This is a postprint version of the following published document:

M. Masoumi *et al*., "Dynamic Online VNF Placement with Different Protection Schemes in a MEC Environment," *2022 32nd International Telecommunication Networks and Applications Conference (ITNAC)*, Wellington, New Zealand, 2022, pp. 1-6, https://doi.org/10.1109/ITNAC55475.2022.9998347

Dynamic Online VNF Placement with Different Protection Schemes in a MEC Environment

Maryam Masoumi
Universidad de Valladolid
Valladolid, Spain
0000-0002-3832-9106

Ignacio de Miguel
Universidad de Valladolid
Valladolid, Spain
0000-0002-1084-1159

Ramón J. Durán Barroso
Universidad de Valladolid
Valladolid, Spain
0000-0003-1423-1646

Lidia Ruiz
Universidad de Valladolid
Valladolid, Spain
0000-0001-6241-5998

Fabrizio Brasca
Architecture
Wind Tree SpA
Milan, Italy

Gianluca Rizzi
Architecture
Wind Tree SpA
Rome, Italy

Noemí Merayo
Universidad de Valladolid
Valladolid, Spain
0000-0002-6920-0778

Juan Carlos Aguado
Universidad de Valladolid
Valladolid, Spain
0000-0002-2495-0313

Patricia Fernández
Universidad de Valladolid
Valladolid, Spain
0000-0001-5520-0948

Rubén M. Lorenzo
Universidad de Valladolid
Valladolid, Spain
0000-0001-8729-3085

Evaristo J. Abril
Universidad de Valladolid
Valladolid, Spain
0000-0003-4164-2467

***Abstract*— The Multi-access Edge Computing (MEC) architecture is made up of geographically distributed edge servers so that computing capabilities are provisioned at the network edge, close to the end users. Network Function Virtualization (NFV), when combined with MEC, provides network services in the form of Service Function Chains (SFC) with low latency. In the design of NFV-based 5G networks, the trade-off between the cost of resource deployment and the effective provisioning of services must be considered. In this work, we analyze the impact of having different MEC locations when considering the provision of SFCs in a dynamic scenario (and thus also address VNF placement). In order to deal with infrastructure failures, it is of great importance to employ robust and resilient network strategies. To safeguard SFCs against failures, various protection techniques can be applied. We use two protection methods, namely, dedicated VNF protection and shared VNF protection, under the assumption of single network failures. The operational performances of different approaches are evaluated in terms of blocking ratio and end-to-end delay, both for the whole network and for different services, and we analyze whether it is better to distribute computing servers among a few MEC sites or among a higher number.**

***Keywords*— NFV, MEC, VNF Placement, Protection.**

I. INTRODUCTION

Software Defined Networking (SDN) is a networking model that eliminates the shortcomings of traditional network infrastructures by decoupling the control plane and data plane from switches and routers. Enabling network control via centralized software controllers, SDN makes network management more efficient, fast, and flexible [1]. Network Function Virtualization (NFV), as an emerging technology, makes new device deployment more flexible and efficient by separating software from traditional hardware devices [2]. Multi-access edge computing (MEC) is another technology that is expected to have great impact in 5G networks in order to meet the ultra-low latency requirements of certain applications and services while reducing transport network load [3]. The telecommunication industry is rapidly moving towards becoming completely virtualized. Therefore, virtualization is also seen as a critical component of 5G, with SDN, NFV, and MEC technologies being used to provide virtual segmentation of mobile radio access, virtual core networks, and network slicing. A virtual network function (VNF) is a software implementation of a network function, such as a firewall, router, load balancer, or mobile core network component. VNFs may be instantiated and executed in the data plane as virtual machines or containers hosted in dedicated infrastructures such as the cloud or MEC sites. Services are usually provided by means of a chain of several VNFs, thus creating a service function chain (SFC).

