the globus alliance
www.globus.org

Managed Object Placement Service

John Bresnahan, Mike Link and Raj Kettimuthu (Presenting)

Argonne National Lab

Topics for discussion

- MOPS overview
- MOPS 0.1
    - Pipelining
    - Gfork
    - GridFTP plugin for Gfork
    - Lotman
- Other GridFTP enhancements
    - Performance
    - Ease of use
- Future directions

Managed Object Placement Service (MOPS)

- Resource management in GridFTP
  - Memory usage limitation
  - Enforce appropriate storage usage
  - Enforce appropriate bandwidth usage
  - Eliminates the potential to *overheat* a system
- Bandwidth and storage reservation
- Transfer scheduling
- MOPS 0.1 is available at http://www.cedps.net/wiki/index.php/Software

MOPS

- MOPS 0.1 includes
  - Optimization for lots of small files transfer
  - Globus fork (Gfork) - inetd like service that allows state to be maintained across connections
  - GridFTP plugin for Gfork - allows for dynamic addition/removal of data movers, limit memory usage
  - Lotman - manage storage
  - Plugin for GridFTP to enforce storage usage policies using lotman

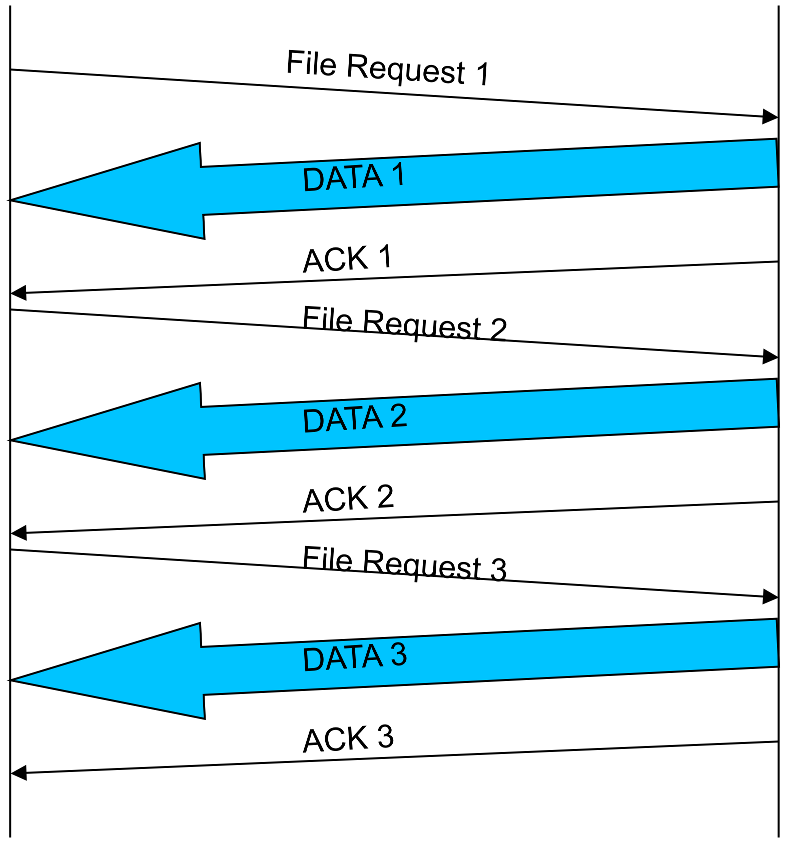
the globus alliance
www.globus.org

Lots of Small Files (LOSF) Problem

- GridFTP and FTP - command response protocols
- A client must wait for a “Finished response” before sending the next command
- Overhead added on a per file basis
- Performance is best on large files
  - Overhead has less impact
- Performance suffers for a large data set partitioned into many small files

