Received: 27 February 2021 | Accepted: 12 March 2021
DOI: 10.1002/itl2.289

CASE STUDY

WILEY

Network slicing for 5G edge services

Michail-Alexandros Kourtis[1] | Thanos Sarlas[2] | Themis Anagnostopoulos[2] | Sławomir Kukliński[3] | Lechosław Tomaszewski[3] | Michał Wierzbicki[3] | Andreas Oikonomakis[2] | George Xilouris[1] | Ioannis P. Chochliouros[4] | Na Yi[5] | Alexandros Kostopoulos[4] | Harilaos Koumaras[2]

[1]ORION Innovations Private Company, Athens, Greece
[2]National Center of Scientific Research "Demokritos", Athens, Greece
[3]Orange Polska, Warsaw, Poland
[4]Hellenic Telecommunications Organization S.A. (OTE), Research Programs Section, Athens, Greece
[5]University of Surrey, Guildford, UK

Correspondence
Michail-Alexandros Kourtis, ORION Innovations Private Company, Athens, Greece.
Email: akis.kourtis@iit.demokritos.gr

Funding information
H2020 European Research Council, Grant/Award Numbers: 814956, 815178

In the current 5G technology domain network slicing already plays an important role as a critical enabler. An industry that 5G aims to disrupt is the vehicular one. In this paper the brief scope of the 5G-DRIVE research project is presented, regarding 5G vehicular research between the EU and China. In the frame of 5G-DRIVE a set of slicing mechanisms are investigated and evaluated in regard to their performance. Firstly, related to slicing mechanisms in the NFV domain the OSM orchestrator is measured in terms of scalability and performance. In the next experimental set related to RAN slicing, the Katana Slice Manager is evaluated and depicts how different slicing configurations can achieve different performance results. Furthermore, the paper showcases how 5G network slicing can be integrated as a key enabler to the stringent demands of a vehicular network environment. Finally, the paper concludes, setting future directions in the related field.

KEYWORDS
5G, edge cloud computing, network function virtualization, network slicing, OSM, V2X

1 | INTRODUCTION

The rapid proliferation of mobile technologies and communications has driven their inclusion in our daily life, across various sectors, "as a whole". It is expected to increase their essential role in various upcoming modern services and facilities. Similarly, mobile devices have become an indispensable accessory in our daily activities. Future International Mobile Telecommunications (IMT) systems and more specifically the 5G initiative plan to support emerging new vertical use cases, including applications that require very high data rate communications, a large number of connected devices, and ultra-low latency or high-reliability applications.

In this scope, 5G has identified three-pillar requirements for future networks: namely enhanced mobile broadband (eMBB), massive machine-type communications (mMTCs), and ultra-reliable low-latency communications (URLLCs). These requirements are vital for the support of upcoming user services that will demand significantly higher data rates and simultaneous multi-connectivity. However, the management and allocation of resources in a network level is still an open issue. The transition beyond 5G telecommunication networks brings more challenges concerning the provision, placement, migration, spectrum management and dynamic resource allocation.

The paradigm of network slicing has been introduced in order to address differentiated resource allocation requirements and bring service provision a step closer, especially for high demand vertical industries.[1] Network slicing will

