Networks and Grids for HENP as Global e-Science

Harvey B. Newman
California Institute of Technology
SC2004, Pittsburgh
November 11, 2004
LHC Data Grid Hierarchy: Developed at Caltech

Tier 1
- 10 - 40 Gbps
- IN2P3 Center
- RAL Center
- INFN Center
- FNAL Center

Tier 0 +1
- CERN Center
- PBs of Disk; Tape Robot
- ~100-1500 MBytes/sec

Tier 2
- Tier2 Center
- ~1-10 Gbps
- Tens of Petabytes by 2007-8. An Exabyte ~5-7 Years later.

Tier 3
- ~1-10 Gbps
- Physics data cache
- Workstations

Tier 4
- 1 to 10 Gbps
- Institute

CERN/Outside Resource Ratio ~1:2
Tier0/(Σ Tier1)/(Σ Tier2) ~1:1:1

Emerging Vision: A Richly Structured, Global Dynamic System
Challenges of Next Generation Science in the Information Age

Petabytes of complex data explored and analyzed by 1000s of globally dispersed scientists, in hundreds of teams

Flagship Applications

- High Energy & Nuclear Physics, AstroPhysics Sky Surveys:
  TByte to PByte “block” transfers at 1-10+ Gbps
- Fusion Energy: Time Critical Burst-Data Distribution;
  Distributed Plasma Simulations, Visualization, Analysis
- eVLBI: Many real time data streams at 1-10 Gbps
- BioInformatics, Clinical Imaging: GByte images on demand

Advanced integrated Grid applications rely on reliable, high performance operation of our LANs and WANs

Analysis Challenge: Provide results with rapid turnaround, over networks of varying capability in different world regions
## HENP Bandwidth Roadmap for Major Links (in Gbps)

<table>
<thead>
<tr>
<th>Year</th>
<th>Production</th>
<th>Experimental</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>0.155</td>
<td>0.622-2.5</td>
<td>SONET/SDH</td>
</tr>
<tr>
<td>2002</td>
<td>0.622</td>
<td>2.5</td>
<td>SONET/SDH DWDM; GigE Integ.</td>
</tr>
<tr>
<td>2003</td>
<td>2.5</td>
<td>10</td>
<td>DWDM; 1 + 10 GigE Integration</td>
</tr>
<tr>
<td>2005</td>
<td>10</td>
<td>2-4 X 10</td>
<td>λ Switch; λ Provisioning</td>
</tr>
<tr>
<td>2007</td>
<td>2-4 X 10</td>
<td>~10 X 10; 40 Gbps</td>
<td>1st Gen. λ Grids</td>
</tr>
<tr>
<td>2009</td>
<td>~10 X 10 or 1-2 X 40</td>
<td>~5 X 40 or ~20-50 X 10</td>
<td>40 Gbps λ Switching</td>
</tr>
<tr>
<td>2011</td>
<td>~5 X 40 or ~20 X 10</td>
<td>~25 X 40 or ~100 X 10</td>
<td>2nd Gen λ Grids Terabit Networks</td>
</tr>
<tr>
<td>2013</td>
<td>~Terabit</td>
<td>~MultiTbps</td>
<td>~Fill One Fiber</td>
</tr>
</tbody>
</table>

Continuing Trend: ~1000 Times Bandwidth Growth Per Decade; Compatible with Other Major Plans (ESNet, NLR; GN2, GLIF)
Problem: Extract “Small” Data Subsets of 1 to 100 Terabytes from 1 to 1000 Petabyte Data Stores

Survivability of the HENP Global Grid System, with hundreds of such transactions per day (circa 2007) requires that each transaction be completed in a relatively short time.

