
Ernesto Fontes Pupo ¹, Claudia Carballo Gonzalez ², Eneko Iradier ², Jon Montalban ², Pablo Angueira ², and Maurizio Murroni ²

¹University of Cagliari
²Affiliation not available

December 7, 2023

Abstract

The upcoming 6G will represent a complete paradigm shift for global communications. Addressing the critical research verticals toward the envisioned 2030 will require a compelling mix of enabling radio access technologies (RAT) and native softwarized, disaggregated, and intelligent conceptions such as the Open Radio Access Network (O-RAN) architecture. Integrating the Multicast/Broadcast Services (MBS) capability is an appealing feature to overcome the ever-growing traffic demands, disruptive multimedia services, massive connectivity, and low-latency applications. This article discusses the insertion of machine learning (ML) based multicasting Radio Resource Management (RRM) solutions in the 6G O-RAN. We review the expected evolution of the MBS capability, including enabling technologies and challenges for the IMT-2030 framework. Moreover, we cover essential aspects at the intersection of MBS, ML-based RRM solutions, and the disaggregated O-RAN architecture, identifying possible scenarios as feature extensions of O-RAN. We present the outcomes of a comprehensive MBS use case simulation oriented to validate our approach, highlighting critical remarks and conclusions.

Ernesto Fontes Pupo, Claudia Carballo González, Eneko Iradier, Jon Montalban, Pablo Angueira, Maurizio Murroni

Abstract—The upcoming 6G will represent a complete paradigm shift for global communications. Addressing the critical research verticals toward the envisioned 2030 will require a compelling mix of enabling radio access technologies (RAT) and native softwarized, disaggregated, and intelligent conceptions such as the Open Radio Access Network (O-RAN) architecture. Integrating the Multicast/Broadcast Services (MBS) capability in 6G O-RAN architecure is an appealing feature to overcome the ever-growing traffic demands, disruptive multimedia services, massive connectivity, and low-latency applications. This article discusses the insertion of machine learning (ML)-based multicasting Radio Resource Management (RRM) solutions in the 6G O-RAN. We review the expected evolution of the MBS capability, including enabling technologies and challenges for the IMT-2030 framework. Moreover, we cover essential aspects at the intersection of MBS, ML-based RRM solutions, and the disaggregated O-RAN architecture, identifying possible scenarios as feature extensions of O-RAN. We present the outcomes of a comprehensive MBS use case simulation oriented to validate our approach, highlighting critical remarks and conclusions.

Index Terms—6G, Machine Learning, Multicast/Broadcast Services, O-RAN, Radio Resource Management

I. INTRODUCTION

The envisioned 6G era will represent a complete paradigm shift for global communications, merging the physical, digital, and virtual worlds through immersive human interaction. The 6G baseline network architecture must be extremely disaggregated, scalable, and virtualized, enabling ultra-secure and resilient communications. It will also embrace a green and sustainable approach with suitable energy efficiency, longer life cycles, and less environmental impact [1].

The evolution towards 6G must sustain a hyperconnected world and meet stringent requirements for groundbreaking use cases surpassing the well-known Enhanced Mobile Broadband (eMBB), Ultra-Reliable and Low-Latency Communication (URLLC), Massive Machine Type Communications (mMTC), and Vehicular-to-Everything (V2X). A significant use case will be immersive and advanced experience-sharing communications, including extended reality (XR), holographic communications, and 3D video delivery. Then, 6G will enable extreme communication applications such as autonomous driving, remote telesurgery, mixing robotic technologies, flexible manufacturing, and seamless interaction with immersive applications. Such a wave of multimedia and experience delivery will align with the connected everything paradigm. 6G massive communication implies a hyperconnected resilient network infrastructure with unprecedentedly diverse end devices (e.g., XR equipment, sensors, and cars). Facing these challenging use cases requires meeting advanced capabilities such as 0.1 ms of latency, users’ throughput of up to 1000 Mbps, mobility at high speeds of up to 1000 km/h, a capacity density of 500 Mbps/km² with connection density of up to 10⁸ devices per km². Regarding energy and capacity efficiency, 6G will improve the 5G numbers by up to five and three times, respectively [1].

