Enabling Multi-Domain Orchestration using Open Source MANO, OpenStack and OpenDaylight

Panagiotis Karamichailidis, Kostas Choumas and Thanasis Korakis
Dept. of ECE, University of Thessaly, Volos, Greece
Email: karamiha, kohoumas, korakis@uth.gr

***Abstract*—In recent years, the rise of Network Function Virtualization (NFV) makes the Network Service (NS) deployment much agile and flexible. The proprietary and custom-made hardware is replaced by a virtual and software-based infrastructure, that is easily exploited in a common way for the NS deployment. One of the most challenging problems in this environment is deploying and organizing in a large-scale and multi-domain infrastructure, which contains geographically distributed but interconnected data-centers. The proposed solution focuses on the extra networking operations required in a NFV infrastructure, managed by Open Source MANO, OpenStack and OpenDaylight. We develop software proxies that collaborate with the aforementioned tools and enhance their functionality. Finally, we implement and evaluate the proposed architecture, using the NITOS experimentation testbed.**

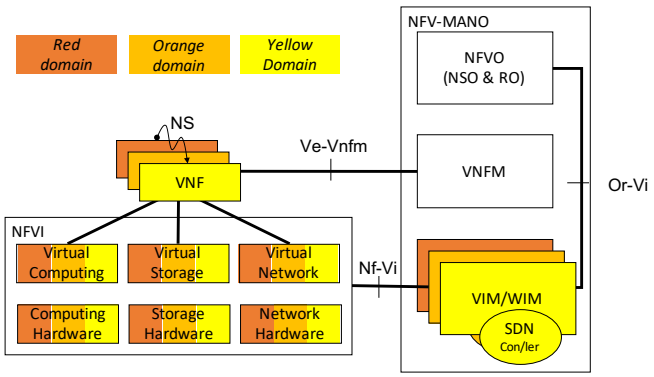


Fig. 1: NFV-MANO

I. Introduction

As virtualization dominates in every aspect of networking and computing, Network Services (NSs) could not be unaffected by this. For example, the Evolved Packet Core (EPC) of LTE is a NS that previously was exclusively supported by proprietary and hardware implemented Network Functions (NFs) and now is not. The proliferation of Network Functions Virtualization (NFV) enables the software implementation of these NFs, that is decoupled from the utilized computation, storage and network resources. In this way, NFV exposes a new set of entities, named Virtualized NFs (VNFs) and NFV Infrastructure (NFVI). The latter entity includes all the resources used for the VNF deployment, management and execution.

According to the ETSI Standardization group, the NFV Management and Orchestration (NFV-MANO) [1] is a challenging task that requires the synergy of several functional blocks, organized in an architectural framework and collaborating through specified reference points, as it is depicted in Figure 1. Except for the VNF Manager (VNFM), that is in charge of the VNF lifecycle, the other two functional blocks of our interest are the NFV Orchestrator (NFVO) and the Virtualized Infrastructure Manager (VIM). The NFVO has two main functions, the NS Orchestration (NSO) and the Resource Orchestration (RO), which implement the lifecycle management of the NSs and the orchestration of the NFVI resources across multiple VIMs respectively. The VIM is responsible for controlling and managing the NFVI resources within a single domain[(1)], leveraging on hypervisors and SDN controllers for the control of the computation/storage and network resources respectively. There are also specialized VIMs, such as the WAN Infrastructure Manager (WIM) that controls only network resources. In this paper, we focus on network and compute domains (or data-centers), controlled by WIMs and VIMs respectively, and we investigate the role of SDN in their operation. The interaction and functionality of the SDN controller within the NFV-MANO architecture is not clearly defined, even in the report on the SDN usage in the NFV-MANO [2], since the SDN controller could either be part of the VIM or the NFVI, among other options. *This paper sheds some light on the functionalities that could be offered by the SDN controller, either as a part of the VIM or constituting the WIM, especially for enabling the multi-domain orchestration.*

For our deployment, we exploit and extend the state-of-the-art open-source software tools, namely Open Source MANO (OSM) [3], OpenStack [4], [5] and OpenDaylight [6], [7], which are for the NFVO/VNFM, the VIM and the SDN control respectively. This triangle is one of the most widely-used options for implementing the NFV-MANO, and from now on is denoted as *Open*^{3}. Although *Open*^{3} enables a remarkable set of operations, it still does not support multi-domain NS deployment. We present a solution that enables the VNF deployment of a NS over multiple compute domains, as well as their interconnection through a network domain, relying

[(1)]A domain could be either administrative, which exists within a single organization/service provider, or technology oriented, based on the type of technology in scope.

