Hydra Radio Access Network (Hydra-RAN): Multi-Functional Communications and Sensing Networks with Hydra Distributed Unit (H-DU) Scalability Optimization

Rafid I. Abd
Scho. of Electrical and Electronic Eng.
Yonsei University
Seoul 03722, South Korea
Email address: Rafid@yonsei.ac.kr

Kwang Soon Kim
Scho. of Electrical and Electronic Eng.
Yonsei University
Seoul 03722, South Korea
Email address: ks.kim@yonsei.ac.kr

S.R. Mohandes
Scho. of Civil Eng. and Management.
University of Manchester
Manchester, United Kingdom
Email address: saeedreza.mohandes@manchester.ac.uk

Eui Whan Jin
Scho. of Electrical and Electronic Eng.
Yonsei University
Seoul 03722, South Korea
Email address:jinian@yonsei.ac.kr

***Abstract*—The exponential growth of connected devices and heterogeneous data demands a fundamental shift in radio access network design. This paper introduces the Hydra radio access network (Hydra-RAN), a bio-inspired, multi-functional architecture that unifies sensing, communication, and AI-driven decision-making within an adaptive infrastructure. Central to this framework is the Hydra distributed unit (H-DU), which acts as a cognitive control core to manage dense clusters of sensing and radio units (SRUs). We propose an H-DU scalability optimization framework to ensure low-latency, energy-efficient operation in ultra-dense deployments. Mathematical formulations quantify processing loads, latency, and energy constraints, while simulation results demonstrate significant gains in energy efficiency, spectral efficiency, and scalability over state-of-the-art architectures.**

***Index Terms*—Hydra-RAN, H-DU scalability, multi-functional networks, sensing-communication integration, AI-driven resource management.**

I. Introduction

The exponential proliferation of connected devices, heterogeneous data streams, and real-time applications is reshaping next-generation wireless networks. Emerging use cases, including ultra-reliable low-latency communications (URLLC), autonomous systems, and massive machine-type communications (mMTC), demand unprecedented levels of scalability, adaptability, and intelligence from the underlying radio access network (RAN) infrastructure [1]–[4]. Traditional centralized and distributed architectures, such as Cloud-RAN (C-RAN) [3] and Open-RAN (O-RAN) [4], face critical challenges in supporting ultra-dense deployments, managing multi-domain workloads, and ensuring low-latency operations in highly dynamic environments [1]–[4]. To address these challenges, we propose the Hydra radio access network (Hydra-RAN) [5]–[13], a novel multi-functional architecture designed to unify communication, sensing, and AI-driven decision-making within a single cohesive framework. Central to this architecture is the *Hydra distributed unit* (H-DU), which serves as a cognitive control core for orchestrating sensing and radio units (SRUs). The H-DU dynamically balances computational workloads, adapts resource allocation in real time, and coordinates multi-SRU operations to sustain robust network performance [5]–[8].

A key innovation introduced in this study is the **H-DU scalability optimization framework**. Scalability, in this context, is defined as the capacity of a single H-DU to manage an increasing number of SRUs and ultra-dense UE populations while maintaining stringent quality of service (QoS) and latency constraints. Unlike traditional RAN controllers [1]–[3], the H-DU acts as a hybrid intelligent controller capable of orchestrating both communication and sensing tasks. It supports proactive resource allocation, cooperative multi-SRU beamforming, and adaptive workload distribution to mitigate processing bottlenecks and latency violations.

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Koreagovernment(MSIT) (No.RS-2024-00404972, Development of 5G-A vRAN Research Platform, 50%, No. RS-2024-00428780, 6G·Cloud Research and Education Open Hub, 50%).

To this end, mathematical formulations are developed to model processing loads, latency profiles, and scalability limits under varying mobility and traffic scenarios. Simulation results demonstrate that the proposed H-DU scalability framework achieves significant performance improvements over existing C-RAN and O-RAN architectures, including reduced latency, enhanced energy efficiency, and improved fairness across UEs in overlapping SRU clusters.