However, provisioning resources to the SFCs remains a challenging problem, particularly addressing latency and resource consumption needs. The provision of resources in the context of SFCs involves addressing multiple sub-issues, including VNF placement [3]. The objective of the VNF placement problem is to find an appropriate location for each VNF in a MEC server. Most of previous works have addressed VNF placement in a static context where the traffic loads, network services, and the number of requests are constant. Existing literature moreover emphasizes placement algorithms in dynamic contexts but still, there are some gaps such as maintaining network performance and latency, as well as additional security threats. In [4], an efficient SFC placement is formulated in a MEC-NFV environment which aims at maximizing resource utilization. However, the users' requirements for latency, failures of network components, and resilience issues are not taken into account. With the goal of reducing end-to-end latency, the dynamic latency-optimal VNF placement problem is considered in [5], but network components' survivability and resiliency aspects are not considered there.

In this paper, we propose a dynamic VNF placement algorithm aiming at addressing these gaps, using two different protection techniques to provide survivability. The scenario considered in this paper assumes several MEC sites distributed in different geographical locations with a constrained computational capacity. It is of paramount importance to provide reliability in a MEC infrastructure, since a failure in the service chain, server, or MEC site can

This work is part of IoTalentum project that has received funding from the EU H2020 research and innovation programme under the MSCA grant agreement No 953442. It is also supported by Consejería de Educación de la Junta de Castilla y León and the European Regional Development Fund (Grant VA231P20), and the Spanish Ministry of Science of Innovation and the State Research Agency (Grant PID2020-112675RB-C42 funded by MCIN/AEI/10.13039/501100011033).

lead to service outage. Hence, dynamic resource provisioning should guarantee reliability in order to deal with failures. This is where the protection schemes, including dedicated and shared backup strategies, come into play. In this work, protection strategies select a backup MEC site so as to cope with failures even when the whole primary MEC site is affected. Thus, in VNF placement in an NFV architecture, the main issues include i) how to find VNF placement to satisfy the delay requirements of the requested services, and ii) how to place backup VNFs in order to guarantee a reliable network if there is any failure in the network.

The key contributions of this paper can be summarized as follows. 1) We consider the impact of different numbers of MEC locations on network performance. 2) We formulate the VNF placement problem without any protection methods. 3) We utilize a dedicated SFC backup strategy to solve the VNF placement problem. 4) We also adopt a shared VNF backup mechanism to guarantee the reliability of network services. In order to evaluate the performance of the proposed backup methods, a real-world network topology is used.

The rest of this paper is organized as follows. In Section II, we review previous works on VNF placement. In Section III, we describe the system model considered in this paper. Section IV explains the VNF placement problem and the proposed reliability-aware algorithm for VNF backup placement. Then, in Section V, we describe the results of simulations conducted to evaluate the performance of proposed protection algorithms. Finally, Section VI concludes the paper.

II. Related Works

The issue of VNF placement has been extensively studied in recent years. Some studies have put a lot of effort into finding the optimal solution to the VNF mapping problem, which is an NP-hard problem [1]. Some approaches use integer linear programming (ILP) techniques to find the best solution, while others rely on heuristic algorithms to solve the problem in a feasible computational time. Many researchers have investigated the issue of VNF placement with the aim of increasing Network Service Providers (NSP) economic gains or lowering capital expenditure. It can be achieved by maximizing network throughput, minimizing the number of active VMs, or reducing the consumption of computational and bandwidth resources for network service deployment while others have studied the VNF placement problem by considering network service delay requirements. In [7], the two traditional 1:N and 1:1 protection schemes are investigated, and backup resources are considered for the MEC servers. In [8], a deep reinforcement learning–based (DRL) online framework is designed for placing VNF automatically.

In [9], the VNF placement problem for Poisson-based traffic is formulated as a 0-1 quadratic fractional programming problem. In [10] and [11] some redundant backup approaches were chosen to ensure network service reliability. With these approaches, each VNF is assigned an independent backup, ensuring the reliability requirements of each request. Qu *et al.* in [10] concentrate on VNF placement with reliability and multipath flow routing for general SFCs and multisource multicast network services. Karimzadeh-Farshbafan *et al.* in [12] considered both VNF and backup VNF placement at the same time. Some other studies [13], [14] have investigated how to share backup VNFs among requests to ensure service reliability. A backup approach is introduced for both computational resources and link bandwidth sharing. However, the request delay requirement was not considered. In practice, services arrive at the system dynamically with varying QoS requirements, and the underlying physical infrastructure's conditions change over time. A static service provisioning strategy is incapable of meeting the dynamic properties of the services and may result in inefficient resource utilization. As a result, dynamic VNF placement algorithms are critical for improving the overall performance of various service provisioning in SDN/NFV-enabled environments. Thiruvasagam *et al.* [6] consider different dedicated backup methods in MEC-enabled networks, focus on planning, and consider the provisioning of SFC in a static context.

