A Framework for Next Generation Cloud-Native SDN Cognitive Resource Orchestrator for IoTs (NG2CRO) *

Hafiza Kanwal Janjua[1[0000-0002-6655-6015]], Ignacio de Miguel[1[0000-0002-1084-1159]], Ramón J. Durán Barroso[1[0000-0003-1423-1646]], Maryam Masoumi[1[0000-0002-3832-9106]], Soheil Hosseini[1[0000-0003-3006-5297]], Juan Carlos Aguado[1[0000-0002-2495-0313]], Noemí Merayo[1[0000-0002-6920-0778]], and Evaristo J. Abril[1[0000-0003-4164-2467]]

Universidad de Valladolid (UVa), Valladolid, Spain
{hafizakanwal.janjua, ignacio.demiguel, ramon.duran, maryam.masoumi, soheil.hosseini, juancarlos.aguado.manzano, noemer, ejabril}@uva.es

Abstract. SDN (Software Define Networking) and NFV (Network Function Virtualization) are the key enablers for 5G systems and also open many doors in the cloud-native application. Besides, it invites new challenges to the efficiency and scalability of resource management. This work aims to provide a cognitive framework for 5G resource and service orchestration in a cloud-native SDN environment. The proposed NG2CRO framework resource orchestrator is designed to adapt the network's self-learning capabilities and dynamicity while taken on to account the network's Markovian properties and diverse service requirements. We consider incorporating AI (Artificial Intelligence) techniques specifically RL (Reinforcement Learning) methodologies because literature has shown that these techniques can efficiently address and comply with the current dynamic behaviors and heterogeneity of 5G services and applications. In conclusion, both benefits and liabilities are discussed of incorporating AI specifically RL into resource orchestration practices that provide us with future research challenges.

Keywords: Next Generation 5G Network · Artificial Intelligence · Network Automation · Software-Define Networking · Cloud-Native SDN · Resource Orchestration.

1 Introduction

SDN orchestration facilitates energy and resource optimization for the Internet of Things (IoT) industry. SDN allows efficient, cost-effective, and dynamic

* This work is part of the IoTalentum project, which 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).

scalability of resources in cloud and edge computing services [1, 12]. The evolution of the IoT industry has raised the demand for cloud data centers where tens of thousands of compute nodes are connected through thousands of networks, creating avenues for manageability and performance issues. Cloud data centers need dynamic provision computing and network resources to adapt to the fluctuating demands of customers [17].

With advents in software-defined clouds (SDC) that integrate SDN and cloud resource management where the programmable controller provides dynamic and autonomic configuration and management of the underlying resources; cognition of such scenario can predict the requirement and increase the efficiency in such scenarios [2, 10]. Despite the increasing popularity of studies for joint resource optimization in the cloud environment with SDN technology, the realization is still limited for developing integrated management platforms providing simultaneous controllability and energy optimization of computing and networking infrastructures. The testing cost and energy efficiency in real-time are required for migration to new technologies [4, 6].

Moreover, 5G networks are complex, large, and diverse with heterogeneous services which require high flexibility and efficiency to meet the SLA and QoS agreed upon by the services provider. However, due to the lack of flexibility in traditional 4G networks, when comes to the dynamics of 5G systems due to the heterogeneous services. Therefore, the current resource optimization techniques are not suitable for the current 5G infrastructure, which, is faced with challenges of SLA and QoS violation due to the incapability to handle the system dynamicity.

Hence, there is a need to have advanced cognitive control methods that can efficiently maneuver and orchestrate the dynamicity of the network demands. Consequently, studies extrapolated that RL-based cognitive approaches are beneficial to learn and exploring the dynamics of the network autonomously which is capable to support service flexibility and can provide an enhanced user experience in terms of SLA and QoS fulfillment [8].

