Original scientific paper

Impact of NFV Resources Distribution on Inter-Subnetwork Latency

Gjorgji Ilievski and Pero Latkoski

***Abstract* — 5G networks are already being implemented around the globe. One of the most important enablers of their penetration are the Software Defined Networking (SDN) technologies and the Network Functions Virtualization (NFV) architecture, which allow the needed flexibility of the network and the composing elements. In such circumstances, the Internet-of-Things (IoT), which has been long awaited, is becoming feasible and economically reasonable. This setup has its challenges, especially due to the network expansion toward the edge, where the number of networking elements and service consumers is rapidly rising. The compute resources and the storage have to be brought in the network proximity of the access network, so that the latency of the service is kept under 1ms, which is one of the base 5G requirements. For our research, we have made an experimental setup of a distributed NFV architecture on a multiple geo-location, with a main objective to review the network latency caused by the architectural distribution of the services that are built in it. The results can be used by researchers and network architects to build reliable and cost-effective distributed services with the lowest possible latency, as well as to plan possible disaster recovery scenarios when some physical location is unavailable.**

***Keywords* — NFV, SDN, Experimental Model, Distributed NFV, MANO.**

I. Introduction

TELECOMMUNICATIONS industry is changing rapidly. The virtualization technologies have evolved from classical virtual machines (VMs) and virtual appliances (VAs) to containerized applications and virtual functions that enable extraordinary flexibility, especially in the virtualization of the networking elements. Software Defined Networking (SDN) is accompanied by the evolution of the Network Functions Virtualization (NFV) architecture and they both make the 5G revolution possible. Capital investments and operational expenses can be reduced, allowing telecommunications and IT companies to create custom services which are complementing the public clouds, but are not dependent on them. These services can be brought close to the end consumers from both geographical and networking point of view, thus allowing the best end-user experience, enhanced security, scalability and agility [1].

The NFV architecture [2] allows complete separation of the network functions from the proprietary purpose-built hardware appliances, using software based virtual network functions that run on generic hardware. The architecture is flexible and agile by design, but one of the less explored directions, both in the industry and in the science communities, is the possibility of building distributed services on multiple physical locations, by leveraging the NFV architecture. The packet latency caused by the distribution of the functions and their placement [3] is one of the main concerns when designing the services. There are many research efforts that are focused on the Distributed NFV (D-NFV) from an analytical point of view, modeling the systems, as well as the services built on them, to predict their behavior [4], [5]. Other authors are developing software components, applications and frameworks in order to improve various aspects of the NFV, such as the Virtual Network Functions (VNF) placement or to improve the management and operations of the system, so that the services running on it are optimized [6], [7], [8]. Another research stream is focused on individual features of the distributed SDN and NFV environments [9], [10].

Our main focus of research is the distribution of the SDN components as well as the NFV architecture on multiple geographical locations, in order to predict the network packet latency caused by it. We have made an experimental environment using two datacenters and a service that spans on both of them. We are making a comparison of the network packet latencies in the distributed service, relative to a service that is built on a single location, having in mind the proximity of the service consumer (the end user) to the locations. The reliability of the service should be of the highest possible level. The experiments that we are making are also valid for a D-NFV that can be used in a disaster recovery scenario, when the primary location fails, in order to predict the latency in the network that will be caused due to the service location shift. This research should enable network architects to build reliable and cost-effective services with predictive behavior.

The experimental environment that we are using is based on a set of tools that is well known for SDN and NFV simulations. Mininet [11] is used for modeling the virtual network, while D-ITG [12] is used as a network

Paper received April 29, 2022; revised December 18, 2022; accepted December 28, 2023. Date of publication August 08, 2023. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Grozdan Petrović.

This paper is revised and expanded version of the paper presented at the 29th Telecommunications Forum TELFOR 2021 [32].

Gjorgji Ilievski is with the IT Security Department of Makedonski Telekom AD Skopje, RN Macedonia (e-mail: ilievskigjorgji@yahoo.com).

