Drivers and Applications of Optical Technologies for Internet Data Center Networks

Vijay Vusirikala, Cedric Lam, Paul Schultz and Bikash Koley
Google, 1600 Amphitheatre Pkwy, Mountain View, CA 94303
vijayvusiri@google.com

Abstract: The rise of large-scale Data Centers to power the Internet infrastructure is driving new architectural directions for optical networking. This paper highlights these architectural options and discusses technology building blocks for scaling inter-Datacenter connectivity.

OCIS codes: 060.4250   Networks; 060.1660   Coherent communications; 060.4256   Networks, network optimization

1. Introduction

Increasing broadband penetration in the last several years last few years has resulted in dramatic growth in innovative, bandwidth-intensive applications that are being embraced by consumers. Coupled with this consumer trend is the migration from local compute/storage model to a cloud computing paradigm. The net result of these two trends is the emergence of Warehouse Scale Computing (WSC) which provides a ubiquitous, interconnected computing platform as a shared resource for many distributed services [1]. In this paper, we describe the drivers behind the network scaling requirements and discuss architectural options for scaling the inter-datacenter connectivity n a cost-effective and operationally streamlined manner.

2.   Long Haul Networks for Datacenter Operators

Figure 1 below shows a high-leveloverview of the network topology of large WSC operators. At the center is a cloud of geographically distributed mega datacenters connected by a high-capacity network. A user accesses the cloud through his/her local network service provider's network, which interconnects with the datacenter network through Internet Points of Presence (POPs). As far as a user is concerned, the datacenter network appears as a single computing infrastructure. The cloud infrastructure, which consists of the network and multiple warehouse-scale datacenters, is transparent to the user.

There are two primary types of traffic traversing this network (a) User Traffic: This traffic is typically user initiated (e.g., mail, video, images etc.) and is served from the datacenters through the carrier networks that are interconnected at various serving PoPs, (b) Machine-to-machine traffic: This machine-machine communication traffic is from one datacenter to another without exiting the datacenter cloud to the end users. Examples of this traffic are data mirroring, redundancy, database sync etc.
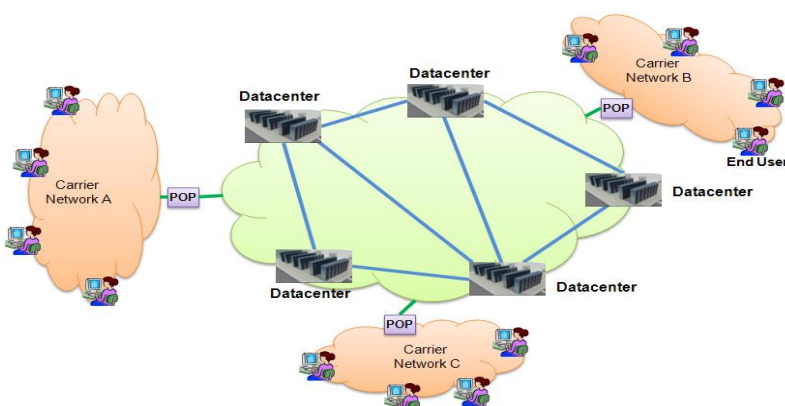

Figure 1 Overview of network topology for large WSC operators

3. Architectural Options

These two traffic types differ in terms of traffic volume and Service Level Agreement (SLA) requirements. We investigated various architectural options for the design optimization of a network consisting of these two traffic types at different relative load factors. The options considered are shown in Figure 2 and described below:
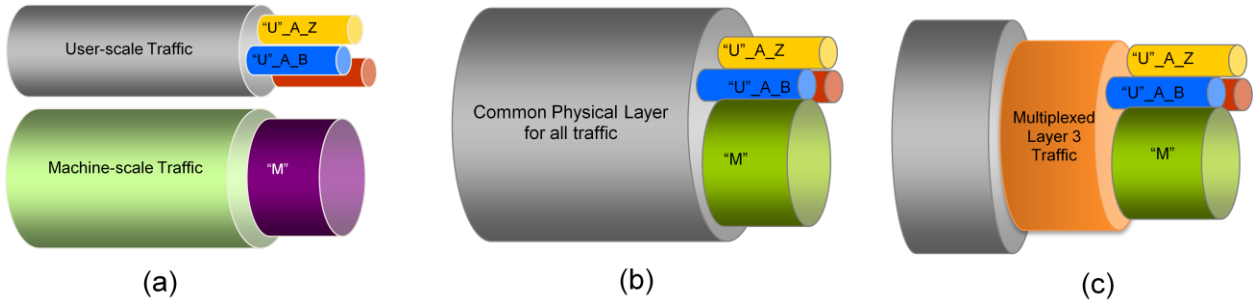
Figure 2 Overview of architectural options considered for WSC networks (a) Overlay, (b) Layer-1 Converged and (c) Layer 1 + Layer 3 Converged