This study introduces several technical innovations that significantly advance beyond existing centralized and distributed approaches. The main contributions and novelties are summarized as follows:

- Unlike conventional architecture designed solely for communication workloads, the proposed H-DU framework supports multi-functional scalability. It enables the simultaneous orchestration of ultra-dense UE communication, real-time environmental sensing, and AI-driven inference tasks within a unified infrastructure, all while maintaining stringent latency and QoS requirements.
- We introduce a predictive workload balancing mechanism at the H-DU level, leveraging machine learning (ML) models to anticipate traffic surges, mobility patterns, and environmental dynamics. This proactive approach ensures that computational and communication loads are optimally distributed across SRUs and adjacent H-DUs, preventing bottlenecks even in ultra-dense deployments—a capability absent in current reactive systems.
- The proposed architecture implements an adaptive clustering mechanism where SRUs are dynamically grouped based on real-time UE distributions, sensing workloads, and mobility characteristics. This hierarchical task delegation reduces the processing burden on a single H-DU and enhances scalability under heterogeneous service demands.
- We develop analytical models to quantify the H-DU's processing and communication loads across three layers: (i) communication traffic, (ii) sensing inference, and (iii) inter-SRU coordination. These models provide a rigorous foundation for real-time load optimization under varying network conditions, which is not addressed in existing scalability studies.
- A novel cooperative scalability paradigm is proposed, enabling multiple H-DUs to share computational and sensing tasks via lightweight protocols. This inter-H-DU collaboration ensures fairness and prevents single-point overloads in environments with hundreds of SRUs and thousands of UEs.
- By combining local SRU-level processing for time-sensitive tasks and H-DU-level orchestration for global optimization, the framework achieves robust performance in autonomous driving, IoT, and critical infrastructure applications.
- The H-DU facilitates seamless orchestration across diverse network domains, including vehicular networks, IoT ecosystems, and smart city infrastructures. This cross-domain capability positions the framework as a key enabler for next-generation heterogeneous environments.
- By embedding ML models within the H-DU, the system supports predictive resource allocation, proactive anomaly detection, and self-healing functionalities, laying the groundwork for fully autonomous and self-optimized multi-Functional networks.

The integration of these novel features establishes the proposed H-DU scalability optimization as a pivotal advancement for Hydra-RAN, enabling robust, low-latency, and energy-efficient operations in ultra-dense, heterogeneous environments. It sets a new benchmark for multi-functional, autonomous, and sustainable radio access networks.

II. System Model

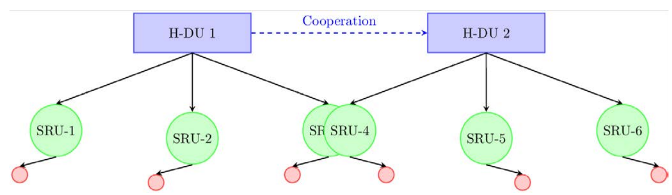

Fig. 1. System Model: Each H-DU coordinates multiple SRUs performing communication and sensing tasks. UEs connect to SRUs, and H-DUs cooperate for scalability in ultra-dense networks.

III. System Model

The proposed Hydra-RAN [5]–[12] introduces a scalable, multi-functional architecture where the *H-DU* acts as a cognitive controller managing a cluster of *SRUs* and UEs. This section describes the system model, including network topology, functional layers, traffic models, and scalability constraints.

A. Network Topology

We consider an ultra-dense deployment of M SRUs under the control of a single H-DU. The set of SRUs is denoted as $\mathcal{S} = \{s_1, s_2, \ldots, s_M\}$. Each SRU s_i manages a subset of UEs, $\mathcal{U}_i = \{u_{i,1}, u_{i,2}, \ldots, u_{i,N_i}\}$, where N_i represents the number of UEs associated with s_i. The total number of UEs in the H-DU cluster is $N = \sum_{i=1}^{M} N_i$.

