Lawrence Berkeley National Laboratory
Lawrence Berkeley National Laboratory

Title
Next Generation Ethernet the time to start is now

Permalink
https://escholarship.org/uc/item/29s051km

Author
Bennett, Mike J.

Publication Date
2006-06-21
Next Generation Ethernet
the time to start is now!

Mike Bennett
Sr. Network Engineer

POFWORLD
June 21, 2006
Discussion Points

• Current “state of the network”
• Why not use aggregated links?
• Challenges with 10G
• Why 100G?
• Questions and wrap-up
Current “state of the network”

- Most sites generally have a mix of 10/100/1000 Mb/s Ethernet connectivity
- Most computers now ship with 1000BASE-T capability
  - Many would argue whether or not it’s necessary on basic host machines
- We’ve seen a significant increase in demand for 1000BASE-T at the host end of the connection
  - No surprise since the computer comes with it
Current “state of the network”

- LBLnet is built in a star topology (feeding 80+ buildings)
Current “state of the network”

- LBLnet is currently upgrading to 10G
  - ISP to Core switch first
    - Security is a challenge (more later)
- Other National Labs have significant 10G networks
  - LLNL has at least 400 10G connections!
  - NERSC is connected to ESnet @ 10G
  - LBNL plans to be connected @ 10G this calendar year
    - ESnet has several 10G MANs
- The question is how do we increase BW until next generation Ethernet comes along?
Why not use Link Aggregation?

- Manageability/`troubleshooting` of multiple physical links for a single logical interface more complex than a serial link
- Traffic flows over aggregated links are not deterministic
  - Force vendors to implement complex hashing functions to try to make flows evenly distributed
    - That’s not free
- Some traffic patterns will always cause inefficient link utilization, including
  - Any traffic that ends up flowing between a few addresses
  - Single interface server with many clients
  - More than 1 Gbps coming from 10 GbE host connections
Why not use Link Aggregation?

• Even with the issues mentioned, many people provision LAG as a “temporary” solution
  o Have to weigh the cost of complexity against the cost of higher BW ports
• Based on a survey of potential users of next generation Ethernet
  o Typical number of channels in an aggregated link is in a range between 2 and 8
  o The difference in cost between 8 aggregated optical 1 GbE links and a single 10 GbE link is not that great
    ~ More cost effective to provision the 10G link
• All things considered, there are still challenges using serial 10G links
Challenges with 10G

• Even though it’s more cost effective to provision the 10G link, there are issues
  o Hard to find security devices, e.g. IDS, that will operated at wire-speed
    ~ They’re out there but not cheap
    ~ For that matter, not all routers and switches operate at wire speed for all frame sizes
  o Price per port is still relatively high
    ~ IEEE “rule of thumb” is 10X increase in BW for 3 X increase in cost
    • We’re not quite there yet
Challenges with 10G

- Current LBLnet security system
Challenges with 10G

- LBLnet 10G security system (phase 1)
Challenges with 10G

- LBLnet 10G security system (phase 2)
Challenges with 10G

- 10GBASE-T was approved for publication on June 9, 2006
  - This should help drive down the cost of 10G
  - This version of copper Ethernet not quite as simple for end users as the previous
    - Multiple cable lengths based on cable type
  - Seems like there may be an opportunity for POF
- Challenges aside, all of this will drive the need for next generation Ethernet
Why 100G?

- First – This is MY OPINION
  - IEEE CFI is next month
  - If a study group is formed, it will be their job to determine the next speed
  - So why do I think it should be 100G?
- Research and education networks support high-bandwidth applications
- Survey of network capacity demands by the research community conducted in 2002 yielded forecast by discipline for 5-year and 10-year windows
- The large dataset size and the geographic distribution of users and resources present major challenges in network bandwidth utilization
Why 100G?

- Researchers forecast
  - 20 major data sites to have better than 10 Gb/s network access in order to keep up with the expected requests for data
  - Most areas of science will generate petabytes of data per year
    - Forecast made in 2002
    - Expected demand between 2007-2012
### Why 100G?

