On the Scalability of Connectivity Services in a Multi-Operator Orchestrator Sandbox

A. Muhammad[1], A. Sgambelluri[2], O. Dugeon[3], J. M. Perez[4], F. Paolucci[2], O. G. De Dios[5], F. Ubaldi[6], T. Pepe[6], C. J. Bernardos[4], P. Monti[1]

[1] *KTH Royal Institute of Technology, Kista, Sweden,* [2] *Scuola Superiore Sant'Anna, Pisa, Italy,* [3] *Orange Labs, Lannion, France,* [4] *Universidad Carlos III de Madrid, Madrid, Spain, ,* [5] *TID Telefonica, Madrid, Spain,*
[6] *Ericsson Research, Pisa, Italy*
Email: ajmalmu@kth.se

Abstract: The paper investigates the performance of a multi-domain orchestrator (MdO) deployed in a real multi-domain European testbed. Results show how the MdO prototype scales well with the number of domains advertised and connectivity services provisioned.

OCIS codes: (060.4258) Networks, network topology; (060.4259) Networks, packet-switched;

1. Introduction

In the next years, infrastructure providers will be asked to deliver new and flexible services such as Infrastructure as a Service (IaaS) and advanced virtualized function chaining 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 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. Chaining/slicing operations may require geographically distributed resources and may need to be orchestrated in a multi-domain orchestration fashion [1]. Another crucial aspect to consider is that service chaining has also to fulfill Quality of Service (QoS) requirements (e.g., latency below a certain threshold or a guaranteed bandwidth). For this reason, provisioning best-effort connectivity is not enough and dedicated network resources need to be advertised, computed, and provisioned with stringent traffic engineering requirements. The orchestration element of each multi-domain network infrastructure, i.e., the Multi-domain Orchestrator (MdO), is emerging as the key element to offer services to a federation of infrastructure providers while proactively monitoring and guaranteeing Service Level Agreements (SLAs) [1]. The workflows and procedures for efficiently establishing multi-provider services is still an open research topic and particularly challenging from the standpoint of the provisioning of connectivity resources. Most of the works in the literature focus on multi-domain lightpath and Label Switched Paths (LSP) establishment and rely on a hierarchical approach based on the Path Computation Element (PCE) concept [2, 3]. However, due to business models and trust issues, inter-operator traffic engineering may not allow a third-party neutral orchestrator, and thus, a peer-to-peer approach may represent a more reasonable and feasible approach. Furthermore, ensuring service availability and continuity increases the importance of scalability and resilience aspects of the orchestrator framework.

This paper evaluates the performance of an inter-operator orchestration framework based on the multi-domain orchestrator (MdO) developed in the context of the European 5G Exchange project [4] from the perspective of connectivity services. Resorting to the 5GEx Sandbox (i.e., a large multi-domain European network connecting 15 different lab premises of operators and research centers [1]) different scalability and resiliency measurements have been 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. Two novel approaches are evaluated in this paper. Advertising operations are carried out via a topology advertising module (i.e., referred to as TADS) based on an extension of the BGP-LS protocol in order to include IT information [5]. QoS-based connectivity services are provisioned via a multi-domain PCE module resorting to the stateful Backward Recursive PCE-based Computation (BRPC) [6] procedure.

2. Connectivity Service Orchestration in a Multi-domain Federation

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 three functional blocks of the 5GEx MdO, i.e., Catalogue and Topology Exchange (CTE), Service Orchestration (SO), and Service Assurance Management (SAM). The CTE block is responsible for exchanging (abstracted) topology information and MdO service level capabilities with other MdOs. The SO block performs the actual service deployment based on the information provided by CTE. Once the service is deployed the SAM block ensures the performance of the service across the different

This research has been supported by the EU H2020-ICT-14 project 5GEx (grant n. 671636) and by the Celtic-Plus sub-project SENDATE-EXTEND funded by Vinnova.

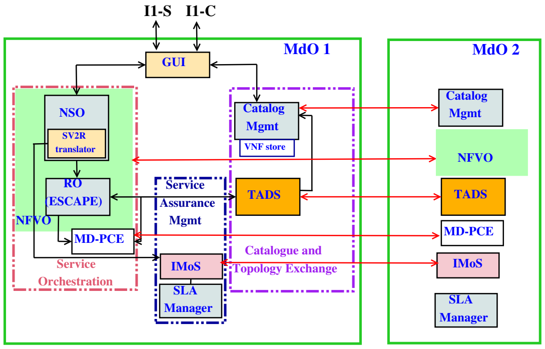

Figure 1: Functional architecture of the 5GEx MdO.

