Berkeley Storage Manager (BeStMan)

Alex Sim

Scientific Data Management Research Group
Computational Research Division
Lawrence Berkeley National Laboratory

What is SRM?

- **SRM : Storage Resource Manager**
  - Well-defined storage management interface specification based on standard
  - Different implementations for underlying storage systems are based on the same SRM specification
  - Provides dynamic space allocation and file management on shared storage components on the Grid
- **Over 300 deployments of different SRM servers in the world**
  - Managing more than 10 PB

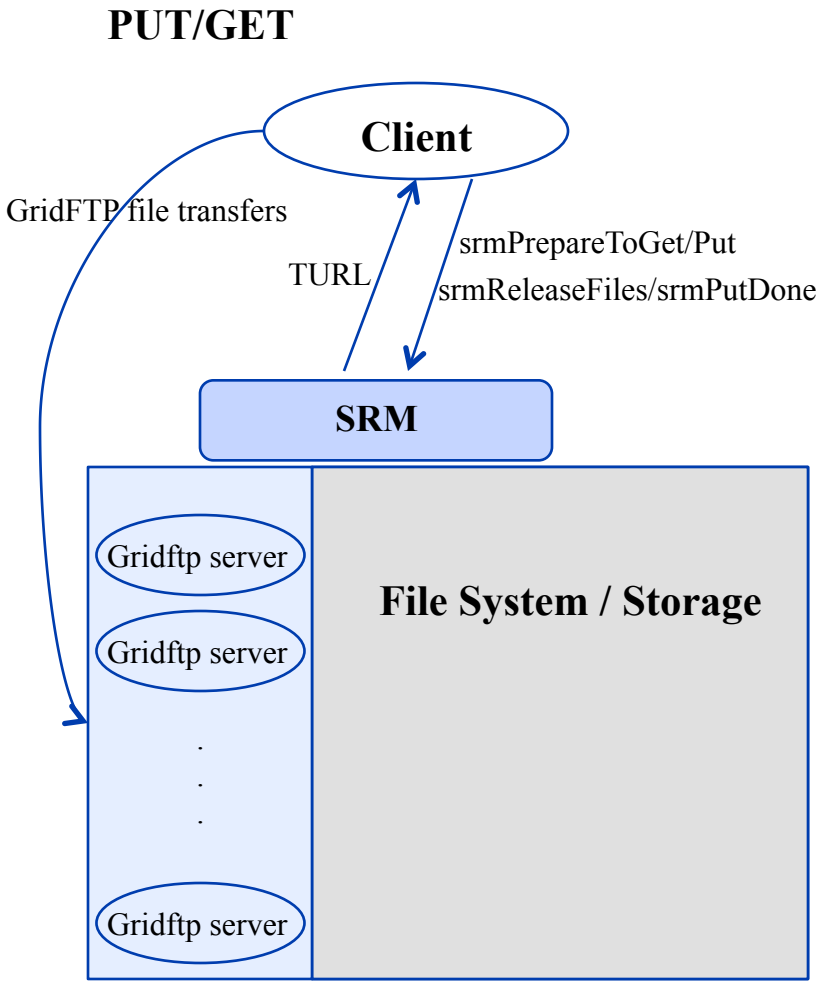
Why do you need SRM?

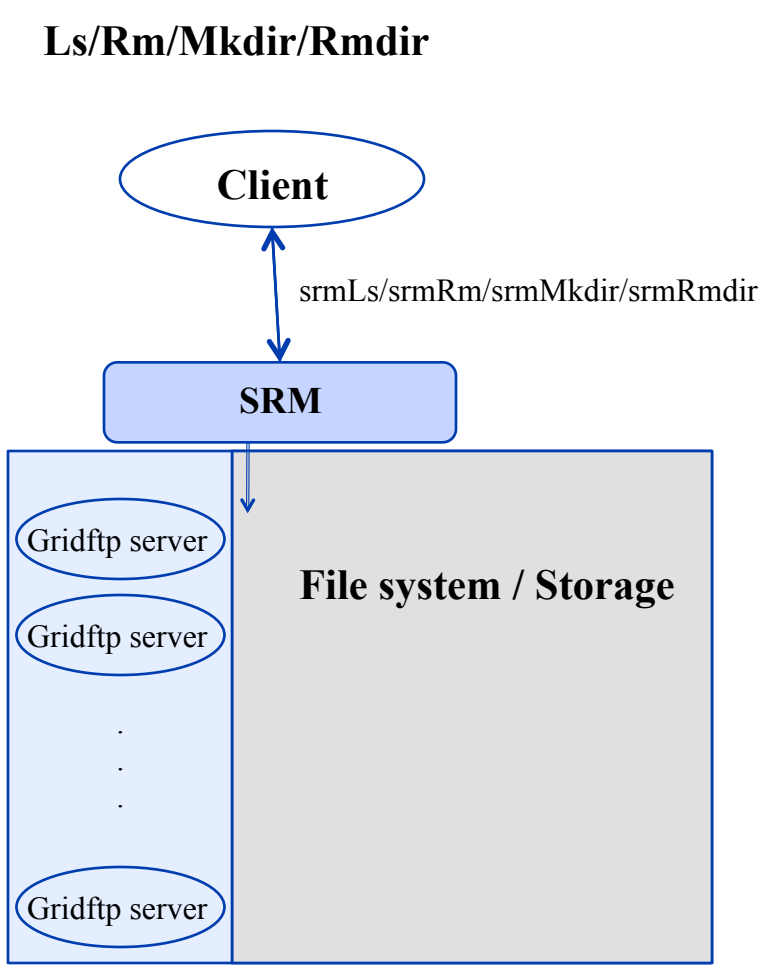
- Suppose you want to run a job on your local machine
  - Need to allocate space, and bring all input files
  - Need to ensure correctness of files transferred
  - Need to monitor and recover from errors
  - What if files don't fit space? Need to manage file "streaming"
    - Need to remove files to make space for more files
  - Need to remove files after the job is done for more jobs
- Now, suppose that the machine and storage space is a shared resource
  - Need to do the above for many users,
    - Need to enforce quotas
    - Need to ensure fairness of space allocation and scheduling
- Now, suppose you want to do that on a Grid
  - Need to access a variety of storage systems
    - mostly remote systems, need to have access privileges
    - Need to have special software to access mass storage systems
- Now, suppose you want to run distributed jobs on the Grid
  - Need to allocate remote spaces
  - Need to copy (or stream) files to remote sites
  - Need to manage file outputs and their transfer to destination site(s)

What does SRM do?

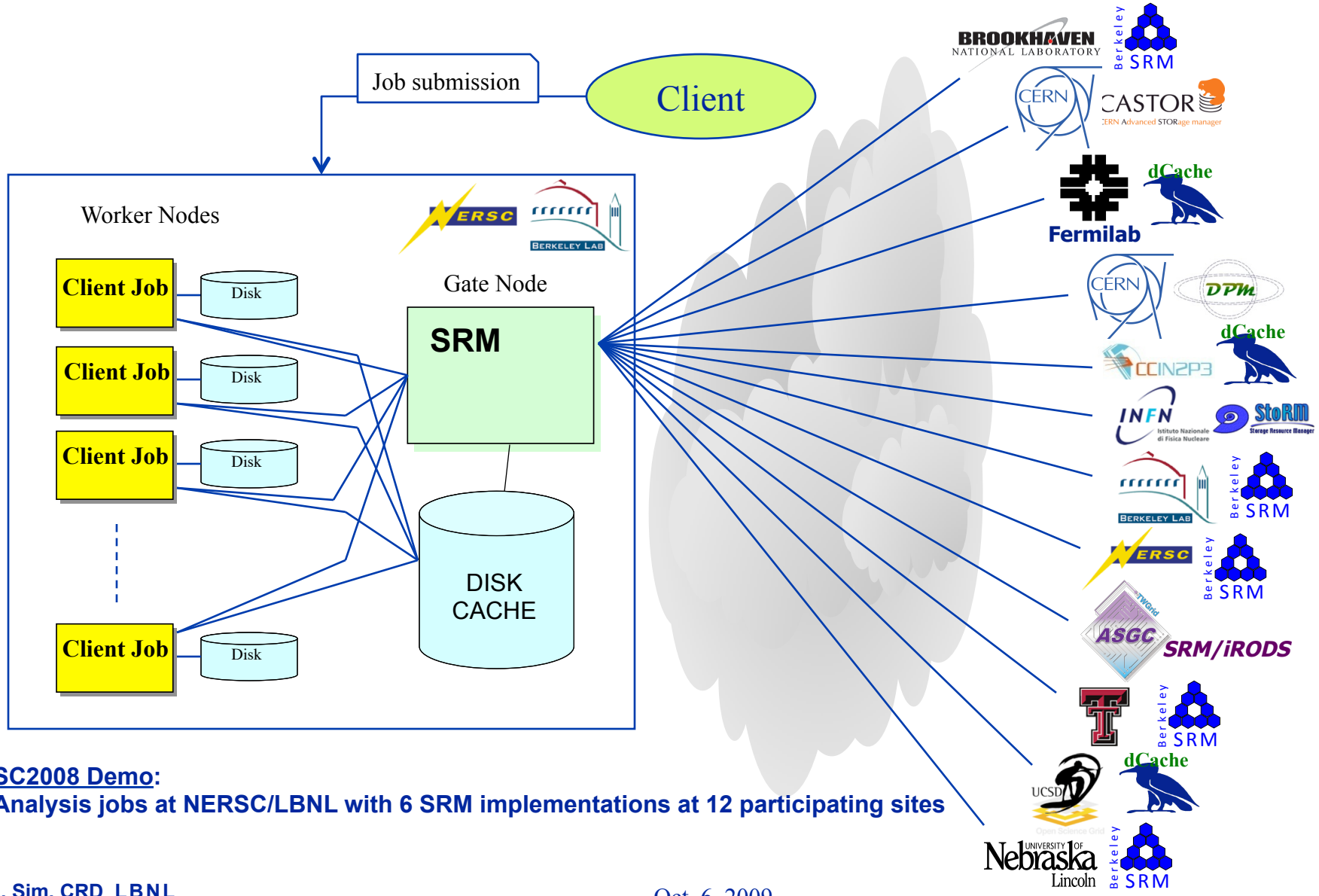
- Support space management for files with lifetime
  - Allocation of space, garbage collection
- Support dynamic space reservation – opportunistic storage
- Support for multiple file transfer protocols
  - Support for transfer protocol negotiation
  - Support for multiple file transfer servers
  - Incoming and outgoing file transfer queue management and transfer monitoring
- Support for asynchronous multi-file requests
- Directory management and ACLs
- Support file sharing and file streaming
- Gives compatibility and interoperability based on standard

What do users do with SRM?





