Efficient Protected VNF Placement and MEC Location Selection for Dynamic Service Provisioning in 5G Networks [⋆]

Maryam Masoumi[1][0000-0002-3832-9106], Ignacio de
Miguel[1][0000-0002-1084-1159], Ramón J. Durán Barroso[1][0000-0002-1084-1159],
Soheil Hosseini[1][0000-0003-3006-5297], Hafiza Kanwal
Janjua[1][0000-0002-6655-6015], Noemí Merayo[1][0000-0002-6920-0778], Juan Carlos
Aguado[1][0000-0002-2495-0313], and Rubén M. Lorenzo[1][0000-0001-8729-3085]

Universidad de Valladolid, Valladolid, Spain
{`maryam.masoumi, ignacio.demiguel, ramon.duran, soheil.hosseini,`
`hafizakanwal.janjua, noemer, juancarlos.aguado.manzano, ruben.lorenzo`
}`@uva.es`

Abstract. Multi-Access Edge Computing (MEC) and Network Function Virtualization (NFV) have emerged as promising technologies for providing low-latency and high-bandwidth services to mobile users through Service Function Chains (SFCs) consisting of Virtual Network Functions (VNFs). However, the efficient utilization of network resources and optimal placement of MECs to host VNFs remain challenging issues in NFV-based dynamic networks. Since the proximity of MEC nodes to end-users and the availability of resources at the location impact the blocking ratio in VNF placement, in this paper we analyze the impact of the location of the set of MEC resources in the network. Thus, starting from the work previously done related to the VNF placement scheme with different protection methods to accommodate services in MEC nodes, we conduct a simulation to evaluate the impact of different MEC locations on the blocking ratio. Our results demonstrate that strategic placement of MEC locations can significantly reduce the blocking ratio.

Keywords: VNF Placement · MEC Location · Blocking Ratio.

1 Introduction

Network Function (NF) refers to a component or application that performs a specific networking task within a larger network infrastructure. Despite the fact

---

[⋆] 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).

that these components are created to enhance network security and efficiency, they come with a high price tag, limited adaptability, and pose operational challenges [9]. Network Function Virtualization (NFV) is introduced to address these challenges by virtualizing and consolidating network functions onto commodity servers, storage, and switches. This approach replaces traditional dedicated network hardware with software-based virtualized network functions (VNFs) running on standard IT infrastructure [10] and [14]. Multi-Access Edge Computing (MEC) is another technology that is predicted to have a significant influence on 5G networks by fulfilling the demand for ultra-low latency in specific applications and services and minimizing transport network congestion [11]. In order to have more services, Service Function Chaining (SFC) can be employed. This involves processing a sequence of VNFs in a specific order to provide a particular service. As a result, a crucial concern is determining the appropriate placement of the VNFs in the predefined sequence to fulfill the service requirements [9].

VNF placement refers to the process of selecting the optimal locations to deploy VNFs within a network infrastructure. To enhance customers' Quality of Experience (QoE), VNFs need to be dynamically relocated between network nodes. One of the challenges concerns the placement of VNFs in edge nodes consisting of selecting the most suitable edge (MEC) nodes to deploy VNFs- SFCs to meet service requests. Reliability is also an important criterion for selecting a service provider, which prompts them to seek NFV deployment algorithms that maintain reliability above specific standards. Sometimes, networks are faced with increased computing loads, hardware failures, or malicious attacks. To ensure a desired level of reliability, protection schemes are commonly employed, with dedicated or shared backup being the two types of protection methods [3]. This is necessary to ensure the reliability of each request. Therefore, in all the protection strategies, a primary SFC is set up to provide the related service in normal conditions and is protected by a backup SFC that has its VNFs located in distinct physical positions [1].

Network operators strive to boost the service acceptance ratio by accommodating as many SFC requests as possible and minimizing the blocking probability. Previous research has investigated different techniques for protecting SFCs for static traffic. However, our research focuses on the challenge of ensuring reliable SFC mapping in dynamic settings. In our previous study [7], we conducted an investigation into the VNF Placement problem, specifically focusing on VNF Placement at the Edge and dynamic control and resource allocation for CPU and RAM. Our approach involved identifying MEC locations that would minimize communication delays between nodes and MEC sites. We then implemented and compared different scenarios with varying numbers of MEC sites. For this aim, we compared 3-MEC, 5-MEC and 7-MEC scenarios.

