End-to-End Demonstration of Optical Multicasting over a Dynamically-Reconfigurable Hybrid Data Center Network Architecture

Howard Wang[1], Yiting Xia[2], Kunwadee Sripanidkulchai[3], Sambit Sahu[4], T.S. Eugene Ng[2], and Keren Bergman[1]
[1]*Department of Electrical Engineering, Columbia University, 500 West 120th Street, New York, New York 10027*
[2]*Department of Computer Science, Rice University, 6100 Main Street, Houston, TX 77005*
[3]*NECTEC, 112 Phahon Yothin Road, Klong Luang, Pathumthani 12120, Thailand*
[4]*IBM T.J. Watson Research Center, 1101 Kitchawan Road, Yorktown Heights, NY 10598*
Email: howard@ee.columbia.edu

Abstract: A hybrid data center network architecture featuring reconfigurable optical functionalities is presented and evaluated on an end-to-end test bed. Physical layer multicast is successfully verified via video streaming and UDP performance measurements.

1. Introduction

Conventional electronic packet-switched networks can no longer adequately support the bandwidth requirements of applications in modern production data centers at reasonable power consumption and costs. To overcome this challenge, recent work has reimagined the data center network through the hybridization of electronic packet-switching with optical circuit switching [1,2]. By leveraging the optical domain to realize reconfigurable high-bandwidth paths between racks, such designs can achieve bisection bandwidths on par with fully-provisioned packet-switched networks, while doing so at significantly less complexity, cost, and power. However, connectivity provided by the MEMS-based optical circuit switches utilized in these architectures are limited to one-to-one matchings between racks. As a result, despite enabling significant improvement for long-lived strictly pairwise communications, the utility of these designs is considerably compromised under the richer connectivity demands of realistic data center traffic patterns [3].

Not limited to the millisecond-scale space switching of MEMS technologies, capabilities ranging from passive wavelength routing to broadband nanosecond-scale switching are available. However, each capability presents its own tradeoffs, ranging from port count limitations to bandwidth scalability. To overcome these tradeoffs, we've proposed an unconventional approach for leveraging these unique photonic capabilities [4]. By dynamically allocating them across the system to better support diverse communication patterns at the optical layer, we achieve a more versatile photonics-accelerated data center network. In this paper, we demonstrate the end-to-end viability of a physical layer multicast as implemented in such a system via both video streaming and UDP throughput measurements.

2. Network Architecture

Due to the view of the data center as a centralized pool of computing resources to be allocated on-demand, system-wide reachability by all servers is an absolute requirement. However, given the rarity of a single application spanning an entire system while requiring full-bandwidth all-to-all communications, there exists an opportunity to utilize the aforementioned optical functionalities in a way that can leverage the respective advantages of each technology at the scale of the data center while maintaining agility in the face of unpredictable traffic demands.

We therefore propose a rethinking of the concept of hybrid data center networks. Instead of merely combining

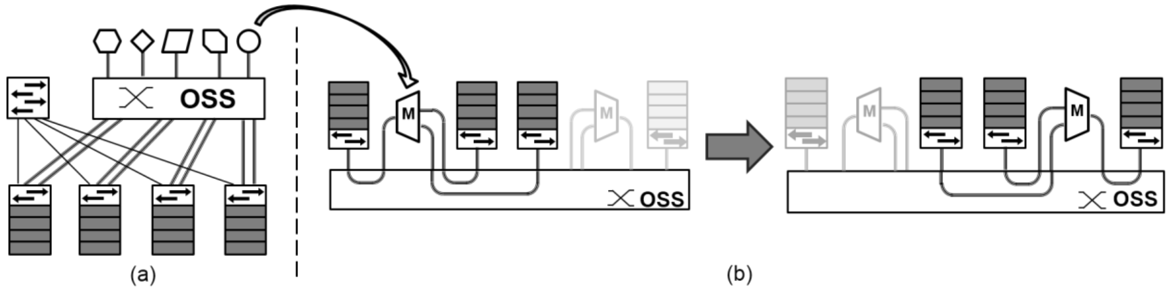

Fig. 1. (a) Hybrid network architecture featuring various photonic capabilities, as represented by the polygons. (b) Utilizing the optical space switch to dynamically deliver optical layer multicast between different racks in a data center.

electronic switches with a single photonic technology, we take the concept further by proposing the hybrid utilization of a reconfigurable pool of a variety of unique photonic functionalities. A high-level representation of our proposed architecture is illustrated in Figure 1(a). Each photonic device delivers a specific physical layer capability, which is then attached to subset of the ports of a high-radix optical space switch (OSS). By treating the OSS as a pool of links delivering system-wide reachability, we can dynamically allocate a veritable library of photonic capabilities across the data center to where they are needed most. Each device and how they are assembled to provide a specific function can be managed by a central controller, which can either accept explicit requests for gadgets or provision them based on demand estimation. By doing this provisioning in a reconfigurable manner, the result is a unique architectural framework for a hybrid network capable of supporting data center traffic in an on-demand capacity while more effectively leveraging the performance, cost, and power advantages of photonic technologies.