1.1 Aims & Objectives

The aim of this project is an association with the IoTalentum project [7] specifically to the topic ESR7: "Cognitive orchestrator of MEC and network resources". Hence, the vision is to design a Cloud-Native SDN joint resource orchestration system for the network and service resources. Also, it should optimize the energy and cost profile via pivoting the ability of self-learning and self-management to guarantee performance in the 5G systems.

Moreover, the integration of Cloud and SDN platforms should leverage the framework functioning at the controller via intelligent network management, optimal resource utilization, load balancing, and service orchestration to provide IaaS (Infrastructure-as-a-Service) and SaaS (Software-as-a-Service) to users that will operate the MEC and the underlying IoT network of IoTalentum architecture.

We aim to leveraged the objectives of ESR7 by integrating the efficient, optimized, and proactive performance-aware Deep Reinforcement Learning techniques to develop the advanced decision-making pattern for energy monitoring, protection, and control mechanisms to deliver optimal SLA and QoS in IaaS and SaaS environments.

We designed some research questions with the consideration of 5G network resources orchestration using artificial intelligence specifically reinforcement learning and Machine learning along with the major problems in current network resource management. Table 1 provides an overview of the research questions in relation to the problems.

Table 1. Problem and Research Questions

Problem (P)	**Research Questions (RQ)**
P1: Dynamicity in traffic demands and users' mobility.	**RQ1:** How AI techniques has performed in literature orchestration in 5G Network?
P2: Dynamic systems render markovian properties due to heterogeneous network devices.	**RQ2:** How RL-based methods can be useful for resource orchestration to to address the dynamicity?
P3: Need for safe orchestration due to constraints of the physical system even in virtualized SDN NFV infrastructure	**RQ3:** What will be use of constraint-aware RL methods having an impact on SLA and QoS of the 5G slicing network scenario? **Safe Orchestration is Required!!!** which is termed as making the resource orchestration policies which should consider the physical system constraint such as Bandwidth, CPU, RAM and Memory

2 Background Studies

In the research area of 5G resource and service orchestration, we will specifically discuss the current state-of-the-art projects related to Cloud and SDN management and orchestration likewise IoTalentum.

2.1 Projects Related to 5G Resource Orchestration

Virtual Network Functions as a Service over virtualized infrastructures (T-NOVA) focuses on network functions as a service (NFaaS) [9]. This project provides a solution for deploying and managing NFaaS, which allows operators to create and deliver new services efficiently on virtualized infrastructure. 5GEx focuses on creating an ecosystem that allows the exchange of resources and services to and from different 5G experimental infrastructures [3]. The aim of 5GEx is to facilitate cross-domain experimentation and validate 5G technologies. Sonata [13] focuses on developing a service-oriented 5G architecture that supports the dynamic creation and management of network services, including smart cities, Industry 4.0, and virtual/augmented reality. Vital [16] is a 5G-related project that focuses on the development of an integrated framework for the design, deployment, and management of 5G networks and services. It addresses the challenges associated with the deployment of 5G networks, including network slicing, orchestration, and automation. 5G Transformer [11] aims to develop a flexible and scalable 5G architecture that can support the dynamic creation and management of network slices. The project's focus is on developing an SDN-based architecture that allows the efficient and flexible orchestration of 5G network resources. 5G Growth presents an ecosystem for the co-creation of 5G services and applications [14]. It provides a platform to allow collaboration between different stakeholders to develop and deploy new 5G services and applications. Inspire 5G Plus [15] project aims to develop an integrated platform to create, deploy, and manage 5G services and applications to enable the seamless integration of several 5G components and services, such as network slicing, edge computing, and AI. 5G Zorro aims to develop a security framework for 5G networks [5]. The project aims to address the security challenges associated with the deployment of 5G networks by proposing a security architecture that can ensure the confidentiality, integrity, and availability of 5G network resources.

