Sgambelluri, A., Dugeon, O., Muhammad, A., Martín-Pérez, J., Ubaldi, F., Sevilla, K., de Dios, O. G., Pepe, T., Bernardos, C. J., Monti, P. & Paolucci, F. (2019). Orchestrating QoS-based Connectivity Services in a Multi-Operator Sandbox. *Journal of Optical Communications and Networking, 11*(2), A196-A208.

DOI: 10.1364/jocn.11.00a196

Orchestrating QoS-based Connectivity Services in a Multi-Operator Sandbox

A. Sgambelluri, A. Muhammad, J. Martín-Pérez, F. Ubaldi, K. Sevilla, O. Dugeon, O. G. De Dios, T. Pepe, C. J. Bernardos, P. Monti, F. Paolucci

***Abstract*—One of the main features of 5G networks is the coordinated orchestration of both IT and connectivity resources. This enables the deployment of a flexible and programmable architecture able to provision end-to-end services with different (and sometimes quite) stringent Quality of Service (QoS) constraints. In this paradigm service orchestration may take place over single and multiple administrative domains. The 5G Exchange (5GEx) project building on the Software-Defined Network (SDN) and the Network Function Virtualization (NFV) concepts, targets the design and the implementation of a multi-domain orchestrator (MdO) prototype for the automatic provisioning of Network Service (NS) across multiple administrative domains.**

This paper presents the architectural solution, designed and implemented in the context of 5GEx Multi-domain Orchestrator (MdO) prototype, that can be used to establish end-to-end connectivity tunnels with QoS constraints (i.e., bandwidth and end-to-end delay) connecting VNFs deployed in remote data-centers controlled by different providers. The proposed solution has been experimentally validated in terms of scalability, reliability and end-to-end workflow. Results show how the designed solution permits the automatic establishment of QoS-based end-to-end tunnels spanning across multi-technology, multi-operator network domains. The orchestration scheme does not present scalability problems neither for the advertisement of resources nor for the provisioning of connectivity services. Moreover, no issues have been identified from the reliability point of view.

***Index Terms*—QoS-based connectyvity services, BRPC, 5GEx Project, Multidomain Orchestrator, 5GEx**

I. Introduction

5G networks are envisioned as connect-and-compute infrastructures equipped with extremely high flexibility and programmability functionalities that allow them to provision services in end-to-end (E2E), application-, service-, and location- aware fashion. 5G networks are also expected to enable new business opportunities by meeting the requirements of a large variety of use cases by means of (i) implementing network slicing in cost efficient way, (ii) addressing both end user and operational services, (iii) supporting softwarisation natively, and (iv) integrating communication and computation operations [2].

In order to accommodate 5G paradigm, infrastructure providers are expected (in a few years) to be able to deliver new services (e.g., Infrastructure as a Service (IaaS)) and advanced virtualized function chaining capabilities that require the joint provisioning of both connectivity and information technology (IT) (i.e., storage and compute) resources. Leveraging on Network Function Virtualization (NFV) techniques and on Software Defined Networking (SDN) orchestration, infrastructure providers will also be able to slice their connectivity and IT resources so that they can be assigned to different tenants for their use [3].

Function chaining and slicing operations may take place within the administrative domain of the same provider (i.e., *single domain* orchestration) or, depending on the tenant request, the provisioning of resources may span over multiple infrastructure providers, i.e.,the chaining/slicing operations may require geographically distributed resources orchestrated in a *multi-domain* fashion [4].

When looking at function chaining and slicing operations, another crucial aspect to consider is the Quality of Service (QoS) requirements. Regardless whether resources are orchestrated in a single or in a multi-domain environment, an operator needs to make sure that the latency, bandwidth and resiliency requirements (i.e., just to name a few) are always met while a given service is in operation.

In a multi-domain environment, the Multi-domain Orchestrator (MdO) [5] plays a key role offering and managing services to a federation of infrastructure providers. More specifically, by interacting among themselves the deployed MdOs (i.e., one in each single network infrastructures provider) orchestrate the service provision phase and guarantee that the Service Level Agreements (SLAs) are met by proactively monitoring and reconfiguring (if/when needed) the running services [4].

The workflows and procedures for efficiently establishing multi-provider services is still an open research topic, which, in turn, is particularly challenging from the standpoint of the provisioning of connectivity resources. Most of the works available in the literature focus on multi-domain lightpath and Label Switched Paths (LSP) establishment and rely on the Hierarchical Path Computation Element (HPCE) concept [6], [7]. However, due to business models and trust issues, inter-operator traffic engineering may not allow a third-party neutral orchestrator. For this reason a peer-to-peer approach may represent a more reasonable and feasible approach. Furthermore, ensuring service availability, continuity, as well as delivering the promised quality to customers in a multi-domain environment increases the importance of scalability and resiliency aspects of the orchestrator framework.

An initial study on the scalability and reliability performance of an inter-operator orchestration framework based on the MdO developed in the H2020 European 5G Exchange project [8] was presented in [1]. On the other hand, this work did not provide any detailed architectural solution for the orchestrator nor it presented or validated any workflow

Manuscript received August 3, 2018. A. Sgambelluri, F. Paolucci are with Scuola Superiore Sant'Anna, Pisa, Italy.

A. Muhammad, P. Monti are with KTH Royal Institute of Technology, Kista, Sweden.

J. Martin-Pérez, C. J. Bernardos are with Universidad Carlos III de Madrid and IMDEA Networks Institute, Madrid, Spain.

F. Ubaldi, T. Pepe are with Ericsson Research, Pisa, Italy.

K. Sevilla, O. Dugeon are with Orange Labs, Lannion, France.

O. G. De Dios is with ID Telefonica, Madrid, Spain.

This paper is an extended version of the work presented in [1].

This work has been supported by the EU H2020 5GEx Project (contract number 671636).

description that could be used to establish QoS-based connectivity services among multiple administrative domains.