The results indicated that concentrating computing resources in a smaller number of MEC sites rather than dispersing them across multiple locations results in decreased probabilities of blocking. To enhance network resiliency against a single failure, we proposed two protection methods: dedicated and shared protection. A dedicated approach can be effective in providing high levels of protection for VNFs, but can also be costly due to the need for additional resources while the shared approach can be more cost-effective than dedicated protection, but may not provide the same level of protection as dedicated resources. Since the location of MECs is of great importance in service blocking, in this current paper we aim at considering the impact of the locations of the MEC on the blocking ratio. Moreover, we focus on the significance of MEC node placement in relation to service blocking. The primary objective of this paper is to analyze and understand how the locations of MECs influence the blocking ratio. In our previous study, we made certain assumptions about the placement of 5 MEC nodes.

In this current research, we are taking a further step by exploring the implications of keeping the same number of MEC nodes but locating them in different positions. This is a contribution compared to our previous work. Additionally, we are also considering a more realistic assumption regarding service chains. Essentially, this paper has a two-fold objective. Firstly, it aligns with the Io-Talentum project [5], specifically addressing one of the research topics focused on integrating NFV in a distributed computing resources in MEC by means of SDN-based control mechanism and determining the optimal locations for MEC nodes. Secondly, it aims to provide updates and enhancements to our previous research, which are being presented in this paper.

The main contributions of this paper can be summarized as follows: Several possibilities for the 5-MEC locations are taken into account and then to assess and compare the blocking rate of different locations for 5MEC configuration, a real-world network topology is employed.

The remainder of this paper is as follows: In Section 2, we provide a background on VNF placement and MEC location problems. Section 3 discusses the problem addressed in our previous paper. In Section 4 an overview of our previous work is given followed by an explanation of VNF placement algorithm in 4.1. We present the main contribution of this work by providing the different assumptions for MEC locations in 4.2. Section 5 is our numerical results obtained by simulation. Finally, Section 6 concludes the paper and explains the potential direction of future research.

2 Background Works

MEC utilizes the characteristics of NFV, SDN, and NS (Network Slicing) to provide delay-sensitive services dynamically as requested. As a result, MEC has become a critical component for future 5G networks and beyond [4]. Due to its importance as a 5G technology enabler alongside NFV and SDN, there has been interest from both industry and academia in addressing issues related to MEC technology. Dash et al in [2] propose an efficient clustering-based approach to locate MEC servers and associate them with RANs. The approach uses a greedy-based algorithm called MEC Placement and Association (MEC PA) to minimize the number of required MEC servers while satisfying the propagation delay constraint of all RANs. Under this approach, each MEC server acts as a cluster head, and each RAN is treated as a member of the cluster that meets the propagation delay constraint. This placement strategy not only meets the propagation delay requirement but also balances the workload.

In [12], an algorithm is proposed for selecting a few MEC locations from a set of base stations in order to establish MEC infrastructure and address the user requirements of delay-sensitives services. In this paper, a set of BS nodes are selected as potential MEC locations based on high Closeness Centrality (CC) values. So, the high-CC value BSs are selected as potential MEC locations. However, the aspects of resiliency and service continuity have not been adequately explored in resource-limited MEC-enabled 5G networks to ensure that latency requirements and service continuity are maintained even in the event of failures. This section will examine previous studies that have focused on the deployment of VNFs in MEC-enabled networks with the aim of enhancing system performance and network efficiency. [6] addresses the problem of VNF placement and resource allocation in MEC networks enabled by NFV/SDN, with the objective of minimizing both the overall placement and resource cost. However, it does not take into account the latency requirements of users, failures of network components, or resiliency aspects.

The issue of VNF placement in a MEC-NFV environment is addressed in [13], with a focus on minimizing costs by taking into account both latency and resource availability. However, the study does not address backup methods or resiliency aspects, which are necessary to ensure service continuity in the case of failures. Furthermore, the study only considers a single VNF placement for slice creation. [8] is the most closely related prior studies to previous and current work. The issue of resilient VNF placement in 5G networks is addressed, with the objective of minimizing the impact caused by network component failures. This paper aims to explore how the blocking ratio is impacted by varying MEC node locations in a network that will have five MEC nodes.

