2023 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN) | 979-8-3503-0254-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/NFV-SDN59219.2023.10329589

Proactive ML-assisted and Quality-driven Slice Application Service Management to keep QoE in 5G Mobile Networks

Felipe S. Dantas Silva, Ayuri Bessa, Sérgio Silva, Samuel Ferino, Pablo Paiva, Marcos Medeiros, Lucas Silva, José Neto, Kevin Costa, Charles Santos, Douglas Maciel, Lucileide Silva, Anderson Inoue, Roger Immich, Eduardo Aranha, Allan Martins, Vicente Sousa, Uirá Kulesza, Marcelo Fernandes, Marcos Salvador, Guilherme Pupio, Ramon Fontes, Augusto Neto
Federal University of Rio Grande do Norte (UFRN), Natal, Brazil
felipe.dantas.046@ufrn.edu.br

***Abstract*—Network slicing is a core feature that 3GPP defines in the 5G realm, allowing operators and providers to control traffic resources more granularly. This is achieved by provisioning different network services as independent and isolated network logical partitions atop a shared network physical infrastructure. A network slice must be managed at each constituent part's granularity to ensure an acceptable Quality of Experience (QoE) over time. The state-of-the-art provides plenty of solutions at the network functions and resources management level. In this demo, we propose a service application management solution that applies Machine Learning (ML) to conduct predictive analysis to yield the anticipated detection of service application quality degradation. The solution enables further, fully automatic adjustments with high assertiveness to the network slice structure to maintain QoE. The validation was conducted through test trials on an emulated testbed designed to provide flexibility and dynamics, enforcing network slicing in the Radio Access Network (RAN) and core network tiers and adapting resource allocation policies according to the slice instance's needs.**

***Index Terms*—5G, slicing, machine learning, QoE, mobility.**

I. Introduction

Network slicing [1] stands out for its ability to deploy 5G infrastructures with computing, storage, and network services in a personalized and elastic way. This capacity is provided through a set of network resource components called slice, which can be extended through physical resource virtualization strategies, network control software, and infrastructure cloudification [2]. Slices are materialized through the capabilities promoted by, among others, the Network Function Virtualization (NFV) and Software-Defined Networking (SDN) paradigms [3]. A network slice is structured by manageable parts, ranging from a set of Virtual Network Functions (VNF) and resources (e.g., computation, storage, and network resources) bound together to lay the traffic data demands of service applications running within slice parts (Slice Application Service – SAS). For instance, operators are allowed to allocate dedicated portions of the underlying network infrastructure to cater to diverse application verticals, such as ultra-low latency, mission-critical services (including autonomous vehicles and smart factories), intense-bandwidth experience-enhanced multimedia services (e.g., holographic calls and immersive video) and ultra-reliable-connected massive Internet of Things (IoT) deployments [4].

Recent research footprints on managing solutions tailored to slicing-defined 5G systems suggest that the joint orchestration of SASs for running inside network slice manageable parts is essential for delivering best-connected and best-served services toward User Equipment (UE) [5]. Therefore, the slicing management plane must deal with the quality levels of SASs to guarantee users' experience perspectives over time. However, the increasingly challenging and dynamic aspects of 5G mobile systems make quality management of SASs a non-trivial task and require advancing beyond existing strategies through specialized and proactive new approaches. We believe new approaches must be designed with intelligent capabilities to enable slicing re-orchestration procedures based on pre-established Service-Level Agreement (SLA) disrupting events predictions.

In a densely populated 5G mobile system, users may experience a handover to a location far from the offloaded content. This could be due to the initial network slicing setup or the lack of transmission rate within the slice caused by the unrestricted admission of users to best-effort slices. In order to prevent user experience degradation, it is essential to implement self-managing, self-healing, and self-optimizing procedures in the running slice with the mobile service consumer without causing any perceivable service interruption and enabling end-to-end network services.

This demonstration stands out over existing solutions by presenting a new quality-level management plan capable of operating at SAS granularity. Our solution relies on a proactive slicing monitoring service, which features real-time cell throughput predictive analysis to carry out slice re-orchestration, enforced by a machine learning-based load-balancing strategy, seeking to avoid the incidence of user Quality of Experience (QoE) violation over time. We validated the proposed scheme in a testbed featuring real-world slicing enablers for delivering programmability and virtualization capabilities with widely employed solutions for managing and orchestrating slicing infrastructures.

To the best of our knowledge, this is the first operational demonstration of a mobile network scenario with network-slicing capabilities enabled by real-world orchestration mechanisms. We present our proposed approach and its integration

with the testbed in Section II and provide details about our planned demonstration in Section III.

II. System Architecture

Figure 1 depicts our proposed approach atop the considered slicing-enabled vehicular network testbed ecosystem. SAS instances are deployed in the form of VNFs, and as such, the NFV Management and Orchestration capabilities implemented by the ETSI-hosted Open Source MANO (OSM)[1] approach are exploited.