Pero Latkoski is with Telecommunications Institute of the Faculty of Electrical Engineering and Information Technologies, Ss. Cyril & Methodius University Skopje, RN Macedonia (e-mail: pero@feit.ukim.edu.mk).

traffic generator. There are hypervisors based on Ubuntu 18.04 and VirtualBOX [13] on two geographically separated locations, which are over 160km apart. The SDN elements are generated and chained in Mininet, while Open vSwitch [14] is used as a virtual switch on the hypervisors. The entire network is relying on OpenFlow (OF) protocol.

The conclusions made can be used in the preparation phase for introducing 5G radio access network (RAN), as well as for the services that will run in such an environment. The 5G ultra reliable low latency communications (URLLC) [15] is one of the key pillars of 5G, which is supposed to support dense grids of endpoints, while end-to-end latency must be as low as 5ms. The experimental results that we use to identify the latency factors in a distributed NFV environment are the main contribution of our paper. The results can have a variety of use cases [16], for services in almost any area, such as transportation [17], [18], healthcare [19], [20], manufacturing [21], [22], energy transmission [23].

This paper is organized in the following order: in the next chapter we discuss the related work, then we explain in more detail the experimental testbed used. Afterwards we analyze the results. The conclusions are at the end.

II. Related Work

The concept of building services in an NFV based architecture is a hot topic in the research community. There are various aspects that are being discussed using different kinds of methods. Some of the authors are focused on the theoretical part, using mathematical models to predict the behavior of such services [24], [25]. Queuing theory and Markovian chains [26], [27] are one of the most used techniques for such an analysis.

Other authors are exploring the possibilities for improvements of the services built upon SDN and NFV using software applications and frameworks [28], [29].

There is research on the subject that is experimental [30], [31]. Similarly, we have made a unique experimental environment in order to measure the packets sojourn time in a D-NFV based services.

E. Fountoulakis et al. [24] use NFV and Multi-access Edge Computing (MEC) to develop a queueing model for the performance analysis of a system that is consisted of both processing and transmission flows and propose a method in order to derive analytical expressions of the performance metrics of interest, such as end-to-end delay, system throughput, task drop rate. They use different scenarios for simulation and analytical calculations in order to provide insights for the decision making on traffic flow control and its impact on critical performance metrics. Our work is purely experimental and is focused on a single metric: the network latency, with similar motivation, to ease the decision-making process when a service chain is developed.

In [25], the main focus is on the network slicing as one of the key features of 5G mobile networks to cope with the diverging network requirements. Network slices are isolated, virtualized, end-to-end networks optimized for specific use cases that share a common physical network infrastructure. A mathematical model for solving the offline Network Slice Embedding Problem formalized as a standardized Mixed Integer Linear Program is presented, where a latency sensitive objective function guarantees the optimal network utilization as well as minimum latency in the network slice communication. We are also considering the latency, and our case is feasible in a 5G environment, but from an experimental point of view.

The authors of [26] also deal with network slicing that has emerged as a promising paradigm of the 5G mobile networks and which is based on NFV and SDN technologies. They propose a network slicing architecture for 5G mobile networks involving a cloud radio access network (C-RAN), mobile edge computing (MEC), and a cloud data center. They model the proposed network slicing system based on queueing theory, which is used to derive the main performance metrics of the network. Similarly, in [27] a softwarized IMS architecture built upon a NFV paradigm is modeled using M/G/c system in order to characterize the behavior of softIMS in terms of failure/repair events, and to derive a set of optimal configurations satisfying a given availability requirement. This research, unlike ours, is theoretical, although some of the results regarding the network latency are comparable to our experimental results.

NFV architecture from a point of software and system modeling is overviewed by M. Sadaf et al. [28]. They use MAPLE, an integrated process modelling and enactment environment with traceability information generation and analysis support built into Eclipse Papyrus to review the automation of network service management in NFV and to define an automated process for the design, deployment, and management of network services. They use UML activity diagrams for process modelling.