In our work, the main focus is on the control and dynamic allocation of SFCs. Service requests arrive dynamically at the network triggering the establishment of SFCs. Moreover, in addition to dedicated protection, shared VNF protection is evaluated here. Thanks to it, it is possible to save computational resources while allowing more requests to be admitted and ensuring their reliability. Furthermore, we also guarantee that any backup SFC can satisfy the request's delay requirement.

III. Problem Statement

This paper considers a 5G network with an optical network backhaul which uses a distributed MEC infrastructure. Each of those MEC sites makes use of virtualization to host VNF chains, and thus handle service requests, but has constraints on computational and storage capacities as well as bandwidth. Different services have specific resource requirements, which affect the number of resources requested. VNF failures interrupt VNF-dependent services, while in the case of MECs, the operation of multiple VNFs is susceptible to being interrupted, which leads to the failure of several services. As a result, it is critical to devise a solution to map dynamic service requests onto MECs, while also considering protection schemes to accommodate services.

We assume that MEC sites are selected so that the average delay when nodes communicate with their closest MEC site is minimized. Then, the formulation of the dynamic MEC resource allocation problem is considered, at first, without any protection methods. Afterwards, protection methods such as dedicated VNF backup and shared VNF backup are applied to guarantee resiliency, security, and availability of the established services. We consider the case of protection against single failure like in most of the works of literature. The objective is to minimize the blocking ratio (i.e., ratio of non-established requests) of a given service request, and the main research question that we address in this manuscript is whether, given a fixed set of computing resources, it is better to place them in a few MEC locations or distribute them among a higher number of MEC sites.

A. Network Model

The network is modeled as an undirected graph $G = (N, L)$, where N and L respectively represent the set of base stations and physical links interconnecting the base stations. To set up MEC locations, a small subset of base stations is selected and denoted by M, where $M \subset N$. There is a finite number of servers S at each MEC location that accommodate services for user requests. The available resource capacity at each server ($s \in S$) is measured in terms of CPU cores and RAM. Following [6], each service request consists of a set of VNFs composing an SFC which can be created on top of the MEC servers to serve a particular request. We assume, as in [6], that the whole SFC must be placed in a single MEC location. Obviously, multiple SFCs can be placed on a single MEC server, as long as there are enough available resources.

Each VNF type necessitates a specific number of resources to process incoming traffic. A user connects to the network via a nearby base station, and requested services are routed through this base station. Service requests arrive over time, and the embedding algorithm should determine whether or not the VNFs within the requested service can be mapped to physical network components. When a request is accepted, the necessary resources are assigned for the establishment of the SFC, and released when the request expires. Three different scenarios are considered: (a) unprotected operation, (b) use of dedicated SFC protection, and (c) use of shared VNF protection.

B. SFC Request Model (Service Request)

We consider services with strict ordering of VNFs constituting the request. We model each SFC request r as (V^r, $\tau^r{}_a$, $\tau^r{}_d$, b^r, d^r, n^r) in which V^r is the set of VNFs in the SFC. The terms $\tau^r{}_a$ and $\tau^r{}_d$ denote the arrival time and the lifetime of the SFC request, respectively; b^r specifies the bandwidth requirement for the service type r, and d^r shows the maximum allowed end-to-end latency for the service type r. Finally, $n^r \in N$ represents the base station to which the user who requires service r is attached. In order to protect SFCs against failures and ensure reliable performance, it is essential to allocate backup resources to each primary SFC.

We assume, like in most of the previous works, protection against a single failure. In this study, two ways of protecting SFCs and VNFs are considered: dedicated SFC protection and shared SFC protection. In dedicated SFC protection, an SFC is protected with a set of backup resources which are protecting only that service chain, that is, a backup SFC only protects one primary SFC. However, a more efficient use of the resources can be done using shared protection. A shared backup VNF can protect multiple primary VNFs (of different SFCs) if the primary VNFs are located in different MEC sites, thus avoiding service unavailability problems in the event of a single failure. When a service request is received, resources must be reserved for the primary SFC, but also for the backup SFC. If it is not possible to reserve resources for either the primary or the backup SFC, then the service is blocked.

