Traffic Engineering Database dissemination for Multi-layer SDN orchestration

O. González de Dios[1], V. López[1], C. Haya[1], Chris Liou[2], Ping Pan[2], Gert Grammel[3], Javier Antich [3], Juan Pedro Fernández-Palacios[1]

[1] Telefónica Investigación y Desarrollo, Global CTO Unit (ogondio@tid.es)
[2] Infinera
[3] Juniper Networks

Abstract *Orchestration between multiple layers requires a standard architecture to enable such coordination in multi-vendor scenarios. This work proposes an architecture to solve multi-layer orchestration and experimentally validates topology dissemination in a multi-vendor environment.*

Introduction

Carrier Networks are based on multiple technologies and multiple vendors, with both horizontal (multi-domain) and vertical (multi-layer) relationships. One of the goals of Transport Software Defined Networking (SDN) is to provide the foundation for unifying and simplifying the control of such heterogeneous environments and truly enabling Multi-Layer (ML) Traffic Engineering Optimization. The optimization can be achieved by means of the Path Computation Element (PCE), which is a standard functional block defined by the IETF to solve the path computation problem in complex environment. It has been appointed as a feasible solution for multi-layer and multi-domain environments. However, PCE alone can not provide a complete solution like an SDN controller. Application-based Network Operations (ABNO) architecture [1] can cope with the requirements for ML use cases. The PCE is a key building block for ABNO or an SDN controller, as it is completely decoupled from the data and control plane technology.

One of the main issues of the SDN orchestration that needs to be solved is the ability to have a unified view of the heterogeneous network resources. This comprises two main challenges, one regarding the modeling of the network, and the other pertaining to the mechanisms for feeding the Multi-layer Traffic Engineering Database (TED) in such heterogeneous environments.

Two mechanisms are studied in this paper. In case of IP/MPLS, the recently proposed North-Bound Distribution of Link-State and TE Information using BGP, known as BGP-LS, is considered. For the converged transport layer (WDM converged with OTN), we examine a model in which data is exchanged in RESTful interface with JSON encoding and following the recent approaches from IETF I2RS WG and ONF to model network topology and resources in YANG.

The paper is organized as follows. First, the use case for multi-layer orchestration is outlined. Next, the architecture for multi-layer SDN orchestration is presented. Later, the two considered models for topology exchanges, BPG-LS and RESTful with JSON encoding of YANG data are explained. Finally, the experimental setup is presented.

Multi-layer use cases

Many network operator deployments are migrating towards an IP/MPLS over an optical transport network. Such transport network can be OTN, WSON or even SSON. Three main use cases can be identified for this type of multi-layer architecture: automatic ML link provisioning, ML restoration and ML network optimization. While in the automatic ML link provisioning use case, the trigger for the process is a request from the NMS (via GUI), ML restoration and ML network optimization are more typically triggered by a failure and a traffic measurement, respectively. There are multiple other flavors of ML use cases, but the key use case is automating ML link provisioning to reduce the turnup time for new network bandwidth, reduce operational costs and enable the other ML use cases.

Architecture for Multilayer Orchestration

When a request for a ML link provisioning between two routers is received, the SDN controller can determine the ML path between both routers (Fig. 1). This can be done thanks to the interaction between L3 PCE, VNTM and L0/L1 PCE, as demonstrated in [2]. SDN controller requests to the L3-PCE (stateless) for a path between both routers. If L3-PCE can not find a path in L3 topology, it asks VNTM for a new link in the VNT. After checking for policy rules, VNTM asks for a path from L0/L1 PCE between two transport nodes. We are assuming stateless PCEs, so this response is returned to the SDN controller.

Based on the PCE response, the SDN controller

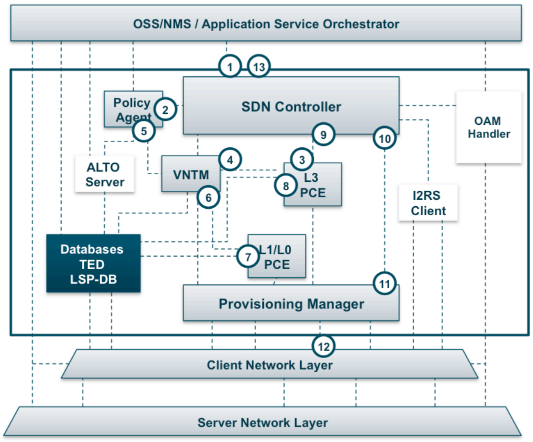

Fig. 1: Automatic ML link provisioning [1]

can then check if the establishment of a transport path is needed. If new transport bandwidth is required, the SDN controller would first configure the transport nodes using PCEP, UNI or OpenFlow. Another option is to configure forwarding between port and LSPs using OpenFlow and utilize the transport layer's control plane using OFConfig.