SRMs Facilitate Analysis Jobs



SC2008 Demo:
Analysis jobs at NERSC/LBNL with 6 SRM implementations at 12 participating sites

Berkeley Storage Manager (BeStMan)

- Light-weight implementation of SRM v2.2
  - Works on existing disk storages with posix compliant file systems
    - E.g. NFS, GPFS, GFS, NGFS, PNFS, HFS+, PVFS, Lustre, Xrootd, Hadoop, Ibrix
  - Supports multiple partitions
  - Adaptable to other file systems and storages
    - Supports customized plug-in for file system access
    - Supports customized plug-in for MSS to stage/archive such as HPSS
  - Easy adaptability and integration to special project environments
- Supports multiple transfer protocols
  - Supports load balancing for multiple transfer servers
- Scales well with some file systems and storages
  - Xrootd, Hadoop
- Works with grid-mapfile or GUMS server
- Simple installation and easy maintenance
- Packaged in VDT using Pacman
- Who would benefit from BeStMan?
  - Sites with limited resources and/or limited admin effort
  - Users can run their own BeStMan server

What can BeStMan do?

- In addition to what SRMs do….
- Dynamic installation, configuration and running
  - If the target host does not have an SRM, BeStMan can be installed, configured, and started with a few commands by the user.
- BeStMan can restrict all user access to certain directory paths through configuration
- A site can customize the load-balancing mechanism for transfer servers through plug-in

Difference between BeStMan Full mode and BeStMan Gateway mode

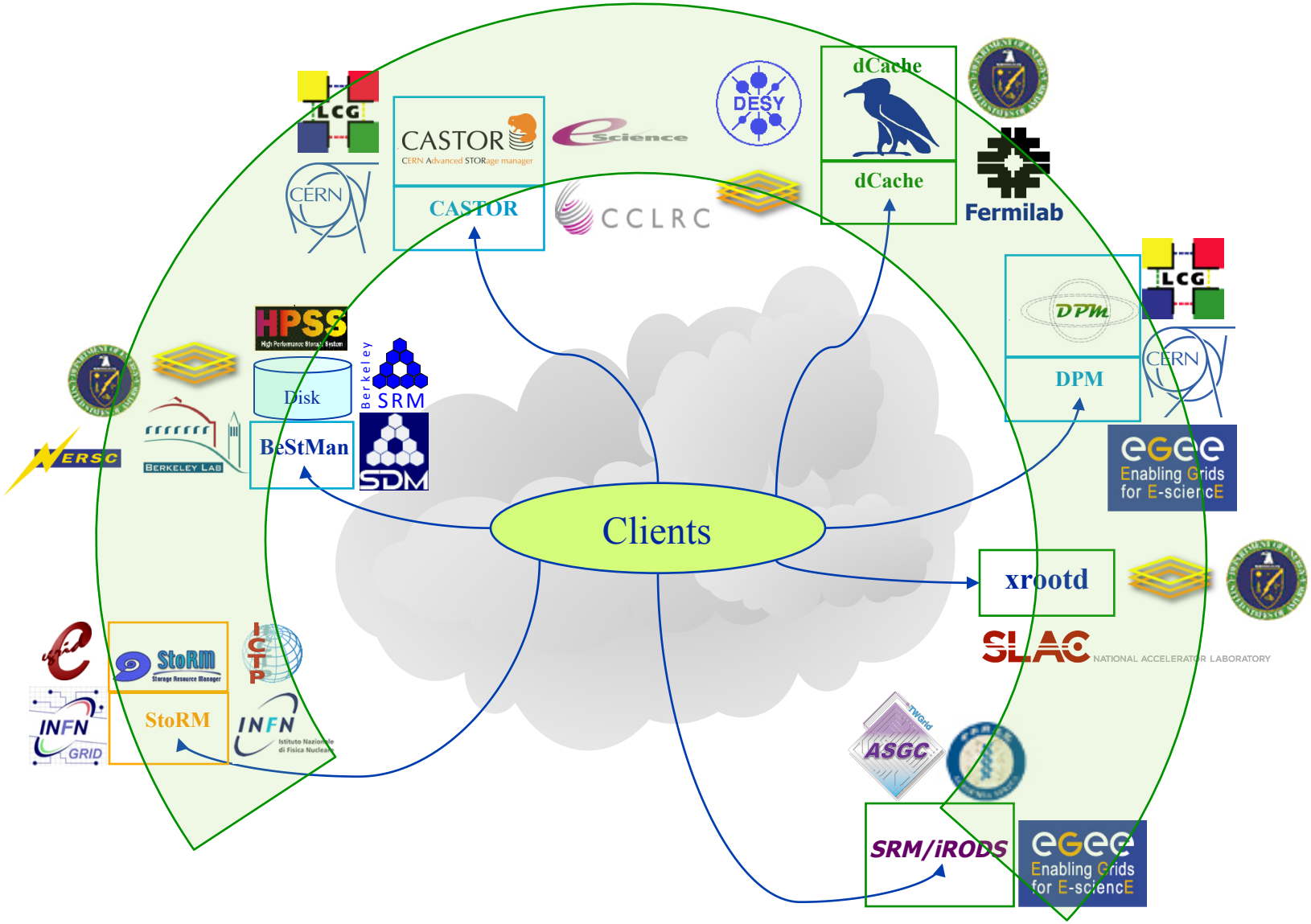
- Full implementation of SRM v2.2
- Support for dynamic space reservation
- Support for request queue management and space management
- Plug-in support for mass storage systems

- Support for essential subset of SRM v2.2
- Support for pre-defined static space tokens
- Faster performance without queue and space management

Some BeStMan Use Cases

- CMS
  - BeStMan Gateway as an SRM frontend for Hadoop at Caltech, UCSD, UNL
  - Passed all the automated CMS tests through EGEE SAM
- ATLAS
  - BeStMan Gateway on Xrootd/FS, GPFS, Ibrix
- STAR
  - Data replication between BNL and NERSC/LBNL
    - HPSS access at BNL and NERSC
    - SRMs in production for over 4 years
  - Part of analysis scenario to move job-generated data files from PDSF/NERSC to remote BNL storage
- Earth System Grid
  - Serving about 13000 users
    - Over a million files and 170TB of climate data
    - from 5 storage sites (LANL disks, LLNL disks, NCAR HPSS, NERSC HPSS, ORNL HPSS)
  - Uses an adapted BeStMan for NCAR's own MSS and HPSS