C. Constraints

There are various types of constraints that must be considered when provisioning service requests. We classify them into three categories:

1) *Bandwidth requirement*: The total bandwidth required to fulfil all service requests should not exceed the available bandwidth at a MEC location.

2) *Latency requirement*: The maximum allowed end-to-end latency for the primary and the backup SFC for each service request (including propagation and processing times) should not exceed the delay requirement of the request. The end-to-end latency for the primary ($d_{e2e-\text{primary}}$) and the backup SFCs ($d_{e2e-\text{backup}}$) are calculated here as follows:

$$d_{\text{e2e-primary}} = 2d_{n,m1} + d_{SFC} = 2\,\frac{l_{n,m1}}{v_g} + V\beta \qquad (1)$$

$$d_{e2e-backup} = 2d_{n,m2} + d_{SFC} = 2\,\frac{l_{n,m2}}{v_g} + V\beta \qquad (2)$$

$d_{n,m1}$ and $l_{n,m1}$ are the communication delay and distance between the base station to which the user is connected and the primary MEC site (in terms of millisecond and kilometers), while $d_{n,m2}$ and $l_{n,m2}$ are delay and distance between base station to which the user is connected and the backup MEC site. v_g (group velocity) is the propagation speed through the physical medium connecting the nodes (e.g., ~$2{\cdot}10^8$ m/s for optical fiber). Moreover, each VNF adds some additional delay in terms of milliseconds whic is denoted by d_{SFC}. It is expressed by β, and V is the number of VNFs composing the SFC.

3) *Resource requirement*: The total amount of resources allocated to each service should be less than the number of available resources at the MEC server.

If any of these constraints cannot be met, the service request is blocked.

IV. MEC Location and VNF Placement Heuristic

As previously mentioned, M out of N network nodes will host MEC resources ($M \subset N$). Those M nodes are selected so that when each node $n \in N$ communicates with the closest node equipped with MEC resources $m \in M$, the average delay in the whole network is minimized.

Regarding service provisioning, the whole SFC is placed in a single MEC site. For the first scenario (unprotected operation), for each request only a primary SFC is established (there is no backup SFC), and the nearest MEC node to the user requesting the service is selected to host the SFC. The first server having sufficient available CPU and RAM capacity within the MEC site is selected to provide the service (first-fit policy), and the required CPU and RAM are reserved during the duration of the service. If the algorithm fails to find a server with available resources in this MEC, it will search among the other nearest MECs to find a server with enough capacity. The request will be blocked if no server can be found with enough resources (or if any of the other constraints described in Section III.C is not met).

In the second and third scenarios (when dedicated and shared protection are applied), the first nearest MEC to the user is selected as the primary MEC and the second closest one is considered to host the backup MEC. Again, if not enough resources are available in the preferred MEC site, the following nearest MEC site will be considered.

When dedicated protection is used, a backup VNF will only protect one primary VNF of a SFC (one service request). The first-fit policy previously mentioned is used here within each MEC site for both the selection of the server hosting the primary VNF (in the primary MEC) and the backup VNF (in the backup MEC site). In contrast, when shared protection is used, a backup VNF can protect multiple primary VNFs as long as those primary VNFs are located in different MEC sites. Thus, when searching how to assign backup resources, the heuristic first searches whether a backup VNF of the same type is already available at the backup MEC site (and it is not protecting any primary VNF located in the same primary MEC site than the requesting one). If that is the case, it can be reused and there is no need to reserve additional CPU or RAM. If the algorithm fails to find a suitable backup VNF for being shared, a new instance of the VNF should be created.

When the lifetime of a request expires, all resources employed by the primary and the backup SFC are released except if shared protection is used. When shared protection is used, the resources employed by a backup VNF are only released when the last SFC which uses that backup VNF for protection is released.

