Computer Networks 113 (2017) 29–45

Contents lists available at ScienceDirect

Computer Networks

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

The stellar transformation: From interconnection networks to datacenter networks

Alejandro Erickson[a,*], Iain A. Stewart[a], Javier Navaridas[b], Abbas E. Kiasari[b]

[a] School of Engineering and Computing Sciences, Durham University, Science Labs, South Road, Durham DH1 3LE, UK
[b] School of Computer Science, University of Manchester, Oxford Road, Manchester M13 9PL, UK

ARTICLE INFO

Article history:
Received 26 June 2016
Revised 5 November 2016
Accepted 2 December 2016
Available online 5 December 2016

Keywords:
Dual-port server-centric datacenter networks
Fault tolerance
Generalized hypercubes
Interconnection networks
Performance evaluation
Routing
Topological properties

ABSTRACT

The first dual-port server-centric datacenter network, FiConn, was introduced in 2009 and there are several others now in existence; however, the pool of topologies to choose from remains small. We propose a new generic construction, the stellar transformation, that dramatically increases the size of this pool by facilitating the transformation of well-studied topologies from interconnection networks, along with their networking properties and routing algorithms, into viable dual-port server-centric datacenter network topologies. We demonstrate that under our transformation, numerous interconnection networks yield datacenter network topologies with potentially good, and easily computable, baseline properties. We instantiate our construction so as to apply it to generalized hypercubes and obtain the datacenter networks GQ*. Our construction automatically yields routing algorithms for GQ* and we empirically compare GQ* (and its routing algorithms) with the established datacenter networks FiConn and DPillar (and their routing algorithms); this comparison is with respect to network throughput, latency, load balancing, fault-tolerance, and cost to build, and is with regard to all-to-all, many all-to-all, butterfly, random, hot-region, and hot-spot traffic patterns. We find that GQ* outperforms both FiConn and DPillar (sometimes significantly so) and that there is substantial scope for our stellar transformation to yield new dual-port server-centric datacenter networks that are a considerable improvement on existing ones.

1. Introduction

The digital economy has taken the world by storm and completely changed the way we interact, communicate, collaborate, and search for information. The main driver of this change has been the rapid penetration of cloud computing which has enabled a wide variety of digital services, such as web search and online gaming, by offering elastic, on-demand computing resources to digital service providers. Indeed, the value of the global cloud computing market is estimated to be in excess of $100 billion [46]. Vital to this ecosystem of digital services is an underlying computing infrastructure based primarily in datacenters [5]. With this sudden move to the cloud, the demand for increasingly large datacenters is growing rapidly [20].

This demand has prompted a move away from traditional datacenter designs, based on expensive high-density enterprise-level switches, towards using commodity-off-the-shelf (COTS) hardware. In their production datacenters, major operators have primarily adopted (and invented) ideas similar to Fat-Tree [3], Portland [36], and VL2 [18]; on the other hand, the research community (several major operators included) maintains a diverse economy of datacenter architectures and designs in order to meet future demand [16,20,22,33,39,42]. Indeed, the “switch-centric” datacenters currently used in production datacenters have inherent scalability limitations and are by no means a low-cost solution (see, *e.g.*, [8,20,21,32]).

One approach intended to help overcome these limitations is the “server-centric” architecture, the first examples of which are DCell [20] and BCube [19]. Whereas in a switch-centric datacenter network (DCN) there are no links joining pairs of servers, in a server-centric DCN there are no links joining pairs of switches. This server-centric restriction arises from the circumstance that the switches in a server-centric DCN act only as non-blocking “dumb” crossbars. By offloading the task of routing packets to the servers, the server-centric architecture leverages the typically low utilisation of CPUs in datacenters to manage network communication. This can reduce the number of switches used in a DCN, the capabilities required of them, and their cost. In particular, the switches route only locally, to their neighbouring servers, and therefore have

---

[*] Corresponding author at: 3-864 Swan St., Saanich, BC, V8X 2Z3, Canada, Tel.: +1 778 350 4976.

E-mail address: alejandro.erickson@gmail.com (A. Erickson).

http://dx.doi.org/10.1016/j.comnet.2016.12.001

no need for large or fast routing tables. Thus, a server-centric DCN can potentially incorporate more servers and be both cheaper to operate and to build (see [37] for a more detailed discussion). Furthermore, using servers (which are highly programmable) rather than switches (which have proprietary software and limited programmability) to route packets will potentially accelerate research innovation [30]. Of course, the server-centric approach is not a panacea as packet latency can increase, with the need to handle routing imposing a computational overhead on the server.

The server-centric paradigm is currently an area of intensive study with numerous new server-centric DCNs having been proposed and scrutinized, although there is still much to be done before server-centric DCNs make it through to production. Since the advent of DCell and BCube, various server-centric DCNs have been proposed, some of which further restrict themselves to requiring at most two ports per server, with FiConn [28] and DPillar [31] being the most established of this genre. This dual-port restriction is motivated by the fact that many COTS servers presently available for purchase, as well as servers in existing datacenters, have two NIC ports (a primary and a backup port). Dual-port server-centric DCNs are able to utilise such servers without modification, thus making it possible to use some of the more basic equipment (available for purchase or from existing datacenters) in a server-centric DCN and thereby reduce the building costs.