3 Problem Description

In this paper, an analysis of a 5G network is presented, which employs a distributed MEC infrastructure with virtualization at each site to handle service requests. However, this network faces limitations in computational and storage capacities as well as bandwidth. VNF failures can lead to disruptions in VNF-dependent services, and multiple VNF operations are vulnerable to interruptions, causing various services to fail. Therefore, the necessity arises to identify a solution for mapping dynamic service requests onto MECs while considering protection schemes. To reduce communication delays between nodes and their nearest MEC site, strategically chosen MEC sites are utilized. The dynamic resource allocation problem is formulated in this paper without protection methods, and then dedicated VNF backup and shared VNF backup protection methods are applied to ensure the resilience, security, and availability of established services. The objective of this paper is to minimize the blocking ratio of a given service request while focusing on protection against single failures.

4 A Summary of our Proposed VNF Placement Method

The network is represented as an undirected graph with base stations and physical links. MEC locations are set up by selecting a small subset of base stations and assigning a finite number of servers at each MEC location to accommodate services for user requests. Each server is equipped with computing capacity with a certain number of CPU cores and RAM. Physical links have a specific bandwidth capacity and introduce latency consisting of propagation and processing delay. Three scenarios are considered: unprotected operation, dedicated SFC protection, and shared VNF protection.

4.1 VNF Placement Algorithm

In this part, we provide a brief explanation of the method for VNF placement that we previously introduced in our work [7]. We assume dynamically arriving of SFC requests. In order to set up a service function chain, we must allocate all of its VNFs on MEC nodes and direct the traffic between them. Each SFC needs a certain amount of bandwidth and has a maximum latency limit. Additionally, the VNFs associated with each SFC require a particular amount of computational resources, which is measured in terms of CPU core usage and RAM. Each SFC is modeled based on some specifications including the source base station from which the user requests service, the number of VNF, the required bandwidth, the maximum allowed end-to-end latency, the time of arrival, and the lifetime of the SFC request.

Our algorithm includes two phases:

1. Phase I: The placement of the whole SFC in a single MEC site is considered for service provisioning. In the first scenario (unprotected operation), a primary SFC is established for each request (with no backup SFC), and the MEC node closest to the user requesting the service is selected to host the SFC. The service is provided by selecting the first server within the MEC site that has sufficient available CPU and RAM capacity, and the required CPU and RAM are reserved during the service duration. If no server is found with available resources in this MEC, the other nearest MECs are considered so as to find a server with enough capacity. The request will be blocked if no server can be found with enough resources (or if bandwidth or delay constraints are not met).
   In the second and third scenarios (when dedicated and shared protection are applied), the primary MEC is selected as the first nearest MEC to the user, and the backup MEC is considered to be the second closest one. If there are insufficient resources in the preferred MEC site, the algorithm considers the other nearest MEC. In dedicated protection, one backup SFC protects one primary SFC in a single MEC site using the first-fit policy. In contrast, shared protection allows one backup VNF to protect multiple primary VNFs located in different MEC sites. If a suitable backup VNF cannot be found for shared protection, a new instance of the VNF is created. When the lifetime

of a request expires, all resources are released except for shared protection, where backup VNF resources are released only when the last SFC using them is released.

2. Phase II: During this phase, the total time it takes for both the primary and backup routes to transmit data is computed, considering both the propagation and processing times. The end-to-end delay for the primary route and the backup route is calculated based on the distance between the source base station and the primary MEC and backup MEC respectively. Moreover, each VNF introduces an additional delay which should be taken into account in each SFC.
   After meeting the latency criteria, the SFC is provisioned and the computational resources are then allocated. The purpose of this mechanism is to ensure that the backup VNF is available for use in case of a failure of any primary VNF in any of the SFC instances that use it.

4.2 Different Assumptions for MEC Locations

In our previous study, we proposed concentrating computing resources in a smaller number of MEC sites, rather than spreading them across a larger number of locations, as it demonstrated better performance. Building upon that research, this paper investigates the influence of different MEC locations on the blocking ratio while utilizing a shared protection mechanism. We present a scenario with 5 MECs and introduce three distinct sets of MEC node locations to evaluate their impact on service availability.

5 Simulation Results

To obtain numerical results we develop a discrete event-driven simulator in Python. We consider a reference metro regional network of Italy similar with 51 nodes, out of which 5 nodes are MEC, each one has 315 servers which are equipped with 512 CPU cores and RAM. In this network, there are 61 bidirectional links enabling communication in both directions. The longest among these links measures 101.46 kilometers in length. The SFCs considered in this paper, their VNFs as well as their latency and bandwidth requirements are depicted in Table 1 while computational requirements of VNFs are shown in Table 2. We consider that incoming SFC requests are selected with the possibility of 25% among those. SFC requests are generated at MEC nodes dynamically.

