Fast Data Transfer

Preliminary Bakeoff Results

Rasan Rasch, NYU

Goals & Background

- NDIIPP – Web at Risk
- Explore Data Transfer / Replication Options to move data between grant partners (CDL, NYU, and UNT)
- Measure Performance, Usability of various tools.

Transfer tools

- SRB
  http://www.sdsc.edu/srb/index.php/Main_Page
- GridFTP
  http://www.globus.org/grid_software/data/gridftp.php
- MogileFS
  http://www.danga.com/mogilefs/
- High Performance SSH
  http://www.psc.edu/networking/projects/hpn-ssh/
- BBFTP
  http://doc.in2p3.fr/bbftp/
- BBCP
  http://www.slac.stanford.edu/~abh/bbcp/
- RSYNC, SCP, SFTP, FTP

Machines

Transfers conducted between NYU and CDL

	NYU	CDL
OS	Fedora Core 5	Suse 10
Kernel	2.6.20	2.6.13
Memory	488 MB	1 G
Disk	33 G	420 G
Max Rcv Window	129024	131071
Max Snd Window	129024	131071
Max mem TCP rcv buffers	131072	174760
Max mem TCP snd buffers	131072	131072

Storage Resource Broker (SRB)

Pros

- Distributed Architecture
- Supports Unix, Windows file systems, HPSS, DBMS, Tape
- Easy to install
- Parallel Streaming
- Fast file transfer
- Jargon (Java API)

Cons

- Not Lightweight
- Configuration Cumbersome

MogileFS

Pros

- No special kernel modules required
- No single point of failure
- File Replication
- File Classes
- Cheap alternative to RAID SAN
- Local Filesystem Agnostic
- Written in Perl

Cons

- Minimal Documentation
- Not that fast transferring Large files
- Written in Perl
- Tricky Install

GridFTP

Pros

- Fast file transfer
- API for adding file transfer capabilities
- Striping (multiple endpoints at src and dest of transfer)
- Tight Security
- Supports Large Files

