The role of Network Digital Twinning in a 6G-oriented Optimization-as-a-Service Platform

J.-M. Martinez-Caro(1), J. Pérez-Romero(2), R. Agüero(3), A. Muñiz (4), E. Atxutegi (5), L. Diez(3), L.-M. Contreras(4), I. González(5), F.-J. Moreno-Muro(1), P. Pavon-Marino(1), O. Sallent(2)
jmmartinez@e-lighthouse.com, jordi.perez-romero@upc.edu, ramon@tlmat.unican.es, alejandro.muniz@telefonica.com, eneko.atxutegi@nemergent-solutions.com, ldiez@tlmat.unican.es, luismiguel.contrerasmurillo@telefonica.com ivan.gonzalez@nemergent-solutions.com, jmoreno@e-lighthouse.com, ppavon@e-lighthouse.com, sallent@tsc.upc.edu
(1) E-lighthouse Network Solutions, Cartagena, Spain.
(2) Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
(3) Universidad de Cantabria, Santander, Spain
(4) Telefónica I+D, Madrid, Spain
(5) Nemergent Solutions, Bilbao, Spain

***Abstract*- The advent of 5G has meant providing users with higher data rates, more connected devices and lower latency, implying the need to redefine traditional network management methods. In addition, 6G is expected to provide much more challenging and restrictive performance, enhancing the idea of redefining network management methodologies, aimed at maximising the use of network resources. Network Digital Twinning (NDT) and Optimization-as-a-Service (OaaS) are two key mechanisms, along with the use of artificial intelligence (AI) and machine learning (ML), to assist in network management procedures. On the one hand, the NDT is a digital replica of the physical network, which permits evaluating the impact of concrete configuration settings prior to its real implementation in the system, helping to optimize the overall network performance. On the other hand, OaaS offers ad-hoc optimized recommendations to third-party applications based on the execution of several optimization algorithms. In this work, we propose an open, flexible and scalable OaaS platform, defined in the OPTIMAIX project, to exploit the NDT potential targeted to support automated planning and optimization of networks. Also, an evaluation of the NDT role within this architecture is provided.**

I. Introduction

Last decades, the proliferation of Internet, its easy accessibility from a mobile device, and new communication technologies have caused remarkable changes in the real world about how we communicate with others, share information, or buy products, among others. In 2000, only 6.5% of the world's population had access to Internet, growing to 66% by 2023 [1]. The use of devices with limited computing capabilities and the increase of existing applications and platforms (e.g., online stores, multimedia services, social networks, etc.) that make use of Cloud Computing architectures, results in the increase of traffic experienced by the network operators [2].

The envisioning of the next network generation (6G) predicts a considerable increase in mobile data traffic. While mobile data consumption reached 62 EB per month in 2020, it is expected to reach up to 5016 EB per month in 2030. Such advanced 6G services will necessarily leverage on existing brownfield networks complemented with novel technologies and supportive tools [3]. The expected traffic increase in the coming years envision new applications for optimization-based tools in order to provide an efficient usage of network resources to cope with the challenging traffic demands. Under this umbrella the concept of Optimisation-as-a-Service (OaaS) allows third-party applications to access to optimization and planning functionalities that enable efficient network service and resource allocation through the use of Artificial Intelligence (AI) and Machine Learning (ML) techniques [4].

The application of OaaS defines, by using specific algorithms, optimized recommendations to be later translated in concrete actions in the network to improve its performance. Novel techniques allows, in advance, to test these actions on a highly accurate digital replica of the physical network, integrating topological information with historical records of the same. This digital replica, called Network Digital Twin (NDT), can be a key element in network management for decision making by operators achieving simple, automatic, resilient, and full life-cycle operation and maintenance. NDT is key in the planning of next-generation technologies (e.g., Industry 4.0) to accelerate technological innovation, reduce costs, and run "what-if" where the trial-and-error nature of these interactions can lead to performance degradation or even to unsafe states. Therefore, the NDT allows an ML/AI-based optimisation algorithm to be trained using data from the NDT and/or observing the impact of the decisions made on the NDT rather than on the real network [5].

The combination of the concept OaaS and NDT in a common architecture (OaaS platform) can became a powerful tool gaining deeper insights into network performance and optimising resource utilization and traffic flows in dynamic environments where network conditions can change rapidly. These two elements are key within the architecture defined in the coordinated project OPTIMAIX [6].