The SFC requests are generated with an inter-arrival rate λ that follows a Poisson distribution and a lifetime according to negative-exponential distribution, with mean lifetime $\mu = 60$ seconds. All the results are obtained with a confidence level of 95% with at most 5% confidence interval on blocking probability. The load is determined based on the average lifetime of each request, the average inter-arrival time, and the number of nodes in the topology as follows:

$$load = \frac{\lambda * \mu}{N(N-1)} \tag{1}$$

Table 1. Services Function Chain Requirements [15].

Services	Service-Chain-VNFs	Banddwidth(Mbps)	Delay(ms)
Cloud Gaming	4	NAT, FW, VOC, WO, IDPS	80
Augmented Reality	200	NAT, FW, TM, VOC, IDPS	2
MIoT	100	NAT, FW, IDPS	5
Video Streaming	4	NAT, FW, TM, VOC, IDPS	100

Table 2. CPU Core and RAM Usage For Various VNFs [15].

VNF	CPU	RAM
NAT	2	4
FW	1	4
IDPS	1	2
TM	1	2
VOC	2	4
WO	1	2

The unit of load is typically expressed as Erlang. In order to measure the latency, the nodes are linked together using optical fibers, which leads to a propagational delay. Moreover, each VNF add a value of 0.050 milliseconds to the total delay.

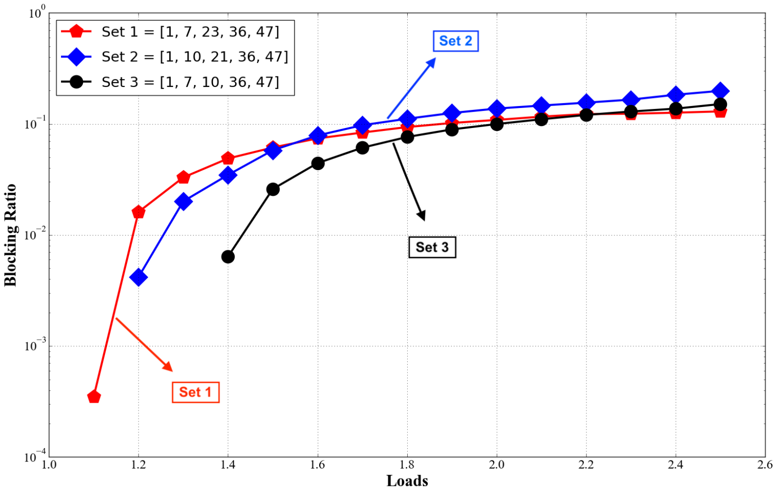

Fig. 1. Comparison in Blocking Ratio between Different Location Sets with Shared Protection

For the purpose of this study, we make the simplifying assumption that the constraints are only due to computing resources and delay, and we do not consider any bandwidth requirements. We consider a scenario with 5 MECs and introduce three sets of MEC node locations, denoted as location set 1, location set 2, and location set 3. Each set consists of 5 MEC locations. In location set 1, the MEC nodes are situated at nodes 1, 7, 23, 36, and 47. Location set 2 places the MECs at nodes 1, 10, 21, 36, and 47, while location set 3 consists of nodes 1, 7, 10, 36, and 47. The performance, specifically the blocking probability, varies depending on the MEC location.

The graph in Fig. 1 illustrates the blocking ratio for these proposed node sets. Location set 3, with MEC nodes placed at nodes 1, 7, 10, 36, and 47, exhibits superior performance in terms of blocking probability. On the other hand, configuring the network according to the location set 1 results in a higher blocking ratio, particularly for lighter loads. This discrepancy can be attributed to the distribution of the loads handled by each server, as MEC nodes in certain locations may serve a larger number of nodes. Location set 3 appears to have a more homogeneous distribution of workloads. As we transition from set 1 to set 2 and finally to set 3, the blocking rate decreases, indicating a more balanced workload in set 3, where MECs are positioned at nodes 1, 7, 10, 36, and 47.

