Poster: Network-Aware Orchestration for QoS-Constrained Application Deployment

Fernando Rego*
José Costa*
fernando.rego@fraunhofer.pt
jose.costa@fraunhofer.pt
Fraunhofer Portugal AICOS
Porto, Portugal

Carlos Resende
João Oliveira
carlos.resende@fraunhofer.pt
joao.oliveira@fraunhofer.pt
Fraunhofer Portugal AICOS
Porto, Portugal

Luís Almeida
André Restivo
lda@fe.up.pt
arestivo@fe.up.pt
FEUP - University of Porto
Porto, Portugal

Abstract

The computing continuum extends computation beyond centralized cloud data centers to edge resources, supporting applications with stringent Quality-of-Service (QoS) requirements. However, the heterogeneity and dynamic nature of edge environments make orchestration challenging, particularly when decisions rely on static network information. This paper proposes a network-aware orchestration architecture that integrates a continuously maintained network representation to support adaptive application placement and management. By combining network and computational awareness, the proposed approach aims to improve orchestration decisions and application performance across the computing continuum.

CCS Concepts

• **Computer systems organization** → **Distributed architectures**; • **Networks** → *Network management*; *Network architectures*.

Keywords

Network-Aware Orchestration, Resource Management, Network Tomography, Edge Computing, QoS

1 Introduction

The adoption of distributed computing paradigms has extended computation beyond centralized cloud data centers toward the network edge, addressing the limitations of cloud-centric architectures in meeting the operational and Quality-of-Service (QoS) requirements of latency-sensitive, energy-constrained, and communication-intensive applications [1].

However, unlike relatively homogeneous cloud environments, edge infrastructures comprise diverse computing resources, networking technologies, energy constraints, and administrative domains. Moreover, resource availability and network conditions can vary considerably over time due to mobility, fluctuating workloads, and changing connectivity patterns, impacting applications with stringent requirements regarding latency, throughput, availability, and resilience. Despite this dynamic environment, orchestration decisions are frequently based on static assumptions or infrequently updated network information, limiting their ability to adapt to evolving operating conditions [1].

Consequently, integrating continuously updated network-state information into application orchestration remains an open research challenge. While notable advances have been made in network monitoring, telemetry, and tomography, the insights produced by these mechanisms are rarely integrated into orchestration processes that jointly reason about computational resources and the underlying network substrate to improve orchestration decisions. To address this gap, we propose a hybrid network monitoring framework and a network-aware orchestrator that leverages continuously collected network metrics to support adaptive application placement and management across heterogeneous networks.

2 Network-Aware Orchestration

The proposed architecture, presented in Figure 1, can improve orchestration and placement decisions by considering application QoS requirements and continuously updated network information. Unlike orchestration solutions that assume homogeneous IP connectivity, such as Diktyo [2], our approach integrates protocol gateways, enabling end-to-end reasoning across heterogeneous communication domains and exposing protocol transitions and communication bottlenecks.

Furthermore, the network is modeled as a graph whose nodes represent computing resources, communication devices, and protocol gateways, and whose edges represent communication links.

The network graph representation is maintained through hybrid monitoring that combines passive observations with selective active measurements [3] and is enriched with topology, resource, and performance information, including latency, bandwidth, packet loss, and link utilization.

The resulting graph is exposed via a topology analytics component (cf. 1, Figure 1) that supports application orchestration (cf. 2, Figure 1). The representation can also support other tasks that require up-to-date network views, such as network configuration orchestration, performance simulation, what-if analyses, and digital twin environments.

Applications are represented as dependency graphs (cf. 3, Figure 1) with QoS requirements such as latency, bandwidth, reliability, and locality constraints. By jointly reasoning over application and network graphs, the orchestrator can place, migrate, or replicate components to improve QoS compliance while balancing resource utilization, energy efficiency, and fault tolerance.

3 Network Simulator for Orchestration

To support the development and evaluation of the proposed architecture, we built a network simulation environment capable of modeling heterogeneous communication infrastructures. This simulator enables topology discovery and metric collection via realistic

*Both authors contributed equally to this research.

