Review of US LHCnet and End-to-End Meetings at FNAL October 23-25

Shawn McKee

US ATLAS Tier2 Meeting

UTA

December 8th 2006

Invitation to US LHC Network Meeting

- You are invited to send a representative to the upcoming meeting of the US LHC Network Working Group at Fermilab October 23-24. October 25 is reserved for the network R&D meeting on end-to-end circuits being organized by Dantong Yu et al. Please also forward the notice to interested parties, as appropriate.

  This meeting is very much in the spirit of the successful meeting co-sponsored by Internet2 and ESnet at Fermilab in October 2005 (see http://networks.internet2.edu/tier2/20051012-i2esnetinfo.html the website includes links to the presentations.

  This year we will focus on recent progress, current developments, and especially milestones and other action items for the next 12 months.

  Fermilab will be hosting the meeting, and will shortly set up a web page and registration.

End-to-End Meeting Description

- US-LHC End-to-End Networking

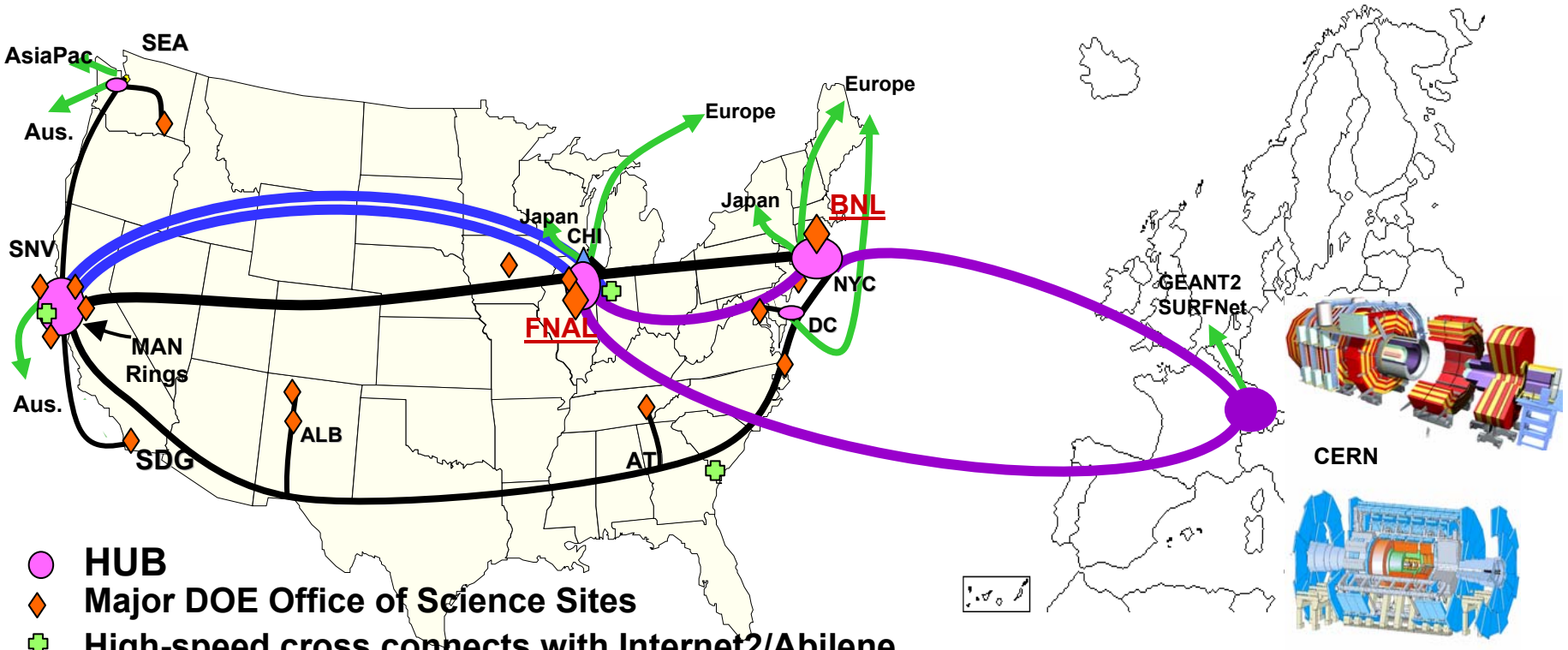
  Meeting **Wednesday 25 October 2006**

  from **08:00** to **19:05**

  at **Fermilab (Wilson Hall 8X "Hornet's Nest")**

- **Description:** Several network research projects funded by DOE MICS and NSF will provide and monitor dynamic lightpaths in various ways. The goal of the meeting will be to examine the networking needs implied by the US LHC experiments' computing models and examine whether and how these projects meet those needs. Participation by USCMS and USATLAS will sharpen the focus.

US LHCNet (From S. Ravot)



- Connections to ESnet Hubs in New-York and Chicago
- Redundant "light-paths" to BNL and FNAL
- Access to USNet for R&D

LHCnet Challenges: Network Infrastructure

"Providing the right cost-effective transatlantic network infrastructure for the HEP community's needs"

- Increased capacity
  - Accommodate the large and increasing amount of data that must reach the HEP DoE Labs.
  - Also ensure data can reach the Tier2s where much, or most of the data analysis will be performed
- High network reliability
  - Essential for distributed applications
  - Availability requirement: 99.95%
- LHC OPN
  - "Virtual Private Circuits" to FNAL and BNL

Physics and the Network

- The methods physicists use to do physics have evolved to become highly dependent upon computers and networks
- All computers have CPUs, memory, long-term storage and a **bus** to interconnect those components
- The trend in computing is to aggregate individual computers into much more powerful virtual computers
