Demonstration of Joint Operation across OpenROADM Metro Network, OpenFlow Packet Domain, and OpenStack Compute Domain

Behzad Mirkhanzadeh[1], Shweta Vachhani[2] , Balagangadhar G. Bathula[2], Gilles Thouenon[3], Christophe Betoule[3], Ahmed Triki[3], Martin Birk[1], Olivier Renais[3], Tianliang Zhang[1], Miguel Razo[1], Marco Tacca[1], Andrea Fumagalli[1]

[1]*Open Networking Advanced Research (OpNeAR) Lab, UT Dallas, TX, USA.* [2]*AT&T Labs, 200 Laurel Avenue South, Middletown, NJ, USA .* [3]*Orange Labs, 2 Avenue Pierre Marzin, Lannion, France*

Email: behzad@utdallas.edu

Abstract: Progress on the recent implementation of OpenROADM MSA functionalities is reported along with a description of the related TransportPCE SDN controller and PROnet multi-domain resource orchestrator software modules. These functionalities enable the described use cases.

OCIS codes: (060.4250) Networks, (060.2310) Fiber optics.

1. Introduction

Edge computing has become an attractive architecture for providing computing resources to many applications that demand specific QoS requirements by offering multiple compute resources at locations close to the application in a cost-effective way. However, the reliability of the edge data centers are still under doubt and any interruptions in the data center operations can bring down an entire business, unless an efficient backup strategy is devised. The use case envisioned in this paper consists of a main data center that can temporarily substitute the edge data centers in the presence of an imminent disaster. To achieve this goal, the PROnet Orchestrator [1] makes concurrent use of three open source platforms: (1) TransportPCE [2] based on OpenROADM [3], (2) OpenDaylight [2], and (3) OpenStack [4]. The OpenROADM platform is used to dynamically establish high-data rate optical circuits (up to 100Gbps) between data center pairs, the OpenDaylight platform is used to create OpenFlow rules for establishing data flows between the top-of-rack (TOR) switches of compute node racks, one at the primary and one at the backup data center, and lastly the OpenStack platform is used to execute the live migration of the virtual machines (VMs) over the newly created optical circuits.

In the remainder of the paper we first describe the current state of the art of OpenROADM and we then elaborate on the settings used to demonstrate feasibility of the proposed disaster recovery use case through commercial equipment, and collected experimental results.

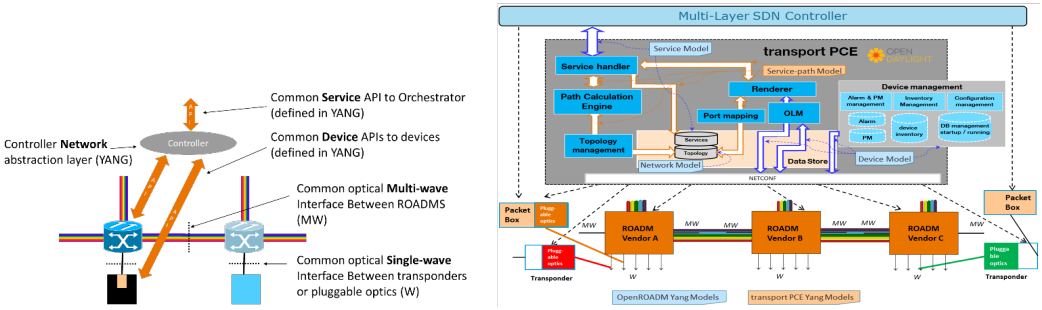
2. OpenROADM

OpenROADM Multi Source Agreement (MSA) defines the interoperability of Reconfigurable Optical Add/Drop Multiplexers (ROADMs). OpenROADM specifications also include transponders, OTN switches, and pluggable optics. There are more than 24 members, evenly split between network operators and equipment manufactures (for the most up to date list, refer to [3]). The OpenROADM MSA publishes optical specifications for data-plane interoperability, as well as open APIs and YANG data models for model-driven design. At a high level, the MSA provides two optical specifications for both single-wave and multi-wave interoperability, and three YANG data models for control plane interoperability (see Fig. 1(a)).

In order to enable plug-and-play of optical components, one approach would be to standardize every optical component in the optical path and their performance. The other approach would be to define optical behavior at interface points. This second approach was chosen here to give room to innovate on functions and yet make sure that everything works when plugged together.