In [29], the authors make a review of NFV and service function chain (SFC) implementation frameworks and present a taxonomy of the current proposals. They have three major categories based on the primary objectives of each of the surveyed frameworks: resource allocation and service orchestration, performance tuning, and resilience and fault recovery. Their accent is also on the NFV slicing which is closely related to the NFV distribution that we are investigating in our research.

J. Vergara-Reyes et al. [30] have created an experimental testbed based on an NFV architecture in which they generate network traffic and analyze it using supervised machine learning algorithms in order to grade the efficiency of the algorithms in traffic classification. The results are important for network performance, monitoring and security in the NFV and SDN based environment where the traffic itself goes mainly in the virtual layer in east-west direction, rarely going through the physical links in north-south direction. Our work is similar to their in regard to the experimental setup of the SDN and NFV architecture, but our goals and research are in a different direction.

The authors of [31] introduce a novel testbed, called 5GIIK, that provides implementation, management, and orchestration of network slices across all network domains and different access technologies. They identify design criteria that are a superset of the features present in other testbeds and determine appropriate open-source tools for implementing them. Their testbed is about slice provision dynamicity, real-time monitoring of VMs and VNF-onboarding to different virtual infrastructure managers (VIM). We are also introducing an NFV testbed, with a main goal to measure the network packet latency, and to make conclusions on the factors that influence it.

III. Experimental Setup

We are creating an environment with a multi-site distributed data plane NFV MANO. Such an environment is practical when it is necessary that the network elements that are forming the user service are closer to the consumer. Multimedia services are one such example. This scenario is also plausible for multiple data centers in one or more geographic regions. Our experiment is limited to a two-site infrastructure, but the results and the conclusions can be widened to a scenario with more sites [32].

The experiment will show the impact of the placement of the VNF elements forming the service chain on the network latency. The service tends to have the lowest possible network latency, but this is bounded by the possibilities of the NFV infrastructure to dynamically create the VNFs on the location that is possible.

Our testbed is based on Ubuntu 18.04 hypervisors and Oracle VirtualBox that are installed on two geographically separated locations. The distance between the locations is 160km. The link between the locations has a bandwidth of 10Gbps. The physical switches in the datacenters are working in a traffic shaping mode. Each hypervisor has a network interface with 1Gbps bandwidth. There is not any kind of Quality of Service (QoS) implemented. As this is a NVF environment, we assume that the system is based on a single network slice, with two subnets, one in each site.

Open vSwich (OVS), which is based on open flow, is used on hypervisors. The network is simulated with Mininet, with multiple hosts, switches and links. The IP addressing in the simulated network is private, and GRE tunneling is used for outside communication. Ryu is used as an SDN controller and together with the Mininet network generator simulates the NFV Management and Orchestration (MANO). A single Ryu controller is used, meaning that it is responsible for the flow tables on both vSwitches. When a flow is going in a single location, it passes only inside the virtual infrastructure, the traffic is completely east-west based. When the service chain is distributed, the traffic leaves the virtual infrastructure, and through the physical switches and the link goes to the other location. The routes in the NFV are completely OF based and the controller always chooses the best possible network path.

To simulate network traffic, we use the distributed Internet Traffic Generator (D-ITG). We generate both TCP and UDP based traffic. D-ITG produces traffic at packet level, replicating appropriate stochastic processes for both IDT (Inter Departure Time) and PS (Packet Size) random variables [12].

In D-ITG we use the multi-flow option to create multiple simultaneous data flows with standard payload information (as it is defined by D-ITG). We set the inter-departure time in the flows to have normal distribution. We conducted 30 iterations for each experiment by using 4, 8, 12 and 16 dataflows. The iterations last from 30 seconds to 10 minutes. The time intervals follow the Poisson distribution.

We are sniffing the network traffic inside 2 virtual machines, each installed on a separate location, using Wireshark and tshark [33]. The number of packets that were captured varied from 7612 to 4029312 packets, with an average number of 1862225 in an iteration. We are calculating the mean packet delivery time of the packets within an iteration and within an experiment.