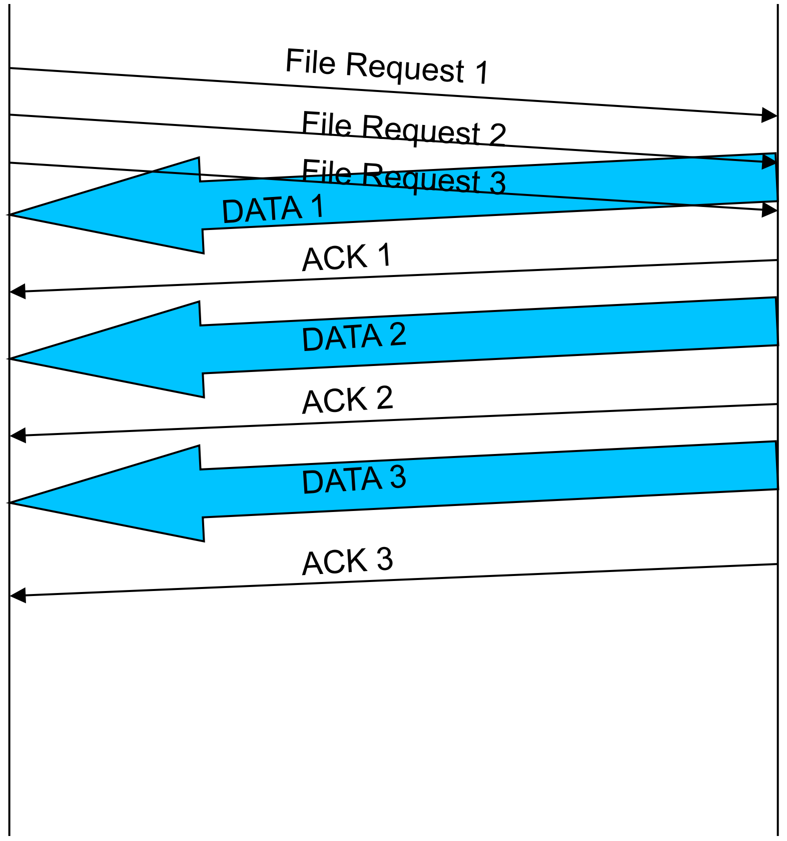
LOSF

- Traditional data transfer pattern



Pipelining

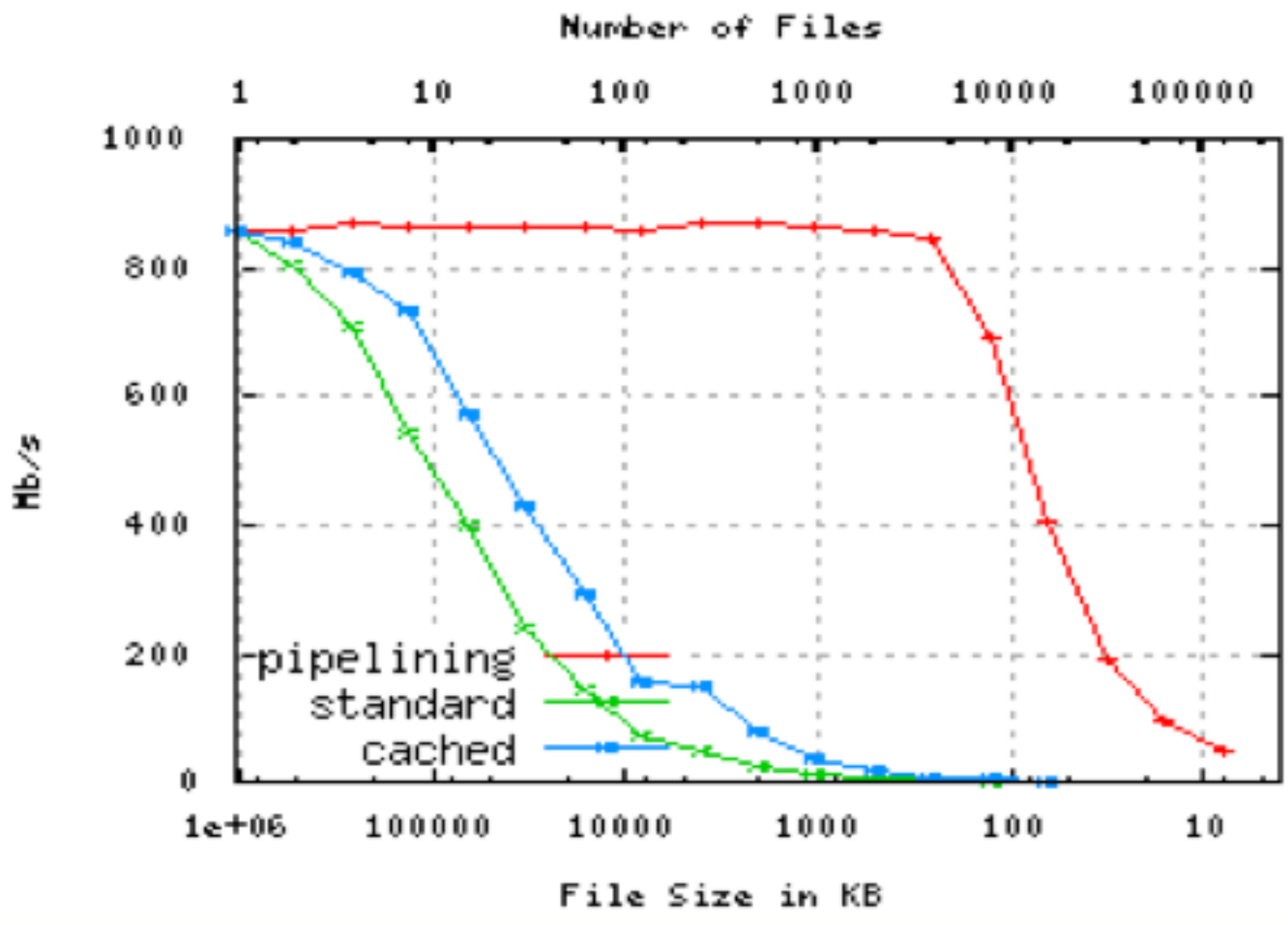
- Data transfer pattern with pipelining



the globus alliance
www.globus.org

Pipelining

WAN - SDSC and ANL with GSI security

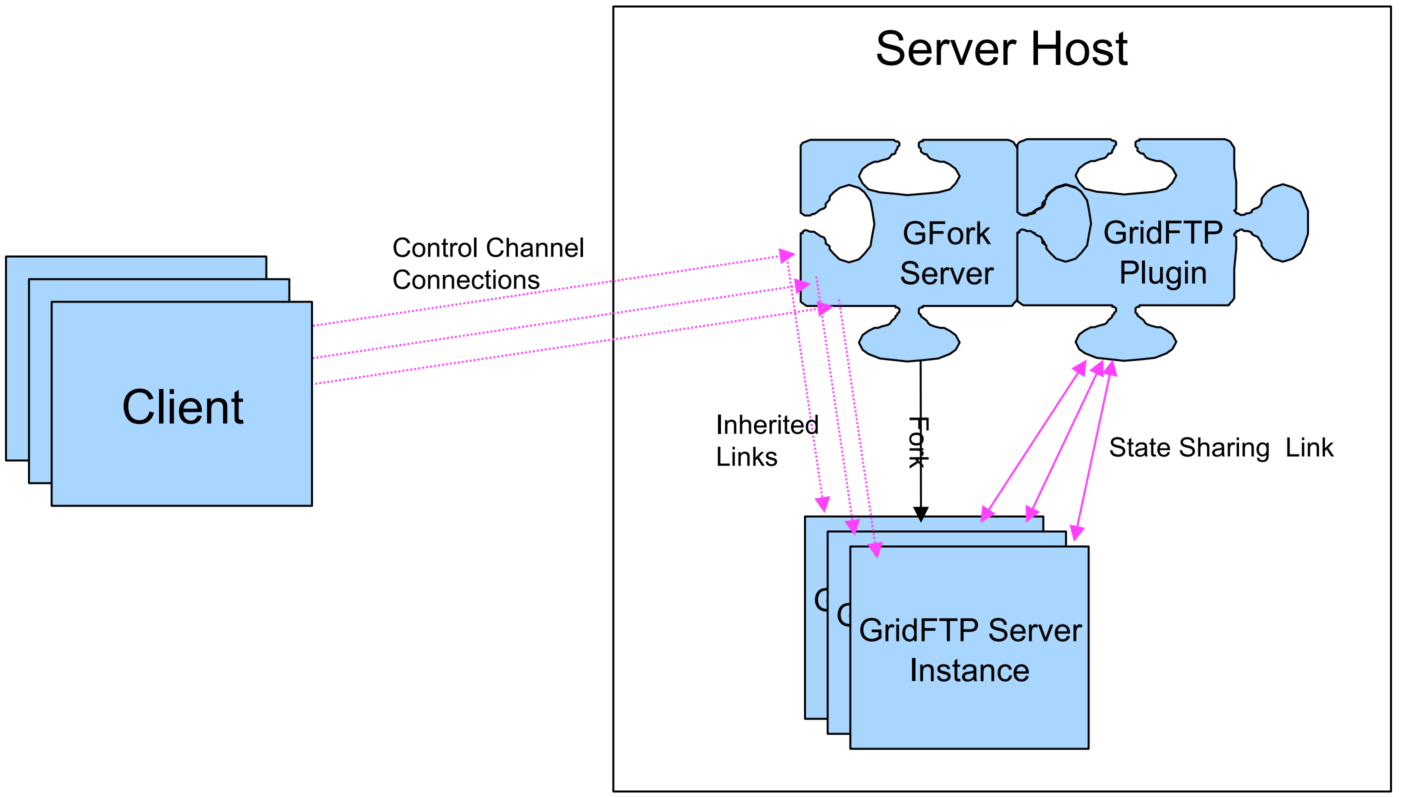


Gfork

- Under extreme loads it is possible that GridFTP servers require more memory than the system has and cause the system to fall over
- Developed a service called gfork to help avoid this situation
- Gfork - a service like inetd that listens on a TCP port and runs a configurable executable in a child process whenever a connection is made
- Allows server state to be maintained across connections

Gfork

- Associated with Gfork is a user defined master process
- Master process runs for the lifetime of the gfork daemon