V. Performance Evaluation

A. Simulation Settings

This section describes the simulation scenarios considered in this work. The purpose of the considered scenarios is to evaluate the performance of the proposed algorithms. All the scenarios were evaluated using simulations implemented in Python and running on a desktop computer with the Windows operating system and the following features: 11th Gen Intel(R) Core (TM) i7-11800H @ 2.30GHz and 16.0 GB of RAM. Libraries such as Networkx, Numpy, matplotlib, and Pandas are used for graph-based and numerical implementations.

The reference topology used for performance assessment is a metropolitan network topology from the northern area of Italy with 51 nodes and 67 optical fiber links (Fig. 1) [17]. The network diameter (distance between the two more distant nodes) is 101.6 km. Each node is assumed to be a base station node. As shown in Table I, we consider three cases in which the number of MEC nodes differs and are equal to 3, 5, and 7. For the scenarios with three and seven MEC sites, the location of those sites minimizes the average communication delay of the nodes with their closest MEC. However, for the case with five MECs, we assumed an evolutionary upgrade of the network, where two new sites were added as an intermediate step towards the final configuration with seven MEC sites.

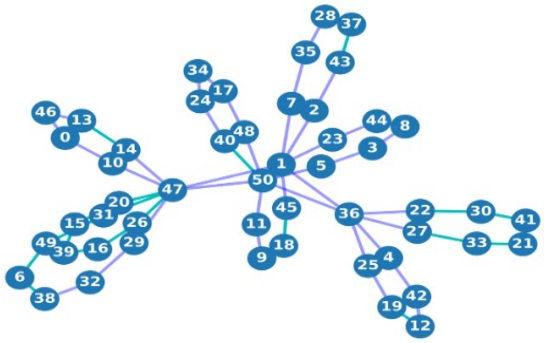

Fig. 1. Network Topology

TABLE I. Location of MEC Sites in the Analyzed Cases

Number of MEC locations	MEC locations (nodes in Fig. 1 hosting MEC servers)	Servers per MEC	CPU per Server	RAM per Server
3 MECs	1, 36, 47	105	256 cores	256 GB
5 MECs	1, 36, 47, 7, 10	63	256 cores	256 GB
7 MECs	1, 36, 47, 7, 10, 21, 23	45	256 cores	256 GB

It should be noted that the three scenarios have an equal number of computing resources. In all cases, the total number of servers is 315, and each server has 256 CPU cores and 256 GB of RAM. Thus, in the 3-MEC case there are 315/3 = 105 servers per MEC, while in the 5-MEC and 7-MEC scenarios there are 63 and 45 servers per MEC, respectively.

We assumed that each user is connected to one base station. Four different types of network services can be requested by the user: Augmented Reality/Virtual Reality (AR/VR), vehicle-to-everything communications (V2X), electronic health (e-health), and 8K TV and Gaming. The corresponding SFC (assuming generic names), bandwidth, and delay requirements for each network service are listed in Table II. Additionally, each instance of a VNF has associated IT requirements in terms of CPU cores and RAM, which are shown in Table III. We assume that the user service request through the base station is uniformly distributed with an equal probability of 25% for the four service types.

TABLE II. Service Chain Requirements [15], [16].

Service	Chained VNF	Bandwidth (Mbps)	Latency (ms)
AR/VR	VNF0, VNF3	200	2
V2X	VNF1, VNF3, VNF4	100	3
e-health	VNF0, VNF3	50	5
8k TV and gaming	VNF0, VNF1, VNF2, VNF3, VNF4	250	10

TABLE III. Hardware Requirements Associated to the VNFs.

VNF	HW Requirements
VNF0	CPU: 1 core, RAM: 4 GB
VNF1	CPU: 2 cores, RAM: 3 GB
VNF2	CPU: 4 cores, RAM: 3 GB
VNF3	CPU: 3 cores, RAM: 4 GB
VNF4	CPU: 2 cores, RAM: 1 GB

Service requests arrive at the network randomly according to a Poisson process. Interarrival time is generated with an exponential distribution with parameter λ. The incoming node for each request is randomly selected using a uniform distribution. Each request has a lifetime which is exponentially distributed with T = 60 s. In this study, we define the load based on some parameters including the average lifetime of each request, the average interarrival time, and the number of nodes in the topology as follows:

$$load = \frac{\lambda T}{N(N-1)} \quad (3)$$

For latency assessment, since links between nodes are optical fibers, the propagation speed (v_g) is $2 \cdot 10^8$ m/s. The value of β (delay introduced per VNF) is equal to 0.050 milliseconds. In this initial work, for the sake of simplicity, we assume that there are no constraints due to bandwidth requirements but only due to computing resources and delay.

