New Trend of Open and Disaggregated Optical Networks

Liang Dou[1], Sai Chen[2], Huan Zhang[2], Jingchi Cheng[2], Fan Gao[2], Boyuan Yan[2], Shuai Zhang[1], Zhao Sun[1], Lei Wang[1], and Chongjin Xie[3]

[1]Alibaba Cloud, Alibaba Group, Beijing, China, [2]Alibaba Cloud, Alibaba Group, Hangzhou, China,
[3]Alibaba Cloud, Alibaba Group, Sunnyvale, USA
Email: douliang.dl@alibaba-inc.com

Abstract: Open and disaggregated systems supporting flexgrid operation and ROADM are being introduced to data center interconnect networks. System integration and network automation are effective to further increase the network efficiency in the future. © 2022 The Author(s)

OCIS codes: (060.2330) Fiber Optics Communications;

Introduction

With growing demands for internet services and computing power, data centers have already become part of fundamental infrastructures of our society. As intra- and inter-data center networks carry most of the internet traffic [1], demands around data centers become one major driving force to optical network technology. Disaggregation in optical transmission systems originates from data center interconnection (DCI) scenario, aiming at breaking out vendor lock-in and encouraging more competition both in cost and performance, and now it is widely deployed in many hyperscalers' networks [2]. Enabled by standard north bound interface protocols and open data models [3][4], devices from different vendors can be managed by the third-party network management system (NMS), which may be provided by some vendors or developed by operators themselves. Among the reported cases, an optical line system (OLS) usually belongs to one vendor while transponder pairs come from several other vendors. This partially disaggregation form in hardware is a tradeoff between integration difficulties and benefits obtaining from decoupling transponders from OLS [5]. In this case, operators are able to enjoy the benefits brought by relatively fast evolving speed of optical transponders.

In this paper, we discuss the recent DCI technologies developed in data center networks. First, fixed grid wavelength-division-multiplexing (WDM) systems are evolving to flexible grid (flexgrid) due to the coexistence of multiple types/speeds of transponders. Second, reconfigurable optical add drop multiplexers (ROADMs) together with a disaggregation method in OLS is imported in metro or long-haul DCI networks to meet the requirement of data center network architectures. To facilitate system integration, connection validation methods among devices from different vendors are introduced. Finally, we show two typical network automation implementations, e.g. spectrum optimization and automatic power adjustment.

From fixed grid to flexgrid systems

Nowadays, two different types of coherent transponder technology are commercially available. One is realized by the most spectrum efficient embedded optics technology, usually based on vendor specific coherent optics and digital signal processing (DSP) chips, such as 600 Gb/s or 800 Gb/s per wavelength with 90 Gbaud+ symbol rates. The other one is focused on interoperability, pluggable property, and low cost. The 400 Gb/s interface has already been defined by standard organizations or multi-source agreements (MSAs), e.g. labeled as OIF-400ZR [6] or OpenZR+ [7], with symbol rate around 60 Gbaud. Currently, OIF is developing next generation 800ZR implementation agreement.

The transponder technology evolves much faster than OLS, and operators always prefer the newest transponders to further reduce the network CapEx. However, a 50/100 GHz fixed grid multiplexer/demultiplexer neither supports 90 GB+ baudrate signals nor achieves a high spectral efficiency for 60 GB baudrate transponders. In order to accommodate various baud rates, a scalable solution is to move to a flexgrid system. Considering insertion loss and port isolation, a wavelength selective switch (WSS) with a high port count is preferred as the best choice to realize a flexgrid multiplexer/demultiplexer (FMD). The highest port count of a commercially available flexgrid WSS today is 32 among reported cases. To support 75 GHz channel spacing, which is compatible with a 400 Gb/s signal, the maximum add/drop port number should not be less than 64. Figure 1 shows a configuration of a 64-port multiplex/demultiplexer composed by a 3-dB coupler and two 32-port WSSes. In this configuration, coordination is required between the two WSSes to avoid frequency conflict. Usually a WSS has a higher insertion loss than an array waveguide grating (AWG), and it is possible to reduce this impact with better designing on the device function blocks, e.g. fully utilizing the two input ports of 2*2 coupler in an optical line protection (OLP) system as shown in Fig. 1 (a), where the diagram of an optical line terminal supporting FMD and OLP is illustrated.