Interoperability with other SRMs

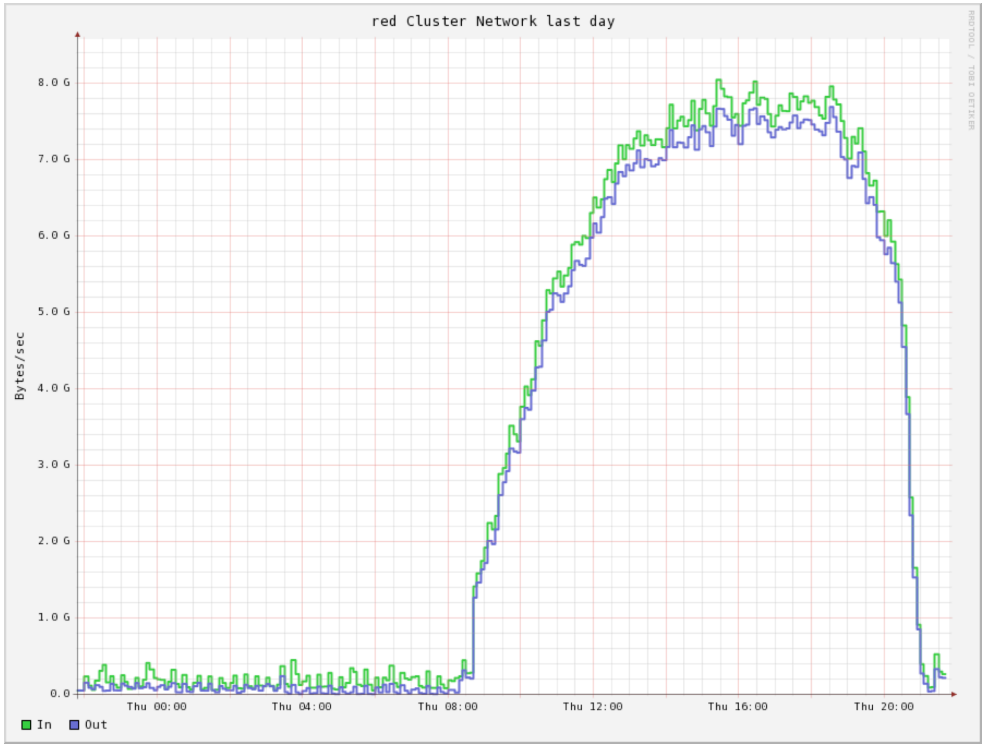


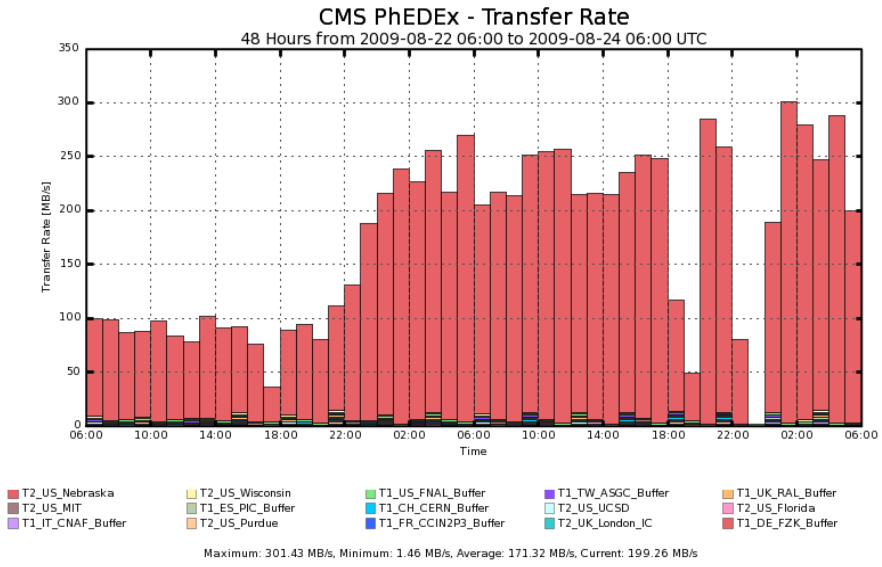
SRM client runs (1)

- **Ping: srm-ping**
  - srm-ping checks the SRM server. In response to the call, SRM server returns the SRM version number as well as other backend information.
  - srm-ping srm://hostname:port/service_handle
- **Put: srm-copy**
  - srm-copy requests to copy files to and from SRM, between SRMs, between SRM and other storage repository, depending on the source and target URLs.
  - srm-copy file:////local_file_path srm://hostname:port/sevice_handler\?SFN=/remotefilepath
- **Get: srm-copy**
  - srm-copy srm://hostname:port/sevice_handler\?SFN=/remotefilepath file:////local_file_path
- **Ls: srm-ls**
  - srm-ls srm://hostname:port/service_handle\?SFN=/file_path
- **Rm: srm-rm**
  - srm-rm srm://hostname:port/service_handle\?SFN=/file_path
- **Mkdir: srm-mkdir**
  - srm-mkdir srm://hostname:port/service_handle\?SFN=/dir_path
- **Rmdir: srm-rmdir**
  - srm-rmdir srm://hostname:port/service_handle\?SFN=/dir_path

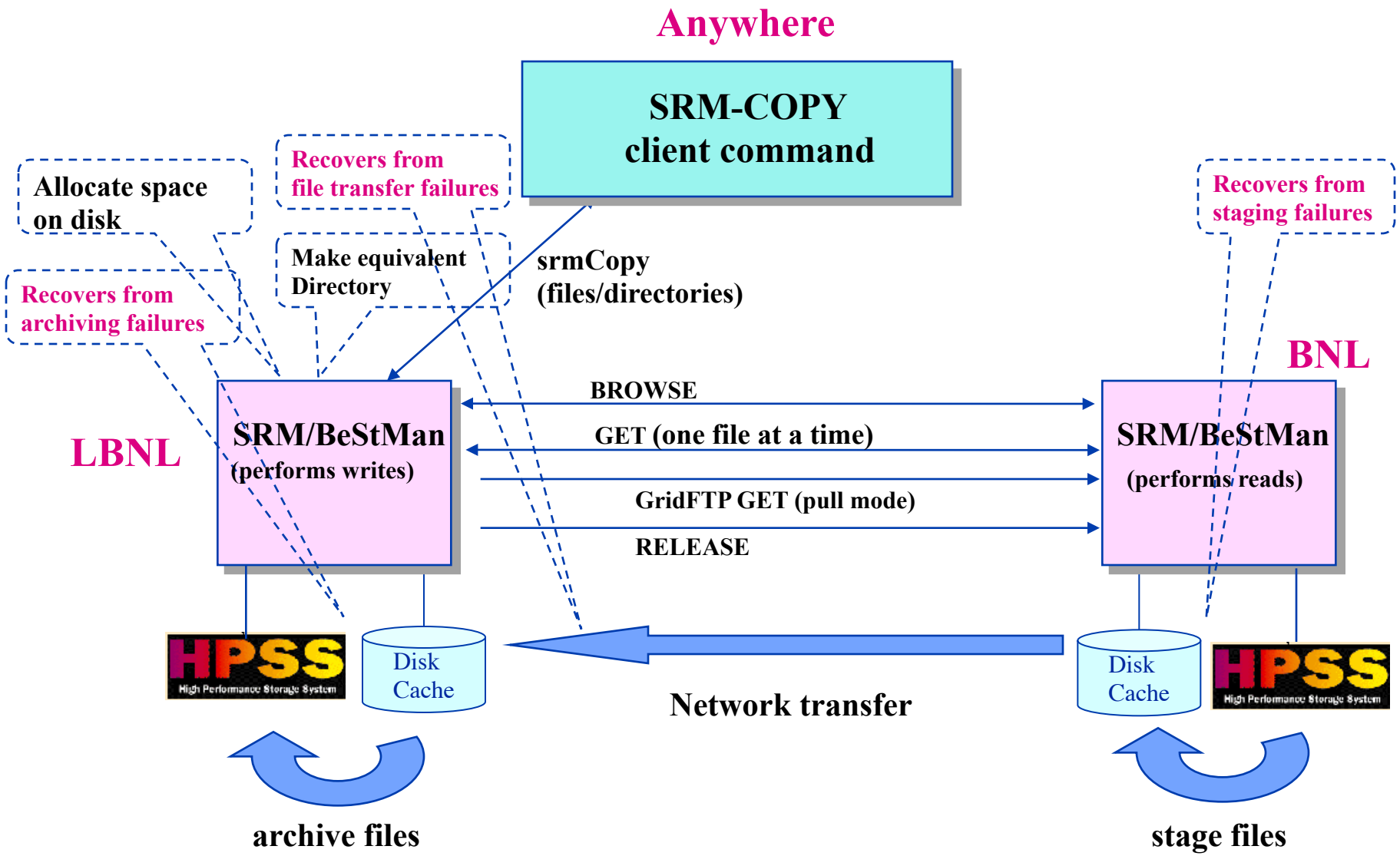
Performance

- Much depending on the hardware
- Scalable up to the network capacity and hardware limit