This paper, on the other hand, presents the architecture and the orchestration procedure, designed and implemented in the context of the 5GEx MdO prototype, that can be used to establish end-to-end connectivity tunnels (with QoS constraints) connecting VNFs deployed in remote data-centers controlled by different providers. In the proposed architecture advertising operations are carried out via a topology advertisement module (i.e., referred to as TADS) based on an extension of the BGP-LS protocol that includes IT information [9]. QoS-based connectivity services are provisioned via a multi-domain PCE module resorting to the stateful Backward Recursive PCE-based Computation (BRPC) [10] procedure. Taking advantage of the 5GEx Sandbox (i.e., a large multi-domain European network connecting 15 different lab premises of operators and research centers [4]) the paper also reports results from the experimental validation of the workflow proposed for the tunnel establishment. The set of results is then completed by scalability and resiliency measurements conducted in terms of the following operations: (i) resource announcement (for both connectivity and IT), and (ii) computation and provisioning of QoS-based connectivity services spanning multiple domains, infrastructure providers, and network operators. Results show that the proposed orchestration scheme does not present scalability problems neither for the advertisement of resources nor for the provisioning of connectivity services. Moreover, no issues have been identified from the reliability point of view.

II. Related Work on Multi-domain Orchestration

There are a number of research works in the literature that look into SDN-based orchestration methods for the management of network infrastructures with heterogeneous resources in a multi-domain scenario. The work in [11] presents a management strategy for a multi-technology transport networks where a centralized controller acts as the hypervisor in charge of the management of Virtual Optical Networks (VONs) serving multiple tenants.

The work in [12] studies methods for the migration of data center resources over a multi-domain transport network. The solution proposed in the paper is based on a multi-controller collaboration scheme that coordinates the resource migration process over a number of wide area networks (WANs). A solution for controlling each one of these WANs, and for properly integrating the WAN controller with the data center controllers is presented in [13]. This latter work presents and validates an architecture that integrates the OpenFlow and the Generalized Multi-Protocol Label Switching (GMPLS) protocols to achieve an interworking solution that relies on the HPCE concept. As a result, the paper delivers an industry-ready architecture with service instantiation and topology discovery capabilities.

There are other works in the literature where the HPCE concept is applied. Some of them focus on Elastic Optical Networks (EONs). In [7] the authors consider a HPCE architecture for EONs with a GMPLS/PCE control plane. They propose a scheme for updating the traffic engineering database at the parent PCEs (pPCEs) using Link State extensions to the Border Gateway Protocol (BGP-LS). Their simulation work evaluated the pPCE control load, the lightpath setup time, and the lightpath blocking probability.

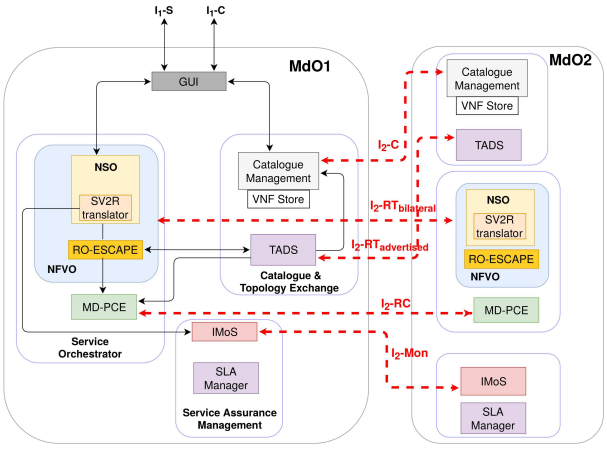

Fig. 1: Functional architecture and interfaces of the 5GEx Multi-domain Orchestrator.

Their results show that the proposed scheme reduces the pPCE control load and achieves lower blocking with respect to standard PCE Protocols and BGP-LS schemes. In [6] the authors implemented, evaluated, and validated a control plane solution (based on a HPCE architecture), for a multi-domain EON. They present experimental results for path computation and LSP setup times. These results represent an important benchmark for future research activities and extensions in the context of multi-domain optical networks.

Along with the management and control of optical transport networks, it is important to make sure that connections established over multiple domains satisfy the required QoS level, e.g., bandwidth, latency, jitter. The work in [4] presents an MdO architecture able to provision services over a federation of service provider while ensuring that the SLAs of the provisioned services are within acceptable limits. In fact, the MdO architecture presented in this work includes a number of components that can be used to proactively guarantee the required SLAs between service providers. Both the work present in [4] and the one presented in this current paper are developed in the context of the H2020 European research project 5G Exchange (5GEx) [8]. More details about the functional architecture and about the the 5GEx MdO features are provided next.

III. 5GEx MdO Architecture and Features

The heart of the 5GEx framework is the multi-domain orchestrator (MdO) that coordinates resources and/or service orchestration at the multi-technology and/or multi-operator level. Fig. 1 shows the 5GEx MdO functional architecture along with the intra- and inter-MdO interfaces. Functionality wise, the 5GEx architecture can be divided into three blocks: Catalogue and Topology Exchange (CTE), Service Orchestration (SO), and Service Assurance Management (SAM) [9]. The CTE block is responsible for exchanging (abstracted) topology information and MdO service level capabilities with other MdOs. In other words, the CTE block gathers the information essential for service deployment. The SO block performs the actual service deployment based on the service requirements and the information provided by the CTE block. Once a service is deployed, the SAM block makes sure that the performance of the service (across the different domains) are within the SLA requirements. Fig. 2

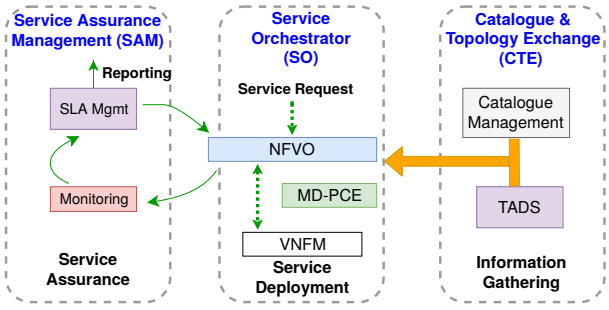

Fig. 2: General operation workflow of the 5GEx Multi-domain Orchestrator.

illustrates a simplified version of the operational workflow of the 5GEx MdO and explains the objective of the different functional blocks, i.e., information gathering, service deployment, and service assurance. The rest of this section describes briefly the main modules within each functional blocks.

A. Catalogue and Topology Exchange

The CTE block comprises the Topology Abstraction and Discovery Subsystem (TADS) and the Catalogue Management Subsystem (CMS). They both actively communicate with intra- and inter-MdO modules to exchange the resource information.