Addressing the challenging research verticals shaping the 6G context will require a compelling mix of enabling radio access technologies (RAT), innovative technical solutions, and native softwarized/intelligent conceptions. According to [2], the native Multicast/Broadcast Services (MBS) support will be crucial, bringing point-to-multipoint (PTM) delivery mechanisms for efficient resource utilization, load balancing, reliability, overhead, and delay reduction. Nevertheless, optimized Radio Resource Management (RRM) solutions are required to offer a cost-effective PTM service with considerable capacity and quality of service (QoS) gain. The dynamic nature of the prominent MBS use cases and the intrinsic complexity of the involved RAT, including spectrum expansion to the Terahertz (THz) bands, make the RRM a daunting challenge. Consequently, the computational complexity must also be considered a Key Performance Indicator (KPI) during the solution-finding process [3].

Moreover, the 6G conception must incorporate intelligence as an endogenous characteristic to manage ultra-dense HetNets where scalability is critical. This multi-RAT environment will complicate even more the optimum radio resource allocation. In this case, traditional approaches for solving problems, such as heuristic algorithms, become impractical due to complex issues related to high-speed mobility, zero-touch latency, and the diverse nature of users and network segments. In contrast, Machine Learning (ML)-based solutions are suitable for facing various optimization problems, making proactive decisions, and dynamically adapting them to constant network/services/user changes. Deep Reinforcement Learning (DRL) is a valuable prospect thanks to the trial-and-error learning process [4]. ML-based RRM solutions will allow engagement with the trade-off between optimal network performance and complexity [3], [5].

Finally, this intelligent native conception will require flexible architectures such as Open Radio Access Network (O-RAN) to add the softwarization and disaggregation expected in
Multicast/Broadcast capability

- Dedicated carrier for MBMS
- Large OFDM symbol

Foundation

FeMBMS

Rel-15

Expansion

T-Broadcast

Rel-16

5G Baseline

2017-2022

2022-2028

5G Advanced

Rel-18/19/20 +

6G

2027-2030...

O-MBS

Rel-21+

- New and enhanced network functions to support MBMS infrastructure
- Shared and individual traffic delivery
- Compatibility with legacy nodes
- NR MBS enhancements
- Energy efficiency
- E2E AI/ML data driven designs
- Decentralized and softwarized-based elements
- MBS via NTNs
- New and enhanced NR MBS (open framework)
- Extreme disaggregation
- Cloud/core/RAN convergence
- Green communications
- Ultra secure communications
- New sharing paradigms

Figure 1. Multicast/Broadcast evolution toward 6G and main features of each Release. FeMBMS: further evolved MBMS; NTNs: non-terrestrial networks; OFDM: orthogonal frequency-division multiplexing.

6G. The O-RAN framework enables an effective ML closed-loop workflow to dynamically conduct several optimization actions directly impacting QoS and user perception [6].

Recent studies surveyed O-RAN [6]–[8], the RRM, and multicast use cases [9], [10] separately. However, little attention has been devoted to discussing the intersection of MBS, RRM solutions, and O-RAN in the 6G context. This work addresses this gap and proposes the insertion of ML-based multicasting RRM solutions in the 6G O-RAN under a softwarized and intelligent vision. The major contributions of this article can be summarized as follows: (i) An overview of the MBS’s upcoming evolution toward the envisioned Open MBS (O-MBS), enabling technologies, and RRM challenges; (ii) A taxonomy of the distributed O-RAN, including novel deployment scenarios for 6G; (iii) Characterization of ML-based multicasting RRM solutions as potential O-MBS use cases and their insertion in the 6G O-RAN framework; (iv) The evaluation of a particular O-MBS use case aided by terrestrial/airborne connectivity, network slicing, and Federated DRL (F-DRL) to satisfy multiple user petitions with differentiated traffic management.

