Architecting Orchestrators in Dynamically Evolving Scenarios: From Network-Aware Micro-Services to Application-Aware Network Slices

Roberto Bruschi
DITEN-University of Genoa
CNIT S2N Nat. Lab
Genoa, Italy
roberto.bruschi@unige.it

Franco Davoli
DITEN-University of Genoa
CNIT S2N Nat. Lab
Genoa, Italy
franco.davoli@unige.it

Diego San Cristobal
Ericsson Spain
Madrid, Spain
diego.san.cristobal.epalza@ericsson.com

Manuel Fuentes
Fivecomm
València, Spain
manuel.fuentes@fivecomm.eu

Laura González
ABB
Madrigalejo del Monte, Spain
laura.gonzales@es.abb.com

Dimitrios Klonidis
Ubitech
Athens, Greece
dklonidis@ubitech.eu

Chiara Lombardo
DITEN-University of Genoa
CNIT S2N Nat. Lab
Genoa, Italy
chiara.lombardo@unige.it

Manuel Lorenzo
Ericsson Spain
Madrid, Spain
manuel.lorenzo@ericsson.com

Miguel Ángel Martínez
YBVR
Toledo, Spain
ma@ybvr.com

Pablo Trelis
Fivecomm
València, Spain
pablo.trelis@fivecomm.eu

Thanos Xirofotos
Ubitech
Athens, Greece
txirofotos@ubitech.eu

***Abstract*—The evolution of mobile networks beyond the fifth generation (5G) and toward the sixth (6G) sees an unprecedented trend in the integration of the fixed and mobile network segments, as well as in network automation and self-organization. In this context, a relevant goal is the simplification and automation of the management of Network Applications (nApps) onto 5G and beyond (B5G) infrastructures. The 5G-INDUCE European Project is pursuing this goal, by creating a multiple orchestration framework with a clear separation of concerns between the cloud-native application and networking domains. The interaction between a Network Application Orchestrator (NAO) and the network Operations Support System (OSS) enables the intent-based creation of application-aware network slices and their dynamic adaptation to application and network changing conditions. The paper describes the design principles of the 5G-INDUCE orchestration framework and presents results of one of the eight Use Cases that are being demonstrated on three different Experimentation Facilities (ExFas).**

***Keywords*—B5G, Network Automation, Orchestration Frameworks, Dynamic Adaptation, Dynamic Slicing**

I. INTRODUCTION

The advent of 5G and its evolution toward the next generation have produced a profound change in the interaction between vertical cloud-native applications and the underlying networking services offered by the Telcos. 6G requirements and foreseen application environments are quite demanding in terms of Key Performance Indicators (KPIs) and network automation (see, e.g., [1] for a comprehensive vision). Orchestrators of micro-services in the vertical applications domain and of Virtual Network Functions (VNFs) in the networking domain play a relevant role in the evolution toward automated and dynamically adaptive network slicing; likewise, experimenting and demonstrating new application environments opens up new perspectives for the future applications landscape.

This paper highlights these two aspects, being pursued in the 5G PPP H2020 European Project 5G-INDUCE [2]. Section II describes the main features of the orchestration platform developed in the framework of the project. Section III reports some results on the experimentation of one of the eight Use Cases (UCs) being tested over three different ExFas, and Section IV contains the conclusions.

II. THE 5G-INDUCE PLATFORM

The 5G-INDUCE platform is specifically conceived for simplifying and automating the management of Network Applications (nApps) onto 5G and beyond (B5G) infrastructures. At a glance, the platform aims to mostly hide

979-8-3503-6595-5/24/$31.00 ©2024 IEEE

the complexity of the 5G environment to application developers and providers and to make the development, deployment and operation of 5G-ready nApps similar to the well-known corresponding processes applied to cloud-native applications in cloud computing environments.

