NANOG 40 – Panel
The Case for 40 Gigabit Ethernet

Richard Steenbergen, nLayer Communications

First the things we can all agree on

- Traffic on the Internet is continuing to grow rapidly
- Ethernet will play an important roll in handling this growth
- Ethernet speed/technology is constantly improving
- Ethernet adoption in LAN, WAN, and Metro is increasing
- 10GbE will not be sufficient for many users “very soon”
- All signs point to continued growth, past 40G ***and*** 100G
- Being ready for future capacity needs is a good thing
- The time to improve/enhance Ethernet is now
- We as users all benefit from Ethernet’s continued evolution and widespread deployment in a variety of roles

Forcing 100GbE is not the right choice

- But careful analysis leads to an obvious conclusion
- Forcing 100GbE does not help us meet these goals!
  - Adopting 100GbE only will ***increase*** costs for end users
  - Adopting 100GbE only will greatly ***slow*** availability of > 10GbE
  - Adopting 100GbE only will result in increased ***incompatibility*** with existing transport infrastructures
- 100G is an arbitrary number and a poor fit for the current technology, solely for the marketing claim of “10x faster”.
- Deciding the future of the industry based of nothing more than logic like “100 bigger than 40” is astoundingly

Reality Check – Optical Technology

- The “sweet spot” for commodity optics today is 10G
- 40G optical technology available, but has disadvantages
  - Currently more expensive than simply using 4x10G
  - Currently not capable of longer distances like 10G
  - Current generation of optics are bulky and unwieldy
  - But many native 40G systems have been deployed already
  - Recent RZ-DQPSK work suggests longer reaches possible soon
- 100G serial optical technology is not available today
  - Chromatic Dispersion increases exponentially with the bit rate
  - This makes 100G 100x harder to implement than 10G
  - It will not be commercially viable for many years to come.

So how would 100GbE be implemented?

- Parallel paths of obtainable lower-speed technology
- Current proposals center around 4x25G waves
- So what's wrong with using parallel optical paths?
  - Inherently, nothing at all!
  - This is a sensible way to meet the need for higher speed
    - Like using multi-core CPUs instead of just increasing clock speeds
    - Already successfully used in technologies like 10GBASE-LX4
  - But the choice of component wavelengths should be based on sensible compatibility with existing infrastructure!
- Ditching compatibility with existing technology and hardware to hit an arbitrary number is a bad idea.

Transport Compatibility

40 Gigabit Ethernet

- 40G transport products have been shipping for many years.
- There is a significant installed base of ITU G.709 OTU3 40G capable DWDM transport platforms.
- These platforms continue to be deployed an increasing rates
- 40G could be transported over carrier DWDM infrastructures without forklift upgrades

100 Gigabit Ethernet

- No existing 100G products
- No existing standards for 100G Optical Transport Network
- No significant availability of 25G components today
- The best case scenario is transport of 10x10G waves
- No compatibility with native 40G optical technology

Reality Check – Router Implementations

- High end routers have done 40G/slot for some time
  - But most are really 2x20G or 4x10G, very few support 40G
    - Cisco CRS-1
    - Juniper T640 with FPC4
  - Parallel paths at the IP level are tricky due to reordering
- The current direction in router evolution is true 40G
  - Router capacity will be doubled by implementing 2x40G
  - Packet Processing ASICs and Fabric must be upgraded
    - But easily accomplished with new fabric modules and linecards.
- 40G is available today or coming soon, true 100G will not be possible let alone implemented for years.

The “benefits” of a 100GbE choice

- No re-use of existing 40G optical research/components
  - Higher prices are the natural and inevitable result.
- No compatibility with existing 40G transport platforms
  - Harm’s Ethernet’s usefulness outside of the LAN.
- No clear path for routers to support 100G in the near term
  - Significant investment in upgrades required when it does happen
- Lack of a modular approach to delivering needed capacity
  - Not everyone needs 100G for their application/architecture
  - ***Many*** uses of 10G today are for aggregation of only 1-3 Gigs
  - Users wanting the benefits of > 10G would be forced to pay for entire 100G, even if they don’t need a 90G jump in capacity.

Scaling capacity with parallel paths

- Today we do this with 802.3ad Link-Aggregate (LAG)
  - Many large networks run 8x10GbE today quite successfully
- But there are significant and fundamental drawbacks
  - Hashing algorithm required to prevent TCP packet reordering
    - Requires deep inspection of packet headers
    - Results in imperfect distribution of traffic across multiple links
    - Designing good hash algorithms for every bundle size is challenging
  - Individual flows are limited to the size of the member channels
  - End user responsible for managing multiple links
    - Increased cable management complexity
    - Increased logical interface management complexity
    - Increased complexity results in more potential for configuration errors
  - Slow software-based negotiation protocol (LACP)

Scaling capacity with parallel paths

- A better solution is aggregation at the physical layer
  - Avoids the limitations of layer 2 LAG protocols
  - End-runs the entire 40GbE vs 100GbE argument completely
    - Develop component channel specifications based on the sensible requirements of optics and ASICs, not political requirements
    - Separate the development of “Ethernet” MAC from the speed of the protocol, quicker delivery of incremental speed increases to market
    - If 40G isn’t fast enough for you, use 80G, 120G, 160G, 320G, etc.
- Some vendors are already delivering proprietary solutions
  - Juniper OC-768c over 4xOC192 with inverse multiplexing
- Some attempts to propose this in the IEEE as well
  - http://grouper.ieee.org/groups/802/3/hssg/public/nov06/frazier_01_1106.pdf

The Arguments of 100GbE Proponents

- "But we can do 4x10GbE today, we need ***revolutionary***!"
  - We've already acknowledge that we can't meet all bandwidth requirements with a single channel solution, 100GbE is no different, it just disguises this fact to those who don't know better.
  - 100GbE does nothing revolutionary to Ethernet, it just makes it incompatible for little to no gain. We need evolutionary progress.
- "We need faster NOW! Look at my up-and-right graphs!"
  - If this is true, can we really afford to wait for 100GbE components to become feasible before upgrading our 10GbE? It'll be YEARS.
- "Working on 40GbE will delay 100GbE significantly"
  - There is absolutely no evidence whatsoever that the only thing holding back 100GbE is a standard. There is time to do both.

The Arguments of 100GbE Proponents

- “40GbE will take just as long to develop as 100GbE”
  - Regardless of the administrivia of the standardization process, 40GbE technology is significantly more fleshed out than 100GbE, and can be delivered quickly with significantly less debate over its implementation.

Final Arguments

- One of the great values of Ethernet is its versatility
  - Whatever speed and technology we decide to adopt, it should be compatible with optical and transport infrastructure standards.
- We should really be working towards supporting both
  - There is a very real future for 100Gbps optical signals
    - 107Gbps DQPSK systems being designed and tested today
    - The same way 40G signals were being tested in 1999
  - But 40GbE is what needs a rush to standardization, it CAN be implemented quickly and has clear advantages in compatibility and deliverability over the next 5 years.
  - Blocking it's development hurts everyone.