Once the transport layer is configured, routers must be updated. To do so, OpenFlow wire protocol looks like the best-suited protocol, as it uses a standard, practically realtime, interface for all routers. As OpenFlow cannot configure all options in routers, another protocol may be required (like NETCONF or OFConfig). To support multiple routers, a translation between YANG information models is required.

To run previous ML provisioning process, the SDN controller has to gather topology and resource information from WDM/ROADM, OTN, MPLS/IP networks, as well as application networks, such as CDN. As such, it is important for the underlying networks to present a consistent view of the nodes and links to the users, with the following key considerations:

1. Virtualization. The data queried by the controller should not contain transport technique-specific parameters. The controller should operate the underlying networks as a pool of bandwidth, regardless of the transmission types (optical, packet etc.)
2. Security. The SDN controller may provide "virtual slices" to different users. The users should not access each other's network and resource information.
3. Compatibility. The control-network interface should be open and supported by a large number of vendors and providers in the community.

Options for Multi-layer TED building

To support the control-network interface, we have adapted the recent approaches from IETF

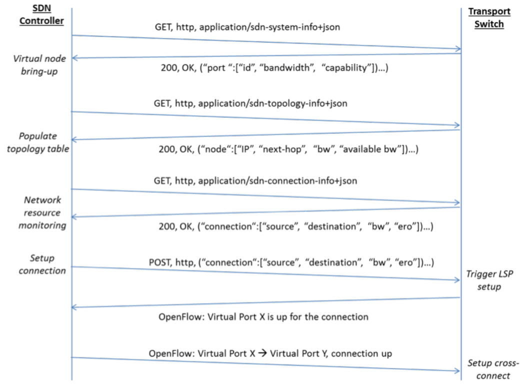

Fig. 2: SDN controller and Switch interactions

I2RS WG, ONF and open source community to model network topology and resources in YANG, and exchange data in RESTful interface with JSON encoding.

Fig. 2 depicts the sequence in gathering system resource and network topology, and setting up a connection from a SDN Controller. The SDN Controller will first initiate the requests to gather the system information, including ports and capabilities. It is followed by a query on network topology, where the network node will export the TE database, with both consumed and available bandwidth information to SDN controller. The controller can also obtain the existing network resource consumption by querying the LSP database. Finally, the controller and the transport node can utilize the existing OpenFlow 1.3 to setup the connection.

Fig. 3 shows a topology data example in JSON. The concept of using JSON is based on the 'Application Layer Traffic Optimization' ALTO architecture has been designed in order to deliver to applications through HTTP/JSON.

However in existing IP and MPLS enabled networks, BGP is the method of choice to disseminate inter-domain information. So an extension of BGP-LS is now used to distribute


```
{
  "links":[
  {
    "id":"urn:ogf:network:domain=infn-demo:node=infn-dtnx-1:port=300",
    "nodeID":"192.23.4.5",
    "localNode":"10.1.2.3",
    "remoteNode":"10.1.7.2",
    "MaxBandwidth":"5000000000",
    "UsedBandwidth":"1000000000",
    "MaxReservableBandwidth":"20000000000",
    "TrafficEngineeringMetric":"100",
    "ProtectionCapability":"yes",
    "SRLG":"10",
    "encodingType":"OTN",
  },
  {
    "id":"urn:ogf:network:domain=infn-demo:node=infn-dtnx-1:port=301",
    "nodeID":"192.23.4.5",
    "localNode":"10.1.2.5",
    "remoteNode":"10.1.52.2",
    "MaxBandwidth":"100000000000",
    "UsedBandwidth":"0",
    "MaxReservableBandwidth":"100000000000",
    "TrafficEngineeringMetric":"200",
    "ProtectionCapability":"no",
    "SRLG":"20",
    "encodingType":"Ethernet",
  }
  ]
}
```