978-1-7281-1434-7/19/$31.00

on the *Open*^{3} tools and our custom-made and open-source software. Our software leverages on the functionalities offered by the OpenDaylight SDN controller. Last but not least, we evaluate our implementation and we showcase the results of our experimentation in the NITOS testbed [8].

The remainder of this paper is organized as follows. In Section II, we briefly summarize the capabilities of *Open*^{3} and the most relevant work. In Section III, we explain the challenges of the multi-domain orchestration and our solution using more SDN. In Section IV, we present an evaluation of our solution, obtained via testbed experimentation. Section V concludes the paper.

II. Related Work

The most representative implementation of the NFV-MANO is the Open Source MANO (OSM), supported by ETSI. The OSM community works on being aligned with the evolution of the ETSI NFV specifications, while these specifications are updated according to the feedback coming from the use of OSM. Although OSM has its own implementation of the VIM, named as OpenVIM, OpenStack is more stable, well-known and used by many projects. OpenDaylight is similarly one of the state-of-the-art SDN controllers, that has already developed the required interfaces and plugins for synergy with OSM and OpenStack.

At the moment of writing this paper, the most updated and stable versions of OSM (*Release FOUR*), OpenStack (*Queens*) and OpenDaylight (*Nitrogen-0.7.2*) enable only the single-domain NS deployment. Although OSM can interact with multiple VIMs in parallel, it uses only one VIM each time a new NS is requested, which has to be declared on the NS request. Thus, OSM cannot use VNFs offered by different VIMs for multi-domain NS deployment, and of course WIMs do not exist in its ecosystem. Moreover, even in the single-domain NS deployment, OpenStack is able to control and manage only the physical machines (computing/storage hardware) hosting the VNFs of the deployed NS, and not the network hardware connecting these machines. OpenDaylight is only and optionally used for the configuration of the Open virtual Switches (OvS) existing in the host machines (or *compute nodes*).

SDN-Assist [9] is an extra plugin offered by OSM, that can be used for the single-domain NS deployment. It is in charge of connecting the VNFs located in different compute nodes, if and only if the compute nodes use physical Network Interface Cards (NICs) with Peripheral Component Interconnect express (PCIe) pass-through capabilities. These NICs exploit the Single-Root Input/Output Virtualization (SR-IOV) functionality to appear to be multiple separated physical devices, each of them mapped to and used by a single VNF. Although SR-IOV enables higher performance, it is not yet compatible with the majority of the utilized software (hypervisors) or hardware (NICs) components. The majority of the existing deployments depend on compute nodes with non-supporting SR-IOV NICs, while the VNFs use virtual NICs (vNICs) connected to virtual bridges (operated by OvS) residing on top of the physical NICs, as it is depicted in Figure 3(a). Moreover, SDN-Assist cannot not be used for multi-domain NS deployment, as the one depicted in Figure 3(b).

