CDF NETWORKING NEEDS

Krzysztof Sliwa / Tufts University

CDF Collaboration Meeting, January 24-25, 2002

(presented by Ben Whitehouse on January 25, 2002)

For details see: "An Estimate of CDF Networking Needs"
CDF/PUB/CDF/PUBLIC/5836 (in preparation)

CDF NETWORKING NEEDS

- A questionnaire (16 questions) has been distributed in mid-October 2001 among Computer Representatives of all groups.
- 20 out of 53 groups responded (you can still send the information to ksliwa@tufts.edu, this will only help to get a more complete picture)
- Most of US university groups (23/25) connected to Internet2
- Most groups have been thinking of using network transfers to get the data and avoid tapes if possible
- some (~10%) of respondents) think that tape may be used as well
- one group (Rutgers) plans on having a system integrating tapes and disks a la Fermilab
- Most groups are prepared to have ~1 TB disk space in the next 6 months

CDF NETWORKING NEEDS

- CDF Run-II is similar (in terms of dataset size) to one year of ATLAS running at LHC. Detailed simulations were performed to model a data handling system that assumes a hierarchical set of Tier-1, Tier-2 and local centers allowing data to be transferred from/to the central site after reconstruction (2 times a year), re-definition of secondary datasets (once a month) and also allowing a large number of analysis jobs utilizing a *bona fide* oo database.

- CDF needs correspond to (in this model) to the first two activities.

- Estimates can be made with simple back-of-the-envelope calculations, or with the help of a modified model built with MONARC tools

CDF NETWORKING NEEDS

- THE BOTTOM LINE:

  For the next 12 months CDF needs would be satisfied with ~0.8 Gbps (gigabits per second) aggregated bandwidth to distribute secondary datasets and selected parts of primary datasets to the collaborating groups. This assumes NO COORDINATION in disk use between the CDF Groups. One could imagine groups allowing their data disks to be afs-mounted (or similar) among the CDF collaborators. This is commonly done in CERN-based collaborations.

  For the next 12 months an OC-48 (2.5 Gbps) link from FNAL to the outside world should be sufficient for CDF, D0 and "standard"needs. The link should be OC-192 capable.

  Beyond that time, FNAL should acticipate a need for upgrading the networking equipment to OC-192

BACK OF THE ENVELOPE ESTIMATES

i)

Extrapolated size of pads for all datasets (CDF4568)

~50-100 TB (terabytes)

4 years of data taking => 12.5-25 TB/y

1 "y" ~ 100 d

1 "d" ~ 100 000 s

=> ~ 125-250 GB/d

=> ~ 1.25-2.5 MB/s

=> ~ 10-20 Mbps

each of 50 sites takes ~ 1/4 of all datasets; at least twice (re-reconstruction)

=> 4-10 Mbps per site

=> 200-500 Mbps

BACK OF THE ENVELOPE ESTIMATES

ii)

Bandwidth needed to transfer 0.5 TB=500 GB in a week to 50 sites:

"week" ~5 d

"day" ~100 000 s

=> 1 MB/s or 8 Mbps per site

=> 400 Mbps

The rule of thumb is that one rarely gets more than 30% of the nominal bandwidth

=> 1.2 Gbps

Two experiments + extra daily activities

=> 2.4 Gbps

=> ***OC-48 (2.5 Gbps) for the next 12 months, or so***

=> ***OC-192 (10 Gbps) in 2003 and beyond***

Recommendation (abstract)

- I have tried to estimate the bandwidth that CDF users would like to have available to/from FNAL in order to access and analyze the data from Run-II in an efficient and timely manner. I have distributed a questionnaire based on which I have a good idea of how much data various CFD groups would like to have at their remote sites.

- Most of CDF institutions are connected to Internet2 already. Most of 11 large cities in US are connected via Abilene network with current speeds of 2.5 Gbps (gigabit-per-second) and their connections will be upgraded to N x OC-192, or N x 10 Gbps (N is an integer larger than one indicating that one may achieve higher bandwidth on a single fiber by multiplexing) in the next 18 months. Connection to CERN is planned to reach the bandwidth of 622 Mbps by April 2002 and 2.5 Gbps by Summer 2002. Compared to those numbers, the ESNET connection to Fermilab, currently at 155 Mbps, is really quite poor.