The design principle along which the platform has been implemented is based on the separation of concerns between the orchestration at the domain of vertical applications and at the domain of the network service providers, already consolidated within the previous MATILDA project [3][4]. The former – Network Application Orchestrator (NAO) – is aimed at easing the task of the application developer in the creation of cloud-native applications enriched with the description of their micro-services' communication requirements (nApps); the latter deals with the creation and lifetime management of network slices on the basis of slice intent declarations, which is managed by the Operation Support System (OSS) in the Telco domain.

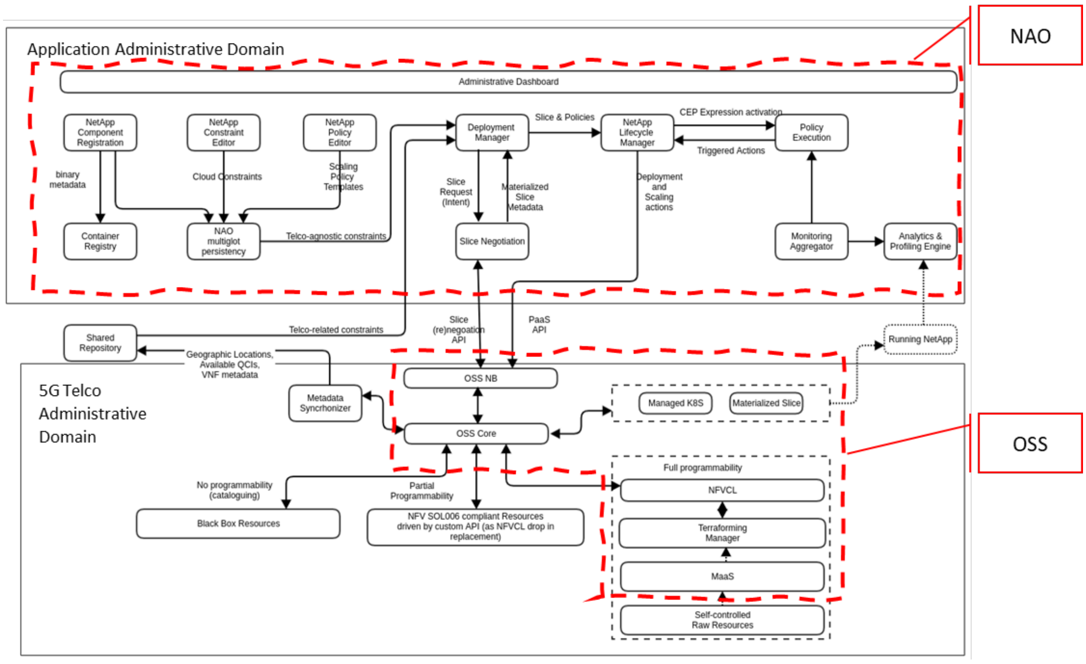

Fig. 1. The 5G-INDUCE architecture.

Such orchestrators obviously need to strictly cooperate to allow the coherent use and configuration of computing, storage and network resources and ensure the correct end-to-end deployment of a 5G-ready application along with the supporting network services. Specifically, the interworking between NAO and OSS is realized through an intent-based interface, called "5G and edge computing slice negotiation interface". Upon nApp deployment requests, through the slice negotiation interface, the NAO can request, negotiate and obtain from the OSS the needed computing resources at the edge facilities where to run nApp components, as well as the networking services to ensure the connectivity among such resources and with the User Equipment (UE). The OSS has the task of analysing operational and performance (soft and hard) constraints expressed by the NAO slice request, and, consequently, selecting the most suitable computing facilities and network services complying with the requirements. The initial slice intent produced by the NAO towards the OSS includes the application graph, annotated with QoS and operational requirements. Following the successful completion of the negotiation, the OSS provides back to the NAO the "materialized slice", which describes the set of needed network services and resources, instantiated and configured, ready to support the nApp.