1) TADS: The TADS is responsible for discovering the MdOs of other providers and for exchanging information about the type (i.e., connectivity/IT) and quantity of resources within the domain administered by a remote MdO. This information exchange takes place via the I2-RTadvertised interface. The information gathered by the TADS is stored and maintained in a dynamic database from which it is sent to other subsystems at the local MdO through different intra-MdO interfaces. More precisely, the TADS database contains: (i) the entry points (i.e., URLs) of the remote MdOs the TADS is in contact with, (ii) the entry points of the MD-PCEs at each remote MdOs, (iii) an updated view of the connectivity resources available at each remote MdOs, and (iv) an updated view of the IT resources (i.e., CPUs, storage, and memory) available at each remote MdOs. The TADS comprises the following set of modules: (i) the Topology Module (TM), (ii) one Traffic Engineering Databases (TEDs) for each remote MdO the TADS is in contact with. Each TED maintains an updated status of the resources (i.e., connectivity and IT) available at the remote MdO, (iii) a BGP-LS Plugin (Speaker) for the I2-RTadverised interface, (iv) a Local Plugin for the intra-MdO interface with MD-PCE at the local MdO, and (v) REST Plugins for communication with the CMS and the NFVO at the local MdO.

2) CMS: The CMS is responsible not only for managing the repository with the list of network services that can be provided locally, but it also collects information from the CMS of each one of the remote MdOs in the provider federation (i.e., to understand what are the network services that are offered by each remote provider). The exchange of catalogue information with remote MdOs takes place over the I2-C interface. The CMS communicates with the local TADS to keep an updated list of the remote MdOs and to understand how to access their local catalogue information. The catalogue inside the local MdO is used for internal management, i.e., composing new network services, combining different items from the catalogue, adding new items coming from remote domains and adapting them to the local domain. The CMS adds elements from a remote catalogues to the local one only after a process that comprises testing, validation, and SLA negotiation for the specific network service functionality to be imported.

B. Service Orchestration

The SO block includes the Network Function Virtualization Orchestrator (NFVO) and the Multi-Domain Path Computation Element (MD-PCE), i.e., the two key modules that are responsible for the deployment of a service.

1) NFVO: The NFVO is composed by the Network Service Orchestration (NSO) and the Resource Orchestrator (RO). The NSO is responsible for handling those network service requested by the customers or by the NSO of a remote MdO. The RO oversees the embedding of the resource requested by a given service into the respective resource domains (local and remote). To effectively perform this role, RO tightly cooperates with TADS to dynamically discover resources at remote MdOs.

2) MD-PCE: The MD-PCE is responsible for establishing QoS-based connectivity services utilizing the multi-domain topology information provided by TADS. To perform this task, the MD-PCE interacts with different subsystems, i.e., with the local RO and TADS with the MD-PCE at remote MdOs. More details about how the MD-PCE sets up a QoS-based connectivity service are provided in Sec. IV.

C. Service Assurance Management

The SAM block comprises the Intelligent Monitoring Subsystem (IMoS) and the SLA Manager, which are responsible for the assurance of the deployed services.

1) IMoS: The IMoS is in charge of the coordinated end-to-end monitoring required for the management and orchestration of services across the multi-domain federation. The IMoS collects monitoring information of every deployed service across the multi-domains. This operation is done via the I2-Mon interface that connects the local IMoS module to IMoS subsystems at the remote MdOs. The monitoring data is processed for aggregation to KPI values that are then stored in a central monitoring database.

2) SLA Manager: The SLA Manager evaluates the selected KPIs for each active network service according to an agreement that is established and signed upon the service instantiation. For this purpose, the SLA Manager retrieves (periodically and/or on demand) the measurements from the central monitoring database of the IMoS. Based on the collected measurements, the SLA Manager checks/calculates whether the SLA requirements (initially set for that service) are satisfied and reports the results to the costumer and/or other MdO management entities, e.g., for reconfiguration.

IV. Use case definition

This section provides a more detailed description of the use-case analyzed in the paper, i.e., the provisioning of a connectivity service over multi- technology and administrative domains. The description includes details on the scenario, the functionalities of the MdO components involved in the orchestration of the service, and the related workflow.

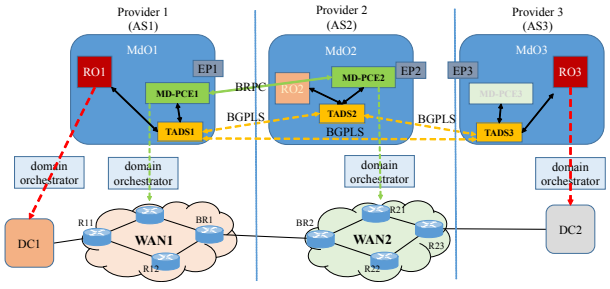

Fig. 3: General scenario with 3 operators.

A. Scenario description

The use-case under exam consists of interconnecting a set of compute resources distributed over different data center (DC) providers via a number of transit domains managed by different network operators (Fig. 3). More specifically, Provider 1 controls one data center (i.e., DC1) and a WAN transport network (i.e., WAN1). Provider 2 is responsible for another WAN transport network (i.e., WAN2), while Provider 3 controls a second data center (i.e., DC2) connected, in turn, to the WAN2.

According to the customers requests, connectivity resources are dynamically deployed over the two transport network domains with specific QoS constraints (i.e., guaranteed bandwidth or end-to-end delay). The inter-operator connectivity, adopted to interconnect the VNFs deployed in the two data-centers (i.e., DC1 and DC2), is done according to both the concepts of Assured Service Quality (ASQ) path and the on-demand Value Added Connectivity (VAC) enterprise [14].

Regarding the implementation of the dataplane interconnectivity, we rely on the existing 5GEx SandBox [14]. For the implementation of the orchestration plane, our focus is on the MD-PCE and TADS, that are the functional blocks devoted to the setup of the QoS-based connectivity services. In fact, the considered use-case demonstrates how the MD-PCE and the TADS are adopted to perform the deployment and setup of dynamic connectivity with QoS constraints.

B. TADS component

As described in Sec. III, the role of the TADS is to discover the resources of the other 5GEx providers connected to the community. By means of the BGP-LS protocol, each TADS exports: (i) an abstracted topology representing the underlying WAN network with, in addition, a representation of inter-domain links, and (ii) an aggregated view of the IT resources available in the local domain, together with the local MdO entry-point. More specifically, when looking at the specific use case in Fig. 3, TADS1 exports to TADS2 and TADS3 the info related to Provider 1: the abstracted topology of WAN1, the overall amount of IT resources in DC1, and the entry-point of MdO1, i.e., EP1. TADS1, on the other hand, receives (from TADS2) the information related to Provider 2 (i.e., abstracted topology of WAN2 and the entry-point EP2) and from TADS3 the information about Provider 3 (i.e., the overall amount of IT resources in DC2 and the entry-point EP3). At the end of the information gathering procedure the three TADSs in the example will have the same global view of the system, including the details of all the three providers. Each TADS, then, exports the abstracted topology information to its local MD-PCE, while the information about the available IT resources at each provider (including the MdO entry point) are sent to the local RO.