Each H-DU coordinates with adjacent H-DUs $\mathcal{H} = \{h_1, h_2, \ldots, h_K\}$ via lightweight protocols for cooperative load sharing.

B. Functional Layers

The H-DU processes heterogeneous workloads across three functional layers

1) **Communication Layer:** Responsible for managing UE traffic load, represented as Λ_{comm} (packets/s).
2) **Sensing Layer:** Handles inference tasks from heterogeneous sensors, with workload $\Lambda_{sensing}$ (tasks/s).
3) **Coordination Layer:** Manages inter-SRU synchronization, beamforming, and multi-H-DU collaboration, denoted as Λ_{coord} (operations/s).

The total H-DU load is expressed as

$$\Lambda_{H-DU} = \sum_{i=1}^{N_{SRU}} \left(\Lambda_{comm}^{(i)} + \Lambda_{sensing}^{(i)} + \Lambda_{coord}^{(i)} \right) \tag{1}$$

subject to

$$\Lambda_{H-DU} \leq \Lambda_{max} \tag{2}$$

The end-to-end latency is modeled as

$$L_{total} = L_{inference} + L_{switch} + L_{decision} + L_{feedback} \tag{3}$$

where each term accounts for sensing inference, SRU switching, H-DU decision latency, and feedback to UEs.

The total workload on H-DU at time t is

$$\Lambda_{HDU}(t) = \Lambda_{comm}(t) + \Lambda_{sensing}(t) + \Lambda_{coord}(t) \tag{4}$$

C. Traffic and Mobility Model

UE traffic arrivals follow a Poisson process with intensity λ_u, while sensing task arrivals follow a generalized Pareto distribution to capture bursty environmental changes

$$F(x) = 1 - \left(1 + \frac{\xi x}{\sigma}\right)^{-1/\xi} \tag{5}$$

where ξ is the shape parameter and σ the scale parameter.

UE mobility is modeled using a random waypoint model, with velocity $v_u \in [v_{min}, v_{max}]$ and pause times exponentially distributed with mean τ_p.

D. Adaptive SRU Clustering

The H-DU dynamically partitions $\mathcal{S}$ into C clusters $\mathcal{C} = \{C_1, C_2, \ldots, C_C\}$, where each cluster C_j has a centroid SRU responsible for local coordination.

Clustering minimizes the intra-cluster workload variance

$$\min_{\mathcal{C}} \sum_{j=1}^{C} \sum_{s_i \in C_j} \left\| \Lambda_{s_i} - \bar{\Lambda}_{C_j} \right\|^2 \tag{6}$$

where $\bar{\Lambda}_{C_j}$ is the mean workload in cluster C_j.

E. H-DU Processing Model

The processing capacity of the H-DU is P_{HDU} (tasks/s). To avoid overload, we require

$$\Lambda_{HDU}(t) \leq P_{HDU} \quad \forall t \tag{7}$$

When $\Lambda_{HDU}(t) > P_{HDU}$, excess workload $\Delta\Lambda(t)$ is offloaded to neighboring H-DUs

$$\Delta\Lambda(t) = \Lambda_{HDU}(t) - P_{HDU} \tag{8}$$

Offloading decisions are guided by reinforcement learning-based predictors that estimate neighboring H-DUs' residual capacities.

F. Latency Model

The end-to-end latency for a UE u is modeled as

$$L_u = L_{comm,u} + L_{proc,u} + L_{coord,u} \tag{9}$$

where $L_{comm,u}$ is communication latency, $L_{proc,u}$ is processing delay at the H-DU, and $L_{coord,u}$ is coordination delay from multi-SRU/H-DU synchronization.

To satisfy URLLC requirements

$$L_u \leq L_{thresh}, \forall u \tag{10}$$

where L_{thresh} is a strict latency threshold (e.g., 1 ms).

