Network performance trade-Off in optical spatial division multiplexing data centers

Network Performance Trade-Off in Optical Spatial Division Multiplexing Data Centers

Li Yan[1], Matteo Fiorani[2], Ajmal Muhammad[2], Massimo Tornatore[3], Erik Agrell[1], Lena Wosinska[2]

[1]*Department of Signals and Systems, Chalmers University of Technology, 41296 Gothenburg, Sweden,*
[2]*Optical Networks Lab, KTH Royal Institute of Technology, 16440 Kista, Sweden,*
[3]*Department of Electronics, Information, and Bioengineering, Politecnico di Milano, 20133 Milano, Italy*

{*lyaa,agrell*}*@chalmers.se,* {*matteof,ajmalmu,wosinska*}*@kth.se, massimo.tornatore@polimi.it*

Abstract: We propose close-to-optimal network resource allocation algorithms for modular data centers using optical spatial division multiplexing. A trade-off between the number of established connections and throughput is identified and quantified.

OCIS codes: (060.1155) All-optical networks; (060.4256) Networks, network optimization.

1. Introduction

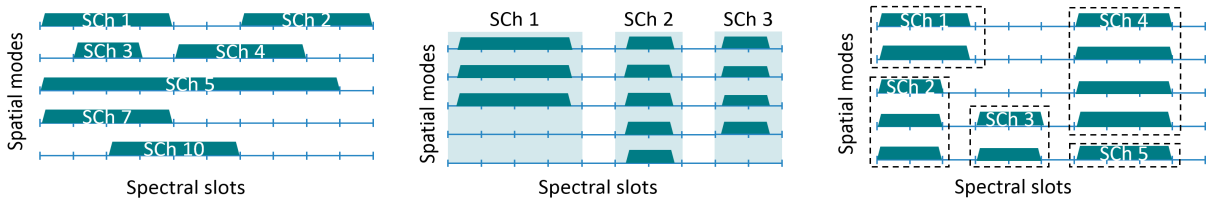
Motivated by the growing popularity of cloud services, large data centers (DCs) are expected to host hundreds of thousands of servers in the near future [1]. Modular DCs [2] based on prefabricated stand-alone modules, referred to as PODs, which are composed of a predefined set of compute, storage, and network resources, are considered as an efficient approach to build large DC facilities. PODs can generate a large amount of traffic and thus require an ultra-high-capacity interconnection network.

Spatial division multiplexing (SDM) enabled by multi-core, multi-mode, or multi-element fibers has recently emerged as a cost-effective and energy-efficient solution for DC networks [3], offering high scalability. Different options to incorporate SDM in optical networks have been proposed [4] and investigated for designing modular DC architectures [5]. These architectures combine SDM with flexible WDM to provide the required capacity for modular DCs. To fully exploit the ultra-high capacity offered by SDM, effective resource allocation algorithms are necessary [5]. In traditional optical networks, a resource allocation strategy that achieves a low connection blocking probability usually leads to a high network throughput as well. This is the reason why most of the recently proposed algorithms for SDM networks mainly focus on optimizing one of the two objectives [6, 7], and do not explicitly improve the other one. However, considering the diverse traffic patterns in DCs, where low-bandwidth mice flows and high-bandwidth elephant flows coexist, the maximization of the number of established connections may inflict high blocking probability on bandwidth-intensive connections and, thus, reducing the network throughput. On the other hand, optimizing the throughput without paying attention to the number of connections might lead to blocking a large number of mice flows. Therefore, a balance between the number of connections and throughput should be achieved to guarantee fairness and efficiency in DC networks.

In this paper, we propose effective heuristics to calculate close-to-optimal resource allocations for three SDM-based modular DC architectures. Both the number of connections and throughput are optimized simultaneously up to a predefined priority. Simulation results indicate that the balance between these two metrics must be chosen carefully to optimize the overall performance and fairness of the network.

2. SDM Schemes and Network Topology

Three SDM-based modular DC network architectures [5] are considered in this paper. The first architecture (A1) is the *uncoupled SDM with flexgrid WDM* as shown in Fig. (1a), where each spatial element operates as an independent flexgrid WDM fiber and multiple independent spectral superchannels can be established. The second architecture (A2) is the *coupled SDM with spectral flexibility*, which is illustrated in Fig. (1b). In this SDM architecture, spectral superchannels are expanded to all the spatial elements to create spectral–spatial superchannels with increased capacity. The third architecture (A3) is the *coupled SDM with spectral and spatial flexibility* as displayed in Fig. (1c), where the unrestricted flexibility in both spectral and spatial domains are exploited to form flexible spectral–spatial superchannels. A simple modular DC network topology [5] is studied in this paper, where the PODs are interconnected through a single optical large port count (LPC) SDM switch. Each POD is connected to the LPC switch with a single bidirectional fiber that supports N spatial elements and M spectral slots per spatial element. Furthermore, thanks to the inherent flexibility of SDM switches, optical superchannels can use different spatial elements at the input and output fiber links to the switch [3, 5].