In the remainder, we cover the intersection of MBS, ML-based RRM solutions, and the disaggregated O-RAN architecture as technical enablers for 6G. First, we comprehensively review the MBS paradigms, enabling O-MBS technologies, and related RRM challenges. Then, we delve into the O-RAN framework and possible scenarios. Third, we characterize the insertion of ML-based multicasting RRM tasks into O-RAN. Eventually, we implement and evaluate an introductory example of an O-MBS use case, including the corresponding discussion and outcomes.

II. Multicast/Broadcast Services: Potentialities & Challenges

The first phases of the 5G standardization (Rel-15 and 16) focus on the solo unicast capability [2]. Then, Rel-16 added the Terrestrial (T)-Broadcast targeting Enhanced Television (EN-TV), and after that, in Rel-17 [9], the 3GPP starts standardizing the novel MBS paradigm for the 5G system architecture. This novel solution allows receiving unicast/multicast/broadcast services simultaneously in the radio resource control (RRC) states connected and idle. It guarantees MBS continuity and lossless handover. Current Rel-18 marks the start of the 5G-Advanced era. It aims to enhance the efficiency in multicast/broadcast resource utilization over heterogeneous networks and sharing scenarios supporting the multicast reception in inactive RRC state [2]. The envisaged New Radio (NR) MBS must include KPIs to reduce computational complexity and increase energy efficiency with ML solutions based on decentralized and softwarized network elements.

The 6G study and conceptualization will start from the time frame of Rel-20 with the first 6G specifications in Rel-21. The 6G era will be characterized by groundbreaking use cases where a novel MBS conception will be prominent in reaching milestones. The envisioned MBS must be aligned with the cloud, core, and RAN convergence through open interfaces (i.e., O-MBS). The new specifications must be oriented to support intelligent solutions in an ultra-dense, disaggregated, and versatile environment. Figure 1 illustrates the expected MBS evolution from the baseline 5G to the upcoming 6G.

A. Enabling MBS Technologies

O-MBS becomes a core element in reaching 6G capabilities through flexible resource utilization, load balancing, reliability, overhead, and end-to-end (E2E) delay reduction. Group-oriented communications allow for efficiently streaming content to large and small areas and offloading popular information to the network edge caching. Multicast/broadcast is identified as the baseline technology for 6G massive vehicular IoT in disseminating early warnings and public safety as a fundamental component of modern transportation systems.

The seamlessunicast/multicast/broadcast convergence as an essential capability of the 6G toolbox requires multiple enabling technologies. These trending features belong to the longer-term new R&D wave toward 6G, summarized in the following:

- Millimeter wave (mmWave) and THz communications with massive MIMO (mMIMO) and beamforming
(BF) is a game-changing technology for delivering high throughput group-oriented services while exploiting the users’ spatial and channel diversity. Moreover, cell-free mMIMO allows extra spatial diversity and co-processing gain by simultaneously and coherently delivering unicast and multicast services through multiple geographically distributed base stations (BSs) [9].

- **Reflective Intelligent Surfaces (RIS)** can reduce blockage effects and improve the reception conditions of the cell-edge or worst channel quality users and the corresponding multicast group (MG) QoS. Furthermore, RIS is considered a sustainable and ecologically friendly solution based primarily on passive components [1].

- **Non-Orthogonal Multiple Access (NOMA)** is essential for future wireless networks’ mixed unicast/multicast/broadcast service delivery [11]. NOMA empowers the network with seamless connectivity, secure transmission strategies, improved spectral efficiency, average fairness, and reduced outage probability.

- **Rate-Splitting Multiple Access (RSMA)** is a flexible and scalable framework to optimize non-orthogonal transmissions and facilitates interference management in heterogeneous environments. It is a promising technology to reduce the latency of future advanced applications [12].

- The proximity technology **Device-to-Device (D2D)** [10] underlaid multicasting (D2DM) is a cost-effective solution for group-oriented communications with users in proximity, reducing latency, handling diversity, enabling alternative links, and extending the coverage.

- The slicing paradigm adds flexibility, prioritization, and isolation by creating several logical network slices (NSs) utilizing a shared physical infrastructure. Based on the application type, the users’ distribution, and network conditions, a shared content flow must be dynamically mapped into unicast/multicast/broadcast slice instances, exploiting network and radio resources economically and efficiently [2], [13].