Recommendation (abstract)

- The most important bottleneck within the current CDF Data Handling system is at the mass storage/disk cache interaction point. I believe that problems would be reduced significantly if the strategy of automated distributing the data from FNAL to the collaborating institutions is adopted. The system would be similar to the one already in existence for software distribution and software updates. ***In the proposed scenario, the requested data could be distributed over the networks to the remote sites BEFORE it is archived onto tapes. This could work for all secondary datasets, and selected subsets of other datasets.***

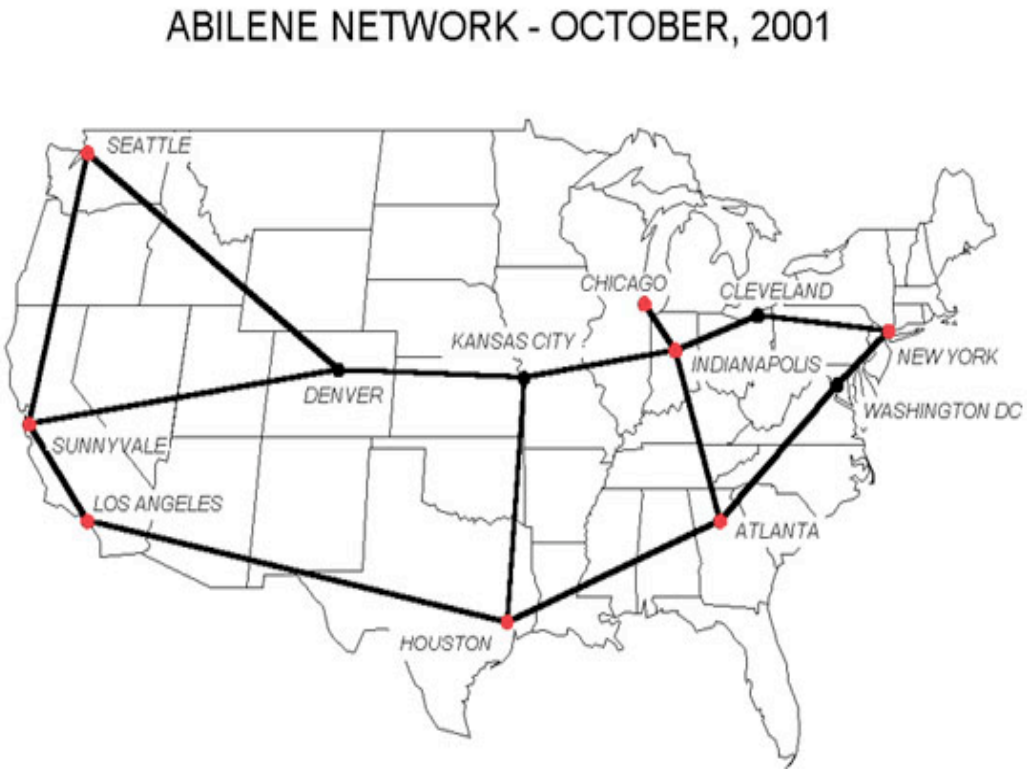
- The bandwidth necessary to support such a system, taking into account that D0 also will try to do similar things, and the fact that there will be other activities on the networks has been estimated to be quite comfortably within OS-48 (2.5 Gbps). This assumes unicast data transmissions, which is the most pessimistic scenario. (Work is under way towards developing multicast protocols for data transfers.)

Recommendation (abstract)