The Single-Wave (SW) interface is defined in the optical specifications published on the OpenROADM MSA [3], currently for 100G (28 Gbaud), 200G, 300G, and 400G (63 Gbaud). 100G uses DP-QPSK modulation and hard-decision staircase forward-error-correction (FEC), as originally proposed in [5] for an interoperable mode. Interoperability at 63 Gbaud has three different possible modulation formats, QPSK (200G), 8 QAM (300G) and 16 QAM (400G) with a soft-decision FEC to get better performance. Instead of re-inventing the wheel, standards were re-used as much as possible, such as OTN framing from ITU-T. The Multi-Wave (MW) interface defines how ROADMs and in-line amplifiers (ILAs) interoperate. OpenROADM MSA supports different suppliers ROADMs to interconnect or ROADMs with different suppliers ILAs in the middle (even a mix of different suppliers ILAs).

The ROADM is defined as a black-box optical model with noise impact for through and add/drop paths. The optical control loops are divided into a local fast component (for optical impairments, such as transients) and a slower global component (for events, such as spectral balancing) that is abstracted into the SDN Controller.



(a) (b)

Fig. 1. (a) OpenROADM high-level overview. (b) TransportPCE architecture.

OpenROADM MSA publishes four major collections of YANG data models: device, service, network, and common models. Common models contain YANG files that are applicable to all three of the other models. The three other models are described in detail in the following paragraph.

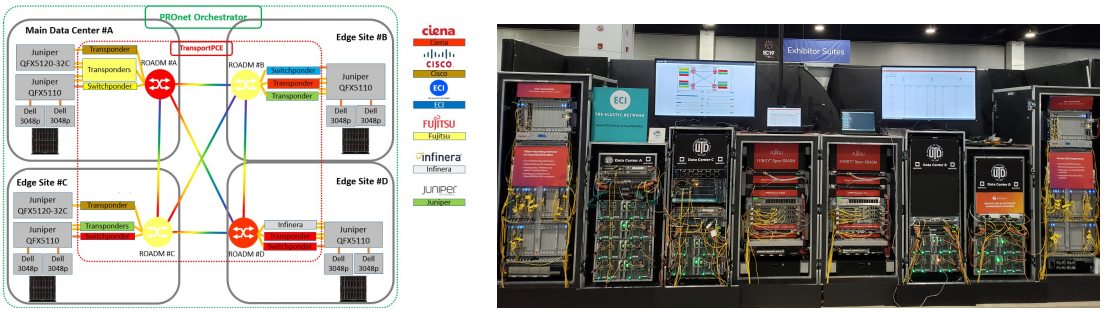
To guarantee plug-and-play compatibility of the OpenROADM devices, a standard interface (called Device Model) to the hardware is published through YANG data models, which are collectively known as the OpenROADM Device YANG models. The device implementation is a NETCONF [6] interface based on the YANG data model. An actual device will be an instance in JSON/XML that meets the YANG data-model rules. Any equipment complying with this standard can be used on a standard-based OpenROADM SDN Controller as the Southbound interface to the hardware. The Service Model describes the northbound interface out of the OpenROADM SDN Controller and is used to provision/delete/modify services. This interface is also described in YANG and most implementations use RESTCONF interface [7]. The Network Model describes the Topology Model inside an OpenROADM SDN Controller. The main goal of this model is to abstract away any hardware-dependent implementation details and present just the minimal amount of details needed for the optical path computation elements (PCEs) or other SDN applications to carry out their tasks. Having this abstraction layer also enables SDN applications to be written independently of the underlying hardware. Thus, switching between different manufacturers of hardware does not require any re-write of the SDN applications.

TransportPCE [8] was created in May 2016 as an incubation project of OpenDaylight. The main goal of the project was to provide the open-source community with a controller for optical infrastructures based on open standards. TransportPCE northbound API (to higher level controllers) relies on the OpenROADM Service Model, which is also used to describe the service topology in the ODL MD-SAL (Model-Driven Service Abstraction Layer). The network topology is maintained in the MD-SAL according to the OpenROADM Network Model. The southbound API to network elements is based on NETCONF and the OpenROADM Device Model. TransportPCE therefore provides a reference implementation for OpenROADM network control that can be reused in third party derived products. ODL Eclipse Public License (EPL) allows the integration of proprietary modules (weak copyleft [9]), which favors the adoption by both operators — that can rely on integrators to customize their automation environment — and equipment manufacturers — which can find a new way to monetize their research efforts while sharing the development effort for the base code. The TransportPCE modular architecture is described in Fig. 1(b) and a more detailed description is available at [8].