### III. RRM Challenge Over MBS

Despite the enormous benefits of O-MBS, the dynamic and heterogeneous nature of the multicast/broadcast use cases and the intrinsic complexity of the involved technologies make the RRM challenging. Multicasting in mmWave and THz is complex because directional beams usually cover a small angular area and must be steered toward the right direction, dynamically adjusting the beamwidth, switching to multiple beams, and managing the beams’ gain and power, s.t the users’ distribution. mmWave and THz communications with mMIMO over cell-free networks bring new challenges related to high propagation loss, severe signal blockage attenuation, and limited coverage. It implies a higher probability of throughput impairment for the user with the worst channel condition and complex mobility behaviors, affecting the overall QoS. On the other hand, selecting the cluster header for D2DM implies a concern about the multicast users’ spatial distribution, energy consumption, end-device remaining battery, and social awareness. RRM must lead with proximity

users’ discovery and interference management. Multicasting RRM solutions must address the available multiuser diversity, fulfilling the service requirements without producing unfair resource allocation.

The complexity analysis cannot be solely faced by the intrinsic computational complexity of these technologies and the convergence time to find the best RRM strategies. How often the RRM must recalculate these solutions impacts the network performance, increasing the control plane delay and overhead and affecting delay-sensitive traffic. The mmWave and THz propagation is a decisive aspect triggering the RRM recalculation because as the frequency rises, increases as well, the variation in the users’ reception conditions, signal-to-noise ratio (SNR), and feedback channel quality information (CQI). Other essential factors directly correlated with the variations in the channel conditions are the mobility behavior and speed of the multicast users, which are highly dependent on the propagation frequency.

Figure 2 shows the MG CQI variation for various frequencies and users’ mobility behaviors based on the approach presented in [3]. It illustrates how, as the speed of the MG users increases for different frequencies, the percentage of users that experience a CQI variation rises almost linearly. The results reveal an average CQI variation increase of 2.5 % as the evaluated frequencies go higher. Moreover, the speed increase adds an average extra 5.5 % of CQI variation. In such conditions, the necessity of constant recalculation and the intrinsic convergence time finding the best RRM strategies could become a critical factor from the complexity point of view, increasing the control plane and E2E delay and inducing a critical extra latency in communication. The results were obtained through numerical link-level simulations for a single-cell NR BS, delivering a multicast multimedia service to 50 users with random directional mobility.

The computational complexity must be considered a KPI during the multicast RRM solution-finding. Consequently, the multicast RRM solutions must effectively cope with the trade-off between optimal network performance and complexity [5]. This becomes essential because the upcoming evolution will be tied to the 3D ultra-dense HetNet [2], differentiated advanced
applications and tight requirements, making managing and exploiting network resources even more complex.

IV. THE 6G O-RAN FRAMEWORK

O-RAN is an ML-native framework that uses virtualized and disaggregated elements to conduct dynamic tasks [7]. The slicing paradigm aids this architecture, mapping multiple service types (e.g., services 1-L) into numerous slices (e.g., NSs 1-M) to manage differentiated traffic and ensure the defined Service Level Agreement (SLA). O-RAN disaggregates the BS functionalities into the Open-RAN Control Unit (O-CU), Distributed Unit (O-DU), and Radio Unit (O-RU). The logical split allows these functional units to be flexibly deployed at different network locations and hardware platforms.

Two critical elements in the O-RAN are the Non-Real Time RAN Intelligent Controller (Non-RT RIC) and the Near-Real Time RIC (Near-RT RIC). The former is a Service Management and Orchestration (SMO) element. It implements optimization actions through microservices termed rApps on a time scale superior to 1 s. It trains and updates ML models that will be executed by structures nearer to the end-user (e.g., Near-RT RIC). Moreover, the Non-RT RIC realizes long-term monitoring via the O1 interface and sends A1 policies to the Near-RT RIC to drive end-to-end SLA assurance. The latter executes ML tasks through microservices termed xApps in control loops between 10 ms-1 s. Additionally, it conducts real-time monitoring tasks (E2 interface) to detect whether the performance is out of the target KPIs.

