Advantages and Control of Hybrid Packet Optical-Circuit-Switched Data Center Networks

John Bowers (1), Atiq Raza (2), Daniel Tardent (2), Jitender Miglani (2)

UCSB, Electrical & Computer Engineering Department, Santa Barbara, CA 93106(1), CALIENT Technologies, Goleta, CA 93117(2)

Abstract: Modern Data Center networks must perform at near 100% utilization in order to economically support the high bandwidth, highly elastic traffic patterns and data flows seen in back-end networks. Hybrid Packet-OCS networks offer excellent CAPEX and OPEX advantages but require coordination of traffic between packet and circuit fabrics. This paper provides analysis of the economic benefits and describes a practical management and control plane architecture.

OCIS codes: (060.4253) Networks, circuit-switched; (060.6718) Switching, circuit; (060.6719) Switching, packet

1. Introduction

Mobile data, video, and social media applications are dramatically escalating traffic in data centers. Cisco’s Global Cloud index report forecasts that data center traffic will triple to reach 7.7 zettabytes annually by 2017, representing 25% CAGR[1]. Global cloud data center traffic, the fastest-growing component of data center traffic, is projected to grow 4.4x, representing 35% CAGR [1].

It is significant that on average 76 percent of data center traffic from 2012 through 2017 is projected to remain inside the data centers. This represents east-west traffic between racks and clusters contributed to by functional separation of application servers, storage, and databases, which generates replication, backup, and read/write traffic traversing the data center. Other major contributing factors to east-west traffic are server virtualization and parallel processing.

Rapid growth in internal data center traffic brings with it a serious set of challenges, not the least of which is the cost to deploy large amounts of additional capacity, particularly as interface rates scale to 40Gbit/s. The power required to switch the traffic is a significant expense and lower power switching is required[2]. With escalating bandwidth demands, the cost of maintaining traditional utilization rates (<50%) is no longer acceptable. Networks must be able to support up to 100% utilization of resources.

A second challenge is that modern cloud-scale datacenters must support a variety of traffic patterns including traditional short-lived (aka “mice”) flows, and more persistent (aka “elephant”) flows. Mice are typically associated with bursty, latency-sensitive applications while “elephant” flows include traffic generated by virtual machine migrations, data migrations, MapReduce, as well as certain real time events. Typically the majority of flows within a data center are mice, but the majority of packets belong to a few highly persistent elephant flows.

Highly persistent TCP flows tend to fill network buffers end-to-end, and this introduces non-trivial queuing delay to anything that shares these buffers. In a network of elephant and mouse flows, this means that the more latency-sensitive mice are adversely affected [3].

Such elasticity in the bandwidth demands requires application level control so that relative priorities and traffic flows (both short-lived and persistent) can be handled and routed appropriately within the network. This is needed to avoid alternative approaches that require heavy and expensive over-provisioning.

2. The Hybrid Packet-OCS Datacenter Network

The cost and performance benefits of hybrid data center networks consisting of both packet and Optical Circuit Switched (OCS) elements have been explored and elaborated on in a number of research projects including Helios and c-Through [4,5]. In such networks (as depicted in Figure 1), the packet network provides relatively lower-capacity any-to-any connectivity for handling short-duration flows of data between top of rack switches (TORS), while the OCS network provides the high capacity links to transport the majority of the traffic, which is contained in more persistent flows.

For the hybrid network to function, a management and control plane is needed to monitor the data center traffic matrix and the specific flow requirements and priorities of applications. Together with a historical view of traffic patterns, this allows the control plane to open and close connections between TORS through the Optical Circuit Switches and update the routing tables in the packet network devices to provide an optimal best-fit network in real time. As an example, highly persistent elephant flows, when detected, can be offloaded from the packet network to...
the OCS network, thereby facilitating faster delivery of the flow while freeing up the packet network to focus on handling the short-lived flows that it services most effectively.

Figure 1 – The Hybrid Packet-OCS datacenter Network

3. Economic and Power Advantages

Optical Circuit Switching is the most cost effective method to add high bandwidth datacenter network capacity. Recent analysis[6,7] shows that the Capex cost of adding 10Gbit/s capacity between TORs with OCS is one half that of a commercial packet based solution. At 40 Gbit/s, the comparison is even more compelling with OCS delivering capacity at one seventh of the cost of packet switching. The power advantage is even more compelling[2,8], which drives lower Opex.

