Edinburgh e-Science TOE

Training Outreach and Education

http://www.nesc.ac.uk/training

NGS National Grid Service

http://www.ngs.ac.uk

GridFTP

Guy Warner,

NeSC Training Team

PPARC

http://www.pparc.ac.uk/

egee

http://www.eu-egee.org/

Acknowledgement

- These slides are slides given by
  Bill Allcock of Argonne National Laboratory at the GridFTP Course at NeSC in January 2005

  With some minor presentational changes

What is GridFTP?

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

Basic Definitions

- Network Endpoint
  - Something that is addressable over the network (i.e. IP:Port). Generally a NIC
  - multi-homed hosts
  - multiple stripes on a single host
- Parallelism
  - multiple TCP Streams between two network endpoints
- Striping
  - Multiple pairs of network endpoints participating in a single logical transfer (i.e. only one control channel connection)

NGS
National Grid Service

Striped Server

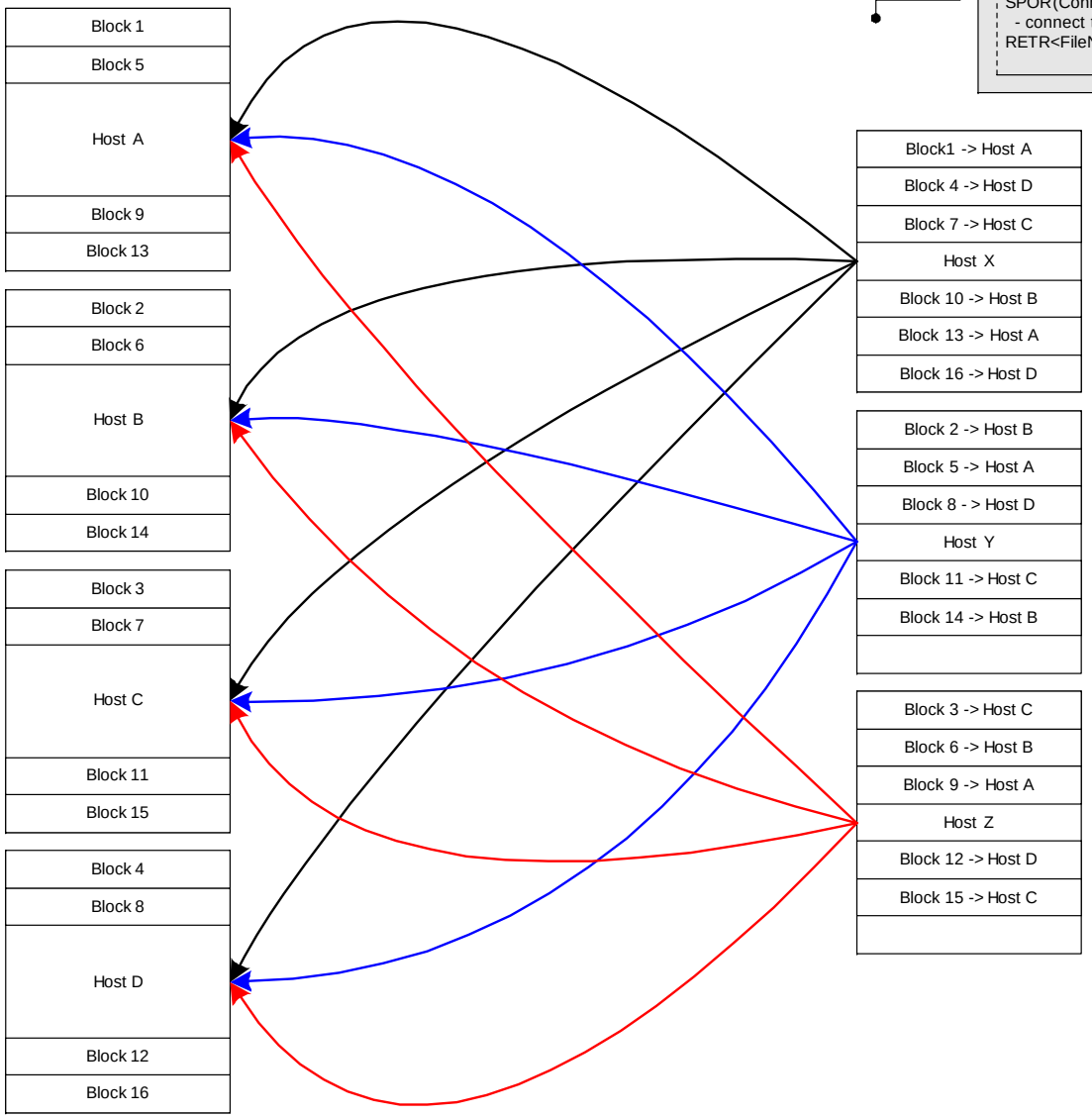
- Multiple nodes work together and act as a single GridFTP server
- An underlying parallel file system allows all nodes to see the same file system and must deliver good performance (usually the limiting factor in transfer speed)
  - I.e., NFS does not cut it
- Each node then moves (reads or writes) only the pieces of the file that it is responsible for.
- This allows multiple levels of parallelism, CPU, bus, NIC, disk, etc.
  - Critical if you want to achieve better than 1 Gbs without breaking the bank