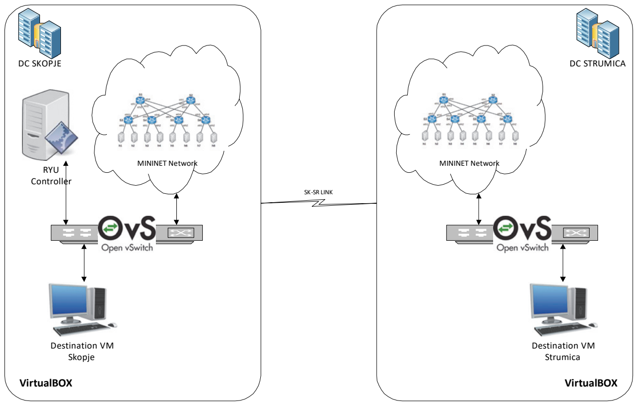

Fig. 1. Experimental environment for D-NFV based system.

IV. Experimental Results

Our first experiment was with a network path simulating a service that spans on a single location. The generated traffic is completely within the primary location, with destination set to the local VM, simulating a consumer in a close proximity to the provider. As expected, the number of concurrent dataflows has an impact on the mean packet delivery time, but the rise in the packet delivery time is small.

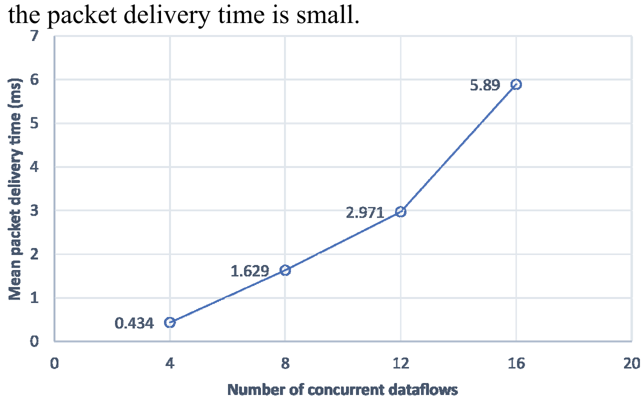

Fig. 2. Mean packet delivery time in a single location service.

Table 1 shows the results of the mean packet delivery time taken from 30 measurement iterations for each experiment, using 4, 8, 12 and 16 concurrent dataflows. Fig. 2 shows the results graphically.

TABLE 1: MEAN PACKET DELIVERY TIME IN A SINGLE LOCATION SERVICE

Flows	*4*	*8*	*12*	*16*
Mean packet delivery time(ms)	0.434	1.629	2.971	5.86

When the service spans across two locations, with some of the VNFs on the primary, and some on the secondary location, the network packets can flow between the two locations multiple times. In this scenario we were measuring the mean packet delivery time by simulating traffic that flows 1, 3, 5 and 7 times through the link. We have done the experiment 4 times, again by using 4, 8, 12 and 16 parallel dataflows. The source of the network traffic is always in the primary location, while the destination (e.g., the consumer) is always in the secondary location. Mean packet delivery time was calculated based on 30 iterations for every situation.

TABLE 2: MEAN PACKET DELIVERY TIME IN A TWO LOCATIONS SERVICE.

Passes through the link	*Concurrent flows*			
	4	*8*	*12*	*16*
1	5.121ms	5.67ms	6.195ms	6.634ms
3	5.72ms	11.21 ms	31.461ms	48.389ms
5	8.432ms	31.961ms	61.31ms	89.455ms
7	19.651ms	58.12ms	112.942ms	193.82ms

As the results from Table 2 show, the impact of the link crossing has a significant impact on the mean packet delivery time, and the impact is bigger for a larger number of concurrent flows, which is expected. Visually this can be seen in Fig. 3.

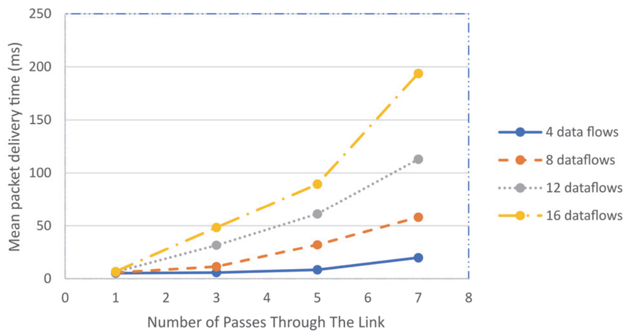

Fig. 3. Mean packet delivery time in a two locations service.