Moreover, the 5G-INDUCE NAO-OSS interface supports the modification and reconfiguration of the slice within its lifecycle and enables advanced operations to deal with UE mobility and dynamic QoS/operational requirements. In more detail, 5G-INDUCE nApps (or some of their components) can be lively scaled and relocated onto computing facilities in new geographical areas or at different network infrastructure aggregation levels in a transparent and smooth fashion. nApp traffic from and to UEs is steered accordingly (by coherently updating the configuration of network slices and related network services), while the platform adapts the nApp geographical scope and properly scales the components and network services on each area depending on the number of hosted nApps, the local workload, and the dynamic QoS/operational requirements.

Figure 1 depicts the general 5G-INDUCE conceptual architecture and highlights the NAO and OSS platform components, which are briefly described in more detail in the remainder of this Section.

A. The Network Application Orchestrator (NAO)

The 5G-INDUCE NAO is the upper part of Fig. 1 and is responsible for the functionalities of registering nApps and all their components, authoring the application deployment and runtime policies, providing application monitoring, negotiating slices and managing the operational state of the nApp within the scope of an allocated slice. The NAO interplays with the lower part of Fig. 1, the OSS, through a slice intent-reply mechanism for the deployment and run-time management of the nApp lifecycle, in order to dynamically enable, scale and modify nApp capabilities according to relevant events: e.g., the bandwidth of the radio link, the scaling of the nApp deployment including alternating the respective required resources, etc.

The NAO provides the interfacing layer with the end user (i.e., the vertical industry expert and the application developer) for managing the deployable applications and their features. Fig. 2 depicts the internal structure of the NAO architecture, highlighting the latest 5G-INDUCE updates on a per module basis. The latest developments are some updated submodules, i.e., monitoring, policies and analytics (in orange) and they are related with the nApps' runtime functionalities. Similarly, the new developments, i.e., slice update, slice dispatch and handler (in green) are related with the slice update feature of the 5G-INDUCE Platform. The arrows that are marked as slice intent/slice are the actions supported by the interface with the OSS.

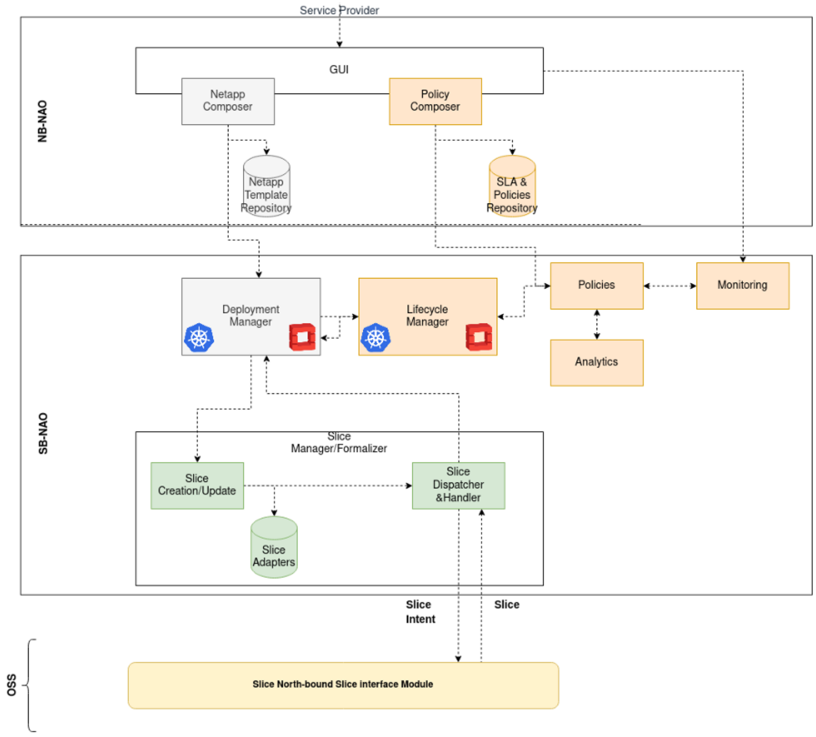

Fig. 2. Overview of the 5G-INDUCE NAO.