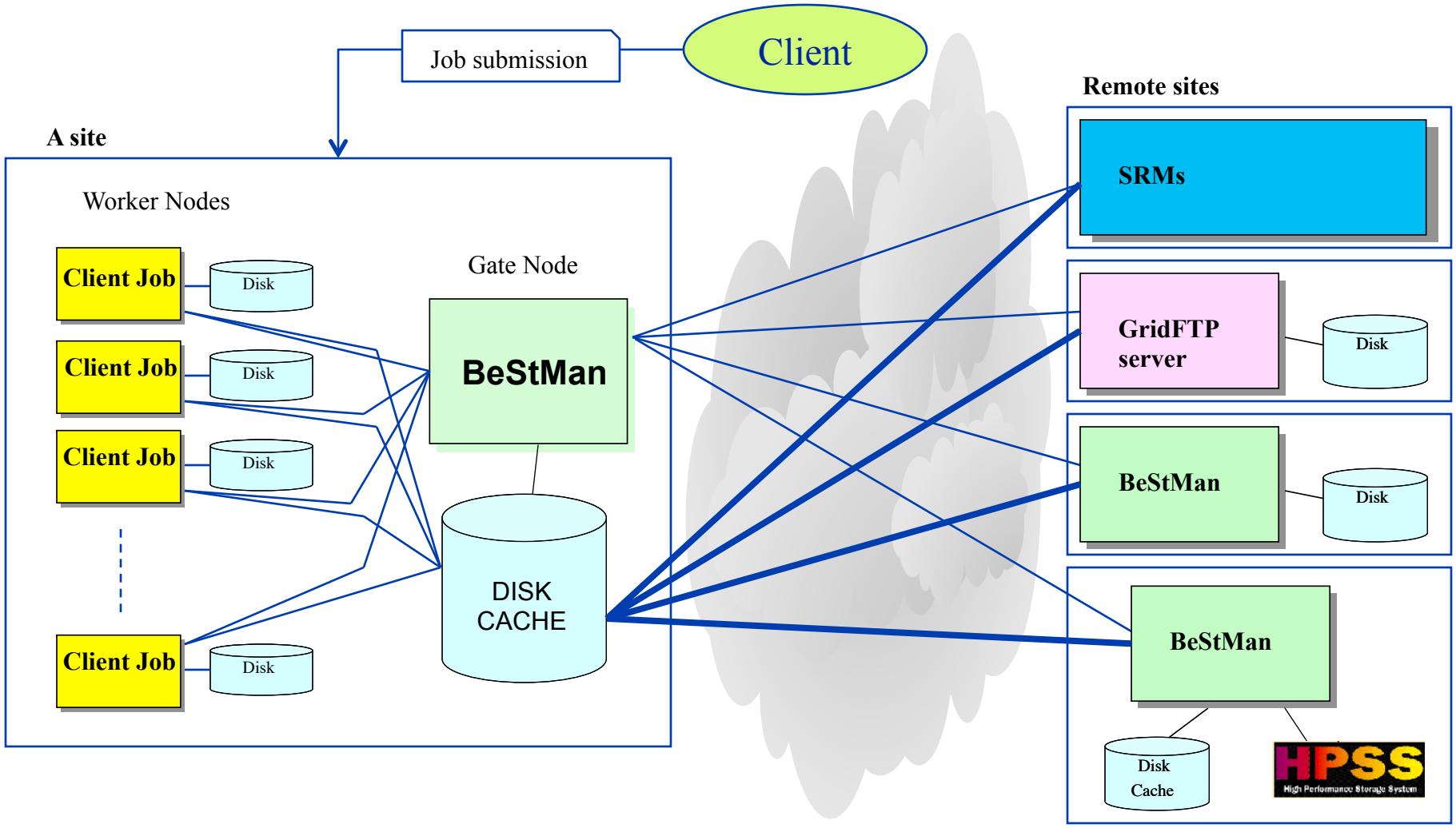




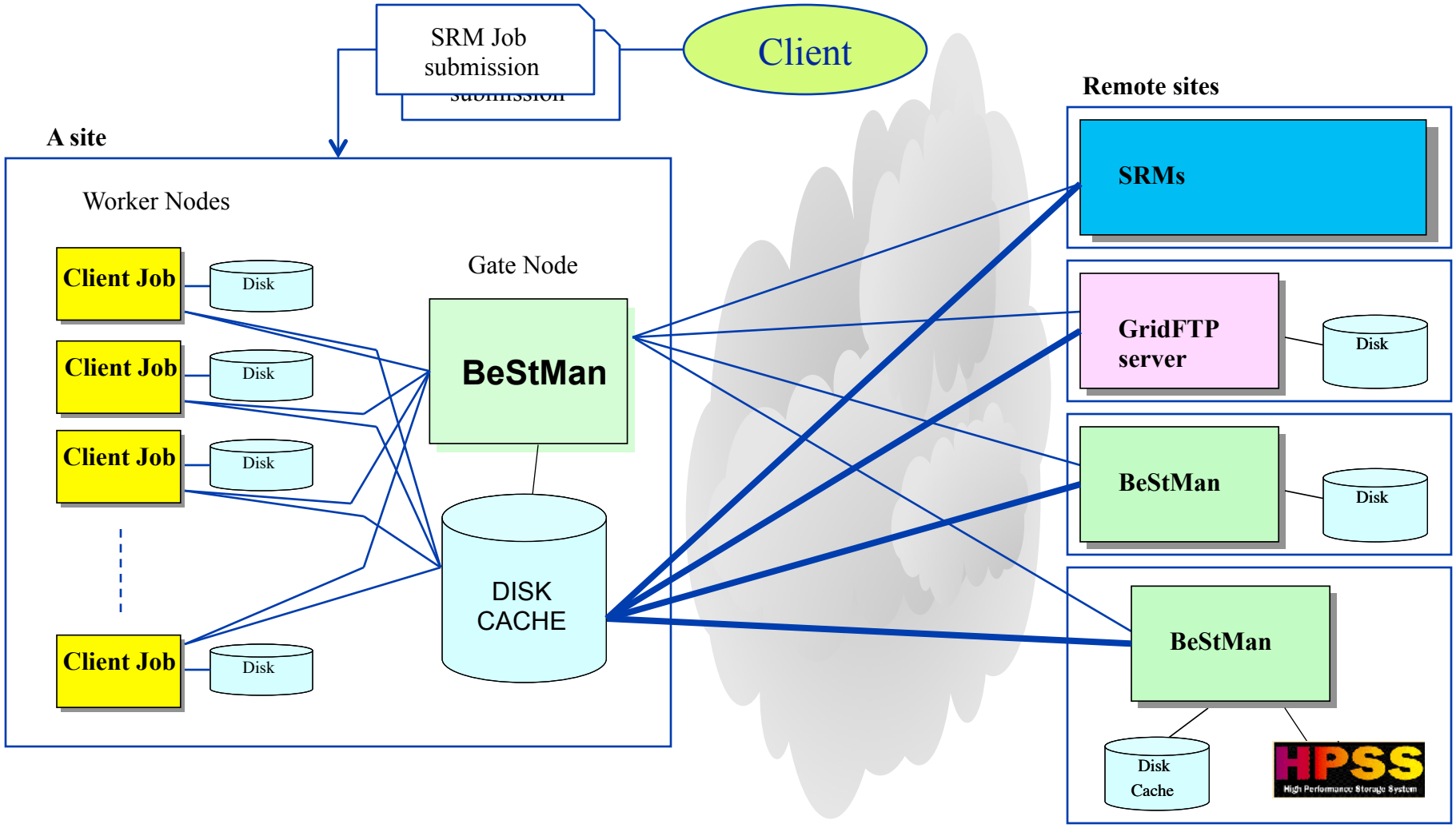
Data Replication in STAR



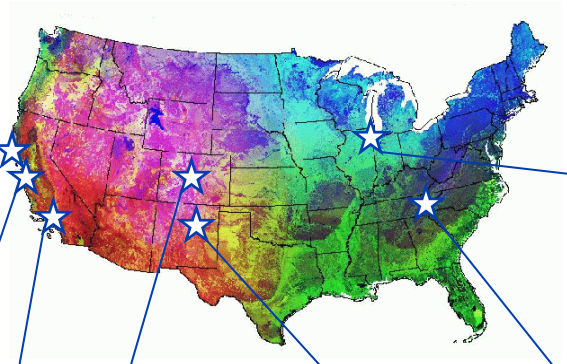
STAR Analysis scenario (1)



STAR Analysis scenario (2)



SRM works in concert with other Grid components in ESG



LBNL
DISK
HPSS
HRM Storage Resource Management
RLS
GridFTP server

ANL
GridFTP service
Globus Security infrastructure

NCAR
ESG Portal
User DB
XML data catalogs
ESG CA
ESG Metadata DB
MyProxy
OPeNDAP-g
RLS
LAHFS
GridFTP server
HRM Storage Resource Management
MSS Mass Torage System
DISK

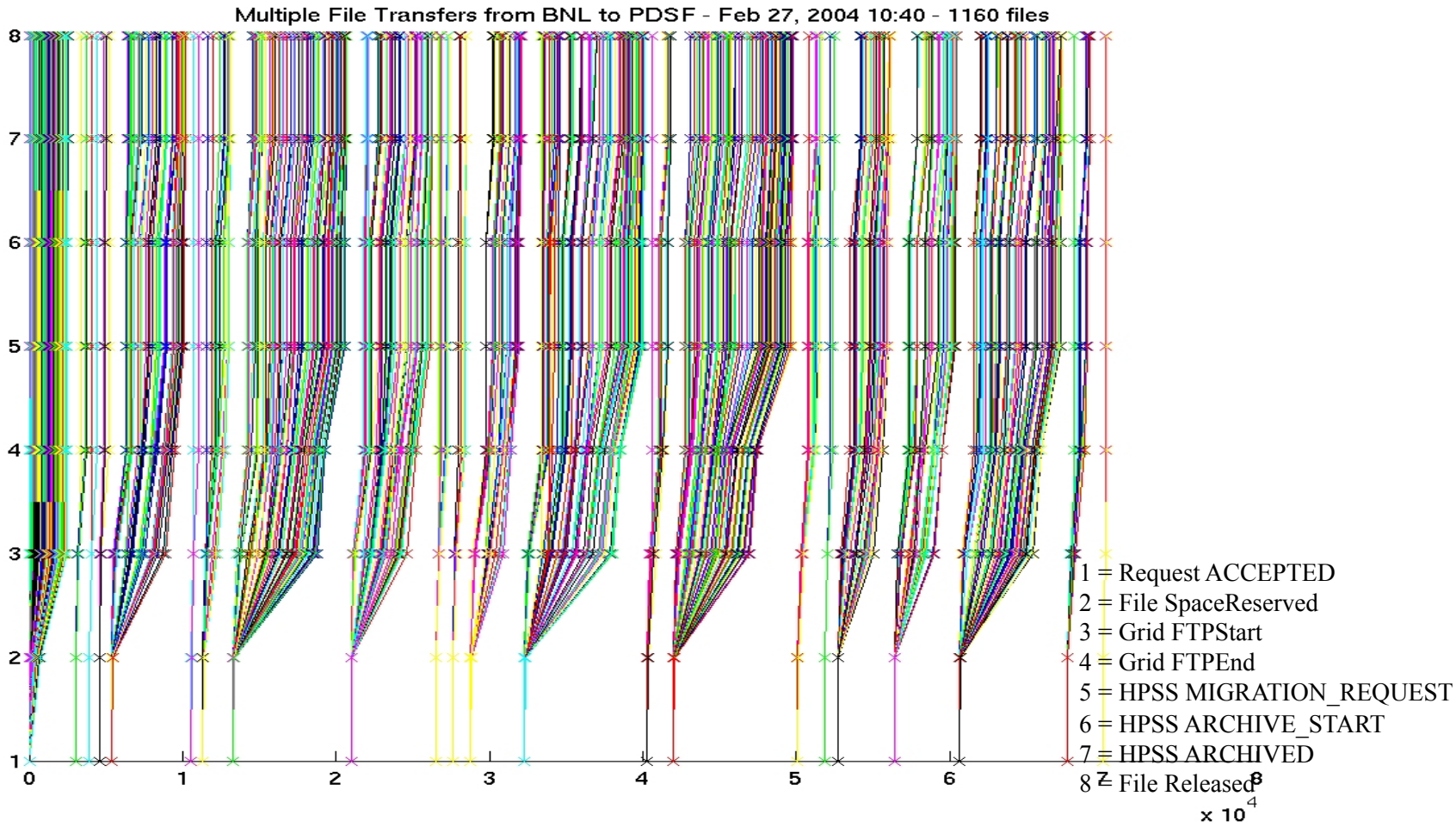
LLNL
DISK
XML data catalogs
IPCC Portal
RLS
DRM Storage Resource Management
GridFTP server
FTP server

