

# Experimental Assessment of ABNO-driven Multicast Connectivity in Flexgrid Networks

Ll. Gifre, F. Paolucci, O. González de Dios, L. Velasco, L. M. Contreras, F. Cugini, P. Castoldi, and V. López

**Abstract**—The increasing demand of internet services is pushing cloud services providers to increase the capacity of their data centers (DC) and create DC federations, where two or more cloud providers interconnect their infrastructures. As a result of the huge capacity required for the inter-DC network, the flexgrid optical technology can be used. In such scenario, applications can run in DCs placed in geographically distant locations and hence, multicast-based communication services among their components are required. In this paper, we study two different approaches to provide multicast services in multi-layer scenarios assuming that the optical network is based on the flexgrid technology: *i*) establishing a point-to-multipoint optical connection (light-tree) for each multicast request, and *ii*) using a multi-purpose virtual network topology (VNT) to serve both unicast and multicast connectivity requests. When that VNT is not able to serve an incoming request as a result of lack of capacity, it is reconfigured to add more resources. A control plane architecture based on the Applications-based Network Operations (ABNO) one, currently being standardized by the IETF, is presented; workflows are proposed and PCEP extensions are studied for the considered approaches. The experimental validation is carried-out on a test-bed set-up connecting Telefonica, CNIT, and UPC premises.

**Index Terms**—Optical multicast, Flexgrid networks, Datacenter interconnection.

## I. INTRODUCTION

THE distributed nature of cloud computing entails that modular applications and services can dynamically distribute their components to be run on servers belonging to datacenters (DCs) spread across distant geographic locations. Huge data transfer is thus needed e.g. to synchronize databases (DB) or to distribute content, e.g. live TV [1], among DCs. The new capacity demands advocate for an evolution towards optical transport-based solutions for DC interconnection [2]. In this area, the flexgrid optical network technology is being extensively investigated because of its inherent spectrum efficiency and connection flexibility [3].

Some applications might require group communication services (one-to-many or many-to-many) among their distributed components. Although point-to-point (p2p) connectivity can be used thus, copying contents to each of the destinations, point-to-multipoint (p2mp) connections might fit better to this purpose.

In the case that high capacity connectivity (e.g. 100Gb/s) is required, p2mp connections can be established on the optical layer (known as *light-trees*). The feasibility of creating light-trees in flexgrid networks was demonstrated in [4]. Authors in [5] compared the performance of using light-trees against p2p optical connections (*lightpaths*) and showed that, although resource utilization is improved using light-trees, its limitation lie in the spectrum continuity constraint [6] and the connections' length. Another approach to improve resource utilization is creating a virtual network topology (VNT) that can be used to serve both p2p and p2mp connections. Authors in [7] compare the performance of providing multicast services on the single layer and the multilayer approach.

This paper extends our previous work in [8] and experimentally demonstrate the above approaches when the control of the flexgrid network resides in a centralized control element following the Software Defined Networking (SDN) concept; in this paper we assume that the Application-Based Network Operations (ABNO) architecture [9] is used. The DCs participating in the communication are considered to be part of a federation. As such, a cloud management system at federation level is in charge of managing intra-DC resources [10]. Typically, one or more core Ethernet programmable switches within each DC are connected to the flexgrid optical core network providing the necessary interconnection among the DCs in the federation. The necessary coordination between the cloud management system in each DC and the ABNO-based transport network control element is performed by a differentiated software element, named as Application Service Orchestrator (ASO), acting as an end-to-end orchestrator and responsible of translating the connectivity requests from the DCs to the ABNO element. A similar scenario was used to demonstrate unicast services on multi-layer scenarios in [11].

The ABNO architecture is based on functional elements defined by the IETF, like the active stateful path computation element (PCE) [12]. As described in [11], most of interfaces among ABNO modules are PCEP [13]. A specialized PCE can be used to perform complex computations, e.g. to perform in-operation planning [14], as demonstrated in [15], [16].

The remainder of this paper is organized as follows. Section II presents the considered approaches to serve large capacity multicast connectivity in multi-layer scenarios. Section III proposes an architecture to serve inter-datacenter multicast connectivity, proposes a workflow for each of the approaches, and studies the support of PCEP. Section IV experimentally validates the workflows and PCEP considerations in a distributed test-bed facility. Finally, Section V draws the main conclusions of the paper.

Manuscript received October 6, 2014.

Ll. Gifre (lgifre@ac.upc.edu) and L. Velasco are with the Optical Communications Group (GCO) at Universitat Politècnica de Catalunya (UPC), Barcelona, Spain.

F. Paolucci, F. Cugini, and P. Castoldi are with Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT), Pisa, Italy.

O. González de Dios, L. M. Contreras, and V. López are with Telefónica Investigación y Desarrollo (TID), Madrid, Spain.

## II. APPROACHES TO SERVE MULTICAST CONNECTIVITY SERVICES

To implement multicast connectivity services in a multi-layer network, a VNT is needed to connect every switch. As a result of the large bitrate required for the multicast connectivity, the VNT can be created ad-hoc for each multicast request and removed when the multicast connection is torn-down.

