A Techno-Economic Evaluation of VNF Placement Strategies in Optical Metro Networks

Leila Askari, Francesco Musumeci, Massimo Tornatore
Department of Electronics, Information and Bioengineering, Politecnico di Milano, Milan, Italy
E-mail: firstname.lastname@polimi.it

***Abstract*—Network Function Virtualization (NFV) has changed the way operators can provision network services. Decoupling network functions from dedicated hardware and running them on software, on top of commodity servers and switches, not only helps operators have more flexible and easy-to-manage networks, but also reduces their capital and operational expenditures. This is especially true for incoming 5G services, characterized by ultra-low latency, high reliability and bandwidth requirements. To satisfy these challenging requirements, multi-layer optical networks based on Optical Transport Network (OTN) over wavelength division multiplexing (WDM) are being deployed in the metro segment to support 5G services. In addition, the possibility to equip metro nodes with computing capabilities, enabled by new paradigms such as CORD (Central Office Re-architected as a Datacenter) is being exploited. In this scenario, an efficient placement of Virtual Network Functions (VNFs) for Service Chain (SC) provisioning within the metro network is needed, and different VNF placement strategies can lead to different costs for network operators. In this paper we analyze the impact of different VNF placement strategies on the optical metro network cost, considering specific Service Level Agreement (SLA) requirements, expressed in terms of service blocking probability. We provide a cost model which takes into consideration both capital and operational expenditures. Through extensive numerical results, we quantify the impact of using a cost-effective VNF placement strategy in decreasing network cost while meeting the desired SLA performance.**

***Index Terms*—NFV, VNF placement, Dynamic Service Chaining, cost analysis, metro network**

I. INTRODUCTION

In the last decade, two new networking paradigms have attracted the attention of researchers and practitioner, i.e., Network Function Virtualization (NFV) and Software Defined Networking (SDN), for their ability to lead to more agile and flexible networks. SDN simplifies network control by decoupling control plane from data plane [1]. NFV enables network operators to achieve flexibility and cost saving by replacing dedicated hardware with software running on top of commodity server and switches [2]. Traditional network functions such as, e.g., firewall, Network Address Translation (NAT), Intrusion Detection System (IDS), etc., can be implemented in software installed on general-purpose hardware, leading to a simplification of the management of network function and reduction of costs. These *softwarized* functions are called Virtual Network Functions (VNFs). Several of future-generation (i.e., 5G) network services are characterized by ultra-low latency, high bandwidth requirements and high availability and they can be supported through an ordered sequence of VNFs, which is called Service Chain (SC).

To satisfy the stringent requirements of 5G services, operators are deploying multi-layer optical networks based on Optical Transport Network (OTN) over Wavelength Division Multiplexing (WDM). Also, the architecture of the metro network nodes is evolving towards the concept of Central Office Re-architected as a Datacenter (CORD) [3], i.e., central offices are now equipped with processing and storage units. In this context, achieving a dynamic and flexible (i.e., network-status-aware) VNF placement for SC provisioning in metro networks is not a trivial task, as a trade-off arises between: i) network transport capacity, ii) processing units, i.e., capacity of network nodes hosting VNFs (called NFV-nodes in the following), and iii) required Service Level Agreement (SLA), e.g., expressed in terms of service blocking probability and/or maximum tolerated service latency. Furthermore, the SC provisioning should be performed by limiting the overall network cost, including both Operational Expenditure (OpEx) and Capital Expenditure (CapEx).

While most of the studies in the literature consider the impact of placement only on the OpEx (e.g., power consumption) [4], in this paper we evaluate how different placement strategies can impact both the OpEx and the CapEx, which we consider as the cost of activating an NFV-node and bandwidth cost, respectively.

Hence, we perform a techno-economic evaluation of different VNF placement strategies considering a realistic metro network topology under different physical network settings, in terms of number of NFV-nodes and number of wavelengths per link.

The reminder of the paper is organized as follows. In Section II we provide an overview of related works. Section III describes the proposed cost model we considered in the paper. Numerical results obtained comparing different VNF placement algorithms in different network settings are then discussed in Section V. Finally, we conclude the paper in Section VI.

II. RELATED WORK

The problem of VNF placement has attracted the interest of many researchers in recent years. Some of the existing works formulate the problem as an optimization model and provide optimal or near-optimal solutions. For example, authors in [5] provide a binary integer programming model for optimal VNF placement with the objective of minimizing expensive

978-1-7281-0875-9/19/$31.00 ©2019 IEEE

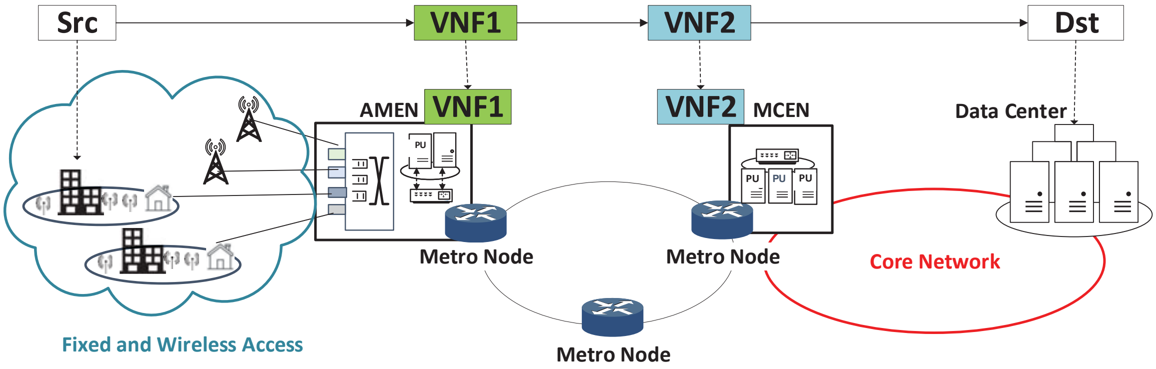


Fig. 1: Overview of the metro network architecture and VNF mapping into the physical network