<table>
<thead>
<tr>
<th>Switch Type</th>
<th>CAPEX per Gb/s (10G)</th>
<th>CAPEX per Gb/s (40G)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Packet-Based</td>
<td>$115</td>
<td>$322</td>
</tr>
<tr>
<td>OSC</td>
<td>$57</td>
<td>$46</td>
</tr>
</tbody>
</table>

The CAPEX comparison was based on comparing a commercial 768 Port 10GbE switch (96 Port 40GbE) equipped with generic optical transceivers with a Calient S320 OCS. Both solutions included the cost of optical transceivers on TORS, with SR optics assumed for the packet solution and LR optics for the OCS solution. All costs are expressed as CAPEX per Gbit/s.

As an extension of the economic analysis, we also compared the Total Cost of Ownership (TCO) of circuits based on OCS and packet solutions. This analysis was conducted for 40Gb/s port rates using a TCO model adapted from Google’s “Datacenter as a Computer” [9], and shows that over a twelve-year data center lifetime, TCO for an OCS based solution is more than 8.8 times lower than that for a packet solution. In addition to lower Capex, key reasons for the much lower OCS TCO include extremely low power consumption (45 watts for an S320 OCS), and complete scalability – meaning that the OCS will scale without upgrade to support 40Gb/s, 100Gb/s and beyond.

<table>
<thead>
<tr>
<th>Basis</th>
<th>40G Packet Solution</th>
<th>40G S320 Solution</th>
<th>Reduction</th>
<th>% Reduction</th>
<th>40G Packet Solution</th>
<th>40G S320 Solution</th>
<th>Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAPEX</td>
<td>$12,890</td>
<td>$1,840</td>
<td>7.00 x</td>
<td>85.7%</td>
<td>$3.71M</td>
<td>$0.53M</td>
<td>$3.18M</td>
</tr>
<tr>
<td>3 Yr TCO</td>
<td>$17,200</td>
<td>$2,340</td>
<td>7.35 x</td>
<td>86.4%</td>
<td>$4.96M</td>
<td>$0.67M</td>
<td>$4.28M</td>
</tr>
<tr>
<td>6 Yr TCO</td>
<td>$34,410</td>
<td>$4,180</td>
<td>8.23 x</td>
<td>87.8%</td>
<td>$9.91M</td>
<td>$1.20M</td>
<td>$8.71M</td>
</tr>
<tr>
<td>12 Yr TCO</td>
<td>$68,820</td>
<td>$7,870</td>
<td>8.84 x</td>
<td>88.6%</td>
<td>$19.82M</td>
<td>$2.27M</td>
<td>$17.55M</td>
</tr>
</tbody>
</table>

Key assumptions of the TCO analysis include: (1) Both Solutions include TOR Optical Transceivers, (2) 15% off list prices on all systems and components, (3) Assumes upgrade replacement of Packet-based switch every 3 years, (4) Assumes upgrade replacement of TOR Optical transceivers every 3 years, (5) S320 OCS systems are not replaced over 12 year lifetime, (6) Cost of electricity ($/kWh): $0.062, (7) interest rate: 12%, (8) Data Center Capex
($/W): 15, (9) Data Center amortization period (years): 12, (10) Data Center Opex ($/kW/mo): $0.04, (11) PUE: 2, (12) Switch Opex (Maintenance): 5% of Capex

4. Management and Control Plane Architecture

A management and control plane is required to manage and move traffic flows between the packet and optical circuit networks in response to the needs of specific applications, data flows, and rack-to-rack traffic patterns. A representative architecture is depicted in Figure 2.

![Figure 2 – Management and Control Plane Architecture](image)

The hybrid control plane constantly monitors and analyzes system wide network traffic from netmon tools or sFlow statistics reporting (generated in the merchant silicon of most TORS). vSwitches in virtualized network scenarios report similar data. When combined with historical analytics or scheduled events, it maintains an accurate view of bandwidth demand versus available link capacities and algorithmically computes the optimal circuit topology configuration to serve the needs of the network at any given time. Algorithms ensure that high-bandwidth highly-persistent traffic is served by circuit switched OCS links, while bursty and lower demand traffic is served by packet switched links.

As the demand changes throughout the network, new optimal circuit topology configurations are identified and orchestrated over and over again via the respective packet and circuit switch managers that coordinate with their switches via well-known southbound APIs (OpenFlow, TL1 or any other vendor specific interface). To minimize the Layer 2 and Layer 3 network convergence times during topology changes, the hybrid packet-OCS controller ensures that interface change events (Notify Down or Notify Up) are flooded throughout the network before and after any new topology change at the OCS layer.

5. Conclusion

In conclusion, Hybrid Packet-OCS Data Center networks offer major CAPEX and OPEX savings while providing flexible bandwidth on demand to support the capacity elasticity needed in modern data centers. In order to realize these advantages, however, an effective management and control plane solution is required. Recent advances in Software Defined Networking and flow analytics support the realization of commercially deployable solutions for this application.

4. References

[7] N. Farrington, N. Huang, Packetcounter