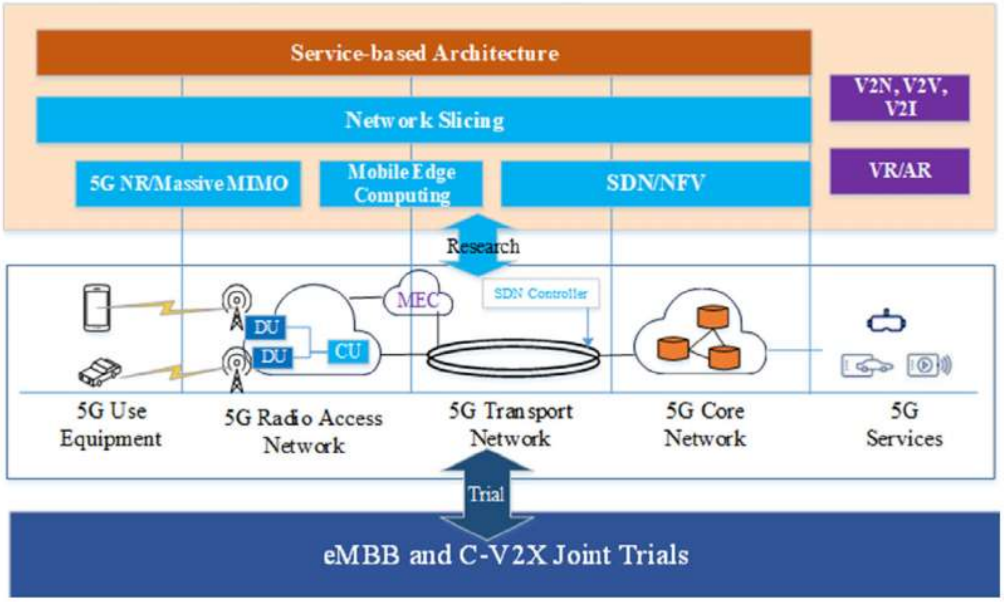

FIGURE 1 Overall architecture of 5G-DRIVE

play a pivotal role in addressing varied vertical applications by enabling dedicated virtualized network slices for each vertical.

In this paper, the 5G-DRIVE[2] approach is introduced, as a 5G enabled platform for vertical vehicular technologies. 5G-DRIVE, among other topics, focuses on the integration of virtualized network services running at the edge, that facilitate vehicular end-users and their respective systems. Additionally, considering critical slicing requirements of such an infrastructure, this paper presents a performance evaluation of the OSM NFV slicing mechanisms and RAN slicing evaluation measurements for Katana slicing framework.[3]

1.1 | Context and aim

Following by the previously described challenges towards a fast and reliable 5G evolution, the 5G-DRIVE project aims to perform a close collaboration between the European Union (EU) and China to synchronize 5G technologies and spectrum issues before the final roll-out of 5G. The project's overall concept is illustrated below, as shown in Figure 1, which shows the three "core" streams and also depicts the flow from research through adaptation into existing testbeds and commercial testbed deployments, to the real-world trials of the 5G radio access network (RAN) and of the more extensive 5G network. The 5G-DRIVE project "brings together" solid research competence, commercial-grade testbeds, and some of the stakeholders who will eventually become significant customers of 5G systems.

For the needs of the 5G-DRIVE experimentation and evaluation activities, a joint trial testbed has been defined with a 5G testbed installation, co-located in OTE and ORION Innovations premises. In the research stream, the project investigates network and RAN slicing, mobile edge computing (MEC), massive multiple-input multiple-output (MIMO) for the 5G NR (New Radio), as well as Software-Defined Network (SDN) & Network Functions Virtualization (NFV) techniques applied to different traffic and load scenarios. Research-centric methods and mechanisms related to NFV, SDN and network slicing of the project are under development and implementation into the testbed. In the scope of this study the testbed was used for the experimental evaluations of two network slicing mechanisms. The first one used the widely known OSM NFV orchestrator and was focused on the NFV domain slicing. The detailed results are depicted in Section 2. For the second part for the slicing measurements the Katana slice manager was used and carried out evaluation experiments for the RAN domain.

2 | OSM NFV SLICING EVALUATION

Current slicing solutions focus on different domains or a combination of thereof. However, at the edge, the research results are still in their infancy, especially regarding resource allocation across different layers of the network environment. Moreover, integration of MEC with network slicing is still an open issue. Therefore, this paper presents two network slicing solutions, along with their performance evaluation for a set of test network services. Firstly, the OSM slicing platform is presented and measured in terms of scalability.

2.1 | OSM slicing evaluation

The performance of slice orchestration is an important and so far, this problem did not receive enough attention from researchers. In this section, we will describe some experiments related to the performance evaluation of the OSM orchestrator. To make the assessment, we have used the methodology proposed in a study for 5G Key Performance Indicators (KPIs).[4] The proposed in the paper KPIs are slice agnostic. The list of the measured KPIs includes:

1. Slice Deployment Time (SDT)—a parameter that describes the interval between the slice deployment request and the moment in which slice is ready for operation. Unfortunately, this parameter depends on the slice template (blueprint) complexity, the performance of the orchestrator, and the time needed for the allocation of virtualized resources by the infrastructure. The slice complexity may deal with the footprint size of VNFs, their interconnection topology, amount of configuration parameters. Therefore, in a generic case, it is impossible to define the required value of SDT. It can be noted that SDT may be critical for some network slices, for example, on-demand or short-lived ones, but much less significant for long-lived slices.
2. Slice Deployment Time Scalability (SDTS)—a measure of scalability of slice deployment operations. To evaluate the scalability, it is possible to send N slice deployment requests of the same slice template and calculate SDTS in the following way:

$$SDTS = \frac{GSDT}{N \times SDT}$$

where GSDT is the overall time for the deployment of N identical slices and SDT is the deployment time of a single slice (as defined above). It is hard to determine the N value *a priori*. If the N value is too big, then there can be a problem with the availability of the requested resources. If it is too small, then the obtained result may not express the scalability of the orchestration well. We have used the SDTS parameter $N = 10$, and the SDTS is expected to be higher than 1.

3. Slice Termination Time (STT)—a parameter that describes the interval between the slice termination request and the moment in which all slice allocated resources are released. If the time is long, it decreases the efficiency of the infrastructure resources usage.

The lifecycle KPIs have been obtained by observing interactions between the OSS/BSS and NFVO. Hence, the OSS/BSS can determine both the beginning and the end of the procedure. It is worth noting that in the case of disturbances of intra-MANO communication (eg, OSS/BSS is notified about the delay of procedure execution due to the need of retrying). The main goal of the experiments was to evaluate the scalability of lifecycle orchestration of OSM. The experiments have been carried out using two identical computers, each of them had Intel Core i7-8700 CPU @ 3.2 GHz and 16 GB memory. On the first machine, OSM MANO was installed, whereas on the second machine the OpenStack used as VIM has been installed. The tests were performed through deployment and termination from 1 to 50 identical instances using *hackfest-basic-vnf* template on CirrosOS image. This minimal Linux distribution was designed for use as a test image on clouds such as OpenStack. To average the results, each test was conducted ten times. We have measured the slice deployment time by deploying 1, 3, 5, 10, 15, 20, 30, 40 and 50 instances simultaneously and one by one using the same template. We have measured the SDT and the GSDT parameters. The normalized slice deployment time has also been calculated. The obtained GSDT and SDTS results are presented in Figure 2A and B, respectively.

In Figure 2A, we have presented the normalized value of SDTS—a deployment time of a single slice when N slice instances are deployed in comparison to a single slice deployment time. In Figure 2B, the obtained measurements for

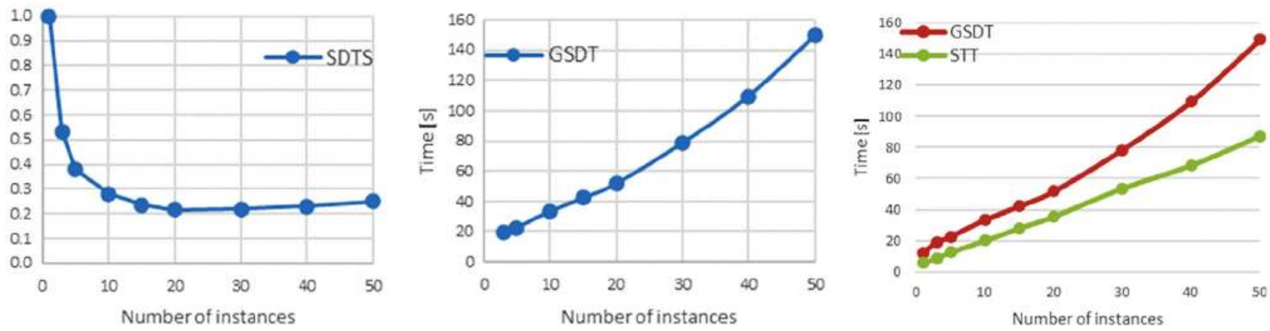

FIGURE 2 SDTS (A), GSDT (B) and GSDT vs STT (C) results for a various number of instances

N = 3-50 are presented. As it can be found the GSDT increases nearly linearly in the tested range. For N = {1,3,5}, we have received significantly higher values than for the larger value of N. It can be justified by the need of the initialization of some libraries and system processes. Linear SDT increase shows good scalability of the OSM orchestrator. The comparison of deployment (GSDT) and termination (STT) times of N slice instances is shown in Figure 2C.

