Task Offloading in Open RAN: Deployment Scenarios, Research Trends

Heejae Park, Seongryool Wee, Seungyeop Song, and Laihyuk Park
Department of Computer Science and Engineering, Seoul National University of Science and Technology, Seoul, 01811, Korea
Email: {prkhj98, holylaw, sysong, lhpark}@seoultech.ac.kr

***Abstract*—By disaggregating base station functions, Open Radio Access Network (O-RAN) enables enhanced interoperability, scalability, and intelligence throughout the RAN infrastructure. However, end devices within the O-RAN environment continue to face constraints in computing power, memory, and energy capacity. To address these limitations, task offloading emerges as a critical solution. This paper investigates representative O-RAN deployment scenarios and explores current research trends.**

***Index Terms*—O-RAN, deployment scenarios, task offloading.**

I. Introduction

Open Radio Access Network (O-RAN) has emerged as a transformative architecture for next-generation mobile networks, aiming to address the limitations of traditional, vendor-locked RAN systems. By disaggregating the base station functions into the O-RAN Central Unit (CU), O-RAN Distributed Unit (DU), and O-RAN Radio Unit (RU), O-RAN fosters interoperability, scalability, and intelligence across the RAN infrastructure [1]. To enhance the flexibility of O-RAN, O-RAN Alliance introduces various deployment scenarios (Scenario A to F in Fig. 1).

Despite these architectural advancements, end devices in the O-RAN environment still suffer from limited computing power, memory, and battery life. As applications such as augmented reality, autonomous driving, and real-time video analytics become increasingly computation-intensive, device-side processing often results in unacceptable latency. To address this challenge, recent studies have increasingly focused on task offloading, recognizing it as a promising solution for alleviating computational burdens and enhancing system performance [2]. In this paper, we examine the various deployment scenarios of O-RAN and research trends.

II. O-RAN Deployment Scenarios

O-RAN is composed of several components, including the Service Management and Orchestration (SMO), O-Cloud, RAN Intelligence Controller (RIC), O-CU, O-DU, O-RU, R1, A1, O1, O2, E2, and O-RAN Fronthaul. As shown in Fig. 1, O-RAN Alliance introduces various deployment scenarios [3].
Scenario A: Near-RT RIC, O-CU, and O-DU are co-located on the same edge cloud platform, enabling low-latency coordination. This setup is ideal for dense urban areas with high fronthaul capacity, allowing centralized baseband processing. An open fronthaul connects the O-DU to the O-RU.

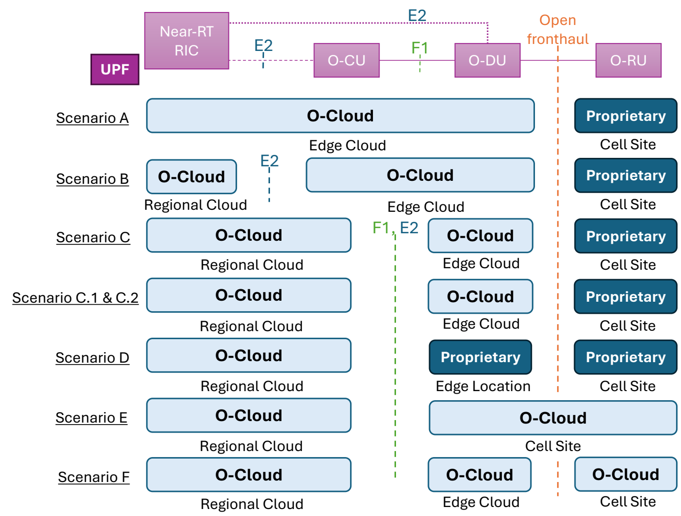

Fig. 1. O-RAN deployment scenarios.

Scenario B: Near-RT RIC is deployed at a regional cloud, while the O-CU and O-DU are co-located at an edge cloud. The E2 interface connects the Near-RT RIC with both O-CU and O-DU across these locations. This setup is designed for areas with limited fronthaul capacity and geographically dispersed O-RUs. It helps reduce latency by placing compute functions closer to the cell sites, while the centralized Near-RT RIC enables coordinated control.
Scenario C: Near-RT RIC and O-CU are deployed on a regional cloud, while the O-DU is placed on an edge cloud. The Near-RT RIC and O-CU share the same cloud platform and are connected to the O-DU via both F1 and E2 interfaces. This scenario is suitable for deployments with limited fronthaul capacity and widely distributed O-RUs. It improves resource efficiency by pooling control-plane functions centrally while maintaining low-latency data-plane processing at the edge. Scenario C.1 and C.2 are two variations to handle network slice instances.
Scenario D: Scenario D is a variant of Scenario C, where the O-DU is replaced by a non-virtualized (physical) O-DU at the edge location. The Near-RT RIC and O-CU remain in the regional cloud, connected to the physical O-DU via F1 and E2 interfaces. This scenario is used when a fully virtualized

O-Cloud deployment is not feasible due to cost, deployment timeline, or implementation constraints.

Scenario E: Both O-DU and O-RU are implemented on the same cloud platform at the cell site, possibly leveraging acceleration hardware if needed. Meanwhile, the Near-RT RIC and O-CU remain at the regional cloud, connected via E2 and F1 interfaces.