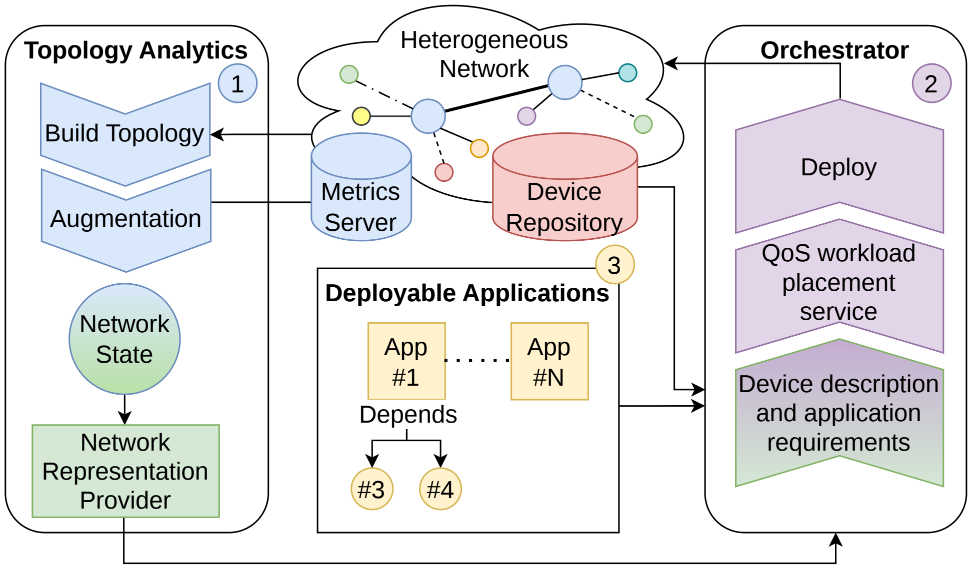

Figure 1: Overview of the proposed architecture.

monitoring and management interfaces. The platform combines ns-3[1] network simulations with virtualized network appliances (e.g., OpenWRT[2]) and end hosts, which can be simulated or virtualized.

Our current objective is to understand which protocols, metrics, and measurement mechanisms can be combined to discover topology, characterize links and devices, and build an orchestration-oriented network representation. Preliminary experiments, using active monitoring probes, indicate that the selected monitoring mechanisms can accurately reconstruct the network topology specified in a static ns-3 scenario and expose resource and performance metrics relevant to orchestration decisions.

By comparing estimated topology and performance metrics with the network state and link characteristics specified by the ns-3 scenario, we can evaluate the accuracy of different discovery, monitoring, and metric-enrichment strategies. These results will guide the design of the topology analytics component and help identify the information most valuable for QoS-aware orchestration in heterogeneous networks.

4 Conclusion

This poster presented a network-aware orchestration architecture that bridges network monitoring and application orchestration in the computing continuum.

While the current metrics collection framework already provides valuable insights into network characteristics and topology, further work is required to reduce the reliance on active network probes through passive data collection techniques and to expand the set of supported network and resource metrics. Future research will focus on evaluating the proposed approach in realistic deployment scenarios, assessing its effectiveness in supporting QoS-aware orchestration, and on deploying best-effort and soft real-time services.

Acknowledgments

This work was funded by EU's Horizon Europe HORIZON-JU-Chips-2024-1-IA, project NeAIxt, GA No 101194172, also supported by France, Italy, the Czech Republic, Germany, Sweden, Denmark, Greece, Spain, the Netherlands, Portugal, Turkey, and Switzerland. Views and opinions expressed are those of the author(s) only and do not reflect those of the EU or CHIPS JU. Neither the EU nor the granting authority can be held responsible for them.

References

[1] Angelo Marchese and Orazio Tomarchio. 2024. Telemetry-Driven Microservices Orchestration in Cloud-Edge Environments. In *2024 IEEE 17th International Conference on Cloud Computing (CLOUD)*. IEEE, Shenzhen, China, 91–101. doi:10.1109/CLOUD62652.2024.00020

[2] José Santos, Chen Wang, Tim Wauters, and Filip De Turck. 2023. Diktyo: Network-Aware Scheduling in Container-Based Clouds. *IEEE Transactions on Network and Service Management* 20, 4 (2023), 4461–4477. doi:10.1109/TNSM.2023.3271415

[3] Yi Sun, Dong Li, and Hongjie Sun. 2007. Network Tomography and Improved Methods for Delay Distribution Inference. In *The 9th International Conference on Advanced Communication Technology*. IEEE, Gangwon-Do, Korea, 1433–1437. doi:10.1109/ICACT.2007.358625

[1] https://www.nsnam.org/
[2] https://openwrt.org