Positioning 40 and 100 GbE in data center inter-switch link applications and 40GbE PMD recommendations

Adam Carter, Cisco
Alessandro Barbieri, Cisco
Data Center inter-switch link applications

- The adoption of 40GE as objectives for 802.3ba will address the server interconnect application.
- However, it is inevitable that the technology will migrate into the next layers of the network and be adopted into the inter-switch links within the data center.
- The drivers for this migration and the requirements of these inter-switch links will be discussed.
List of supporters

- Dan Dove, ProCurve Networking by HP
- Chris Cole, Finisar
- Pete Anslow, Nortel
- Steve Trowbridge, Alcatel-Lucent
- Rick Rabinovich, T-PACKETS
### Ethernet architectures in the data center

<table>
<thead>
<tr>
<th>Prevalent Media</th>
<th>Three-tier architecture</th>
<th>Switch Topology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metro Single Mode Fiber</td>
<td>Core</td>
<td>High Density Modular Chassis</td>
</tr>
<tr>
<td>Multimode fiber structured cabling 150m @ 95% coverage (kolesar_01_0906.pdf)</td>
<td>Distribution</td>
<td>High Density Modular Chassis (End of Row)</td>
</tr>
<tr>
<td>Rack server: UTP Copper Blade Server: copper backplane</td>
<td>Modular</td>
<td>Top of Rack (1U) Embedded Blade Switch</td>
</tr>
<tr>
<td>ToR</td>
<td>Blade Switches</td>
<td>Servers</td>
</tr>
</tbody>
</table>

---

**Why do we need layers?**

1. Accommodate high-density server deployments.
2. Accommodate network growth (easier than a mesh topology to bring a new rack or row coming online).
3. Improve scalability of L2 network protocols - network design best practice.
The uplink speed at each switching tier is determined by two factors:
1. The downlink speed connecting to one layer below.
2. The number of downlinks being aggregated.
Ethernet speeds positioning in the data center for 1 GbE and 10 GbE optimized access

1 GbE Server Access

Three-tier architecture

10 GbE Server Access

40/100 GbE

40 GbE

10 GbE blade switches (16 servers)
or 10/40 GbE ToR (~30-40 servers)

1 GbE

Top of Rack (1U - 48 ports)

Blade Switch (24-32 ports)

Modular Chassis

100 GbE

40 GbE blade switches

40/100 GbE ToR

10 GbE

For power, size/density and cost reasons 40G is likely to dominate access-to-distribution switch uplinks for several years after the IEEE 802.3ba standard will be ratified.
The transition of port speeds in switches

For example: 10M to 100M

- On switches, as technology matures and costs decrease, the ratio of high speed to low speed ports alters.

- It took 4 years for 100M ports to match 10M ports shipped.

- Beyond the 1:1 point, ratio takes off (i.e. low speed virtually disappears).

- This transition is driven by technology cost.

Source: Dell'Oro
(high speed : low speed shipped ports ratio)
Higher-Speed Ethernet adoption dynamics on servers and switches

- Mass adoption of Higher Speed Ethernet (HSE) on server ports is driven by the availability of HSE on LAN on Motherboard (LOM).

- Although HSE on server ports eventually drives HSE adoption on switch ports, the transition is not nearly as fast on the switch side!

Note: Switch port data are not broken down by data center and wiring closet.
The need for solutions optimized for 1 GbE aggregation is here to stay well beyond 2010

High/low speed switch port units ratio

Source: Dell’Oro Q3’07 (high speed : low speed shipped ports ratio)

Note: Switch port data are not broken down by data center and wiring closet.
## 40G PMD Fiber Requirements in the Data Center

<table>
<thead>
<tr>
<th>Aggregation</th>
<th>Distance and media requirements for 40G</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OM-3 ribbon</strong></td>
<td><strong>OM-3 duplex</strong></td>
</tr>
<tr>
<td><strong>Single mode ~ 10km</strong></td>
<td><strong>OM-3 ribbon</strong></td>
</tr>
<tr>
<td><strong>OM-3 ribbon</strong></td>
<td><strong>40 - 100G</strong></td>
</tr>
<tr>
<td><strong>10 - 40G</strong></td>
<td><strong>Legacy</strong></td>
</tr>
</tbody>
</table>

A multi-fiber PMD should represent the most cost effective interconnect solution for 40 GbE, hence it is expected that a 40G PMD supporting ribbon fibers, will be instrumental to drive high-volume adoption of 40G.
40G Requirements Summary

- **Drivers for 40G on OM-3 ribbon:**
  - Access to distribution uplinks in greenfield data centers
  - Access to distribution in legacy data center future-proofed for 100 G and/or 40/100 mix (blade switches vs. ToR switches)

- **Drivers for 40G on duplex OM-3:**
  - Access to distribution in legacy data centers with cable infrastructure designed for 10 GbE

- **Drivers for 40G on SMF:**
  - Core links - for SMF we need a minimum of 10km to be backward compatible with lower data rates.
Recommendations

Now that 40 GbE has been adopted as part of the 802.3ba Task Force, there is a need to consider inter-switch links applications at 40 GbE. Our recommendation is that the 802.3ba Task Force should consider the following:

1. The DC inter-switch link application should be considered to be part of the 40G MMF reach discussion.

2. Consider as part of the 40G MMF objective both a multi-fiber and a duplex fiber.

3. Adopting a new objective for 40G SMF with minimum reach of 10km.