6G involves an ultra-dense heterogeneous environment with massive data. The input/output ML design and how often this data is collected depend on specific optimization problems. Data preparation, including pre-processing, cleaning, and transformations, is an efficient mechanism to homogenize the ML algorithm’s inputs [14]. Following the O-RAN specifications to prevent poor RAN performance or outages, the ML model that has not been previously trained and validated offline cannot be deployed [6]. Once these phases are successfully finished, the resulting trained model is published in the ML Catalog. The ML Catalog also includes under which specific conditions the ML-trained model delivers the best performance. Then, the ML model is deployed into the inference host and fed with data from the environment to execute specific tasks online. During the implementation phase, the ML model previously trained can be fine-tuned and updated based on architectural changes or inefficiencies detected. Continuous operation in the 6G dynamic environment is crucial to improve the previously trained ML models online.

The O-RAN Alliance defines five scenarios detailing the allocation of ML training and inference entities [6]:

- **Scenario 1.1**: Non-RT RIC performs ML training and inference functions.
- **Scenario 1.2**: Non-RT RIC performs ML training, and Near-RT RIC conducts the ML inference.
- **Scenario 1.3**: SMO performs ML training, and Non-RT RIC executes ML inference functionalities.
- **Scenario 1.4**: ML training is performed due to the collaboration between Non-RT RIC and Near-RT RIC. The ML inference is located in the Near-RT RIC.
- **Scenario 1.5**: Non-RT RIC performs ML training, and O-CUs/O-DUs assume ML inference (for further studies).

As the 6G requirements comprise very low latency, high speeds, and massive heterogeneous data, there are certain cases where the ML execution cannot be carried out within the timescale supported by the RICs. Collecting sensitive data through O1 and E2 can negatively impact latency, overhead, and privacy. Then, elements nearer to the end-user must be capable of executing various ML inference tasks in a distributed manner at a timescale below 10 ms (e.g., **Scenario 1.5**). Moreover, as future extensions of the O-RAN specifications, the Near-RT RIC could assume ML training in cases where extensive data arrives from the E2 interface, which may not be available via O1. Thus, we identify three new scenarios that will be fundamental for the future 6G deployment:

- **Scenario 1.6**: Near-RT RIC performs ML training, and O-CUs/O-DUs conduct the ML inference.
- **Scenario 1.7**: ML training is performed due to collaboration between Near-RT RIC and O-CUs/O-DUs. The ML inference is located in the O-CUs/O-DUs.
- **Scenario 1.8**: O-CUs/O-DUs perform ML training and inference functionalities.

Figure 3 details the described O-RAN components, their interfaces, and scenarios deployment. Although not currently supported in O-RAN, the ML inference must also be considered in the O-RU for assignments.
like beamforming. Consequently, the nearer entities to the end user must be dotted with enough resources and open interfaces to assume ML tasks efficiently from different vendors. The power consumption, security, computational complexity, and storage capacity are critical concerns for future 6G ML solutions, sensibly impacting deployment costs and algorithm performance.

V. Multicasting with RRM into the 6G O-RAN Framework

The expected O-MBS solutions can benefit from the O-RAN framework by exploiting enriched environment knowledge and interactions among the architectural elements. O-RAN introduces multiple potentialities for heterogeneous systems, adapting to various use cases, actions, and state spaces. Through O-RAN, multiple multicast tasks can be dynamically disaggregated and executed based on ML and the NS paradigm. ML-aided multicasting RRM can handle complex environments with high mobility, CQI variation, and challenging service constraints where resource optimization is critical with an adequate computational complexity balance. Additionally, the ML life-cycle management guarantees constant monitoring and evaluation of the environment, updating the algorithm if necessary to preserve the long-term system performance.

Table I identifies some primary tasks required for optimum multicasting data delivery and efficient RRM in a THz frequency band. The different functions have been related to the possible ML types [4] for performing these tasks, general inputs/outputs data, and the corresponding O-RAN scenario. This division and task identification is critical for successfully integrating MBS services in an O-RAN framework for the future 6G networks. The proposed considerations do not exclude other solutions where some tasks are divided into multiple subtasks. The final selection depends on specific service providers’ considerations, such as latency, computational complexity, data movement cost, and privacy issues.

