QoE-Driven Optimization in 5G O-RAN Enabled HetNets for Enhanced Video Service Quality

Bharat Agarwal[1], Mohammed Amine Togou[1], Marco Ruffini[2], and Gabriel-Miro Muntean[1]

[1]School of Electronic Engineering, Dublin City University, Ireland
[2]Computer Science and Statistic Department, Trinity College Dublin, Ireland
[1]bharat.agarwal2@mail.dcu.ie, mohammedamine.togou@dcu.ie, gabriel.muntean@dcu.ie
[2]marco.ruffini@scss.tcd.ie

***Abstract*—Many innovative applications are projected to be supported by 5G networks across three verticals: enhanced mobile broadband, ultra-reliable low latency communication, and massive machine-type communication. Given the constraints of the current Radio Access Networks (RANs), accommodating all these applications, considering their Quality of Service and Quality of Experience (QoE) requirements, is not practical. Open-RAN is a new architecture touted as the most viable next-generation RAN solution. It promotes a software-defined component, labelled RAN Intelligent Controller (RIC), that governs and supplies intelligence to optimize radio resource allocation, implement handovers, manage interference, and balance load between cells. RIC has two parts: Non-Real-Time (RT) and Near-RT. This article introduces a novel QoE Enhancement Function (QoE^2F) xApp to enhance the functionality of Near-RT RIC through providing efficient resource provisioning to users requesting high-resolution video services. It deploys an innovative Adaptive Genetic Algorithm to perform optimal user association along with resource and power allocation in HetNets. Simulation results demonstrate superior QoE^2F xApp performance in terms of VMAF and MoS for two different resolution videos and diverse numbers of users.**

***Index Terms*—O-RAN, 5G, HetNets, Genetic Algorithm, QoE.**

I. INTRODUCTION

The mobile industry is approaching a critical juncture. Currently, video traffic accounts for 66% of all mobile data traffic, with that figure anticipated to climb 77% by 2026 [1]. The planned increase in network capacity until 2024 lays the groundwork for higher-quality video services, Augmented Reality/Virtual Reality (AR/VR), cloud gaming, and linked consumer wearables. Some applications, on the other hand, necessitate low latency and real-time data processing. As a result, accommodating a plethora of applications with distinct needs entails a flexible network that deploys efficient resource provisioning techniques. Unfortunately, the current Radio Access Networks (RANs) are unable to do so, necessitating network upgrades [2].

Building unique networks that can meet the requirements of particular applications is one possible solution to address this issue. Yet, this is not economically viable from network operators' perspective. Recently, academics and industrial partners have joined hands to build cost-effective mobile networks that are software-driven, virtualised, adaptable, intelligent, and energy-efficient [3]. One promising way to do so is to split the RAN into different network slices, depending on functionalities required. This is known as the Open-RAN architecture, which promotes a software-oriented framework that allows networks to act differently depending on the applications' Quality of Service (QoS) requirements.

A real-life scenario including an Open-RAN 5G Heterogeneous Network (HetNet) is illustrated in Fig. 1. It includes various services such as emergency health (e-Health), entertainment (*e.g.,* high-resolution videos, AR/VR), 5G Connected Homes, and smart wearables. These services have different requirements, including high throughput, low latency, and low energy consumption. Fig. 1 also shows several user equipments (UE) units dispersed at random throughout the city due to temporary events or daily commuting. Periodically, traffic hotspots with many users are formed, putting a significant strain on mobile data access capacity. To meet the high demands, a dense deployment of femtocell base stations (FBS), designated as low power nodes, within the coverage area of macrocell base stations (MBS), labelled as high power nodes, is performed. Note that instead of FBSs, any other non-macro BSs could be used. With different QoS requirements, the network must be flexible to support future upgrades and should be compatible with existing devices to enable HetNets with Open-RAN.

Several studies have been proposed in the literature to improve the users' QoE. For instance, a joint Caching-Power-and-Association approach based on hidden monotonicity was presented in [4]. A Genetic Algorithm (GA)-based discrete power allocation in conjunction with resource block assignment and UE association was proposed in [5] to optimize the QoE of Device-2-Device pairings while still meeting the users' QoE requirements. A heuristic, greedy, low-complexity, and QoE aware resource allocation algorithm was described in [6] that maximizes the mobile network operators' profit while providing high QoE. Similarly, a heuristic-based user association algorithm was presented in [7], which makes decisions based on the capability of BSs to offer acceptable QoE levels to UEs, while maximizing the mobile network operators' profit.