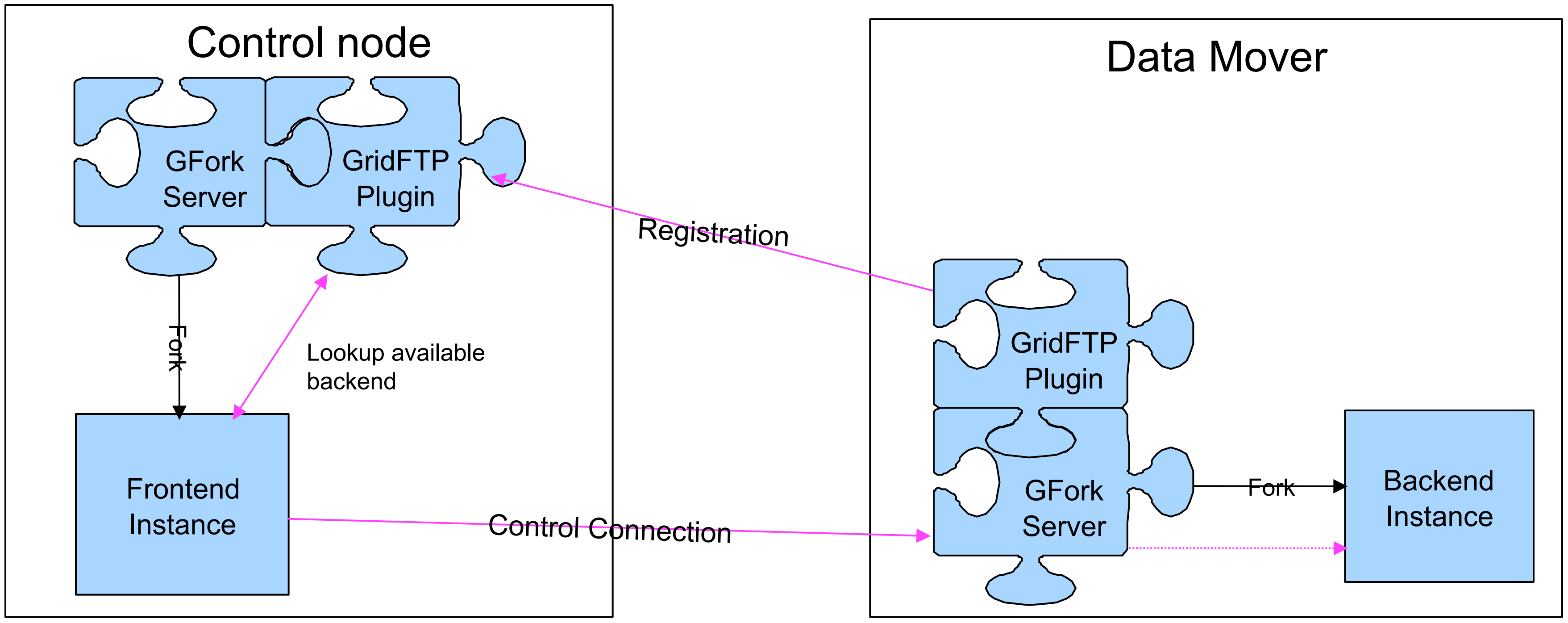
Dynamic data movers

- GridFTP can be run as a striped server - there is a control process and several data movers.
- The data movers run in tandem to transfer files faster by tying together many NICs.
- When the control process is run out of inetd the list of possible data movers had to be statically configured.
- But data movers tend to come and go.
  - Sometimes data movers fail
  - Sometimes data movers are added to a pool

Dynamic data movers

- GridFTP plugin for Gfork allows for dynamic addition/removal of data movers
- GridFTP plugin on the control node can be configured to listen for data mover registrations
- GridFTP plugin on the data movers can be configured to register with the plugin on the control node

Dynamic data movers



- Multiple data movers register with plugin
  - Plugin maintains the list of available DMs
- Control process instance selects N DMs for use
- If any one DM fails another can be used
- DM pool can grow and shrink

Memory Management

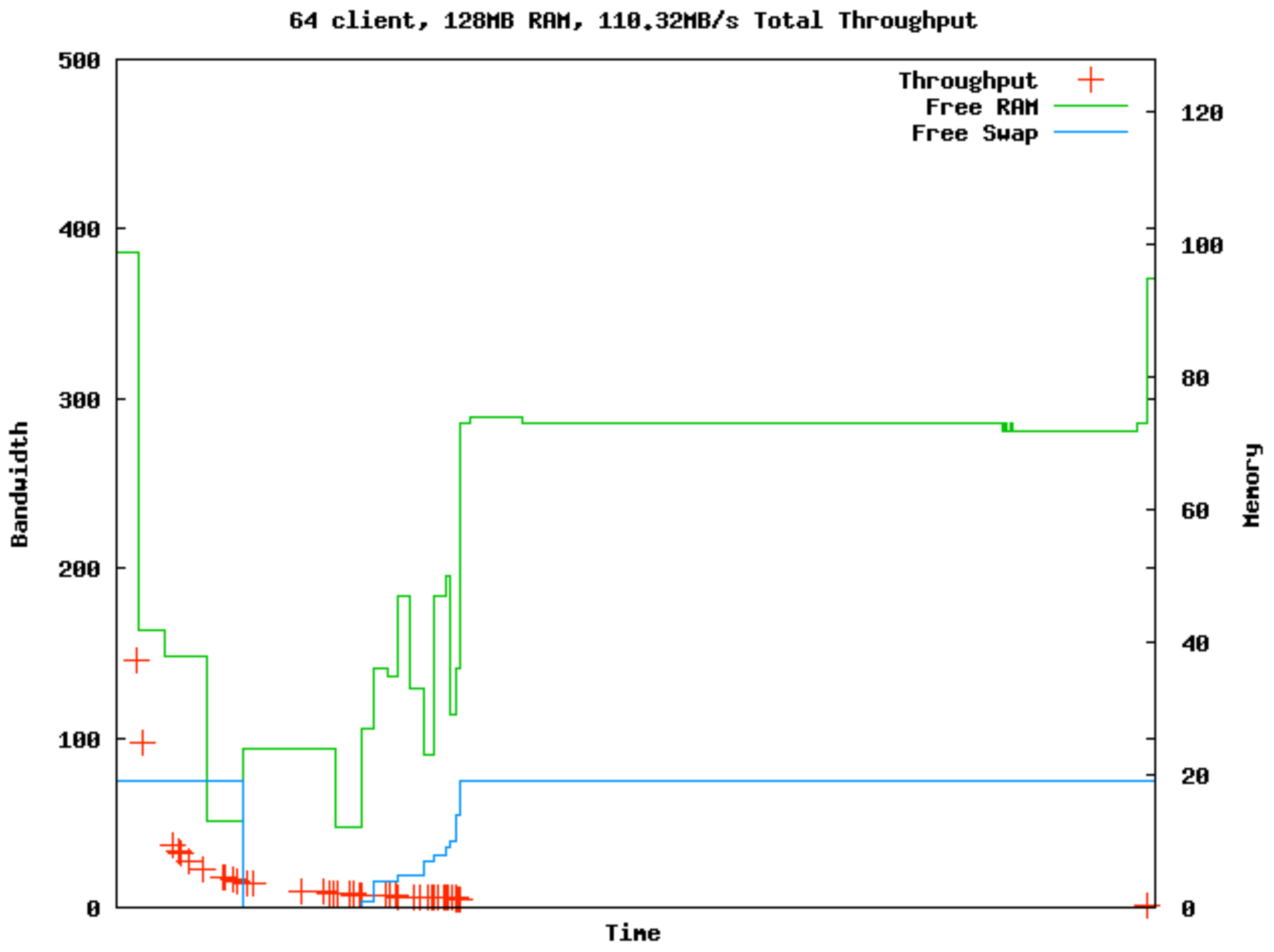
- GridFTP plugin for Gfork has a memory limiting option
- Limit memory usage to a given value or to the maximum amount of RAM in the system.
- Most of the memory is given to the first few connections
- When the plugin detects that it is overloaded, each session is limited to half the available memory.