Example: Take 800 secs to complete the transaction. Then

<table>
<thead>
<tr>
<th>Transaction Size (TB)</th>
<th>Net Throughput (Gbps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
</tr>
<tr>
<td>100</td>
<td>1000 (Capacity of Fiber Today)</td>
</tr>
</tbody>
</table>

Summary: Providing Switching of 10 Gbps wavelengths within ~2-4 years; and Terabit Switching within 5-8 years would enable “Petascale Grids with Terabyte transactions”, to fully realize the discovery potential of major HENP programs, as well as other data-intensive research.
Judged on product of transfer speed and distance end-to-end, using standard (TCP/IP) protocols, Across Production Net: e.g. Abilene


IPv4 with Windows & Linux: 6.6 Gbps Caltech-CERN (15.7 kkm; “Grand Tour of Abilene”) June 2004

Exceeded 100 Petabit-m/sec

7.48 Gbps X 16 kkm (Linux, 1 Stream) Achieved in July

11 Gbps (802.3ad) Over LAN in Sept.

Concentrate now on reliable Terabyte-scale file transfers

Note System Issues: CPU, PCI-X Bus, NIC, I/O Controllers, Drivers

SC04: 6.9 Gbps X 26 kkm 11/08

SC04 BW Challenge: 101.1 Gbps
<table>
<thead>
<tr>
<th>Science Areas</th>
<th>Today <em>End2End Throughput</em></th>
<th>5 years <em>End2End Throughput</em></th>
<th>5-10 Years <em>End2End Throughput</em></th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Energy Physics</td>
<td>0.5 Gb/s</td>
<td>100 Gb/s</td>
<td>1000 Gb/s</td>
<td>High bulk throughput</td>
</tr>
<tr>
<td>Climate (Data &amp; Computation)</td>
<td>0.5 Gb/s</td>
<td>160-200 Gb/s</td>
<td>N x 1000 Gb/s</td>
<td>High bulk throughput</td>
</tr>
<tr>
<td>SNS NanoScience</td>
<td>Not yet started</td>
<td>1 Gb/s</td>
<td>1000 Gb/s + QoS for Control Channel</td>
<td>Remote control and time critical throughput</td>
</tr>
<tr>
<td>Fusion Energy</td>
<td>0.066 Gb/s (500 MB/s burst)</td>
<td>0.198 Gb/s (500MB/20 sec. burst)</td>
<td>N x 1000 Gb/s</td>
<td>Time critical throughput</td>
</tr>
<tr>
<td>Astrophysics</td>
<td>0.013 Gb/s (1 TByte/week)</td>
<td>N*N multicast</td>
<td>1000 Gb/s</td>
<td>Computations steering and collaboration</td>
</tr>
<tr>
<td>Genomics Data &amp; Computation</td>
<td>0.091 Gb/s (1 TBy/day)</td>
<td>100s of users</td>
<td>1000 Gb/s + QoS for Control Channel</td>
<td>High throughput and steering and collaboration</td>
</tr>
</tbody>
</table>

See [http://www.doecollaboratory.org/meetings/hpnpw/](http://www.doecollaboratory.org/meetings/hpnpw/)
Transition beginning now to optical, multi-wavelength Community owned or leased “dark fiber” (10 GbE) networks for R&E

National Lambda Rail (NLR): www.nlr.net

- Coming Up Now
- Initially 4 10G Wavelengths
- Northern Route LA-JAX by 4Q04
- Internet2 HOPI Initiative (w/HEP)
- To 40 10G Waves in Future

Initiatives in: nl, ca, pl, cz, uk, ko, jp
- + 18 US States (CA, IL, FL, IN, ...)

NLR
21 NLR Waves: 9 to SC04
High Speed TeraByte Transfers for Physics

- Joint Caltech, FNAL, CERN, SLAC, UF, SDSC, BR, KR, ....
- 10 10 Gbps waves
- HEP on show floor
- Bandwidth challenge: aggregate throughput of 101.13 Gbps achieved
- FAST TCP
High-speed cross connects with Internet2/Abilene
Major DOE Office of Science Sites

ESnet Beyond FY07 (W. Johnston)

MANs
Qwest – ESnet hubs
NLR – ESnet hubs
High-speed cross connects with Internet2/Abilene
Major DOE Office of Science Sites

MANs
Qwest – ESnet hubs
NLR – ESnet hubs
High-speed cross connects with Internet2/Abilene
Major DOE Office of Science Sites

Production IP ESnet core
High-impact science core
Lab supplied
Major international

2.5 Gbs
10 Gbs
Future phases

10G
30Bc
40G
Integrated hybrid experimental network, leveraging Transatlantic R&D network partnerships; packet-switched + dynamic optical paths

- 10 GbE across US and the Atlantic: NLR, DataTAG, TransLight, NetherLight, UKLight, etc.; Extensions to Japan, Taiwan, Korea, Brazil