Fig. 3: Topology data example in JSON

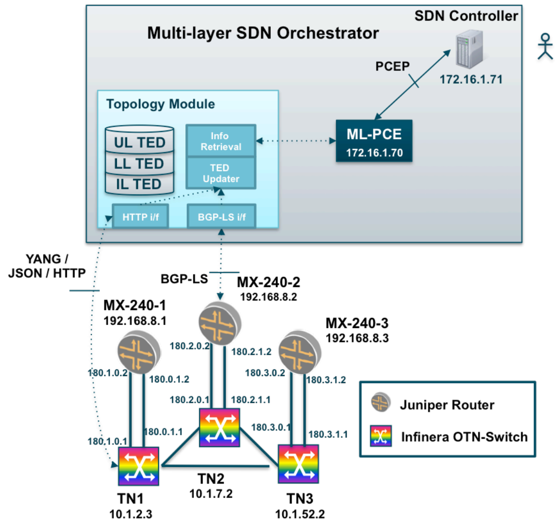

Fig. 4: Reference scenario

Link State and TE information which is needed to optimize topological decisions. BGP-LS is a way to collect Link State and TE information from existing networks and share it with external components by using the BGP routing protocol [3]. To enable this, a new BGP Network Layer Reachability Information (NLRI) encoding format is introduced. The mechanism is applicable to physical and virtual.

It builds on the fact that a router maintains one or more databases for storing link-state information about nodes and links in any given area. Link attributes stored in these databases include: local/remote IP addresses, local/ remote interface identifiers, link metric and TE metric, link bandwidth, reservable bandwidth, per CoS class reservation state, preemption and Shared Risk Link Groups (SRLG). The router's BGP process can retrieve topology from these LSDBs and distribute it to a client, either directly or via a peer BGP Speaker. In our case the PCE consumes the topological information in order to compute SDN controller requests.

Experimental validation

In order to validate the proposed architecture, a topology module has been developed to offer a common TED model to the ML-PCE, which was implemented by the authors in [4]. As shown in Fig. 4, TED updater module is in charge of receiving information from multiple interfaces and updating TEDs that are in the topology module. As shown in previous sections, Upper-Layer (UL) TED is feed based on the information from BGP-LS protocol, while Lower-Layer (LL) TED is updated with JSON-HTTP messages. Inter-Layer (IL) TED is not updated dynamically, but it is configured with the information about the port interconnected between the routers and the transport equipment. Information retrieval module sends the topology to the ML PCE on-demand.

Fig. 5 shows some details about OPEN and UPDATE BGP-LS messages that are received by the topology module. According to the preliminary AFI/SAFI values of BGP-LS in IANA register are set to 16388/71 respectively. BGP UPDATE message is an example of the information between MX-240-1 and MX-240-2, which is include in NLRI object.

Information disseminated with JSON-HTTP messages is displayed in Fig. 3. The information model is inside the JSON object, so no more information can be displayed with pcap traces.

The PCEP response with a ML Explicit Route Object (ERO) is composed by information of the equipment in the IP/MPLS layer and in the transport layer. To separate information between layers SERVER LAYER INFO object as shown in [2]. For transport nodes, Unnumbered interfaces are used with the IP address of the node and the port ID for the label, while IPv4 prefixes are used for routers.

Conclusions

This work presents a realistic problem to deploy SDN solutions TED dissemination. BGP-LS and HTTP/JSON approaches appear as two solutions, which can help to solve this problem. BGP-LS extends current functionalities in network elements to disseminate TED beyond AS reachability. This improves IGP solutions proposed now to feed PCE. On the other hand, HTTP/JSON simplifies protocol development for interoperability and, just it is required an agreement in the information model.

References

[1] D. King and A. Farrel, draft-farrkingel-pce-abno-architecture-03
[2] O. Gonzalez de Dios et al. IPOP 2012.
[3] Gredler et al. draft-ietf-idr-ls-distribution-02
[4] S. Martinez, OW4G.1, OFC 2013.

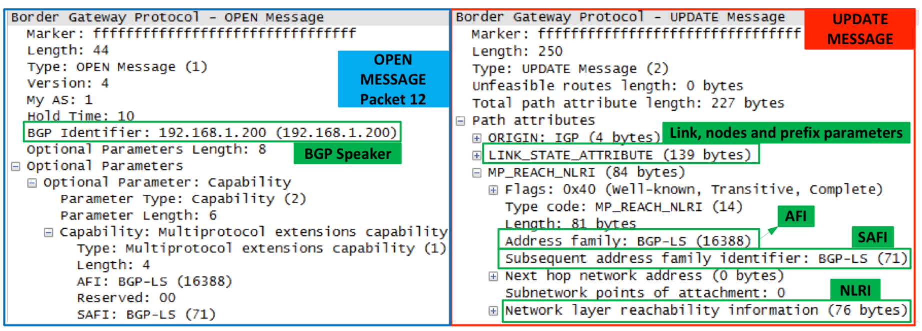

Fig. 5: Wireshark Capture of messages in the Topology Module