ARTICLE IN PRESS
YJCSS:3105

Journal of Computer and System Sciences ••• (••••) •••–•••

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

On the combinatorial design of data centre network topologies ☆,☆☆

Iain A. Stewart

School of Engineering and Computing Sciences, Durham University, Science Labs, South Road, Durham DH1 3LE, UK

ARTICLE INFO

Article history:
Received 28 April 2016
Received in revised form 26 January 2017
Accepted 29 May 2017
Available online xxxx

Keywords:
Data centre networks
Switch-centric data centre networks
Fat-Trees
Combinatorial designs
Bipartite graphs
Path diversity

ABSTRACT

The theory of combinatorial designs has recently been used in order to build switch-centric data centre networks incorporating a large number of servers, in comparison with the popular Fat-Tree data centre network. We clarify and extend these results and prove that in these data centre networks: there are pairwise link-disjoint paths joining all the servers adjacent to some switch with all the servers adjacent to any other switch; and there are pairwise link-disjoint paths from all the servers adjacent to some switch to any identically-sized collection of target servers where these target servers need not be adjacent to the same switch. In both cases, we always control the path lengths. Our constructions and analysis are undertaken on bipartite graphs with the applications to data centre networks being easily derived. Our results show the potential of the application of results and methodologies from combinatorics to data centre network design.

1. Introduction

1.1. The data centre network context

Data centres are expanding both in terms of their size and their importance as computational platforms for cloud computing, web search, social networking, and so on. There is an increasing demand that data centres incorporate more and more servers but so that overall computational efficiency is not compromised through excessive traffic. A key factor as to the eventual performance of a data centre is the *data centre network* (*DCN*); that is, the interconnection fabric of the servers and switches within the data centre. As we strive to incorporate more and more servers, new topologies are being developed so as to cope with the increase in scale and best utilize the additional computational power. It is with topological aspects of DCNs that we are concerned in this paper.

The traditional design of a DCN is *switch-centric* so that the routing intelligence resides amongst the switches, with the servers behaving only as computational nodes. In switch-centric DCNs, there are no direct server-to-server links; only server-to-switch and switch-to-switch links. Switch-centric DCNs are traditionally tree-like with servers located at the 'leaves' of the tree-like structure. Examples include ElasticTree [1], VL2 [2], HyperX [3], Portland [4], and Flattened Butterfly [5], although the dominating switch-centric DCN is Fat-Tree [6]. Whilst it is generally acknowledged that tree-like, switch-centric

☆ This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant 'Interconnection Networks: Practice unites with Theory (INPUT)' [grant number EP/K015680/1].

☆☆ A preliminary version of this paper appeared as an extended abstract in the *Proceedings of 20th International Symposium on Fundamentals of Computation Theory* (A. Kosowski, I. Walukiewicz, eds.), Gdansk, Poland, August 17–19 2015, Lecture Notes in Computer Science, Volume 9210, Springer, 2015, 283–295.

E-mail address: i.a.stewart@durham.ac.uk.

http://dx.doi.org/10.1016/j.jcss.2017.05.015

DCNs have their limitations when it comes to, for example, scalability, due to the size of routing tables at the switches, switch-centric DCNs remain popular and can usually be constructed from commodity hardware. A more recent paradigm, namely the *server-centric* DCN, has emerged so that deficiencies of the tree-like, switch-centric DCNs might be ameliorated. Server-centric DCNs reflect that the routing intelligence resides within the servers with switches operating only as dumb crossbars. In server-centric DCNs there are only server-to-switch and server-to-server links. However, server-centric DCNs also suffer from deficiencies such as packet relay overheads caused by the need to route packets within the server; moreover, server-centric DCNs have yet to make it into the commercial mainstream (the reader is referred to [7] for an overview of the state of the art as regards DCN architectural design). It is with the construction of switch-centric DCNs that we are concerned here.