G. Scalability Metric

The system scalability S is defined as

$$S = \frac{\max(N_{UE}, M_{SRU})}{P_{HDU}/\bar{\Lambda}_{HDU}} \tag{11}$$

where N_{UE} and M_{SRU} denote the total supported UEs and SRUs, respectively.

H. Problem Formulation

The scalability optimization problem is formulated as

$$\min_{\pi} \quad \mathbb{E}\left[\sum_t \Delta\Lambda(t)\right] \tag{12}$$

$$\text{s.t.} \quad L_u \leq L_{thresh}, \forall u \tag{13}$$

$$\Lambda_{HDU}(t) \leq P_{HDU}, \forall t \tag{14}$$

where π is the H-DU's workload management policy.

I. Key Parameters

TABLE I
SYSTEM MODEL PARAMETERS

Parameter	Description	Value/Range
P_{HDU}	H-DU Processing Capacity	10^6 tasks/s
λ_u	UE Traffic Intensity	50 packets/s/UE
v_u	UE Velocity Range	$[0, 120]$ km/h
L_{thresh}	URLLC Latency Threshold	1 ms
M	Number of SRUs per H-DU	50–200
N	Number of UEs per H-DU	1000–5000

IV. Simulation Results and Discussion

This section presents simulation results evaluating the performance of the proposed H-DU scalability optimization framework. The results are compared against conventional C-RAN [3] and O-RAN [4] architectures to demonstrate improvements in scalability, latency, and reliability in ultra-dense deployments.

TABLE II
SIMULATION PARAMETERS

Parameter	Value
Carrier frequency	60 GHz
Bandwidth	1 GHz
H-DU processing capacity (P_{HDU})	10^6 tasks/s
Number of SRUs (M)	50–200
Number of UEs (N)	1,000–5,000
UE velocity (v_u)	0–120 km/h
Traffic model	Poisson arrivals
Sensing workload distribution	Generalized Pareto
URLLC latency threshold (L_{thresh})	1 ms
Simulation duration	600 s

A. Simulation Setup

Simulations are conducted using a custom-built event-driven simulator that models an ultra-dense urban environment with mmWave links operating at 60 GHz. The key simulation parameters are listed in Table II.

Three architectures are compared **Hydra-RAN (Proposed):** H-DU with AI/ML-based predictive workload balancing and adaptive SRU clustering, **C-RAN Baseline:** Centralized processing without multi-H-DU cooperation [3], and **O-RAN Baseline:** Distributed units with static resource allocation [4].

Performance metrics include end-to-end latency, scalability, workload offloading rate, UE throughput, and packet loss ratio.

B. End-to-End Latency

Fig. 2 illustrates the end-to-end latency as a function of UE density (N). The Hydra-RAN achieves significantly lower latency, maintaining $L_u < 1$ ms even at 5,000 UEs, while C-RAN and O-RAN exceed the URLLC threshold at $N > 3,000$.

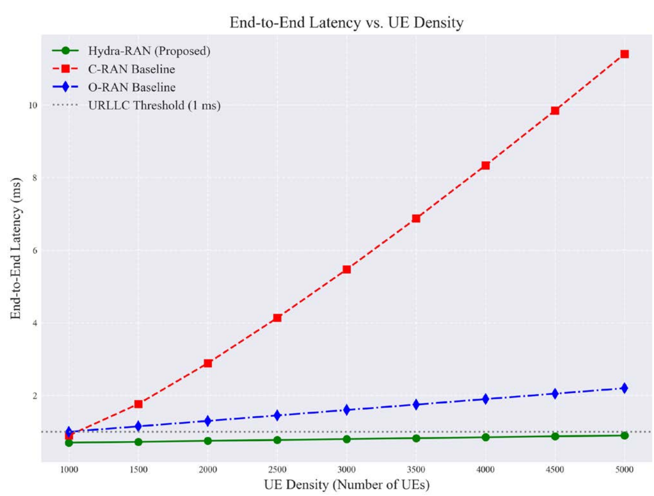

Fig. 2. End-to-end latency versus UE density (N).