optical-electrical-optical conversions. Since the problem of VNF placement is proven to be NP-hard [6], they also provide a heuristic algorithm to solve the problem of optimal VNF placement. Ref. [7] proposes an Integer Linear Programming (ILP) model for VNF placement that considers latency requirements of SCs with the objective of minimizing resource utilization. Authors in [8] propose a heuristic algorithm for VNF placement to consider VNF interference caused by VNF consolidation. Their algorithm tries to maximize the throughput of provisioned service requests. Ref. [9] provides two heuristic algorithms for scalable VNF placement to be able to accommodate more dynamically arriving user requests. Ref. [4] provides two algorithms for VNF placement and allocation problem aiming at minimizing number of VNFs deployed while satisfying all the data flows in the network.In [10] authors provide meta-heuristic solution, with the objective of maximizing the utilization of nodes hosting VNFs and minimizing the number of nodes that host VNFs in the network. Ref. [11] proposes a Mixed Integer Linear Programming model for VNF placement that considers QoS requirements of the VNFs and tries to optimize the resource utilization.

A number of existing works have considered deployment cost as a constraint in placing VNFs. For example, Ref. [12] provides two algorithms to perform VNF placement considering both processing capacity limitation of the nodes hosting VNFs and budget constraint. In [13] authors provide an ILP model that satisfies the reliability requirements of the SC with the objective of minimizing SC orchestration cost.

In addition, there have been some efforts performing cost analysis of NFV. For example Ref. [14] provides a techno-economic analysis of a 5G network infrastructure based on SDN/NFV. Authors in [15] present analysis on performance and cost of a cloud network system based on NFV. To the best of our knowledge, none of the existing works has evaluated the impact of the VNF placement strategy in combination with the SLA requirements on the network cost.

III. Network Model

In this paper we consider a metro network architecture as defined in the context of Metro-HAUL project [16] and depicted in Fig. 1. In this optical metro-network there are three categories of nodes: the Metro Core Edge Nodes (MCENs) that are gateways towards core network, Access Metro Edge Nodes (AMENs), that constitute the interfaces between the metro network and heterogeneous access networks (e.g., fixed and/or mobile access), and Metro Nodes (MNs), that represent transport metro nodes and, unlike MCENs and AMENs are not equipped with processing units (PU).

A. Service chaining model

A SC is composed of different VNFs (virtual nodes) connected together using virtual links in a specific order. In order to provision a SC we need to deploy its VNFs on the NFV-nodes, that we assume are chosen among AMENs and MCENs.

In such an optical metro network, we focus on a dynamic traffic environment where SCs are dynamically generated at forwarding nodes, which constitute the source of the SC. Based on the SC type, an NFV-node is chosen as the destination of SC. Moreover, according on the SC type, a specific end-to-end maximum latency and a total required bandwidth characterize the SC being provisioned. As depicted in Fig. 1, to provision a SC, its constituting VNFs are mapped (i.e., deployed) to NFV-nodes and SC traffic traverses the various VNFs in a predefined order, while satisfying the latency and bandwidth requirements of the SC.

Figure 2 shows an example of how two different SC requests are provisioned in the physical network. In this paper we consider an Optical Transport Network (OTN) over Wavelength Division Multiplexing (WDM) architecture, where each node (either MCEN, AMEN or MN) is constituted by a Digital Cross Connect (DXC) over an Optical Cross Connect (OXC). This architecture is being used by China Mobile for its first deployment of 5G metro aggregation networks [17]. The figure shows an upper SC layer, where the sequence of VNFS for the two SCs are highlighted, and the lower OTN/WDM layer.

Assuming the aforementioned OTN/WDM physical architecture, at each transit node the transported traffic is converted from the optical to the electronic domain (OE), it is electronically processed by a DXC and, if necessary, it is groomed (respectively, degroomed) with traffic inserted (resp.

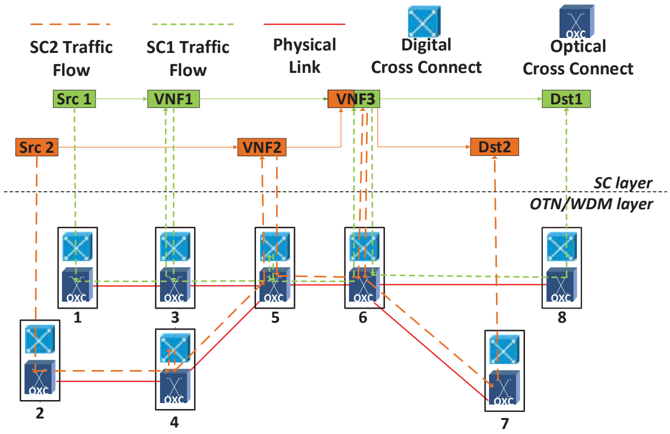


Fig. 2: Node architecture

dropped) locally. Then, it is converted back into the optical domain (EO) and switched towards the next node by an OXC[1]. Moreover, all the nodes are assumed as having full wavelength conversion capability.