Cons

- Heavyweight (Globus Toolkit)
- Very difficult install and configure

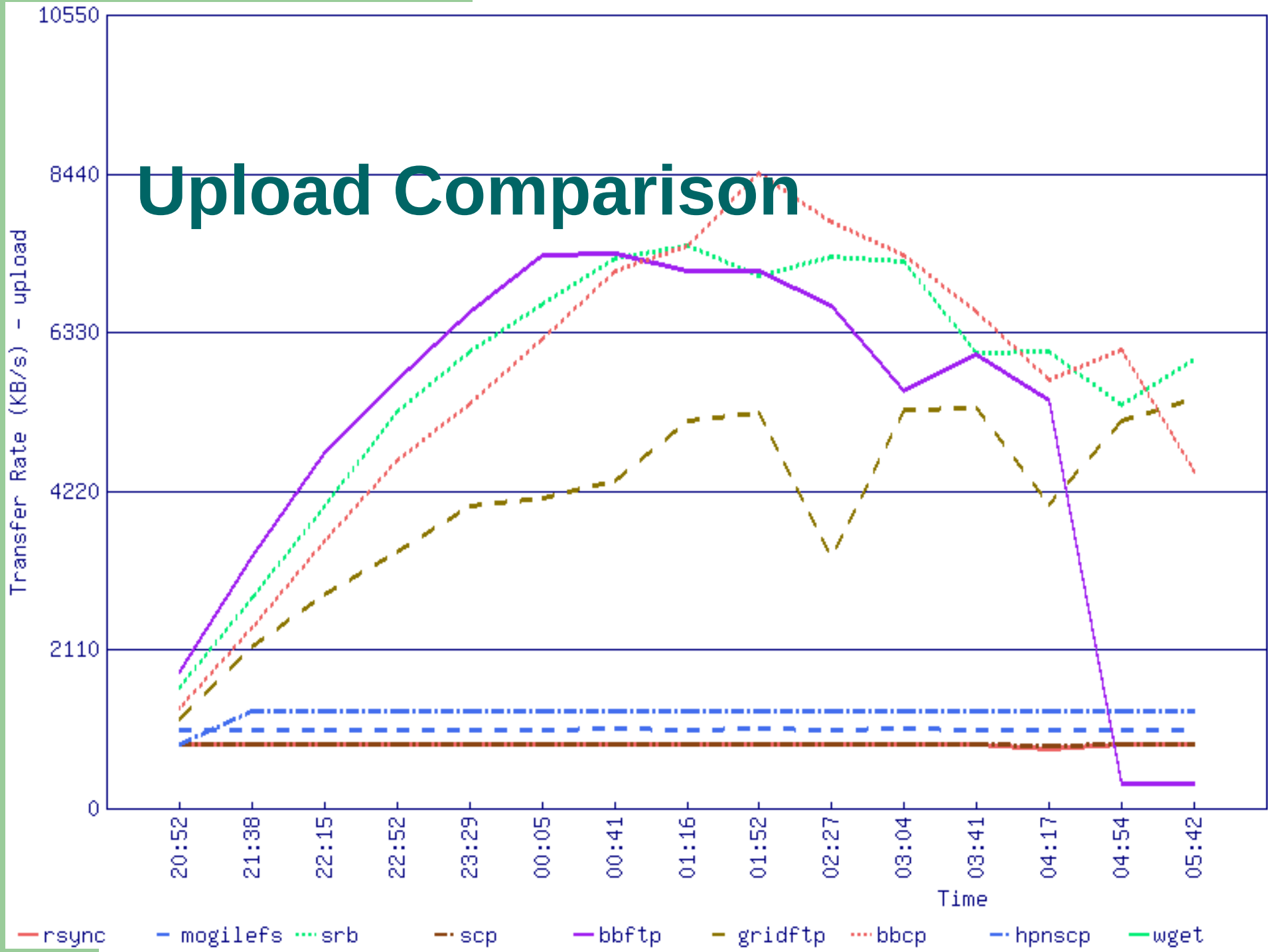
BBFTP & BBCP

- Very Fast
- Simple and Lightweight
- Usage similar to most ftp cmd line clients
- Parallel File Transfers
- Easy Install – one binary on each host

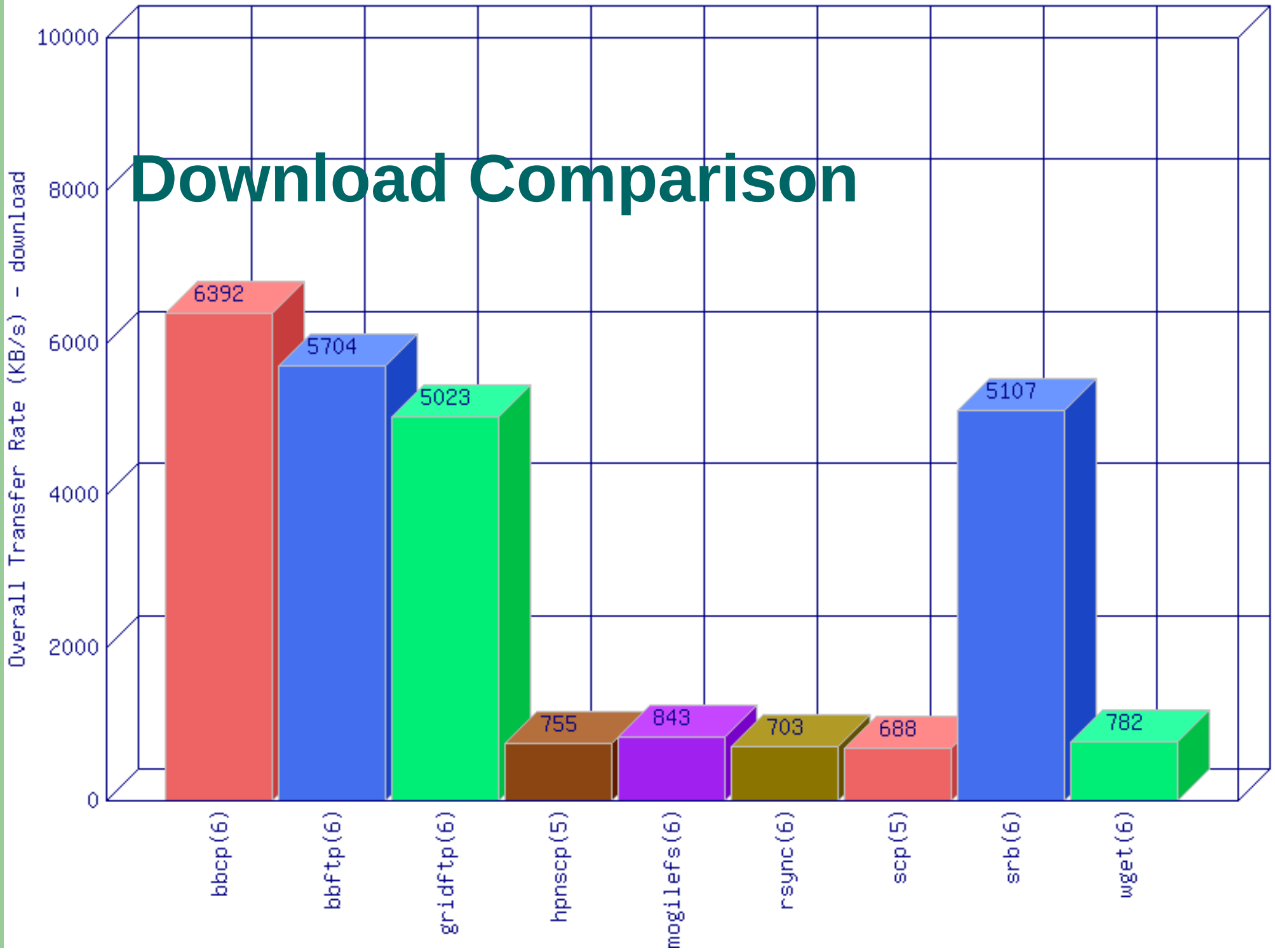
High Performance SSH

- Works by increasing buffer sizes in OpenSSH
- Easily patches

Upload Comparison



Download Comparison



Conclusion

- Tools which offer parallel transfer fastest.
- BBCP and BBFTP easiest to use.
- SRB has most features / flexibility but there is a configuration learning curve.

Contact

Contact:
Rasan Rasch
Developer - NYU Digital Library
rasan@nyu.edu

Testing scripts available at:
https://raschr01.acs.its.nyu.edu/websvn/