Computer Communications 000 (2017) 1–12

Contents lists available at ScienceDirect

Computer Communications

journal homepage: www.elsevier.com/locate/comcom

Optical switching based small-world data center network

Dongxu Zhang, Hongxiang Guo, Tingting Yang, Jian Wu*

Institute of Information Photonics and Optical Communication, Beijing University of Posts and Telecommunications, P.O. Box 55, Xitucheng Rd. #10, Haidian District, Beijing 100876, China

ARTICLE INFO

Article history:
Received 23 December 2015
Revised 26 December 2016
Accepted 2 January 2017
Available online xxx

Keywords:
Network architecture
Data center network
Optical networks
Optical switching
Small world topology

ABSTRACT

Redesigning data center networks, particularly taking advantages of optical switching technologies to reduce cost-per-bandwidth and energy consumption, has spawned interests of researchers worldwide in recent years. In this paper, we introduce and analysis our recently proposed optical switching technology powered data center network architecture, which is inspired from the small world data center network topology yet has additional flexibility brought by optical lightpath's reconfigurability. Specifically, logically full-meshed optical burst switching rings are employed to compose a lattice substrate, which provide dense connectivity for small groups of ToR switches to accommodate highly clustered and dynamic regional data traffic, while a reconfigurable wavelength circuit switching plane offers direct connections among the racks to effectively reduce overall network diameter. A centralized control plane is employed to realize traffic-adaptive optical lightpath scheduling mechanisms. We thoroughly study the throughput and latency performance of the proposed architecture through numerical simulations, and particularly reveal the impact of traffic patterns. It turns out that, the proposed network architecture can achieve high throughput with significantly reduced cost, and exploiting its traffic-adaptive reconfiguration capability can effectively improve throughput and reduce latency under varying traffic distributions.

1. Introduction

With the prosperity of cloud services, data centers (DCs) are becoming the most fundamental IT infrastructures that accommodate huge amount of data and provide tremendous data processing capacities. According to a recent report from Cisco [1], the global DC traffic volume will exceed 15.3ZB in 2020, and more than 70% of which would be within DCs. As such huge amount of data communication strains intra-DC network, rethinking the network architecture and utilizing optical interconnections to improve network performance have become an emerging and attractive research field in both industry and academic communities [2,3,13–36].

Among the requirements of data center network (DCN) design, scalability is the first and foremost challenge to be faced [2]. Scalability is basically about balancing performance and costs as network size grows. Costs include not only equipment acquisition cost but also energy consumption, cooling, and management costs. The costs of traditional DCN architectures (e.g. FatTree [4]) scale super-linearly with the number of servers, imposing a ceiling on the maximum economically-viable datacenter dimension.

A plenty of innovative solutions [5–12] have been proposed to deal with DCN's scalability problem by replacing the canonical multi-tier tree-like network architecture. Among them, the unstructured topology or small world topology based proposals [8–12] provided some valuable clues for constructing DCNs with certain extent of asymmetry or randomness. These researches demonstrated that irregular topologies can outperform traditional symmetric topologies in terms of throughput and robustness, and more importantly have relatively low cost as DCs scale out. However, constructing those irregular topologies with pure electrical switches will require very complicated physical wiring among servers or switches, which may entail error-prone and costly manual intervenes for system maintenance and upgrading. Considering that optical switching technologies are advantaged in exploiting multi-dimension resources (time, frequency, space, etc.) to form elastic and automatically reconfigurable logical topologies, we can expect to design an optical switching fabric that not only realizes the benefits of those unstructured network topologies, but also be very neatly wired without confronting head-scratching cabling problems, and at the same time brings down latency and energy consumptions as traffic loads are offloaded to the optical layer as much as possible.

Despite that there have been quite a number of researches pioneered in introducing optical switching into DCN [13–36], we argue that few of the existing solutions are satisfying for both adequately exploiting the capabilities of optical switching and being technically mature to implement. Concretely speaking, some tended to conservatively use very simple optical networks to just

---

* Corresponding author.
E-mail address: jianwu@bupt.edu.cn (J. Wu).

http://dx.doi.org/10.1016/j.comcom.2017.01.001
0140-3664/© 2017 Elsevier B.V. All rights reserved.

complement existing electrical networks, which fail in making the most of optical switching's elasticity at the system level, whereas there are a lot others attempting to build end-to-end all-optical switching fabrics which are typically either difficult to scale out or too aggressive in adopting advanced optical technologies that are years away from commercially ready (in Section 2 we review those related work). Previously, we also proposed our own design, namely the OpenScale architecture [37,38], which exploits optics' multi-dimensional elastic-bandwidth switching ability yet with all commercially available technologies. By elaborately combining optical wavelength and timeslot switching with electrical packet switching, OpenScale architecture is designed to be a topology-reconfigurable inter-rack network, which in particular conforms to the small world DCN principles but achieves better flexibility than purely electrical switch based implementations. Moreover, OpenScale architecture supports incremental expansion, and it even can be upgraded gradually from traditional electrical switching DCNs in a plug-and-play manner.