In the example of Fig. 2, the two SCs have different source/destination end nodes (i.e., nodes 1-8 for SC1 and nodes 2-7 for SC2), but share the processing units for one of their VNFs, namely, VNF3. As shown in the figure, after the insertion of traffic at the two source nodes, traffic is OE and EO converted in all the transit nodes, also including the ones where no grooming is performed, i.e., nodes 4 and 3, where the traffic of SC1 is also processed at the SC layer by VNF1. Moreover, note that at node 5 the two traffic flows are groomed as they share the physical link between node 5 and 6, and additionally the traffic of SC2 is sent towards the SC layer to perform traffic processing at VNF2. Finally, when arriving at node 6, traffic flows are firs OE converted and processed by the shared VNF3, then they are EO converted and inserted into two different lightpaths to be sent towards the two different destinations.

B. Cost model

In this section we provide a model representing the cost of VNF placement for service chaining. We consider two main contributions to the total cost, i.e.:

- *Active NFV-nodes*: deploying a VNF instance to an NFV-node requires a certain amount of virtual machines to be used. Each of these virtual machines is equipped with limited processing capacity and the VNF deployment leads to operational costs [18], e.g., due to energy consumption and/or software license usage.
- *Bandwidth*: in order to provision a SC, traffic needs to be routed through the network, hence, a certain amount of transport capacity (i.e., a certain number of wavelengths per link) is required for SC provisioning [19].

Therefore the cost related to the provisioning of one SC and the corresponding VNFs placement can be formulated as follows:

[1]Note that traffic is OE and EO converted at any node also in case no traffic needs to be added/dropped at that node

$$C_{SC} = C_{nodes} + C_{wl} \tag{1}$$

where C_{node} represents the cost related to the utilization of NFV-nodes (e.g., due to power consumption, software licences, etc., [20]), and C_{wl} represents the cost of the wavelengths required for traffic transport, namely, due to the transponders to be installed at the nodes [21]. Note that, in this paper, we do not consider the capital expenditures due to the network and computing equipment (i.e., switches, routers and servers) as we assume it is not affected by the VNF placement strategy being adopted. Conversely, as we will detail in the following, the VNF placement strategy has a strong impact on the amount of active NFV-nodes and utilized bandwidth.

IV. Dynamic VNF placement strategy

In this section we briefly describe the VNF placement strategy introduced in our previous work [22] and used for our cost evaluation. We assume dynamically arriving SC requests, where each SC is characterized by its source/destination nodes, the SC type (i.e., the ordered sequence of VNFs to be used, its bandwidth requirement and maximum latency) and its duration. Upon the arrival of a given SC requests, the Dynamic VNF Placement (DVNFP) algorithm will consider current network state (i.e., current set of deployed SCs with the corresponding provisioned lightpaths and VNFs) and try to deploy the incoming request at a minimum cost, while meeting its bandwidth and latency constraints.

The DVNFP algorithm consists of two phases, as described in the following (the reader is referred to [22] for further details). A high level flowchart of the DVNFP algorithm is depicted in Fig. 3.

1) VNF placement. In the first phase, the algorithm performs the VNF placement for each of the VNFs in the SC, which are analysed in order. For each VNF, the algorithm first tries to reuse an already-active VNF instance on an NFV-node which is the closest NFV-node both to the source and destination of the SC request. If such VNF instance is not found, the algorithm tries to activate a VNF instance on one of the NFV-nodes located along the shortest path between source and destination nodes of the SC request (namely, $SP_{s,d}$).

In the case that no NFV-node is available on $SP_{s,d}$, based on the SC requirements, an NFV-node is chosen. In other words, if the SC requires high computational capacity (e.g., for Augmented Reality SC) the NFV-nodes closer to the core nodes are selected as these nodes are more likely to have large computational capacity. However, if the SC has stringent latency requirements, the NFV-node closer to the source of SC will be selected to host the required VNF instance. The above-mentioned step will be repeated for each VNF of SC until an appropriate NFV-node is chosen for all the VNFs in the SC. If it is not possible to find any NFV-node to place VNFs of the SC, the SC request is blocked.

2) Lightpaths provisioning and VNF adjustment. In this phase, the end-to-end latency of the path traversing all the VNFs of the SC in the required order (namely, L_{e2e}) is calculated, considering propagation and switching latency contributions as in [22]. If L_{e2e} is higher than the latency

