ESnet Network Upgrade Plans: ESnet6

Joe Metzger
Team Lead
ESnet Network Planning Team
Lawrence Berkeley National Laboratory

HEPIX
Hungarian Academy of Sciences
April 24-28 2017
ESnet

- ESnet is a mission network funded by the U.S. Department of Energy to support DOE Science research.
  - **ESnet's Mission** is
    - To enable and accelerate scientific discovery by delivering unparalleled network infrastructure, capabilities, and tools.
  - And **ESnet's Vision** is summarized by these three points.
    - Scientific progress will be completely unconstrained by the physical location of instruments, people, computational resources, or data.
    - Collaborations at every scale, in every domain, will have the information and tools they need to achieve maximum benefit from scientific facilities, global networks, and emerging network capabilities.
    - ESnet will foster the partnerships and pioneer the technologies necessary to ensure that these transformations occur.

- As part of this mission, ESnet is supporting the LHC community with 100G (or faster) connections to CERN, FERMI, BNL, and the large US LHC Universities.

- Over half ESnet traffic is HEP related.
ESnet's Network across the US and Europe
Historic ESnet Traffic Rates from my.es.net

Traffic Volume

<table>
<thead>
<tr>
<th>Year</th>
<th>Total Traffic</th>
<th>OSCARS</th>
<th>LHCONE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>12.88PB</td>
<td>21.9%</td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>22.37PB</td>
<td>38.1%</td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td>23.49PB</td>
<td>40.0%</td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td>23.49PB</td>
<td>40.0%</td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td>23.49PB</td>
<td>40.0%</td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>23.49PB</td>
<td>40.0%</td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td>23.49PB</td>
<td>40.0%</td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td>23.49PB</td>
<td>40.0%</td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td>23.49PB</td>
<td>40.0%</td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td>23.49PB</td>
<td>40.0%</td>
<td></td>
</tr>
</tbody>
</table>

March 2017

<table>
<thead>
<tr>
<th>Category</th>
<th>Bytes</th>
<th>Percent of Total</th>
<th>One Month Change</th>
<th>One Year Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSCARS</td>
<td>12.88PB</td>
<td>21.9%</td>
<td>+30.4%</td>
<td>+87.5%</td>
</tr>
<tr>
<td>LHCONE</td>
<td>22.37PB</td>
<td>38.1%</td>
<td>+7.54%</td>
<td>+111%</td>
</tr>
<tr>
<td>Normal traffic</td>
<td>23.49PB</td>
<td>40.0%</td>
<td>+15.5%</td>
<td>+18.9%</td>
</tr>
<tr>
<td>Total</td>
<td>56.74PB</td>
<td>40.0%</td>
<td>+15.1%</td>
<td>+57.9%</td>
</tr>
</tbody>
</table>
ESnet6

• We are in year 2 of a ~7 year project to build the next generation of the ESnet network to support the DOE science network needs from 2020 through 2025, and possibly 2030.

• Three key design goals
  – **Capacity** - Handle exponential traffic growth at reasonable cost
  – **Reliability and Resiliency** - distributed science facilities, computing, data – scientists depend on the network for their science research to work
  – **Flexibility** - support changing compute models, near real-time analysis, ‘superfacility’, etc.
ESnet6 Project Management

• The US Department of Energy has rigorous methodology for managing large projects
• This is the first time an ESnet network refresh will be managed via this process
• The major milestones are "Critical Decisions" where DOE program management gives permission to proceed with the next step of the project.