Table 2. Projects related to 5G resource Orchestration

Project	Domain	Resource Orchestration Goals
T-NOVA [9]	Cloud, NFV, SDN	Enable end-to-end service provisioning via NFVaaS, VNFaaS, NFVIaaS
5GEx [3]	Cloud, NFV, SDN	Enable cross-domain orchestration of 5G services
Sonata [13]	Cloud, NFV SDN	Develop an integrated service platform
Vital [16]	Cloud, NFV, SDN, IoT	Develop a framework for IoT and smart city
5G Transformer [11]	Cloud, NFV, SDN	Develop SDN service-oriented 5G network architecture
5G Growth [14]	NFV	Develop 5G infrastructure and services for rural areas
Inspire 5G Plus [15]	NFV, SDN, DLTs, IoT	Develop 5G infrastructure and services for IIoT
5G Zorro [5]	DLTs	Develop an integrated platform for 5G services

3 Proposed NG2CRO Framework

We designed a cloud-native SDN-based framework addressing cognitive capabilities using RL methods. Also, the proposed architecture is designed to achieve an intelligent, autonomous, and self-managed network and resource orchestration system. The design of the proposed system is separated into four logical layers that are; the user plane (UP), the data plane (DP), and the control plane (CP) (further CP is subdivided into the following: management, orchestration, and provisioning), management plane (MP) and application plane (AP). On the top of figure 1 we described the use cases of NG2CRO. Also, we aimed to solve the problems related to network orchestration addressed in section **??** by separating each part of the network (NVF-I, VNF-O, and slice orchestration & management) by separating each module in master-slave architecture. Here in master-slave architecture, each part will do its resource management on its own but get the updates from the main controller, which can lower the delays and latency when operations are directly performed by the main controller.

The main aim of the four logical layers is to adapt the convergence of DP into CP and MP that leads towards the AP. The data plane performs data transfer to upper abstract layers in terms of infrastructure and configuration information. The control plane performs the following tasks that are, monitoring of network function virtualization management (NFVI-M), Virtual network function orchestration (VNF-O), VNF life cycle management, and Slice management. Resource provisioning and orchestration are the capabilities of CP, but in contrast, the convergence of CP operations must be adaptable to DP as the MEC environment is dynamic and heterogeneous. Moreover, to maintain service continuity the CP convergence to DP must be addressed in terms of infrastructure adaptability, service flexibility, dynamic updates, and resource provisioning.

Following, the process flow of NG2CRO is depicted in figure1 where the flow is labeled with numbers. At 1. The compatible application scenarios are discussed and also the application services related to the scenarios are given. 2. The enabling fronthaul and backhaul technologies are given for 5G applicability. 3. Next the information flows goes toward and controller in the control plane. Here the information flow will be in back fourth manner between the controller and applications. 4. Here we have the cloud-native SDN controller to support the cloud-native 5G applications.

Further on, the resource orchestration operations are carried out by the controller, here in step number 10 the policy optimization, and model aggregation of AI methods is performed. The data fed into the main brain is acquired from the modules which are VIM (5) and NFVO (7). At level 10 NG2CRO will make use of RL techniques to build the resource orchestration policies and send these policies to NFV and VNf-O orchestrator components. Next, it incorporates the TL (transfer learning) techniques also to build future predictions, but here we can face the problems with data privacy issues which is an area to address.

At level (9) bandwidth management is carried out among 5G slices, we have carried out the experimentation of this module here [8]. Our future work includes further experimentation of modules 7 and 6. Lastly, the main brain is a main controller that manages all of the operations based on the given states as traffic information and sends back the required action based on the current state and optimal policies.

The Proposed NG2CRO is the subpart of the IoTalentum [7] project and it aimed to solve the issue in current state-of-the-art resource orchestration projects by adhering to the standards used by existing projects. Furthermore, we plan to perform the experimentation of each module of NG2CRO with RL techniques and integrate all of them with the main controller with the aim of optimizing the RL methods compliant with network management.

3.1 Discussion

The platforms given in the table 2 have distinct goals and sets of capabilities, but they all seek to make it possible to manage and orchestrate SDN resources effectively in a scalable and flexible way. However, certain platforms could have restrictions or gaps in their SDN-specific functionality, or they might not support setups with many vendors. For instance, certain systems could lack sophisticated SDN-specific capabilities like effective flow management, network slicing, or network function chaining.

