Computer Communications 69 (2015) 40–49

Contents lists available at ScienceDirect

Computer Communications

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

Dynamic resource allocation in hybrid optical–electrical datacenter networks☆

Dinil Mon Divakaran*, Soumya Hegde, Raksha Srinivas, Mohan Gurusamy

Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583, Singapore

ARTICLE INFO

Article history:
Available online 17 July 2015

Keywords:
Datacenter
Bandwidth
Optical
Embedding
Virtual network

ABSTRACT

A promising development in the design of datacenters is the hybrid network architecture consisting of both optical and electrical elements, in which end-to-end traffic can be routed through either an electrical path or an optical path. The core optical switch is used to dynamically create optical paths between pairs of electrical edge-switches in such a datacenter network. In this context, the joint problem of bandwidth allocation and VM-placement poses new and different challenges not addressed yet in hybrid datacenter. In particular, we foresee two issues: (i) the number of edge-switches that can be simultaneously reached using optical paths from an edge-switch is limited by the size of the optical switch, (ii) the dynamic creation of optical paths can potentially establish a constrained optical network topology leading to poor performance. In this work, we abstract the requests of tenants as virtual networks, and study the problem of embedding virtual networks on a hybrid datacenter. We formulate the problem as a non-linear optimization problem and analyze its complexity. We develop and analyse three algorithms for embedding dynamically arriving virtual network demands on a hybrid optical–electrical datacenter. Through simulations, we demonstrate the effectiveness of not only exploiting the already established optical paths, but also of using electrical network in embedding requests of virtual networks.

1. Introduction

Datacenters today are hosting increasing number of services and applications, which in turn generate tremendous amounts of traffic. The global datacenter traffic growth rate is estimated to be approximately 25% per year till 2017; and the annual global datacenter IP traffic is estimated to reach 7.7 zettabytes (10^{21} bytes) by the end of 2017 [2]. The annual traffic growth rate between datacenters as well as within datacenters is also predicted to be ≈ 30%. To meet such traffic growth trends, optical switching based on WDM (wavelength division multiplexing) technology has recently been proposed as a promising approach to connect a datacenter [3–5]. Optical networks provide not only huge bandwidth, but also reduce the cabling complexity and power consumption in comparison to electrical networks.

An important feature of optical networks is its ability to dynamically reconfigure optical paths between any pair of (electrical) switches connected using an optical switch. We can leverage on this capability to solve one important challenge in datacenters—VM (virtual machine) placement problem. As optical paths between edge-switches (top-of-rack switches, to which server machines are connected) can be created on-demand, there is more flexibility in placing VMs of a request, than in an electrical datacenter network. However, building an all-optical datacenter that provides simultaneous connectivity between every pair of edge-switches is expensive and impractical for large datacenters hosting tens of thousands of servers. This, along with the fact that electrical network is better suited for multiplexing short and bursty traffic, makes a hybrid optical–electrical network architecture the right choice for future datacenters [3,4,6]. A hybrid datacenter gives flexibility in connecting edge-switches with high communication demands dynamically using optical network, while maintaining connections between edge-switches with bursty traffic using electrical network.

Fig. 1 illustrates a hybrid optical–electrical datacenter network (similar to Helios [3]) considered in this work. In the hybrid architecture, both electrical and optical networks coexist; traffic from an edge-switch to another can be routed either through the electrical switch(es) or the optical switch. The optical switch in the core connects the electrical edge-switches using optical fibers. The number of optical fibers from an edge-switch is limited by design, and defines the *reachability factor k*. Traffic from a fixed number of ports (on different wavelengths) of an edge-switch is multiplexed into an optical fiber, and switched through the optical switch to any other

☆ This article is an extended version of the paper published in IEEE ICCCN 2014 [1].

* Corresponding author. Tel.: +65 90554094.
E-mail addresses: divakarand@i2r.a-star.edu.sg, dinil.d@gmail.com (D.M. Divakaran), soumya.hegde@hp.com (S. Hegde), raksha.srinivas@finisar.com (R. Srinivas), elegm@nus.edu.sg (M. Gurusamy).

http://dx.doi.org/10.1016/j.comcom.2015.07.008

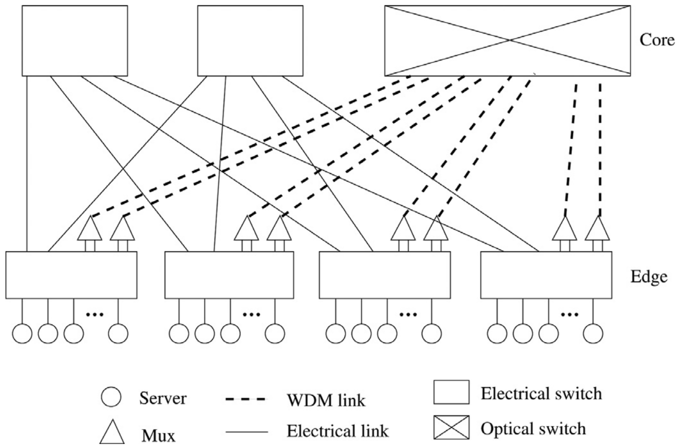

Fig. 1. Illustration of a hybrid optical–electrical datacenter network.

edge-switch, where it is demultiplexed. For a given number of ports at an edge-switch, the value of k determines a trade-off between the size of the optical switch (in number of ports) and the number of edge-switches that can be simultaneously reached. The higher the value of k, the larger the size of the optical switch and more the number of simultaneous optical paths from an edge-switch. In Fig. 1, k is set to two; hence the maximum number of edge-switches that can be reached from any edge-switch using (one hop) optical paths at a given time is limited to two. This is a cost-effective simple optical network which can connect large number of edge-switches using one or a few optical switches. For example, for k equal to four, an optical switch with 400 ports can connect 100 edge-switches. Although an edge-switch can reach only k other edge-switches simultaneously using the optical paths, due to the dynamic reconfiguration capability of the optical switch it can reach different sets of edge-switches at different times. As the architecture is hybrid, the edge-switches can also reach other edge-switches through the electrical core switches (as shown in the figure).