- End-to-end monitoring; Realtime tracking and optimization; Dynamic bandwidth provisioning

- Agent-based services spanning all layers of the system, from the optical cross-connects to the applications.
Research Networking in Latin America: Just Taking Off in 2004

- **AmPath** Provided connectivity for some Latin American countries
  - Argentina, Brazil, Chile, Mexico, Venezuela
- **New CHEPREO Sao Paolo-Miami Link at 622 Mbps Starting This Month**
- **New: CLARA (Funded by EU)**
- **Regional Network Connecting 19 Countries:**
  - Argentina, Dominican Republic, Panama
  - Brasil, Ecuador, Paraguay
  - Bolivia, El Salvador, Peru
  - Chile, Guatemala, Uruguay
  - Colombia, Honduras, Venezuela
  - Costa Rica, Mexico, Nicaragua
- **155 Mbps Backbone with 10-45 Mbps Spurs; 4 Mbps Satellite to Cuba; 622 Mbps to Europe**

Also WHREN NSF Proposal: 2.5G to US

Brazilian HEPGrid: Rio, Sao Paolo etc.
OC3 circuits Moscow-Chicago-Beijing since January 2004

OC3 circuit Moscow-Beijing July 2004 (completes the ring)

Rapid traffic growth with heaviest US use from DOE (FermiLab), NASA, NOAA, NIH and 260+ Univ. (UMD, IU, UCB, UNC, UMN… Many Others)

Plans for Central Asian extension, with Kyrgyz Gov’t

GLORIAD 5-year Proposal (with US NSF) for expansion to 2.5G-10G
Moscow-Amsterdam-Chicago-Pacific-Hong Kong-Pusan-Beijing early 2005
10G ring around northern hemisphere 2007;
Multi-wavelength hybrid service from ~2008-9

Aug. 8 2004: P.K. Young, Korean IST Advisor to President Announces
◆ Korea Joining GLORIAD as a full partner

> 5TBytes now transferred monthly via GLORIAD to US, Russia, China
International ICFA Workshop on HEP Networking, Grids and Digital Divide Issues for Global e-Science

Dates: May 23-27, 2005
Venue: Daegu, Korea

Dongchul Son
Center for High Energy Physics
Kyungpook National University
ICFA, Beijing, China
Aug. 2004

Approved by ICFA
August 20, 2004
Workshop Goals

- Review the current status, progress and barriers to effective use of major national, continental and transoceanic networks used by HEP.
- Review progress, strengthen opportunities for collaboration, and explore the means to deal with key issues in Grid computing and Grid-enabled data analysis, for high energy physics and other fields of data intensive science, now and in the future.
- Exchange information and ideas, and formulate plans to develop solutions to specific problems related to the Digital Divide in various regions, with a focus on Asia Pacific, as well as Latin America, Russia and Africa.
- Continue to advance a broad program of work on reducing or eliminating the Digital Divide, and ensuring global collaboration, as related to all of the above aspects.
Networks and Grids for HENP and Global Science

◆ Network backbones and major links used by HENP and other fields are advancing rapidly
  ◦ To the 10 G range in < 3 years; much faster than Moore’s Law
  ◦ New HENP and DOE Roadmaps: a factor ~1000 BW Growth per decade
◆ We are learning to use long distance 10 Gbps networks effectively
  ◦ 2004 Developments: to 7 - 7.5 Gbps flows with TCP over 16-25 kkm
◆ Transition to community-operated optical R&E networks (us, ca, nl, pl, cz, sk, kr, jp ...); Emergence of a new generation of “hybrid” optical network

◆ We Must Work to Close to Digital Divide
  ◦ To Allow Scientists in All World Regions to Take Part in Discoveries
  ◦ Removing Regional, Last Mile, Local Bottlenecks and Compromises in Network Quality are now On the Critical Path
◆ Important Examples on the Road to Progress in Closing the Digital Divide
  ◦ CLARA, CHEPREO, and the Brazil HEPGrid in Latin America
  ◦ Optical Networking in Central and Southeast Europe
  ◦ APAN Links in the Asia Pacific: GLORIAD and TEIN
  ◦ Leadership and Outreach: HEP Groups in Europe, US, Japan, & Korea