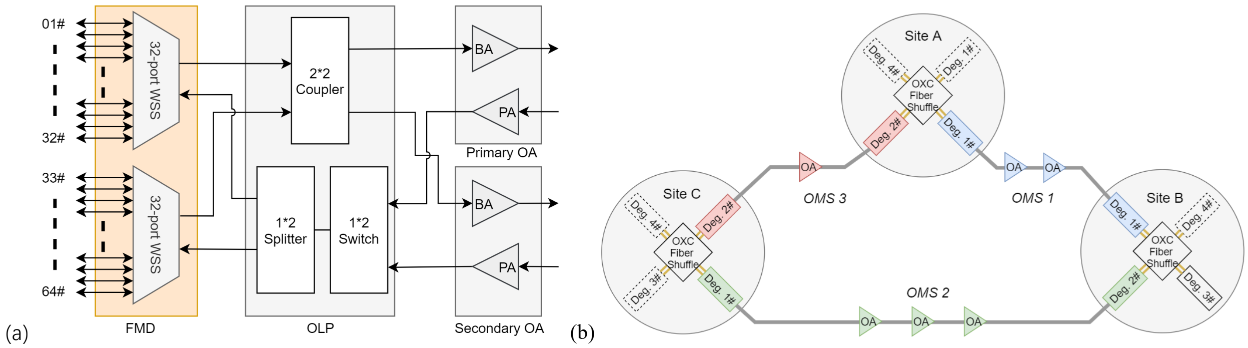

Fig. 1. (a) Optical line terminal supporting flexgrid operation. FMD: flexgrid multiplexer/demultiplexer, OLP: optical line protection, OA: optical amplifier, BA: booster amplifier, PA: pre-amplifier. (b) Network diagram with disaggregated ROADMs. Deg: degree, OXC: optical cross-connect, OMS: optical multiplex section. Different colors denote devices coming from different vendors.

From point-to-point to mesh networks

Although DCI may cover long haul or even subsea transmission, most common applications of DCI focus on metro distances, which is around 100 km. In an area with a few mega data centers, point-to-point topology is widely used for metro-DCI networks. In some areas, ROADM-based star or mesh topology may be more appropriate [8]. For example, in a metropolitan area, it is hard for hyperscalers to build large enough data centers to house massive number of servers due to land restriction, so high computation power has to be shared among several physically discrete middle data centers. Due to the limitation on fibers, it is difficult to fully connect every two data centers in a region, especially when the number of data centers are large. ROADM can switch traffic remotely at the wavelength layer without converting optical signals to electronic signals, therefore both CapEx and OpEx can be reduced.

In point-to-point disaggregated system, OLS part is relatively simple and usually comes from one vendor. Single-vendor may not be appropriate for a ROADM based network as it may restrict the network from growing at scale. OpenROADM MSA defines the interoperability specifications on the device level with the goal to realize fully disaggregated systems. Another method is to divide a network by optical multiplex sections (OMS), and each OMS can be a closed sub-system, with device coming from a single vendor [9]. As a result of OMS-level disaggregation, each ROADM node should be disaggregated by degree, as shown in Fig. 1 (b). To make this technology work, we designed an optical cross-connect (OXC) fiber shuffle to connect each WSS with vendor specific fiber interface. Figure 1 (b) shows three sites with OXC fiber shuffle, and each of them has two degrees, which are from different vendors. In this way, it is easy to scale a mesh network as business grows without vendor restrictions.

System integration and connection validation

As every coin has two sides, disaggregated systems usually include devices from more than one vendor, and operators may take more responsibilities on device level testing and system integration. During the device integration stage, functionality verification and performance evaluation based on specifications are necessary. Another important issue is to check the north bound interface according to the pre-defined YANG models. This activity used to take a lot of testing efforts since there are tens of models and hundreds of leaves. To accelerate the testing process, we developed an automatic testing tool [10]. It can generate reports on the read and write properties of the suspected leaves automatically and save a lot of R&D efforts from doing repetitive work.

During the construction stage, devices installation may be completed by different vendors, so it is important to validate fiber connections in such disaggregated systems. In a fixed grid system, as tributary ports of AWG are wavelength sensitive, it is easy to verify the connection between transponders and OLS just by observing the spectrum of optical channel monitor on the planned wavelength. While in flexgrid system, media channels in WSS are blocked by default or even not created, so it is hard to use the similar way to check the connection between transponders and WSS tributary ports. To solve this problem, each input port of our FMD has a photo detector and a short period of waveform is recorded continuously. If the transponder's output power is changed intentionally, the operator can locate the corresponding connected FMD port by checking all the waveforms on FMD's ports. Obviously, this method is vendor agnostic and can also be used to check the degree connection in ROADM nodes.

Network automation

To operate DCI networks growing at scale, more automation functions are needed to limit the growth of OpEx. There are three levels of automatic operations. The first level is purely inside the device, such as setting the target channel power in WSS, and it usually acts fast. The second level is within a sub-system, such as automatic power reduction between upstream and downstream optical amplifiers, and it relies on vendor specific protocol. The third level is at the network scale, such as automatic service provision and automatic power optimization. In a disaggregated system, network automation can only be achieved by an operator through network wide programmability with direct access to equipment from NMS.