Memory Management

- To measure the effectiveness of the memory limiting functionality, we performed experiments using xen VM.
- Clients are run on a 2GHz processor with 512MB of RAM and the server was run in a xen VM with either 64MB or 128MB of RAM with a 2GHz AMD 64bit processor.
- The machines were connected to each other via a 1Gb/s network with iperf measured max speeds of 600Mb/s.

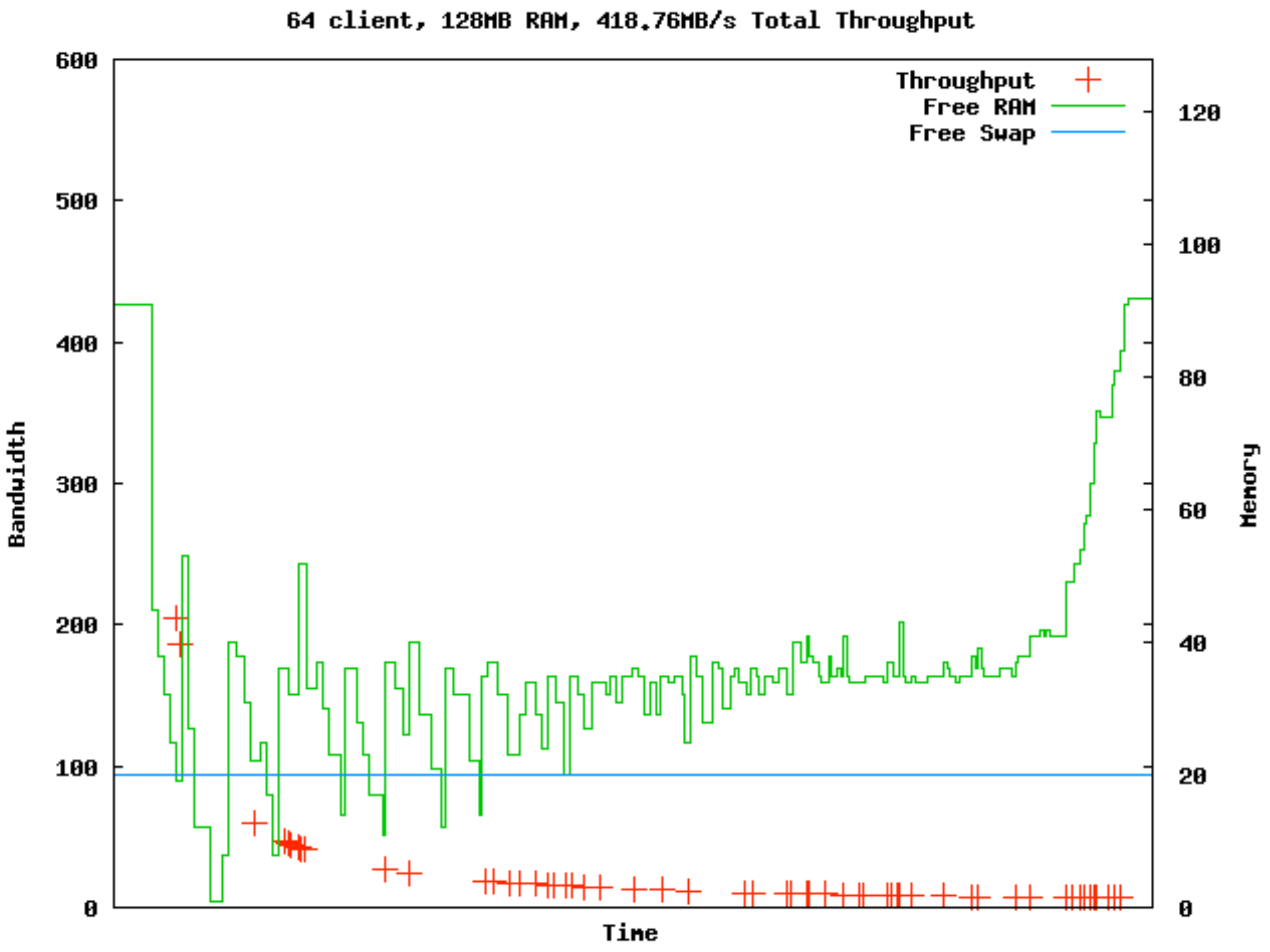
the globus alliance
www.globus.org

No memory limiting



the globus alliance
www.globus.org

With memory limiting

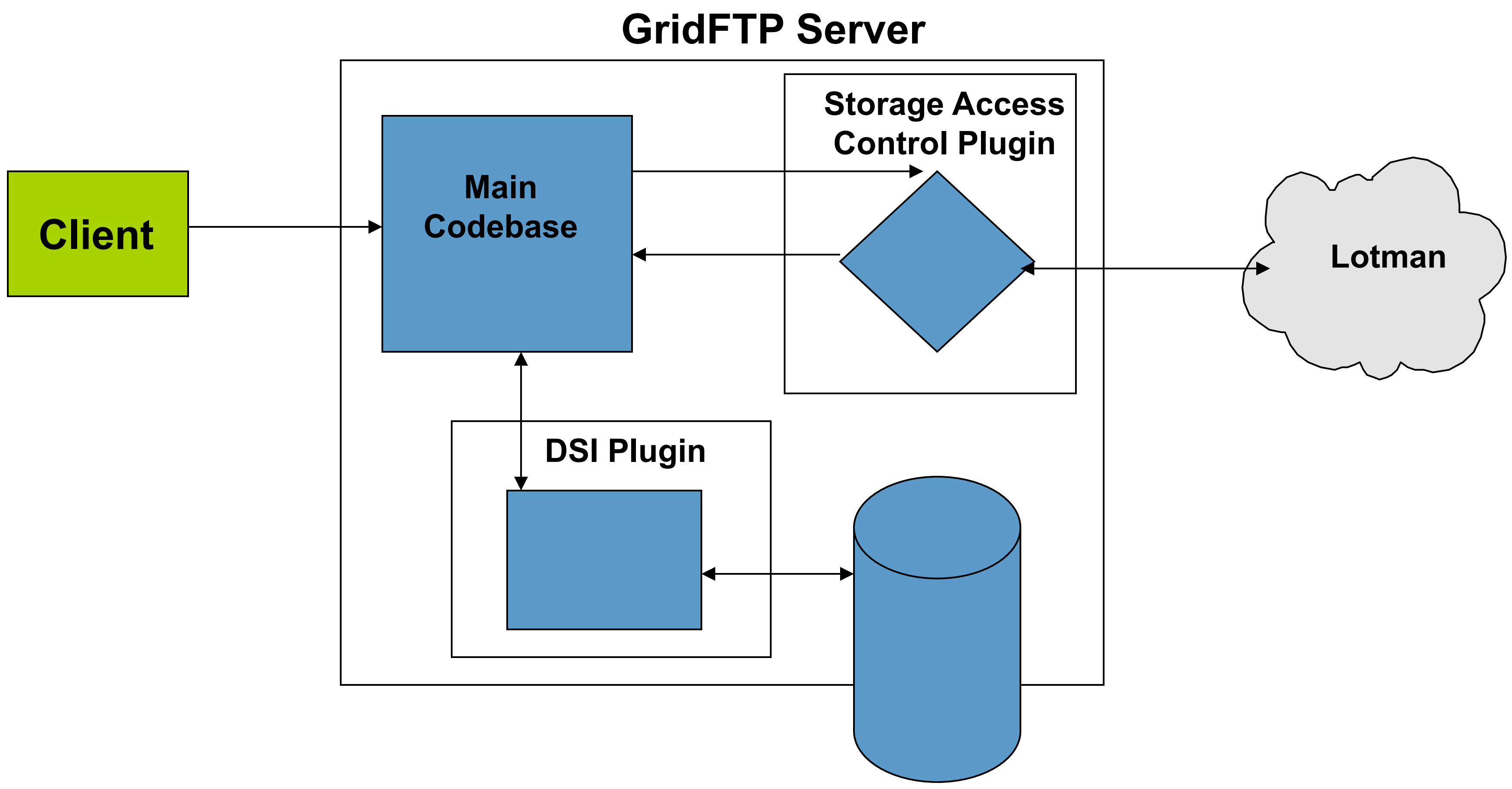


Lotman

- Manage disk space - interface to create disk space called lot
- Lot defined by four characteristics
  - Owner, capacity, duration, files
  - Owner - client that is allowed to use the lot
  - Capacity - amount of data that can be stored
  - Duration - time a lot is guaranteed to exist
  - Lot contains a set of files

the globus alliance
www.globus.org

GridFTP with lotman



the globus alliance
www.globus.org

Other Enhancements

- Performance enhancement
  - GridFTP over UDT
- Ease of Use enhancements
  - Alternate security mechanism
  - GridFTP Where there’s FTP

GridFTP over UDT

- UDT is an application-level data transport protocol that uses UDP to transfer data
- Implement its own reliability and congestion control mechanisms
- Achieves good performance on high-bandwidth, high-delay networks where TCP has significant limitations
- GridFTP uses Globus XIO interface to invoke network I/O operations