A platform's capacity to serve complex, multi-vendor setups may also be constrained, which can make integration and administration more difficult. It's crucial to assess the specific requirements of the network and consider the advantages and disadvantages of each platform before selecting the best SDN management and orchestration platform. we aimed to design a platform that tried to suit the needs of current and help to efficiently manage and orchestrate the SDN resources by carefully considering these criteria.

The integration of RL into SDN and NFV resource orchestration has the potential to have beneficial as well as detrimental impacts on the interoperability, flexibility, and scalability of the infrastructure. Also, we strive to integrate RL algorithms into our work. the initial part to integrate RL into this work is been published in [8].

On the plus side, adding RL to SDN and NFV resource orchestration can increase the infrastructure's automation and intelligence, which can further boost interoperability and adaptability. For instance, RL algorithms may be applied to forecast traffic patterns, optimize resource allocation, and detect abnormalities in real time, allowing for more effective use of network resources.

Additionally, RL can provide a dynamic response to shifting network settings, enhancing the infrastructure's adaptability. RL algorithms may make SDN and NFV resource orchestration more sensitive to changes in traffic patterns, network topology, and user behavior by continually learning and adjusting to new scenarios.

The integration of RL into SDN and NFV resource orchestration might hinder scalability and interoperability, which is a serious drawback. For instance, for RL algorithms to be effective, huge volumes of data must be collected and processed instantly. This can make scaling difficult, especially in large-scale networks where there is a lot of data to manage.

The lack of a generally adopted standard for RL integration into SDN and NFV resource orchestration might also cause compatibility and interoperability problems. This can make it challenging for network operators to combine various