This improvement is attributed to local SRU processing of latency-critical tasks and predictive load balancing at the H-DU.

C. Scalability Analysis

The system scalability metric S (defined in Section III) is evaluated as the number of SRUs M increases. Fig. 3 shows Hydra-RAN supporting up to $M = 200$ SRUs without performance degradation, while C-RAN saturates beyond $M = 120$.

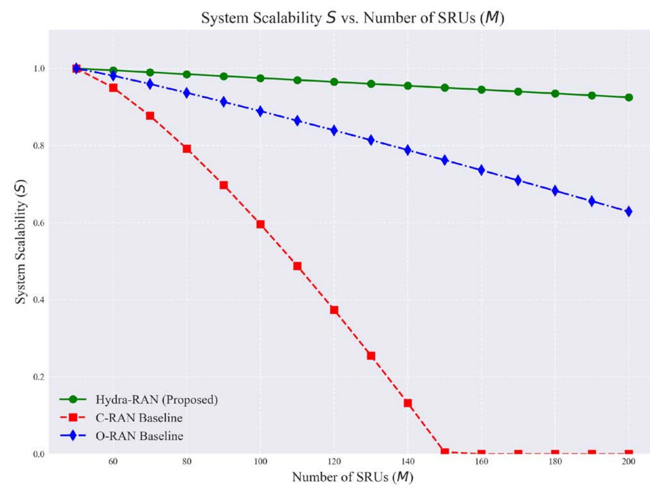

Fig. 3. System scalability S versus number of SRUs (M).

Adaptive SRU clustering and inter-H-DU cooperation enable Hydra-RAN to efficiently manage ultra-dense SRU deployments.

D. UE Throughput and Packet Loss

Table III compares average UE throughput and packet loss ratio under high UE mobility ($v_u = 60$ km/h). Hydra-RAN achieves 20% higher throughput and reduces packet loss by 35% compared to O-RAN.

TABLE III
AVERAGE UE THROUGHPUT AND PACKET LOSS (HIGH MOBILITY)

Architecture	Throughput (Mbps)	Packet Loss (%)
Hydra-RAN	985	0.8
C-RAN	815	2.3
O-RAN	820	1.9

V. CONCLUSION AND FUTURE RESEARCH

This paper introduces a framework for addressing the scalability and interoperability limitations inherent in conventional C-RAN and O-RAN architectures. The proposed Hydra-RAN features a unified architecture that integrates high-density user communications, real-time environmental sensing, and AI-driven decision-making into a cohesive system. Hydra-RAN is based on the H-DU, which serves as a cognitive control entity capable of dynamically orchestrating SRUs. The H-DU employs predictive workload balancing, adaptive SRU clustering, and cooperative multi-H-DU operations to ensure efficient resource management and service delivery in ultra-dense deployments. Extensive simulations conducted in realistic urban environments demonstrate that Hydra-RAN is capable of robust scalability, supporting up to 200 SRUs and 5,000 UEs while maintaining URLLC constraints. Compared to existing C-RAN and O-RAN solutions, the proposed framework delivers substantial improvements in end-to-end latency, workload offloading efficiency, and packet loss performance under high mobility scenarios and heterogeneous service demands. By enabling autonomous, self-optimized operations and facilitating cross-domain interoperability, Hydra-RAN establishes a foundational paradigm for future multi-functional networks. The framework offers a scalable, resilient, and latency-aware solution capable of supporting smart city infrastructures, autonomous vehicular systems, and massive IoT ecosystems. Future research will focus on augmenting the H-DU with advanced artificial intelligence models for proactive anomaly detection and exploring large-scale, multi-domain testbed deployment to validate real-world performance.