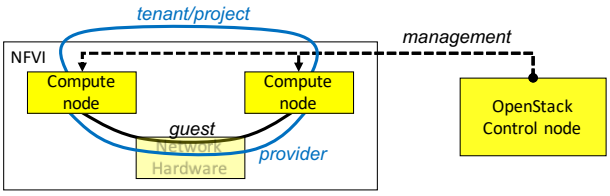


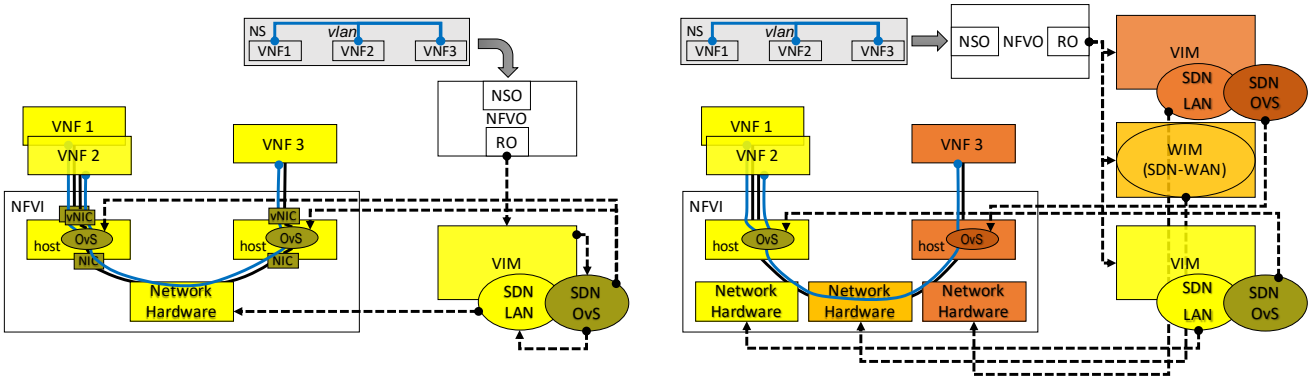
Fig. 2: OpenStack networks

There are also some works proposing solutions for multi-domain NS deployment. The main disadvantage of these approaches is their lack of support for VNF interconnection inside the domains, since they focus only on the WIM implementation and not the SDN controller needed as part of each VIM [10], [11], [12]. Moreover, they use their own proprietary software for the orchestration (*FROGv4*, *T-NOVA* and *TeNOR* respectively), instead of the widely-used and well-tested OSM. 5GEx [13] is one of the most well known and successful projects for enabling cross-domain orchestration of services over multiple administrations or over multi-domain single administrations, however, it relies on a proprietary multi-domain orchestrator. To the best of our knowledge, this is the first paper extending OSM for multi-domain orchestration.

III. Extending *Open*^{3} Networking

We present *Open*^{3}++, which extends the networking functionality of *Open*^{3} by enabling the interconnection of VNFs belonging to the same NS in more scenarios, either if they are collocated inside a single domain or spread to multiple domains.

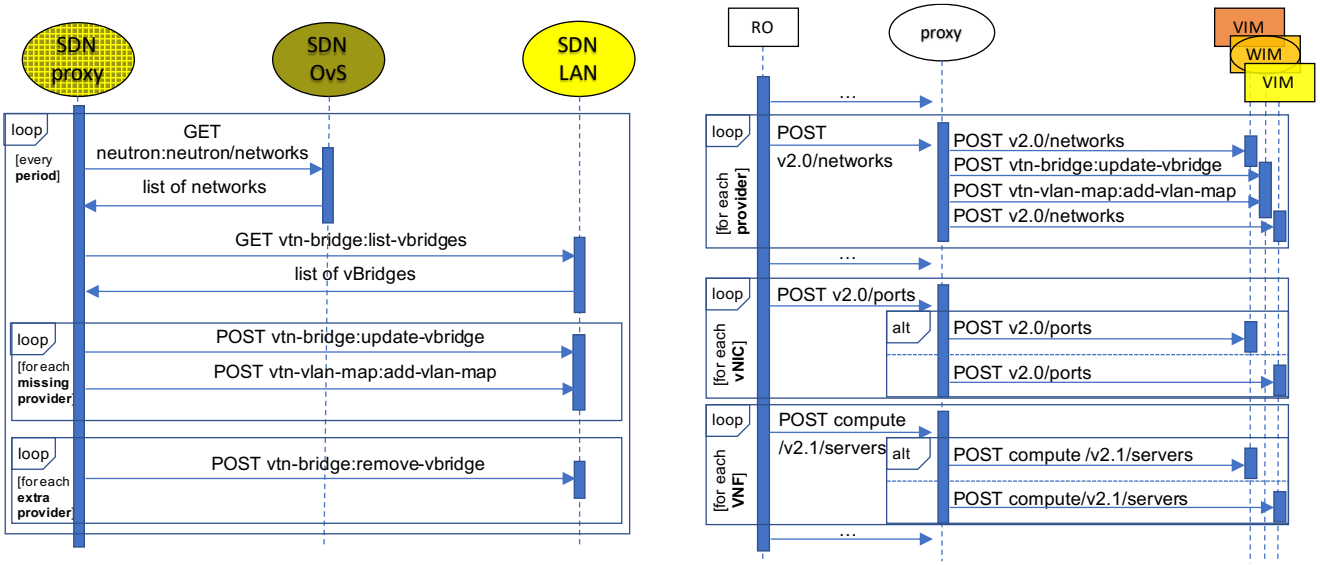
Before proceeding, it will be helpful to clarify the terminology and the classification of the networks followed by OpenStack, since our software extending *Open*^{3} to *Open*^{3}++ is mainly handling these networks with use of OpenDaylight. In a single compute domain managed by OpenStack, the VNF interconnection is built either through tunnels over the OpenStack *management* network, which are named *tenant* or *project* networks, or using overlay networks on top of the OpenStack *guest* network, which are called *provider* networks. Figure 2 depicts these networks. Our focus is on the latter case, since the guest network can be physically connected to other network domains and the provider networks may be used for interconnecting VNFs of different compute domains. In *Open*^{3}, the provider networks exist "out there" and OpenStack simply interfaces them. But in *Open*^{3}++, they are deployed on demand by exploiting the SDN-LAN and SDN-WAN controllers with use of our extensions. The following sections give more details on the SDN-LAN and SDN-WAN controllers, as well as their usage for the creation of *flat* or *vlan* provider