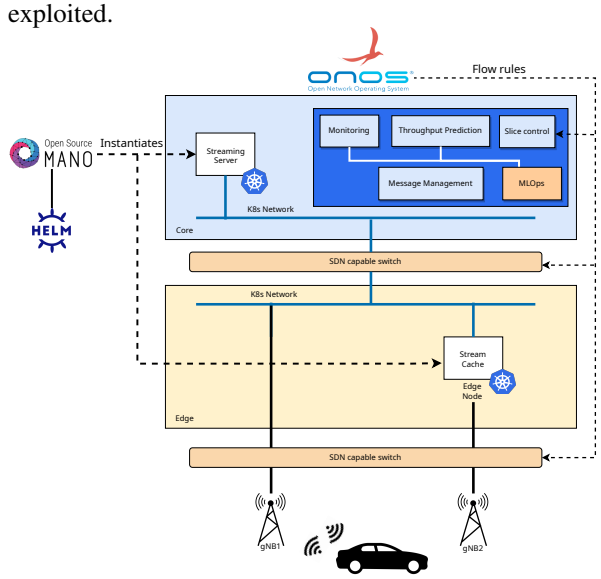

Fig. 1. Positioning of the slicing-aware network services in the proposed slicing-enabled vehicular network testbed ecosystem.

The main functionalities provided by the VNFs are described as follows:

A. Monitoring

Gathers and shares data on network traffic and Key Quality Indicators (KQIs) to assist in predicting SAS throughput. The approach comprises distributed agents enabled with application knowledge requirements (e.g., required/transmitted bitrate and tolerable packet loss rate) for on-site slice-part collecting.

B. MLOps

MLOps manages the lifecycle operations related to the throughput prediction performed by the ML, such as model deployment, maintenance, and performance monitoring. To achieve this goal, we build a dataset from data collected in our mobile network testbed to train the ML model with essential network traffic data characteristics, including relevant features such as throughput and packet loss rate.

C. Throughput prediction

This VNF utilizes a feed-forward Artificial Neural Network (ANN) to estimate network overload by analyzing slice throughput. It consists of two layers, the first containing 20 neurons and the second with ten neurons. The Rectified Linear Unit (ReLU) activation function introduces non-linearity into the model, thereby improving its predictive power. The learning rate is set at 0.001, and epsilon is set at 1e-8 to ensure efficient convergence during training. Thanks to a Python object-loading system, the Adam optimizer is employed to avoid frequent model retraining. The VNF receives monitoring data from the monitoring service as input for the machine learning-based predictor. If it predicts potential slice throughput overload and a potential user QoE violation, it will activate the slice control service to implement measures aimed at preventing network strain and protecting the user experience.

D. Slice control

Responsible for enforcing slicing isolation and Quality of Service (QoS) guarantees through dedicated transmission rates per slice. Upon receiving a slice create or update request, it deploys the appropriate slicing configuration along all network elements (e.g., switches and gNBs), which composes the datapath to the serving VNF (e.g., video streaming service). The slice control service implements customized slicing features through the centralized WAN Infrastructure Manager (WIM) approach provided by the Open Network Operating System (ONOS[2]) SDN controller.

On receiving a slice overload prediction triggered by the throughput prediction service, it calls the internal procedures in charge of the slice re-orchestration to instantaneously apply appropriate measures for ensuring the user's QoE. The slice control service handles flow rules centrally through the ONOS northbound API to provide service continuity.

III. Demonstration

This demonstration showcases our ML-assisted network-slicing orchestration with QoE control scheme in a 5G mobile network scenario with mobility demands employing a vehicle moving between wireless cells while consuming real-time video streaming. For a more detailed overview of our planned presentation, we present the deployment setup with information about the testbed and then describe the use case scenario.

A. Testbed deployment setup

The testbed is based on an infrastructure deployed in the REGINA Lab premises at the Federal University of Rio Grande do Norte (UFRN), Brazil, which enables orchestration capabilities with real-world slicing enablers. Figure 2 depicts the primary components implemented in the testbed.

A cloud computing infrastructure is built on an OpenStack[3] setup. The cloud's primary goal is to provide cloudification capabilities to achieve flexible and scalable resource allocation for deploying essential orchestration services such as Kubernetes[4], which acts as a Virtual Infrastructure Manager (VIM), OSM for deploying the service VNFs through the

[1]https://osm.etsi.org/

[2]https://opennetworking.org/onos/

[3]https://www.openstack.org/

[4]https://kubernetes.io/

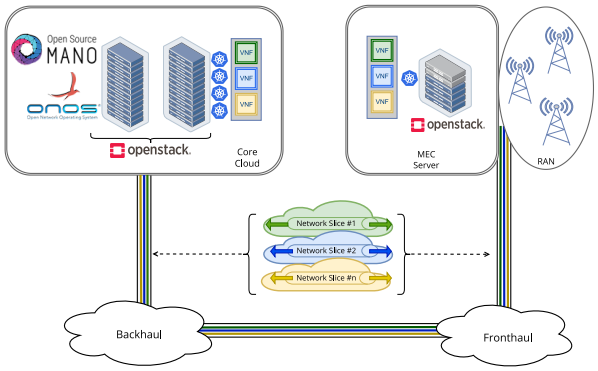

Fig. 2. Testbed deployment setup.

