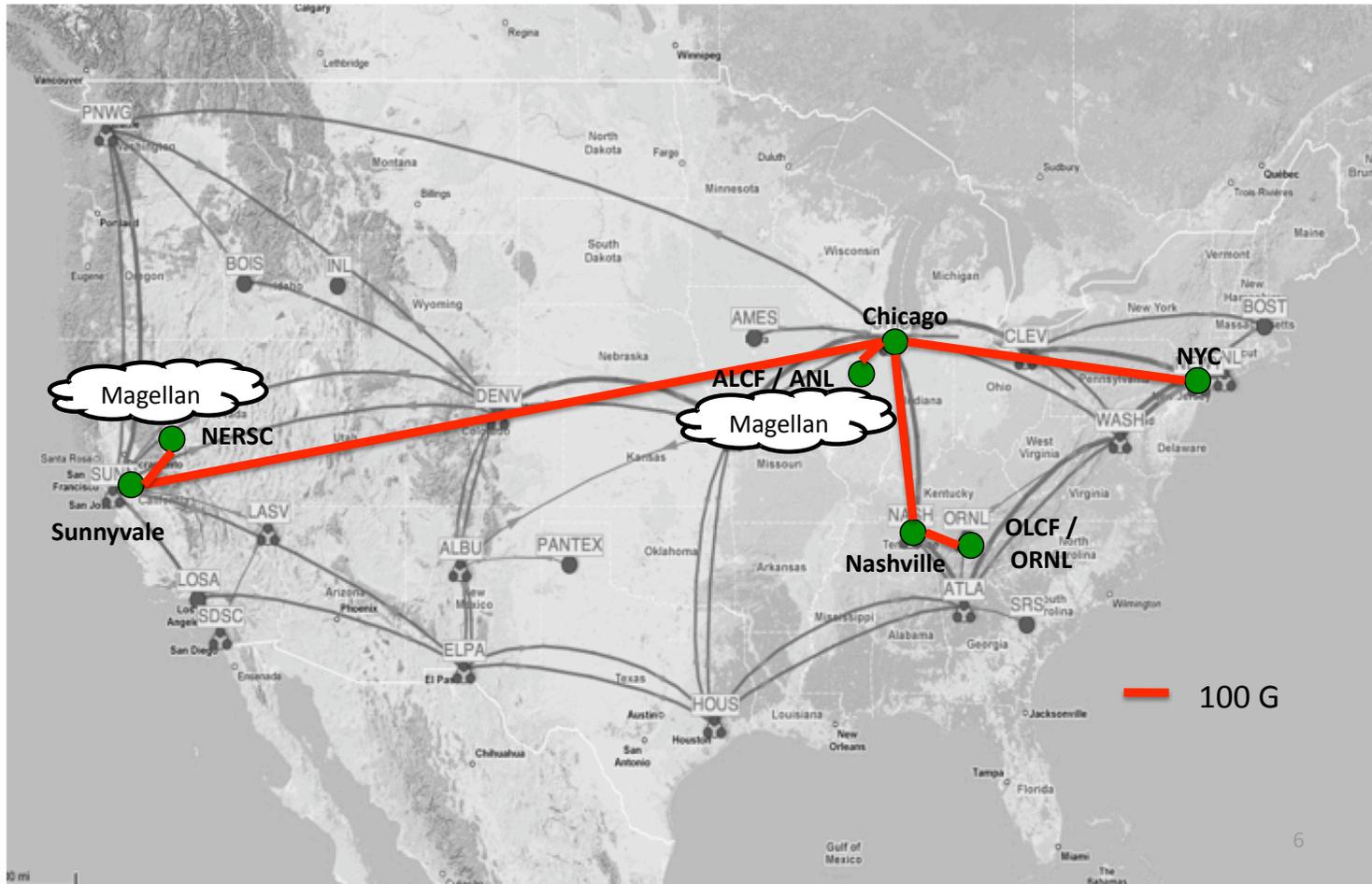


Globus Connect Multi-User





100G Network



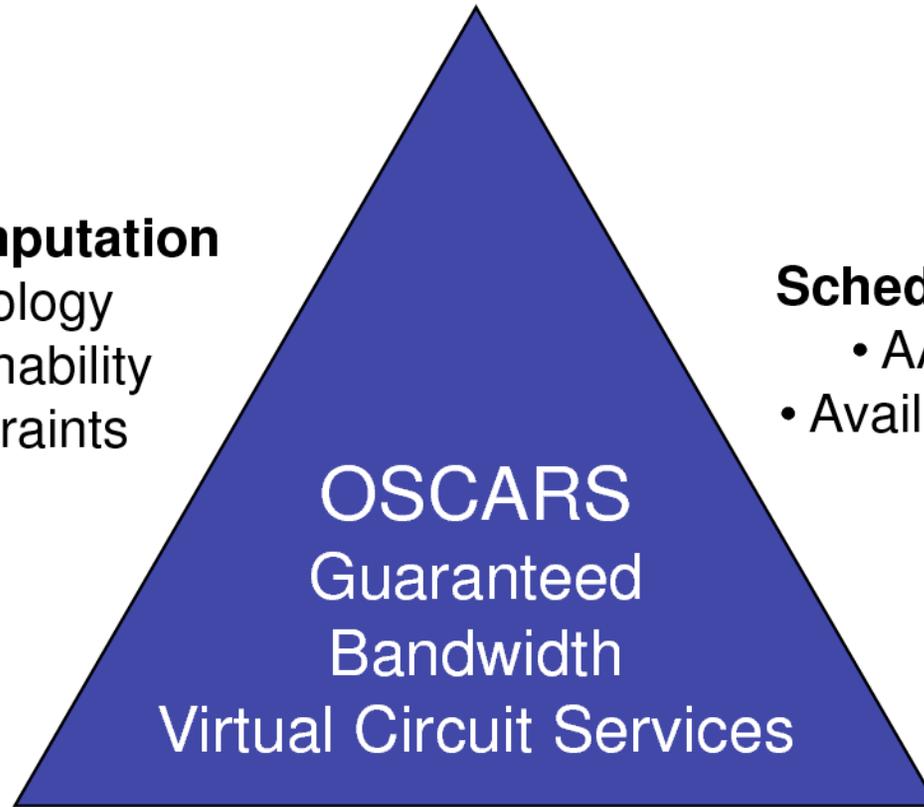
OSCARS Overview

Path Computation