3. Experimental Results

This section describes the live demonstration of our disaster recovery scheme using commercial equipment, which was held in the UTD booth at (Supercomputing) SC'19 conference. The topology schematic is shown in Fig. 2(a) and Fig. 2(b) depicts the booth hosting the following equipment: four ROADM nodes provided by Ciena (6500) and Fujitsu (1FINITY), eight 100Gbps transponder blades (for a total of 16 wavelengths), and five switchponders (line rate of 100Gbps and client rates of 1Gbps and 10Gbps) — provided by Ciena, Cisco, ECI, Fujitsu, Infinera, and Juniper — are controlled by the TransportPCE plugin. Six Juniper QFX OpenFlow-enabled Ethernet switches are controlled by an OpenDaylight controller. A total of forty re-purposed Stampede compute nodes are controlled through OpenStack. We divided these gears to form three edge data centers and one main data center. The PROnet SDN Orchestrator is interfaced with the TransportPCE, OpenFlow ODL controller, and OpenStack and automatically executes the live VM migration procedure by sequentially first provisioning a wavelength circuit (at 100Gbps) between the selected edge data center (affected by the imminent disaster) and the the main data center,

then creating end-to-end Ethernet flows between the servers in the two data centers, requesting the live migration of the VMs from the edge data center to the main data center, and finally relinquishing the network resources (optical circuit and Ethernet flows), when they are not any longer used.



(a) (b)

Fig. 2. (a) Topology used in the SC'19 demo and (b) OpenROADM equipment at SC'19.

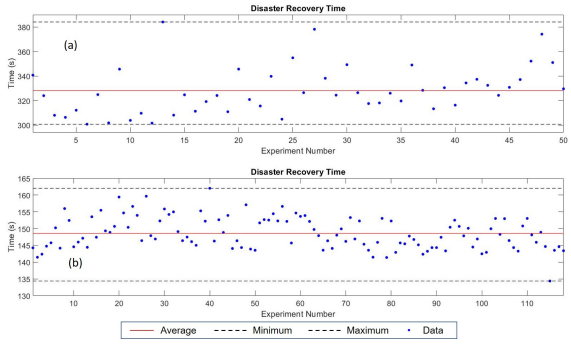


Fig. 3. Total disaster recovery time.

Fig. 3 reports the experimental results collected from a number of trials depicting two different scenarios. In the first scenario (top chart) 80 VMs (2TB of data) are offloaded from one of the edge data centers using a pair of dedicated 100Gbps transponders. In the second scenario (bottom chart) only 40 VMs (1TB of data) are offloaded using a pair of switchponders. The primary reason for the wide distribution of task completion times is the non-deterministic nature of the live migration procedure performed by OpenStack using pre-copy method which is highly depended on the VM load and the CPU utilization of the compute nodes involved. Network resources do not present any bottleneck during both procedures as each data center does not require more than 8Gbps of data transfer.

4. Conclusion

An OpenROADM demonstration consisting of certified equipment from six equipment vendors is described in this paper. The resources in the OpenROADM domain — controlled by TransportPCE — are leveraged to demonstrate a disaster recovery use case which also requires the orchestration of other resources, namely Ethernet flows (by using an OpenFlow controller) and compute resources (by using OpenStack). The PROnet SDN Orchestrator is used to sequentially provision resources and perform live migration of active VMs from one of three edge data centers to a main data center that is used as backup facility.

Acknowledgment

This work is supported in part by NSF grants CNS-1405405, CNS-1409849, ACI-1541461, and CNS-1531039T.

References

1. B. Mirkhanzadeh *et al.*, "An SDN-enabled multi-layer protection and restoration mechanism," Opt. Switch. Netw. **30**, 23 – 32 (2018).
2. "OpenDaylight," [online]., https://www.opendaylight.org/.
3. "OpenROADM MSA," [online]., http://OpenROADM.org.
4. "OpenStack," [online]., https://www.openstack.org/.
5. A. Mattheus *et al.*, "Black link versus cutting edge transceivers: A comparison for next generation WDM optical networks," in *2016 Optical Fiber Communications Conference and Exhibition (OFC),* (2016), pp. 1–3.
6. Internet Engineering Task Force (IETF), "Network configuration protocol (netconf)," in *RFC 6241,* (2011).
7. Internet Engineering Task Force (IETF), "Restconf protocol," in *RFC 8040,* (2017).
8. "Transport PCE Wiki," [online]., https://wiki.opendaylight.org/view/TransportPCE:Main.
9. "Copyleft," [online]., https://en.wikipedia.org/wiki/Copyleft.