BNL Service Challenge 3 Status Report

Xin Zhao, Zhenping Liu, Wensheng Deng,
Razvan Popescu, Dantong Yu and Bruce Gibbard

USATLAS Computing Facility
Brookhaven National Lab

BROOKHAVEN
NATIONAL LABORATORY

Services at BNL

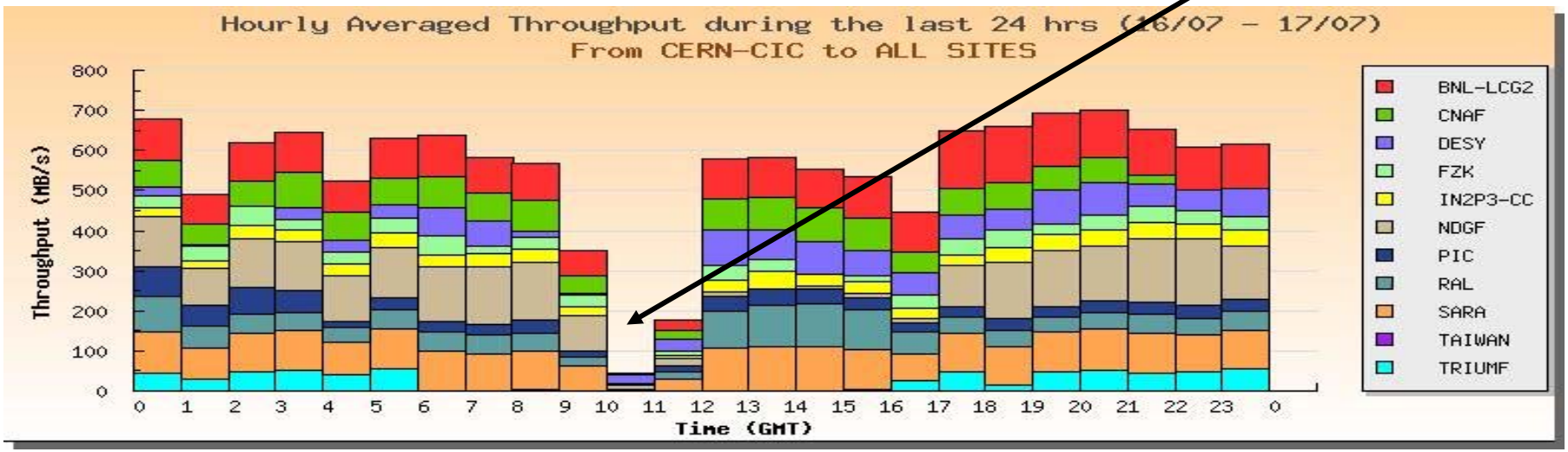
- FTS (version 2.3.1) client + server and its backend Oracle and myproxy servers.
  - Reliable in controlling CERN to BNL data transfer, after several rounds of bug fixes
    - Short timeout value caused excessive failures
    - Incompatibility with dCache/SRM.
  - Still does not support DIRECT data transfer between CERN and BNL dCache data pool servers (dCache SRM third party data transfer)
    - Data transfers go through a few dCache GridFTP door nodes, a scalability issue
    - Had to move these door nodes to non-blocking network ports to distribute traffic
  - Both BNL and RAL discovered that the number of streams per file could not be more than 10, (Is this intended ?)
- Network to CERN:
  - Network for dCache was upgraded to 2*1Gb/s in June.
  - Shared link to CERN with Round Trip Time: >140 ms
    - RTT for European sites to CERN: ~ 20ms.
  - Occasional packet losses observed along BNL-CERN path
  - 1.5 Gb/s aggregated bandwidth observed by iperf with 160 TCP streams.