The internal modules of the NAO in Fig. 2 are grouped into the North-Bound (NB) and South-Bound (SB) modules. The NB modules of the NAO provide the interface with end users, specifically, nApp registration, creation and management, nApp monitoring and policies editing. For these functionalities, there are dedicated Graphical User Interfaces (GUIs) that combine several panels, in order to project the collected information from the deployment and lifecycle of every deployed nApp. The SB modules of the NAO are the "back-end" for the functionalities of the NB, plus the orchestration functionalities that are the initial deployment of an application, along with the lifecycle operations to maintain its operational state. Additionally, the creation and reception of the slice from the OSS are also SB functionalities of the NAO. The NAO NB Services include the nApp Composer and the Policy Composer; the SB Services comprise Monitoring, Deployment Manager, Lifecycle Manager, and Slice Manager.

B. The Operation Support System (OSS)

The 5G-INDUCE Operation Support System (OSS), depicted in Figure 3, is in charge of managing all functions and operations required for the nApp placement over edge computing facilities and for its connection to a (properly configured) network slice, as well as maintaining the information on all the data regarding the deployed applications, network services, and available infrastructure resources.

The OSS is designed according to a highly modular architecture: all the software services are state-of-the-art cloud-native software, i.e., stateless services (or more precisely services with a state maintained in an external database; namely, MongoDB [5] and Prometheus [6]), inherently parallelizable. The 5G-INDUCE OSS architecture is organized in a suite of five main software services, grouped into two main modules: the North-Bound OSS (NB-OSS) and the South-Bound OSS (SB-OSS). The former module is meant to front-facing the NAO by managing slice negotiations for nApps, and to maintain metadata (e.g., coverage area served, operational capabilities, etc.) of one or multiple onboarded SB-OSS modules. The SB-OSS is meant once per each different administrative network/computing resource domain onboarded onto the OSS. To reflect the different programmability levels exposed by such administrative domains (e.g., the various ExFa testbeds), the SB-OSS has been designed as a chain of software services that can be selectively activated to gain access to various programmability levels, passing from a simple catalogue of available resources in case of no programmability, up to the complete terraforming of the physical infrastructure in case of full programmability. Even if not made explicit in the following, in order to maximize the flexibility of the approach, the SB-OSS has the capability to maintain different programmability levels for edge computing and network (service/slice) resources in the same administrative domain.

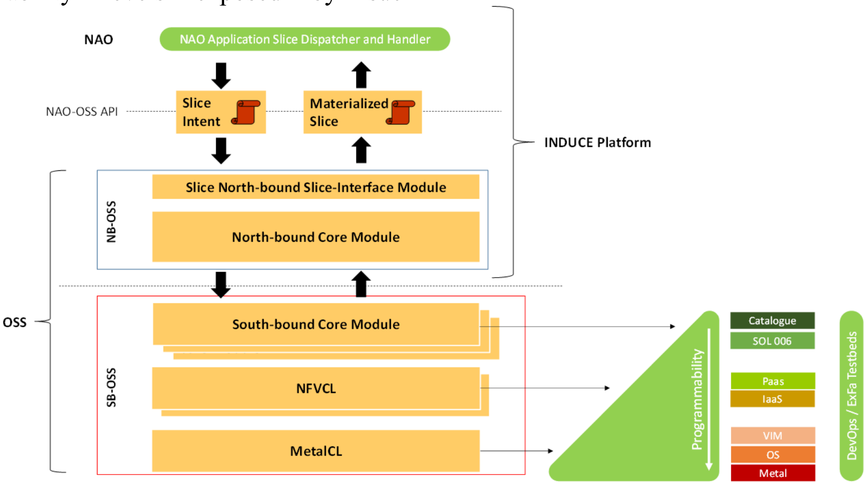

Fig. 3. The OSS architecture.