In this paper we thoroughly evaluate the performances of OpenScale network in throughput and latency with varying traffic load characteristics. We adopt "throughput" rather than "bisection bandwidth" as the metric of performance evaluation, because it has been proven by Jyothi et al. that bisection bandwidth is an inadequate (or even wrong) metric to reflect the performance of a certain topology [39–41]. However, their work mainly focused on revealing a topology's lower-bound throughput by applying specially designed bandwidth-straining traffic matrix. We argue that, for small world topologies whose strong local-clustering feature intrinsically supports regional traffic, investigating the impact of flows' spatial distribution is more meaningful than just inspecting the lower bound with the worst-case traffic matrix. In fact, DCN traffic can exhibit regional-clustered feature [42,43], and in some cases more than 70% of the traffic volume terminates inside one rack due to the elaborately deployment of applications. Since it is feasible to achieve more regional-clustered traffic by deploying correlated applications close to each other, the small world topology could perform better in providing higher throughput and lower latency because of its high clustering coefficient. Therefore, to evaluate the network performance under different traffic distribution, in this paper we define a traffic clustering factor which reflects the flows' spatial distributions, and study its influences on network throughput and latency.

The contributions and text organization of this paper are as the following: first, along with a brief review of related work (Section 2), the basic idea and principle of constructing an optical small-world network architecture (OpenScale) is described, including detailed discussions on its topology features and node structures as well as some control plane considerations supporting traffic-adaptive topology adjustments (Section 3); second, we comprehensively study the throughput performance of the reconfigurable optical small-world network (Section 4), and in particular, verify its advantage in supporting "regional clustered" traffic which may normally happen in real data centers. Third, we further investigate the network's latency performance (Section 5), and use numerical simulations to demonstrate the latency reduction benefits brought by OpenScale's topology reconfiguration capability. Moreover, before concluding this paper (Section 6), the limitations and possible variations of OpenScale architecture are also discussed (Section 7).

2. Related work

Refs. [5–11] are all proposals of innovative DCN architectures which exploited diverse topologies different from FatTree yet all using pure electrical packet switches. Particularly, authors of [8–12] designed their solutions with heterogeneity or randomness in the topologies, and Jyothi et al. discussed and proved the advantages of these kinds of irregular topologies against conventional symmetric topologies in [37–39], which says irregular topologies can support higher throughput with lower cost and easier incremental expansion. But the common deficiencies of these proposals are that they need complicated wiring among switches, and they do not have topology reconfiguration flexibility which can be provided by optical switching technologies.

Researches [13,14] are among the earliest explorations of introducing optical circuit switching into DCN, and similarly [44] employed wireless interconnections instead of optical lightpaths, which are all solutions that try to augment tree-like electrical DCNs with a separate set of optical or wireless connections. The Mordia [15,16] architecture uses a fast optical switching ring to realize a logically full-meshed and bandwidth elastic inter-rack network, but it faces scalability problem because optical layer's physical impairments limit the number of nodes in a ring. Our proposed OpenScale architecture, to be exact, is partially based on the elastic optical switching rings and can be regarded as an extension solution to Mordia. Similar to Mordia, [17] also uses a single optical ring but with different switching paradigms, and scalability remains as the biggest issue. Authors of [18,19] proposed different solutions to scale-out fast optical switching ring networks with paralleled fiber rings, but the rings cannot be boundlessly stacked up without confronting devices' physical limitations (port counts, insertion loss, etc.) and high complexity of optical contention resolution.

Proposals in [20–26] can be classified into one category for that they are based on one or more high-port-count optical circuit switches directly connecting multiple server racks. OSA [20] is a relatively straightforward solution in which all the racks are equally connected to a big optical circuit switch, which essentially produces a degree-bounded random graph inter-rack topology [10] with optical reconfigurability. Lugones et al. [21] decomposes the big centralized switch into several parts to gain flexibility. To enhance the network with fine-grained optical burst or packet switching as well as the all-optical multicast capabilities, authors of [22–24] proposed to add additional optical function modules on top of the big wavelength circuit switch, thus traffic flows can be dynamically guided according to applications' requirements to the specific switch ports connecting the desired optical network modules. Similarly, Rofoee et al. [25] and Peng et al. [26] also introduced this kind of enhancements based on optical circuit switching, which they named "architecture on demand" system. The proposal in [27,28] described a little different design which simply combines optical packet and circuit switching in a paralleled manner, but it still relies ideally on high-port-count switching fabrics to optically interconnect multiple racks. Proposals mentioned above [20–28] all employ symmetric physical structures based on optical switches with large port count. Ji et al. came up with an architecture which uses an arrayed waveguide grating router (AWGR) in combination with spectrum tunable light sources instead of optical switches to perform data switching [29], but the problem is the same: scalability of these architectures remains questionable as the port count of a single optical switch or AWGR always has its limit. Even if we adopt hierarchical extension, the performance and feasibility of highly dynamic scheduling are still the conundrums hindering practical application.

Zhu et al. [30] introduced an optical DCN architecture which consists of distributed wavelength switches rather than centralized high-port-count switches. The switches are connected into a Torus topology, making the system difficult to be incrementally expanded. Refs. [31–34] are all solutions with elaborately fabricated optical switches enabling system-wide optical packet (or burst) switching. However, the large scale optical packet switching

Please cite this article as: D. Zhang et al., Optical switching based small-world data center network, Computer Communications (2017), http://dx.doi.org/10.1016/j.comcom.2017.01.001