- Topology
- Reachability
- Constraints

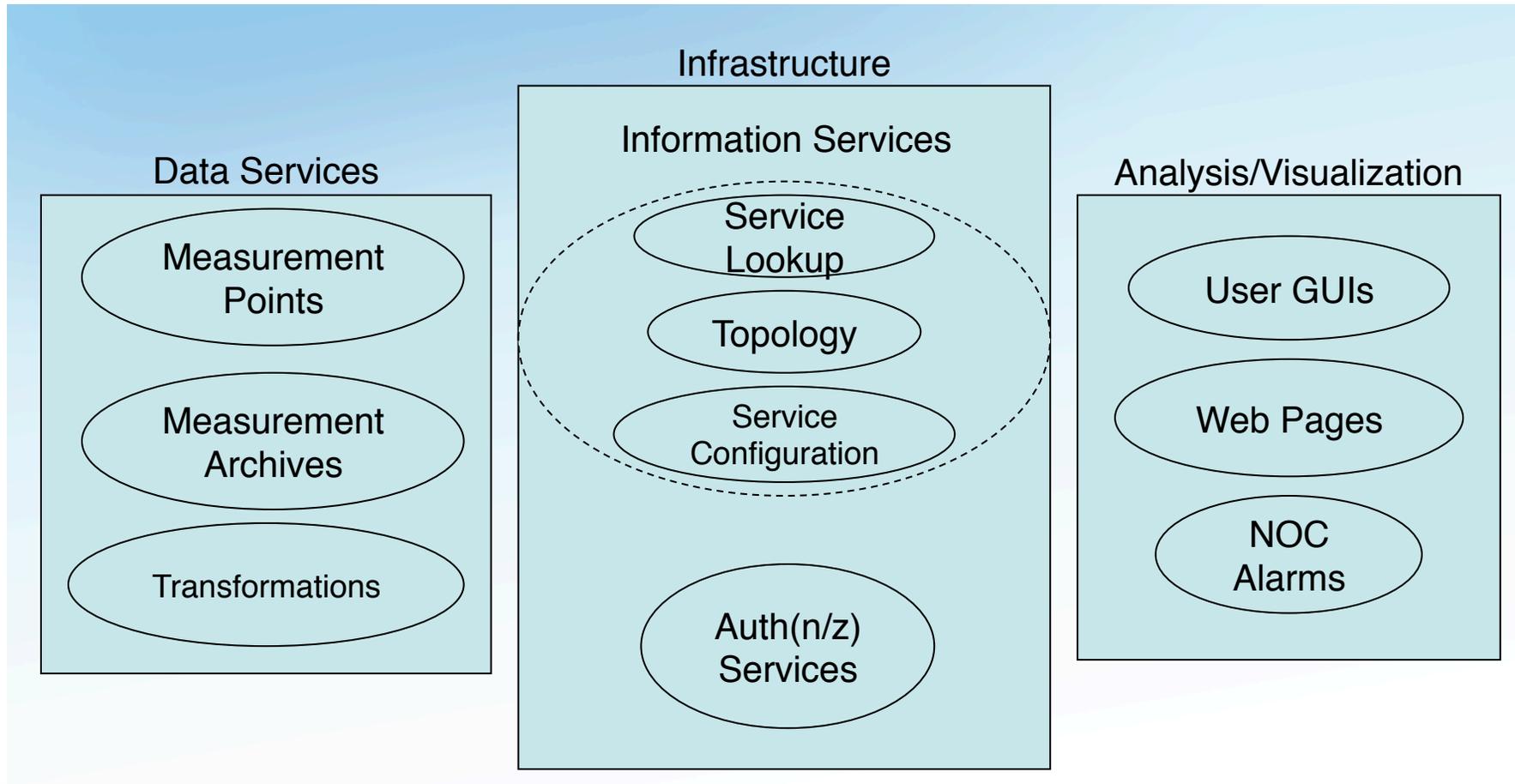
Scheduling

- AAA
- Availability



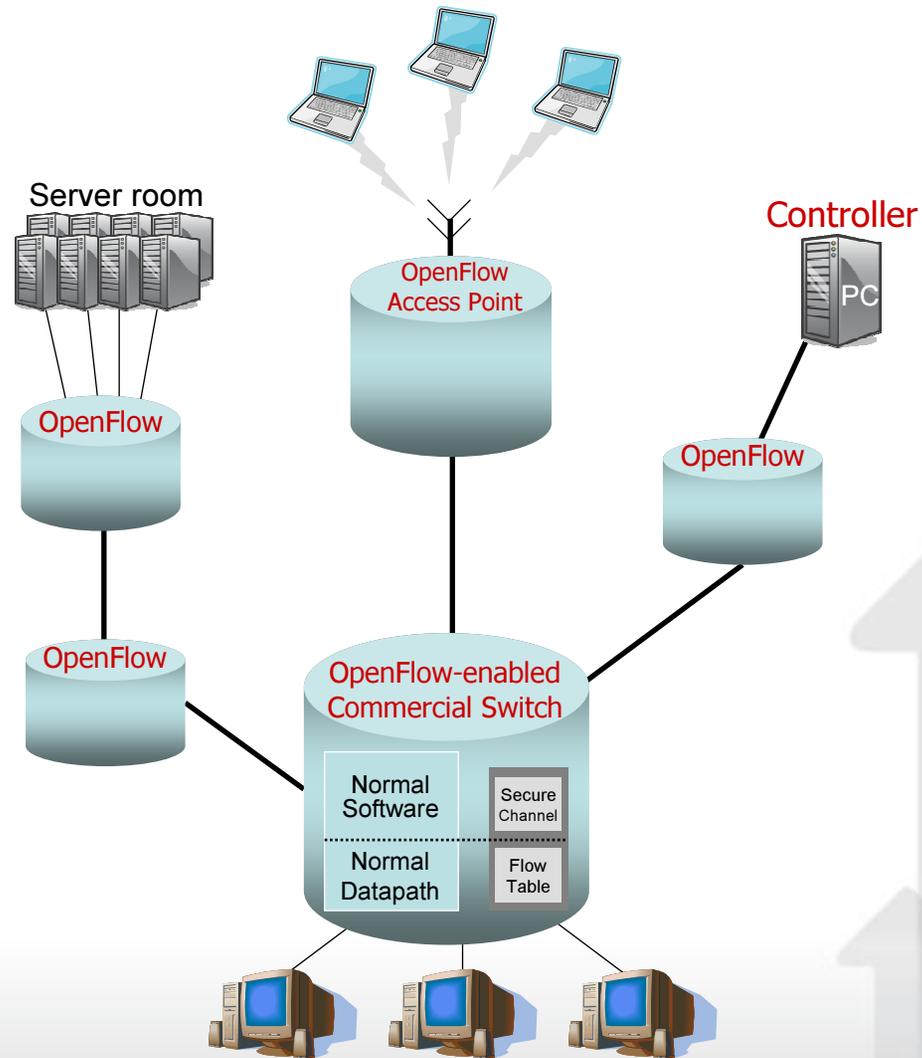