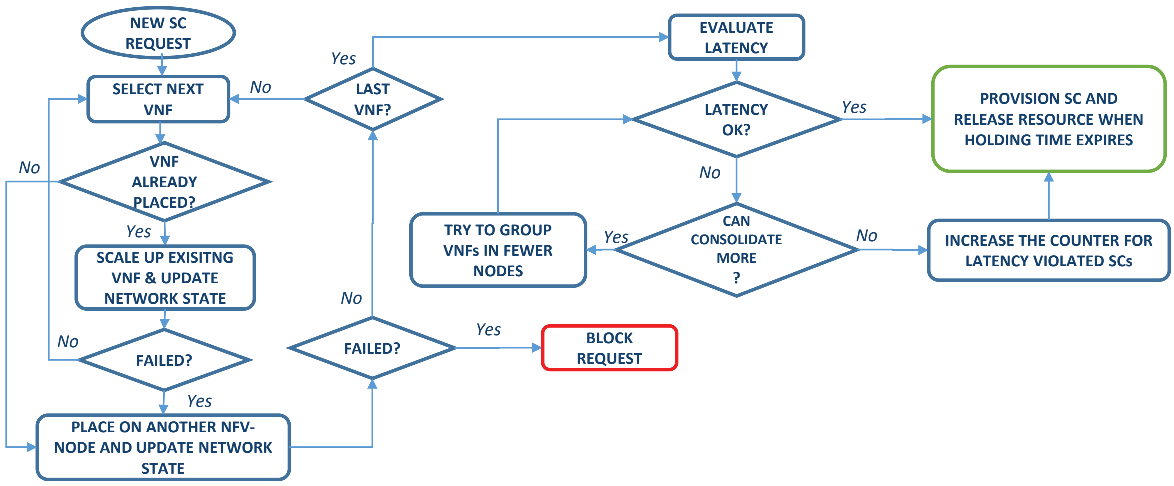


Fig. 3: DVNFP algorithm flowchart

requirements of the SC (L_{SC}), the algorithm tries to shorten the end-to-end path by consolidating (i.e., placing in the same NFV-node) one or more VNFs, until L_{e2e} is lower than or equal to L_{SC}. When all the VNFs of the SC are consolidated on one NFV-node algorithm calculates the L_{e2e} and, if it is still higher than L_{SC}, a counter for latency-violation is increased while the SC is provisioned. Then the SC is provisioned, i.e., its resources (bandwidth and used virtual machines) are allocated. The SC resources are then released after the SC holding time expires.

V. Numerical results

A. Case study and simulation settings

To perform our analysis, we developed a discrete-event-driven simulator in C++. We considered a full OTN network topology as the one in Fig. 4, including 52 nodes and 72 bidirectional WDM links. As depicted in the figure, three types of nodes are present in this network, i.e., Metro Core Backbone (MCB), Metro Core (MC) and Metro Aggregation (MA) nodes. AMEN nodes are chosen among MA nodes and MCEN nodes are chosen among MBC nodes. Each NFV-node is equipped with 64 CPU cores (except the MCB, which is assumed as an NFV-node with unlimited processing resources) and each WDM links supports W wavelengths with 10 Gbit/s capacity. The number of wavelengths per link W is a tuned parameter in this paper, as we vary it to evaluate the bandwidth cost required to support a certain traffic amount with a given SLA. More specifically, in all the following simulations (i.e., for each different VNF placement strategy and in each different NFV-node scenario) the number of wavelengths per link (i.e., the value of W) is set as the minimum number able to support the required SLA.

The types of SCs considered in our evaluation are shown in Table I, where we detail the required VNFs as well as their latency and bandwidth requirements. In addition, the computational requirements of the various VNFs are depicted in Table II in terms of percentage of CPU cores. The considered VNFs are Network Address Translation (NAT), Intrusion Detection

TABLE I: SC and corresponding VNFs, bandwidth and latency characteristics

Service Chain	Service Chain VNFs	Bandwidth	Latency
Augmented Reality	NAT-FW-TM-VO-IDS	100 Mbps	1 ms
MIoT	NAT-FW-IDS	100 Mbps	5 ms
Smart Factory	NAT-FW	100 Mbps	1 ms

TABLE II: Percentage of CPU core usage for various VNFs

VNF Name	NAT	FW	VO	TM	IDS
CPU Core	0.0184	0.018	0.108	0.266	0.214

System (IDS), Firewall (FW), Video Optimizer (VO) and Traffic Monitor (TM).

We simulate the dynamic arrival of SC requests, where the arrival instants are randomly generated considering a Poisson distribution with mean inter-arrival $\lambda = 40$ SC requests per second, while the holding time for each SC is generated according to a negative-exponential distribution with mean μ=1 second[2].

For an incoming SC request, the SC type is randomly selected among these SCs with equal probability, whereas the source node of the SC requests is randomly selected among MA nodes. Moreover, based on the SC type, the destination node can be either among MCB nodes or nodes closer to the source of the SC request. In other words, if the SC has stringent latency requirements the destination node is chosen as close as possible to the source node of the SC request.

All the results are obtained with a confidence level of 95% with at most 5% confidence interval on blocking probability.

B. Evaluation metrics

In order to perform cost analysis, we consider that SLA requirements defined for the users require that the blocking probability (defined as the number of blocked SC request out of total SC requests) is below a certain threshold, initially

[2]This traffic intensity has been chosen as, for the considered network topology and SC characteristics, it can be supporter with a maximum blocking probability of around 10^{-3}.

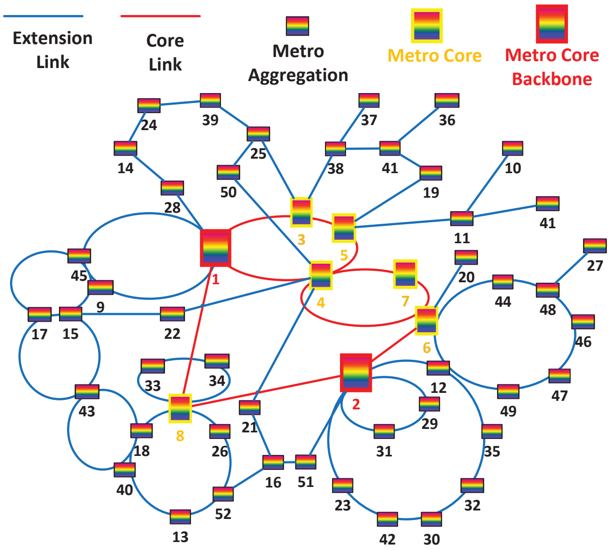


Fig. 4: Network topology

set to 10^{-3}. We consider three different metrics to evaluate the performance of our algorithm with respect to other VNF placement strategies, namely:
i) average number of active NFV-nodes (N_{avg}), that is calculated as follows:

$$N_{avg} = \sum\nolimits_{SC \in SC_{prov}} \frac{N_{SC} \cdot t_{SC}}{t_{tot}} \tag{2}$$

