100G versus '40G and 100G'
or
Single Rate versus Dual Rate

Gary Nicholl – Cisco
IEEE 802.3 Higher Speed Study Group
Geneva , 28-31 May 2007

Supporters

- Andy Moorwood – Extreme Networks
- John Jaeger – Infinera
- Drew Perkins - Infinera
- Thomas Fischer – Nokia Siemens Networks
- Henk Steenman – AMS IX
- Alan Judge – Amazon
- Troy Sprenger – EDS
- Jay Moran – AOL
- Mark Nowell – Cisco
- Donn Lee – Google
- Ted Seely – Sprint
- Bill Trubey - Time Warner Cable
- Mark Kortekaas - CBS Interactive
- Mike Bennett – LBNL
- Joe Lawrence – Level 3
- Vik Saxena – Comcast
- Greg Hankins – Force10 Networks
- Peter Schoenmaker, NTT America
- Adam Bechtel – Yahoo
- Frank Chang – Vitesse
- Dan Dove – ProCurve Networking by HP
- Bill Woodruff – Aquantia
- Med Belhadj – Cortina
- Brad Booth – AMCC
- Wenbin Jiang – JDSU
- Bill Ryan – Foundry Networks
- Shashi Patel – Foundry Networks
- Michael Krause – HP ESS

Presentation Motivation

- There has been a lot of discussion within the HSSG on the subject of 40G and 100G
- There appears to be general consensus that a 100G rate is definitely required, so the debate is really around whether 40G should be included in addition to 100G
- The decision boils down to does the group want to move forward with a Single rate (100G only) or a Dual rate (100G and 40G) solution.
- This presentation reviews the implications of moving forward with a Dual Rate approach

Outline

- HSSG Objectives Recap
- 40G and 100G Applications Recap
- Single or Dual Rate Path
- Dual Rate approach
  - Advantages and Disadvantages
  - Standard's impact
  - Industry impact
  - Historical Perspective
- Summary and Recommendations

HSSG Objectives Recap

- Nine objectives have been adopted by the Study Group:
  - Support full-duplex operation only
  - Preserve the 802.3/Ethernet frame format at the MAC Client service interface
  - Preserve min and max FrameSize of current 802.3 Std
  - Support a BER better than or equal to 10-12 at the MAC/PLS service interface.
  - Support a speed of 100 Gb/s at the MAC/PLS service interface
    - Support at least 100 meters on OM3 MMF.
    - Support at least 10km on SMF.
    - Support at least 40-km on SMF.
    - Support at least 10m over a copper cable assembly.
- The decision to also include a 40 Gb/s rate objective is a critical one, and could cause the SG effort to be delayed or deadlocked.

100G and 40G Applications Recap

- 100G Applications
  - Data center interconnect
  - HPC (High Performance Computing)
  - Aggregation and Core interconnect
  - Enterprise campus interconnect
  - Server NIC cards (~ 2018)
- 40G Applications
  - Server NIC cards (~ 2013)
  - Server to switch connections
    - Pedestal servers
    - Rack servers
    - Blade servers

Single or Dual Rate Path

- 100G is ‘locked and loaded’
  - see dove_01_0507 for supporting details
- 40G still requires additional effort
- even 40G proponents do not want to hold up 100G (e.g. unanimous straw poll in Ottawa)

Advantages and Disadvantages

- Advantages
  - A more optimal solution for server NIC applications (10G->40G->100G)
- Disadvantages (Risks)
  - Fragmentation of R&D efforts (lack of critical mass on either 40G or 100G initially)
  - Industry confusion on "application versus rate"
  - Interoperability concerns (some vendors elect to implement 40G initially, whereas others implement 100G)
  - Potential to delay 100G. 100G project is ready to move forward. 40G still requires additional work.
  - "Muddies the waters"

Standard's Impact

- The standard's impact could be minimal (muller_01_0407.pdf):
  - MAC is (sort of) bit rate independent …
  - 100G PCS/CTBI proposal could be easily scaled to support either rate, however ….
- The standard's impact could possibly be more significant – depending on the Task Force directions taken:
  - IF the MMF PHY chooses the 12x10G path with 8b/10b to more closely align with Infiniband – then not sure what the 40G MMF PHY would entail?
  - IF the 100G copper path is 4x25G, then would a 4x10G PHY be an independent effort – although easier if the same distance is chosen?
  - Backplane – does not exist in the current 100G path, this effort is incremental
- So there is still uncertainty if the PHY/PMD work is trivial for 40G or essentially twice the work of 100G only
- ALSO there is no precedent for IEEE developing two new significant ethernet rates on the same timeline
- BUT .. the standard's effort is not the major concern here

Industry Impact

- Although the requirement for dual rates may initially be driven be the needs of two ‘distinct applications’, having two rates will ultimately force component and equipment vendors to support BOTH.
- Switches which connect to servers will require both rates upfront (40G downlink, 100G uplink)
- Domino effect. There are two ends to every link. 40G on one box forces 40G on other boxes.
- Requires the industry to develop 2 x MACs, 2 x PCS chips, 2 x PMA (serdes) chips, 2 x N PMDs.
- Maybe it is Triple rate ?? (40G LAN, 40G WAN, 100G)

Historical Perspective

- History has shown that standardizing two solutions simultaneously, is not a successful (or at least efficient) approach. One solution ultimately dominates the other (although both take the same amount of effort to develop initially)
- 10GE: LANPHY versus WANPHY
  - WANPHY was standardized as a WAN friendly rate for 10GE
  - component / system companies “forced” to develop WANPHY interfaces
  - but LANPHY ultimately won in the WAN as well, for cost/volume reasons
  - LANPHY shipments ~ 1M , WANPHY shipments significantly less
  - Result: a lot of wasted time and effort
  - WANPHY support still being added to new designs – and so it continues !!
- IETF VPLS (Virtual Private LAN Service)
  - IEFT ‘standardized’ two solutions at the same time
  - initially resulted in interoperability issues between vendors (who choose to implement only one of the solutions)
  - end users ultimately forced vendors to implement both solutions
  - Result: a lot of wasted time and effort.

Historical Perspective (cont'd)

- 10G EFEC (ITU G.975.1)
  - multiple, incompatible high gain FEC algorithms 'standardized'
  - no interoperability
  - not a big issue for initial application (long haul DWDM transport)
  - BUT .. lack of a single 'standard' FEC algorithm is slowing the adoption of integrated DWDM interfaces on client equipment (routers, switches, etc)

Summary

- The current 40Gb/s debate boils down to a decision on a Single Rate versus a Dual Rate path.
- This decision must not be taken lightly. There are a lot of ramifications (we have to live with our ‘mistakes’ for a very long time).
- While defining two (or more rates) initially and ‘letting the market decide’ is one (easy) approach, it is far from being the most efficient.
- Ethernet has been successful by defining the minimum solution set for a broad range of applications (thus providing clarity and focus for the industry).
- Recommendation:
  - HSSG proceed with a single 100G Rate PAR
  - If there is a distinct need/application for 40GE, then it should be addressed in a separate CFI/SG/PAR