We compare the scenarios, when the service chain is on a single location, and when the service chain is on two locations, passing multiple times through the link. This is shown in Fig. 4. We can conclude that the mean packet delivery time rises in both cases, but in the two locations case, the network latency is much bigger, going almost up to 200ms, while in the single location scenario the latency is significantly smaller. This delay is caused mainly due to the traffic shaping setup of the switches in the link, but this setup is common when both TSP and UDP traffic are going through the physical switches and the links. While traffic policing is about dropping or reclassifying packets, traffic shaping tries to make traffic conform to a certain rate by delaying the packets in a buffer and sending them out as "space" becomes available. This proves that the proximity of the VNFs that are part of the service and the end consumer has a significant impact on the network latency, thus confirming that network architects must carefully consider the placement of the networking elements.

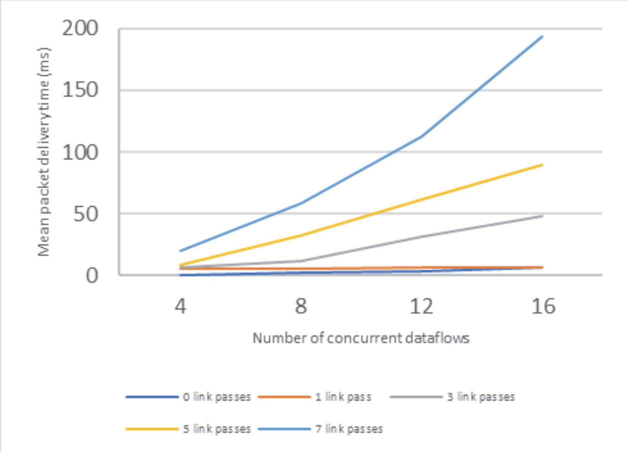

Fig. 4 Single location vs. multiple passes through the link

As a third setup, we are investigating a situation in which the packet source is always in the primary location, while the destination can be in the primary or in the secondary location. Our iterations are based on a specific percentage of the traffic going to the secondary locations, measuring the mean packet delivery time when 10%, 20% and up to 100% of the flows are going to the secondary location. We were using 40 concurrent dataflows. When all the traffic was going through the link, the network interfaces of the hypervisors were utilized up to 97%. The results can be seen in Table 3 and Fig. 5.

TABLE 3: MEAN PACKET DELIVERY WHEN PERCENTAGE OF THE TRAFFIC GOES THROUGH THE LINK

% of flows through the link	*10%*	*20%*	*30%*	*40%*	*50%*	*60%*	*70%*	*80%*	*90%*	*100%*
Mean Packet delivery time (ms)	5.86	9.82	16.13	22.07	28.91	35.87	50.05	61.81	100.12	191.58

We can use the results of the previous experiment for a disaster recovery scenario, where we will assume that the primary location fails and the secondary location is the only location providing the VNFs for the service. The consumer (the destination) is in close proximity to the primary location. We assume that the link is up, which can be treated as a situation in which the link is not impacted from the disaster, or a secondary link exists and is in function. We will not take into consideration the time needed for the link re-establishment as this can vary from a particular situation and it can be treated as a constant that will introduce an additional network latency. It is a scenario in which 100% of the network flows go from the secondary, back to the primary location.

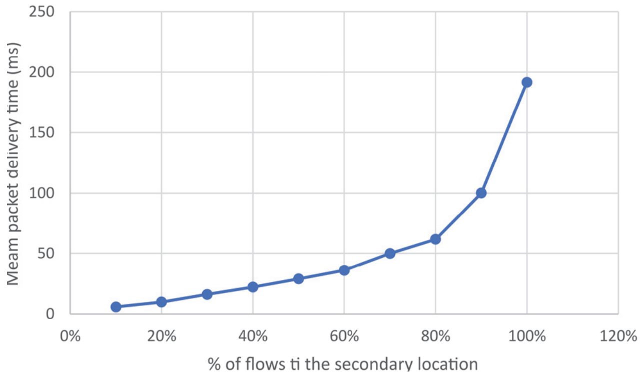

Fig. 5. Mean packet delivery time relative to % of packets through the link.