C. MD-PCE component

The MD-PCE needs both intra-domain and inter-domain topology information in order to be able to compute and enforce QoS-based connectivity paths. The MD-PCE can directly acquire details on the intra-domain topology by receiving BGP-LS messages from a speaker in the local domain (i.e., one of the routers designated to export the topology information). To collect information on the inter-domain topology, the MD-PCE relies on the local TADS, that provides such information through a dedicated intra-MdO BGP-LS session. The received inter-domain topology includes the abstract view of the local WAN domain, the full details of the inter-domain links (i.e., BGP routers and inter-domain links) and finally the information related to MD-PCEs at remote MdOs. This latter information is used by the MD-PCE to establish I2RC session (i.e., PCEP session implementing BRPC) with MD-PCEs at remote MdOs. On the other hand, the other pieces of information are used to compute the AS PATH, i.e., select which domain will be involved in the end-to-end connectivity.

D. Workflow for establishing QoS-based connectivity

This section discusses the detailed workflow for the establishment of QoS-based end-to-end connectivity in the use case described in Fig. 3.

As shown in Fig. 4, when a new inter-connectivity request with QoS constraints arrives at MD-PCE1 (i.e., via a dedicated REST API) (step1), the MD-PCE1 performs the AS path computation, according to the information received by TADS1 and selecting a path (among the ones available) matching required QoS parameters (bandwidth and/or delay). At step 2 the stateful Backward Recursive Path Computation (BRPC) procedure (i.e., as per RFC 5441 [10]) is activated involving MD-PCE1 and MD-PCE2 for the computation of the end-to-end path connecting ingress and egress points. In turn, MD-PCE1 receives the MD-PCE2 part of the end-to-end path and then merges this part with its local path computation to obtain the overall end-to-end path. Once the path is computed, MD-PCE1 activates the instantiation of the computed path. More specifically, MD-PCE1 sends a PCInitiate message to MD-PCE2 (step3), that interacting (step 4) with the ingress router (i.e., BR2) activates the path related to WAN2 and selects the stitching label. A PCReport message with the selected stitching label is sent back to MD-PCE2 (step 5), that recursively sends a PCReport message with the selected stitching label to MD-PCE1 (step 6). Once the message is received, MD-PCE1 sends a PCInitiate message to the ingress router of the connectivity with its local path and the stitching label (step7). The instantiation of the end-to-end inter-domain connectivity is completed by sending (step 8) a PCReport to MD-PCE1. Finally, the connectivity acknowledgement is sent to the customer over the REST interface (step9).

V. Performance Assessment

This section presents the considered performance assessment, including both scalability and resiliency aspects and the complete end-to-end validation,.

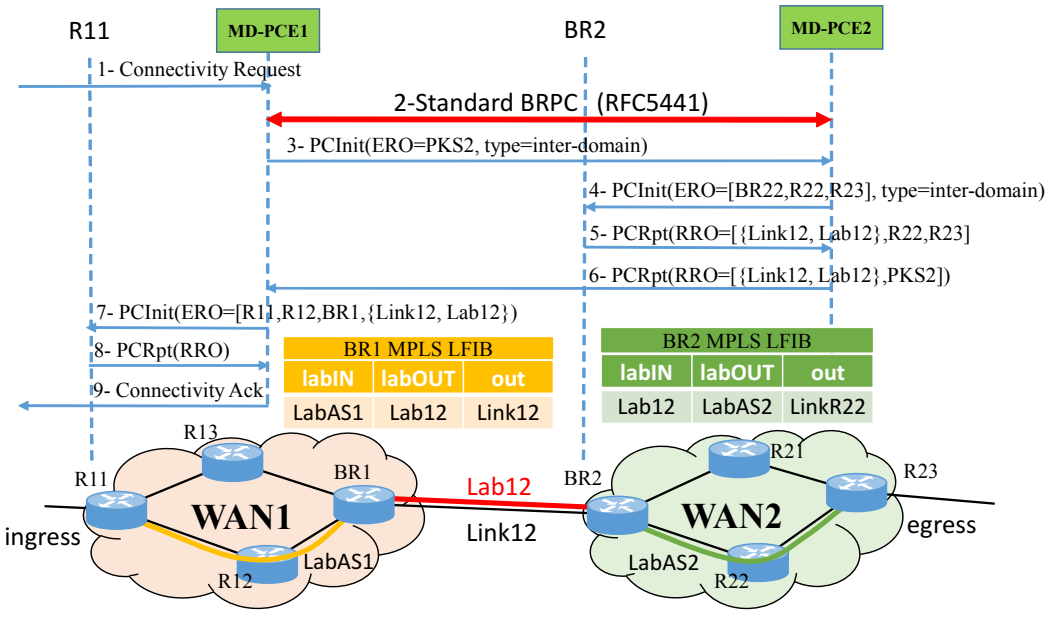

Fig. 4: Workflow for the establishment of QoS connectivity.

A. *Components Scalability*

1) TADS: To measure the scalability performance of the resource advertisement operations we tested TADS on a subset of the islands of the 5GEx Sandbox (Fig. 5). Besides the TADS, the test environment includes a topology generator that creates artificial topologies of different dimensions according to the BRITE-based Waxman methodology [15]. To test scalability with respect to the overall topology dimension, the nodal degree of topologies generated artificially is set to 6 while the number of nodes is varied from 10 up to 250. The generated topologies are exchanged with other TADS using BGP-LS in a chain configuration. The TADS scalability is evaluated by measuring the time required to distribute multi-domain resource information (i.e., discovery procedure) in the Sandbox, i.e., the time needed by a remote TADS to fully synchronize in its database the topology created and advertised by the generator with respect to the first TADS of the chain. The TADS update timer is set to 5s while the maximum round trip time (RTT) for pinging TADS2 and TADS3 from TADS1 is measured to be 60ms. Fig. 6 shows the results for the TADS convergence time as a function of the size of the topology exported to other TADS peers. The experimental set up considers a two (Pisa - Stockholm) and a three (Pisa - Stockholm - Madrid) cascade domain scenarios, thus emulating the three-level Tier-based hierarchical AS internet topologies. Considering the dimension of the 5GEx Sandbox (15 operator domains), if each local domain (node) is designed as a full mesh connected with 6 other nodes, we can reach close to 100 nodes. Taking this into account, on average, the number of domains traversed by the resource information will be less than 3. The results in Fig. 6 indicate that the time needed to advertise the full 5GEx topology in the Sandbox environment is shorter than 10s. However, for a realistic scenario where the domains send few messages to advertise updates about the critical nodes and links, the convergence time is very short, i.e., below 1s. Furthermore, the TADS convergence time rises exponentially as the number of nodes increases, especially for 3 domains scenario. This is