(a) Single-domain NS deployment.

(b) Multi-domain NS deployment.

Fig. 3: Interactions between the functional blocks for the single/multi-domain NS deployment.



(a) Messages from/to the SDN-proxy

(b) Messages from/to the Orchestration-proxy.

Fig. 4: Diagrams of the messages sent or received by the developed proxies.

networks[(2)] connecting the compute nodes of single or multiple domains.

A. *SDN-LAN & SDN-proxy*

Let's consider the single-domain scenario illustrated in Figure 3(a). In this scenario, all compute nodes belong to the "yellow" domain, which is managed by a single VIM. In *Open*^{3}, OpenStack implements the VIM and optionally leverages on the OpenDaylight instance, named SDN-OvS in our examples, for configuring the OvS bridges running in the compute nodes (usually named as *br-int*). When OpenStack receives from the OSM's RO a request for three VNFs of a new NS, connected through a provider network, it chooses the compute nodes on which these VNFs will be deployed according to its scheduling policy. Then, OpenStack deploys the VNFs and informs SDN-OvS to configure the bridges of the chosen compute nodes. The bridge configuration is sufficient for connecting the VNFs located in the same compute node (e.g. VNF 1 and VNF 2), or to export the traffic to the physical NIC, when it is directed from a VNF to another non-collocated one (e.g. from VNF 2 to VNF 3). However, the traffic forwarding between the NICs of the compute nodes requires the appropriate configuration of the OpenStack guest network connecting them. This is the task of the extra OpenDaylight instance, which we call SDN-LAN.

Open^3 does not include any interface to SDN-LAN, assuming that it is standalone and proactively builds and keeps active the provider networks, even in time periods that they are not used. On the other hand, *Open*^{3}++ relies on a

[(2)] *flat* provider networks forward untagged traffic, while *vlan* provider networks expect for VLAN tagged traffic to forward.

software daemon, named SDN-proxy[3], which repetitively checks SDN-OvS and prompts dynamically SDN-LAN to deploy and keep only the needed provider networks on top of the domain's underlying network. More specifically, SDN-proxy checks SDN-OvS for the required provider networks and forces SDN-LAN to form an isolated overlay network for each provider network, which functions as an abstract layer-2 switch. This function is completed with the assistance of the *Virtual Tenant Network (VTN) Manager* plugin, which is used by both OpenDaylight instances implementing SDN-OvS and SDN-LAN. VTN-Manager enables the creation of a virtual bridge (*vBridge*) for each provider network and *VLAN mapping* is used for assigning the VLAN traffic of the provider network to the respective vBridge (VLAN 0 corresponds to the untagged flat network).

Figure 4(a) shows the sequence of REST/HTTP messages exchanged between our daemon and the OpenDaylight instances, SDN-OvS and SDN-LAN. SDN-proxy periodically i) gets from SDN-OvS the provider networks (first GET message), ii) gets from SDN-LAN the existing vBridges (second GET message) iii) forces SDN-LAN to create a new vBridge for each missing provider network (first POST message), iv) maps each vBridge to the VLAN of the related provider network (second POST message) and v) removes the vBridges corresponding to non-existing provider networks (last POST message). The latter three messages are in loops since they are repeated for every vBridge that has to be added or removed.

The disadvantage of this SDN-proxy implementation is that SDN-OvS is periodically checked, thus SDN-LAN gets updated even after a period since SDN-OvS has changed. Another solution is to make OpenStack logging these changes (added or removed provider networks) in a file, using the logging mechanism of the OpenStack *Neutron* component, while SDN-proxy periodically reads this file and tracks the changes. The time period of this process is significantly lower. The two first GET messages of Figure 4(a) are replaced with the tracking of the log file, while the two inner loops are repeated again, for every added or removed network respectively. On the other hand, the disadvantage of this approach is that SDN-proxy has to be collocated with OpenStack, and the logging mechanism of its Neutron component must be activated.