From the results we can see that although the network packets number rises linearly, the network latency rises in an exponential manner, as it can be seen in the figures. The results can be important to network architects that can benefit from the possibility of the service distribution to multiple geographically separated locations, having in mind that the proximity of the networking elements to the end service consumer has a great impact on the network performance. But if the traffic of a single service is routed to multiple locations, the network latency is highly impacted. With this, an optimal scenario from a point of network latency performance, but also from a point of price of the infrastructure can be designed.

V. Conclusion

We have made an experimental testbed to simulate a system based on an SDN and a Distributed NFV in order to evaluate the network latency that will occur under different circumstances. We used two geographically distant locations connected with a carrier grade link with a bandwidth of 10Gbps. Physical switches that connect the locations use traffic shaping policy and the link has no QoS set up. We were calculating the mean packet delivery time by performing 30 iterations of every experiment.

The main motivation for our research is to draw conclusions for the impact that a distribution of a service built on an NFV architecture has on the network packet latency. We compare the results with a service that is built on a centralized NFV environment, having in mind the stability of the system in case of a network congestion caused by filled network links.

The objective of our research is to guide network architects when designing services in an NFV architecture-based environment, so that the most optimized and cost-effective design is chosen, regarding the network packet latency.

As expected, the experiments proved the conclusion that when the service spans across multiple locations, the link between them has a significant impact on the network packet latency. The number of passes of the packets through the link makes an exponential rise in the overall packet delivery time. By introducing a distribution in the NFV infrastructure, the service chain can be kept in close network proximity if the service consumer, thus providing the best possible service quality. The latency that is introduced by the links is significant and the distribution of the NFV makes sense in scenarios where the network packet delays must be minimal. On the other hand, introducing multiple locations can have a major influence on the complexity of the network topology. The approach that we use, by using SDN and NFV, minimizes this problem.

The results that we analyze are in-line with other researchers, that are making analytical and experimental investigations on the network latency within NVF, using the results for different purposes, such as placement of the VNFs, calculating the optimal data-paths, scaling the infrastructure [34], [35], [36].

The conclusions that we made can be used by researchers and network architects to plan the infrastructure and to predict the latency of the services. Bringing the networking elements closer to the consumer is essential if the latency is of great importance. These scenarios are important when using 5G network for the services. Dividing the VNFs in multiple locations can have financial sense if the latency caused by it is carefully calculated and does not impact the service.

References

[1] J. H. Cox, J. Chung, "Advancing Software-Defined Networks: A Survey," *IEEE Access*, vol 5, pp 25487 - 25526, 2017.

[2] ETSI - Standards for NFV - Network Functions Virtualisation, NFV Solutions, https://www.etsi.org/technologies/nfv, cited 04.04.2022

[3] M. J. F. Alenazi, A. Almutairi, S. Almowuena, A. Wadood and E. K. Çetinkaya, "NFV Provisioning in Large-Scale Distributed Networks With Minimum Delay," *IEEE Access*, vol. 8, pp. 151753-151763, 2020.

[4] Z. Zhang, Z. Li, C. Wu and C. Huang, "A Scalable and Distributed Approach for NFV Service Chain Cost Minimization," *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, 2017, pp. 2151-2156, doi: 10.1109/ICDCS.2017.210.

[5] A. Huff, G. Venâncio, V. F. Garcia and E. P. Duarte, "Building Multi-domain Service Function Chains Based on Multiple NFV Orchestrators," *2020 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, 2020, pp. 19-24, doi: 10.1109/NFV-SDN50289.2020.9289849.