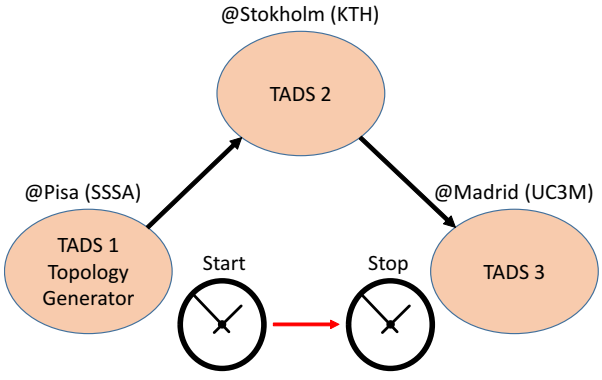

Fig. 5: TADS convergence time setup.

because of the relatively high value (i.e., 5s) of the update timer. This convergence time can be significantly reduced by employing a throttling-based approach, i.e., sending rate-limited updates (to other TADS) upon reception of resource information instead of a fixed timer based strategy. The reported convergence times demonstrate the importance of network and resource abstraction in large multi-domain environments and motivate its adoption for scalability reasons.

2) MD-PCE: The scalability performance of the MD-PCE component has been evaluated considering two different aspects: *(i)* the overall time required to setup an LSP tunnel and *(ii)* the time required to setup a full-mesh of tunnels between all the PE routers present in the topology of Fig. 7. In the latter case, we repeated 100 times the setup of all the LSPs, recording the required deployment time (minimum, maximum and, average). The testbed topology used for performance and scalability measurement of the MD-PCE, is composed by P (i.e., transit) and PE (i.e., Provider Edge) routers from different vendors (e.g., Cisco and Juniper) that are able to act as a Path Computationl Client (PCC) and establish a PCE Protocol (PCEP) session with the MD-PCE.

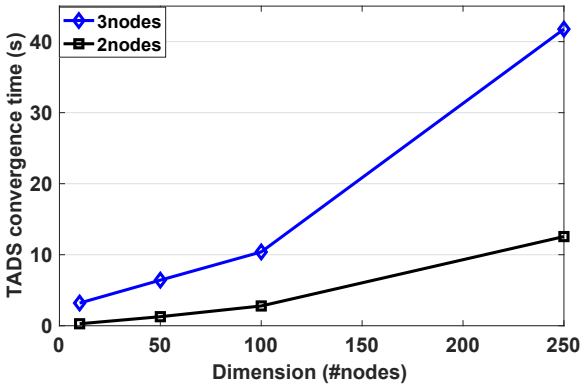

Fig. 6: TADS convergence time scalability.

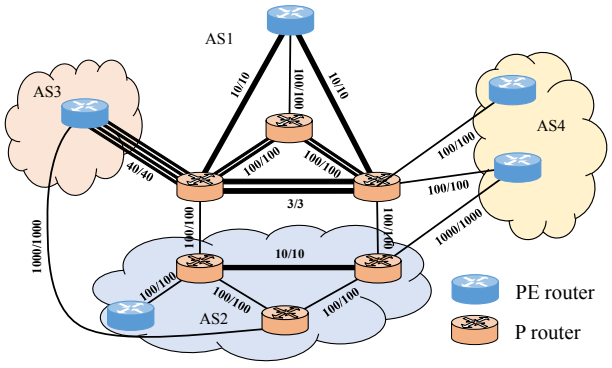

Fig. 7: MD-PCE scalability testbed.

The first evaluation has been performed at the North Bound interface of the MD-PCE, using the HTTP REST API designed to receive connectivity requests. The measurement is done between the HTTP REST POST message and the HTTP 200 OK answer.

The system takes a little bit more than 100 ms to perform the deployment of an LSP. More specifically, the MD-PCE contribution is around 4 ms including 2 ms for the Path computation, and 2 ms needed for storing the LSP in the OpenDayLight DataStore. Then, the PE router takes 2 ms to trigger the RSVP PATH message after the reception of the PCInitiate message from the MD-PCE. The larger time contribution is due to the RSVP process, directly dependent to the path selected for the LSP. To complete the procedure the PE router sends two PCEP PCReport messages: the first one indicates that the tunnel is deployed and it is administratively up, while the second one indicates that it is operationally up. The MD-PCE sends the HTTP 200 OK answer to the HTTP REST POST request once the first PCReport message is received.

In order to avoid that PE routers become bottleneck for the setup of several LSP tunnels, an ad-hoc python script has been developed to send HTTP REST POST request to the MD-PCE in a reverse order, i.e., identifying the egress PE, we asked all ingress PE routers to setup a tunnel to it. We repeated 100 times, the setup of full mesh LSP tunnels (i.e., 2500 LSPs in total, resulting intoto 25 tunnels per PE router). The total average time was around 189 seconds, the minimum time was 150 seconds while the maximum was 252 seconds, resulting into around 13,2 LSPs deployed in a second. During the 100 test, no significant variation in

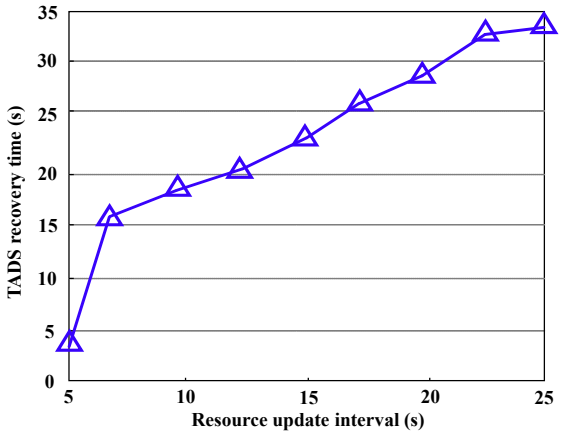

Fig. 8: TADS recovery time.

memory consumption and CPU utilization has been detected at the machine dedicated to the MD-PCE during the establishment of the 2500 LSPs. After each test we deleted all the instantiated LSPs. The complete deletion process took less than 20 seconds (i.e., deletion is 10 times faster compared to the creation). In fact, when PE routers delete an LSP tunnel, they reply to the MD-PCE immediately with a PCReport message, without waiting for the answer related to the RSVP TEAR message.