(a) **Overlay Architecture:** In this architecture, different physical and routing layers are used for the two traffic types - User ("U") and Machine ("M") types of traffic. The different SLA requirements between the two traffic types enabled a cost-performance optimization in both the transport and routing networking equipment independently in this architecture, while the converged architecture models described below necessitate the use of common equipment (at transport and/or routing layers) for both traffic types.

(b) **Converged Architecture at Layer 1 only:** In this architecture, a common physical layer (Layer-1) is used for both traffic types but different routing equipment is used for the two traffic types.

(c) **Converged Architecture at Layer 1 and Layer 3:** In this architecture, common physical and routing layers are used for both traffic types.

Models (a) and (c) assumed the use of low-cost Label Switch Routers (LSR) to take advantage of cost reduction to switch traffic that is transiting through an intermediate node. We assumed that the LSR used in the model has integrated DWDM optics to avoid the cost of client-client optical connectivity. Based on industry standard pricing for optical transport/switching and routing equipment, we performed a cost comparison of the three options for different relative ratios of the "U" and "M" scale traffic.

The results from the modeling are summarized in Figure 3 below for two relative traffic loads (a) Baseline for "U" scale and "M" scale traffic, and (b) Same "U" scale traffic but with 2X the "M" scale traffic. This ratio of M-traffic scaling relative to U-traffic scaling is representative of WSC networks where machine-to-machine traffic is growing faster than the User scale traffic.
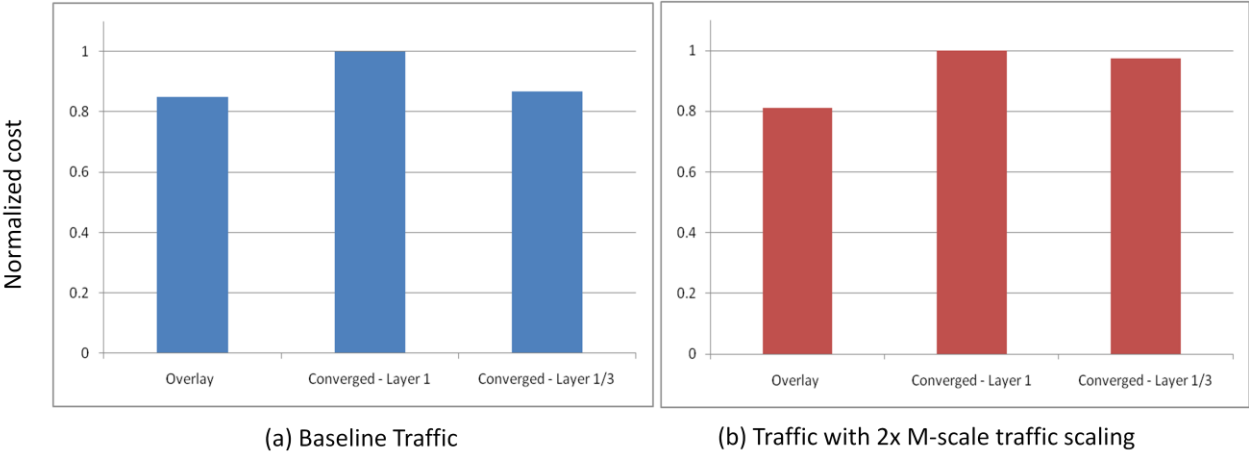

Figure 3 Results showing the normalized cost comparison for the three architectural models considered for two traffic scaling patterns.