Here, $SC \in SC_{prov}$ represents a generic SC in the set of provisioned (i.e., non-blocked) SC, N_{SC} is the number of active NFV-nodes that have at least one running VNF instance at each time instant, t_{SC} is the time between arrival of two consecutive SC requests and t_{tot} is the total simulation time. Hence, the formula calculates the number of active NFV-nodes weighted by the amount of time each NFV-node is serving; *ii) Number of wavelengths per link*, W; *iii) latency violation ratio*, calculated as the ratio between provisioned SC requests with violated latency out of the total number of provisioned SC requests.

In addition to the above-mentioned metrics, we also considered total network cost for SC provisioning, which is obtained based on the following equation:

$$C_{tot} = \alpha N_{avg} + W \cdot L \tag{3}$$

where $L = 72$ is the total number of links in the network and parameter α captures the relative costs of bandwidth and active NFV-nodes, i.e., the higher α, the higher is the importance of N_{avg} in the overall cost.

C. Benchmark VNF placement strategies

To perform our cost evaluation we also consider two benchmark VNF placement algorithms [22], i.e.:

- *Centralized*: this strategy is used to evaluate the case with the lowest possible number of active NFV-nodes. Specifically, we assume that the network has only one NFV-node with unlimited computational capacity and located at the MCB (node2), so that all the VNFs are embedded at that node. Therefore, in this case, all the SC requests are directed towards a single NFV-nodes which will result in higher bandwidth requirements, due to typically longer routes, and higher blocking, due to potential bottlenecks in links closer to the NFV-node.
- *Distributed*: in this algorithm, the main objective is to reduce network blocking by deploying VNFs along the shortest paths between SCs' source/end nodes. This comes at the cost of higher number of active NFV-nodes, as VNF instances are typically activated in all NFV-nodes.

TABLE III: NFV-node selection scenarios

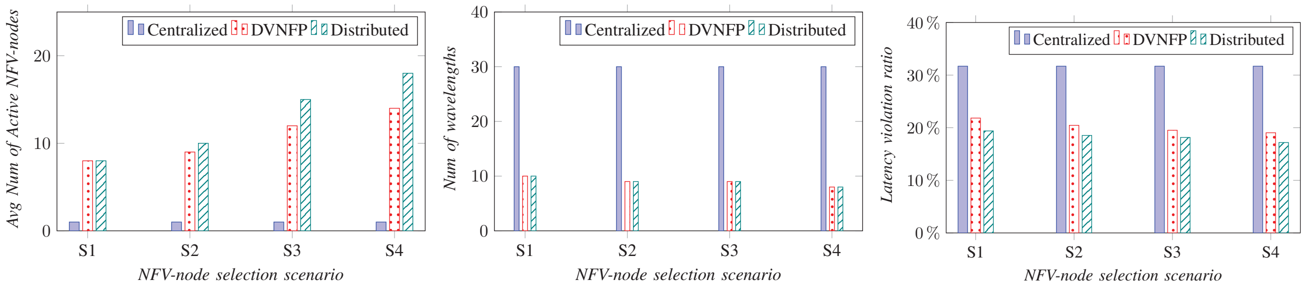
Scenario	S1	S2	S3	S4
% of MCENs	100%	75%	50%	25%
% of AMENs	0%	25%	50%	75%
# of AMENs	0	5	11	17
# of MCENs	6	4	3	2
# of Datacenters	2	2	2	2
Total # of NFV-nodes	8	11	16	21

D. Discussion

We conduct the experiments considering four scenarios, in which a different number of NFV-nodes are present in the network for the cases of *DVNFP* and *Distributed* algorithms. In the first scenario (S1) we considered that all the MCENs are NFV-nodes. The second scenario (S2) indicates that 75% of MCENs and 25% of AMENs are NFV-nodes. Third (S3) and forth (S4) scenarios, respectively, represent the case where half of MCENs and half of AMENs are NFV-nodes and the case where 25% of MCENs and 75% of AMENs are NFV-nodes. It is worth mentioning that the MCB nodes are always considered as NFV-nodes (i.e., they are assumed as datacenters with unlimited computational capacity). We summarize these scenarios in Table III.