GridFTP over UDT

- XIO framework presents a standard open/close/read/write interface to many different protocol implementations
  - including TCP, UDP, HTTP -- and now UDT
- The protocol implementations are called drivers.
  - A driver can be dynamically loaded and stacked by any Globus XIO application.
- Created an XIO driver for UDT reference implementation
- Enabled GridFTP to use it as an alternate transport protocol

the globus alliance
www.globus.org

GridFTP over UDT

	Argonne to NZ Throughput in Mbit/s	Argonne to LA Throughput in Mbit/s
Iperf – 1 stream	19.7	74.5
Iperf – 8 streams	40.3	117.0
GridFTP mem TCP – 1 stream	16.4	63.8
GridFTP mem TCP – 8 streams	40.2	112.6
GridFTP disk TCP – 1 stream	16.3	59.6
GridFTP disk TCP – 8 streams	37.4	102.4
GridFTP mem UDT	179.3	396.6
GridFTP disk UDT	178.6	428.3
UDT mem	201.6	432.5
UDT disk	162.5	230.0

the globus alliance
www.globus.org

Alternate security mechanism

- GridFTP traditionally uses GSI for establishing secure connections
- In some situations, preferable to use SSH security mechanism
- Leverages the fact that an SSH client can remotely execute programs by forming a secure connection with SSHD
  - The client (globus-url-copy) acts as an SSH client and remotely executes a Globus GridFTP server
  - All of the standard IO from the remote program is routed back to the client.

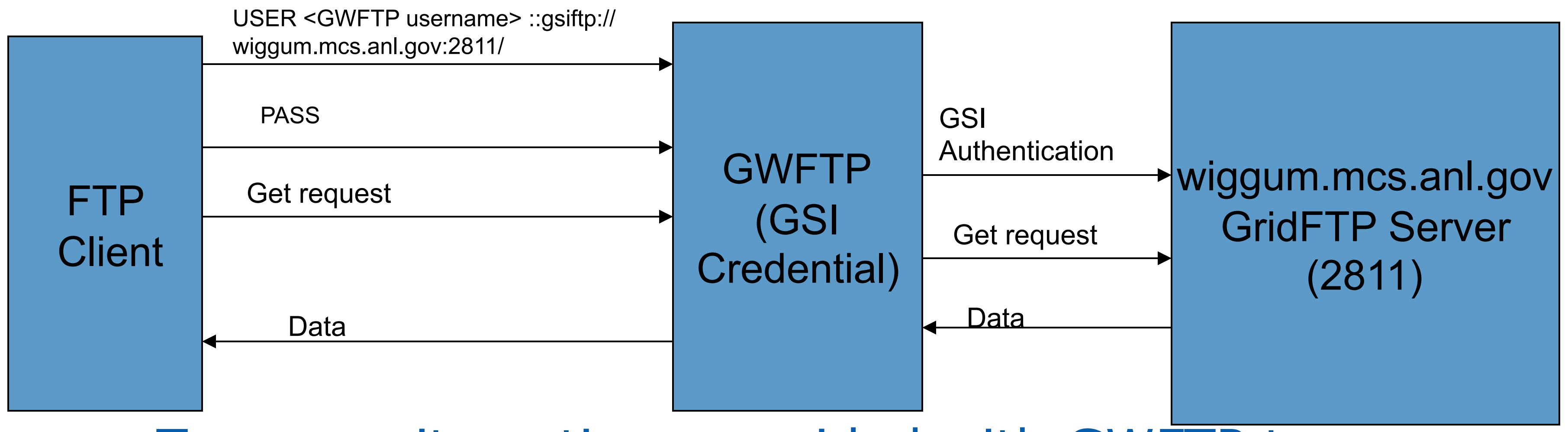
SSH security mechanism

- Client support for using SSH is automatically enabled
- On the server side (where you intend the client to remotely execute a server)
  - setup-globus-gridftp-sshftp -server
- In order to use SSH as a security mechanism, the user must provide urls that begin with sshftp:// as arguments.
  - globus-url-copy sshftp://<host>:<port>/<filepath> file:/<filepath>
  - <port> is the port in which sshd listens on the host referred to by <host> (the default value is 22).

GridFTP Where there's FTP (GWFTP)

- GridFTP has been in existence for some time and has proven to be quite robust and useful
- Only few GridFTP clients available
- FTP has innumerable clients
- GWFTP - created to leverage the FTP clients
- A proxy between FTP clients and GridFTP servers