The OaaS platform provides a set of tools for assisting in a number of use cases classified as i) application timeline, ii) network segment application, and iii) service type. The network carriers typically conduct periodically (e.g., twice a year) capacity planning and network dimension tasks using intensive and potentially slow algorithms, and capacity-on-demand activities with dynamic and fast algorithms that react in near real-time. These actions calculate the Key Performance Indicators (KPI) of the network configuration or deployments to evaluate them from a performance perspective. The existing network infrastructure is splitted into different segments (e.g., fixed-access, radio-access, metro-aggregation, core, etc.) that are commonly managed by diverse network operators, which face particular technical needs. Lastly, the carrier services aim to maximize the telco profit with different services into the same data flow using techniques such as traffic aggregation, which combine traffic from multiple users instead of isolated user-level traffic.

The heterogeneity of use cases for an OaaS system, demands for i) a flexible and extensible architecture, which permits integrating multiple optimization algorithms, with arbitrary inputs and outputs, computational requisites, optimization targets, and application scopes; ii) a scalable scheme, applicable from small to carrier-size networks; and iii) an open system, based on the OpenAPI standard, that permits the integration of contributions (e.g., optimization algorithms, or computation resources), coming from third-parties, which can retain the authorship and economical profit (if any) of their contributions to the system.

The rest of this article is organized as follows: Section II, introduces the main features of the OaaS-NDT architecture. Section III analyses the design, APIs and schema of the NDT exposing a typical workflow, and Section IV concludes this article by painting a picture about the relevance of NDT in actual networks planning and optimization activities.

II. The OaaS Platform Architecture

Fig. 1 illustrates an overview of the proposed 6G-oriented OPTIMAIX architecture where the OaaS platform is presented. On the one hand, the left side represents a particular example of the target network of the OaaS platform. This network infrastructure consists of a Radio Access Network (RAN) composed of several Points of Presence (PoPs), that include different resources (e.g., radio units, transport nodes, computational and storage resources, etc.) interconnected with transport links. Those PoPs handle many cells that provide radio coverage to the User Equipment (UE) in specific geographical areas. The upper-left part represents different services to be allocated in the network infrastructure, consisting in this example of network slices, with their corresponding Service-Level-Agreements (SLAs). Both the network infrastructure and the supported services are described according to defined network models such as 3GPP TS 28.541 for radio access network [7], IETF RFC 8344 for network topologies [8], IETF RFC 8345 for IP management [9], and 3GPP TS 23.501 for network slicing [10].

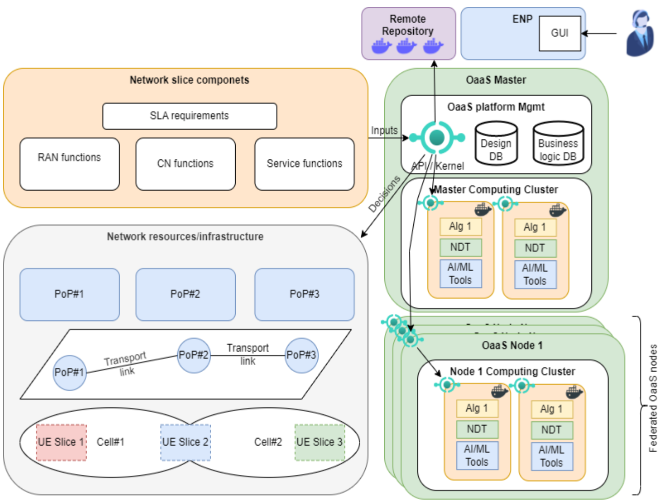

Fig. 1. Overview of the OPTIMAIX OaaS architecture

On the other hand, the right side shows an overall perspective of the OaaS platform architecture. It is conceived as a hybrid system where one logically centralized entity (OaaS Master) and multiple federated (OaaS Nodes) ones coexist. This solution has the advantage to cover two different profiles of third-party institutions/users: i) small scale contributors may be interested in integrating a small number in the OaaS Master; and ii) special/large scale contributors, willing to have full control not only on the algorithms/reports, but also on the computing facilities where they will be run, and how the containers are stored and secured. The OaaS Master includes a key module for management (OaaS platform Mgmt) that consist of the platform kernel with the API server and two databases, the Design DB which stores network information (e.g., SLA requirements, topology components, etc.), and Business logic DB which stores platform metadata (e.g., where the algorithm/NDT is deployed, paths, etc.). The API server presents the API description via REST APIs with messages based on a human-readable format as the JavaScript Object Notation (JSON) format, pervasively used in these contexts. Lightweight implementations, robustness, fast processing, and easiness of documentation and testing process are their main advantages. The initial approach of the OaaS architecture describes four key APIs: i) design handling; ii) control the federated OaaS Nodes; iii) repository management; and iv) functionality management for algorithms and NDTs.

