MANTRA-6G: Management and Orchestration for Cloud- and AI-Native 6G Networks

Marios Avgeris*, Nitinder Mohan†, Gergely Pongracz‡, Chrysa Papagianni*, and Fernando A. Kuipers†
*Informatics Institute, University of Amsterdam, Amsterdam, the Netherlands
†Department of Software Technology, Delft University of Technology, Delft, the Netherlands
‡Ericsson Research, Budapest, Hungary
(m.avgeris, c.papagianni)@uva.nl, (n.mohan, f.a.kuipers)@tudelft.nl, gergely.pongracz@ericsson.com

***Abstract*—Realizing the envisioned capabilities of 6G requires flexibility, programmability, and intelligent automation. To support zero-touch 6G network service management, we propose MANTRA-6G, an integrated cloud-native, AI-native orchestration and management framework that is designed for seamless, robust, scalable, and intelligent service deployment and management across heterogeneous data planes.**

I. Introduction

The evolution towards 6G networks comes with ambitious goals including hyper-reliable, low latency, massive communication, and ubiquitous intelligence; thus, existing architectures need to adapt and scale efficiently and with minimal human intervention. These challenges are imposed by the complex requirements of emerging 6G applications –including industrial automation, augmented reality, intelligent transportation systems, and integrated non-terrestrial networks– which require advanced solutions for autonomous network service management and orchestration [1]. To address these challenges, the Dutch government has launched the national 6G R&D program Future Network Services[1], bringing together over 60 partners, including ICT companies, mobile operators, semiconductor manufacturers, and research institutions. Within this program, we propose a 6G network service management and orchestration framework that extends the hierarchical orchestration and management capabilities of Oakestra [2] with an enhanced version of the DESIRE6G Infrastructure Management Layer (IML) [3] serving as an Accelerator Abstraction Layer, following the O-RAN architecture [4]. The result is a cloud and AI-native framework, called MANTRA-6G (MANagemenT & oRchestration for Autonomous 6G), that enables flexible, scalable service deployment and intelligent automation. MANTRA-6G targets end-to-end (E2E) management and orchestration, from the Radio Access Network (RAN), to the core, effectively addressing the need for a common Service Management and Orchestration (SMO) platform as indicated by the nextG Research Group within the O-RAN alliance [5].

[1]https://futurenetworkservices.nl

II. MANTRA-6G Architecture

MANTRA-6G follows a multi-layer hierarchical structure (fig. 1). At the bottom layer sits the physical infrastructure, comprising geographically distributed sites, equipped with diverse compute and networking resources. These include programmable accelerators (e.g., IPUs, DPUs, etc.) alongside non-programmable hardware, such as legacy radio equipment or general-purpose computing platforms. In the following, we introduce the different layers of our framework.

1) Root Orchestrator: The cloud-native *Root Orchestrator* (RO) serves as the E2E SMO layer across the RAN and core network domains. Building upon traditional NFV orchestration, the RO handles lifecycle management of network services, network slicing, and resource coordination across distributed infrastructure sites. With AI at its core, the RO integrates a Non-Real-Time Intelligent Controller (Non-RT IC) supporting MLOps, i.e., the deployment of ML/AI pipelines to optimize and automate network operations. In this framework, network services are defined as logical service graphs composed of interconnected NFs, annotated with deployment constraints, SLA parameters, configuration dependencies, and end-to-end network requirements. To mitigate the complexity and security risks introduced by programmable data planes, the RO incorporates Network Program Synthesis capabilities, to dynamically auto-generate and configure network software tailored specifically to the tenant's or operator's requirements.

Furthermore, it supports the distributed nature of 6G infrastructures through a 6G Peering component that orchestrates secure, dedicated data-plane connectivity between local facilities and remote infrastructure. This extends the telco's geographical footprint beyond its traditional boundaries, enabling seamless interconnection with third-party networks (e.g., PN-NPN), cloud providers, and edge computing sites, while maintaining security and performance guarantees.