ORNL
RLS
GridFTP server
HRM Storage Resource Management
HPSS
DISK

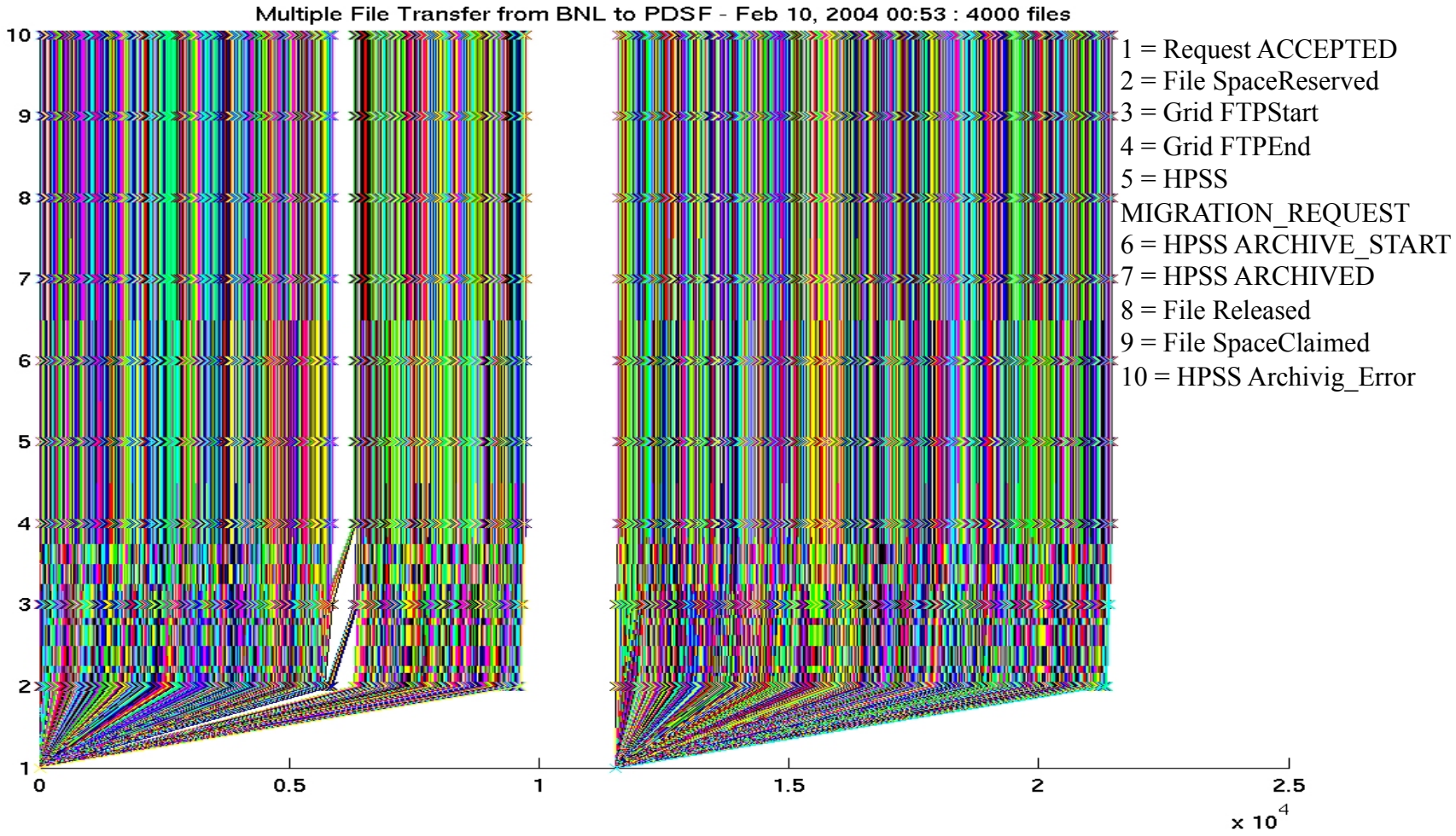
ISI
MCS Metadata Cataloguing Services
RLS Replica Location Services
Monitoring Discovery ervices

LANL
DISK
RLS
DRM Storage Resource Management
GridFTP server

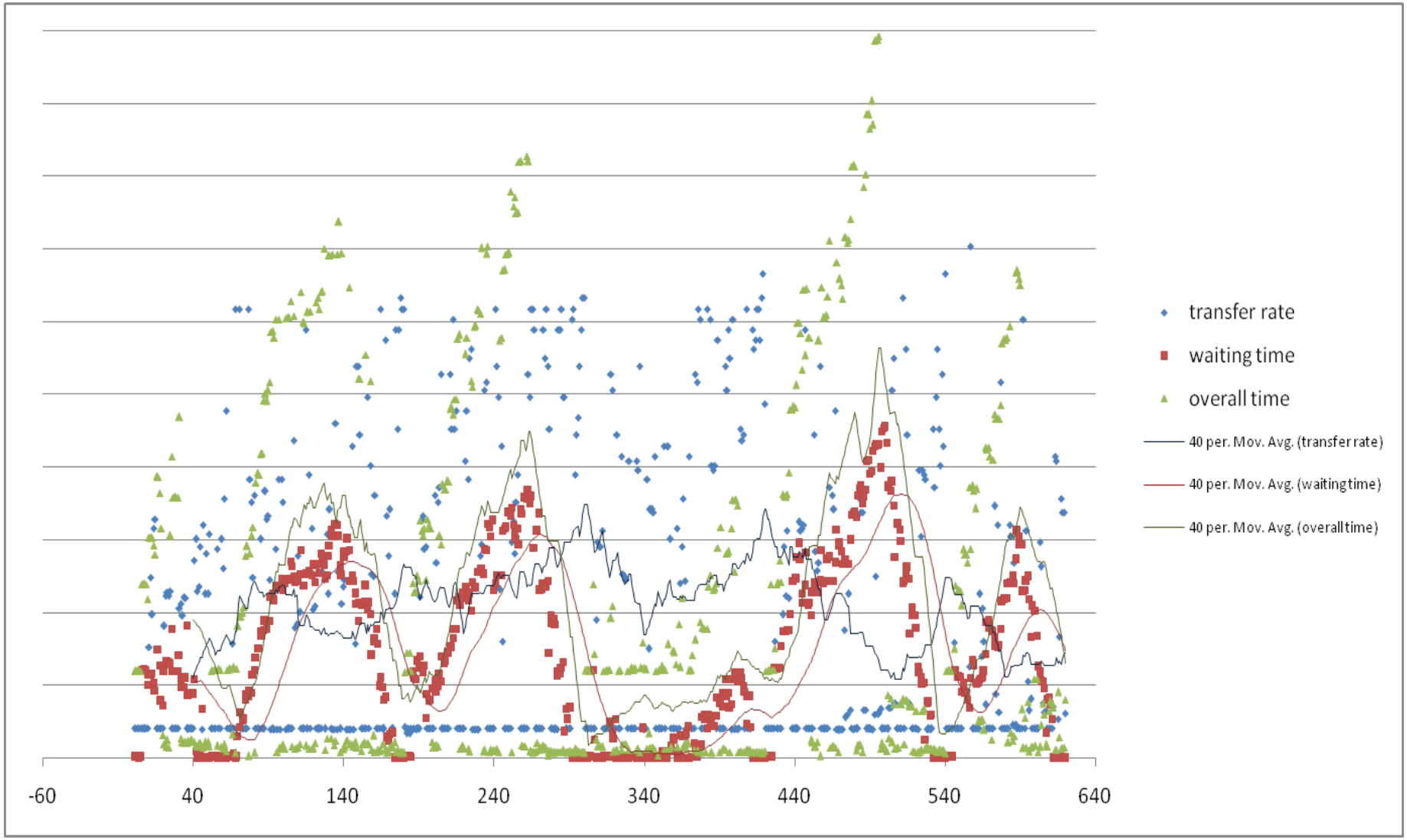
Multi-file Transfer plot from BNL to LBNL (27/02/04)