Automatic service provision in flexgrid networks is a challenge due to the existence of unregular spectrum fragmentations. Defined by ITU-T G.694.1, the fine granularity for a spectrum width is 12.5 GHz. As spectra is allocated and deallocated dynamically with time, spectrum fragmentations occur and lead to spectrum utilization ratio reduction. Several defragmentation strategies are proposed for flexgrid optical networks. Push-and-pull method is a promising way since it can optimize the spectrum without service interruption [11]. Noticing that most of the commercial transponders only support *dark* tuning or very limited *bright* tuning range, e.g. +/-6.0 GHz, it is hard to deploy this method in field. To overcome this limitation, the tuning range of an integrable tunable laser assembly (ITLA) is extended to +/-40 GHz by refining the calibration process of the phase control circuitry. Service provision is automatically processed by a spectrum optimization algorithm in NMS. After checking every vacant slot in spectrum, all the possible candidate solutions are examined. The selected strategy is automatically executed by NMS and the real-time spectrum allocation is captured and shown in Fig. 2 (a).

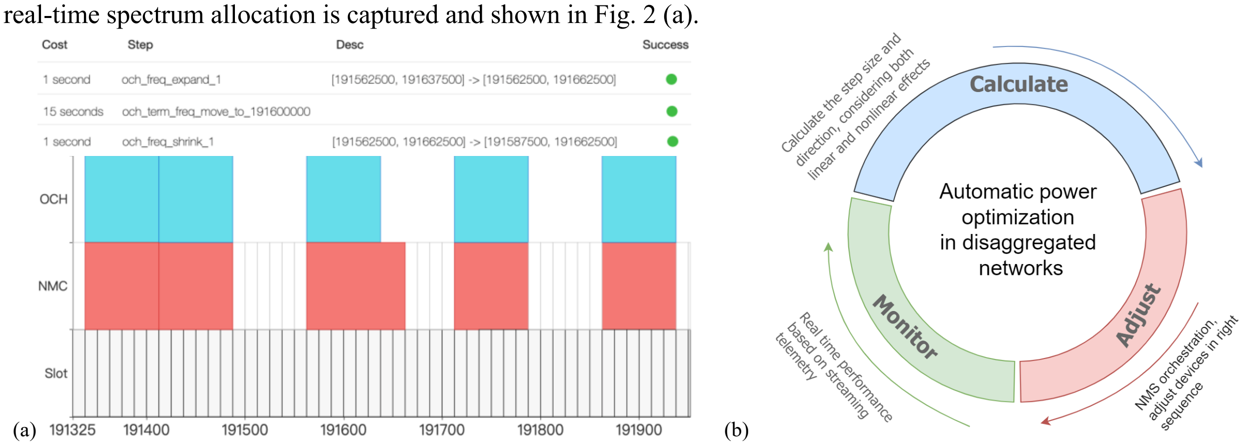

Fig. 2. (a) Spectrum optimization process with the operation log and the spectrum allocation of optical channel (OCH) and network media channel (NMC) layers. (b) Automatic power optimization process in disaggregated networks.

Automatic power optimization in disaggregated networks are implemented in NMS with three parts [9], as shown in Fig. 2 (b). First, target power values at every point along a route are pre-calculated by an engineering planning tool. During the calculation process, both linear and nonlinear effects are considered. Second, NMS iteratively configures devices in a margin-guaranteed manner with an adaptive step size, which is limited by checking real-time generalized optical signal-to-noise ratio margins generated from pre-forward error correction (pre-FEC) bit error rate (BER) of receivers. These key performance indices come from performance monitor part of NMS, and real time streaming telemetry is a proper technology to tradeoff between implementation complexity and data update frequency.

Conclusions

Open and disaggregated optical networks are becoming more complicated with transponders and network architectures evolving in data center networks. Flexgrid devices and ROADM become necessary in many hyperscalers networks. Several technologies related to system integration and network automation are introduced, and more are expected to improve the network efficiency further.

Reference

[1] Cisco Global Cloud Index: Forecast and Methodology, 2016–2021.
[2] C. Xie *et al*., "Open and Disaggregated Optical Transport Networks for Data Center Interconnects," J. Opt. Commun. Netw. 12, C12-22, 2020.
[3] http://www.openconfig.net/ [4] http://openroadm.org/
[5] E. Riccard *et al.*, "An Operator view on the Introduction of White Boxes into Optical Networks," JLT, v. 36, n. 15, pp. 3062-3072, 2018.
[6] https://www.oiforum.com/technical-work/hot-topics/400zr-2/ [7] https://openzrplus.org/
[8] L. Dou *et al.*, "Demonstration of Open and Disaggregated ROADM Networks Based on Augmented OpenConfig Data Model and Node Controller," OFC2020, paper M3Z.12, 2020.
[9] H. Zhang *et al.*, "Demonstration of a Disaggregated ROADM Network with Automatic Channel Provisioning and Link Power Adjustment," ECOC2021, paper demo 1, 2021.
[10] Y. Yin *et al.*, "Standardized Northbound Interface Testing Automation on the Open and Disaggregated Optical Transport Equipment," OFC2019, paper M3Z.3, 2019.
[11] K. Sone, *et al.*, "First Demonstration of Hitless Spectrum Defragmentation using Real-time Coherent Receivers in Flexible Grid Optical Networks," ECOC2012, paper Th.3.D.1, 2012.