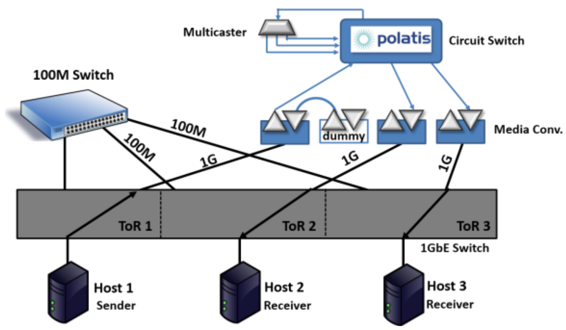

Fig. 2. End-to-end experimental testbed.

Table 1. UDP Multicast Throughput

	At Sender	At Receiver
GbE Packet Switch	826 Mb/s	825 Mb/s
Optical Multicaster	832 Mb/s	831 Mb/s

3. End-to-End Demonstration

We evaluated the physical layer viability of our proposed architecture in [4]. To demonstrate the end-to-end functionality of our proposed system, we constructed a small-scale end-to-end prototype of our architecture enabling multicasting in the optical domain. Our experimental test bed is depicted in Figure 2.

First, a Gigabit Ethernet switch is logically partitioned into three distinct segments, modeling the functionality of three separate top-of-rack (ToR) switches. Next, uplink ports on each ToR switch is connected to both a commodity 100-Mb Ethernet switch and a Polatis piezoelectric beam-steering optical space switch. These uplink ports interface with the optics through media converters attached to SFP transceiver modules, each operating at 1548.51 nm (C36). At a subset of the optical switch's ports, we attach a 1×3 balanced optical splitter, which serves as our optical multicast device. Finally, the OSS is configured to map the input and two of the outputs of the multicast device to each of our three ToRs. In order to ensure link establishment, the Ethernet PHY requires a signal to be incident on its receiver before allowing message transmission on its transmitter. We successfully address this issue by inserting a dummy transmitter, satisfying the aforementioned requirement at the sender. This is not an issue for transceivers at the output of the multicaster.

At Host 1, we run an instance of VLC server and transmit a video stream to IP multicast address 228.8.8.8. At Hosts 2 and 3, we run instances of the VLC client and confirm that the video stream is multicasted correctly. Furthermore, we evaluate the end-to-end performance of our system by measuring the throughput performance of a UDP multicast via *iperf*, a command-line network performance measurement tool, over the optical multicast path in comparison to a 1Gb/s electronic packet switch baseline, the results of which are listed in Table 1. The effective saturation of the sender's output interface in both cases verifies that the throughput performance of our system is on par with or better than a purely electronic packet-switched alternative.

4. References

[1] Farrington, N., Porter, G., Radhakrishnan, S., Bazzaz, H. H., Subramanya, V., Fainman, Y., Papen, G., and Vahdat, A. “Helios: a hybrid electrical/optical switch architecture for modular data centers.” *SIGCOMM Comput. Commun. Rev.* 40, 4 (Aug. 2010), 339-350

[2] Wang, G., Andersen, D. G., Kaminsky, M., Papagiannaki, K., Ng, T. E., Kozuch, M., and Ryan, M. “c-Through: part-time optics in data centers.” *SIGCOMM Comput. Commun. Rev.* 40, 4 (August 2010), 327-338.

[3] Bazzaz, H.H., Tewari, M., Wang, G., Porter, G., Ng, T.S.E., Andersen, D.G., Kaminsky, M., Kozuch, M.A., and Vahdat, A. “Switching the Optical Divide: Fundamental Challenges for Hybrid Electrical/Optical Datacenter Networks,” In *Proc. of SOCC’11: ACM Symposium on Cloud Computing*, Cascais, Portugal, Oct. 2011.

[4] Wang, H., Chen, C., Sripanidkulchai, K., Sahu, S., and Bergman, K. “Dynamically reconfigurable photonic resources for optically connected data center networks.” In *Proc. of OFC/NFOEC’12*, OTu1B.2 (Mar 2012).