As it can be seen, the termination time of a single slice is nearly constant for $N = 10$-50 and much higher for the termination of a single slice. The obtained results show the excellent scalability of the OSM orchestrator again. A longer time for a single instance can be explained by a time needed to delete some processes that are common for all slice instances. The time of slices termination is only two times shorter than the time required for their deployment.

3 | KATANA RAN SLICING EVALUATION

In this section a set of experimental tests is conducted for an assessment of the differentiation in latency performance is exhibited by certain slice configurations. The assessment is performed on the 5G network infrastructure based on Amarisoft 5GC and NR implementations configured in Standalone Mode (Option 2). The slice manager was appropriately extended in order to communicate and modify various parameters at the RAN and Core configuration of the 5G system and enforce certain slicing policies through a dedicated custom build EMS.

The slicing is enforced and measured on the platform in two steps. The first step includes the manipulation of the RAN layer numerology (ie, configuration of waveform parameters), in order to minimize the transmission windows, by fine tuning the Scheduling Request Period (srPeriod) and slot Period parameters of the gNB. The approach aims to minimize the round-trip time (RTT) in the RAN domain, by modifying the Time Division Duplex (TDD) slot to Downlink-Uplink (DL-UL) pattern and reduce the scheduling request period. Furthermore, the srPeriod is a special physical layer message for UE to request UL Grant in order to transmit in Physical Uplink Shared Channel (PUSCH). Regarding, slot Period the main idea is based on TDD where the UL-DL channels transmit simultaneously, and by tailoring the time periods of the lots the performance can be tailored to our needs.

The second step involves slicing enforcement at the 5G Core domain, by allocating different PLMNs or APNs to different services per group in order to differentiate traffic and be able to enforce the necessary policies.

There is currently no support at the RAN for multi-slicing, although 3GPP standards provide the relevant provisions, at the RAN level using different types of slices (ie, URLLC/eMBB/mMTC). Since there are no devices that can utilize more than one slice at the same time (limitation due to APN dependency and SIM card structure) parallel slicing can be achieved via allocating an eMBB slice with specific bandwidth at the backhaul and Core network and perform the URLLC slice enforcement at the RAN level by modifying the srPeriod and slot Period parameters at the gNB. This paradigm of concurrent slicing is critical to the operation and robustness of vehicular edge communications, as it maps different service requirements to slice instances of variant bandwidth and latency.

Most of the currently available slicing mechanisms depend heavily in NFV/SDN capabilities plus network virtualization and QoS/traffic prioritization. Katana for the needs of the proposed platform enforces policies both on the Core and the RAN domain of the infrastructure. Based on this implementation a set experimental tests were performed which measured the latency for various packets sizes of a 5G system, after having allocated an eMBB slice and then enforcing different URLLC slices over it.

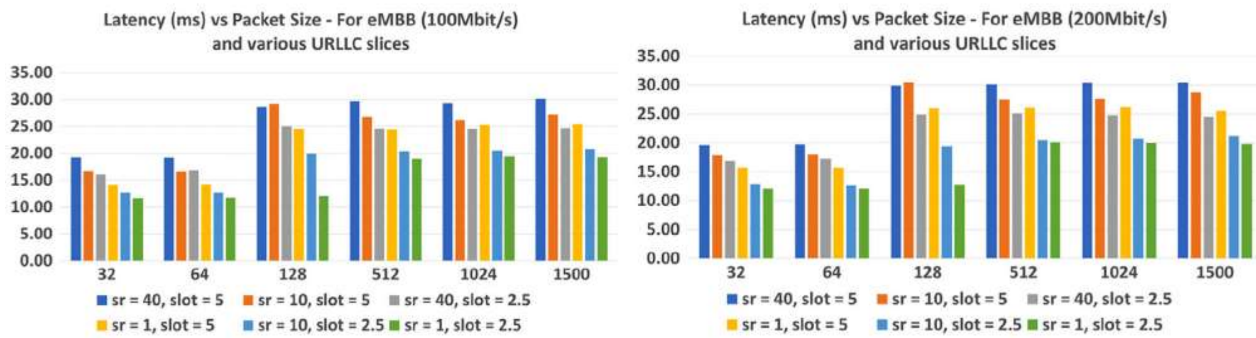

FIGURE 3 Latency results for a 100(Mbits/s) and 200(Mbits/s) eMBB slice and 6 different uRLLC slices