Although alternative approaches can be followed (see e.g. [5]), in this paper we assume that the ad-hoc VNT is based on one single light-tree connecting the source switch to each of the destination switches. An example of the light-tree-based approach is depicted in Fig. 1, where a 100 Gb/s light-tree is set-up in Fig. 1a connecting the switch in DC-A to the switches in DCs B, C, and D. It is clear that, similarly as for p2p connections [6], specific p2mp routing and spectrum allocation (RSA) algorithms, similar to the one proposed in [7], are needed to compute the route from the source to every leaf in the connection and allocate a contiguous and continuous spectrum slot. The resulting VNT consists in a single p2mp virtual link connecting the source switch to all the leaves (Fig. 1b). The multicast connection request can be served on the just created ad-hoc p2mp VNT (Fig. 1c). The advantage of this approach include the reduction of switching capacity in the switches, since multicast is performed by the optical layer.

Although the p2mp virtual link created for the multicast connection request could be shared, the probability of another

multicast request with the same source and set of destination switches as that virtual link arrive is clearly scarce. Thus, in case of some capacity would remain unused after serving the multicast request, as a result of a mismatch between the requested bitrate and the optical layer granularity, it would be unlikely reused.

To foster capacity sharing, a single VNT can be created to serve any kind of connection request, including unicast and multicast connections. In this approach, connection requests are served on the VNT, which is updated to add extra capacity, e.g. new virtual links, in case an incoming connection cannot be served, and releasing virtual links when any connection is using them. Fig. 2 shows an example of the multi-purpose VNT approach. In Fig. 2a the current multi-purpose VNT has not enough capacity to serve an incoming 100 Gb/s multicast request from the switch in DC-A to the switches in DCs B, C, and D. Therefore, the VNT is updated adding a new virtual link [7]. Taking advantage of traffic grooming to improve resource utilization, the VNT can be supported on large capacity, e.g. 400 Gb/s, lightpaths; then a new virtual link connecting switch DC-C to switch DC-D with 400Gb/s of remaining capacity is now available in the VNT, as shown in Fig. 2b. Finally, the requested multicast connection can be served on the multi-purpose VNT, as illustrated in Fig. 2c.

In the next section, multicast connectivity services are provisioned among a set of DCs. A control and management architecture is presented and workflows for both, the light-tree-based and the multi-purpose VNT approach are proposed.

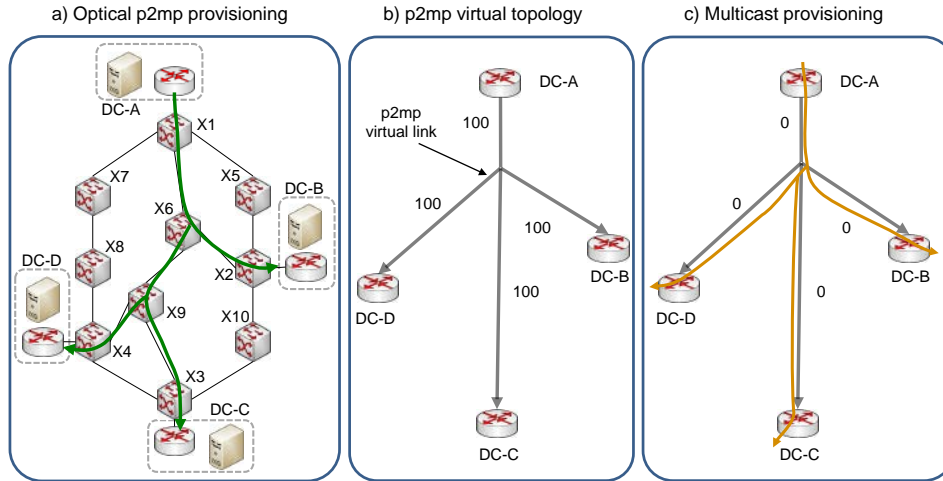


Fig. 1. Light-tree-based VNT to serve an incoming multicast request.

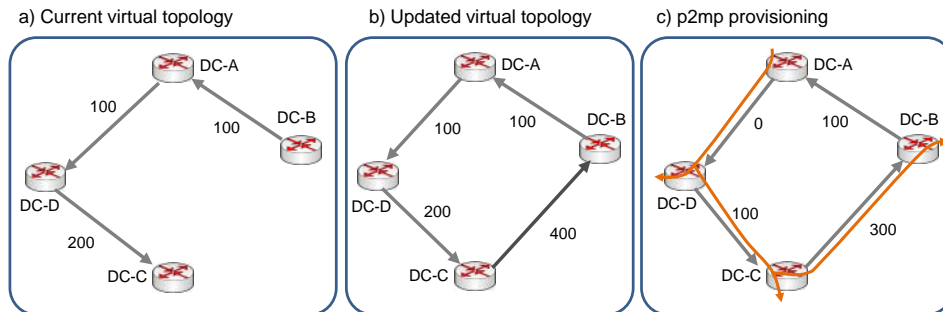


Fig. 2. Multi-purpose VNT updating to serve an incoming multicast request.

### III. PROVISIONING INTER-DATACENTER MULTICAST CONNECTIVITY

#### A. Datacenter management

The delivery of distributed cloud services implies the configuration of the networks involved in the service (both in the distributed DCs and in the domains connecting them). In the case of federated DCs, each DC is operated and managed separately. Service provisioning in such distributed scenario requires a tight coordination to ensure the consistency of the service delivery. This problem space is covered by the SDN framework, which is seen as the facilitator of this capability.

For this work, three main software components are considered to govern and control the network connectivity service: *i*) the Cloud computing manager, which handles the computing resources in each data center. Each intra-DC network can include a local SDN controller to configure resources within the DC. Such controllers are out-of-the-scope of this paper; *ii*) the ASO, which maintains the network information from the DCs requests and interacts with the network orchestrator; and *iii*) the inter-DC network orchestrator, based on the ABNO architecture, which configures the Ethernet switches and the flexgrid network.