18-Nov-03

GridFTP Striped Transfer

MODE E
SPAS (Listen)
- returns list of host:port pairs
STOR<FileName>

MODE E
SPOR (Connect)
- connect to the host-port pairs
RETR<FileName>



NGS National Grid Service

globus-url-copy: 1

- Command line scriptable client
- Globus does not provide an interactive client
- Most commonly used for GridFTP, however, it supports many protocols
  - gsiftp:// (GridFTP, historical reasons)
  - ftp://
  - http://
  - https://
  - file://

globus-url-copy: 2

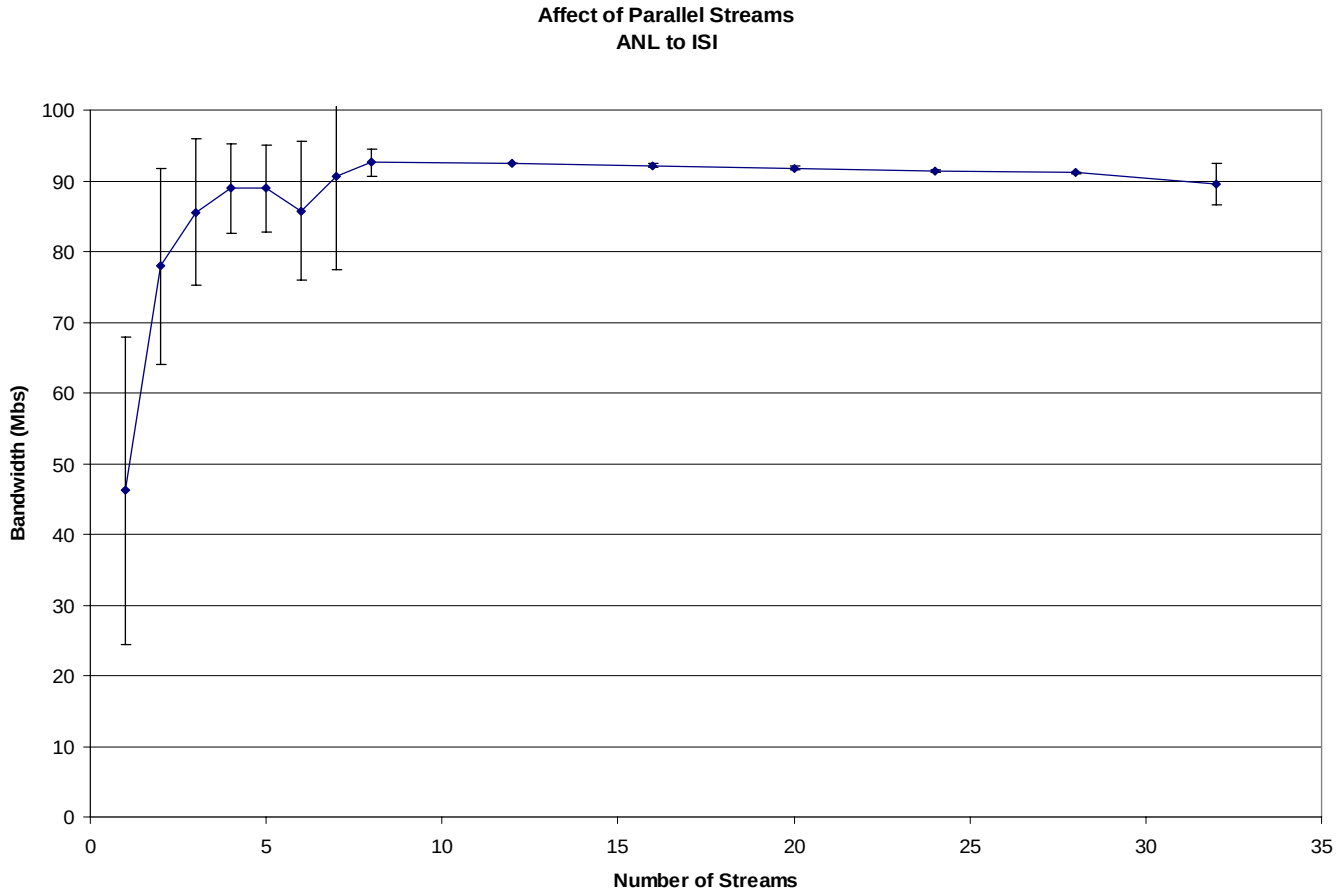
- globus-url-copy [options] srcURL dstURL

Important Options

- -p (parallelism or number of streams)
  - rule of thumb: 4-8, start with 4
- -tcp-bs (TCP buffer size)
  - use either ping or traceroute to determine the Round Trip Time (RTT) between hosts
  - buffer size = BandWidth (Mbs) * RTT (ms) *(1000/8) / P
  - P = the value you used for –p
- -vb if you want performance feedback
- -dbg if you have trouble

NGS
National Grid Service

Parallel Streams



NGS
National Grid Service

BWDP

- TCP is reliable, so it has to hold a copy of what it sends until it is acknowledged.
- Use a pipe as an analogy
- I can keep putting water in until it is full.
- Then, I can only put in one gallon for each gallon removed.
- You can calculate the volume of the tank by taking the cross sectional area times the height
- Think of the BW as the cross-sectional area and the RTT as the length of the network pipe.

NGS
National Grid Service

Other Clients

- Globus also provides a Reliable File Transfer (RFT) service
- Think of it as a job scheduler for data movement jobs.
- The client is very simple. You create a file with source-destination URL pairs and options you want, and pass it in with the –f option.
- You can “fire and forget” or monitor its progress.

TeraGrid Striping results

- Ran varying number of stripes
- Ran both memory to memory and disk to disk.
- Memory to Memory gave extremely high linear scalability (slope near 1).
- Achieved 27 Gbs on a 30 Gbs link (90% utilization) with 32 nodes.
- Disk to disk - limited by the storage system, but still achieved 17.5 Gbs