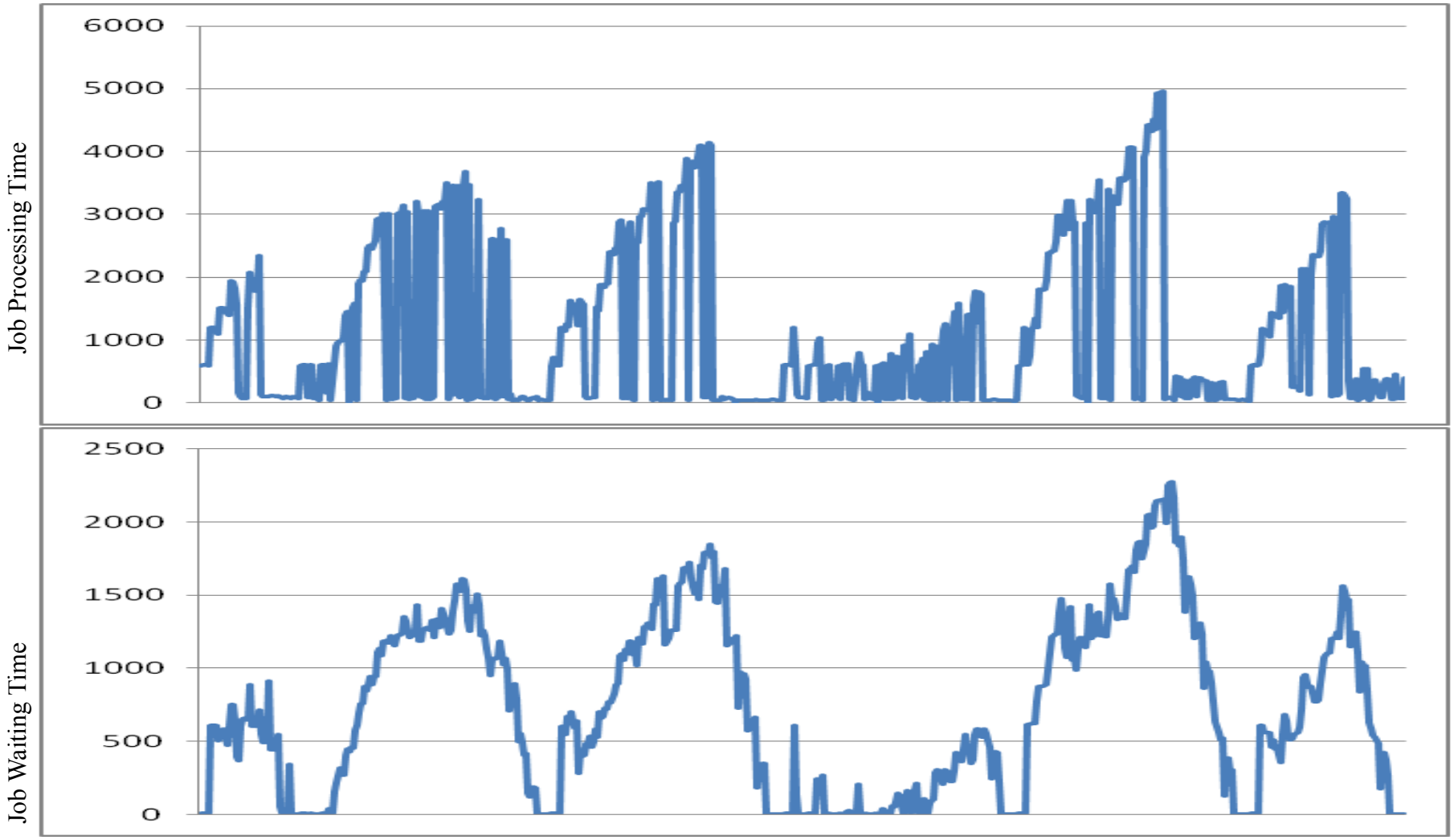
Multi-file Transfer plot from BNL to LBNL (10/02/04)



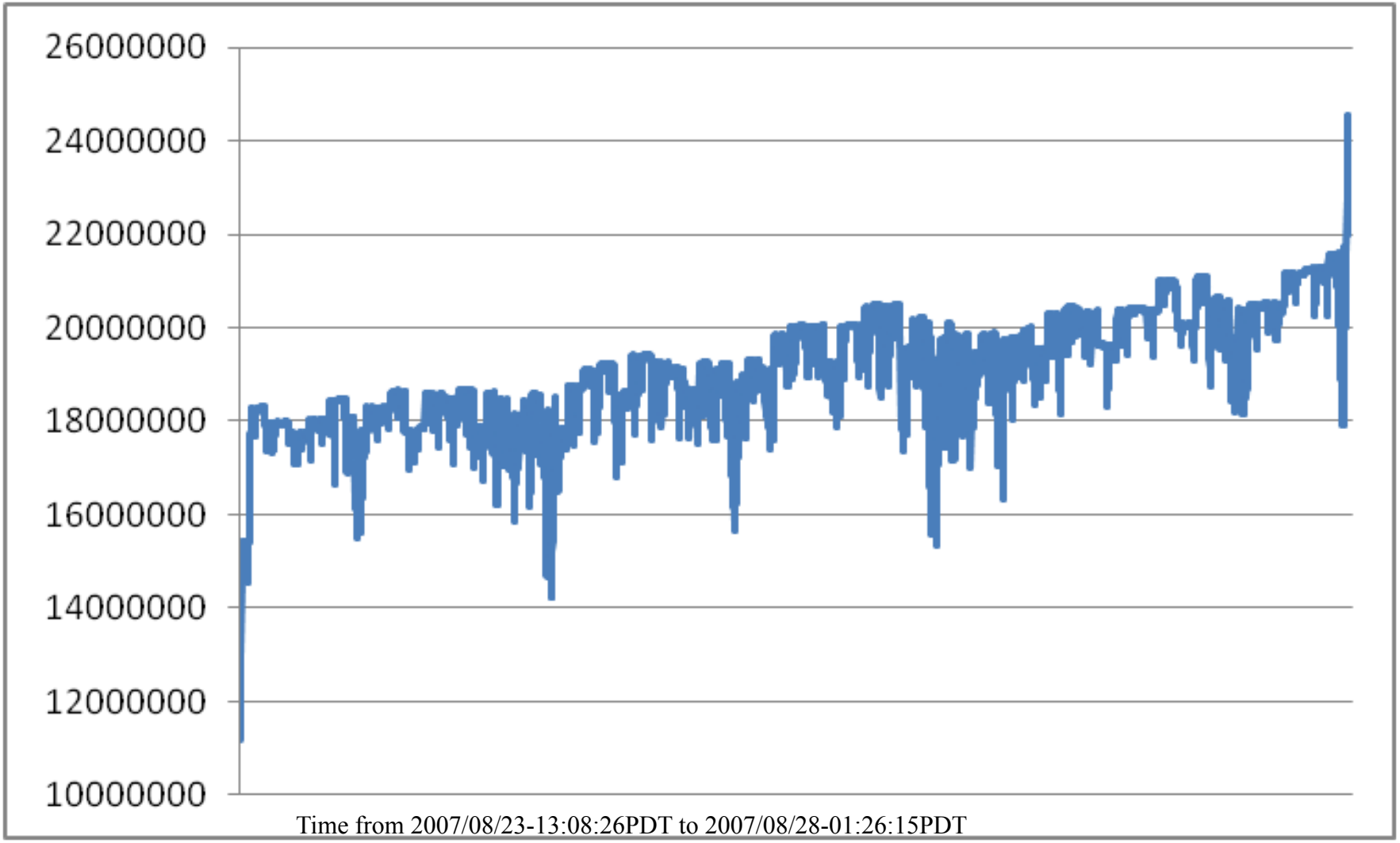
BeStMan File Jobs on 8/26/2007



BeStMan File Job Processing Time on 8/26/2007

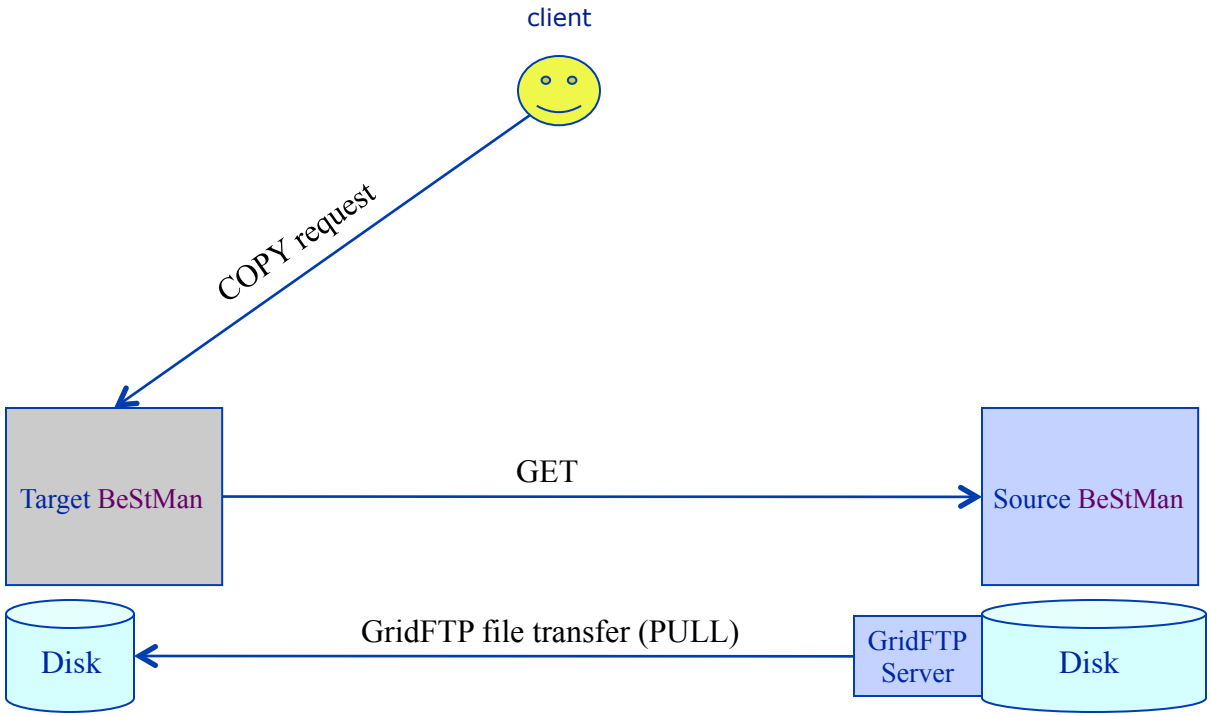


Memory usage of the BeStMan over a week (8/23/2007-8/28/2007) handling about 4000 files