In more detail, the NB-OSS is composed of two main services: the Slicing-Interface and the North-Bound Core services. The Slicing-Interface service is meant to implement the OSS Application Programming Interfaces (APIs) for the interface with the NAO. The NB Core service is in charge of two main tasks: onboarding SB OSS instances, and suitably processing and propagating slicing requests/replies between the Slicing-Interface service (and then the NAO) and the relevant SB-OSS(es).

The SB-OSS includes three “chained” services: the South-Bound Core service, the NFV Convergence Layer (NFVCL), and the Metal Convergence Layer (MetalCL). The South-Bound Core service is the only mandatory element in the SB-OSS, and it is devoted to process the slice instantiation/modification/de-instantiation requests and related resources. This service is the key component for providing adaptive programmability: if the NFVCL and the MetalCL services are available, the SB Core can request them the setup or the change of new or existing network slices/services and of infrastructure resources (e.g., of OpenStack Virtual Infrastructure Manager (VIM) instances and of the hosting physical servers). In case that bare-metal or virtualization programmability levels in an administrative domain are not exposed to the 5G-INDUCE platform, the SB-Core can dynamically request an external NFV framework for the needed slices/configurations, or simply catalogue the pre-configured resources (e.g., a 5G network slice) statically dedicated to the 5G-INDUCE platform.

The role of the NFVCL within the SB-OSS is to manage the lifecycle of NFV services to provide suitable connectivity to nApp components and UEs in fully automated and zero-touch fashion. If not provided by the bare metal layer, the NFVCL is also in charge of providing and maintaining cloud-native computing frameworks at edge facilities (i.e., realizing Kubernetes clusters as NFV services).

The MetalCL is the service dedicated to managing and terraforming bare-metal resources (i.e., physical servers and hardware network equipment) to create Infrastructure-as-a-Service/Platform-as-a-Service (IaaS/PaaS) environments compliant with the 5G-platform needs. Also in such case, this service allows the dynamic Day-0 to -N lifecycle management of operating systems in the servers, of configurations in network equipment, an of complex distributed applications like OpenStack and Kubernetes.

III. USE CASE EXPERIMENTAL RESULTS

The 5G-INDUCE project is conducting experimental KPI measurements on eight industrial UCs, described in detail in [7], over three ExFas (in Greece, Italy and Spain, respectively) with different capabilities. We briefly report here some results concerning one of them; namely, UC3 “VR Immersion and AGV Control”, conducted on the Spanish ExFa. UC3 targets the use of 360°-video as a troubleshooting tool enabled by the 5G infrastructure for industrial environments, by adopting Virtual Reality (VR) to allow the industrial operator getting a quick high-quality interactive view of what is happening on the surrounding environment of each Automated Guided Vehicle (AGV) in a fleet.

The operator views the scene through a VR headset (Oculus). Upon operator's request on the player app menu, the application opens the current live video including the real-time data overlay (Fig. 4). In the following, we report some results of the service-level validation and KPI measurements.

Fig. 4. Video + Overlay in Oculus screenshot.

A. Glass2glass Latency

The glass-to-glass latency is defined as the time that elapses since the camera takes a frame until that frame is represented in the client device (glasses or tablet). Several measurements (taken visually with a tablet in the vicinity of the camera) provided an average latency of 1.2 s, well below the KPI target level of 2 s.

B. Displayed data latency

The KPI was measured by aggregating the measured average latencies of: i) opening the overlay HTML webpage (50 ms); ii) getting data from the database (100 ms); iii) generating the screenshot (downloading the HTML webpage and saving it on a file – 120 ms); iv) downloading the screenshot (50 ms). The total time of 320 ms is below the target KPI of 500 ms.