In a datacenter supporting multi-tenancy, guaranteeing bandwidth is important to deliver predictable performance to applications running on the VMs [7]. Recent research works addressed this problem in all-electrical datacenter networks; the corresponding problem in all-optical network is also getting attention [8]. But there are new challenges posed by a hybrid optical–electrical datacenter architecture not yet addressed. While the dynamic creation of optical paths gives flexibility in placing VMs, the reachability factor limits the number of edge-switches that can be reached simultaneously from one edge-switch (using optical paths). Besides, the topology that gets established dynamically may also pose constraints.

We abstract a request from a tenant in the form of a *virtual network*, where a node of a virtual network corresponds to a set of VMs (VM-cluster), and the weight of an edge connecting two nodes gives the bandwidth required between the two corresponding VM-clusters of a tenant. This is a natural abstraction for most applications in datacenters, such as mapreduce, communicating tasks (where each task is carried out by a set of VMs), etc. [9]. We focus on the problem of embedding of virtual networks on hybrid datacenter networks, which translates to the joint problem of bandwidth allocation and placement of VM-clusters such that topology constraints of the virtual network are satisfied.

An intuitive way to solve this problem is to create a new optical path for each edge of the virtual network (until no more can be created), and then explore the existing optical paths and electrical network for embedding the remaining edges. But such an approach can potentially create a constrained optical network topology, which may not suit well for future virtual network demands. To investigate this, we develop an algorithm based on the above approach called NLFE (New-Link-First Embedding). In addition, we develop another embedding algorithm called ELFE (Existing-Link-First Embedding) that embeds edges on the existing optical paths and electrical network, and only in the worst case will create new optical paths. We also develop a greedy algorithm, called GLE (Greedy Link Embedding), that is faster than the above two algorithms in selecting the edges of an input virtual network to be mapped on to the hybrid datacenter network. We define a control parameter for flexible control of the proportion of edges that can be embedded on the electrical network of a datacenter. Using simulations, we evaluate these algorithms with different topologies for dynamically arriving virtual networks. Our results provide interesting insights. One, ELFE and GLE algorithms consistently outperform NLFE, demonstrating the effectiveness of exploiting existing optical paths. Two, partial (but limited) embedding of edges on electrical links can decrease the rejection ratio further while still increasing the utilization of optical network capacity. Finally, the greedy algorithm GLE performs as good as ELFE, reinforcing that the restricting factor in embedding virtual networks is the topology constraints and not the capacity.

We carried out preliminary work and developed the NLFE and ELFE heuristics in [1]. In this article we extend this work further and make new contributions. We explicitly formulate the problem of embedding a virtual network on a hybrid datacenter network, satisfying the constraints in both the input and the datacenter, as an optimization problem, and present its complexity. We develop a new algorithm based on greedy approach called Greedy Link Embedding (GLE). Algorithm GLE is computationally faster than NLFE and ELFE algorithms and differs in the way edges in an input virtual network are selected for mapping to the hybrid optical–electrical network.

After discussing the related works in Section 2, we define and formulate the problem in Section 3. The operations involved in solving the problem of embedding a virtual network are described in Section 4, where we also define the three embedding algorithms—NLFE, ELFE and GLE. Performance studies are carried out in Section 5.

2. Related works

2.1. Bandwidth allocation in all-electrical datacenter networks

The importance of allocating bandwidth for requests in all-electrical datacenter was recently highlighted by the research community. In this direction, Seawall enforces link-bandwidth allocation to competing VMs based on the weights assigned to VMs [10]; whereas Gatekeeper achieves bandwidth-sharing among competing tenants [11]. Other works consider advance reservation of bandwidth, given input demands. SecondNet [12] is a solution for datacenter network virtualization. It assumes a matrix specifying bandwidth demands between every VM pair as part of the input. The proposed embedding of a tenant request proceeds by first locating a cluster based on the server resource requirement, and then solving the matching problem of a bipartite graph formed of VMs and servers based on both the server resource and bandwidth requirements. Oktopus [7], a similar work that allocates both server and network resources, abstracted bandwidth demands between VMs of a tenant into two topologies. While one topology was a single cluster of VMs requiring a star communication pattern, another topology abstracted the communication demand of clusters of VMs (or VM-clusters) as a tree. Our previous work [13,14] proposes a solution for the joint problem of bandwidth allocation and VM-placement in all-electrical network by considering a matrix of bandwidth demands between every VM pairs as input. But assuming the bottleneck to be in the core network connecting the switches of a datacenter, the results showed that, for efficient allocation VMs can be grouped to form small number of VM-clusters (often four, and not more than six), and bandwidth allocation on these core links is dictated only by the traffic demands between VM-clusters [14].