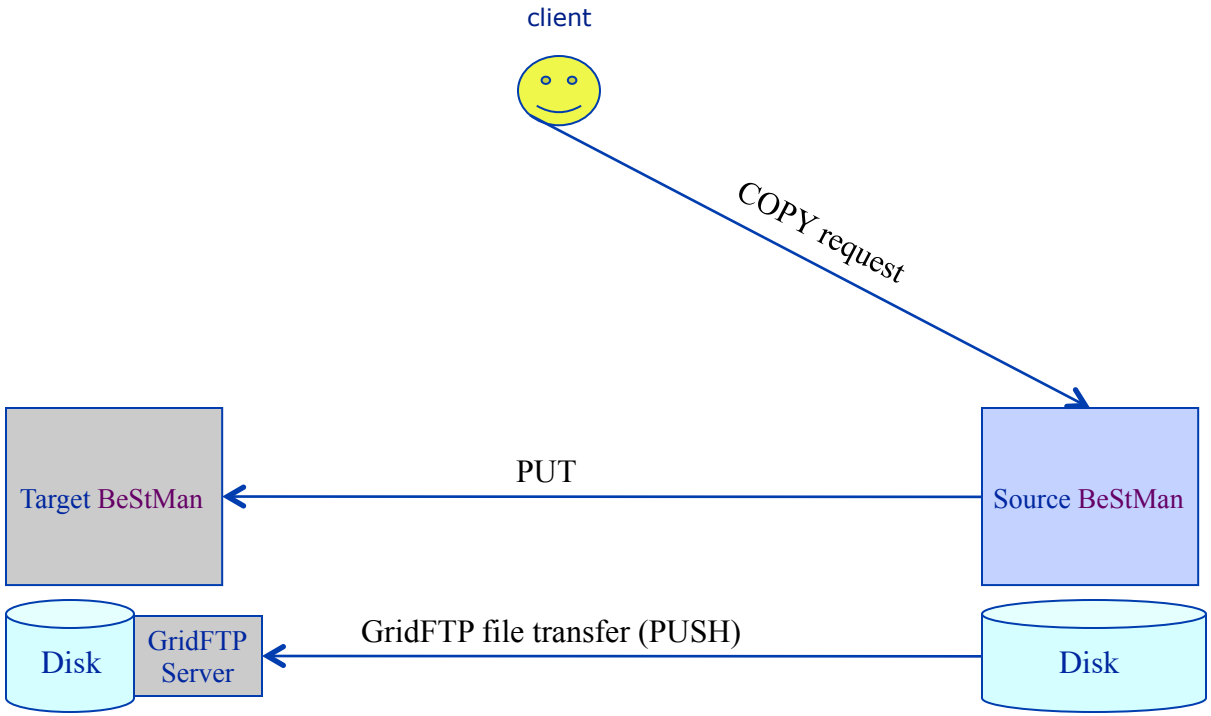
How can this be used with applications?

Use Case 1 - Copy/PULL



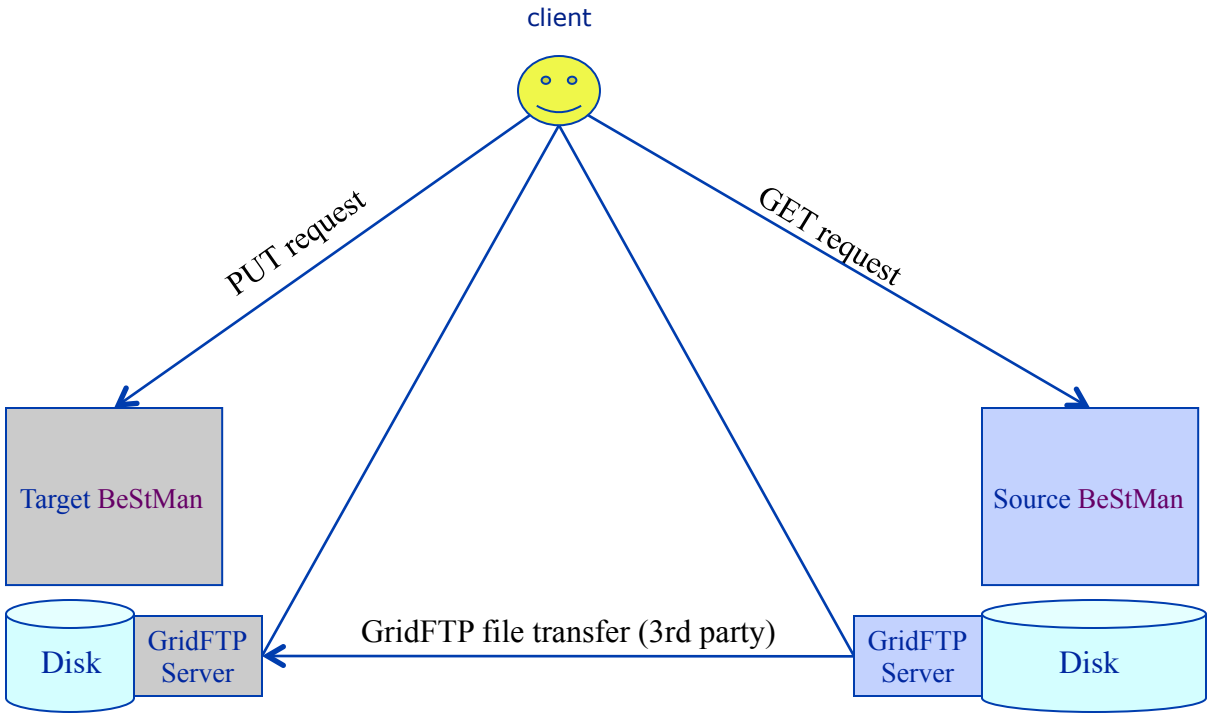
srm-copy srm://SOURCE:port/srm/v2/server\?SFN=/filepath srm://TARGET:port/srm/v2/server\?SFN=/filepath

Use Case 2 – Copy/PUSH



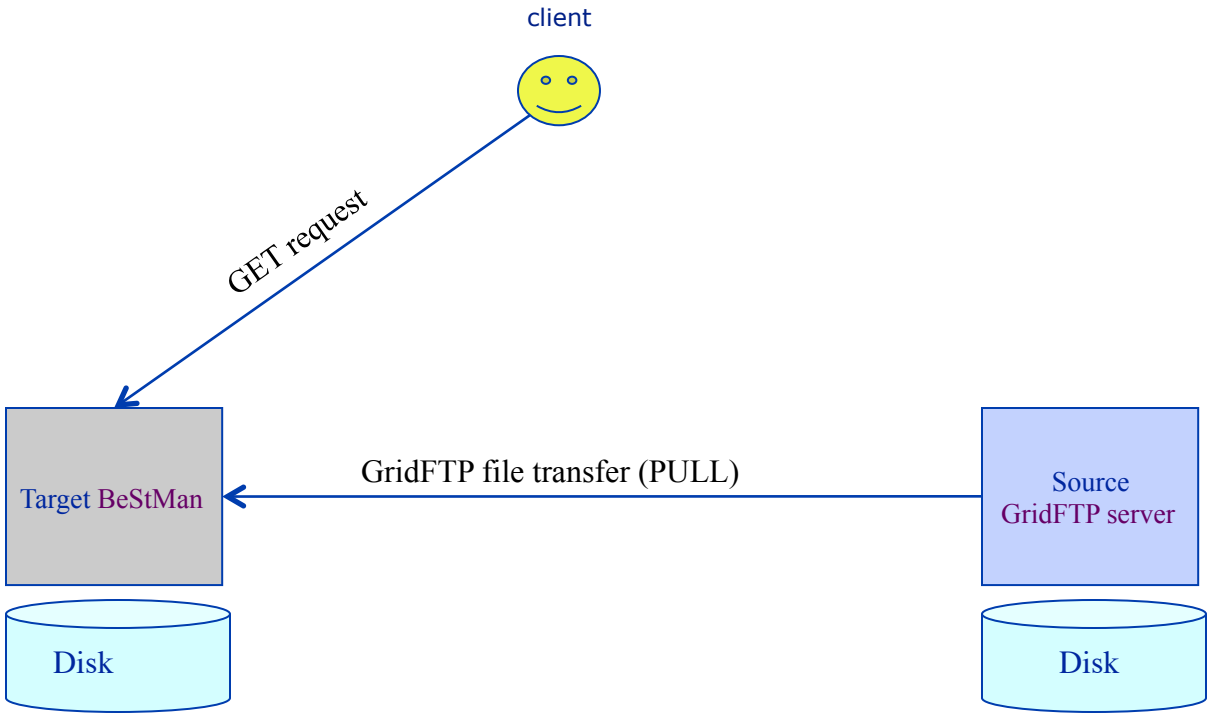
srm-copy srm://SOURCE:port/srm/v2/server\?SFN=/filepath srm://TARGET:port/srm/v2/server\?SFN=/filepath -pushmode