(a) A1: Uncoupled SDM and flexgrid WDM. (b) A2: Coupled SDM with spectral flexibility. (c) A3: Coupled SDM with spectral–spatial flexibility.

Figure 1: Different architectures for combining SDM and flexgrid WDM.

3. Resource Allocation Algorithms

3.1. Optimization Objective

The objective is to achieve the optimal trade-off between blocking probability and throughput, by assigning appropriate spatial and spectral resources to each feasible traffic demand. In the optimization objective, we linearly combine the number of established connections C and throughput T as $C+\beta T/t_{\text{ave}}$, where β is a weighting factor controlling the priority of T relative to C, and t_{ave} is a normalizing factor and is equal to the average data rate per connection.

3.2. Proposed Algorithms

Mixed integer linear programs (MILPs) are formulated for all the SDM architectures, which are the modified versions of the routing, spectrum, and core allocation (RSCA) problem [8]. In the following, we briefly describe the MILPs for A1, A2 and A3 without providing the mathematical details due to space limitations. In A1, each feasible traffic demand is assigned a set of spectral slots satisfying its data rate requirement and one spatial element on the input and output links to the SDM switch, respectively. The spectral continuity and contiguity constraints [8] are imposed. A2 can be viewed as a special case of A1 where the fiber has a single spatial element with N times more capacity per spectral slot than A1. In A3, due to the extra flexibility provided by spectral–spatial superchannels, in addition to the constraints in A1, we also have constraints assuring contiguity and nonoverlapping in the spatial domain.

Given the high complexity of the MILPs, we also developed low-complexity heuristics for A1, A2, and A3. The heuristics decompose the MILPs in two subproblems and tackle them separately. First, the spatial element assignment is carried out by relaxing the spectral continuity constraint in the MILP for each architecture. The resulted solution is a set of traffic demands that can be potentially served, each of which is assigned spatial elements on the input and output fiber links to the SDM switch. The optimal spatial element assignment provides an upper bound for the resource allocation problem, which will be used in the numerical analysis to evaluate the optimality of the proposed heuristics. Secondly, the spectral slots are assigned to the potentially feasible traffic demands, by using an ensemble of first-fit (FF) algorithms [4, 6–9], where a variety of FFs (e.g., SpeF/SpaF, ascending/descending FF, DPH, and SPSA) are computed and the best one is chosen. The reason for trying several FFs is because the weighting factor β in the objective is adjustable and, thus, a single FF is not sufficient to generate good results in all circumstances. Consequently, by adding many computationally efficient FFs into the ensemble, the robustness of the spectral element assignment increases at the cost of little complexity.

4. Numerical Analysis

We assume that the traffic pattern in the modular DC varies slowly with time [5] and that the resources are allocated periodically at predefined intervals. The number of traffic demands generated per POD is a random integer uniformly distributed in the range $[0, \lfloor L(N_p-1)\rfloor]$, where N_p is the number of PODs in the DC, $L=0.2$ is the network load, and $\lfloor x\rfloor$ is the floor function. The data rates of traffic demands are random numbers in the discrete set $[1,10,100,200,400,1000]$ Gbps with the probabilities $[0.16,0.01,0.14,0.19,0.34,0.16]$, respectively [5]. The physical layer impairments are not considered. Dual-polarization quadrature phase shift keying is used for all traffic demands. The bandwidth of a spectral slot is 12.5 GHz. Each SDM fiber has $N=3$ spatial elements and $M=80$ spectral slots per spatial element.

We first investigate the trade-off between the two objectives by using the proposed heuristics. The number of PODs is set to $N_p=100$. The value of β is varied from 0 to 2000 gradually and the resulted Pareto curves are plotted on the connection–throughput plan for all the considered architectures in Fig. (2a)–(2c). The Pareto curves reveal an interesting trade-off between the number of established connections and throughput. When $\beta=0$, the number of connections is maximized with a comparatively low throughput. As β begins to increase, the throughput increases significantly at the expense of slightly growing connection blockage. If β continues to increase after the "elbow points", which are shown as green circles in Fig. (2a)–(2c), the number of connections degenerates dramatically