<table>
<thead>
<tr>
<th>CD Phases for ESnet6</th>
<th>Fiscal Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD-0, Approve Mission Need</td>
<td>FY2016</td>
</tr>
<tr>
<td>CD-1/3a, Approve Alternative Selection and Cost Range, Approve Long Lead Procurement</td>
<td>FY2018</td>
</tr>
<tr>
<td>CD-2/3b, Approve Performance Baseline, Approve start of construction</td>
<td>FY2019</td>
</tr>
<tr>
<td>CD-4, Approve Project Completion</td>
<td>FY2022</td>
</tr>
</tbody>
</table>
Some of the Challenges Designing a Network for 2020-2025

• User trends are changing
  – Computing models & data movement patterns
  – Clouds
  – Automated workflow systems & data tools like Globus are penetrating more science communities
  – Exascale

• Industry is moving quickly
  – Lots of changes are being driven by the Data Center Interconnect market
    • DCI hardware, software, and philosophies are disrupting the industry
    • "Web Scale" boxes with high density 100G colored optics are hitting GA
  – SDN eco system is continually evolving

• Timing
  – On the one hand, it is great to be out-in front of an upgrade cycle and planning several years in advance.
  – On the other hand, almost every technology we are looking at will have multiple major revision cycles between now and when we actually build.
Our first forecast of ESnet 2020 link capacities required to meet demand extrapolated from historic traffic levels and assuming historic services, loads & protection mechanisms.
Our first forecast of ESnet 2020 link capacities required to meet demand extrapolated from historic traffic levels and assuming historic services, loads & protection mechanisms.

- **600 Gbps Rings in Western US**
- **2 Tbps Ring in Eastern US**
- **2 Tbps Trans Atlantic**
- **1 Tbps LHC Sites?**

*Provider Edge: N (where N denotes aggregate bandwidth to customers connected to node, includes redundant bandwidth to sites) [Jan 2020]*

**Optical Add/Drop: N (where N denotes add/drop bandwidth on EVERY degree of node, including connection between nodes at 2-node sites) [Jan 2020]*

NB: Some sites without PE routers on local premises are not shown in this diagram.
Forecasting Link Capacity Requirements

- Our first ESnet6 capacity forecast was based on Monthly Average Utilization info.
  - We have long-term historic data about total bytes accepted on links so this was fairly straight-forward

- We have very limited useful quantitative historical data about burstiness of different workflows & use cases.

- I wanted to get a sense of the burstiness and how warm multiple other organizations in the LHC community are loading their links so I looked at the LHCOPN links, and how this workload might work with the Average based analysis.
ESnet Europe to US Traffic October 2016
Utilization sorted by Time
(Stacked graph of the 4 different circuits)
Same data sorted by Utilization instead of Time
- Shaded area is used capacity
- Unshaded area is unused capacity
Same data sorted by Utilization instead of Time
- Shaded area is used capacity
- Unshaded area is unused capacity

For this example, a capacity of 90 Gbps was sufficient to meet the needs 95% of the time.
Peak Utilization Drives Network Operations & Planning

Increased utilization during the 0 to 80 Percentile intervals can be supported in many cases with no incremental costs & does not impact other users!

90+ Percentiles drive link size requirements, and therefore costs.
Many networks capacity planning triggers are based on 95th percentiles.
Observations

• Two LHCOPN paths (CNAF & PIC) were running in January at the load levels we anticipated in the first round of ESnet6 capacity forecasts.
  – The rest were significantly more lightly loaded (in Jan 2016)
  – This may be due to 'rounding up' to the next possible link size, monthly variations, or many other issues.

• How loaded a network can run before users experience degradation is very dependant on the workload.
  – The evolution of the LHC workflows from mostly bulk data transfer towards more remote IO is making LHC workflows more sensitive to other load.

• Load shaping
  – The experiments have well developed mechanisms to identify different classes jobs that are used by the workflow and storage management systems to optimize resource utilization.
  – There is almost no use of mechanisms to distinguish between high priority and low priority network traffic.
  • Queue depths are in the 3 to 100 millisecond range.
Conclusions

• Evolution in how the LHC community uses the network will have a big impact on network providers

• Changes in the ESnet and the other NRENs might have a big impact on LHC community

• There is room for improvements in the LHC workflow network efficiency
The End

Questions?