In order to evaluate and compare the simulation with and without protection methods, we measure the following performance metrics:

i) Total blocking ratio and blocking ratio of each service: It measures the ratio of the blocked services due to lack of resources or latency violations.

ii) Average end-to-end delay and end-to-end delay of each service.

C. Results Evaluation

The simulations have been carried out considering the three previously mentioned scenarios: unprotected operation, operation with dedicated backup, and operation with shared backup. In all the scenarios, for each load, a total number of 110,000 requests have been dynamically generated (the first 10,000 used to warm-up the simulator and the remaining 10^5 to analyze performance).

The graph in Fig. 2 illustrates the comparison in blocking ratio, which is related to the different protection cases and the use of 3, 5, or 7 MEC locations. Obviously, the unprotected approach leads to lower blocking probabilities as all available resources are completely devoted to establishing primary SFCs. In contrast, dedicated protection is the worst approach in this sense, as it requires reserving resources for backup SFCs (and does not allow any resource sharing). Regarding the impact of the number of MEC sites, it can be seen that by decreasing the number of MEC sites, better performance is obtained, especially when the traffic load is low. Thus, the use of 3 MEC sites leads to the lowest blocking ratio (with and without protection methods), and the use of 7 MEC sites leads to the highest blocking ratio. Therefore, concentrating the computing resources in a lower number of sites leads to lower blocking probabilities, at least in the metropolitan networking scenario considered in this paper.

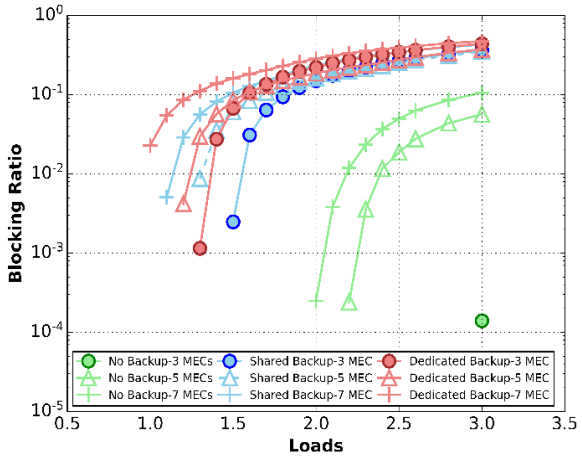

Fig. 2. Comparison in Blocking Ratio between Scenarios with Different Numbers of MECs

Fig. 3 compares the blocking ratio of different services in no protection, dedicated SFC protection, and shared VNF protection scenarios for the 3-MEC sites case. The blocking ratio for 8k TV and Gaming service is significantly higher than other services since it possesses the longest chain of VNFs and thus requires a higher number of resources. In particular, for the unprotected scenario, that is the only service that has some blocking events for a traffic load = 3, and for that reason it is the only unprotected service represented in that figure.

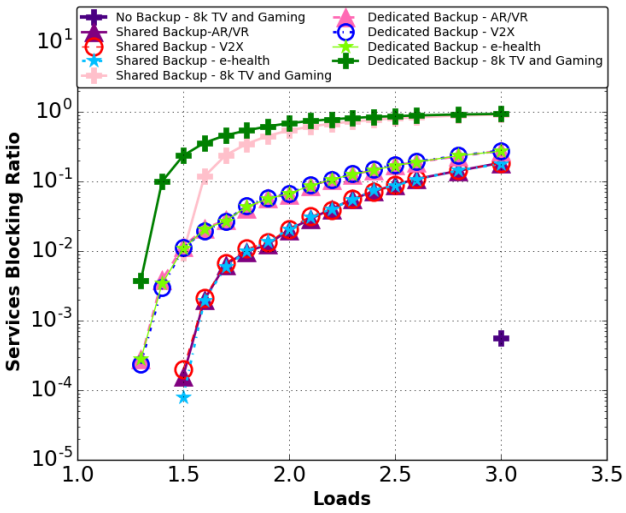

Fig. 3. Comparison in Blocking Ratio for Different Services (3-MEC)

Then, Fig. 4 compares the average end-to-end delay of the services, again for the case with 3 MEC sites. All the primary SFCs, in both protected and non-protected approaches, have the same average end-to-end delay, which is less than its value for the backup SFCs. Moreover, the average end-to-end delay is independent of the traffic load. Finally, in Fig. 5, the average end-to-end delay for the different services is presented, showing again higher values for the backup SFCs than for the primary SFCs, and higher delays for the least latency-demanding service, the 8k TV and Gaming service.