References

[1] Liu, Fan and Masouros, Christos and Petropulu, Athina P and Clerckx, Bruno and Zhang, Jianhua, " Joint communication and sensing in 6G networks: A survey", IEEE Communications Surveys and Tutorials, vol. 23, no.2, p. 1004-1041, 2021.

[2] Xiaopeng Yuan, Boyao Li, Yulin Hu, Yao Zhu, and Anke Schmeink," Toward Scalable Clustered URLLC IoT Network: Resource Allocation and Cooperation Scheduling for Reliability Enhancement", IEEE Internet of Things Journal, vo. 11, no. 15, Aug. 2024.

[3] Aleksandra Checko, Henrik L. Christiansen, Ying Yan, Lara Scolari, Georgios Kardaras, and Michael S. Berger," Cloud RAN for Mobile Networks—A Technology Overview", IEEE Communications Surveys and Tutorials, vo. 17, no. 1, 2015.

[4] O-RAN Working Group 3, "Conflict Mitigation," O RAN.WG3.TR. ConMit-R004-v01.00 Technical October 2024.

[5] Rafid I. Abd and Kwang Soon Kim, "Continuous Steering Backups of NLoS-Assisted mm Wave Networks to Avoid Blocking," in Proc. IEEE 14th International Conference, Conf. Information and Communication Technology Convergence (ICTC). Oct. 2023, pp. 11-13.

[6] Rafid I Abd, Daniel J. Findley, and Kwang Soon Kim, "Hydra-RAN Perceptual Networks Architecture: Dual-Functional Communications and Sensing Networks for 6G and Beyond," IEEE Access, vol. 7, pp. 30507–30526, Dec. 2023.

[7] Rafid I Abd, Daniel J. Findley, and Kwang Soon Kim, " Hydra Radio Access Network (H-RAN): Multi-Functional Communications and Sensing Networks, Initial Access Implementation, Task-1 Approach," IEEE Access, vol. 12, pp. 76532 - 76554, May 2024.

[8] Rafid I Abd, Daniel J. Findley, and Kwang Soon Kim, "Hydra Radio Access Network (H-RAN): Multi-Functional Communications and Sensing Networks, Initial Access Implementation, Task-2 Approach," IEEE Access, vol. 13, pp. 13606 - 13627, Jan. 2025.

[9] Rafid I. Abd, Daniel J. Findley, Somayeh Mohammady, Masoud Ardakani, and Kwang Soon Kim, "Hydra-RAN: Multi-Functional Communications and Sensing Networks Applications: Intelligent Parking Systems", IEEE Open Journal of the Communications Society, Early Access, 2025.

[10] Rafid I. Abd and Kwang Soon Kim, "Hydra-RAN: Multi-Functional Communications and Sensing Networks for Collaborative-Based User Status," in Proc. IEEE 14th International Conference, Conf. Information and Communication Technology Convergence (ICTC). Jan. 2025, pp. 11-14.

[11] Rafid I. Abd and Kwang Soon Kim, "Hydra Radio Access Network (H-RAN): Adaptive Environment-Aware Power Codebook," in Proc. IEEE 15th International Conference, Conf. Information and Communication Technology Convergence (ICTC). Jan. 2025, pp. 11-14.

[12] Rafid I Abd, and Kwang Soon Kim, "Hydra Radio Access Network (H-RAN): Accurate Estimation of Reflection Configurations (RCs) for Reconfigurable Intelligent Surfaces (RIS)," in Proc. IEEE 15th International Conference, Conf. Information and Communication Technology Convergence (ICTC). Jan. 2025, pp. 11-14.

[13] Rafid I. Abd, and Kwang Soon Kim, "Hydra Radio Access Network (H-RAN): Multi-Functional Communications and Sensing Networks, Adaptive Power Control, and Interference Coordination," in Proc. IEEE 15th International Conference, Conf. Information and Communication Technology Convergence (ICTC). Jan. 2025, pp. 11-14.