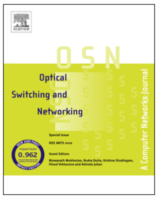
Optical Switching and Networking 16 (2015) 46–51

Contents lists available at ScienceDirect

Optical Switching and Networking

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



A scalable AWG-based data center network for cloud computing

Gang Wu [a,c], Huaxi Gu [a,*], Kun Wang [b], Xiaoshan Yu [a], Yantao Guo [c]

[a] State Key Laboratory of ISN, Xidian University, 710071 Xi'an, China
[b] School of Computer Science and Technology, Xidian University, 710071 Xi'an, China
[c] Science and Technology on Information Transmission and Dissemination in Communication Networks Laboratory, 050000 Shi Jiazhuang, China

ARTICLE INFO

Article history:
Received 9 June 2014
Received in revised form
16 October 2014
Accepted 6 December 2014
Available online 16 December 2014

Keywords:
Data center networks
Arrayed Waveguide Grating
Optical interconnection
Multi-path routing

ABSTRACT

With the development of cloud computing and other online applications, the traffic for data center network (DCN) has increased significantly. Therefore, it is extremely important for DCNs to support more and more servers and provide high scalability, high throughput and low latency. Some current topologies for data centers have such inherent problems as poor scalability, lack of path diversity, cabling complexity, etc. This paper proposes a scalable AWG-based optical interconnection network for data centers, which is called OIT. OIT possesses good scalability and path diversity and benefits from the inherent parallelism and high capacity of WDM and AWG, which makes it a suitable candidate topology for data centers in the cloud computing era. A multi-path routing algorithm is also designed to utilize OIT's parallel links and distribute the load more evenly. The simulation results show that the packet latency and network throughput performance of OIT is better than that of fat tree topology under uniform random distribution or 50%, 80% intra pod traffic distribution and different packet sizes.

1. Introduction

Over the recent years, data centers are facing an exponential increase of the network traffic due to the rise of cloud computing and other emerging online applications. Many of these applications are data-intensive and require high interaction between servers, which imposes greater pressure on the interconnection and communication schemes of the data center [1]. However, traditional DCNs based on tree shaped topology can hardly meet such requirements as the traffic aggregate in the top of the tree and the root switch becomes the bottleneck. Also, its scalability is severely limited by the performance of the root switch.

Some new topologies based on electrical switching have been proposed. Fat tree [2] is a pod based topology which can deliver large bisection bandwidth and has widely been adopted, but it faces the problems of limited scalability and downlink's inflexibility. DCell [3], BCube [4] and MDCube [5] are recently proposed network architectures for modular data centers. DCell is a recursively defined, high network capacity structure with mini-switches to interconnect servers. But due to its structural features, DCell has some inherent defects. The irregular network topology makes it difficult to deploy a cabling solution and the network traffic in DCell is nonuniformly distributed (most traffic is concentrated in the lower level). BCube is another server centric network structure that is built with multiple network ports. However, BCube adopts too many mini-switches, thus makes it difficult for enterprises to build large data centers, and deploying a cabling solution is also

---

[*] Corresponding author. Tel.: +86 13359252600.
E-mail address: hxgu@xidian.edu.cn (H. Gu).

http://dx.doi.org/10.1016/j.osn.2014.12.001
1573-4277/

complex in BCube. MDCube, built recursively with BCube containers, deploys optical fibers to interconnect multiple BCube containers by using the high speed interfaces of COTS switches. But MDCube also has some defects such as large network diameter and complex cabling solution.

To mitigate the effects brought by huge traffic and meet the requirements of cloud computing era, optical interconnects emerged as promising solutions that can provide high bandwidth with reduced power consumption [6,7]. Some schemes are based on Micro-Electro-Mechanical Systems Switches (MEMS switches) such as c-through [8], Helios [9,10], Proteus [11] and OSA [12]. However the reconfiguration of the MEMS switch requires several milliseconds, thus these themes are not appropriate for delay sensitive applications in the cloud computing environment. Another class of schemes is based on an all-optical switching fabric called Arrayed Waveguide Grating (AWG) that has been proven in telecom applications to scale to petabit/second aggregate switching capacity [13–15], such as DOS [16] and LIONS [17]. Nevertheless, AWG suffers from deviation of passband center frequencies and the crosstalk, both of which becomes very large as the port number of AWG increases [18,19]. Considering the factor of the deviation and crosstalk, the current port count of AWG could only reach about 128 [19], thus the scalability of these schemes is severely limited. Some researchers have proposed solutions such as cascading small AWGs to form a large scale switch [19,20], or applying AWG to Clos topology to settle the problem [21,22]. But the complexity and redundant cost of DCN network construction increases significantly as the network size increases.