C. CPU load.

The nApp is not demanding in terms of CPU load, because it just re-streams the video coming from the AGV to the clients requesting the video; there are no video re-encoding tasks. The CPU usage of the container during part of the testing, shown in Fig. 5, is around 3.5%, fulfilling by far the KPI of 80%.

A. Other metrics

Fig. 6 reports measurements of the TCP download latency, which is also within the desired limits.

IV. CONCLUSIONS

We have presented the design choices and the structure of the 5G-INDUCE orchestration platform, which has been realized on the basis of the principle of separation of concerns between the vertical applications domain and the networking domain. The architecture of the orchestrators is oriented toward fostering network automation and zero-touch configuration, in line with B5G trends. Experimental results of one among the eight Use Cases have been also reported.

ACKNOWLEDGMENT

This work has been supported by the Horizon 2020 5GPPP Innovation Action 5G-INDUCE (Grant Agreement no. 101016941).

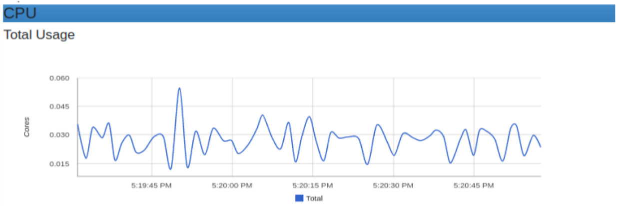

Fig. 5. UC3 MEC CPU usage.

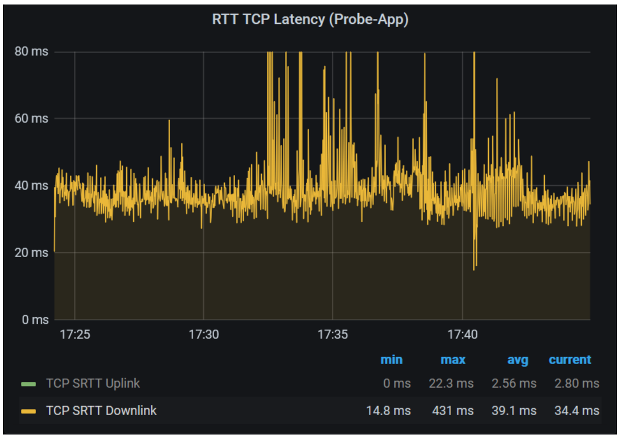

Fig. 6. Download TCP latency.

REFERENCES

[1] I. F. Akyildiz, A. Kak, S. Nie, "6G and Beyond: The Future of Wireless Communications Systems," IEEE Access, vol. 8, pp. 133995-134030, July 2020.

[2] 5G-INDUCE - Open Cooperative 5G Experimentation Platforms for the Industrial Sector nApps, https://www.5g-induce.eu/.

[3] P. Gouvas, A. Zafeiropoulos, E. Fotopoulou, T. Xirofotos, R. Bruschi, F. Davoli, A. Bravalheri, D. Simeonidou, "Separation of Concerns among Application and Network Services Orchestration in a 5G Ecosystem", *Presented at. EuCNC Workshop on "From Cloud ready to Cloud Native Transformation: What It Means and Why It Matters"*, Ljubljana, Slovenia, June 2018; [online] Available: https://zenodo.org/record/3699013/files/MATILDA%20WhitePaper%20-%20Approach.pdf.

[4] R. Bruschi, F. Davoli, C. Lombardo, J. F. Pajo, "Managing 5G Network Slicing and Edge Computing with the MATILDA Telecom Layer Platform", *Computer Networks*, vol. 194, pp. 1-14, July 2021.

[5] MongoDB, URL: www.mongodb.com/.

[6] Prometheus, URL: https://prometheus.io/.

[7] 5G-INDUCE, Deliverable D2.2, Targeted Use Cases and NetApp Related Requirements, Jan. 2022; [online] Available: https://www.5g-induce.eu/.