GWFTP

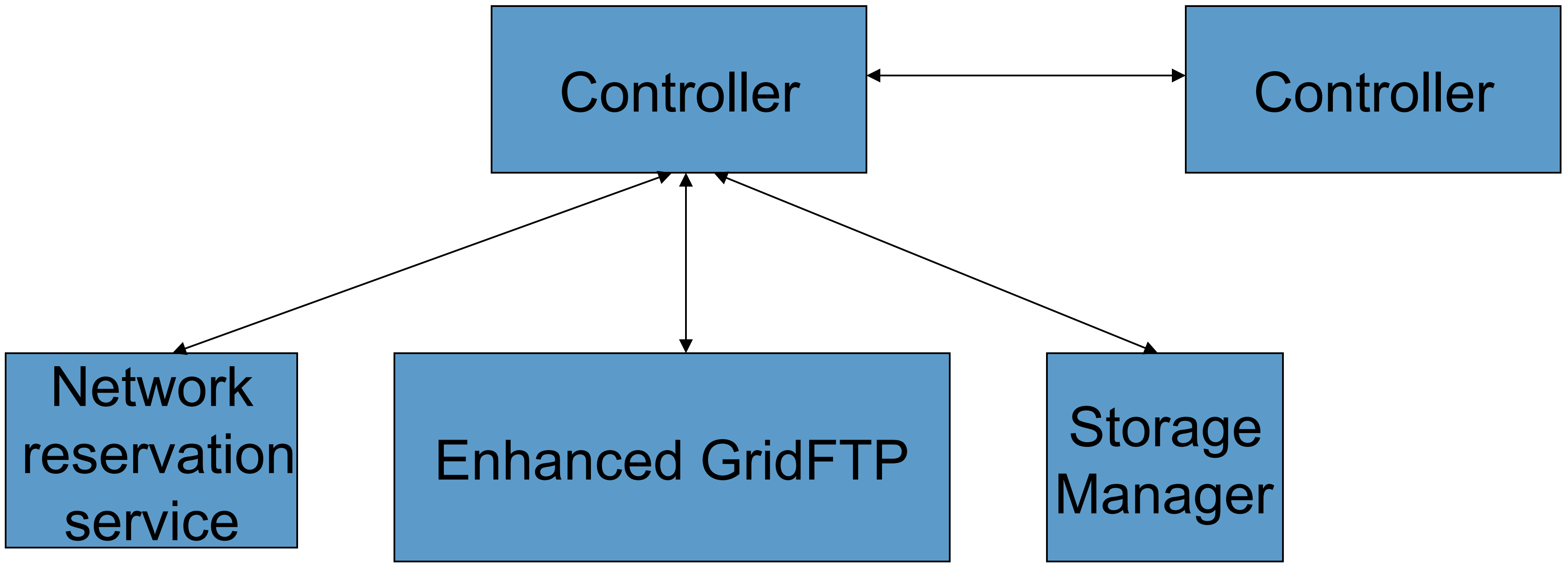


FTP Client
USER <GWFTP username> ::gsiftp://wiggum.mcs.anl.gov:2811/
PASS
Get request
Data
GWFTP (GSI Credential)
GSI Authentication
Get request
Data
wiggum.mcs.anl.gov GridFTP Server (2811)

- Two security options provided with GWFTP to authenticate its client
  - Password based and host based

Future direction

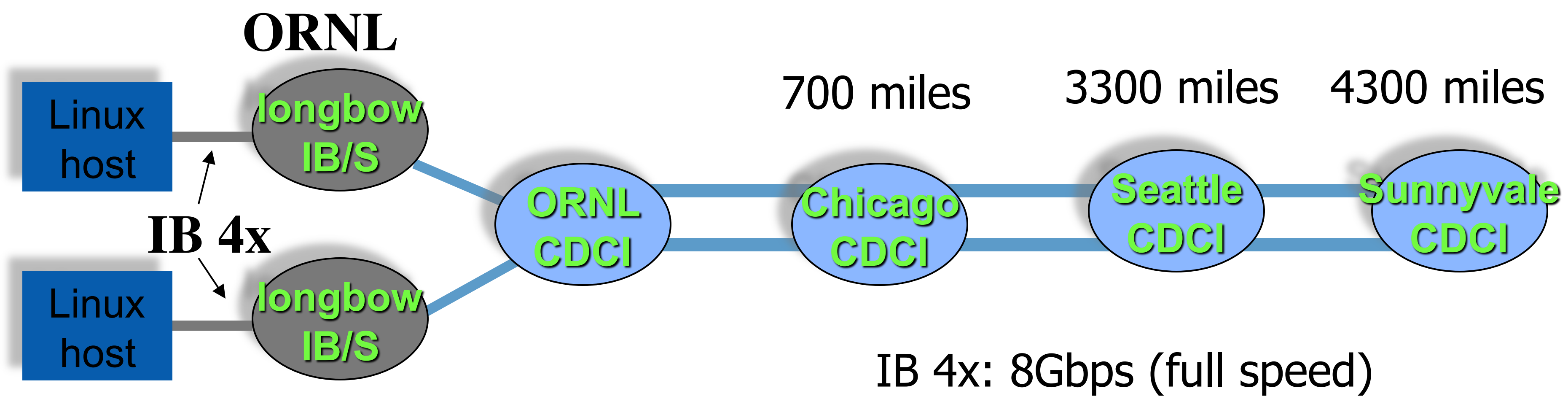
- Enhance GridFTP servers to expose status as WS resource properties
- Take advantage of the network provisioning services
    - Collaborating with Terapath and LambdaStation projects



Infiniband over SONET

Need specialized hardware: Obsidian longbow

1. IB over SONET/Ethernet – frame conversion
2. Buffer-based termination of IB flow control



IB 4x: 8Gbps (full speed)
Host-to-host local switch:7.5Gbps

ORNL loop -0.2 mile: **7.5**Gbps

ORNL-Chicago loop – 1400 miles: **7.46**Gbps

ORNL- Chicago - Seattle loop – 6600 miles: **7.23**Gbps

ORNL – Chicago – Seattle - Sunnyvale loop – 8600 miles: **7.20**Gbps

the globus alliance
www.globus.org

GridFTP over Infiniband

- Can use infiniband through Sockets Direct Protocol (SDP)
- SDP provides a socket interface for Open Fabrics software stack (a standard implemented by infiniband and iwarp)
    - No kernel bypass
- User level verbs to interface directly with infiniband hardware
    - Develop a XIO driver for verbs interface