When federated DCs need connectivity services they are requested to the ASO, who asks ABNO to create unicast and multicast connections between the involved DCs (Fig. 3). Unicast IP services were demonstrated for multi-layer scenarios in our previous work [11]. In this paper, Ethernet services over flexgrid are provided instead of IP/MPLS, but the orchestration process for p2p services is similar.

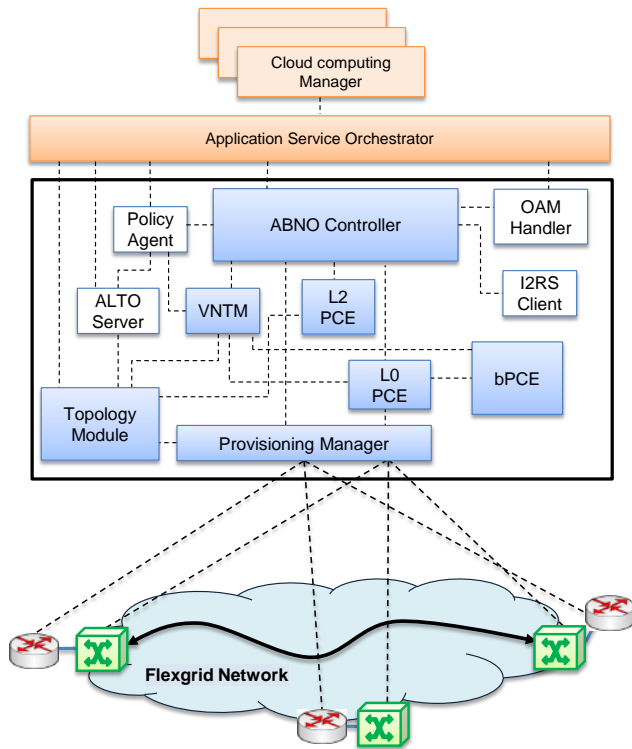


Fig. 3. Control and management architecture.

#### B. Carrier SDN Architecture

The ABNO architecture includes *i*) a controller, responsible for implementing workflows orchestrating operations among ABNO modules; *ii*) a Layer 0 active stateful PCE (L0 PCE) with label switched paths (LSP) initiation capabilities [17], responsible for path computation on the optical topology. The LSP-DB stores information about the LSPs that are provisioned and operational; *iii*) a Virtual Network Topology Manager (VNTM) [18], responsible for maintaining a virtual topology between the DCs using resources in the optical topology; *iv*) a stateless Layer 2 PCE (L2 PCE), which computes paths on the virtual topology; *v*) a Provisioning Manager (PM) dealing with the configuration of the network elements (switches or optical nodes) and *vi*) a Topology Module (TM) maintaining the Traffic Engineering Database (TED). In addition, a back-end PCE (bPCE) capable of performing computationally intensive tasks, such as solving the p2mp RSA algorithm or finding the optimal reconfiguration of the VNT, is available within ABNO.

Next subsections propose workflows for the considered multicast approaches and highlight meaningful PCEP issues.

#### C. Proposed Workflows

The ASO module is responsible for managing multiple DC networks as a single entity; it requests the ABNO controller for unicast and multicast services between Ethernet switches. Fig. 4 for the light-tree-based VNT approach and Fig. 5 for the multi-purpose VNT approach show the proposed workflow to serve multicast connection requests.

Upon a multicast connection request arrives in the ABNO controller, it requests a L2 p2mp path computation to the L2 PCE (*Path Computation Request* (PCReq) message 1 in Fig. 4 and Fig. 5). In case of a VNT with the same source and destination switches and with enough capacity is available, the L2 PCE would reply with the route for that request. However, this is unlikely to happen so the L2 PCE returns a *Reply* (PCRep) message (2) with a NO-PATH object. The controller delegates to the VNTM module updating the VNT, possibly adding more resources to serve the L2 p2mp request. To that end, an *Initiate* (PCInit) message (3) is sent containing the end points of the requested p2mp connection.

At this point the VNTM will implement either the light-tree-based or the multi-purpose VNT approach. Regardless of the approach followed, the VNTM sends back a *Report* (PCRpt) message (9) reporting the results. In case of success, the ABNO controller requests a L2 p2mp path computation to the L2 PCE (10), which now finds a feasible route and returns it to the ABNO controller (11). Since L2 PCE is not active, the ABNO controller module delegates connection set-up to the PM (12) that configures the appropriate rules in each switch.

##### 1) Light-tree-based VNT approach

In the case the light-tree-based VNT approach is followed, upon reception of message (3) the VNTM sends a PCReq message (message 4a in Fig. 4) to the L0 PCE to create optical connectivity among the specified end points; in this particular approach, a p2mp optical connection needs to be created.

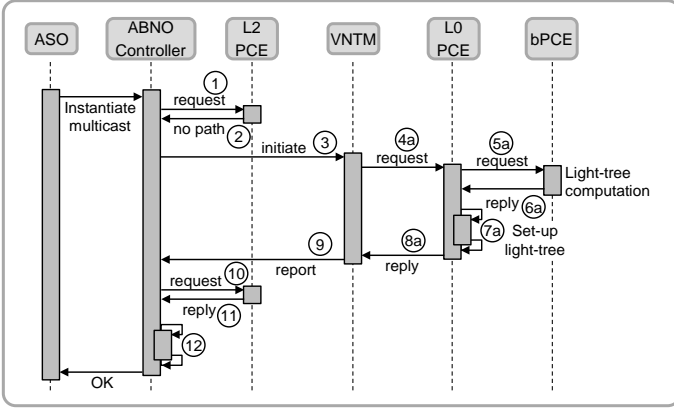


Fig. 4. Workflow for the light-tree-based VNT approach

Because L0 p2mp path computation might take long time, L0 PCE delegates it to the specialized bPCE (5a). When the computation ends, the bPCE sends back the solution to the L0 PCE (6a). When the PCE receives the solution, it sends the appropriate commands to the underlying data plane (7a). When the L0 PCE receives the confirmation from the data plane, a PCRep message (8a) is sent back to the VNTM.