- The new virtual computer "bus" is the *network*…

Motivations: Wide-Area Physics

- Physicists require the network for a number of tasks:
  - Moving or accessing data (raw or analyzed) (gridftp, nfs, afs)
  - Remote collaboration and videoconferencing
  - Control of remote instruments
  - Searching for, or accessing remote information (http)
- The developments in grid computing are placing new requirements on the network

- The goal is to have the WAN enable WAP (Wide-Area Physics) - transparent high-performance access to remote resources required by physicists

US LHCnet/End-to-End URLs

- US LHCnet meeting in Indico at https://indico.fnal.gov/conferenceDisplay.py?confId=417
- End-to-End meeting in Indico at https://indico.fnal.gov/conferenceDisplay.py?confId=420
- Les Cottrell took notes: http://www-iepm.slac.stanford.edu/about/status/trip-lhc-fnal-oct06.htm
- Numerous talks concerning US LHC networking, research projects, USLHCnet itself and the LHC computing models

Summary of US LHCnet Meeting

- Many important presentations (see links in last slide)
- Primary discussion around USLHCnet and the LHC computing models:
  - What services should US LHCnet be providing?
  - Which traffic should flow on US LHCnet (non-US Tier2 to US Tier1, Tier2 to Tier2, LHC OPN, general HEP transatlantic traffic…)?
  - What should the AUP (Acceptable Use Policy) for US LHCnet be?
  - How should we monitor, manage and integrate US LHCnet (and high-performance networking in general) into the LHC computing models?
- Differences in network assumptions and planning in ATLAS and CMS arose: CMS supports any Tier2 connecting to any Tier1, ATLAS has restricted connections via hierarchy. Issues around efficiency, capability, overloading resources, etc.

Further Details: Grid Computing and End-to-End work

- A fundamental concept of grid computing is to transparently use local and remote resources to minimize time-to-completion for a "job"
  - Intrinsic in this concept is a guarantee of network performance
- Jobs are scheduled based upon the location of CPU/Storage/Data to maximize efficient use of these resources
- An inability to access data at a guaranteed rate causes numerous problems (I/O starvation, incorrectly chosen data copies, inefficient resource allocation, etc.)
- It is critical to match the network capabilities with the end-site capabilities: makes no sense to reserve 500Mbits/sec if end-site can only source/sink 100 Mbits/sec…

Big Pipes are *not* Enough!