The results above demonstrate that the architectural choice is very dependent on the relative traffic mix between the U-scale and M-scale traffic. As the M-scale traffic grows in proportion, it becomes cost advantageous to consider an overlay network.

4. Technology Building Blocks for Scaling, Flexibility and Optimization

Cost-effective scaling of high-capacity optical transport links is a critical need for WSC computing infrastructure. As mega-datacenters can be widely distributed geographically and access to locations for optical regeneration can be limited, there is a strong need for high-capacity as well as ultra-long haul unregenerated reach capability. Various emerging technology building blocks offer the potential for this cost-effective capacity scaling as described below:

a) **Higher capacity per fiber:** Optical transport solutions that increase the maximum capacity per fiber beyond today's commercially available 8 Tb/s (based on 80 channels of 100Gb/s transmission in C-band). Published literature has shown a roadmap to continued fiber capacity scaling using a number of approaches for increasing the spectral range and spectral efficiency [2,3]. These include higher data rates, higher-order modulation, OFDM, multiple transmission bands etc.

b) **Unregenerated reach:** As the transmission data rates increase, the unregenerated reach typically decreases due to the higher OSNR required. The use of techniques such as Soft-Decision FEC will help bridge the gap. In addition, techniques for maximizing optical link OSNR across the transmission spectral range such as optimized Raman amplification, tilt control, spectral equalization, per-span launch power adjustment, can be used to increase the maximum unregenerated reach.

c) **Variable Rate Optics:** With coherent optical transmission systems, it is possible to have a variable transmission rate that is based on the link quality and condition. For example, for shorter links or links with "good" fiber types, the additional optical link margin can be used to transmit higher data rates. This type of variable rate transmission has to be tightly integrated with the packet layer and managed at the system/network level to realize the overall throughput maximization.

d) **Flexible grid ROADM:** Current commercially available ROADMs are based on 50GHz or 100GHz ITU grid spacing. These fixed grid ROADMs become a limitation for future capacity scaling and network flexibility. Emerging flexible grid or gridless ROADMs [4,5], which provide the ability to arbitrarily determine spectral pass bands and spacing between pass bands, enable two key functionalities (i) Support for higher data rates in a spectrally efficient way by packing the wavelengths in a manner determined by the spectral content of the waves rather than the limitations of the ITU grid, and (ii) Flexibility for arbitrary add/drop of wavelengths in a manner independent of the underlying data rate or modulation scheme.

e) **Large Core Fibers:** With the ability of coherent systems to compensate for fiber impairments such as chromatic dispersion and polarization mode dispersion, the major remaining fiber impairments that limit transmission are the fiber attenuation and fiber non-linearities. Recent advances in large-core ($\sim$110 um^2), low-attenuation (<0.17 dB/km) fibers demonstrate the capability to increase transmission distance for a given fiber capacity [6]. The large effective area enables a lower power density which helps alleviate penalties from fiber non-linearities.

5. References

[1] L. A. Barroso and U. Hölzle, The Datacenter as a Computer - An Introduction to the Design of Warehouse-Scale Machines, Morgan & Claypool, 2009. http://www.morganclaypool.com/doi/pdf/10.2200/S00193ED1V01Y200905CAC006

[2] René-Jean Essiambre, Gerhard Kramer, Peter J. Winzer, Gerard J. Foschini, and Bernhard Goebel, "Capacity Limits of Optical Fiber Networks," J. Lightwave Technol. 28, 662-701 (2010)

[3] K. Roberts, "Digital Coherent Optical Communications beyond 100 Gb/s," in Signal Processing in Photonic Communications, OSA Technical Digest (CD) (Optical Society of America, 2010), paper JTuA1

[4] C. F. Lam and W. I. Way, "A System's View of Metro and Regional Optical Networks," Photonics West, San Jose, CA, Jan. 29, 2009.

[5] M. Jinno et al., IEEE Commun. Mag., Nov. 2009, pp. 66–73.

[6] R. Chen, M. O'Sullivan, C. Ward, S. Asselin, and M. Belanger, "Next Generation Transmission Fiber for Coherent Systems," in Optical Fiber Communication Conference, OSA Technical Digest (CD) (Optical Society of America, 2010), paper OTuI1.