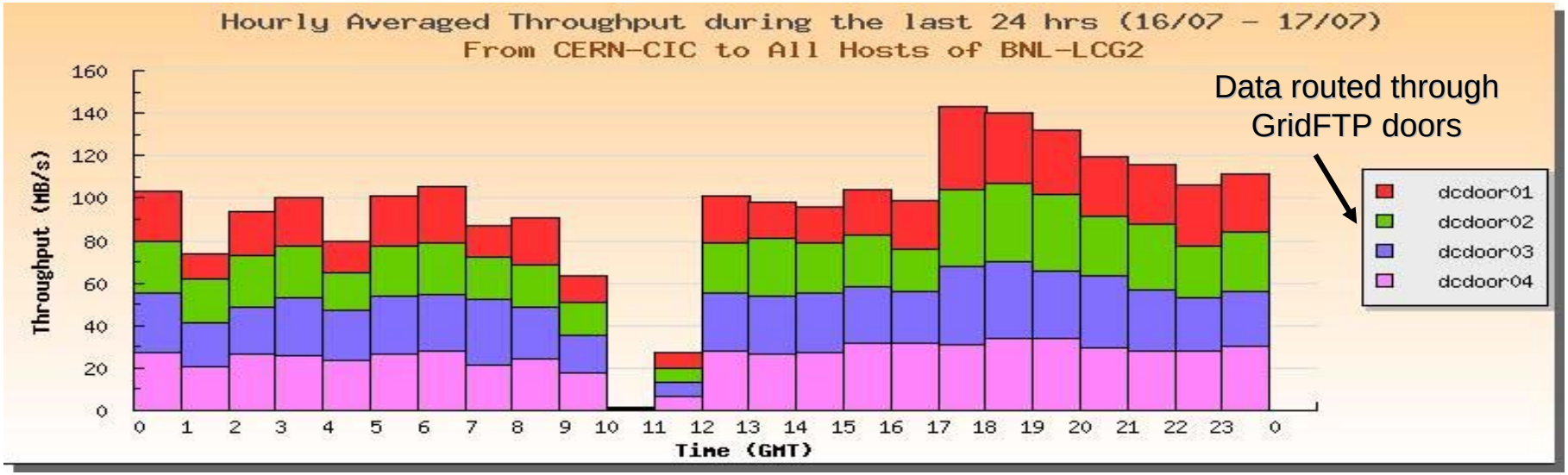
Services at BNL

- dCache/SRM (V1.6.5-2, with SRM 1.1 interface, Total 332 nodes with about 170 TB disks, Multiple GridFTP, SRM, and dCap doors): USATLAS production dCache system.
  - Experienced high load on write pool servers during high rate data transfers
    - Fixed by replacing the EXT file system with XFS file system.
  - Core server crashed once. Reason was identified and fixed
  - Small buffer space (1.0TB) for data written into dCache system
- LFC (1.3.4) client and server was installed at BNL
  - Server was installed and basic functionalities tested: lfc-ls, lfc-mkdir etc.
  - Will populate LFC with the entries from our production Globus RLS server
- DPM is being deployed internally at BNL and under evaluation.
- An agent box was setup for ATLAS experiment software