## 2) Multi-purpose VNT approach

In this approach, upon reception of message (3) the VNTM needs to reconfigure the VNT, e.g. by adding some new virtual links. Similarly as in the previous case with the p2mp RSA problem, since computing a solution for VNT reconfiguration might take long time, the VNTM delegates it to the specialized bPCE, sending a PCReq message (message 4b in Fig. 5). When the computation ends, the bPCE sends back the solution to the VNTM (5b) specifying the lightpaths to be set-up including the route and spectrum allocation for each of them.

When the VNTM receives the solution, it delegates its setting-up to the L0 PCE (6b), which in turn sends the appropriate commands to PM module to establish the computed lightpaths in the underlying data plane (7b). When the L0 PCE receives the confirmation from the PM module, a PCRep message (8b) is sent back to the VNTM.

## D. PCEP issues

In this section we focus on analyzing the differences between PCReq and PCRep messages (5a) and (6a) in the light-tree-based VNT approach and (4b) and (5b) in the multi-purpose VNT approach.

In the light-tree-based VNT approach, PCEP messages follow PCEP standards for p2mp LSPs [19]. PCReq message (5a) requests a single L0 p2mp LSP. The end points can be specified using an ENDPOINTS class 3 (p2mp IPv4) object, which includes the IP addresses of the source and the leaves.

PCRep message (6a) specifies the computed RSA for the p2mp LSP using explicit route (ERO) and secondary explicit route (SERO) objects; one single ERO object is used to define the route and spectrum allocation from the source node to one of the leaves while additional SERO objects define the route and spectrum allocation for each of the rest of leaves; the starting node in each SERO object can be whatever node

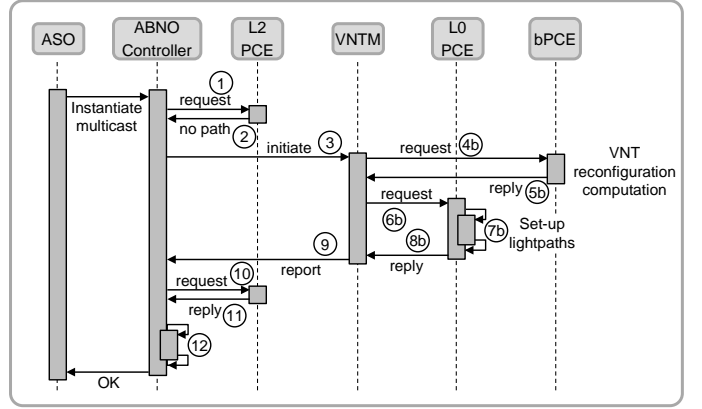


Fig. 5. Workflow for the multi-purpose VNT approach

already in the ERO or SERO objects.

In contrast, standardization under the multi-purpose VNT approach is currently scarce and incomplete. Note that in this approach the VNTM needs to reconfigure the VNT to serve the incoming multicast request. Nonetheless, the exact L0 LSPs to be created to increase the capacity of the VNT are the result of the optimization algorithm running in the bPCE. Hence, the PCReq message (4b) includes the multicast request with some additional objects adapted from [20]. The INTER-LAYER object indicates whether inter-layer path computation is allowed, whereas the SWITCH-LAYER object specifies the layers to be considered, in our case L0 as well as L2.

TABLE I. PCREQ MESSAGE (4b) CONTENTS

```

<PCReq Message> ::= <Common Header>
                    <request-list>

where:
  <request-list> ::= <request>
                    [ <request-list> ]

  <request> ::= <RP>
               <END-POINTS>
               <BANDWIDTH>
               <of-list>
               <INTER-LAYER>
               <SWITCH-LAYER>

where:
  <of-list> ::= <OF> [ <of-list> ]
  
```

TABLE II. PCREP MESSAGE (5b) CONTENTS

```

<PCRep Message> ::= <Common Header>
                    <response-list>

where:
  <response-list> ::= <response>
                     [ <response-list> ]

  <response> ::= <RP>
                [ <NO-PATH> ]
                [ <ERO> ]
                [ <server-layer path-list> ]

  <server-layer path-list> ::=
    <server-layer path>
    [ <server-layer path-list> ]

  <server-layer path> ::= <ERO>
                        <BANDWIDTH>
                        <SERVER-INDICATION>
  
```

Additionally, two objective function (OF) objects are included; the first OF object specifies the objective function for L2, whilst the second one refers to L0 and includes a TLV with the bandwidth of the L0 LSPs to be created. Table I depicts the contents of the PCReq message (4b).

Regarding the PCRep message (5b) (Table II), it is worth highlighting that ERO objects can be used to point out paths computed on the requested layer and in server layers. In the latter case, a SERVER-INDICATION object is used to specify the layer in which a path has been computed among those allowed in the incoming SWITCH-LAYER object.