Despite their creativity, these and other existing techniques confront three significant problems. First, they fail to evaluate

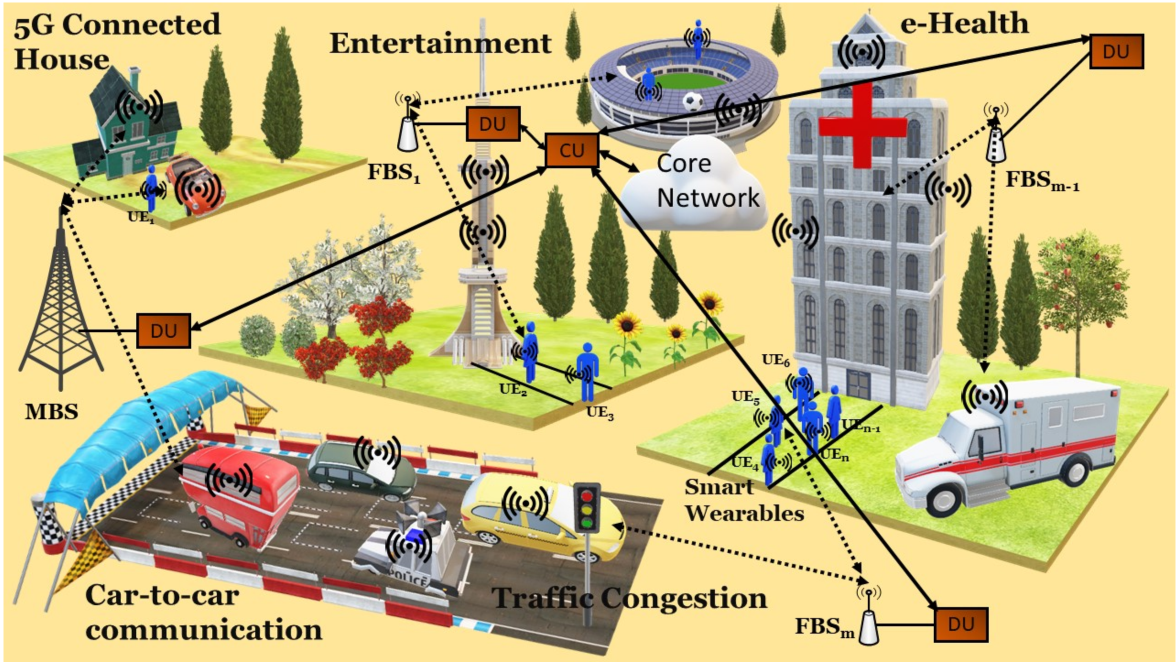


Figure 1. **Open-RAN 5G HetNets:** Corresponds to densification of BSs (MBS, FBS) to support the significant rise in smartphones and various apps with dis-aggregated RAN, i.e., with DU and CU split for the support of different QoS requirements for various applications such as e-Health and entertainment.

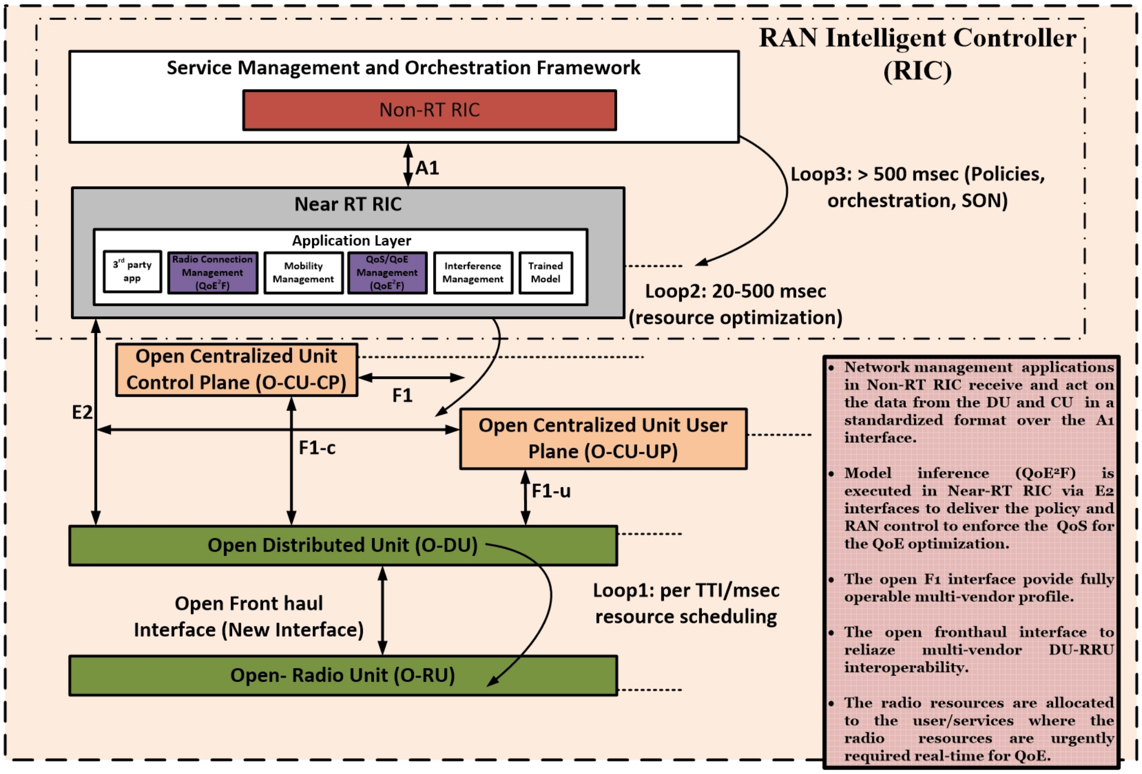
important context variables. Indeed, different context criteria such as maximum BS capacity, transport block size (TBS) index and total BS power capacity should be considered due to rising urbanization and massive UE needs. Second, They incur high computational complexity. As UE demands rise, BS densification rises exponentially, causing the solution space to expand. Third, none of the existing solutions addresses optimal resource management from the Open-RAN's standpoint. Therefore, the main contribution of this article is the design and integration of the QoE^2F application within the Open-RAN architecture to improve users' QoE of video services. The QoE^2F application solves the user association-resource allocation-power allocation (UA-RA-PA) problem with a novel adaptive genetic algorithm (AGA) as a dual '0/1' multiple knapsack problem (MKP).