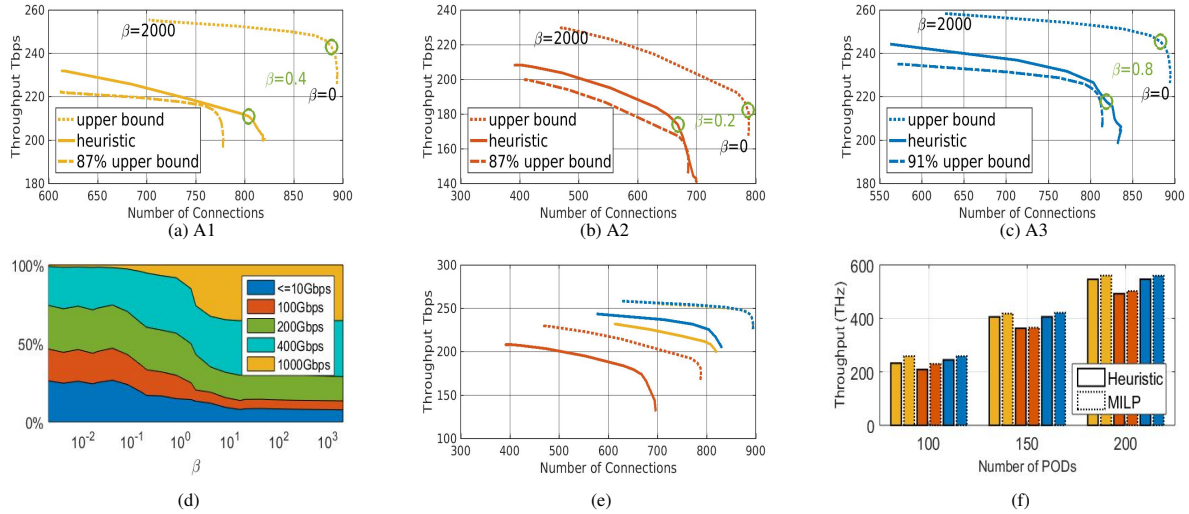


Figure 2: (a)–(c): The dotted lines are upper bounds calculated by the spatial element assignment, the solid lines are given by the spectral element assignment, and the dash-dotted lines are the maximum fractions of the upper bounds that can lower-bound the heuristic curves. (d): The percentages of allocated number of connectionss for different traffic classes as functions of β. (e): Pareto curves for all the architectures, their colors correspond to those in (a)–(c), respectively. The upper bound of A1 is not visible because it is identical to the upper bound of A3 and completely covered. (e): The results of throughput optimizations.

with relatively little throughput improvement. Consequently, at the "elbow points", the best balance between the two objectives is achieved. Note that different architectures achieve the "elbow points" with different β values. The larger value of β, the better the architecture can improve the throughput without affecting the number of connections severely. A3 is the most flexible architecture and for this reason reaches the "elbow point" with the highest value of β ($\beta = 0.8$), while A2 is the least flexible one and reaches the "elbow point" with the lowest value of β ($\beta = 0.2$). As illustrated in Fig. (2d), by increasing β, the low-bandwidth connections (less than 200 Gbps) gradually make room for high-bandwidth demands, which lead to the increase in throughput. In Fig. (2e), the Pareto curves for different architectures are shown together. The upper bounds of A1 and A3 overlap and their heuristic curves are close (the gap between them is approximately 4% of their upper bound), indicating that the extra flexibility in A3 only brings marginal benefit with respect to A2 [5, 7]. The Pareto curves have large horizontal gaps and relatively small vertical distances between each other, implying homogeneous throughput performances among the architectures. The trend is also verified in Fig. (2f), where pure throughput optimizations are carried out for different DC sizes by both the MILPs and the heuristics. As illustrated, all the architectures have similar performances, with A1 and A3 outperforming A2 by approximately 10%, and the gap decreases as the DC sizes grow.

5. Conclusion

This paper addresses the optimal resource allocations in three SDM network architectures for modular DCs. A trade-off between the number of established connections and throughput is revealed and studied by proposing both optimal MILP approaches and low-complexity heuristics. Results show that, if the priority is chosen carefully, the proposed heuristics can achieve close-to-optimal trade-off between the two important network performance metrics. We also show that, even if A2 has lower flexibility than A1 and A3, it provides similar performance when the objective is only to maximize the throughput.

References

1. L. Dittmann *et al.*, "A roadmap for evolving towards optical intra-data-center networks," in *Proc. ECOC*, M.2.F.1, Düsseldorf, Germany (2016).
2. J. Hamilton, "Architecture for modular data centers," *arXiv:cs/0612110 [cs.DB]* (2006).
3. Y. Liu *et al.*, "Comparison of SDM and WDM on direct and indirect optical data center networks," in *Proc. ECOC*, M.1.F.2, Düsseldorf, Germany (2016).
4. D. Klonidis *et al.*, "Spectrally and spatially flexible optical network planning and operations," *IEEE Commun. Mag.*, vol. 53, no. 2, pp. 69-78 (2015).
5. M. Fiorani *et al.*, "Optical spatial division multiplexing for ultra-high-capacity modular data centers," in *Proc. OFC*, Tu2H.2, Anaheim, CA (2016).
6. S. Fujii *et al.*, "On-demand spectrum and core allocation for reducing crosstalk in multicore fibers in elastic optical networks," *J. Opt. Commun. Netw.*, vol. 6, no. 12, pp. 1059-1071 (2014).
7. P. S. Khodashenas *et al.*, "Comparison of spectral and spatial super-channel allocation schemes for SDM networks," *J. Lightw. Technol.*, vol. 34, no. 11, pp. 2710-2716 (2016).
8. A. Muhammad *et al.*, "Routing, spectrum and core allocation in flexgrid SDM networks with multi-core fibers," in *Proc. ONDM*, pp. 192-197, Stockholm, Sweden (2014).
9. S. Shirazipourazad *et al.*, "On routing and spectrum allocation in spectrum-sliced optical networks," in *Proc. INFOCOM*, OW3A-5, Turin, Italy (2013).