#### IV. EXPERIMENTAL VALIDATION

In this section, we experimentally assess the proposed workflows, including the PCEP messages. We start with a brief description of the deployed distributed set-up including the implemented modules. Next, the set-up is used to run the workflows; the network topology depicted in Fig. 1a was used for the experiments. Protocol captures show the exchanged PCEP messages.

##### A. Scenario

The experimental validation was carried out on a distributed field trial set-up connecting premises in Telefonica (Madrid, Spain), CNIT (Pisa, Italy), and UPC (Barcelona, Spain) (see Fig.6). All three locations are connected through IPsec tunnels and PCEP sessions are established on top of them.

The Cloud computing manager is OpenStack Grizzly [21]. Neutron plugin is in charge of providing the local overlay networks. As Neutron is a technology dependent plugin, a customized Neutron plugin has been developed to manage the inter-data center connectivity using an SDN approach; the plug-in interacts with the local SDN controller and the ASO. ASO and each of the OpenStack instances exchange information via the Neutron plugin, so the ASO is aware of which VMs in each DC belong to the same network.

OpenStack and Neutron plugin, most of the ABNO modules –the controller, the VNTM, the PM, and the L2 PCE– and the L2 data plane are located in Telefonica’s premises. The ABNO controller has been developed in Java and supports multiple workflows (e.g. connection set-up or re-optimization); the specific workflow to be executed is defined in the incoming request [11]. The L2 data plane consists of HP 5406zl Ethernet switches; the PM module configures the switches using OpenFlow via a Floodlight [22] controller.

The CNIT’s active stateful L0 PCE has been implemented in C/C++ for Linux. The front-end PCE includes a PCEP Server module, a Path Solver module and the databases (i.e., the TED and the LSP-DB). The TED and the LSP-DB are kept updated by means of PCNtf and PCRpt messages sent by LSP source nodes. The PCEP Server has been extended to enable back-end computation, functionally separated from PCEP sessions established locally with the data plane nodes.

The L0 data plane includes four programmable spectrum selective switches (SSS); to complete the network topology, six node emulators are additionally deployed. Nodes are handled by co-located GMPLS controllers running RSVP-TE with flexgrid extensions [23]. GMPLS controllers communicate with the L0 PCE by means of PCEP.

Finally, the UPC’s bPCE (PLATON) [24] has been developed in C++ for Linux and is organized into 4 main building blocks. The first block is responsible for managing communications with other ABNO modules using standard protocols. The second block, the controller, manages PLATON execution: when incoming messages arrive, the controller decides the action to be taken among updating either the local network topology or the state of network resources, and running optimization algorithms. The third block contains the databases (TED and LSP-DB) for both, L2 and L0. The last block handles optimization algorithms, which are deployed as dynamically linked libraries to allow that third party algorithms can be easily added into PLATON.

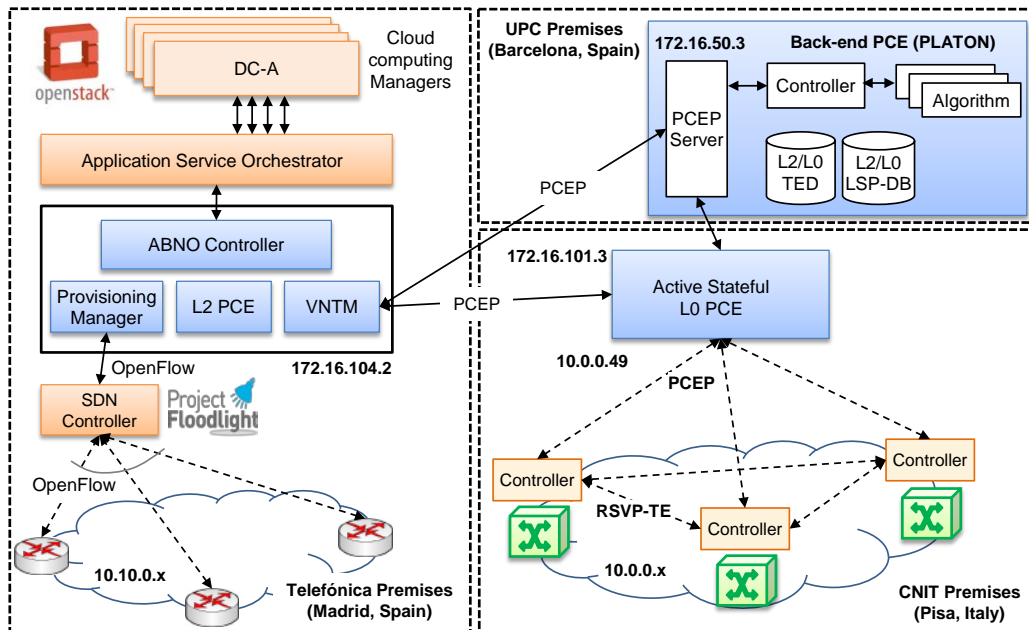


Fig.6. Distributed test-bed set-up. IP addresses are shown.

