Challenges and Research Directions in O-RAN Radio Resource Management

Byeongjin Kim
Dept. of Electrical Engineering
UNIST
Ulsan, Republic of Korea
alt30@unist.ac.kr

Minkwon Lee
Dept. of Electrical Engineering
UNIST
Ulsan, Republic of Korea
minkwon1114@unist.ac.kr

Hyoil Kim
Dept. of Electrical Engineering
UNIST
Ulsan, Republic of Korea
hkim@unist.ac.kr

***Abstract*—The future 6G network necessitates its RAN architecture to guarantee rigorous service- and user-centric performances, while seamlessly integrating communication, computing, and AI in the combined domains of cloud, edge, and user terminals. Open RAN (O-RAN) offers a promising platform to meet such requirements by enabling open interfaces, data-driven control, and the orchestration of heterogeneous resources. This paper examines the principal challenges and key research directions in O-RAN based radio resource management (RRM). We first review deep learning based RRM approaches and their applicability within the O-RAN near-RT and non-RT RIC frameworks. Next, we highlight the limitations of current 5G interfaces and discuss potential enhancements to facilitate the joint optimization of radio and O-Cloud resources. Finally, we outline future research directions, including channel prediction based RRM, user-centric scheduling, and service provider oriented optimization. These insights indicate that O-RAN, when integrated with intelligent RRM, can form a foundational basis for realizing the service-driven performant 6G network.**

***Index Terms*—6G, O-RAN, O-Cloud, RRM**

I. Introduction

One of 6G's main goals is to strictly satisfy the task and service requirements of individual users, rather than only enhancing the average performance from the network's perspective. To achieve this, 6G should integrate communication with computing, sensing, storage, control, and AI, across the entire cloud-edge-terminal domains. In addition, 6G networks should pursue a service-centric and user-centric architecture that understands user preferences and adjusts resource allocation based on QoE (Quality of Experience). To fulfill the vision, 6G's core KPIs (Key Performance Indicators) need to embrace more than traditional metrics (e.g., transmission rate, latency) and reflect user satisfaction (e.g., User Satisfaction Ratio, QoE) and personalized service quality [1].

For 6G RRM (Radio Resource Management), it remains an essential challenge to distribute radio resources appropriately among the end-users so as to maximize the number of users/sessions with their target KPIs satisfied. In particular, 6G RAN with heterogeneous traffic (e.g., eMBB, URLLC, XR) should deal with a diverse range of RRM-related tasks such as scheduling, power/PRB allocation, beamforming, and slicing, thus facing high optimization complexity. In this regard, recent research has shown that deep learning (DL) based RRM, particularly DRL, provides both policy exploration and adaptability in complex wireless environments, going beyond the performance of existing heuristics [2].

Open RAN (O-RAN) is particularly well suited to the RRM requirements. O-RAN enables data-driven closed-loop control through near-RT/non-RT RIC running xApp/rApp and programmable control of RAN through open interfaces such as A1, E2, O1, and O2. Additionally, by integrating lifecycle, resource, and observation management of CU/DU/RIC with SMO and O2/O1 interfaces, it provides a structure that can orchestrate wireless resources and cloud resources altogether. Designed based on the premise of radio-computing integrated intelligence, O-RAN is a strong candidate as an execution platform for service-oriented 6G [3].

In such a vein, this paper presents key elements for implementing RRM in an O-RAN environment. First, we summarize the recent trends in DL-based RRM to gain insights into the models and workflows suitable for O-RAN RIC. Next, it points out the limitations of 5G interfaces from the RRM's perspective and suggests a possible direction of O-RAN standard evolution for concurrent optimization of radio and O-Cloud resources. Finally, the paper presents potential research tasks including channel prediction based RRM, user-centric RRM, and RRM from service provider's perspective.

This paper is organized as follows. Section II introduces the elements necessary for implementing O-RAN RRM, and Section III presents research directions for designing O-RAN RRM problems. Then, the paper concludes with Section IV.