B. SDN-WAN & Orchestration-proxy

Let's now focus on the multi-domain NS deployment, illustrated in Figure 3(b), where the three VNFs belong again to the same NS, but they are deployed to different compute domains, the "yellow" and the "red" one, which are connected through the "orange" network domain. In *Open*^{3}++, we developed a daemon called Orchestration-proxy[4], that is a proxy between the RO and the VIMs/WIMs. It pretends to be a VIM instance for the RO and a RO for the underlying VIMs/WIMs. In particular, Orchestration-proxy presents to the RO as a single VIM responsible for a single domain and receives the requests for the NS resources. Then, it "breaks" the set of resources and assigns the subsets to various compute domains, interacting with their VIMs. For these VIMs, it is behaving as the RO, requesting the VNF deployment to their compute domains. Except for this task, the Orchestration-proxy is responsible to interact with the WIM(s) managing the network domain(s) interconnecting these compute domains, in order to build the VNF connections. Each WIM is actually an SDN controller, called SDN-WAN, responsible for a network domain used for the interconnection of other domains. *Open*^{3}++ uses OpenDaylight and the VTN-Manager plugin to implement SDN-WAN, which again creates one vBridge for each provider network connecting VNFs, and maps this vBridge to the VLAN traffic of the provider network. The goal of both SDN-LAN and SDN-WAN is to create overlay layer-2 networks on top of their domains, using VLAN for their isolation. The networks of the same VLAN but of different domains are stitched together.

Figure 4(b) depicts the REST/HTTP messages exchanged between Orchestration-proxy and the other functional blocks, when a new NS is deployed. Among other messages, we distinguish the following: i) the RO requests from Orchestration-proxy the utilization of a provider network (first POST message from the RO), and Orchestration-proxy, in turn, requests from each VIM/WIM the same network (four first POST messages from the Orchestration-proxy). These requests are either OpenStack POST messages to the VIMs of the compute domains (similar to the request from the RO) or OpenDaylight POST messages to the WIM of the network domain (similar to the two first POST messages in Figure 4(a)). Then, ii) the RO requests for a new vNIC (or port) for each VNF included in the NS (second POST message from the RO). This request is copied to one of the VIMs (first alt). Finally, iii) the RO requests the creation of a new VM (or server) for each VNF (third POST message from the RO), which again is copied to a single VIM (second alt).

IV. Evaluation

Both SDN-LAN and SDN-WAN are OpenDaylight instances that utililize the VTN-Manager plugin, in order to build the provider networks. VTN-Manager exploits the Link Layer Discovery Protocol (LLDP) to discover the underlying network and retrieve the shortest path between each pair of switches. Each link is weighted with a cost, which can be modified through the VTN-Manager northbound interface, enabling the flexible redefinition of the shortest paths. VTN-Manager reactively configures flows, which means that the first packet sent from a VNF to another need some extra time to be forwarded, until the controller configures the flows for this. Each flow is specific enough to connect a particular VNF couple, matching their MAC addresses.

The shortest path connecting a VNF couple is chosen by VTN-Manager during the forwarding process of the very first ARP request. If the first packet is forwarded from VNF 1 to VNF 2, then the corresponding ARP request has the MAC of

[3]The repository of SDN-proxy is given in this URL: http://repo.nitlab.inf.uth.gr/karamiha/odl-project/tree/test

[4]The repository of the Orchestration-proxy software is given in this URL: http://repo.nitlab.inf.uth.gr/karamiha/orchestrator