6 Conclusion

To conclude, this paper provides an overview of the approach proposed in our previous work for dynamically placing VNFs in a MEC-enabled environment. The approach aims to minimize communication delays and introduces two protection methods to enhance network resiliency against single failures. We concluded 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. In our previous work, we focused on the blocking rate with 5 MEC nodes. However, in this present paper, we investigate the placement of 5 MECs using realistic VNF specifications. We recognize the significance of the MEC location and its relationship with load distribution. Our next step involves analyzing the underlying factors contributing to different blocking rates, leveraging the static context, and ultimately identifying optimal methods for locating MECs. Furthermore, we are going to analyze the workload and traffic patterns of the network to identify the optimal MEC locations for specific use cases and traffic profiles.

References

1. Askari, L., Tamizi, M., Ayoub, O., Tornatore, M.: Protection strategies for dynamic vnf placement and service chaining. In: 2021 International Conference on Computer Communications and Networks (ICCCN). pp. 1–9. IEEE (2021)
2. Dash, S., Khan, A.U., Swain, S.K., Kar, B.: Clustering based efficient mec server placement and association in 5g networks. In: 2021 19th OITS International Conference on Information Technology (OCIT). pp. 167–172. IEEE (2021)

3. Ghazizadeh, A., Akbari, B., Tajiki, M.M.: Joint reliability-aware and cost efficient path allocationfig and vnf placement using sharing scheme. Journal of Network and Systems Management **30**, 1–28 (2022)
4. Giust, F., Verin, G., Antevski, K., Chou, J., Fang, Y., Featherstone, W., Fontes, F., Frydman, D., Li, A., Manzalini, A., et al.: Mec deployments in 4g and evolution towards 5g. ETSI White paper **24**(2018), 1–24 (2018)
5. IoTalentum: Sdn for network virtualization in iotalentum infrastructure. iotalentum.eu/esrs (Last accessed 8 May 2023)
6. Kiran, N., Liu, X., Wang, S., Yin, C.: Vnf placement and resource allocation in sdn/nfv-enabled mec networks. In: 2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW). pp. 1–6. IEEE (2020)
7. Masoumi, M., De Miguel, I., Barroso, R.J.D., Ruiz, L., Brasca, F., Rizzi, G., Merayo, N., Aguado, J.C., Fernández, P., Lorenzo, R.M., et al.: Dynamic online vnf placement with different protection schemes in a mec environment. In: 2022 32nd International Telecommunication Networks and Applications Conference (ITNAC). pp. 1–6. IEEE (2022)
8. Mohan, P.M., Gurusamy, M.: Resilient vnf placement for service chain embedding in diversified 5g network slices. In: 2019 IEEE Global Communications Conference (GLOBECOM). pp. 1–6. IEEE (2019)
9. Ren, H., Xu, Z., Liang, W., Xia, Q., Zhou, P., Rana, O.F., Galis, A., Wu, G.: Efficient algorithms for delay-aware nfv-enabled multicasting in mobile edge clouds with resource sharing. IEEE Transactions on Parallel and Distributed Systems **31**(9), 2050–2066 (2020)
10. Scazzariello, M., Ariemma, L., Di Battista, G., Patrignani, M.: Megalos: A scalable architecture for the virtualization of network scenarios. In: NOMS 2020-2020 IEEE/IFIP Network Operations and Management Symposium. pp. 1–7. IEEE (2020)
11. Tabatabaee Malazi, H., Chaudhry, S.R., Kazmi, A., Palade, A., Cabrera, C., White, G., Clarke, S.: Dynamic service placement in multi-access edge computing: A systematic literature review. IEEE Access **10**, 32639–32688 (2022). https://doi.org/10.1109/ACCESS.2022.3160738
12. Thiruvasagam, P.K., Chakraborty, A., Murthy, C.S.R.: Latency-aware and survivable mapping of vnfs in 5g network edge cloud. In: 2021 17th International Conference on the Design of Reliable Communication Networks (DRCN). pp. 1–8. IEEE (2021)
13. Yala, L., Frangoudis, P.A., Ksentini, A.: Latency and availability driven vnf placement in a mec-nfv environment. In: 2018 IEEE Global Communications Conference (GLOBECOM). pp. 1–7. IEEE (2018)
14. Yang, S., Li, F., Trajanovski, S., Yahyapour, R., Fu, X.: Recent advances of resource allocation in network function virtualization. IEEE Transactions on Parallel and Distributed Systems **32**(2), 295–314 (2020)
15. Zhang, Q., Liu, F., Zeng, C.: Online adaptive interference-aware vnf deployment and migration for 5g network slice. IEEE/ACM Transactions on Networking **29**(5), 2115–2128 (2021)