Provisioning

- Signalling
- Security
- Resiliency/Redundancy



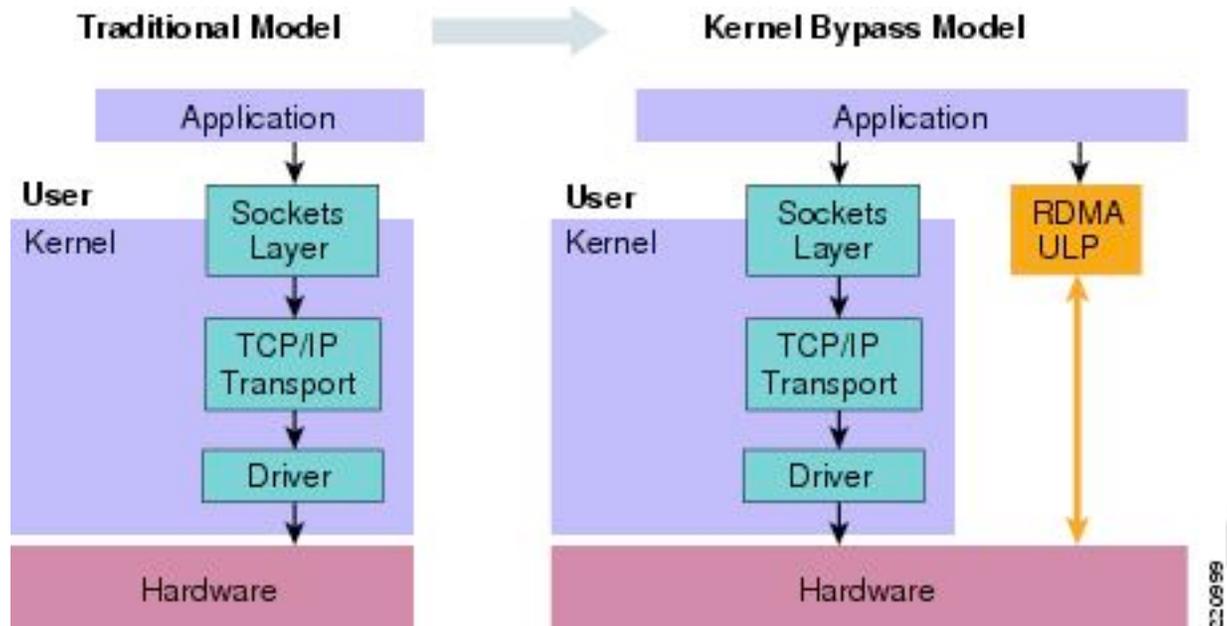


OpenFlow





RDMA over WAN



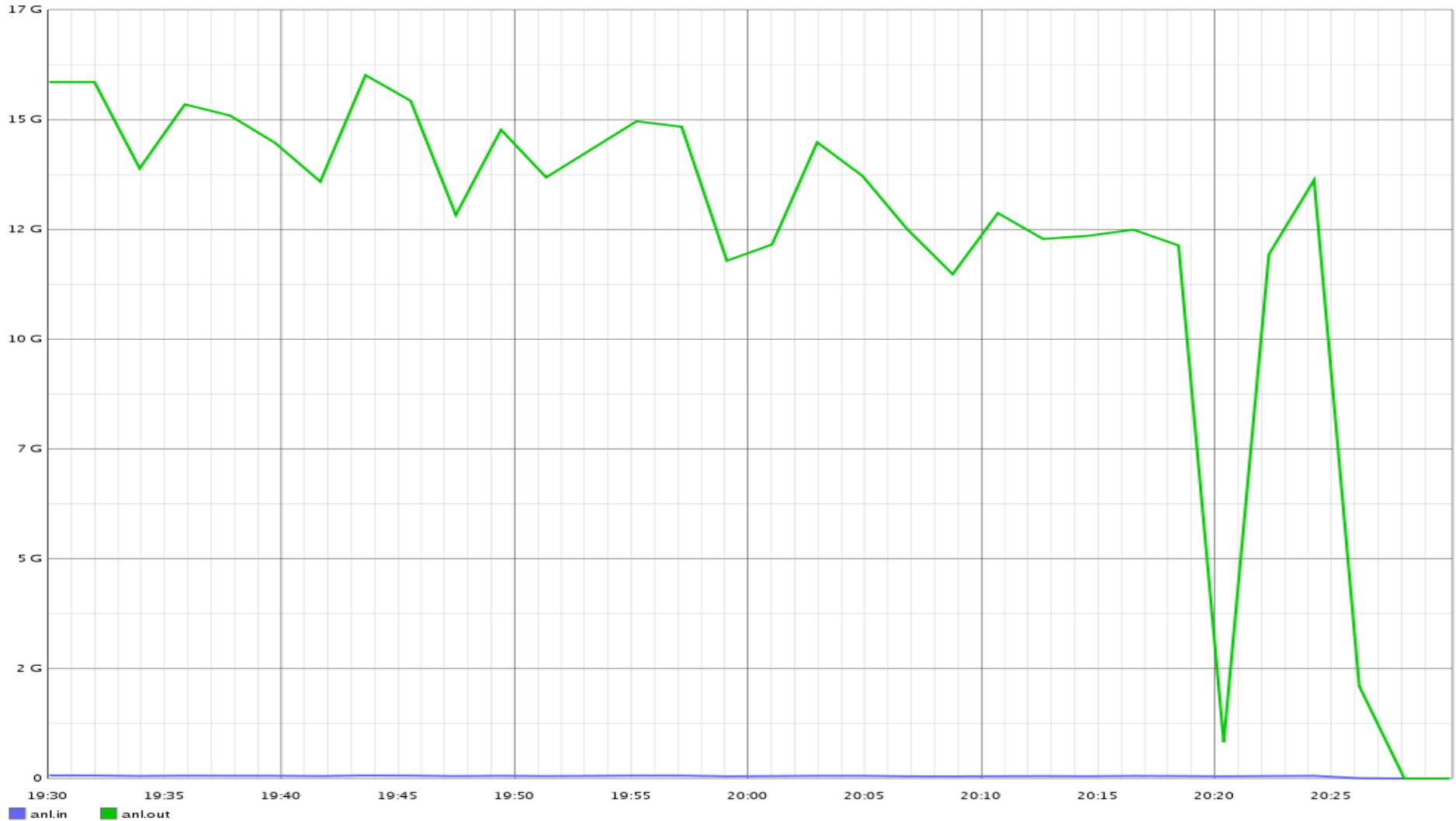


Challenges

- Growing disparity between network speeds and file system characteristics
- Parallel file systems typically optimized for massively parallel local access
- End systems with increasing core counts
- Effectively utilize programmability in the network
- Highly parallel, coordinated data movement