The BS/NS selection for shared MBS traffic delivery is one of the ML tasks in Table I. Future wireless networks will evolve into highly complex and ultra-dense heterogeneous systems. Integrating TNs, unmanned aerial vehicles (UAVs), high-altitude platforms (HAPs), and satellite constellations is crucial to achieving the global connectivity promised by 6G systems [2]. This infrastructure can enable multiple multicast/broadcast services for widely distributed users, reaching cost-effectively remote areas. These services can be mapped into numerous NSs to achieve differentiated traffic management. Thus, selecting the most suitable BS/NS in the user’s coverage area for shared MBS traffic delivery is essential to satisfy the SLA and optimize resource utilization.

The task mentioned above can be solved through supervised and DRL solutions, the second option the most often used in the literature. Moreover, combining the Federated Learning (FL) paradigm and DRL (F-DRL) allows multiple local agents to cooperate in building an ML global model based on individual experiences collected [4]. The central unit only receives the ML local parameters and aggregates them to enhance the global model, preserving data privacy and reducing communication overhead compared with traditional ML approaches. This is an appropriate solution for interoperability among telecommunication operators to provide multicast/broadcast services in 6G networks without sharing sensitive information.

In the O-RAN framework, we visualize the BS/NS selection
task inserted into scenarios 1.4 or 1.7. Nevertheless, selecting the specific scenario depends on the network entities’ use case characteristics and computational resources. For example, industrial automation, remote surgery, and XR applications are extremely sensitive to latency [9]. Thus, Scenario 1.7 is recommended for the above use cases to meet tight QoS requirements. The solution must consider the users’ channel/spatial information (i.e., CQI report, mobility behavior), SLA (e.g., user priority, maximum and minimum tolerable QoS values per service), BS/NS parameters, and network load.

VI. O-MBS EVALUATION AND DISCUSSION

This section proposes a framework for integrating MBS services with RRM based on O-RAN. In particular, it presents an MBS use case over a heterogeneous scenario comprising three BSs operating at 28 GHz: one macro-BS and two UAVs acting as aerial BSs extending service coverage. Each BS supports two multimedia multicast services with diverse QoS constraints mapped into different NSs. As shown in Figure 4, each NS has a specific minimum and maximum throughput constraints, i.e., \( NS_1 \): \( Th_{\text{min}} = 30 \) Mbps, \( Th_{\text{max}} = 100 \) Mbps; \( NS_2 \): \( Th_{\text{min}} = 65 \) Mbps, \( Th_{\text{max}} = 340 \) Mbps. We consider 50 pedestrian users randomly distributed in the service area and requesting one of the available services.

The implemented algorithm selects the best BS/NS combination and resource allocation strategy for differentiated multicast traffic delivery to satisfy multiple users simultaneously. The proposal aims to maximize the long-term QoS of all users in the network. Specifically, the optimization targets are throughput, delay, and energy consumption metrics. Our ML solution is based on multicast (M) F-DRL and inserted into the O-RAN framework. Based on the delay requirements of the analyzed services, we consider the O-RAN Scenario 1.4.

The ML local models (one for each BS) are xApps co-located within the Near-RT RIC, which can extract local knowledge through the E2 interface and make individual decisions. Security, Subscription Management, and Conflict Mitigation modules are critical in the independent xApps’ operation and isolation and the accurate Near-RT RIC decision-making [6]. During training, the ML local models cooperate on building an ML global model in the Non-RT RIC without transmitting user-related or safety-critical data. At each federated round, the ML local models are sent to the Non-RT RIC, which aggregates them via the Federated Average (FedAvg) method to obtain an enhanced global model. Next, the updated ML global model parameters are sent back to the local agents, so knowledge earned by all the agents is leveraged for the individual action selection.

We use two benchmark solutions to evaluate the proposed algorithm performance: unicast (U) F-DRL and Max-SINR. The baseline Max-SINR criterion selects the BS that provides the highest reception conditions and considers the NS accessibility without evaluating QoS parameters and available resources. Results were achieved by averaging 30 simulation runs to ensure a 95% confidence interval.