To optimize algorithms and NDT simulations, precise network and infrastructure information is required in near real-time. This information is available in the Design DB using the previously presented data models. With the Design handling API, new designs can be created, a list of designs can be returned, and existing designs can be retrieved, updated, or deleted by ID.

The OaaS platform design supports both centralized and federated OaaS platforms, and the API allows for the deployment and enrolment of new OaaS Nodes, update of OaaS Node details, and deletion of existing ones. Each OaaS entity (Master or Node) includes a computing cluster capable of instantiating Docker containers with isolated or combined algorithms and/or NDTs. The Docker API performs basic operations to successfully instantiate a Docker image, retrieve and update container details, or delete a container instance.

The enrolled Docker images implement a set of functionalities, such as algorithms, NDTs, or both. Once a Docker container is instantiated, it exposes an API to list available functionalities in the design. Additional paths return information for a specific resource (algorithms and NDTs) and start/stop the execution of the running workflow.

Computing Clusters are present in the OaaS platform architecture to accommodate the Docker instances, both in OaaS Master and OaaS Nodes, including a variable number of algorithms and NDTs with AI/ML assisting tools. These resources are clustered in repositories and will be managed as microservices instantiated via docker-compose. The included algorithms will address network problems based on input parameters such as SLA, network slice components, network segments, current load and traffic in different parts of the network, etc. The solution of these algorithms generates specific actions to be applied into network devices. These algorithms require a training stage to accurately adjust the existing models and enough network observation times to acquire all the needed data. These algorithms solve optimisation problems for different use cases such as end-user use case (e.g., mission critical services), dimensioning use cases (e.g., RAN placement and functional split configuration), and operational use cases (e.g., capacity sharing for RAN slicing).

III. Network Digital Twin

The incorporation of NDT provides greater flexibility and dynamism to existing networks by forecasting their behaviour and improving their efficiency. The successful implementation of an NDT requires several key elements: i) an accurate model of the physical network to enable accurate simulation and prediction of its behaviour; ii) real-time data from the physical network to update the virtual model; iii) scalability to handle large-scale systems with numerous devices, for instance like in an Internet of Things (IoT) environment; and iv) interoperability to integrate data from diverse sources. In order to design an NDT, it is crucial to establish a set of necessary procedures, such as component development, container packaging, building, deployment, and testing. This is necessary to facilitate simulations of the virtual network for pre-defined use cases. These use cases may encompass the whole network or certain segments; thus, the NDT must be tailored to that specific division.

The process of implementing an NDT begins with the development of the system. This includes the use of polyglot programming, utilizing various programming languages and frameworks to achieve greater flexibility and innovation. The NDT module (Fig. 2) is designed with a standardized interface to ensure interoperability with the rest of the OaaS platform and other NDT instances. The mapping sub-module creates a virtual network representation using input parameters and establishes relationships between entities in the NDT through either one-to-one or one-to-many synchronization, allowing for simulations of different network segments. Also, the core of the NDT is responsible for management functions, utilizing AI/ML techniques for simulations and diagnostics of network resources. Finally, the NDT estimates the pre-defined KPIs based on the targeted network objectives, for instance, including availability, performance (throughput, latency, etc.), security, efficiency (energy consumption, resource utilization, etc.), and Quality of User Experience (QoE). If the KPIs of the proposed solution in the NDT simulation exhibit an improvement over the KPIs of the existing physical network configuration, then the new configuration is tagged as recommendation to be applied into the real physical network. The adopted solution is then rigorously overseen through various tests and simulations to ensure that it is functioning correctly.

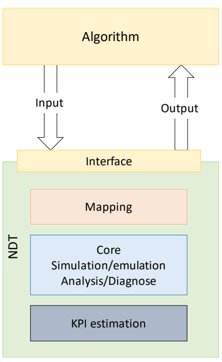

Fig. 2. Key modules of an NDT

Ensuring reliability and minimizing errors is crucial for the successful integration of a module. Therefore, following the DevOps and Continuous Integration (CI) principles, before being built into a Docker container in a repository, the module functionality is thoroughly tested. The packaging process involves building the component with necessary dependencies and configuration settings to create a Docker image. This image can be uploaded to a remote repository to be later instantiated in any computing cluster within the OaaS platform. After being instantiated, the NDT exposes an API to handle external requests.