In this paper, we propose a scalable AWG-based optical network named as OIT (optical interconnect topology). OIT is a cluster based topology which adopts low-radix AWGs and ToR switches to form clusters, and then these clusters are interconnected by multiple WDM fibers to construct a large network. By adopting small AWGs, OIT can still easily scale out to hundreds of thousands of servers. The structural features ensure OIT with good path diversity. A multi-path routing algorithm is also designed according to OIT's multi-path characteristic. Theoretical analyses show that OIT achieves good scalability while keeps network diameter at a constant low value. Simulation results demonstrate that OIT has much saturation bandwidth improvement over fat tree topology under different traffic distributions and packet sizes.

The rest of the paper is organized as follows. Section 2 gives a full description of the addressing, interconnection, scaling pattern and path diversity of OIT. Section 3 provides the communication mechanism of OIT. Section 4 presents the theoretical analysis and performance evaluation of OIT. Finally, Section 5 briefly concludes the paper.

2. The topology of OIT

2.1. The interconnection rules

Fig. 1 gives an overview of OIT(N, m) structure. It composes of $N+1$ clusters and each cluster consists of two layers of AWGs and one layer of server racks. Specially, in one cluster there are $2 \times N$ AWGs and $N \times N$ server racks ($N \geq 2$), and a server rack is made up of a ToR switch and m servers. WDM fibers are adopted to connect different clusters and devices in each cluster. We denote each layer-1 server rack with a 3-tuple [*clusterid*, *layerid*, *rackid*]. The *clusterid* defines the cluster number and takes values from 1 to $N+1$. The *layerid* is defined as 1. The *rackid* represents the server rack number and takes value from 1 to $N \times N$ from left to right. Then we mark each layer-2 AWG with a 3-tuple [*clusterid*, *layerid*, *awgid*]. The *layerid* is defined as 2 and *awgid* takes value from 1 to N from left to right. Similarly, each layer-3 AWG is also

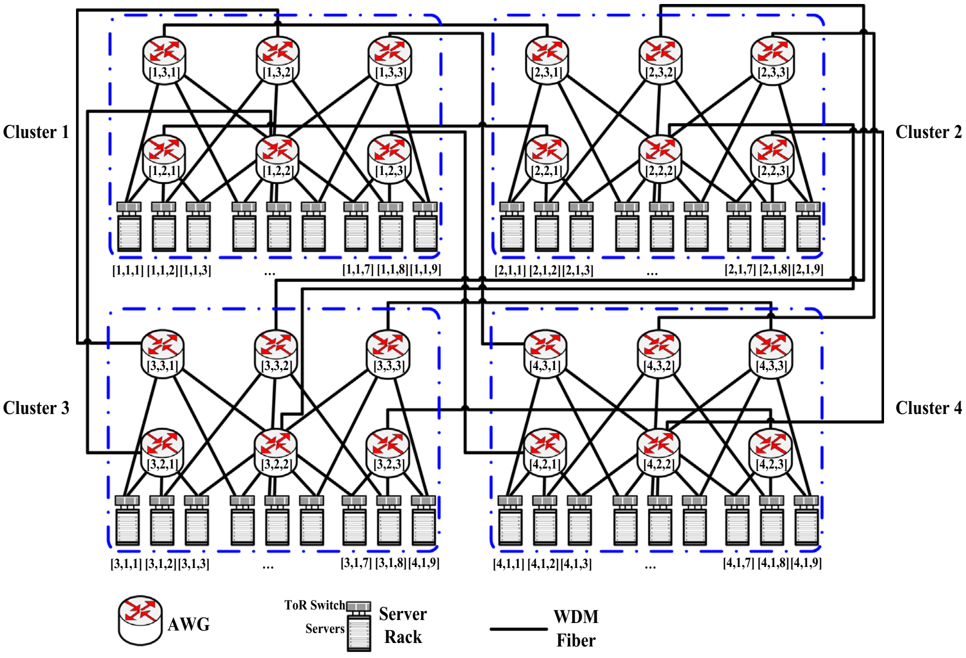


Fig. 1. The OIT($N=3$) topology.