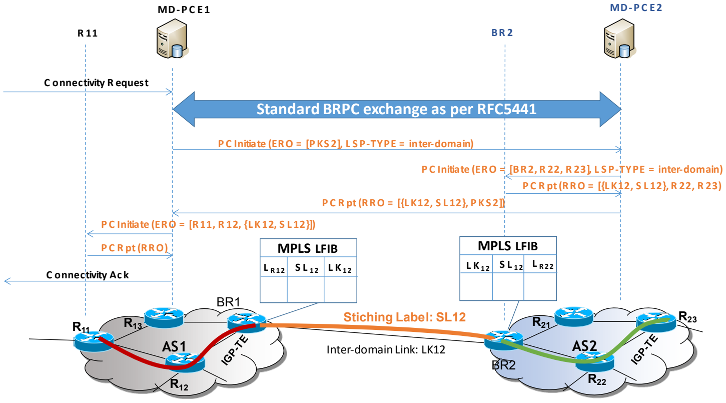
MD-PCE1
MD-PCE2
R11
BR2
Connectivity Request
Standard BRPC exchange as per RFC5441
PCInitiate (ERO = [PKS2], LSP-TYPE = inter-domain)
PCInitiate (ERO = [BR2, R22, R23], LSP-TYPE = inter-domain)
PCRpt (RRO = [{LK12, SL12}, R22, R23)
PCRpt (RRO = [{LK12, SL12}, PKS2])
PCInitiate (ERO = [R11, R12, {LK12, SL12}])
PCRpt (RRO)
Connectivity Ack
MPLS LFIB
L_{R12} SL_{12} LK_{12}
MPLS LFIB
LK_{12} SL_{12} L_{R22}
BR1
R_{11}
R_{13}
R_{12}
AS1
IGP-TE
Stiching Label: SL12
Inter-domain Link: LK12
R_{21}
R_{23}
R_{22}
AS2
BR2
IGP-TE

Figure 2: Inter-domain connectivity workflow.

domains as per the SLA agreement. More details about the MdO general architecture and subsystems can be found in the deliverable 3.1 available at [4]. Inside the CTE, the Topology Abstraction and Discovery Subsystem (TADS) is responsible for maintaining a database of the networking and IT resources available at other MdOs. This information is provided, within the SO block, to the Resource Orchestrator (RO) and the Multi-Domain Path Computation Element (MD-PCE) for the deployment of inter-domain services. More specifically, the MD-PCE is responsible for establishing QoS-based connectivity services utilizing the multi-domain topology information provided by TADS. Fig. 2 provides an example of the procedure used by the MD-PCE to sets up a QoS-based connectivity service. The procedure works as follows: when the MdO receives a connectivity service request MD-PCE is instructed to instantiate a path between the two endpoints specified in the service request. This requires a collaborative effort between the MD-PCEs to find the optimal inter-domain path. A novel distributed approach, i.e., stateful BRPC [6], is adopted for this purpose. In addition to standard BRPC, the stateful approach implement a chain provisioning exploiting PCInitiate and PCReport messages within each domain including the label stitching information between the domains through dedicated PCRpt message. Fig. 2 describes the different call flow for establishing an inter-domain connectivity. More information on this procedure is available in [6]. In the context just described it is crucial that TADS keeps consistent information of domains' resources even in the presence link failures. In other words, TADS should be able to understand if the neighbor from which it was listening to is no more reachable thus switching to another TADS to receive the required resource update messages. This is accomplished by implementing a dynamic database inside TADS, where resource information is deleted if it is not updated within a particular time frame. This procedure works as follows. A time stamp with "value" and "information source" attributes is added to every resource information stored in TADS database. This time stamp is refreshed every time TADS receives an update message from the same source. When the time stamp of a specific information expires then that information is removed from the database making TADS open to acquire and store information from other sources. This allows TADS to retrieve resource information of other MdOs even in the presence of connectivity failures. More details about the MdO general architecture and subsystems can be found in the deliverable 3.1 available at [4].

3. Scalability and robustness tests

To measure the scalability performance of the resource advertisement operation we tested TADS on a subset of islands (local testbeds) of the 5GEx Sandbox as shown in Fig. 3(a). The test environment consists of a topology generator that creates topologies of different dimensions according to the BRITE-based Waxman methodology [7]. For the generated topologies the nodal degree is set to 6 while the number of nodes is varied from 10 to 250. The generated topologies are exchanged with other TADS using BGP-LS in a chain mode. The TADS scalability is evaluated by measuring the time required to distribute multi-domain resource information (i.e., discovery procedure) which is computed in the sandbox. The 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. 4(a) shows the results for TADS convergence time as a function of topology size exported to other TADS peers. The experimental set up considers a two (Pisa - Stockholm) and a three (Pisa - Stockholm - Pisa) domain scenarios. Considering the dimension of 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 into account that, on average, the number of domains traversed by the resource information will be less than 3, the results in Fig. 4(a) indicates 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 the updates about the

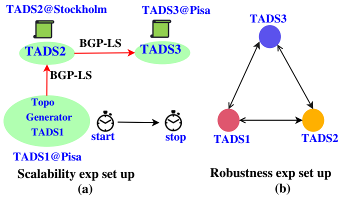

Figure 3: Testing environment.