2) Infrastructure Management Layer: Each site hosts the cloud-native IML, serving as the abstraction and automation layer that bridges logical and physical network functions. The IML specifically targets programmable data-plane (PDP) resources (e.g., SmartNICs, DPUs, etc.) abstracting physical resource management, enabling seamless deployment, scaling, and aggregation/disaggregation of network functions, following a cloud-native approach for the networking elements. For example, it enables the dynamic deployment of specialized infrastructure network functions (*InfraNFs*), such as load balancers, tunneling adapters, and monitoring components, and robustly manages data-plane resources towards service assurance. Data-plane programmability allows the IML to leverage real-time in-network AI/ML techniques to assist its AI-native functionality. Building upon the capabilities of the current IML version, MANTRA-6G enhances the IML with inherent intelligence (e.g., for NF placement, offloading decisions etc.) to promote its AI-native operation, investigating performance trade-offs. To support data-plane multi-tenancy (i.e., aggregation/disaggregation of functions), the IML incorporates formal verification methods, ensuring the correctness/completeness of the respective program.

3) Cluster Orchestrator: The IML serves as a wrapper to the *Cluster Orchestrator* (CO), that handles the deployment, lifecycle management, and optimization of non-PDP resources at the local site level, such as CPUs and GPUs. It employs virtualization technologies such as containers, VMs, and unikernels, extending the respective Oakestra component with additional runtime environments (e.g., WebAssembly). It supports optimized resource allocation and scheduling of the workloads in the non-PDP hardware, gathers real-time resource and monitoring information, and reports aggregated metrics to the RO, thereby maintaining granular administrative control over local site resources. In conjunction with the IML the CO will autonomously establish load balancing, routing policies, and packet tunneling mechanisms to optimize NF-App communication across the infrastructure.

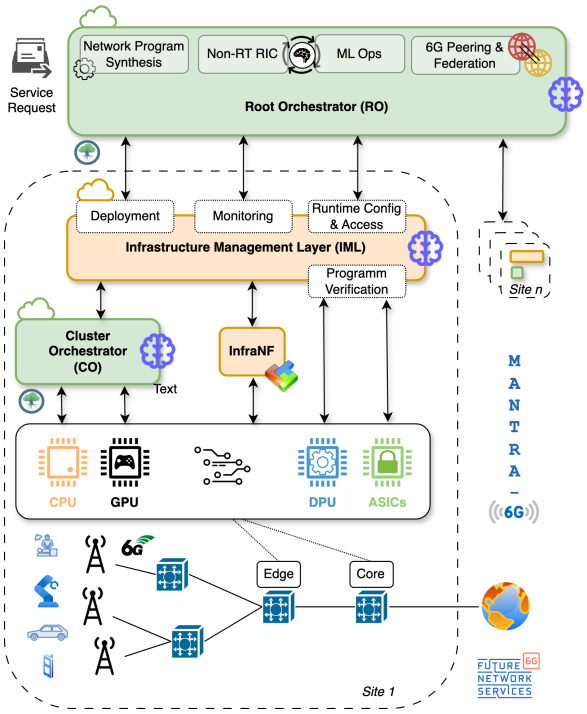

Fig. 1: MANTRA-6G architecture overview.

III. Future Directions

We plan to showcase the capabilities of this framework; this will accelerate the adoption of 6G technologies and pave the way for new business opportunities by illustrating the viability and benefits of advanced network management solutions in 6G-related market scenarios.

References

[1] C.-X. Wang *et al.*, "On the road to 6G: Visions, requirements, key technologies, and testbeds," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 905–974, 2023.

[2] G. Bartolomeo *et al.*, "Oakestra: A lightweight hierarchical orchestration framework for edge computing," in *USENIX ATC 23*, 2023, pp. 215–231.

[3] G. Pongrácz *et al.*, "D2.2: DESIRE6G Functional Architecture Definition," Jul. 2024. [Online]. Available: https://doi.org/10.5281/zenodo.12784579

[4] M. Polese *et al.*, "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1376–1411, 2023.

[5] O-RAN Alliance, "O-RAN Towards 6G," January 2023, accessed: 2025-03-14. [Online]. Available: https://mediastorage.o-ran.org/ngrg-rr/nGRG-RR-2023-01-O-RAN-Towards-6G-v1_3.pdf

Acknowledgement

This research was supported by the NGF Dutch 6G flagship "Future Network Services" and the JU SNS DESIRE6G (grant no. 101095890) projects.