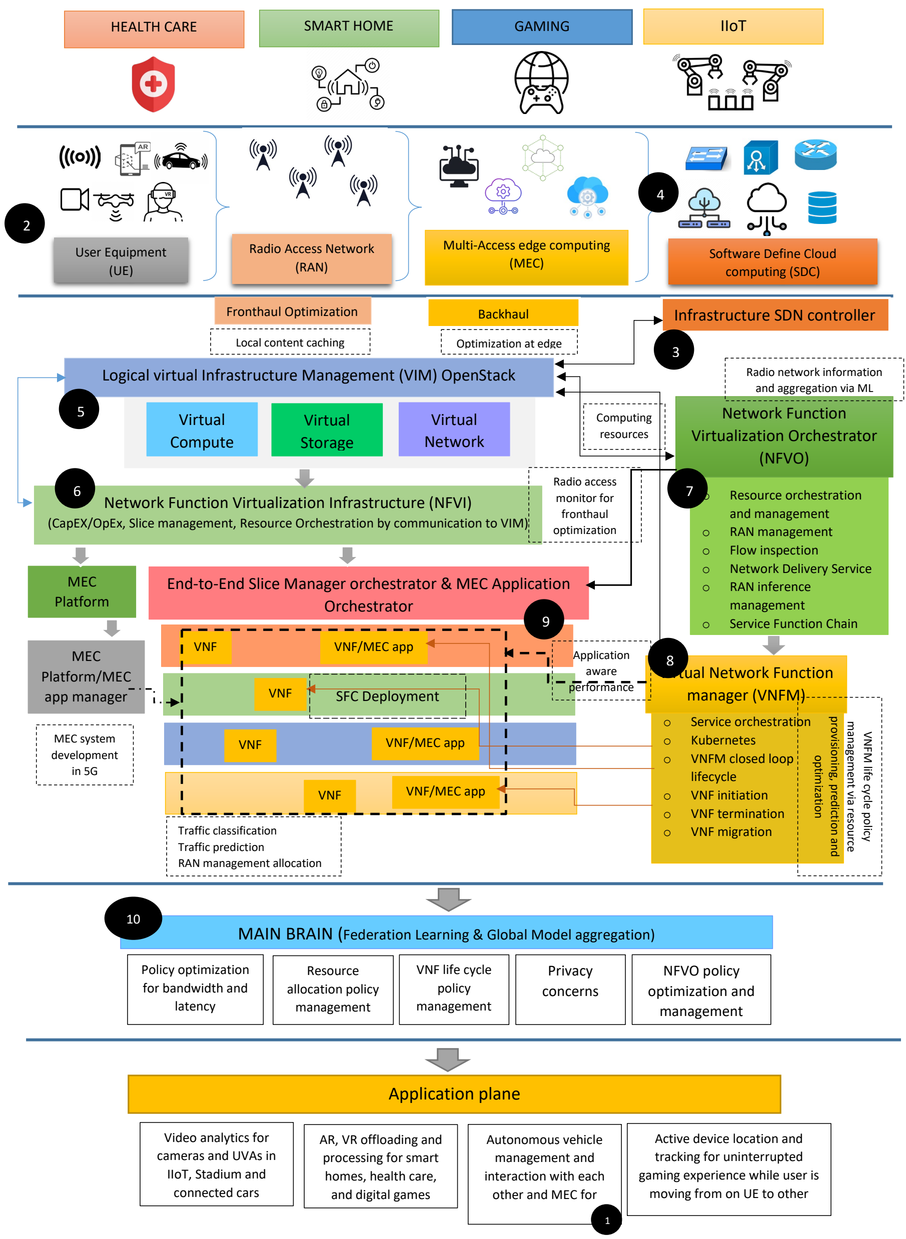

Fig. 1. Proposed NG2CRO Framework

systems and parts, which can affect the infrastructure's overall interoperability and adaptability.

Overall, integrating RL into SDN and NFV resource orchestration has the potential to have beneficial as well as adverse impacts on scalability, flexibility, and interoperability. It is crucial to carefully weigh the possible advantages and challenges of integrating RL into SDN and NFV resource orchestration as well as to build standards and best practices to guarantee interoperability and scalability.

3.2 Conclusions

In this work, we adress the the cognitive SDN resource orchestrator of IoTalentum architecture and proposed a framework named NG2CRO. We discuss the existing state-of-the-art projects related to 5G network resource management and orchestration. We discuss and analyze the existing listed projects that address network management using cognitive techniques across the cloud and SDN domains. Moreover, a brief discussion is given providing a curated list of benefits and limitations for an AI-enabled cloud-native SDN controller for the resource orchestrator. Such as RL techniques will be very useful in terms of improved efficiency and automation, real-time analytics and insights, and predictive maintenance. But also, this invites crucial points to consider which are privacy issues, AI explainability, Adversarial AI, and lack of standardization regarding RL resource orchestration frameworks. We aimed to address the complexity and cost-incurring problem of implementing RL techniques for resource orchestration and the proposed NG2CRO framework is the initial step to this path.

References

1. Bega, D., Gramaglia, M., Perez, R., Fiore, M., Banchs, A., Costa-Pérez, X.: Ai-based autonomous control, management, and orchestration in 5g: From standards to algorithms. IEEE Network **34**(6), 14–20 (2020). https://doi.org/10.1109/MNET.001.2000047
2. Benzaid, C., Taleb, T.: Ai-driven zero touch network and service management in 5g and beyond: Challenges and research directions. IEEE Network **34**(2), 186–194 (2020). https://doi.org/10.1109/MNET.001.1900252
3. Bernardos, C.J., Gerö, B.P., Di Girolamo, M., Kern, A., Martini, B., Vaishnavi, I.: 5gex: realising a europe-wide multi-domain framework for software-defined infrastructures. Transactions on Emerging Telecommunications Technologies **27**(9), 1271–1280 (2016)
4. Camps-Mur, D., Gavras, A., Ghoraishi, M., Hrasnica, H., Kaloxylos, A., Anastasopoulos, M., Tzanakaki, A., Srinivasan, G., Antevski, K., Baranda, J., Schepper, K., Casetti, C., Chiasserini, C., Garcia-Saavedra, A., Guimares, C., Kondepu, K., Li, X., Magoula, L., Malinverno, M., Cogalan, T.: Ai and ml – enablers for beyond 5g networks (05 2021). https://doi.org/10.5281/zenodo.4299895
5. Chainho, P., Fernandez-Fernández, A., Fernández Hidalgo, J., Gil Pérez, M., Jorquera Valero, J.M., Martínez Perez, G., Mifsud, J.M., Huertas Celdran, A., Lekidis, A., Sanchez Sánchez, P.M., et al.: Overview of the security and trust mechanisms in the 5gzorro project. EUCNC 2021 (2021)