- There is a continuing debate concerning over-provisioning versus network management techniques to meet application needs.
- For networked applications that require certain network characteristics, big pipes are no guarantee of meeting those needs.
- Another important point: big pipes do not currently extend to the desktop and most difficulties arise near the endpoints rather than on the backbones…
- Some mechanism is required to insure the expected behavior *end-to-end* in the network or certain applications will not function correctly or effectively.

Problems with End-to-End Network Management

- There are a number of perceived problems with End-to-end Network Management:
  - The performance impact on the network is too large
  - It requires new hardware to implement
  - There is a significant retraining of network engineers required
  - It won't scale to WAN's (policy/admin/etc) and multiple security domains
- To what extent are the perceptions correct, especially considering the more modest scope that HEP/LHC represents?

- We are all working to implement something in the real network to find out…

Tier-1 Issues

- Tier1 sites have a number of responsibilities:
  - They are the primary custodian of data for the experiments
  - They must reprocess the raw data for which they are custodians
  - They must serve the data they keep to the collaboration
  - They provide some level of analysis capability for the collaboration
- Most Tier1 sites are embedded in existing shared facilities. For example BNL has RHIC. Network access must be shared and managed.
- Tier-0 to Tier-1 data transfer is very high-priority and must be protected.
- The ability of the site to prioritize and allocate the network can be critical in unexpected situations: major software error requiring recreating of ESDs and AODs, hardware failure recovery, interesting physics signal discovery, etc.

Tier-2 Issues

- Tier2 sites have two primary responsibilities:
  - They are the primary simulation site for the collaboration
  - They must serve a significant number of analysis users and groups
- Tier2 sites typically have limited manpower and certainly cannot devote any significant fraction of an FTE to managing the network.
- Insuring that required network connections are available to meet the highest priority tasks is a very important consideration.
- As for Tier-1s, the ability of the Tier-2 to prioritize and allocate the network are also critical in unexpected situations: major software error requiring recreating of ESDs and AODs, hardware failure recovery, interesting physics signal discovery, etc.

Major Issues for E2E

- Equipment: Do Tier-1 and Tier-2 sites have the required capabilities in their existing or planned network equipment?
- Management: Is the end-to-end system manpower intensive (or requires a high level of local expertise)?
- Scalability: Is the end-to-end system scalable (feasible) for LHC HEP level deployment?
- Effectiveness: Is the end-to-end system useful for the mission of the Tier-1 or Tier-2?

Action Item From End-to-End Meeting

- A primary outcome of the End-to-End meeting: a plan to deploy/demonstrate some of the US funded e2e research projects in the context of ATLAS's and CMS's Tier-1/Tier-2s
- Relevant projects/technologies (alphabetically):
  - Lambda Station (FNAL/Caltech, Traffic Routing via Optical Paths)
  - OSCARs (ESnet/MPLS)
  - Terapaths (BNL/SLAC/UMich, QoS LAN Mgmt)
  - UltraLight (VINCI/LISA/ApMon/MonALISA)
- We need to identify a set of Tier2's in CMS and ATLAS to participate:
  - CMS: *Caltech, Nebraska, Wisconsin?*
  - ATLAS: *MWT2, SLAC?, SWT2?*
- The milestones will include
  - Documenting the network infrastructure at and between sites
  - Selecting, installing and testing the end-to-end path technologies (Two or more per site?)
  - Running data movement/access tests in the presence of competing traffic with/without the end-to-end technologies
- Timescale is to complete this by **April 2007**
- We need a responsible "pusher" for each project to test, and we need to arrange some time windows with the experiments as early as possible to use their facilities.

Summary

- Physics computing models are strongly dependent upon the network
- The Tier-1 and Tier-2 centers have specific obligations that any end-to-end system must help them to deliver.
- We need to begin deployment, testing and integration now to insure that the bandwidth and network services required will be in place on timescales commensurate with our needs
- The End-to-End projects discussed at this meeting are a very important step toward delivering such a capability for LHC and HEP.
- Lets see if we can test these within USATLAS to see how (if?) this research can help us.