Politecnico di Torino
ScuDo
Scuola di Dottorato – Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Design and Orchestration of Machine Learning-Driven Network Services in O-RAN Architecture

Federico Mungari

In recent decades, mobile networks have played a pivotal role in the transition of modern society towards an interconnected, data-centric, and automated world. Innovative, data-intensive and mission critical applications challenging the traditional mobile network architecture have emerged across various industries. To ensure a seamless expansion into new markets or services, it is essential to open the Radio Access Network (RAN). Open RAN deployments are grounded on disaggregation, virtualization, and software-based RAN components connected through open and standardized interfaces. By embracing Open RAN, the O-RAN alliance develops the O-RAN architecture, which exposes data and analytics, unlocking the integration of intelligent, data-driven closed-loop control within the RAN. In this direction, machine learning (ML)-based near-real-time applications and network services are vital to optimally enhance RAN operations. In this work, we explore the application of the Open RAN paradigm and its fundamental concepts to advance the current state of designing and orchestrating ML-driven near-real-time RAN services. We undertook a comprehensive exploration of the entire spectrum in O-RAN services deployment, spanning from design and development to management of O-RAN services, with a particular focus on ML-based ones.

The first central contribution of this work is the design of a near-real-time application, CAREM, a reinforcement learning framework for dynamic radio resource allocation in heterogeneous vRANs. CAREM optimally selects the best available link and transmission parameters for packet transfer, meeting the KPI requirements. CAREM enables efficient radio resource allocation under different settings and traffic demands, as experimental results derived throguh a proof-of-concept testbed demonstrates. When compared to the closest existing scheme based on neural networks and the standard LTE, CAREM exhibits an improvement of one order of magnitude in packet loss and latency. Furthermore, it provides a 65% latency improvement relative to the contextual bandit approach.

The second crucial aspect of our work involves understanding and profiling the resource requirements of near-real-time applications. We deployed an Edge-based DT solution for remote control of robotic arms in an experimental testbed. This solution, in compliance with the NFV paradigm, has been segmented into virtual network functions VNFs. Our research evaluates the entanglement among overall service performance, VNFs resource requirements, and the number of robots consuming the service. Experimental profiles highlight critical DT features, such as the novel trade-off between computing and time resources requirements and trajectory guarantees based on the industrial processes and the abstraction level of robot commands. These results offer crucial insights for designing network service scaling and resource orchestration frameworks dealing with RAN.

Along these lines, a significant contribution of this document is the introduction of an innovative orchestration framework designed to operate within the non-real-time RIC. We presented OREO, an O-RAN xApp orchestrator, engineered to maximize the offered services. OREO's key idea is to enable services to share xApps when they correspond to semantically equivalent functions, and the xApp output meets the service requirements. Leveraging a multi-layer graph model capturing all system components, OREO implements an algorithmic solution selecting the best service configuration, maximizing the number of shared xApps, and efficiently allocating resources to them. Numerical results and experimental tests, performed using our proof-of-concept implementation, demonstrate that OREO closely matches the optimum solution, outperforming the state of the art by deploying up to 35% more services with fewer xApps and a similar reduction in resource consumption.

In summary, this dissertation contributes a holistic approach to advancing Open RAN principles, offering practical solutions for near-real-time RAN services. The presented frameworks, CAREM and OREO, showcase the potential for efficient resource utilization and orchestration, paving the way for enhanced RAN operations and laying a robust foundation for future mobile services in the telecommunications landscape.