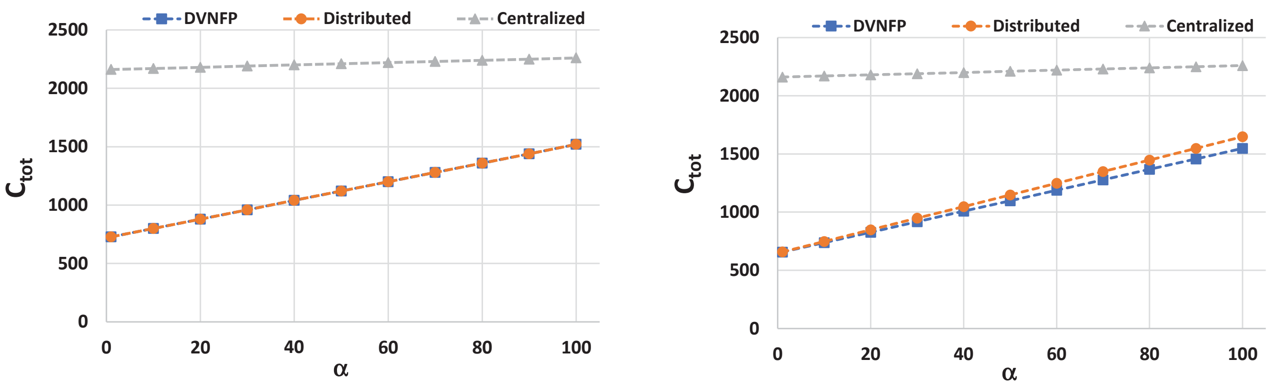
Figure 5 illustrates the comparison between the different VNF placement strategies considering a target blocking probability equal to 10^{-3}. The average number of active NFV-nodes is plotted in Fig. 5(a) for the three strategies. Note that for *DVNFP* the number of active NFV-nodes is up to 22%less than *Distributed*. This is due to the fact that *DVNFP* tries to reuse the already active NFV-nodes as much as possible and demonstrates the importance of an effective VNF placement strategy.

On the other hand, as depicted in Fig. 5(b), in order to satisfy the SLA requirements (i.e., the maximum service blocking probability), for all the scenarios S1 to S4, the *DVNFP* and *Distributed* algorithms require the same number of wavelengths per link W. This means that, for the considered traffic scenarios, the DVNFP algorithm is able to reduce the cost of active NFV-nodes without impacting on the required bandwidth. From the figure it is evident that the *Centralized* algorithm has the lowest performance in terms of bandwidth cost, as it requires 30 wavelengths per link, i.e., three times more than *DVNFP* and *Distributed* cases. This is due to the fact that, with the *Centralized* strategy, only one NFV-node is used to perform VNF placement, therefore a higher number of wavelengths is needed to avoid congestion at the links in its proximity.

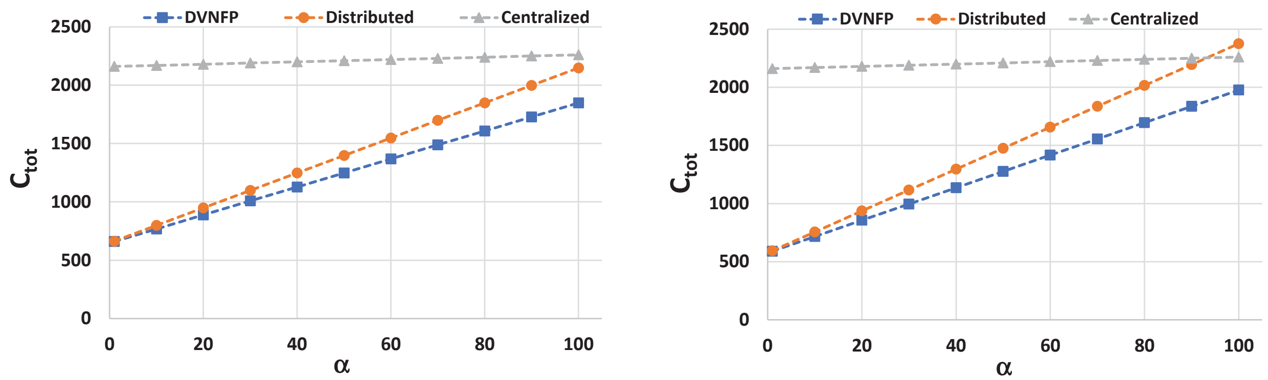


(a) Average number of active NFV-nodes (b) Number of wavelengths per link (c) Latency violation ratio

Fig. 5: Comparison among different strategies



(a) NFV-nodes S1 scenario (b) NFV-nodes S2 scenario



(c) NFV-nodes S3 scenario (d) NFV-nodes S4 scenario

Fig. 6: Total cost comparison for different VNF placement strategies and NFV-nodes scenarios

Latency violation ratio is shown in Fig. 5(c) for the various cases. As shown in the figure, the *Distributed* strategy achieves the best performance, due to the fact that SC requests are provisioned deploying VNFs closer to the source node of the request. Therefore, the shortest path between source and destination of the SC request is often used for traffic routing, which leads to the lower latency violation ratio. Finally, the violation ratio obtained in the *DVNFP* case is slightly higher than the *Distributed* case (around 1-2% higher), and is independent from the NFV-node selection scenario.

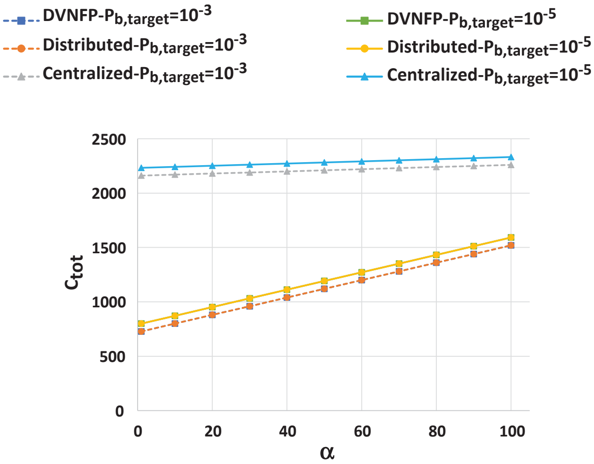
Now we evaluate the impact of the different strategies on the total network cost described in section V-B. In Fig. 6 we show how the total network cost, as defined in eq. 3, is affected by the different VNF placement strategies, tuning the parameter α from 1 to 100 to capture the importance of the two cost contributions in eq. 3.