II. Radio Resource Management in O-RAN

When implementing RRM in the O-RAN environment, the following aspects are necessary to be incorporated, as shown in Fig. 1.

A. *Designing O-RAN-compliant RIC with a suitable DL model*

In the literature, several AI-based methods have been investigated for intelligent RRM. In [4]–[6], model-free RL techniques such as DQN and offline RL have been applied to intelligent resource management for joint network resource

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (RS-2024-00405128).

Hyoil Kim is the corresponding author.

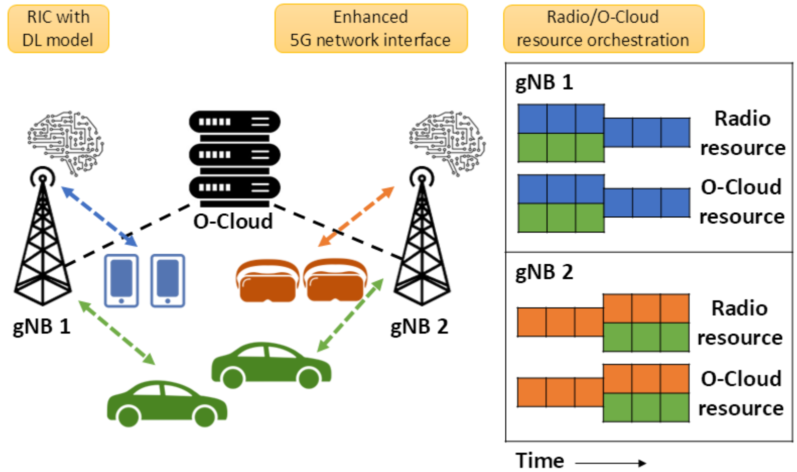

Fig. 1: O-RAN architecture with RRM support

allocation and scheduling. In [7], a dueling DDQN-based dynamic bandwidth allocation scheme is proposed for multi-slice multi-user RAN slicing to maximize spectrum efficiency and SLA (Service Level Agreement) satisfaction. On the one hand, in [8], [9], transformer and LSTM models have been integrated into the state processing stage of the DRL agent to effectively capture historical, sequential, and temporal correlation information of network conditions. Nonetheless, the aforementioned approaches did not consider the implementation issues in the O-RAN environment.

Meanwhile, there also have been research conducted regarding DL-based intelligent RRM while taking into account realistic O-RAN environments. In [10], the authors utilized the O-RAN E2 interfaces to obtain real-time UE/gNB metrics and transmit trained DRL policies, for maximizing throughput and minimizing latency. [11] presented an O-RAN compliant DRL-based two-level scheduling RRM framework that minimizes SLA violations and outperforms conventional methods. Recently, [12] proposed an O-RAN slicing framework that enhances QoS for heterogeneous services and reduces resource costs by jointly optimizing radio and computing resources via the parallel hierarchical DRL-based resource allocation algorithm. Such research trends imply that DRL-based RRM is well-suited for adapting to the complex wireless environment of O-RAN and the policy exploration, compared to traditional heuristics.

B. Enhancing the existing 5G network interfaces

While numerous 5G interfaces exist already, some modifications may be necessary to ensure more enhanced performance within the 6G RAN. For instance, the QoS flow-to-DRB mapping rule is provided by the 5GC to the UE via the NGAP interface. In scenarios where pricing policies are in effect, specifically when UEs paying higher fees are assured higher priority and bit rate guarantees, the mapping rules may need to be applied differently, e.g., QoS flow from high-fee users should be allocated to dedicated DRBs, separate from other UEs. Therefore, it is essential to modify the rules to align with each UE's status/requirements and to revise the related interfaces and their messages to accommodate such changes.