Use Case 3 – Copy/3rdParty



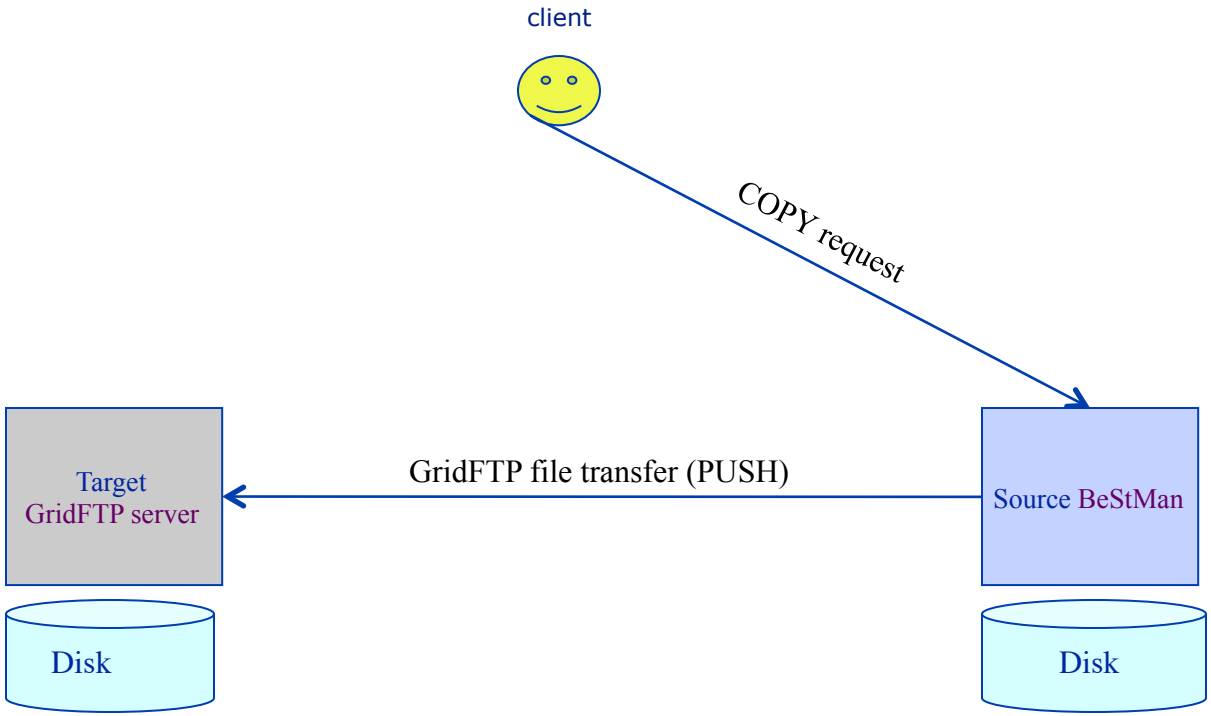
srm-copy srm://SOURCE:port/srm/v2/server\?SFN=/filepath srm://TARGET:port/srm/v2/server\?SFN=/filepath -3partycopy

Use Case 4 - Copy/PULL



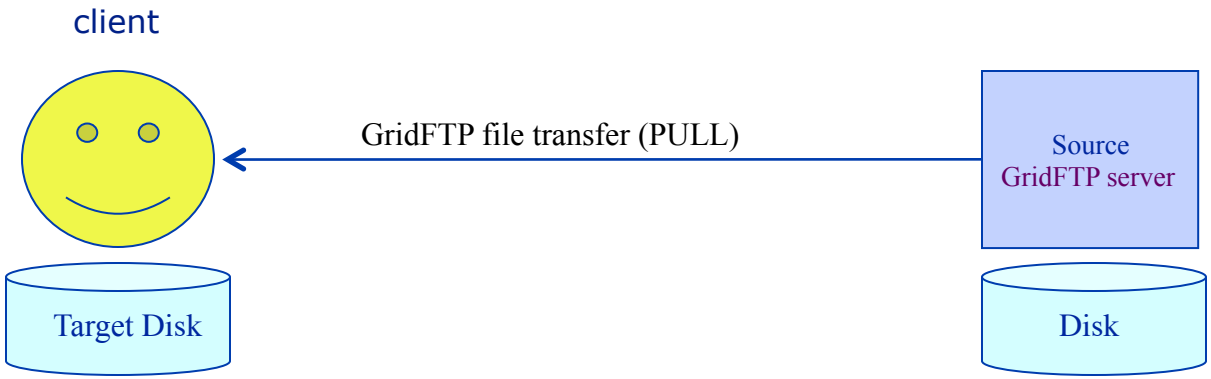
srm-copy gsiftp://SOURCE:port//filepath srm://TARGET:port/srm/v2/server\?SFN=/filepath

Use Case 5 – Copy/PUSH



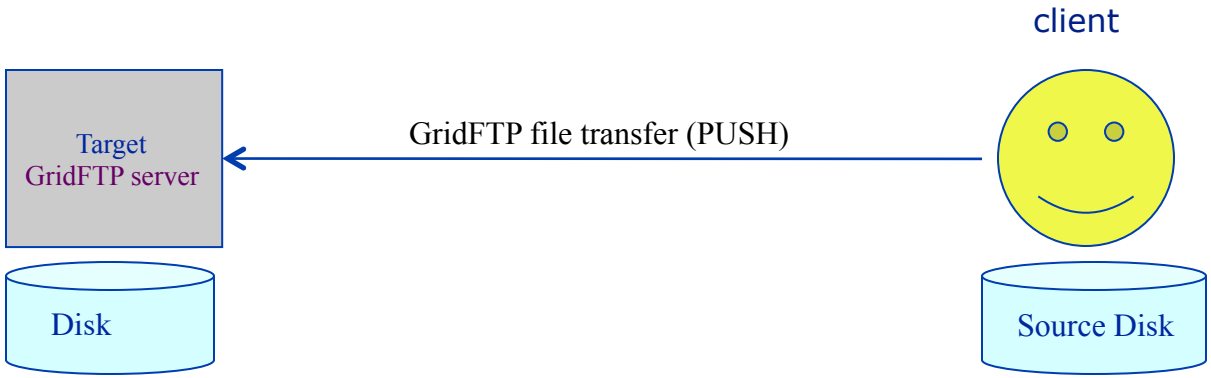
srm-copy srm://SOURCE:port/srm/v2/server\?SFN=/filepath gsiftp://TARGET:port//filepath -pushmode

Use Case 6 - Copy/PULL



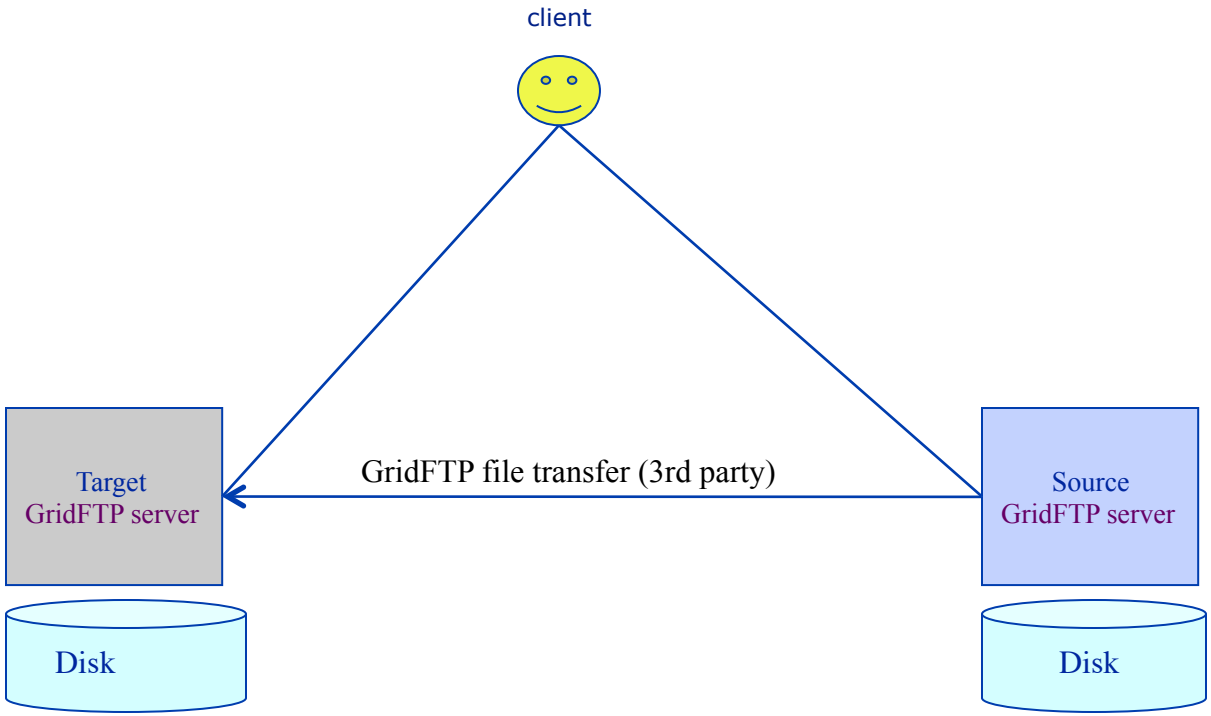
srm-copy gsiftp://SOURCE:port//filepath file:////filepath
globus-url-copy gsiftp://SOURCE:port//filepath file:////filepath

Use Case 7 – Copy/PUSH



srm-copy file:////filepath gsiftp://TARGET:port//filepath
globus-url-copy file:////filepath gsiftp://TARGET:port//filepath

Use Case 8 – 3rd Party



srm-copy gsiftp://SOURCE:port//filepath gsiftp://TARGET:port//filepath
globus-url-copy gsiftp://SOURCE:port//filepath gsiftp://TARGET:port//filepath

Summary

- **BeStMan is an implementation of SRM v2.2**
  - Great for disk-based storage and file systems
  - BeStMan Gateway mode gives scalable performance on some file systems and storages
  - BeStMan full mode gives reliability and management on user requests
  - Easy installation and maintenance through VDT or tar file
  - Works with other SRM v2.2 implementations
    - Servers: CASTOR, dCache, DPM, StoRM, SRM/SRB, …
    - Clients: PhEDEx, FTS, glite-url-copy, lcg-cp, srm-copy, srmcp, …
    - In OSG, WLCG/EGEE, ESG, …

Documents and Support

- OSG Storage documentation
  - https://twiki.grid.iu.edu/twiki/bin/view/Documentation/WebHome
- BeStMan
  - http://sdm.lbl.gov/bestman
  - https://twiki.grid.iu.edu/bin/view/Documentation/BestmanGateway
  - https://twiki.grid.iu.edu/bin/view/Documentation/BestmanGateway-Xrootd
- Xrootd and XrootdFS
  - http://xrootd.slac.stanford.edu
  - http://wt2.slac.stanford.edu/xrootdfs/xrootdfs.html
- SRM Collaboration and SRM Specifications
  - http://sdm.lbl.gov/srm-wg
- Contact and support
  - SRMSUPPORT@LBL.GOV