It is evident that in most cases the *Centralized* strategy provides the highest overall network cost, whereas *DVNFP* one is in general the most cost-effective solution. However, for increasing α, i.e., when the cost of active nodes becomes more relevant than the wavelengths cost, the difference between *DVNFP/Distributed* and *Centralized* strategies is reduced. In general (i.e., except for the S1 scenario), this reduction is quicker for the *Distributed* case than for the *DVNFP* one, due to the difference in the number of active nodes between the two cases.

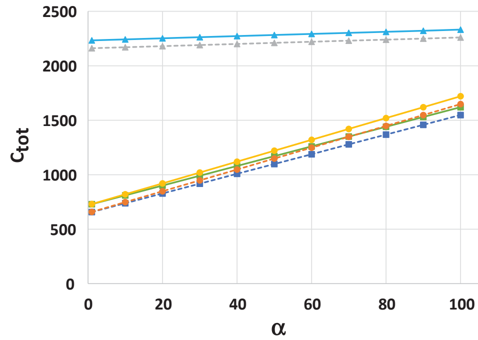
Moreover, the difference between the *Centralized* and other VNF placement strategies becomes lower as we move from S1 to S4. In particular, as depicted in Fig. 6(a) for the S1 scenario, i.e., when only eight NFV-nodes are deployed, the total cost for *DVNFP* and *Distributed* strategies is the same, even for increasing α. This is due to the equal values of N_{avg} and L required in the S1 scenario for the two strategies, as we observed in Fig. 5. However, increasing the number of available NFV-nodes, i.e., changing the network deployment from scenario S1 towards S4, we can see the difference between *DNVFP* and *Distributed* strategies increases up to 16% for S4 scenario when the relative cost of active nodes is high ($\alpha = 100$). This is due to the fact that *DVNFP* is capable of satisfying the same SLA (i.e., guarantee the same maximum blocking probability) as *Distributed* using the same number of wavelengths per link but using less active NFV-nodes. It is noteworthy that the *Centralized* strategy has almost always the highest total network cost (up to 66% higher than the other strategies in the S1 scenario and with $\alpha = 1$), since it requires almost three times more wavelengths per link with respect to the *DNVFP* and *Distributed* strategies. However, for S4 and for higher values of α (e.g., for $\alpha = 100$), since *Distributed* activates on average 18 NFV-nodes and uses 8 wavelengths per link, the total network cost for *Centralized* (which requires 1 NFV-node and 30 wavelengths per link), is lower if compared to the *Distributed* case. It is worth noting that, for the *Centralized* strategy, the various NFV-nodes scenarios provide almost no impact on the total cost, even for increasing α, as the most relevant cost contribution in this case is constituted by the number of wavelengths in the network.

Comparing the various costs obtained for the *DVNFP* strategy, the maximum difference between the various NFV-nodes scenarios and for the lowest value for α (i.e., $\alpha = 1$) is equal to 18% and occurs between S1 and S4. This demonstrates that a higher number of available NFV-node, in general, does not always lower the network cost, although it may guarantee higher flexibility in VNFs placement and thus higher efficiency in network capacity utilization, due to the fact that shorter routes can be used for SC provisioning. The cost gap between S1 and S4 cases for the *DVNFP* strategy increases up to 23% for the highest value of α considered ($\alpha = 100$). The reason is that, for increasing α, the cost of activating one NFV-node has higher impact, as in the S4 scenario more NFV-nodes are activated with respect to S1.

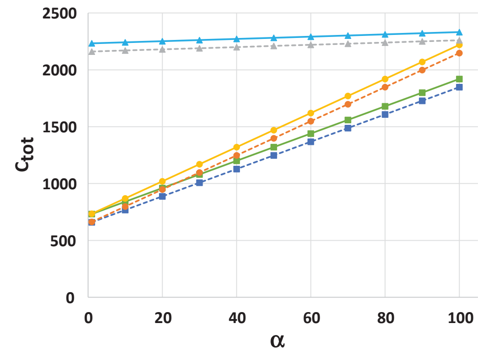
We perform a similar analysis considering a more stringent SLA requirement, namely, a maximum blocking probability target of 10^{-5}. Results are shown in Fig. 7 for the various cases. As shown in the figure, a more stringent SLA produce a fixed increase in the overall cost for all the strategies and in all the NFV-nodes scenarios, even for increasing values of alpha. This means that the SLA variation impacts only the number of wavelengths per link, without impacting the



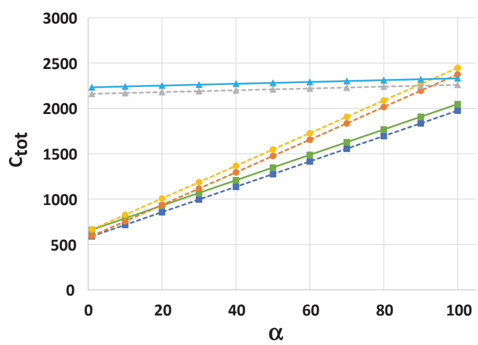
(b) NFV-nodes S1 scenario



(c) NFV-nodes S2 scenario



(d) NFV-nodes S3 scenario



(e) NFV-nodes S4 scenario

Fig. 7: Impact of SLA on total network cost

average number of active nodes.