Moreover, certain messaging protocols could be streamlined or omitted to facilitate AI-driven radio resource optimization. In the case of admission-allocation joint control optimization, intelligent RRM needs to simultaneously determine whether a UE should be admitted and how much resources need to be allocated to it. Nevertheless, the current 5G network involves too many messaging processes between the admission control and the resource allocation decision, e.g., RA (Random Access) procedure, RRC request, and BSR (Buffer Status Report). As a result, it is necessary to further simplify the interface messages to expedite the process of intelligent RRM policies.

C. Orchestration of heterogeneous resources

An orchestration strategy is essential for the concurrent optimization of radio resource allocation and O-Cloud resource management within the O-RAN environment. O-Cloud is a cloud computing platform defined by the O-RAN [13], with resources comprising computing, storage, and network components. In addition to radio resources in the RAN, O-Cloud resources are another key consideration in O-RAN's resource management to meet UEs' KPIs and ensure the QoS of heterogeneous services. Thus, jointly optimized utilization of radio and O-Cloud resources would become one of the most critical challenges in 6G RRM.

III. Further Challenges in O-RAN RRM

There also exist several research challenges in designing O-RAN RRM problems, as introduced below.

A. Channel prediction based RRM

The 6G environment experiences instantaneous and continuous fluctuations in resource demand, due to highly heterogeneous services (e.g., eMBB, URLLC, XR), constantly changing user mobility patterns, and frequent handover due to cell departure and re-entry, and changes in service usage rates over time. Moreover, 6G's diverse components dynamically interact with each other, such as wireless link status (e.g., SINR, CSI), UE's traffic demand, and O-Cloud server load. Hence, RRM will always remain reactive and suboptimal if the aforementioned non-steady and time-varying conditions are not addressed properly and timely.

In this regard, a key enabler of 6G RRM must be channel prediction driven resource management. The purpose of channel prediction is to prepare scheduling, slicing, and orchestration decisions in advance by estimating (i) wireless link status and PRB requirements, (ii) traffic requirements for each network slice, and (iii) O-Cloud load and deployment requirements. As more accuracy and uncertainty-aware predictions become available, RRM could proactively ensure QoS/QoE for more user sessions while simultaneously improving the efficiency of radio and cloud resource utilization.

B. User-centric RRM

In order to design resource optimization problems from a user's perspective, it is necessary to assume that performance requirements vary with service. For example, URLLC imposes a stricter requirement on the maximum end-to-end latency and minimum bandwidth, compared to mMTC [14]. In addition, new services might emerge with a new set of KPIs, e.g., cloud gaming with input lag and the percent of session freezes [15].

One possible goal of user-centric RRM is to maximize the number of users with target QoE/QoS fulfilled. To achieve this, resource scheduling policies can be implemented by considering available wireless and O-Cloud resource constraints and service priorities. To evaluate the performance of such policies, we can adopt the metrics like session-based QoE, request fulfillment rate, and user satisfaction rate, where they are utilized as closed-loop feedback to repeatedly adjust a given scheduling policy.

In addition, User Admission Control (UAC) responds to network condition changes by determining whether to accept or reject new service sessions, based on demand forecasts and service priorities. UAC can be evaluated using various indicators such as the percentage of accepted sessions meeting or violating their requirements, so as to contribute to reliable improvement of user experience in a multi-service multi-user highly-heterogeneous environment.

C. Service provider oriented RRM

It is natural for a service provider to try to maximize its profit by offering services to as many UEs as possible. However, since radio and O-cloud computing resources are limited, it is inevitable to incorporate UE admission decision in such an optimization process. Furthermore, due to diverse subscription plans and service requirements among UEs, maximizing profit while fulfilling all individual service demands (i.e., SLA) becomes a significant challenge. Therefore, the primary objective for a service provider should be to maximize its expected total revenue by admitting and allocating PRBs to UEs, while ensuring compliance with each UE's SLA.

IV. CONCLUSION

This paper identified the key building blocks for implementing RRM in O-RAN to assure QoS/QoE, and delineated research challenges related to the O-RAN RRM problem. Guided by these insights, our future work will design and evaluate practical RRM suitable for deployment in the operational O-RAN system.