**Table D.1. High-Energy Physics Requirements Summary**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Characteristics that Motivate Advanced Infrastructure</th>
<th>Vision for the Future Process of Science</th>
<th>Anticipated Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Frame</td>
<td></td>
<td></td>
<td>Networking</td>
</tr>
<tr>
<td><strong>Near-term</strong></td>
<td>• Instrument based data sources</td>
<td>• The ability to analyze the data that comes out of the current experiment</td>
<td>• Gigabit/sec</td>
</tr>
<tr>
<td></td>
<td>• Hierarchical data repository</td>
<td>• Remote collaborative experiment control</td>
<td>• End-to-end QoS</td>
</tr>
<tr>
<td></td>
<td>• Hundreds of analysis sites</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>• 100 gigabytes of data extracted from a 100 terabyte data store and transmitted to the analysis site in 10 minutes in order not to destabilize the distributed processing system with too many outstanding data requests</td>
<td></td>
<td>• Deadline scheduling for bulk transfers</td>
</tr>
<tr>
<td></td>
<td>• Improved quality of videoconferencing capabilities</td>
<td></td>
<td>• Policy based scheduling / brokering for the ensemble of resources needed for a task</td>
</tr>
<tr>
<td></td>
<td>• Cross-site authentication/authorization</td>
<td></td>
<td>• Automated planning and prediction to minimize time to complete task</td>
</tr>
<tr>
<td><strong>5 years</strong></td>
<td>• 100 terabytes of data extracted from a 100 petabyte data store and transmitted to the analysis site in 10 minutes in order not to destabilize the distributed processing system with too many outstanding data requests</td>
<td>• Worldwide collaboration will cooperatively analyze data and contribute to a common knowledge base</td>
<td>• 100 gigabit/sec</td>
</tr>
<tr>
<td></td>
<td>• Global collaboration</td>
<td>• Discovery of published (structured) data and its provenance</td>
<td>• Lambda based point-to-point for single high-bandwidth flows</td>
</tr>
<tr>
<td></td>
<td>• Compute and storage requirements will be satisfied by optimal use of all available resources</td>
<td></td>
<td>• Capacity planning</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Network monitoring</td>
</tr>
<tr>
<td><strong>5+ years</strong></td>
<td>1000s of petabytes of data</td>
<td>1000 gigabit/sec</td>
<td></td>
</tr>
</tbody>
</table>
Why 100G?

• For 1 site, 100GB in 10 minutes:
  ◦ 10 GB per minute = 1.33 Gbits per second
• For 20 sites that’s 26.6 Gb/s in 2002.
  ◦ ESnet provides several 10Gb/s links to sites today
• In 2007 it scales by a factor of 10, so for 20 sites, that’s 266 Gb/s
• In 2012, assuming the same scaling factor, that’s 2.66 Tb/s aggregate traffic.
• This is just to support Physics!
  ◦ There are many other branches of research that will consume bandwidth
Why 100G? Observed Network Traffic

ESnet is currently transporting 600 to 700 terabytes/month and this volume is increasing exponentially (approximately 10x every 46 months)

ESnet Monthly Accepted Traffic
February, 1990 – December, 2005
Why 100G? Observed Network Traffic

ESnet traffic has increased by 10X every 45 months, on average, since 1990.

R² = 0.9891
Standards Development Timeline

IEEE Std 802.3u-1995 100Mb/s Ethernet
- Standard approved
- IEEE P802.3u PAR Approved

IEEE Std 802.3z-1998 1Gb/s Ethernet
- Standard approved
- IEEE P802.3z PAR Approved

IEEE Std 802.3ae-2002 10Gb/s Ethernet
- Standard approved
- IEEE P802.3ae PAR Approved

High Speed Study Group (HSSG) formed
- IEEE P802.3u 100Mb/s Ethernet Task Force
- IEEE P802.3z 1Gb/s Ethernet Task Force
- IEEE P802.3ae 10Gb/s Ethernet Task Force
Why 100G?

- The standard will likely not be finished till 2010 which will be just in time to meet this demand.
  - This is assuming a study group is formed next month
  - Again, in my opinion, assuming the next speed is 100G
    - ∼40G is too little too late
- It takes roughly 3 to 4 years to complete standard and 40G is already available
  - Costs vendors several million $ to build next generation equipment, 100G will yield better ROI
Why 100G?

Projected Bytes Transported

- Gigabit Ethernet Completed
- 10 Gigabit Ethernet Completed
- Next Gen. Ethernet Completed?
- Fast Ethernet Completed

Projected Bytes Transported over time:

- Gigabytes per month
- Years 1985 to 2020
- Projected Bytes Transported from $10^0$ to $10^8$
Why 100G?

- Consider that most of this discussion is based on a fraction of the Research and Education market
Why 100G?

- As you can see, there is an increasing demand for next generation Ethernet
  - If we start now, this will be ready for 2012
    ~ Standard should be done in 2010
    ~ First generation hardware *affordable* in 2012
  - Early adopters will buy in 2010-2012 timeframe
- Otherwise network operators will have to aggregate a LOT of links
  - Complexity leads to support problems
    ~ This translates to higher operational costs
    ~ You and I get to pay for it!
Questions and wrap-up

• Sources
  ○ High-Performance for High-Impact Science
  ○ A Vision for DOE Scientific Networking Driven by High Impact Science
    ~ Johnston, Kramer, Leighton, Catlett, March 15, 2002
  ○ ESnet Observed Traffic
    ~ Eli Dart, Network Engineering Services Group (ESnet)
  ○ Ethernet standards development timeline
    ~ David Law, IEEE 802.3 Working Group Vice Chair
  ○ End User Survey
    ~ Steve Garrison, VP Corporate Marketing, Force10 Networks

• Questions?
Thanks!

• If you think of a question after this session
  o mjbennett@lbl.gov

• Thanks for your time!