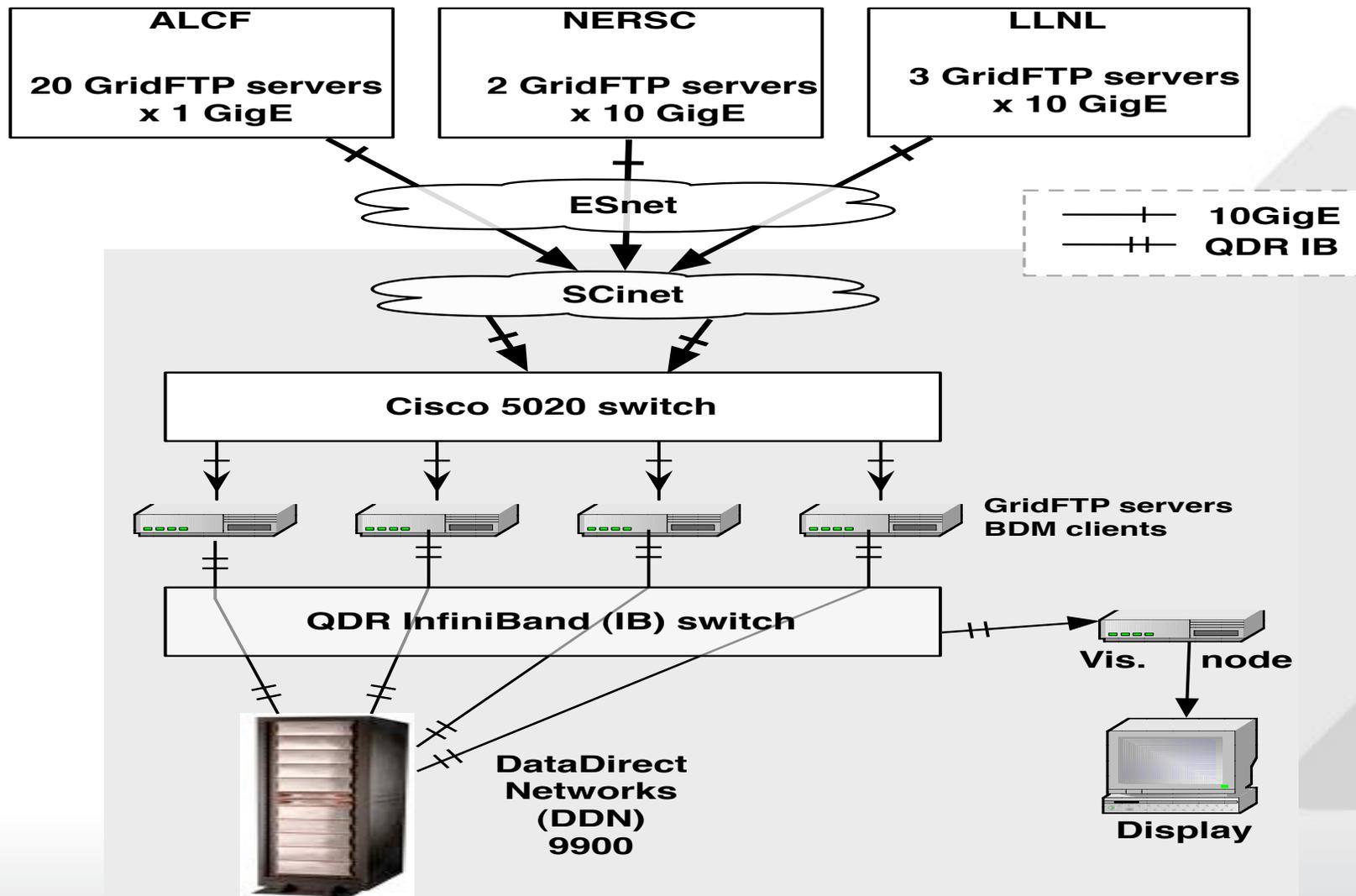


Data Movement in 2009





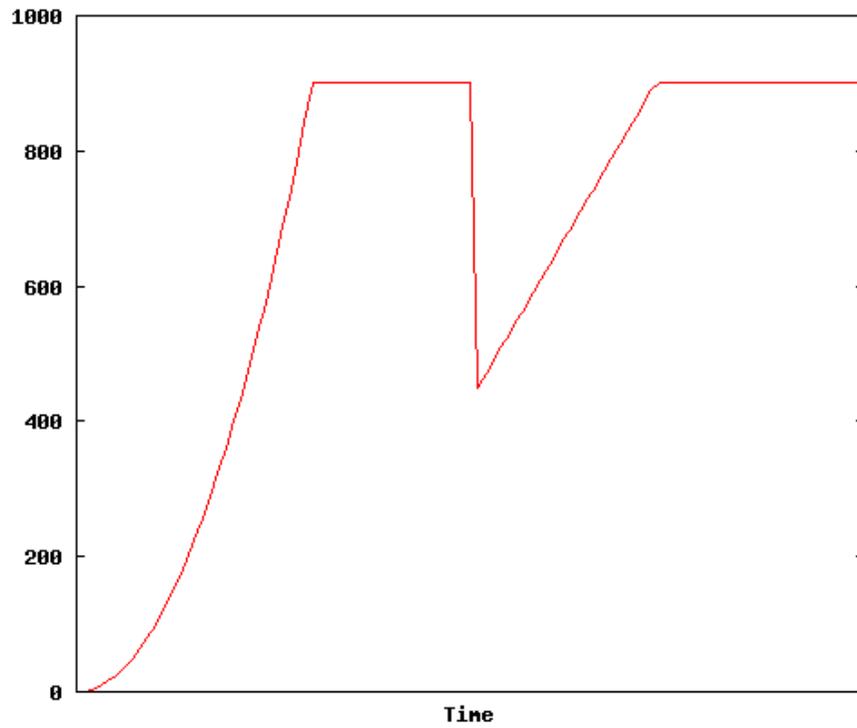
End System Disparity



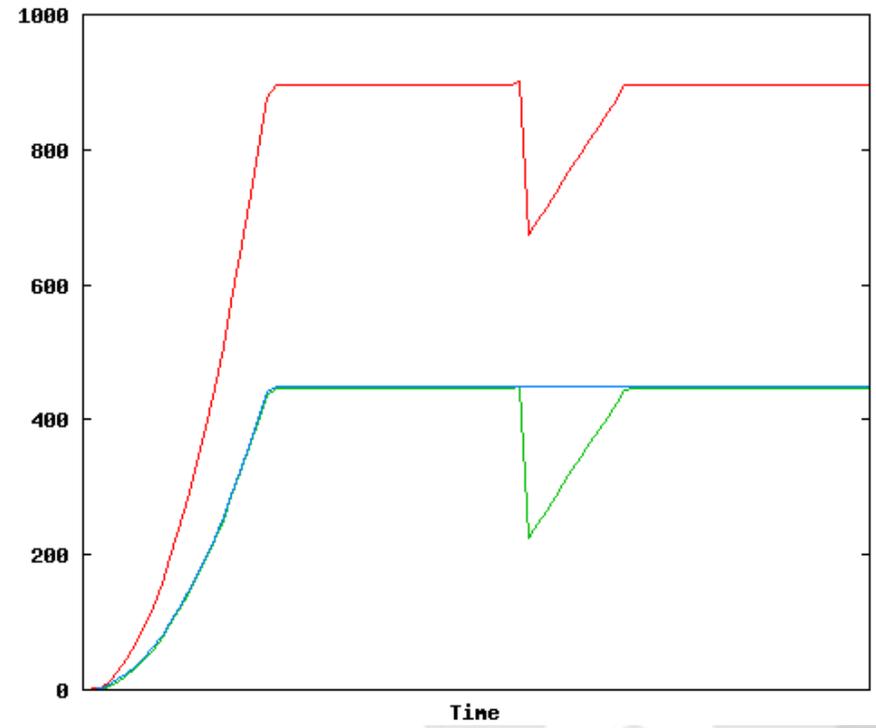


Parallel Streams

One Stream

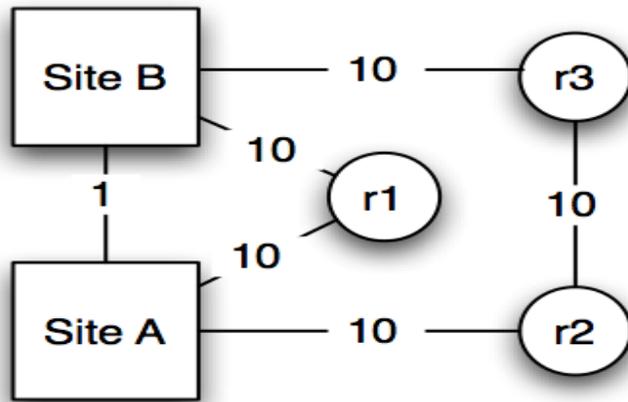


Two Streams



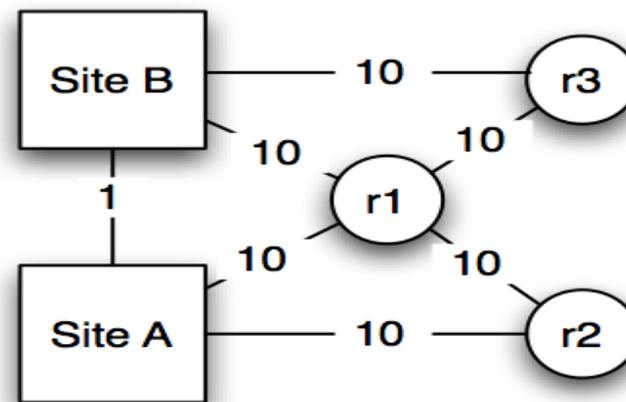


Multi-pathing



1 - 1 Gbps link
10 - 10 Gbps link

(a) Network Example 1

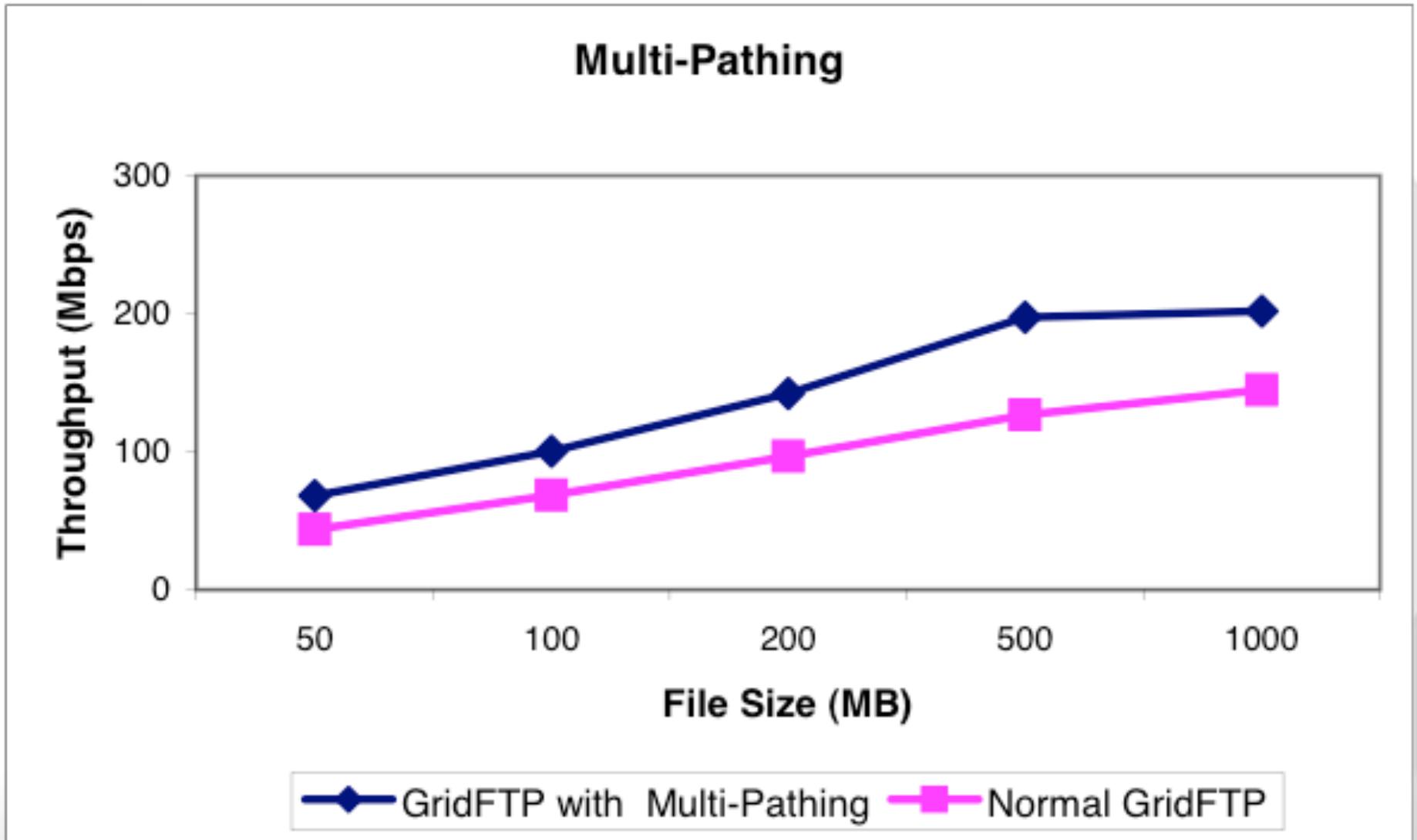


1 - 1 Gbps link
10 - 10 Gbps link

(b) Network Example 2



Multi-Pathing





Protocol Selection

- Protocols are different for a reason
 - Some designed for dedicated networks and are aimed at the greedy acquisition of bandwidth;
 - Others are designed to coexist with the traffic of multiple users in a shared network
- Identify the best protocol for a given transfer
 - Network type (shared or dedicated, IP or non-IP)
 - Network activity (under-utilized or over-utilized)
 - Capabilities of endpoints (network interface etc)
 - Application behavior



Auto-tuning

- Network condition change rapidly
- Optimization parameters have to be changed accordingly
 - Number of parallel streams
 - Number of data movers
 - Buffer sizes
- Even change the transport protocol itself
- Sophisticated auto-tuning/decision making algorithms



End Systems

- **Interrupt affinity**
 - Interrupt processing done by processor to which the interrupt is physically bound
- **Thread affinity**
 - Application thread bound to processor where Interrupt processing of network traffic occurs.
- **Memory affinity**
 - Memory used by an application thread is allocated on the memory bank with the lowest access latency



Small files

- Even though application data sets continue to increase, most file sizes are small
 - The average file size on the OLCF parallel file systems is only 14.8 megabytes
 - Dark Energy Survey expects to have a median file size of approximately 150 KB
 - Climate datasets are characterized by tens of thousands of small files
- As network speed increases, these files are going to look even smaller
 - Sophisticated techniques to get transfer rates close to network speeds for small files.



Questions?