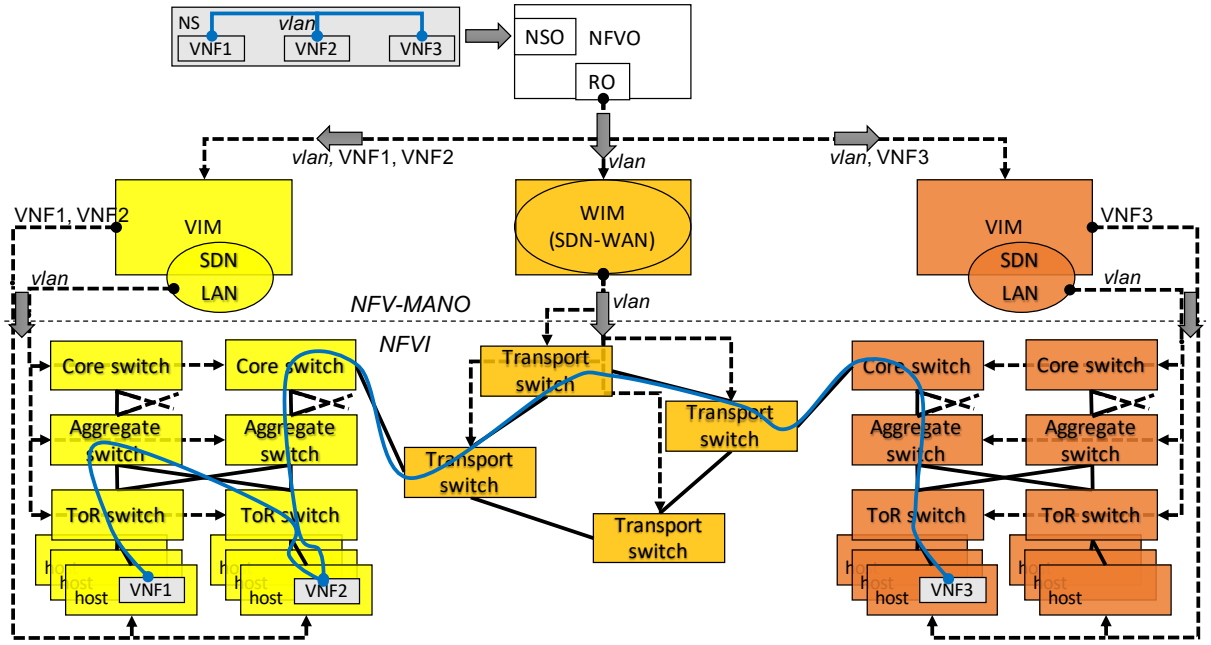


Fig. 5: Simple scenario illustrating the deployment of a NS with 3 VNFs over two compute domains (or data-centers), the yellow and the red one, interconnected through a network domain, the orange one.

VNF 1 as source address and the broadcast MAC as destination address, thus, the SDN controller learns the ingress switch and the port that VNF 1 is attached to and pushes the packet to all other switches. The other switches, in turn, forward this packet to all ports except for the ones discovered by LLDP, delivering the packet only to VNFs and not to switches. Once VNF 2 responds with an ARP reply, the SDN controller learns the ingress switch and the port of VNF 2 and calculates the shortest path between the ingress switches of the two VNFs. Then, it deploys the flows to the switches participating in this path, matching only the packets sent from VNF 1 to VNF 2.

Although *Open*^{3}++ benefits from the dynamic deployment or removal of provider networks comparing to *Open*^{3}, its delay performance may be worse at the first packets exchanged between the VNF couples, due to its reactive flow configuration. This evaluation is to estimate the average time needed for the flow configuration connecting a new VNF couple, when these VNFs are i) in the same compute domain or ii) in different compute domains.

A. *Intra-domain Delay*

Let's assume an SDN network with multiple switches connected in a three-tier fat-tree network topology. This is a typical scenario in a compute domain or data-center with multiple interconnected compute nodes, using access or Top of Rack (ToR), aggregate and core layer switches [14], as it is depicted in Figure 5. In general terms, the compute nodes of a data-center are grouped to racks, each one equipped with a ToR switch connecting its compute nodes. Each ToR switch is connected to one or more aggregate switches, which are connected to one or more core switches.

In the example of Figure 5, there is a NS with three sequentially connected VNFs, named VNF 1, 2 and 3. VNFs 1 and 2 are deployed in the yellow compute domain, in two different compute nodes belonging to different racks. Although the ToR switches connecting these two VNFs are unique, the aggregate switch connecting the two ToR ones is not exclusive. In this example, the left yellow aggregate switch is chosen by the VTN-Manager of SDN-LAN, instead of the right yellow one. If the ToR switches cannot be physically connected through an aggregate switch, then a longer path is used, including two aggregate switches and a core one. As follows, two VNFs of the same compute domain are either i) directly connected, if they are hosted on the same compute node, or connected through ii) a single ToR switch, or iii) two ToR and an aggregate switch or iv) two ToR, two aggregate and a core switch. From our experimentation in NITOS, we checked that the ping delay for the first packet sent from VNF 1 to VNF 2 is between $8-12$ msec, including the delay of the ARP and ICMP forwarding, as well as the delay of the reactive flow configuration. The upper and lower delay limits are not significantly affected by the path length, except for the case that the VNFs are directly connected. The compute nodes are NITOS nodes, featuring Intel Core i7-2600 CPU at 3.40 GHz and 8M Cache, while the switches are emulated by executing Mininet in a separate NITOS node. All NITOS nodes are interconnected through an OpenFlow HP 3800 switch.