VIM, and ONOS which will serve as the network WIM. The network infrastructure is arranged in a classical leaf-spine datacenter topology, where spines represent the Core Network (CN) and the leafs identify edge-cloud domains, which provide connectivity to the Radio Access Network (RAN). The CN, edge-clouds, and the RAN are emulated within a joint setup between Containernet [6] and Mininet-WiFi [7]. The vehicular scenario is integrated into the network infrastructure with the Simulation of Urban Mobility (SUMO) [8].

B. Use case

Figure 1 illustrates the use case scenario adopted for the demonstration. It was built following the definitions described in Section III-A and according to Figure 2. In the depicted scenario, an OSM service function chain offers live-streaming in a virtualized content delivery network slicing. The streaming service comprises a streaming server placed in the CN, which feeds a streaming cache placed in a far edge cloud (i.e., edge cloud #2).

The use case involves a vehicle moving across an urban predefined trajectory while consuming the video streaming from the service chain, configured to seamlessly deliver data flows directly from the streaming server placed in the CN. According to Figure 1, the vehicle is first associated with the gNB1. The vehicle slicing connectivity is attached to the best-effort slice profile, configured with a shared transmission rate between all the users connected to the same slice. At the creation of the slice, all the services VNFs are properly instantiated to perform their respective tasks, as described in Section II.

At a specific time, the vehicle reaches the edge of the wireless coverage area of gNB1 and initiates a handover to gNB2. After some time, the throughput prediction service detects an overload event caused by the unpredictability of resource consumption within the best-effort slice type. Upon detecting the overload event, the slice control is triggered, and the re-orchestration procedure enforces the slicing resources by updating the slicing definitions, thus ensuring QoS guarantees meet the vehicle's QoS requirements for video consumption. Additionally, the slice control manages the necessary flow rules in the network elements, ensuring that the cache node maintains the continuity of video transmission.

In summary, the proposed ecosystem comprises the infrastructure that interoperates to maintain an acceptable video streaming QoE. This is achieved by orchestrating network slices with computational intelligence to anticipate events that could degrade communication. The following are the key features and contributions of this demo:

- Introducing a network fully featuring programmability and service automation;
- Proposing multiple virtualized and isolated logical networks to be built atop a physical network, meeting differentiated SLA requirements;
- Potentially prevent overload situations as well as improve throughput performance.

ACKNOWLEDGMENT

This research was partially funded by Lenovo, as part of its R&D investment under Brazilian Informatics Law, and by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

REFERENCES

[1] Ibrahim Afolabi et al. Network slicing and softwarization: A survey on principles, enabling technologies, and solutions. *IEEE Communications Surveys Tutorials*, 20(3): 2429–2453, 2018. doi: 10.1109/COMST.2018.2815638.

[2] Felipe S. Dantas Silva et al. Necos project: Towards lightweight slicing of cloud federated infrastructures. In *2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft)*, pages 406–414, 2018. doi: 10.1109/NETSOFT.2018.8460008.

[3] Stuart Clayman et al. The necos approach to end-to-end cloud-network slicing as a service. *IEEE Communications Magazine*, 59(3):91–97, 2021. doi: 10.1109/MCOM.001.2000702.

[4] Felipe S. Dantas Silva et al. Network slicing mobility aware control to assist handover decisions on e-health 5g use cases. In *2022 International Wireless Communications and Mobile Computing (IWCMC)*, pages 1034–1039, 2022. doi: 10.1109/IWCMC55113.2022.9825010.

[5] Alcardo Alex Barakabitze et al. Qoe management of multimedia streaming services in future networks: A tutorial and survey. *IEEE Communications Surveys & Tutorials*, 22(1):526–565, 2020. doi: 10.1109/COMST.2019.2958784.

[6] Manuel Peuster et al. Containernet 2.0: A rapid prototyping platform for hybrid service function chains. In *2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft)*, pages 335–337. IEEE, 2018.

[7] Ramon R Fontes et al. Mininet-wifi: Emulating software-defined wireless networks. In *2015 11th International Conference on Network and Service Management (CNSM)*, pages 384–389. IEEE, 2015.

[8] Michael Behrisch et al. Sumo–simulation of urban mobility: an overview. In *Proceedings of SIMUL 2011, The Third International Conference on Advances in System Simulation*. ThinkMind, 2011.