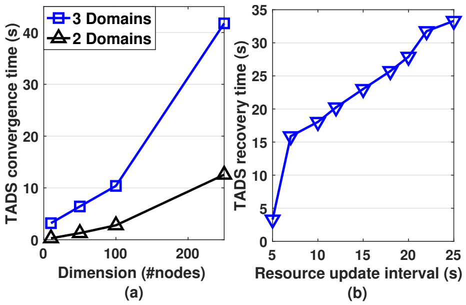

Figure 4: (a) Convergence time; (b) Recovery time.

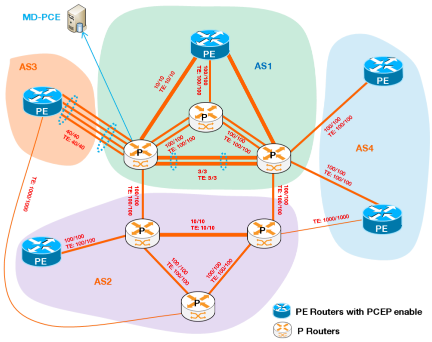

Figure 5: MD-PCE testbed

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 domain scenario. This is because of the fixed and relatively high (5s) value for the update timer. This convergence time can be significantly reduced by employing the throttling-based approach, i.e., sending rate-limited updates (to other TADS) upon reception of resource information instead of the fixed timer based strategy. As already mentioned, resource advertisement operations need also to be robust against link failure scenarios. This robustness feature has been tested in the set up presented in Fig. 3(b). During the test, we interrupt the connectivity between TADS1 and TADS3, and measure how much time is required at TADS1 to detect the link failure and to recover the information about the resources information of TADS3 via TADS2. Fig. 4(b) depicts the value of the measurements results. They show a significant growth in the recovery time with the increase in the time gap between the update messages sent by each TADS. This highlights the importance of setting a proper time interval (i.e., between update messages) so that we can achieve a good tradeoff between network management overhead and fast recovery of domains' resources information.

To assess the scalability of the MD-PCE component, we used the experimental setting depicted in Fig. 5. Thanks to the OpenDayLight BGPCEP project [8] that already implements the BGP and PCE protocols we only had to implement the others features of the MD-PCE (e.g., Path and AS Path computation algorithm, Topology Graph, etc.) as a new bundle of ODL that depends of the BGPCEP bundles. In the experiment we set up a full-mesh of tunnels between the PE routers (i.e. router acting as Path Computation Client and able to receive a PCInitiate message from the MD-PCE). This bulk of tunnels is repeated 100 times. We measure the time necessary to set up all bulk, in addition to minimum, maximum, and average time to set up one bulk. Then we let the MD-PCE manage the numerous LSPs before removing all of them. For the scalability and performance test, a python script was developed to send HTTP REST POST requests to the MD-PCE. We run several test which establish n times a full-mesh of tunnels. Running it 100 times, we could set up 2500 LSPs (25 per PE router) in our testbed. We could create 2500 LSPs in 189.20 [s] with a minimum, average, and maximum time values equal to 1.50 [s], 1.89 [s], and 2.52 [s] respectively. This corresponds to around 13,2 LSPs/[s]. We observe no significant variation in CPU and memory consumption fo the MD-PCE during the establishment of the LSPs. It was then possible to get the list of all LSP tunnels using the standard HTTP REST API of OpenDayLight BGPCEP project. Removing all the tunnels is 10 times faster compated to the creation. This is mostly due to the PE routers. Indeed, it sends back PCReport to the MD-PCE immediately and send only a RSVP TEAR message without waiting for the answer, which is the normal procedure to remove an RSVP-TE tunnel.

4. Conclusions

This paper presents a performance assessment study on the scalability and robustness of the MdO prototype developed in the 5GEx project. 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 how well the MdO scales well with the number of domains to be advertised and QoS-based connectivity services to be provisioned.

References

1. A. Sgambelluri et al., "Orchestration of Network Services Across Multiple Operators: The 5G Exchange Prototype," *EuCNC*, 2017.
2. O. Gonzalez De Dios et al., "Multipartner demonstration of BGP-LS-Enabled multidomain EON control and instantiation with H-PCE," *IEEE/OSA JOCN*, vol.7, no.12, 2015.
3. A. Giorgetti et al., "Proactive hierarchical PCE based on BGP-LS for elastic optical networks," *OFC, Th1A.3*, 2015.
4. 5GEx webpage http://www.5gex.eu/.
5. A. Sgambelluri et al., " Software prototype documentation and user manual," *D 3.5 available at* http://www.5gex.eu/.
6. O. Dugeon et al., "draft-dugeon-brpc-stateful-00," *IETF*, March 2017.
7. A. Medina et al., "BRITE: An approach to universal topology generation," *IEEE/ACM MASCOTS*, 2001.
8. OpenDayLight BGPCEP project: https://wiki.opendaylight.org/view/_LS_PCEP:Main