The first set of results are shown in Figure 3A, where an eMBB slice of 100Mbits/s was allocated at the backhaul, and a set of URLLC slices were enforced and measured in terms of latency, for srPeriod values of {1,10,40} and slot Period values of {2.5, 5}. As it can be deducted the RAN modifications can greatly affect the latency of the system for all packet sizes, and in some cases an improvement of up to 57% in the case of srPeriod = 40, slot Period = 5 and srPeriod = 1, slot Period = 2.5 for packet size of 128 bytes.

Furthermore, in the second set of experimental tests an eMBB slice of 200 Mbits/s was allocated and the same set of RAN uRLLC slice parameters were enforced. The results as depicted in Figure 3B, indicate a similar behavior to the eMBB slice results of 100 Mbits/s, with the case of 128 byte packets to demonstrate again the largest improvement.

It is evident from both set of results that the case of the URLLC slice of srPeriod = 1 and slot Period = 2.5 display the best performance in terms of latency for a URLCC slice, with an average minimum latency of ~12 ms. These values can be indicated as the best achievable latency performance measured in the proposed 5G platform for end-to-end measurements.

4 | CONCLUSION

The paper presented the 5G enabled prototype architecture for edge services of the 5G-DRIVE project, regarding the current open challenges in the field of 5G network slicing but also presented scalability evaluation tests for the OSM NFV slicing mechanism and RAN multi-slicing measurements tests with the Katana slice manager, a slicing mechanism.

Furthermore, the presented Katana slicing mechanism will also be evolved in order to support more dynamic slice resource allocation in the RAN, as was investigated in cloud-enabled small cell systems[5] and end-to-end 5G testing framework.[6]

ACKNOWLEDGMENTS

The research leading to these results has been supported by the H2020 5G-DRIVE ("HarmoniseD Research and trIals for serVice Evolution between EU and China") project, (no. 814956) and the 5GENESIS H2020 5G-PPP (no. 815178).

DATA AVAILABILITY STATEMENT

Data available on request from the corresponding authors. Extended datasets, methodology and extensive results will be included in the public deliverables of the 5G-DRIVE project. In the official site there is a dedicated section for all the public deliverables of 5G-DRIVE, which include extensive results of the experimental procedures. https://5g-drive.eu/resources-and-results/project-deliverables/

ORCID

George Xilouris https://orcid.org/0000-0003-2695-4534

REFERENCES

1. Elayoubi SE, Ben Jemaa S, Altman Z, Galindo-Serrano A. 5G RAN slicing for verticals: enablers and challenges. *IEEE Commun Mag*. 2019;57:28-34.

2. Chochliouros IP; Spiliopoulou AS, Arvanitozisis D, et al. 5G Promotive actions based upon enhanced mobile broadband (EMBB) communication trials between the EU and China. 2020 5th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), IEEE; 2020.
3. Katana. https://github.com/medianetlab/katana-slice_manager
4. Kukliński S, Tomaszewski L. Key performance indicators for 5G network slicing. IEEE Conference on Network Softwarization (NetSoft), Paris, France, 2019. https://ieeexplore.ieee.org/document/8806692. Last accessed: 19-3-2021.
5. Kourtis M.- A., Blanco B., Perez-Romero J. et al. A Cloud-Enabled Small Cell Architecture in 5G Networks for Broadcast/Multicast Services. *IEEE Transactions on Broadcasting*. 2019;65(2):414-424. https://dx.doi.org/10.1109/tbc.2019.2901394.
6. Gardikis G, Papadakis N, Perentos A, et al. The 5GENESIS testing facility as an enabler for integrated satellite/terrestrial 5G experimentation. 2019 IEEE Wireless Communications and Networking Conference Workshop (WCNCW), IEEE; 2019.

How to cite this article: Kourtis M-A, Sarlas T, Anagnostopoulos T, et al. Network slicing for 5G edge services. *Internet Technology Letters*. 2021;e289. https://doi.org/10.1002/itl2.289