B. Resiliency

1) TADS: Handling up-to-date information in the TADS is crucial, in order to provide consistent information to the orchestration functional modules. Each TADS periodically refreshes the information distributed among the TADS in the community, by sending BGP-LS updates to its peers. At any given TADS, if no updates are received for a specific resource, the information expires and it is deleted from its database. To test the reliability of TADS, the scenario in Fig. 3 is considered. Each TADS exchanges the information about its local domain with the other two TADSs. We then cut the peering session between TADS1 and TADS3. After a predefined amount of time the information related to Provider 3 domain is deleted from the TADS1 database because it has become obsolete. In this new scenario the information exported by TADS3 is propagated via TADS2 to TADS1, in order to make sure that TADS1 still has the correct full view of the system. Fig. 8 presents the time required by TADS1 to receive TADS3 domains information via TADS2. This information is reported as a function of the time interval between resource update messages. The results show that increasing the period of the BGP-LS update messages significantly impacts the time required to recover the information, demonstrating the importance of properly set the time gap between the BGP-LS update messages. The results also highlight an interesting tradeoff between network management overhead and information recovery time.

2) MD-PCE: MD-PCE has been designed to be resilient, since it is based on the PCE architecture. In fact, when choosing a multi-instance behaviour it is possible to set a master MD-PCE and and slave MD-PCE. In normal condition, edge routers report tunnels information to both MD-PCE, but only the master MD-PCE is allowed to manage

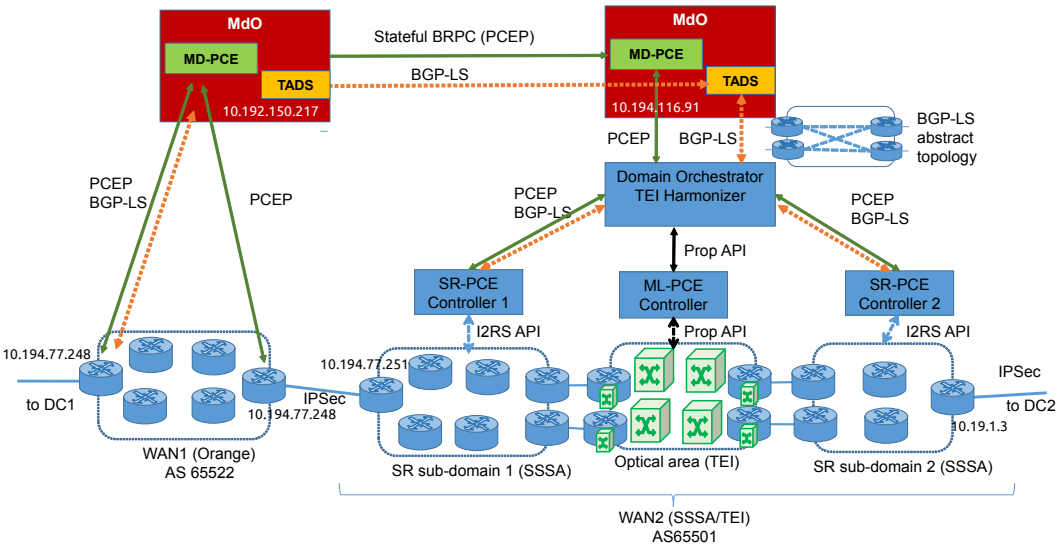

Fig. 9: Experimental Multi-domain scenario with Orange and SSSA/TEI islands.

the tunnels. In case the PCEP session between a router and master MD-PCE fails, the slave MD-PCE receives the tunnels delegation, becoming able to manage the tunnels. In addition, considering the MD-PCE is implemented in OpenDayLight, the clustering behaviour can improve the MD-PCE reliability. If case of failure of one MD-PCE, another instance of the cluster will take the leadership. Moreover, by using a floating IP address for the cluster, the leadership change is completely transparent for the edge routers. In order to evaluate the resiliency of MD-PCE, we deployed two instances of the MD-PCE (i.e., master and slave), configuring the edge routers to establish a PCEP session to both servers. Then, we deactivated the link used to connect the master MD-PCE to the testbed, verifying that all edge routers switched to the slave MD-PCE and that the slave instance of MD-PCE could also manage the tunnels created by the master MD-PCE.

C. End-to-end QoS-based connectivity validation

The workflow depicted in Sec. IV-D has been experimentally validated involving two islands of the 5GEx sandbox, the Orange island and the SSSA/TEI island.

The Orange island includes an IP/MPLS WAN domain, based on commercial Juniper and Cisco routers, configured with legacy WAN protocols: IS-IS-TE routing protocol, RSVP-TE signalling protocol, Segment Routing, and PCEP protocol. Edge routers are configured as Path Computation Client. The MD-PCE, based on OpenDayLight controller, runs as Path Computation Element collecting topology information and setting up PCEP sessions.

The SSSA/TEI island is a multi-technology, multi-AS domain composed by three sub-domains: two edge IP/MPLS domains equipped with Juniper metro-core routers encompassing Segment Routing functionalities [16]–[18], and one internal multi-layer MPLS-over-optical domain made of Polatis switches, handled by a distributed commercial GMPLS control plane, including OSPF-TE and RSVP-TE.

The whole experimental scenario for the two islands involved is depicted in Fig. 9.

The Orange island is handled by a dedicated MdO [4], that includes TADS and MD-PCE (IP address 10.192.150.217). The MD-PCE performs two specific functionalities. From one side it acts as a controller running the Path Computation Element, that receives intra-domain topology information via BGP-LS from the dataplane router and sets up PCEP sessions with the PE routers. From the other side the MD-PCE receives the multi-domain topology by the TADS through BGP-LS.