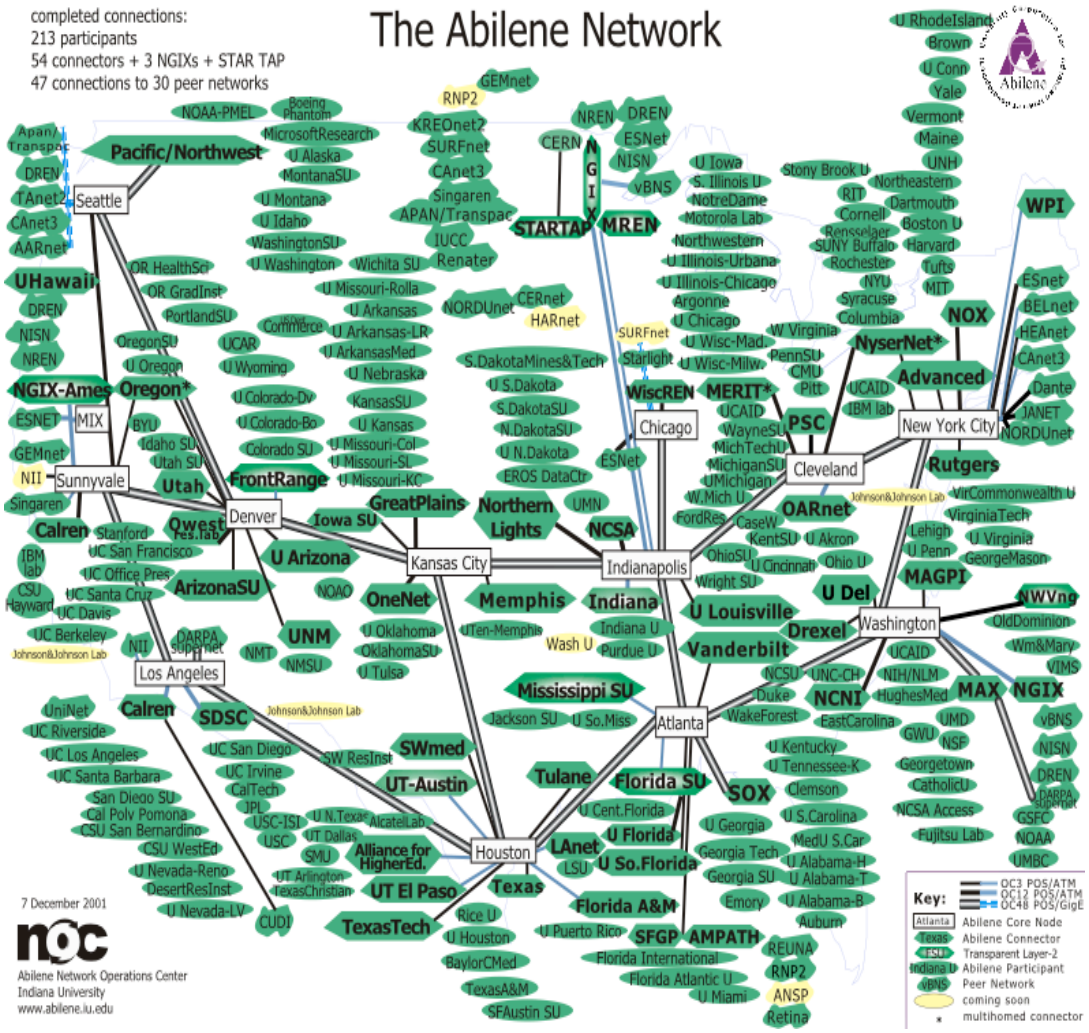
- I recommend that Fermilab obtains a dark fiber connection capable of running 10 Gbps, and equip it with hardware allowing OC-48 connections (2.5 Gbps). Given that, at present, the ESNET has inferior bandwidth in their connections, a Fermilab-Internet2
- (Abilene) connection should be established. It seems that on the time-scale of 2 years, Fermilab may need to upgrade their connection to OC-192, which by then will be a standard and common thing.

- I have run simulations using MONARC tools of 30 groups served by the proposed automatic Fermilab-based data distribution system. All results will be summarized in a forthcoming note, my conclusion is supported by the simulation results.

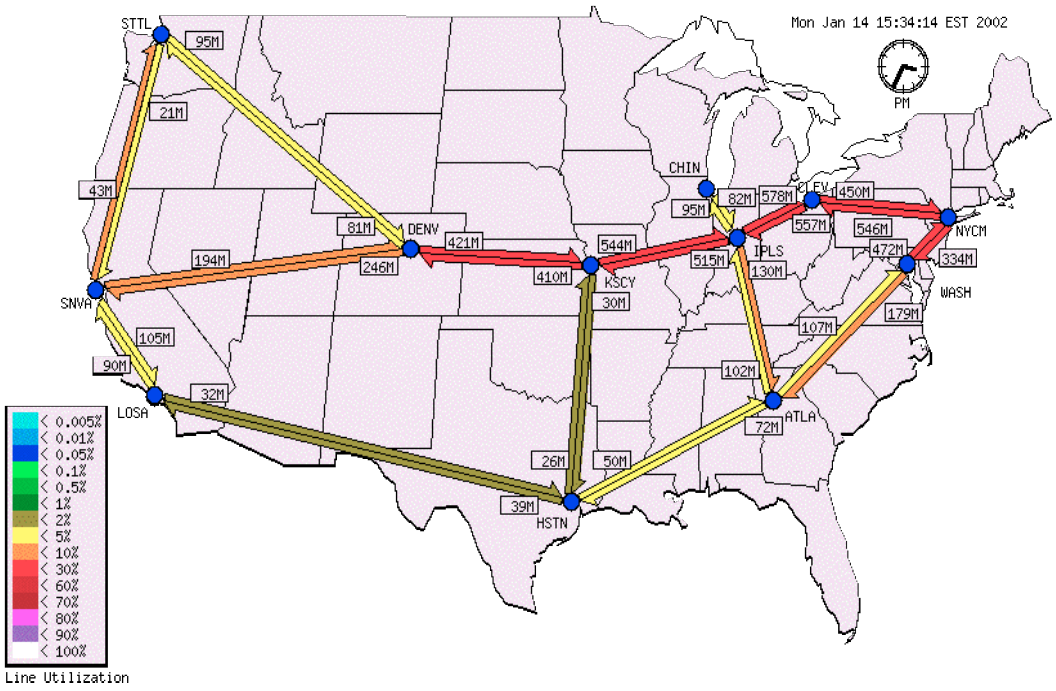
CDF NETWORKING NEEDS



CDF NETWORKING NEEDS



CDF NETWORKING NEEDS



CDF NETWORKING NEEDS