|    | Source       | Destination  | Info                             |
|----|--------------|--------------|----------------------------------|
| ①  | 172.16.104.2 | 172.16.104.2 | PATH COMPUTATION REQUEST MESSAGE |
| ②  | 172.16.104.2 | 172.16.104.2 | PATH COMPUTATION REPLY MESSAGE   |
| ③  | 172.16.104.2 | 172.16.104.2 | INITIATE MESSAGE                 |
| ④a | 172.16.104.2 | 172.16.101.3 | PATH COMPUTATION REQUEST MESSAGE |
| ⑤a | 172.16.101.3 | 172.16.50.3  | PATH COMPUTATION REQUEST MESSAGE |
| ⑥a | 172.16.50.3  | 172.16.101.3 | PATH COMPUTATION REPLY MESSAGE   |
| ⑦a | 10.0.0.49    | 10.0.0.1     | INITIATE MESSAGE                 |
| ⑧a | 10.0.0.1     | 10.0.0.49    | REPORT MESSAGE                   |
| ⑨a | 172.16.101.3 | 172.16.104.2 | PATH COMPUTATION REPLY MESSAGE   |
| ⑩  | 172.16.104.2 | 172.16.104.2 | REPORT MESSAGE                   |
| ⑪  | 172.16.104.2 | 172.16.104.2 | PATH COMPUTATION REQUEST MESSAGE |
| ⑫  | 172.16.104.2 | 172.16.104.2 | PATH COMPUTATION REPLY MESSAGE   |
| ⑬  | 172.16.104.2 | 172.16.104.2 | INITIATE MESSAGE                 |
| ⑭  | 172.16.104.2 | 172.16.104.2 | REPORT MESSAGE                   |

Fig. 7. Relevant PCEP Messages for the light-tree-based VNT approach

```

⑤a END-POINT object
  Object Class: END-POINT OBJECT (4)
  Object Type: P2MP IPv4 (3)
  Leaf Type: New Leaves to add (1)
  Source IPv4 address: 10.0.0.1
  Destination IPv4 address 1: 10.0.0.2
  Destination IPv4 address 2: 10.0.0.3
  Destination IPv4 address 3: 10.0.0.4

⑫ EXPLICIT ROUTE object (ERO)
  >SUBOBJECT: Unnumbered Interface ID: 10.0.0.1:6
  >SUBOBJECT: Label n=-4, m=4
    Label: 6a 00 ff fc 00 04 00 00
  >SUBOBJECT: Unnumbered Interface ID: 10.0.0.6:2
  >SUBOBJECT: Label Control
  >SUBOBJECT: IPv4 Prefix: 10.0.0.2/32
  >SECONDARY EXPLICIT ROUTE object (SERO)
  >SUBOBJECT: Unnumbered Interface ID: 10.0.0.6:9
  >SUBOBJECT: Label Control
  >SUBOBJECT: Unnumbered Interface ID: 10.0.0.9:3
  >SUBOBJECT: Label Control
  >SUBOBJECT: IPv4 Prefix: 10.0.0.3/32
  >SECONDARY EXPLICIT ROUTE object (SERO)
  
```

Fig. 8. Details of p2mp objects in PCEP Messages.

All the modules in ABNO except L0 PCE and bPCE are configured with the same IP address: 172.16.104.2. L0 PCE runs in IP 172.16.101.3 and bPCE runs in IP 172.16.50.3.

### B. Light-tree-based VNT approach

Fig. 7 shows the relevant PCEP messages for the light-tree-based VNT workflow. Each message is identified with the same sequence used to describe the workflow. The details of messages (5a) and (6a) are given in Fig. 8. PCReq message (5a) requests a single L0 p2mp LSP. The end points are specified using an ENDPOINTS object, which includes the IP addresses of source (X1) and leaves (X2, X3, and X4).

PCRep message (6a) contains the computed RSA for the L0 LSP. In line with [19], aiming at describing p2mp routes in an efficient way, we use one single ERO object and additional SERO objects to define the route and spectrum allocation. To illustrate that, the route defined by the ERO in Fig. 8 is for leaf X2 and includes X6 as intermediate node. The first SERO object is for leaf X3 and starts in X6.

As defined in [3], spectrum allocation is described using the tuple  $\{n, m\}$ , where  $n$  is the number of slices (positive, negative or 0) from a reference frequency (193.1 THz), and  $m$  is the number of slices at each side of the central frequency. Since we computed transparent light-trees, the spectrum allocation must be continuous alongside the whole light-tree.

### C. Multi-purpose VNT approach

Fig. 9 shows the relevant PCEP messages for the multi-purpose VNT workflow. The details of messages (4b) and (5b)

|    | Source       | Destination  | Info                             |
|----|--------------|--------------|----------------------------------|
| ①  | 172.16.104.2 | 172.16.104.2 | PATH COMPUTATION REQUEST MESSAGE |
| ②  | 172.16.104.2 | 172.16.104.2 | PATH COMPUTATION REPLY MESSAGE   |
| ③  | 172.16.104.2 | 172.16.104.2 | INITIATE MESSAGE                 |
| ④b | 172.16.104.2 | 172.16.50.3  | PATH COMPUTATION REQUEST MESSAGE |
| ⑤b | 172.16.50.3  | 172.16.104.2 | PATH COMPUTATION REPLY MESSAGE   |
| ⑥b | 172.16.104.2 | 172.16.101.3 | PATH COMPUTATION REQUEST MESSAGE |
| ⑦b | 10.0.0.49    | 10.0.0.3     | INITIATE MESSAGE                 |
| ⑧b | 10.0.0.3     | 10.0.0.49    | REPORT MESSAGE                   |
| ⑨b | 172.16.101.3 | 172.16.104.2 | PATH COMPUTATION REPLY MESSAGE   |
| ⑩  | 172.16.104.2 | 172.16.104.2 | REPORT MESSAGE                   |
| ⑪  | 172.16.104.2 | 172.16.104.2 | PATH COMPUTATION REQUEST MESSAGE |
| ⑫  | 172.16.104.2 | 172.16.104.2 | PATH COMPUTATION REPLY MESSAGE   |
| ⑬  | 172.16.104.2 | 172.16.104.2 | INITIATE MESSAGE                 |
| ⑭  | 172.16.104.2 | 172.16.104.2 | REPORT MESSAGE                   |