Scenario F: O-RU is deployed at cell site, while O-DU is deployed at Edge cloud. The O-CU and Near-RT RIC remain at the regional cloud.

III. Research Trends

The authors in [4] address the problem of minimizing end-to-end task delay in Open RAN by jointly optimizing task offloading and fronthaul segment routing. Unlike traditional systems with fixed paths, the disaggregated structure of Open RAN allows dynamic routing through multiple intermediate nodes like DU pools. The authors propose a Deep Q-Learning (DQL) framework to make offloading and routing decisions based on real-time task and network states. The state space includes current local computation, cloud computation, and offloading using wireless and fronthaul states. The action space consists of offloading decisions, communication resource allocation and routing decisions, and computation resources allocation decisions. Reward is formulated to minimize total delay. To ensure scalability and data privacy, they integrate Federated Learning (FL) so that each node trains its local Q-network and only shares model updates.

The authors in [5] propose an Age of Processing (AoP)-based offloading strategy for autonomous vehicles in a Multi-RAT O-RAN environment. AoP refers to the duration between task generation and the reception of the corresponding computation result. The system consists of vehicles generating computation task, which can be processed locally or offloaded via multiple RATs (e.g., LTE, 5G NR) to edge servers managed under Open RAN. This paper considers 4 scenarios: 1) perform task computation locally, 2) offload the task directly to the nearest edge cloud, 3) offload the task to another edge cloud, and 4) offload the task to the regional cloud through an edge cloud. To solve AoP minimization problem, authors transform the surrogate problem into an unconstrained optimization problem using the Lagrangian method, followed by the application of the dual decomposition technique.

The authors in [6] focus on intelligent task offloading and resource allocation in O-RAN. They use a Proximal Policy Optimization (PPO)-based solution which is one of the reinforcement learning algorithms. The paper distinguishes between URLLC and mMTC service types by assigning two different reward functions tailored to their QoS needs:

- For URLLC service, the reward function prioritizes latency minimization and SLA violation.
- For mMTC service, the reward function focuses on energy efficiency, allowing for more relaxed latency constraints.

The state space consists of slice type, number of bytes, number of CPU cycles, acceptable latency, remaining communication and computational resources, CPU frequency of user, and CPU frequency of each computational resource. The action space includes number of communication resource blocks and computation resources allocated to user.

The authors in [7] proposes Oranits, a unified optimization framework for mission assignment and task offloading in O-RAN-based Intelligent Transportation Systems (ITS). The system integrates O-RAN and Mobile Edge Computing (MEC) to enable real-time cooperation among autonomous vehicles, with a focus on mission dependencies, offloading costs, and adaptive coordination. The main objective is to maximize the number of missions completed before their deadlines, while minimizing delay and offloading cost. Authors propose metaheuristic algorithm, Chaotic Gaussian-based Global ARO (CGG-ARO), Multi-agent Double Deep Q-Network (MA-DDQN) to solve the problem.

IV. Conclusion

This paper examines the deployment scenarios of O-RAN and research trends. Aforementioned, approaches such as DRL, federated learning, and dual decomposition methods demonstrate the growing potential of intelligent scheduling in disaggregated network environments. Future research could explore graph neural network–based approaches that effectively capture the spatial characteristics of O-RAN.

Acknowledgment

This research was partly supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2025-RS-2022-00156353) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation) and National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2025-16070295).

References

[1] M. Polese, M. Dohler, F. Dressler, M. Erol-Kantarci, R. Jana, R. Knopp, and T. Melodia, "Empowering the 6g cellular architecture with open ran," *IEEE Journal on Selected Areas in Communications*, vol. 42, no. 2, pp. 245–262, 2023.

[2] Q. Wang, W. Qi, J. Ling, J. Zong, Y. Shen, and D. Liu, "Energy-efficient resource allocation in leo-assisted open ran architecture towards 6g," in *2024 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. IEEE, 2024, pp. 1–6.

[3] O-RAN Alliance, "O-ran cloud architecture and deployment scenarios for o-ran virtualized ran 8.01. available online: https://www.o-ran.org/specifications," Tech. Rep., 2025.

[4] A. Ndikumana, K. K. Nguyen, and M. Cheriet, "Federated learning assisted deep q-learning for joint task offloading and fronthaul segment routing in open ran," *IEEE Transactions on Network and Service Management*, vol. 20, no. 3, pp. 3261–3273, 2023.

[5] A. Ndikumana, K. K. Nguyen, and M. Cheriet, "Age of processing-based data offloading for autonomous vehicles in multirats open ran," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 21 450–21 464, 2022.

[6] A. Ebrahimi and F. Afghah, "Intelligent task offloading: Advanced mec task offloading and resource management in 5g networks," in *2025 IEEE Wireless Communications and Networking Conference (WCNC)*, 2025, pp. 1–6.

[7] N. H. Nguyen, N. Van Thieu, Q.-T. Luu, A. T. Nguyen, S. Wanasekara, N. C. Luong, F. Kavehmadavani, and V.-D. Nguyen, "Oranits: Mission assignment and task offloading in open ran-based its using metaheuristic and deep reinforcement learning," *arXiv preprint arXiv:2507.19712*, 2025.