REFERENCES

[1] Y. Yang, M. Ma, H. Wu, Q. Yu, X. You, J. Wu, C. Peng, T.-S. P. Yum, A. H. Aghvami, G. Y. Li *et al.*, "6G network AI architecture for everyone-centric customized services," *IEEE Network*, vol. 37, no. 5, pp. 71–80, 2022.

[2] Y. Sun, M. Peng, Y. Zhou, Y. Huang, and S. Mao, "Application of machine learning in wireless networks: Key techniques and open issues," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3072–3108, 2019.

[3] M. Polese, L. Bonati, S. D'oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1376–1411, 2023.

[4] F. Al-Tam, N. Correia, and J. Rodriguez, "Learn to Schedule (LEASCH): A Deep Reinforcement Learning Approach for Radio Resource Scheduling in the 5G MAC Layer," *IEEE Access*, vol. 8, pp. 108 088–108 101, 2020.

[5] S. Fu, F. Yang, and Y. Xiao, "AI inspired intelligent resource management in future wireless network," *IEEe Access*, vol. 8, pp. 22 425–22 433, 2020.

[6] K. Yang, C. Shi, C. Shen, J. Yang, S.-P. Yeh, and J. J. Sydir, "Offline reinforcement learning for wireless network optimization with mixture datasets," *IEEE Transactions on Wireless Communications*, vol. 23, no. 10, pp. 12 703–12 716, 2024.

[7] Y. Jia, C. Zhang, N. Li, Y. Huang, and T. Q. Quek, "Hierarchical intelligence enabled joint RAN slicing and MAC scheduling for SLA guarantee," *IEEE Transactions on Communications*, vol. 73, no. 8, pp. 6081–6094, 2024.

[8] S. Wang, S. Bi, and Y.-J. A. Zhang, "Deep reinforcement learning with communication transformer for adaptive live streaming in wireless edge networks," *IEEE Journal on Selected Areas in Communications*, vol. 40, no. 1, pp. 308–322, 2021.

[9] R. Li, C. Wang, Z. Zhao, R. Guo, and H. Zhang, "The LSTM-based advantage actor-critic learning for resource management in network slicing with user mobility," *IEEE Communications Letters*, vol. 24, no. 9, pp. 2005–2009, 2020.

[10] L. Bonati, S. D'Oro, M. Polese, S. Basagni, and T. Melodia, "Intelligence and learning in O-RAN for data-driven NextG cellular networks," *IEEE Communications Magazine*, vol. 59, no. 10, pp. 21–27, 2021.

[11] J. Dai, L. Li, R. Safavinejad, S. Mahboob, H. Chen, V. V. Ratnam, H. Wang, J. Zhang, and L. Liu, "O-RAN-enabled intelligent network slicing to meet service-level agreement (sla)," *IEEE Transactions on Mobile Computing*, vol. 24, no. 2, pp. 890–906, 2024.

[12] K. Qiao, H. Wang, W. Zhang, D. Yang, Y. Zhang, and N. Zhang, "Resource allocation for network slicing in open RAN: A hierarchical learning approach," *IEEE Transactions on Cognitive Communications and Networking*, vol. 11, no. 4, pp. 2584–2600, 2025.

[13] O.-R. W. G. 6, "Cloud architecture and deployment scenarios for O-RAN virtualized RAN 7.0," Technical Specification, June 2024.

[14] S. A. AlQahtani, "Cooperative-aware radio resource allocation scheme for 5G network slicing in cloud radio access networks," *Sensors*, vol. 23, no. 11, p. 5111, 2023.

[15] C. Baena, O. S. Peñaherrera-Pulla, R. Barco, and S. Fortes, "Measuring and estimating key quality indicators in cloud gaming services," *Computer Networks*, vol. 231, p. 109808, 2023.