B. *Inter-domain Delay*

In the same example of Figure 5, VNFs 2 and 3 are deployed in different compute domains, connected through a network domain. When VNF 2 pings VNF 3, the yellow SDN-LAN

pushes the right yellow core switch to forward the ARP request of VNF 2 to the left orange switch. For the orange SDN-WAN, the left orange switch is the ingress switch of this ARP request, thus repeating the same process with before, the packet is forwarded from the right orange switch to the left red core switch. Finally, the red SDN-LAN receives this ARP request and similarly pushes it to the left red ToR switch to be forwarded to VNF 3. The duration of the ARP request forwarding from VNF 2 to VNF 3 is expected to be almost three times higher comparing to the previous case, between VNF 1 and VNF 2. If the involved controllers were more, having more network domains between the two compute ones, the duration would be respectively higher. This is also verified by our experimentation in NITOS, since the delay of the ping between VNF 2 and VNF 3 is between $20 - 33$ msec. The switches of each domain are emulated executing Mininet in separate NITOS node, one for each domain.

V. Conclusion

This paper presents an implementation of multi-domain NS deployment, which is based on OSM, OpenStack and OpenDaylight, as well as our SDN-proxy and Orchestration-proxy. There are many challenges, especially in programming Orchestation-proxy, which can be modeled as NP-hard optimization problems, such as Location-Routing Problem and Virtual Network Embedding. The focus of this work is to present the framework for applying the solutions, relying on widely-used software tools. Our work is continuously expanding in many directions, including the optimization of the VNF deployment at the compute domains, considering either the compute load of the data-centers or the network load of their connections. We also plan to exploit other OpenDaylight features, for enhancing SDN-WAN and SDN-LAN.

Acknowledgment

This work has been financially supported by the EU Horizon 2020 project 5G-PICTURE under grant agreement No 762057. The European Union and its agencies are not liable or otherwise responsible for the contents of this document; its content reflects the view of its authors only.

References

[1] ETSI GS NFV-MAN 001 v1.1.1 (2014-12), "Network Functions Virtualisation (NFV); Management and Orchestration".
[2] ETSI GS NFV-EVE 005 v1.1.1 (2015-12), "Network Functions Virtualisation (NFV); Ecosystem; Report on SDN Usage in NFV Architectural Framework".
[3] ETSI OSM, "Open Source MANO", https://osm.etsi.org.
[4] O. Sefraoui, M. Aissaoui and M. Eleuldj, "OpenStack: toward an open-source solution for cloud computing", International Journal of Computer Applications, vol. 55, no. 3, pp. 38-42, 2012.
[5] "OpenStack", https://www.openstack.org.
[6] J. Medved, et al, "Opendaylight: Towards a Model-Driven SDN Controller architecture", IEEE WoWMoM, 2014.
[7] "OpenDaylight", https://www.opendaylight.org.
[8] "Network Implementation Testbed using Open-Source platforms", https://nitlab.inf.uth.gr.
[9] "EPA and SDN assist", https://osm.etsi.org/wikipub/index.php/EPA_and_SDN_assist.
[10] R. Bonafiglia, G. Castellano, I. Cerrato and F. Risso, "End-to-end service orchestration across SDN and cloud computing domains", IEEE NetSoft, 2017.
[11] J. Caparinha, et al, "Deployment of Virtual Networks Functions over multiple WAN interconnected PoPs", IEEE NFV-SDN, 2017.
[12] J.F. Riera, et al, "TeNOR: Steps Towards an Orchestration Platform for Multi-PoP NFV Deployment", IEEE NetSoft, 2016.
[13] 5G Exchange (5GEx), http://www.5gex.eu/.
[14] M. Al-Fares, A. Loukissas and A. Vahdat, "A Scalable, Commodity Data Center Network Architecture", ACM SIGCOMM, 2008.