The SSSA/TEI island is managed by a single MdO (including TADS and MD-PCE with IP address 10.194.116.91) and, differently from the Orange island, by the Harmonizer Domain Orchestrator provided by TEI. In fact, according to the 5GEx architecture, each sub-domain is handled by a local controller. In particular, the SR-PCEs developed by SSSA handles two SR sub-domains, while an industrial commercial PCE is in charge of controlling the optical sub-domain. All controllers are connected to the TEI Harmonizer Domain Orchestrator by means of BGP-LS and PCEP protocols for topology exchange notifications and path computation procedures, in a hierarchical fashion (i.e., exploiting the hierarchical PCE architecture [6]). The exception is the TEI optical domain that uses a proprietary interface given by the commercial GMPLS control plane. The Harmonizer provides an abstracted BGP-LS vision of the whole domain as a single domain to the MD-PCE of the SSSA/TEI MdO. In particular, a full mesh of abstract links connecting the SSSA/TEI edge nodes is advertised to the MD-PCE via BGP-LS protocol. Each abstract link encloses QoS parameters (i.e., delay, loss, jitter) that help the MD-PCE to select an assured service quality path inside the domain. The whole hierarchical PCEP infrastructure supports stateful computation with instantiation capabilities. In this way, PCEP is used not only for computing paths but also for triggering the path management and instantiation (i.e., setup, teardown and update). Interface to the Routing System (I2RS) API [19] is adopted at each SR-PCE controller in order to provide the computed Segment List and install the label actions directly in the router configurations. The same API is also exploited at the beginning of the experiment in order to provide SR forwarding scheme derived from routing information, provided by the OSPF protocol running inside the SR domains. An

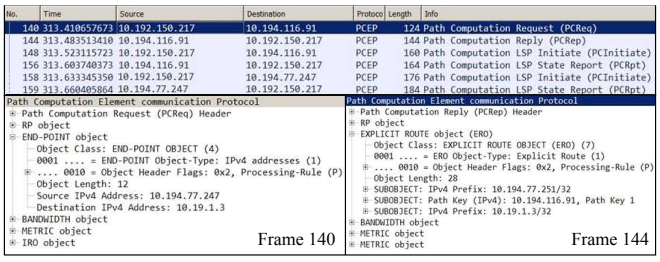

Fig. 10: Orange MD-PCE: capture of the PCEP messages.

optical domain with GMPLS control plane is used to show the compatibility of the system with a legacy domain of a different technology. In fact, the control plane of this sub-domain is an old system currently adopted in production networks by operators.

During the considered experimental validation, we requested to the Orange MD-PCE (with IP 10.192.150.217) a multi-domain path from node 10.194.77.247 (at the ingress of the Orange backbone) to node 10.19.1.3 (at the egress of SSSA/TEI backbone) with 10Mbps of bandwidth. In order to present the interaction among the MD-PCEs involved in the deployment of a QoS-based multi-domain path, in Fig. 10 the capture of PCEP messages collected at the Orange MD-PCE is shown. After receiving the request and computing the AS path, the Orange MD-PCE activates the BRPC procedure, sending a PCReq message to the SSSA/TEI MD-PCE. The message (i.e., frame 140) is exploded in the bottom-left part of the figure highlighting the details of the request. After around 70ms, the SSSA/TEI MD-PCE replies with a PCRep message. The details of the reply message are exploded in the bottom-right part of the figure. More specifically, the selected path, identified with Path Key 1, presents the node 10.194.77.251 as source and node 10.19.1.3 as destination. Then, Orange MD-PCE computes its intra-domain path (from node 10.194.77.247 to node 10.194.77.248) and obtains the end-to-end path by merging local path to the one coming from SSSA/TEI MD-PCE in around 40ms. To activate the computed path, Orange MD-PCE sends the first PCInit message to SSSA/TEI MD-PCE (frame 148) and then, after receiving the PCRep message (frame 156) form SSSA/TEI MD-PCE, it sends a PCInit to the ingress router (i.e., 10.194.77.247) of local path (frame 158). Receiving the PCRep message from of the router, the procedure is complete. The overall time required to coordinate the end-to-end deployment takes around 250 ms.

In the rest of the section, the overall procedure for the deployment of the end-to-end QoS-based connectivity is described, including the detailed intra-domain workflow of the two islands. The Orange MD-PCE receives the connectivity request via REST interface, then:

1) The Orange MD-PCE computes the AS path and activates the BRPC procedure sending PCEP `Request` message to the SSSA/TEI MD-PCE.
2) When the BRPC request arrives at the SSSA/TEI MD-PCE, a PCEP `Request` message is sent to the Harmonizer to compute the path between two selected edge nodes;
3) The Harmonizer replies with a PCEP `Reply` message to the SSA/TEI MD-PCE providing the details of the computed path.
4) The SSSA/TEI MD-PCE continues the BRPC procedure, sending a PCEP `Reply` to the Orange MD-PCE to complete end-to-end path computation.
5) The Orange MD-PCE receives the PCEP `Reply` and, acting as intra-domain PCE, computes the intra-domain path, that together with the SSSA/TEI path realizes the overall end-to-end connectivity with QoS constraints.
6) The Orange MD-PCE sends to SSSA/TEI MD-PCE the PCEP `Initiate` message in order to activate the deployment of the tunnel.
7) The SSSA/TEI MD-PCE sends a PCEP `Initiate` message to the Harmonizer to setup an LSP between two selected edge nodes;
8) Harmonizer splits path computation in three sub computations, identifying the edges of each sub-domain and computing label assignment for each incoming and outcoming label. Then, it issues three `Initiate` messages to the controllers.
9) Controllers perform path computation in their local domain. In case of success, they send a preliminary PCEP `Report` message with the computed `ERO` and the LSP operational flag set to *Going Up*.
10) Controllers perform LSP instantiation in their respective sub-domains, using dedicated APIs (i.e., I2RS in the SR-domains, and PCEP/GMPLS with RSVP-TE in the optical domain). Once the configuration is successfully performed, the controllers send a PCEP `Report` message to the Harmonizer with the `RRO` of the actual path and the LSP operational flag set to *Up*.
11) The Harmonizer merges the three LSP segments attributes and sends a PCEP `Report` message to the MD-PCE providing all the information of the instantiated path (e.g., `ERO`, `RRO`, `LSP` fields such as identifiers, name, PLSP_ID and incoming MPLS label) with the operational flag set to *Up*, meaning that the path is ready to be utilized by service traffic.
12) The SSSA/TEI MD-PCE sends a PCEP `Report` to the source domain (i.e., Orange MD-PCE) to complete end-to-end connectivity instantiation.
13) The Orange MD-PCE sends a PCEP `Initiate` message to the ingress router, providing also the stitching label to be used at the border router over the inter-domain link.
14) The ingress router activates the RSVP RESV to the border router. Once the RSVP session is finished, the ingress router sends the PCEP `Report` to the MD-PCE, to inform that the end-to-end path is established.
15) Finally the MD-PCE send the HTTP 200 OK reply to the customer over the REST API.