Figure 4 shows the average \( Th \) per slice and different numbers of users. Our solution outperforms the baselines, dedicating only 80 resource blocks (RBs) per BS to deliver the multicast services. The proposed algorithm satisfies all requests, ensuring the \( Th_{\text{max}} \) to 50 users. It maximizes the clients’ perception through effective resource management that profits from the multicast potentialities. In contrast, the evaluated benchmarks using 500 RBs per BS present a similar behavior to our algorithm while they have enough resources to satisfy the users in the network. However, after a certain number of users, these algorithms must apply a load-balancing strategy. They split the resources among active users and guarantee to serve the total service petitions at the expense of affecting the average \( Th \), always superior to the \( Th_{\text{min}} \) according to the SLA. Additionally, using only 50 RBs per BS for multicast, our proposal presents less average \( Th \) than the benchmarks for fewer users. Nevertheless, our algorithm surpasses the baselines as the number of clients increases, demonstrating the effectiveness of multicast service delivery with considerably fewer resources.

Figure 5 shows the cumulative energy consumption (\( Ec \)), expressed in Joule (\( J \)), per slice and for different numbers of users. For \( NS_1 \) and 50 users in the network, our proposal guarantees 46 % and 62 % less system’s \( Ec \) than F-DRL unicast and the Max-SINR criterion, respectively. For \( NS_2 \) and 50 users, our algorithm presents 60 % and 76 % less system’s \( Ec \) than F-DRL unicast and Max-SINR, respectively. Furthermore, even with ten times fewer RBs (i.e., 50 per BS) than the baselines, our proposal surpasses them and ensures...
less $ Ec$ for 50 users in the network.

Results demonstrate that our algorithm correctly learns during multiple trial-and-error interactions with the environment to satisfy numerous users’ petitions according to the SLA. A HetNet deployment with multicast service delivery maximizes the network capacity and optimizes resource utilization, ensuring a superior performance regarding solo unicast capabilities. Moreover, employing an F-DRL algorithm inserted in the envisioned O-RAN framework is a suitable solution to dynamically handle a complex environment with multiple requests and diverse service requirements.

VII. Conclusions

This article discussed the insertion of ML-based multicasting RRM solutions into the disaggregated 6G O-RAN framework. We analyzed the envisioned O-MBS solutions and the importance of a native decentralized, softwarized, and intelligent conception. Moreover, we identified possible scenarios to insert trending multicast ML solutions as feature extensions of the O-RAN specifications. The results of the recreated HetNet use case aided by terrestrial/airborne connectivity proved the advantages of integrating F-DRL, the slicing paradigm, and MBS, enabling efficient differentiated traffic management in the context of future 6G networks.

It becomes crucial for the research community, industry, and network/service operators to work on effectively integrating multicasting ML-RRM into the envisioned O-RAN. These technologies are essential to comply with the stringent requirements of future 6G networks. However, they present several challenges regarding computational complexity, power consumption, security, and storage/processing resources that must be effectively handled.

References


Ernesto Fontes Pupo (e.fontespupo@studenti.unica.it) is a Ph.D. student in the Department of Electrical and Electronic Engineering (DIEE/UdR CNIT), University of Cagliari.

Claudia Carballo González (claudia.carballogonz@unica.it) is a Ph.D. student in the Department of Electrical and Electronic Engineering (DIEE/UdR CNIT), University of Cagliari.

Eneko Iradier (eneko.iradier@ehu.eus) is a Ph.D. Assistant Professor in the Department of Computer Languages and Systems, UPV/EHU.

Jon Montalban (jon.montalban@ehu.eus) is a Ph.D. Associate Professor in the Department of Electronic Technology, UPV/EHU.

Pablo Angueira (pablo.angueira@ehu.eus) is a Ph.D. Full Professor in the Department of Communications Technology, UPV/EHU.

Maurizio Murroni (murroni@unica.it) is a Ph.D. Associate Professor in the Department of Electrical and Electronic Engineering (DIEE/UdR CNIT), University of Cagliari.