Fig. 9. Relevant PCEP Messages for the multi-purpose VNT approach

```

> PATH COMPUTATION REQUEST MESSAGE Header
> RP object
> END-POINT object
  Object Class: END-POINT OBJECT (4)
  Object Type: P2MP IPv4 (3)
  Leaf Type: New Leaves to add (1)
  Source IPv4 address: 10.10.0.1
  Destination IPv4 address 1: 10.10.0.2
  Destination IPv4 address 2: 10.10.0.3
  Destination IPv4 address 3: 10.10.0.4
> BANDWIDTH object
  Object Class: BANDWIDTH OBJECT (5)
  Bandwidth: 100,000000
> OF object
  OF-Code: L2 VNT Reconfig (41)
> OF object
  OF-Code: L0 Path Comp (42)
  > TLV BANDWIDTH (502)
    Bandwidth: 400,000000
> INTER-LAYER object
> SWITCH-LAYER object
  
```

Fig. 10. Details of PCReq message (4b)

```

> PATH COMPUTATION REPLY MESSAGE Header
> RP object
> EXPLICIT ROUTE object (ERO)
  Object Class: EXPLICIT ROUTE OBJECT (ERO) (7)
  >SUBOBJECT: IPv4 Prefix: 10.10.0.1/32
  >SUBOBJECT: IPv4 Prefix: 10.10.0.2/32
  >SUBOBJECT: IPv4 Prefix: 10.10.0.3/32
  >SUBOBJECT: IPv4 Prefix: 10.10.0.4/32
  > EXPLICIT ROUTE object (ERO)
  Object Class: EXPLICIT ROUTE OBJECT (ERO) (7)
  >SUBOBJECT: Unnumbered Interface ID: 10.0.0.3:10
  >SUBOBJECT: Label Control
  >SUBOBJECT: Unnumbered Interface ID: 10.0.0.10:2
  >SUBOBJECT: Label Control
  >SUBOBJECT: IPv4 Prefix: 10.0.0.2/32
  > BANDWIDTH object
  Object Class: BANDWIDTH OBJECT (5)
  Bandwidth: 400,000000
  > SERVER-INDICATION object
  
```

Fig. 11. Details of PCRep message (5b)

are depicted in Fig. 10 and Fig. 11, respectively.

PCReq message (4b) includes an ENDPOINTS object specifying the L2 end-points, i.e., source (DC-A) and leaves (DC-B, DC-C, and DC-D) L2 switches. In addition, the BANDWIDTH object specifies the requested bandwidth for the multicast service (100 Gb/s).

Two OF objects are included in message (4b). The first one refers to the same layer than the request (L2) and specifies that a VNT reconfiguration is requested. The second OF object targets L0 and contains in an embedded TLV the bandwidth for the path computation at that layer; in our experiments, 400 Gb/s L0 LSPs are requested to support the VNT thus, favoring resource utilization. The INTER-LAYER object with the flag I set forces inter-layer path computation, and the SWITCH-LAYER object specifies L2 as well as L0 layers to be considered in the computation. For each layer to be traversed

the encoding and switching type has to be defined: Ethernet encoding and Layer-2 Switch Capable, respectively for L2, and Lambda encoding and Lambda-Switch Capable, respectively for L0.

PCRep message (5b) includes one ERO object for the L2 multicast request. We decided to use loose hops here, since the results of the L2 path computation will be discarded; a subsequent path computation will be requested by the controller to the L2 PCE (messages 10 and 11 in the workflows). After this first ERO object for L2, the list of paths in L0 are included. In the example, one single L0 LSP needs to be created to reconfigure the VNT as shown in the capture in Fig. 11. The path includes a SERVER-INDICATION object specifying L0 and an ERO object with the route (X3-X10-X2) and the spectrum allocation for the 400 Gb/s lightpath to be established.

Finally, the control plane contribution to the provisioning process under both approaches (i.e., messages exchange and path computation algorithms at bPCE) lasted around 150ms.

## V. CONCLUSIONS

Two approaches were proposed to serve large capacity (e.g. 100 Gb/s) multicast connectivity services in a multi-layer scenario where a set of federated DCs is connected through a flexgrid optical network. In the first approach, a VNT was created for each multicast request by establishing light-trees in the flexgrid network. In the second approach, multicast services are served on a multi-purpose VNT supported by 400 Gb/s lightpaths, thus favoring resource utilization.

The feasibility of implementing both approaches using standardized control plane protocols has been studied. In particular, the ASO module was responsible for managing multiple DC networks as a single entity interacting with the ABNO architecture in charge of controlling the interconnection network. PCEP was used among the modules within the ABNO architecture. Workflows were proposed for the considered approaches. Multicast connections (including light-trees) are supported in PCEP using p2mp extensions. However, VNT reconfiguration, entailing multilayer computation is not currently supported; an IETF draft was used as a guide.

Experimental assessment was carried out in a distributed field trial set-up connecting Telefonica (Madrid, Spain), CNIT (Pisa, Italy), and UPC (Barcelona, Spain) premises.

## ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement n° 317999 IDEALIST project.

## REFERENCES

[1] A. Asensio, L.M. Contreras, M. Ruiz, V. López, L. Velasco, "Scalability of Telecom Cloud Architectures for Live-TV Distribution," in Proc. OFC, 2015.  
 [2] L. Velasco, A. Asensio, J. Ll. Berral, E. Bonetto, F. Musumeci, V. López, "Elastic Operations in Federated Datacenters for Performance

and Cost Optimization," Elsevier Computer Communications, vol. 50, pp. 142-151, 2014.  
 [3] "Spectral grids for WDM applications: DWDM frequency grid," ITU-T Rec. G.694.1, 2012.  
 [4] N. Sambo, G. Meloni, G. Berrettini, F. Paolucci, A. Malacarne, A. Bogoni, F. Cugini, L. Potì, and P. Castoldi, "Demonstration of data and control plane for optical multicast at 100 and 200 Gb/s with and without frequency conversion," IEEE/OSA Journal of Optical Communications and Networking (JOCN), vol. 5, pp. 667-676, 2013.  
 [5] M. Ruiz and L. Velasco, "Performance Evaluation of Light-tree Schemes in Flexgrid Optical Networks," IEEE Communications Letters, vol. 18, pp. 1731-1734, 2014.  
 [6] L. Velasco, A. Castro, M. Ruiz, and G. Junyent, "Solving Routing and Spectrum Allocation Related Optimization Problems: from Off-Line to In-Operation Flexgrid Network Planning," (Invited Tutorial) IEEE/OSA Journal of Lightwave Technology (JLT), vol. 32, pp. 2780-2795, 2014.  
 [7] M. Ruiz and L. Velasco, "Serving Multicast Requests on Single Layer and Multilayer Flexgrid Networks," accepted in IEEE/OSA Journal of Optical Communications and Networking, 2015.  
 [8] Ll. Gifre, F. Paolucci, J. Marhuenda, A. Aguado, L. Velasco, F. Cugini, P. Castoldi, O. Gonzalez de Dios, L.M. Contreras, and V. López, "Experimental Assessment of Inter-datacenter Multicast Connectivity for Ethernet services in Flexgrid Networks," in Proc. ECOC, 2014.  
 [9] D. King and A. Farrel, "A PCE-based Architecture for Application-based Network Operations," IETF draft, work in progress, 2014.  
 [10] L. Velasco, A. Asensio, J.Ll. Berral, V. López, D. Carrera, A. Castro, and J.P. Fernández-Palacios, "Cross-Stratum Orchestration and Flexgrid Optical Networks for Datacenter Federations," IEEE Network Magazine, vol. 27, pp. 23-30, 2013.  
 [11] A. Aguado, V. López, J. Marhuenda, O. Gonzalez de Dios, and J. Fernández-Palacios, "ABNO: a feasible SDN approach for multi-vendor IP and optical networks," in Proc. OFC, 2014.  
 [12] E. Crabbe, J. Medved, I. Minei, and R. Varga, "PCEP Extensions for Stateful PCE," IETF draft, work in progress, 2014.  
 [13] JP. Vasseur, and JL. Le Roux, "Path Computation Element (PCE) Communication Protocol (PCEP)," IETF RFC 5440, 2009.  
 [14] L. Velasco, D. King, O. Gerstel, R. Casellas, A. Castro, and V. López, "In-Operation Network Planning," IEEE Communications Magazine, vol. 52, pp. 52-60, 2014.  
 [15] Ll. Gifre, F. Paolucci, A. Aguado, R. Casellas, A. Castro, F. Cugini, P. Castoldi, L. Velasco, and V. López, "Experimental Assessment of In-Operation Spectrum Defragmentation," Springer Photonic Network Communications, vol. 27, pp. 128-140, 2014.  
 [16] Ll. Gifre, F. Paolucci, L. Velasco, A. Aguado, F. Cugini, P. Castoldi, and V. López, "First Experimental Assessment of ABNO-driven In-Operation Flexgrid Network Re-Optimization," (Invited Paper) IEEE/OSA Journal of Lightwave Technology (JLT) DOI: 10.1109/JLT.2014.2343157, 2014.  
 [17] E. Crabbe, I. Minei, S. Sivabalan, and R. Varga, "PCEP Extensions for PCE-initiated LSP Setup in a Stateful PCE Model," IETF draft, work in progress, 2014.  
 [18] K. Shiomoto, D. Papadimitriou, JL. Le Roux, M. Vigoureux, D. Brungard, "Requirements for GMPLS-Based Multi-Region and Multi-Layer Networks (MRN/MLN)," IETF RFC 5212, 2008.  
 [19] Q. Zhao, et al., "Extensions to the Path Computation Element Communication Protocol (PCEP) for Point-to-Multipoint Traffic Engineering Label Switched Paths," IETF RFC 6006, 2010.  
 [20] E. Oki, T. Takeda, J-L Le Roux, A. Farrel, F. Zhang, "Extensions to the Path Computation Element communication Protocol (PCEP) for Inter-Layer MPLS and GMPLS Traffic Engineering," IETF draft work-in-progress, 2014.  
 [21] OpenStack. <http://www.openstack.org/>  
 [22] Floodlight <http://www.projectfloodlight.org/>  
 [23] F. Cugini, F. Paolucci, G. Meloni, G. Berrettini, M. Secondini, F. Fresi, N. Sambo, L. Potì, and P. Castoldi, "Push-pull defragmentation without traffic disruption in flexible grid optical networks," IEEE/OSA Journal of Lightwave Technology (JLT), vol. 31, pp. 125-133, 2013.  
 [24] Ll. Gifre, L. Velasco, N. Navarro, and G. Junyent, "Experimental Assessment of a High Performance Back-end PCE for Flexgrid Optical Network Re-optimization," in Proc. OFC, 2014.