VI. Conclusion

The contribution of the paper is twofold. It first presents an architectural solution able to establish end-to-end services with QoS constraints over multiple administrative domains. More specifically, the paper proposes an orchestration architecture (i.e., the 5GEx MdO) that is able to deploy end-to-end connectivity tunnels with QoS constraints (i.e., bandwidth and end-to-end delay) connecting VNFs deployed in remote data-centers. Second, the paper presents a performance assessment study on the scalability and the reliability of the proposed MdO prototype. Two types of operation were considered: *(i)* resource announcement, and *(ii)* QoS-based

connectivity service provisioning. Results obtained in a real multi-domain European testbed show that the MdO scheme does not present scalability problems nor any issues have been identified from the reliability point of view.

REFERENCES

[1] A. Muhammad, A. Sgambelluri, O. Dugeon, J. Martin-Perez, F. Paolucci, O. G. D. Dios, F. Ubaldi, T. Pepe, C. J. Bernardos, and P. Monti, "On the scalability of connectivity services in a multi-operator orchestrator sandbox," in *Optical Fiber Communication Conference*. Optical Society of America, 2018, p. M2A.2. [Online]. Available: http://www.osapublishing.org/abstract.cfm?URI=OFC-2018-M2A.2

[2] G. G. A. WG, "5g white paper views on 5g architecture," June 2016. [Online]. Available: https://5g-ppp.eu/white-papers/

[3] M. R. Raza, M. Fiorani, A. Rostami, P. Öhlen, L. Wosinska, and P. Monti, "Dynamic slicing approach for multi-tenant 5g transport networks [invited]," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 10, no. 1, pp. A77–A90, Jan 2018.

[4] A. Sgambelluri, F. Tusa, M. Gharbaoui, E. Maini, L. Toka, J. M. Perez, F. Paolucci, B. Martini, W. Y. Poe, J. M. Hernandes, A. Muhammed, A. Ramos, O. G. de Dios, B. Sonkoly, P. Monti, I. Vaishnavi, C. J. Bernardos, and R. Szabo, "Orchestration of network services across multiple operators: The 5g exchange prototype," in *2017 European Conference on Networks and Communications (EuCNC)*, June 2017, pp. 1–5.

[5] A. Sgambelluri, A. Milani, J. Czentye, J. Melian, W. Y. Poe, F. Tusa, O. G. de Dios, B. Sonkoly, M. Gharbaoui, F. Paolucci, E. Maini, G. Giuliani, A. Ramos, P. Monti, L. M. Contreras-Murillo, I. Vaishnavi, C. J. B. Cano, and R. Szabo, "A multi-operator network service orchestration prototype: The 5g exchange," in *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, March 2017, pp. 1–2.

[6] O. G. de Dios, R. Casellas, R. Morro, F. Paolucci, V. López, R. Martínez, R. Muñoz, R. Vilalta, and P. Castoldi, "Multipartner demonstration of bgp-ls-enabled multidomain eon control and instantiation with h-pce [invited]," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 7, no. 12, pp. B153–B162, Dec 2015.

[7] A. Giorgetti, F. Paolucci, F. Cugini, and P. Castoldi, "Proactive hierarchical pce based on bgp-ls for elastic optical networks," in *2015 Optical Fiber Communications Conference and Exhibition (OFC)*, March 2015, pp. 1–3.

[8] "5g exchange," http://www.5gex.eu/.

[9] "Deliverable d3.7 5gex," 2018. [Online]. Available: http://www.5gex.eu/

[10] O. D. et al., "draft-dugeon-brpc-stateful-01," March 2017.

[11] R. V. et al., "Network virtualization controller for abstraction and control of openflow-enabled multi-tenant multi-technology transport networks," in *2015 Optical Fiber Communications Conference and Exhibition (OFC)*, March 2015, pp. 1–3.

[12] Y. Yu, Y. Lin, J. Zhang, Y. Zhao, J. Han, H. Zheng, Y. Cui, M. Xiao, H. Li, Y. Peng, Y. Ji, and H. Yang, "Field demonstration of datacenter resource migration via multi-domain software defined transport networks with multi-controller collaboration," in *OFC 2014*, March 2014, pp. 1–3.

[13] R. Casellas, R. Muñoz, R. Martínez, R. Vilalta, L. Liu, T. Tsuritani, I. Morita, V. López, O. G. de Dios, and J. P. Fernández-Palacios, "Sdn based provisioning orchestration of openflow/gmpls flexi-grid networks with a stateful hierarchical pce," in *OFC 2014*, March 2014, pp. 1–3.

[14] "Deliverable d2.2 5gex," 2018. [Online]. Available: http://www.5gex.eu/

[15] A. M. et al., "Brite: An approach to universal topology generation," in *IEEE/ACM MASCOTS*, 2001.

[16] A. Sgambelluri, F. Paolucci, A. Giorgetti, F. Cugini, and P. Castoldi, "Experimental demonstration of segment routing," *Lightwave Technology, Journal of*, vol. 34, no. 1, pp. 205–212, 2016.

[17] A. Giorgetti, A. Sgambelluri, F. Paolucci, F. Cugini, and P. Castoldi, "Segment routing for effective recovery and multi-domain traffic engineering," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 9, no. 2, pp. A223–A232, Feb 2017.

[18] F. Paolucci, "Network service chaining using segment routing in multi-layer networks," *J. Opt. Commun. Netw.*, vol. 10, no. 6, pp. 582–592, Jun 2018. [Online]. Available: http://jocn.osa.org/abstract.cfm?URI=jocn-10-6-582

[19] A. Sgambelluri, F. Paolucci, F. Cugini, L. Valcarenghi, and P. Castoldi, "Generalized SDN control for access/metro/core integration in the framework of the interface to the routing system (I2RS)," in *2013 IEEE Globecom Workshops (GC Wkshps)*, Dec 2013, pp. 1216–1220.