The server-centric DCN architecture provides a versatile design space, as regards the network topology, evidenced perhaps by the sheer number of fairly natural constructions proposed from 2008 to the present. On the other hand, this pool is small relative to the number of interconnection networks found in the literature, *i.e.*, highly structured graphs with good networking properties. One of the challenges of identifying an interconnection network suitable for conversion to a DCN topology, however, lies in the fact that the literature on interconnection networks is focused primarily on graphs whose nodes are homogeneous[1], whereas in both a switch-centric and a server-centric DCN we have server-nodes and switch-nodes which have entirely different operational roles. Some server-centric DCN topologies arise largely from outside the interconnection network literature, *e.g.*, DCell and FiConn, whilst others arise from transformations of well-known interconnection networks, *e.g.*, BCube and DPillar.

The transformations used to obtain BCube and DPillar take advantage of certain sub-structures in the underlying base graphs of the interconnection networks in question (generalized hypercubes and wrapped butterfly networks, respectively) in order to create a server-centric DCN that inherits beneficial networking properties such as having a low diameter and fault-tolerant routing algorithms. The limitation, of course, is that not every prospective base graph has the required sub-structures (cliques and bicliques, respectively, in the cases of BCube and DPillar). New methods of transforming interconnection networks into server-centric DCNs may therefore greatly enlarge the server-centric DCN design space by lowering the structural requirements on potential base graphs.

It is with the construction of new dual-port server-centric DCNs that we are concerned in this paper. In particular, we provide a generic methodology to systematically transform interconnection networks, as base graphs, into dual-port server-centric DCNs, which we refer to as *stellar* DCNs. The stellar transformation is very simple and widely applicable: the edges of the base graph are replaced with paths of length 3 involving two server-nodes each, and the nodes of the base graph become the switch-nodes of the stellar DCN (see Fig. 3). By requiring very little of the base graph in the way of structure, the stellar construction greatly increases the pool of interconnection networks that can potentially serve as blueprints to design dual-port server-centric DCN topologies.

We validate our generic construction in three ways: first, we prove that various networking properties of the base graph are preserved under the stellar transformation; second, we build a library of interconnection networks that suit the stellar transformation; and third, we empirically evaluate GQ*, an instantiation of a stellar DCN whose base graph is a generalized hypercube, against both FiConn and DPillar, and we also compare GQ* and its routing algorithm (inherited from generalized hypercubes) against what might be optimally possible in GQ*. This latter validation demonstrates that not only does our methodology allow us to transport properties from interconnection networks to dual-port DCNs in general, but also that a specific application of it yields a very competitive dual-port DCN in comparison with other well-established dual-port DCNs.

Our empirical results are extremely encouraging. We employ a comprehensive set of performance metrics so as to evaluate network throughput, latency, load balancing capability, fault-tolerance, and cost to build, within the context of all-to-all, many all-to-all, butterfly, random, hot-region, and hot-spot traffic patterns, and we show that GQ* broadly outperforms both FiConn and DPillar as regards these metrics, sometimes significantly so. Highlights of these improvements are as follows. In terms of aggregate bottleneck throughput (a primary metric as regards the evaluation of throughput in an all-to-all context), our DCN GQ* improves upon both FiConn and DPillar (upon the former markedly so). As regards fault-tolerance, our DCN GQ*, with its fault-tolerant routing algorithm *GQ*-routing* (inherited from generalized hypercubes), outperforms DPillar (and its fault-tolerant routing algorithm *DPillarMP* from [31]) and competes with FiConn even when we simulate optimal fault-tolerant routing in FiConn (even though such a fault-tolerant routing algorithm has yet to be exhibited). Not only does *GQ*-routing* (in GQ*) tolerate faults better than the respective routing algorithms in FiConn and DPillar, but when we make around 10% of the links faulty and compare it with the optimal scenario in GQ*, *GQ*-routing* provides around 95% connectivity and generates paths that are, on average, only around 10% longer than the shortest available paths. When we consider load balancing in GQ*, FiConn, and DPillar, with their respective routing algorithms *GQ*-routing, TOR*, and *DPillarSP* and under a variety of traffic patterns, we find that the situation in GQ* is generally improved over that in FiConn and DPillar. As we shall see, DPillar performs particularly poorly except as regards the butterfly and hot-region traffic patterns; indeed, for the hot-region traffic pattern, it performs best. The improved load balancing in GQ* in tandem with the generation of relatively short paths translates to potential latency savings.

However, we have only scratched the surface in terms of what might be possible as regards the translation of high-performance interconnection networks into dual-port server-centric DCNs in that we have applied our generic, stellar construction to only one family of interconnection networks so as to achieve encouraging results. In addition to our experiments, we demonstrate that there are numerous families of interconnection networks to which our construction might be applied. Whilst our results with generalized hypercubes are extremely positive, we feel that the generic nature of our construction has significant potential and scope for further application.

To summarise, the contributions of this paper are as follows:

- We propose the star-replaced server-centric DCN construction as a generic methodology in order to automatically convert graphs and interconnection networks into ‘stellar’ dual-port server-centric DCNs;
- We demonstrate how the properties of the base graph or interconnection network translate so that similar properties are

---

[1] We disregard the terminal nodes of indirect networks, which are not intrinsic to the topology.