[6] A. Khalid and F. Esposito, "Optimized Cuckoo Filters for Efficient Distributed SDN and NFV Applications," *2020 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, 2020, pp. 77-83, doi: 10.1109/NFV-SDN50289.2020.9289870.

[7] X. Fu, F. R. Yu, J. Wang, Q. Qi and J. Liao, "Performance Optimization for Blockchain-Enabled Distributed Network Function Virtualization Management and Orchestration," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 6, pp. 6670-6679, June 2020.

[8] H. Feng, J. Llorca, A. Tulino, D. Raz, A. Molisch, "Approximation Algorithms for the NFV Service Distribution Problem," 2021. Preprint, unpublished.

[9] W. Chiang and J. Wen, "Design and Experiment of NFV-Based Virtualized IP Multimedia Subsystem," *2018 3rd International Conference on Computer and Communication Systems (ICCCS)*, Nagoya, Japan, 2018, pp. 397-401, doi: 10.1109/CCOMS.2018.8463235.

[10] A. Jain, Sadagopan N S, S. K. Lohani and M. Vutukuru, "A comparison of SDN and NFV for re-designing the LTE Packet Core," *2016 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN)*, Palo Alto, CA, USA, 2016, pp. 74-80, doi: 10.1109/NFV-SDN.2016.7919479.

[11] M. Team, 2017 Mininet: An instant virtual network on your laptop (or other pc) - mininet. [Online]. Available: http://mininet.org

[12] A. Botta, A. Dainotti, A. Pescapè, "A tool for the generation of realistic network workload for emerging networking scenarios,"

Computer Networks (Elsevier), 2012, Volume 56, Issue 15, pp 3531-3547

[13] Oracle VirtualBox. 2019 [Online]. Available: https://www.virtualbox.org

[14] M. V. Bernal, I. Cerrato, F. Risso and D. Verbeiren, "Transparent Optimization of Inter-Virtual Network Function Communication in Open vSwitch," *2016 5th IEEE International Conference on Cloud Networking (Cloudnet)*, Pisa, Italy, 2016, pp. 76-82, doi: 10.1109/CloudNet.2016.26.

[15] M. Eiman, "Minimum Technical Performance Requirements for IMT-2020 Radio Interface(s)," Presentation. 2018. [Online] Cited 2022-04-04.

[16] Lei, W. et al. (2020). From 4G to 5G: Use Cases and Requirements. In: 5G System Design. Springer, Cham. https://doi.org/10.1007/978-3-030-22236-9_1.

[17] A. Gohar and G. Nencioni, "The Role of 5G Technologies in a Smart City: The Case for Intelligent Transportation System," *Sustainability*, vol. 13, no. 9, p. 5188, May 2021, doi: 10.3390/su13095188.

[18] Jose F. Monserrat, Adam Diehl, Carlos Bellas Lamas, and Sara Sultan. (2020). Envisioning 5G-Enabled Transport. 10.1596/35160.

[19] T. Neumuth, C. Bulitta, S. Hamm, F. Edelmann, A. Mittermaier, M. Rockstroh, C. Thuemmler, (2020). 5G Health - The need for 5G technologies in healthcare. 10.13140/RG.2.2.10915.27687.

[20] A. Ahad, M. Tahir and K. -L. A. Yau, "5G-Based Smart Healthcare Network: Architecture, Taxonomy, Challenges and Future Research Directions," *IEEE Access*, vol. 7, pp. 100747-100762, 2019, doi: 10.1109/ACCESS.2019.2930628.

[21] J. Lee, M. Azamfar, M. Miller. (2020). 5G and Smart Manufacturing.

[22] A. Chehri, A. Zimmermann. (2021). 5G Assisted Smart Manufacturing and Industrial Automation. In: Leitner, C., Ganz, W., Satterfield, D., Bassano, C. (eds) Advances in the Human Side of Service Engineering. AHFE 2021. Lecture Notes in Networks and Systems, vol 266. Springer, Cham. https://doi.org/10.1007/978-3-030-80840-2_44.