A. *NDT Workflows*

NDTs are key components of the OaaS platform that enable the simulation or emulation of network conditions specified by the network operator. The main objective of this development is to encourage the creation of highly dynamic modules that allow for almost any type of interaction. Fig. 3 illustrates two basic workflows within the OPTIMAIX OaaS architecture showing two different implementation schemas. In Fig. 3a), the standalone schema assumes the instantiation of the entire OaaS platform system in one single server, only playing the role of the OaaS Master. Thus, the OaaS Mgmt orchestrates the entire workflow in one single entity. First, it calls the API of the instance executing the algorithm, sends all the necessary information for its performance, and validates the solution in an NDT. The calculated KPIs are then sent to the OaaS Mgmt module to determine whether it is necessary to apply the proposed changes in the previous solution, and then completing the working loop. In contrast, Fig. 3b) introduces complexity and more entities are assumed, creating several OaaS nodes to accommodate, in a federated way, the instances of the algorithms and NDTs by consuming the available computational resources of the nodes. Specifically, after executing two algorithms on a federated OaaS Node, the adopted solution is validated by using an NDT located in the OaaS Master. It is worth mentioning that the modules utilized, including the algorithms and NDTs, may be developed by the same partner(s) or could be sourced from a third-party module available on the OaaS platform.

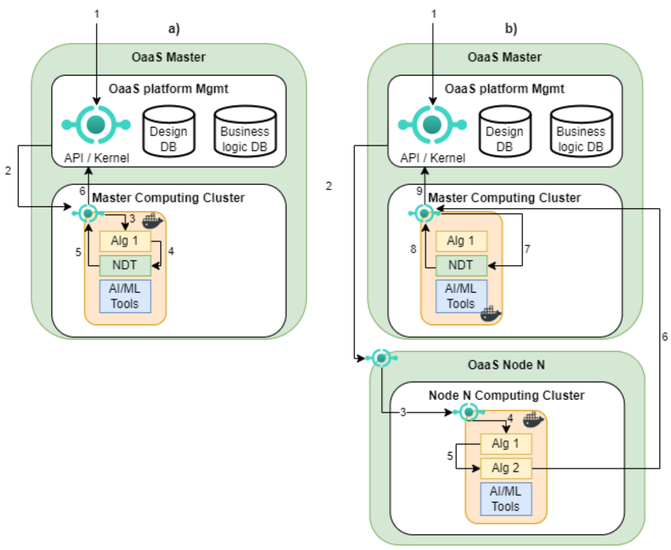

Fig. 3. Examples of OaaS Platform implementations: a) centralized and standalone; b) federated and multi-node.

B. *Applicability example*

To illustrate an example of applicability of the NDT, we consider an algorithm to address the capacity sharing optimisation in a sliced RAN. The algorithm is in charge of deciding the amount of capacity allocated to each RAN slice in each cell in accordance with the traffic demands of the involved slices and their requirements. The specific solution, as detailed in [11] is based on Reinforcement Learning (RL). Thus, the RL model needs to be trained to learn the optimum policy that selects the best action (capacity share assigned per slice and cell) for each possible situation (i.e. state). The training process consists in iteratively getting the state of the network, deciding an action based on the current policy (balancing between the exploration of new actions and the exploitation of the policy) and then getting a reward that is used to progressively improve the policy.

An NDT is used to train the algorithm instead of performing the training on the real network. In this way, it is avoided that the selection of wrong actions during the training impacts negatively on the performance of the real network. Following the structure of Fig. 2, the capacity sharing optimisation algorithm will include an RL agent that will send as input to the NDT an action consisting of the allocated capacities per RAN slice in each cell. Then, the NDT will emulate the operation of the network with this configuration and will estimate the KPI metrics needed to compute the reward, which will be sent back as output to the RL agent in conjunction with the metrics characterising the state to be used for making subsequent decisions.

The core module of the NDT represented in Fig. 2 will include a set of models for simulation/emulation that should represent the different elements and functionalities of the real RAN. A modular design is devised to improve the programmability and the agility of operation and deployment of the NDT instance to suit different use cases. The mathematical/statistical models may be fed by data collected from the real RAN to update their parameters according to the real RAN behaviour. The set of considered models include: i) scenario topology with detailed maps with buildings, streets, etc.; ii) gNB model with the configuration parameters of the base stations (e.g. position, frequency, bandwidth, etc.); iii) User Equipment (UE) model that characterises parameters of the mobile terminal such as the height, the position in accordance with a certain mobility model and the traffic generation model in accordance with a certain temporal and spatial traffic distribution; iv) channel model that characterises the propagation loss, interference and noise for the different links between UEs and gNBs; v) network slice model that characterises the parameters of the slice such as the Service Level Agreement to be guaranteed; and vi) RAN management and optimization models used to enforce the per-slice capacity allocated by the RL agent and the resource allocation that makes the corresponding allocation of resource blocks to UEs following the per-slice capacity.