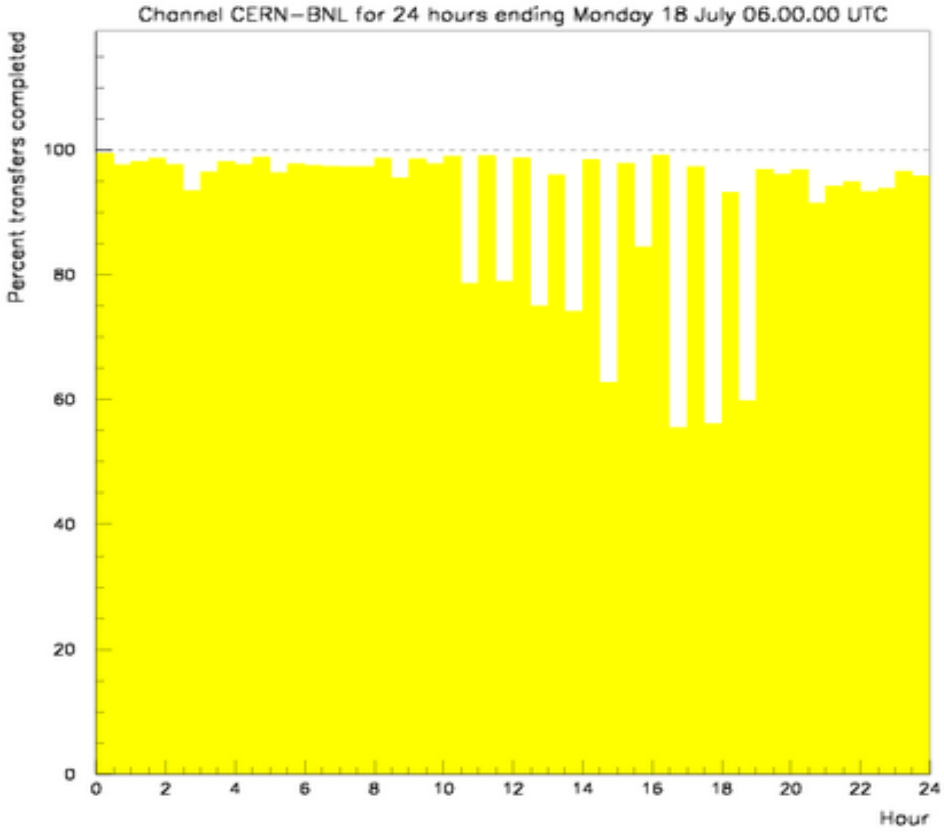
Transfer Plots

Castor2 LSF plug-in problem



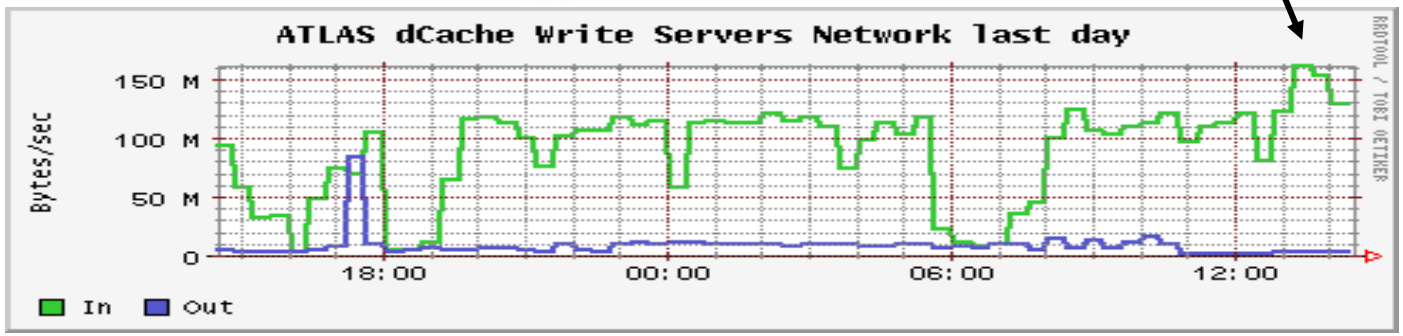


Data Transfer Status



❄After stabilization of data transfer components, reasonable rates of successful transfer completions are being achieved, as shown on the left.

❄We have attained 150 MB/sec rate for about one hour with large numbers (> 50) of parallel file transfers. CERN FTS limit of 50 files per channel is not really enough to fill CERN⇔ BNL data channel



BROOKHAVEN NATIONAL LABORATORY

Some Issues

- Service Challenge Effort
  - Activities have included tuning network pipes, optimizing the configuration and performance of BNL production dCache system and its associate OS & file systems, installation and integration of other SC3 components
  - Major effort was required to stabilize newly deployed FTS, dCache and network infrastructure.
  - Effort level does seem to be decreasing as services became stable.
- Resources are shared by experiments and users.
  - CERN SC3 infrastructure is shared by multiple Tier 1 sites
    - Due to variations between Tier 1's, data transfer needs to be individually optimized based on a Tier 1 site's characteristics, network RRT, packet loss rates, other experiment specific requirements, etc.
  - At BNL, network and dCache are also used by production users.
    - Need to closely monitor the SRM and network to avoid impacting production activities.
- At CERN, James Casey seems almost single handedly to be answering email, setting up systems, reporting problems and running data transfers
  - Does he need more help?

Plans for Remainder of Throughput Phase

- Tier 2 SC3 Participation proceeding slowly
  - Continue to evaluate DPM at BNL for US ATLAS Tier 2 centers
  - Limited work on going at selected Tier 2 sites
  - Still hope to transfer data from/to two Tier 2 sites with BNL FTS instance: Boston University (GridFTP server), University of Chicago (GridFTP/DPM?)
- SC3 Tape Transfer
  - Small scale transfer test between Tier 0 ~ BNL dCache/HPSS with BNL FTS. (July 18~July 22)
  - Plan to borrow tape resources for actual data transfer exercise