[23] H. C. Leligou, T. Zahariadis, L. Sarakis, E. Tsampasis, A. Voulkidis and T. E. Velivassaki, "Smart Grid: a demanding use case for 5G technologies," *2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, Athens, Greece, 2018, pp. 215-220, doi: 10.1109/PERCOMW.2018.8480296.

[24] E. Fountoulakis, Q. Liao and N. Pappas, "An End-to-End Performance Analysis for Service Chaining in a Virtualized Network," *IEEE Open Journal of the Communications Society*, vol. 1, pp. 148-163, 2020, doi: 10.1109/OJCOMS.2020.2966689.

[25] A. Fendt, C. Mannweiler, L. C. Schmelz and B. Bauer, "A Formal Optimization Model for 5G Mobile Network Slice Resource Allocation," *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Vancouver, BC, Canada, 2018, pp. 101-106, doi: 10.1109/IEMCON.2018.8615049.

[26] S.A. AlQahtani, W.A. Alhomiqani, "A multi-stage analysis of network slicing architecture for 5G mobile networks," *Telecommun Syst*, vol. 73, pp. 205–221, 2020, doi: 10.1007/s11235-019-00607-2

[27] M. D. Mauro, G. Galatro, F. Postiglione and M. Tambasco, "Performability of Network Service Chains: Stochastic Modeling and Assessment of Softwarized IP Multimedia Subsystem," *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 5, pp. 3071-3086, 1 Sept.-Oct. 2022, doi: 10.1109/TDSC.2021.3082626.

[28] S. Mustafiz, O. Hassane, G. Dupont, F. Khendek, M. Toeroe, "Model-driven process enactment for NFV systems with MAPLE," *Softw Syst Model*, vol. 19, pp. 1263–1282, 2020.

[29] H. U. Adoga and D. P. Pezaros, "Network Function Virtualization and Service Function Chaining Frameworks: A Comprehensive Review of Requirements, Objectives, Implementations, and Open Research Challenges," *Future Internet*, vol. 14, no. 2, p. 59, Feb. 2022, doi: 10.3390/fi14020059.

[30] J. Vergara-Reyes, M. C. Martinez-Ordonez, A. Ordonez and O. M. Caicedo Rendon, "IP traffic classification in NFV: A benchmarking of supervised Machine Learning algorithms," *2017 IEEE Colombian Conference on Communications and Computing (COLCOM)*, Cartagena, Colombia, 2017, pp. 1-6, doi: 10.1109/ColComCon.2017.8088199.

[31] A. Esmaeily, K. Kralevska and D. Gligoroski, "A Cloud-based SDN/NFV Testbed for End-to-End Network Slicing in 4G/5G," *2020 6th IEEE Conference on Network Softwarization (NetSoft)*, Ghent, Belgium, 2020, pp. 29-35, doi: 10.1109/NetSoft48620.2020.9165419.

[32] G. Ilievski, P. Latkoski, "Experimental evaluation of network packet latency within a distributed NFV Infrastructure," *29th Telecommunications Forum (TELFOR)*, 2021, pp. 1-4. doi:10.1109/TELFOR52709.2021.9653395.

[33] Wireshark, 2006 [Online]. Avalable: https://www.wireshark.org/

[34] H. Huang, W. Miao, G. Min, J. Tian and A. Alamri, "NFV and Blockchain Enabled 5G for Ultra-Reliable and Low-Latency Communications in Industry: Architecture and Performance Evaluation," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 8, pp. 5595-5604, Aug. 2021, doi: 10.1109/TII.2020.3036867.

[35] D. de Freitas Bezerra, G. L. Santos, G. Gonçalves, et al., "Optimizing NFV placement for distributing micro-data centers in cellular networks," *J Supercomput* vol. 77, pp. 8995–9019, 2021. doi:10.1007/s11227-021-03620-y

[36] G. Hu, Q. Li, S. Ai, T. Chen, J. Duan, Y. Wu, "A proactive auto-scaling scheme with latency guarantees for multi-tenant NFV cloud," *Computer Networks*, vol. 181, 2020. doi:107552. 10.1016/j.comnet.2020.107552.