6. Duan, Q.: Intelligent and autonomous management in cloud-native future networks—a survey on related standards from an architectural perspective. Future Internet **13**, 42 (02 2021). https://doi.org/10.3390/fi13020042
7. IoTalentum: Cognitive orchestrator of mec and network resources. iotalentum.eu/esrs (Last accessed 8 May 2023)
8. Janjua, H.K., de Miguel, I., Durán Barroso, R.J., de Dios, G., Aguado, J.C., Álvarez, N.M., Fernández, P., Lorenzo, R.M.: Efficient optimization of actor-critic learning for constrained resource orchestration in ran with network slicing. In: 2023 26th Conference on Innovation in Clouds, Internet and Networks and Workshops (ICIN). pp. 100–104 (2023). https://doi.org/10.1109/ICIN56760.2023.10073489
9. Kourtis, M.A., McGrath, M.J., Gardikis, G., Xilouris, G., Riccobene, V., Papadimitriou, P., Trouva, E., Liberati, F., Trubian, M., Batallé, J., et al.: T-nova: An open-source mano stack for nfv infrastructures. IEEE Transactions on Network and Service Management **14**(3), 586–602 (2017)
10. Latah, M., Toker, L.: Artificial intelligence enabled software defined networking: A comprehensive overview. IET Networks **8**, 79–99 (2018)
11. Mangues-Bafalluy, J., Baranda, J., Pascual, I., Martínez, R., Vettori, L., Landi, G., Zurita, A., Salama, D., Antevski, K., Martín-Pérez, J., et al.: 5g-transformer service orchestrator: design, implementation, and evaluation. In: 2019 European Conference on Networks and Communications (EuCNC). pp. 31–36. IEEE (2019)
12. Manias, D.M., Shami, A.: The need for advanced intelligence in nfv management and orchestration. IEEE Network **35**(1), 365–371 (2021). https://doi.org/10.1109/MNET.011.2000373
13. Martrat, J.: Sonata approach towards devops in 5g networks. In: SDN World Congress (2017)
14. Nguyen, T.H., Yoo, M.: A vnf descriptor generator for tacker-based nfv management and orchestration. In: 2018 International Conference on Information and Communication Technology Convergence (ICTC). pp. 260–262. IEEE (2018)
15. Ortiz, J., Sanchez-Iborra, R., Bernabe, J.B., Skarmeta, A., Benzaid, C., Taleb, T., Alemany, P., Muñoz, R., Vilalta, R., Gaber, C., et al.: Inspire-5gplus: Intelligent security and pervasive trust for 5g and beyond networks. In: Proceedings of the 15th International Conference on Availability, Reliability and Security. pp. 1–10 (2020)
16. Trichias, K., Landi, G., Seder, E., Marquez-Barja, J., Frizzell, R., Iordache, M., Demestichas, P.: Vital-5g: innovative network applications (netapps) support over 5g connectivity for the transport & logistics vertical. In: 2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit). pp. 437–442. IEEE (2021)
17. Zhang, S., Zhu, D.: Towards artificial intelligence enabled 6g: State of the art, challenges, and opportunities. Computer Networks **183**, 107556 (12 2020). https://doi.org/10.1016/j.comnet.2020.107556