It should be remarked that the figure represents average values, but the VNF placement assignment heuristic employed ensures that all successfully established SFCs comply with the latency requirements of the service.

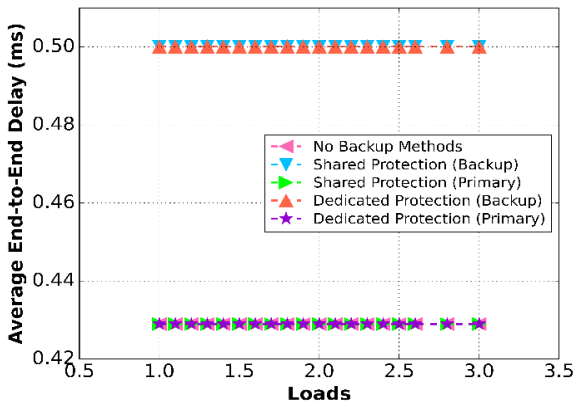

Fig. 4. Average End-to-End Delay (3-MEC)

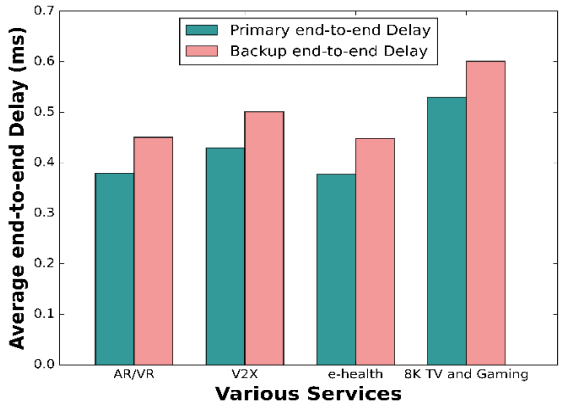

Fig. 5. Average End-to-End Delay of Various Services (3-MEC)

VI. Conclusion

In this study, we have investigated the VNF Placement problem by taking into consideration the existence of edge computing infrastructures (VNF Placement at the Edge) and dynamic control and resource allocation (CPU and RAM). After finding MEC locations with the aim of minimizing communication delays between the nodes and the MEC sites, we have implemented and compared different scenarios with various numbers of MEC sites. Then, we have proposed two different protection methods including dedicated and shared protection to enhance resiliency against a single failure in the network. The results suggest that locating computing resources in fewer MEC sites rather than distributing them in a higher number of MEC locations shows better performance in terms of blocking ratio, at least in the analyzed scenario, corresponding to a metropolitan network topology from the northern area of Italy with 51 nodes and 101.6 km network diameter. Nevertheless, in future work, we will further analyze this issue by considering network topologies with different features, as well as considering bandwidth constraints (which were omitted in the simulation part of this manuscript) and different types of SFCs.

References

[1] O. S. Al-Heety, Z. Zakaria, M. Ismail, M. M. Shakir, S. Alani, and H. Alsariera, "A Comprehensive Survey: Benefits, Services, Recent Works, Challenges, Security, and Use Cases for SDN-VANET," IEEE Access, vol. 8, pp. 91028–91047, 2020, doi: 10.1109/ACCESS.2020.2992580.

[2] M. Liyanage, P. Porambage, A. Y. Ding, and A. Kalla, "Driving forces for Multi-Access Edge Computing (MEC) IoT integration in 5G," ICT Express, vol. 7, no. 2, pp. 127–137, Jun. 2021, doi: 10.1016/J.ICTE.2021.05.007.

[3] I. Alam et al., "A Survey of Network Virtualization Techniques for Internet of Things Using SDN and NFV," ACM Computing Surveys, vol. 53, no. 2. Association for Computing Machinery, Jun. 01, 2020. doi: 10.1145/3379444.