IV. CONCLUSIONS

The continuous increase in network traffic and its complexity due to the proliferation of new applications and services has driven the need for network optimization. The combination of OaaS and NDT novel concepts using AI/ML techniques can achieve efficient network services and resource allocation. The proposed OaaS platform architecture integrates the OaaS system and NDTs, providing a comprehensive range of tools to assist in various use cases. The NDT is a digital replication of the physical network which facilitates testing operations such as what-if. This capability serves to curtail the occurrence of physical network downtime, ultimately leading to heightened levels of efficiency and decision-making. The presented architecture demands a flexible, scalable, and open system to permit integration of multiple optimization algorithms with arbitrary inputs and outputs, computational requisites, optimization targets, and application scopes.

ACKNOWLEDGEMENTS

This work is funded by the Spanish Ministry of Economic Affairs and Digital Transformation and the European Union - NextGenerationEU under projects OPTIMAIX_OaaS (Ref. TSI-063000-2021-34) and OPTIMAIX_NDT (Ref. TSI-063000-2021-35).

REFERENCES

[1] "ITU-T Facts and Figures 2022," International Telecommunication Union, 2022. [Online]. Available: https://www.itu.int/itu-d/reports/statistics/facts-figures-2022/. [Accessed 31 03 2023].

[2] Hung V. Dang, Mallik Tatipamula and Huan X. Nguyen, "Cloud-Based Digital Twinning for Structural Health Monitoring Using Deep Learning," *IEEE Transactions on Industrial Informatics,* vol. 18, no. 6, pp. 3820 - 3830, 2022.

[3] W. Jiang, B. Han, M. A. Habibi and H. D. Schotten, "The Road Towards 6G: A Comprehensive Survey," *IEEE Open Journal of the Communications Society,* vol. 2, pp. 334 - 366, 2021.

[4] M. Garrich, C. S. N. Martínez, F.-J. M. Muro, M.-V. B. Delgado and P. P. Mariño, "Network Optimization as a Service with Net2Plan," in *European Conference on Networks and Communications (EuCNC),* Valencia, 2019.

[5] X. Yang, C. Zhou, X. Duan, D. Lopez, A. Pastor, Q. Wu, M. Boucadair and C. Jacquenet, "Digital Wtin Network: Concepts and Reference Architecture (draft-irtf-nmrg-network-digital-twin-arch-02)," Internet Engineering Task Force (IETF), 24 10 2022. [Online]. Available: https://www.ietf.org/id/draft-irtf-nmrg-network-digital-twin-arch-02.html. [Accessed 31 03 2023].

[6] "OPTIMAIX project," [Online]. Available: https://optimaix.upc.edu. [Accessed 31 03 2023].

[7] "Management and orchestration; 5G Network Resource Model (NRM); Stage 2 and Stage 3 (Release 18)," 3GPP TS 28541 v18.1.2, September 2022. [Online]. Available: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3400. [Accessed 31 03 2023].

[8] M. Bjorklund, "A YANG Data Model for IP Management," Internet Engineering Task Force (IETF), 03 2018. [Online]. Available: https://datatracker.ietf.org/doc/rfc8344/. [Accessed 31 03 2023].

[9] A. Clemm, J. Medved, R. Varga, N. Bahadur, H. Ananthakrishnan and X. Liu, "A YANG Data Model for Network Topologies," Internet Engineering Task Force (IETF), 03 2018. [Online]. Available: https://datatracker.ietf.org/doc/rfc8345. [Accessed 31 03 2023].

[10] "System architecture for the 5G System (5GS); Stage 2 (Release 17)," 3GPP TS 23 501 V17.6.0., September 2022. [Online]. [Accessed 31 03 2023].

[11] I. Vilà, J. Pérez-Romero, O. Sallent and A. Umbert, ""A Multi-Agent Reinforcement Learning Approach for Capacity Sharing in Multi-tenant Scenarios"," *IEEE Transactions on Vehicular Technology,* vol. 70, no. 9, pp. 9450-9465, 2021.