VI. Conclusion

In this paper we presented a techno-economic analysis of different VNF placement strategies for SC provisioning, considering different NFV-nodes deployments and SLAs under realistic optical metro network topology and traffic assumption. For the considered traffic, results show that an efficient placement strategy can reduce the cost of service provisioning up to 16% or 23%, according to the various NFV-nodes deployments.

Acknowledgment

The work leading to these results has been supported by the European Community under grant agreement no.761727 Metro-Haul project.

References

[1] J. Matias, J. Garay, N. Toledo, J. Unzilla, and E. Jacob, "Toward an SDN-enabled NFV architecture," *IEEE Communications Magazine*, vol. 53, no. 4, pp. 187–193, 2015.

[2] Network functions virtualisation. [Online]. Available: https://www.etsi.org/technologies/nfv

[3] L. Peterson, A. Al-Shabibi, T. Anshutz, S. Baker, A. Bavier, S. Das, J. Hart, G. Palukar, and W. Snow, "Central office re-architected as a data center," *IEEE Communications Magazine*, vol. 54, no. 10, pp. 96–101, 2016.

[4] Y. Sang, B. Ji, G. R. Gupta, X. Du, and L. Ye, "Provably efficient algorithms for joint placement and allocation of virtual network functions," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE, 2017, pp. 1–9.

[5] M. Xia, M. Shirazipour, Y. Zhang, H. Green, and A. Takacs, "Network function placement for NFV chaining in packet/optical datacenters," *Journal of Lightwave Technology*, vol. 33, no. 8, pp. 1565–1570, 2015.

[6] F. Bari, S. R. Chowdhury, R. Ahmed, R. Boutaba, and O. C. M. B. Duarte, "Orchestrating virtualized network functions," *IEEE Transactions on Network and Service Management*, vol. 13, no. 4, pp. 725–739, 2016.

[7] D. B. Oljira, K.-J. Grinnemo, J. Taheri, and A. Brunstrom, "A model for QoS-aware VNF placement and provisioning," in *2017 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*. IEEE, 2017, pp. 1–7.

[8] Q. Zhang, F. Liu, and C. Zeng, "Adaptive interference-aware VNF placement for service-customized 5G network slices," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019, pp. 2449–2457.

[9] W. Zhou, Y. Yang, M. Xu, and H. Chen, "Accommodating dynamic traffic immediately: A VNF placement approach," in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*. IEEE, 2019, pp. 1–6.

[10] S. Tavakoli-Someh and M. H. Rezvani, "Multi-objective virtual network function placement using NSGA-II meta-heuristic approach," *The Journal of Supercomputing*, pp. 1–37, 2019.

[11] F. B. Jemaa, G. Pujolle, and M. Pariente, "QoS-aware VNF placement optimization in edge-central carrier cloud architecture," in *2016 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2016, pp. 1–7.

[12] G. Sallam and B. Ji, "Joint placement and allocation of virtual network functions with budget and capacity constraints," in *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*. IEEE, 2019, pp. 523–531.

[13] X. Zhong, Y. Wang, and X. Qiu, "Cost-aware service function chaining with reliability guarantees in NFV-enabled Inter-DC network," in *2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM)*. IEEE, 2019, pp. 304–311.

[14] C. Bouras, P. Ntarzanos, and A. Papazois, "Cost modeling for SDN/NFV based mobile 5G networks," in *2016 8th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*. IEEE, 2016, pp. 56–61.

[15] M. Ananth and R. Sharma, "Cost and performance analysis of network function virtualization based cloud systems," in *2017 IEEE 7th International Advance Computing Conference (IACC)*. IEEE, 2017, pp. 70–74.

[16] The Metro-Haul project deliverables. [Online]. Available: https://metro-haul.eu/deliverables/

[17] C. Zhang. (2018) "Optical Networking in the Cloud and 5G Era," Keynote talk at Optical Fiber Communications Conference and Exhibition (OFC). [Online]. Available: https://www.ofcconference.org/en-us/home/ news-and-press/ofc-video-library

[18] B. Addis, G. Carello, F. De Bettin, and M. Gao, "On a virtual network function placement and routing problem: properties and formulations," 2017.

[19] A. Leivadeas, M. Falkner, I. Lambadaris, G. Kesidis, C.-H. Lung, and M. Ibnkahla, "Considerations for a successful network service chain deployment," in *2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft)*. IEEE, 2018, pp. 288–292.

[20] A. Greenberg, J. Hamilton, D. A. Maltz, and P. Patel, "The cost of a cloud: research problems in data center networks," *ACM SIGCOMM computer communication review*, vol. 39, no. 1, pp. 68–73, 2008.

[21] P. Yi, H. Ding, and B. Ramamurthy, "Cost-optimized joint resource allocation in grids/clouds with multilayer optical network architecture," *IEEE/OSA Journal of Optical Communications and Networking*, vol. 6, no. 10, pp. 911–924, 2014.

[22] L. Askari, A. Hmaity, F. Musumeci, and M. Tornatore, "Virtual-network-function placement for dynamic service chaining in metro-area networks," in *2018 International Conference on Optical Network Design and Modeling (ONDM)*. IEEE, 2018, pp. 136–141.