It is extremely difficult to design computationally efficient (switch-centric) DCNs so as to incorporate large numbers of servers as there are many additional considerations to take into account. For example, switches and (especially) servers in data centres have a limited number of ports with a consequence being that the more servers there are, the greater the average or worst-case link-count between two distinct servers; hence, there is a packet latency overhead to be borne. Also, so as to better support routing, fault-tolerance, and load-balancing, we would prefer that there are numerous alternative paths within the DCN joining any two distinct servers; that is, that there is *path diversity*. Irrespective of the DCN paradigm within which one works, there are many other design parameters to bear in mind relating to, for example, (incremental) scalability, throughput, cost, oversubscription, energy consumption, latency, and security (see, for example, [8,9] for an overview). The upshot is that the DCN designer has to simultaneously secure a number of performance characteristics, some of which are competing against each other; this makes the DCN design space complex and difficult to work in.

1.2. Using combinatorial designs to build DCNs

A recent proposal in [10] advocated the use of *combinatorial design theory* in order to design switch-centric DCNs; these DCNs have beneficial properties as regards incorporating more servers and possessing path diversity yet it is possible to limit the worst-case link-length of server-to-server shortest paths (and so, ultimately, achieve better control over packet latency in a DCN). The use of combinatorial designs within the study of general interconnection networks is not new and originated in [11] where the targeted networks involved processors communicating via buses (the reader is referred to [12] for a range of applications of combinatorial design theory within computer science). A hypergraph framework was developed in [11] where the hypergraph nodes represent the processors and the hyperedges the buses. Likewise, an analogous framework was developed in [10] but where the hypergraph nodes and edges both represent switches so that the pendant servers 'hang off' some of the switches (we present a detailed description of this framework in Section 3.3). In [10], the ubiquitous switch-centric Fat-Tree DCN from [6] was used as a yardstick against which to compare the new DCN designs developed in [10] under the normalization that all DCNs are to have the same worst-case link-length of server-to-server shortest paths, namely 6, as this equals the worst-case link-length of server-to-server shortest paths in the Fat-Tree DCN. It was shown that more servers can be incorporated within the new DCNs yet, crucially, the resulting DCNs have good path diversity. It is the algebraic properties (relating to symmetry and balance) possessed by transversal designs that enable the constructions and analysis as described in [10]. One slight difficulty with the original and novel approach taken in [10] is that some of the path diversity results derived there are incorrect (as we explain later in Section 4.1). Not only has combinatorial design theory featured as regards the design of interconnection networks but other aspects of algebra have too; indeed, there has been recent work on the relevance of Cayley graphs, Hamming graphs, and hyperbolicity to DCN design (see, e.g., [13–15]).

1.3. Our contribution

In this paper we return to the framework of [10] and formulate and prove path diversity results for the switch-centric DCNs constructed using the methods of that paper. As our concern is entirely with topological properties of DCNs, henceforth we abstract our DCNs as undirected graphs where the nodes are to represent servers and switches and the edges point-to-point links. The crux of the construction in [10] is (essentially) to build a bipartite graph using a systematic method, called the 3-step method, involving a different 'base' bipartite graph and a transversal design, and to convert the resulting bipartite graph into switch-centric DCNs (in a variety of ways). After explaining how hypergraphs and transversal designs can all be considered as bipartite graphs in Section 2, in Section 3 we provide a detailed description of the 3-step framework from [10] and explain how the bipartite graphs constructed are converted into switch-centric DCNs. Next, we revisit the results from [10]. In particular, in Section 4 we correct and extend the analysis in [10] and affirm that using the 3-step method from [10], we can build switch-centric DCNs: with many more servers than the Fat-Tree DCN yet so that, like the Fat-Tree, every server-to-server shortest path has length at most 6; and so that (assuming some numeric conditions on the base bipartite graph and the transversal design) we can find pairwise link-disjoint paths from all of the servers adjacent to a particular switch to all of the servers adjacent to any other switch. Moreover, we provide an upper bound on the lengths of the paths constructed in terms of the diameter of the base bipartite graph (see Theorem 4). We also deal with a scenario missing from [10] (see part (*b*) of Theorem 4). As we explain, the general situation is more subtle than was assumed in [10].

The DCN path diversity, as we have described it above, comes about from building bipartite graphs (which are subsequently converted to DCNs) so that given any two distinct nodes, there are numerous node-disjoint paths joining these two

Please cite this article in press as: I.A. Stewart, On the combinatorial design of data centre network topologies, J. Comput. Syst. Sci. (2017), http://dx.doi.org/10.1016/j.jcss.2017.05.015