[4] M. Wang, B. Cheng, W. Feng and J. Chen, "An Efficient Service Function Chain Placement Algorithm in a MEC-NFV Environment," 2019 IEEE Global Communications Conference (GLOBECOM), 2019, pp. 1-6, doi: 10.1109/GLOBECOM38437.2019.9013235.

[5] R. Cziva, C. Anagnostopoulos and D. P. Pezaros, "Dynamic, Latency-Optimal vNF Placement at the Network Edge," IEEE INFOCOM 2018 - IEEE Conference on Computer Communications, 2018, pp. 693-701, doi: 10.1109/INFOCOM.2018.8486021.

[6] P. K. Thiruvasagam, A. Chakraborty, and C. S. R. Murthy, "Resilient and Latency-Aware Orchestration of Network Slices Using Multi-Connectivity in MEC-Enabled 5G Networks," IEEE Transactions on Network and Service Management, vol. 18, no. 3, pp. 2502–2514, Sep. 2021, doi: 10.1109/TNSM.2021.3091053.

[7] H. D. Chantre, H. D. Chantre, and N. L. Saldanha Da Fonseca, "The location problem for the provisioning of protected slices in NFV-Based MEC infrastructure," IEEE Journal on Selected Areas in Communications, vol. 38, no. 7, pp. 1505–1514, Jul. 2020, doi: 10.1109/JSAC.2020.2986869.

[8] Y. Mu, L. Wang, and J. Zhao, "Energy-Efficient and Interference-Aware VNF Placement with Deep Reinforcement Learning," Jun. 2021. doi: 10.23919/IFIPNetworking52078.2021.9472805.

[9] J. Sun, F. Liu, H. Wang, M. Ahmed, Y. Li, and M. Liu, "Efficient VNF Placement for Poisson Arrived Traffic," IEEE Transactions on Network and Service Management, vol. 18, no. 4, pp. 4277–4293, Dec. 2021, doi: 10.1109/TNSM.2021.3102583.

[10] L. Qu, C. Assi, M. J. Khabbaz, and Y. Ye, "Reliability-Aware Service Function Chaining with Function Decomposition and Multipath Routing," IEEE Transactions on Network and Service Management, vol. 17, no. 2, pp. 835–848, Jun. 2020, doi: 10.1109/TNSM.2019.2961153.

[11] L. Qu and C. Assi, "Reliability-Aware Multi-Source Multicast Hybrid Routing in Softwarized Networks," IEEE Access, vol. 8, pp. 113331–113341, 2020, doi: 10.1109/ACCESS.2020.3003697.

[12] M. Karimzadeh-Farshbafan, V. Shah-Mansouri, and D. Niyato, "Reliability Aware Service Placement Using a Viterbi-Based Algorithm," IEEE Transactions on Network and Service Management, vol. 17, no. 1, pp. 622–636, Mar. 2020, doi: 10.1109/TNSM.2019.2959818.

[13] M. Wang, B. Cheng, and J. Chen, "Joint Availability Guarantee and Resource Optimization of Virtual Network Function Placement in Data Center Networks," IEEE Transactions on Network and Service Management, vol. 17, no. 2, pp. 821–834, Jun. 2020, doi: 10.1109/TNSM.2020.2978910.

[14] L. Qu, M. Khabbaz, and C. Assi, "Reliability-Aware Service Chaining in Carrier-Grade Softwarized Networks," IEEE Journal on Selected Areas in Communications, vol. 36, no. 3, pp. 558–573, Mar. 2018, doi: 10.1109/JSAC.2018.2815338.

[15] A. Reznik et al., "Cloud RAN and MEC: A Perfect Pairing," 2018. [Online]. Available: www.etsi.org

[16] Q. Zhang, F. Liu, and C. Zeng, "Online Adaptive Interference-Aware VNF Deployment and Migration for 5G Network Slice," IEEE/ACM Transactions on Networking, Oct. 2021, doi: 10.1109/TNET.2021.3080197.

[17] L. Contreras, "D1.2- Final 5G/Crosshaul System Design and Economic Analysis," 2017, pp. 84–87, [Online]. Available: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5b77d8419&appId=PPGMS.