The remainder of this article is structured as follows. First, we describe the HetNets in 5G along with challenges in 5G HetNets. Next, we present Open-RAN architecture. Then, we introduce the QoE^2F applications and illustrate how it addresses the UA-RA-PA problem using the AGA algorithm. Afterwards, we evaluate QoE^2F performance using a case study focusing on video streaming applications. Finally, we offer our findings and research directions for the future.

II. HETNETS IN 5G

In 5G wireless communications, wireless data speeds, bandwidth, coverage, and connectivity increase and round trip latency and energy consumption decrease. Different forms of communications will have to be enabled by 5G networks, and diverse specifications coming from a wide range of use cases will have to be addressed. There have been many opinions in recent years about the ultimate shape that 5G technology can take. In particular, two views on what 5G wireless technology should be included: 1) Hyper-Connected Vision, promoting a world where unrestricted connectivity enhances people's lives, redefines business, and ushers in a more sustainable future ,and 2) Multi-gigabit per second peak data speeds, ultra-low latency, improved dependability, vast network capacity, increased availability, and a more consistent user experience for a larger number of users will be the foundations of Radio-Access Technology (RAT) in the next decade. For substantial progress to be made, it is important that a definition of the targeted technology is to be agreed on first. In order to satisfy the needs of both the market and the customer, all criteria within the definition process must be met, ensuring that the final definition matches the needs of the majority of users without being overly demanding as in such a case, no framework will function. The following collection of 5G specifications is gaining market recognition by accounting for the majority of current and near-future needs: 1) with real networks, 1-10 Gbps data rates, 2) 1-millisecond latency, 3) 1000x capacity per unit area, 4) Up to 100x the number of connected devices per unit area, 5) 99.999% availability, 6) 100% coverage, 7) 90% network energy savings, 8) Up to 10-year battery life for low-power IoT devices

UEs and BSs are players in 5G HetNets with opposing objectives. On the one hand, each UE must obtain the highest data rate possible to meet its QoS objectives. On the other hand, BSs seek to meet UEs' QoS demands while remaining within their capacity and transmit power limits. The time-variable fading channels in wireless communication and overwhelmed BSs can degrade UEs' QoS (*e.g.,* data rate). To tackle these challenges, decentralized methods are often utilized, allowing UEs to interact with BSs that provide the best channel



- **Network management applications in Non-RT RIC receive and act on the data from the DU and CU in a standardized format over the A1 interface.**
- **Model inference (QoE²F) is executed in Near-RT RIC via E2 interfaces to deliver the policy and RAN control to enforce the QoS for the QoE optimization.**
- **The open F1 interface povide fully operable multi-vendor profile.**
- **The open fronthaul interface to reliaze multi-vendor DU-RRU interoperability.**
- **The radio resources are allocated to the user/services where the radio resources are urgently required real-time for QoE.**

Figure 2. QoE^2F integration with Open-RAN [8].

conditions and meet the minimum QoS requirements. These schemes can also reduce power consumption as they spread users across small cells with high QoS and excellent user QoE.

As demonstrated in Fig. 1, the deployment of various types of small-cell networks, such as FBSs, can be represented as a multi-tier HetNet. Although FBS networks are already in use in present cellular network setups, the future 5G HetNet will have a higher density of these small cells, perhaps allowing each building room to have its own FBS. This is required to meet 5G's high capacity and enormous connectivity requirements. Furthermore, to enable numerous applications, the new HetNet is expected to mix multiple RAT such as 2G, 3G, LTE-Advanced (LTE-A), WiFi, and Device-to-Device communications.

Future mobile devices will be outfitted with various radio interfaces, allowing users to use these RATs and shift between them with ease. However, there are certain caveats in terms of implementation. For example, the 5G HetNet Radio Resource Management (RRM) is expected to be more sophisticated than the previous LTE-A, enabling efficient resource management. In addition, the interference control of HetNets becomes more challenging as privately held buildings or residential properties deploy uncoordinated femtocell networks. For instance, a UE in a HetNet may encounter interference from MBSs, other UEs, and FBSs on a separate tier. If the deployed FBS shares the same spectrum as the MBS, this will exacerbate the problem. Other equally critical difficulties include PA, UA, fairness, allocated capacity, and computational complexity.

III. OPEN-RAN ARCHITECTURE WITH QOE ENHANCEMENT FUNCTION

RANs connect users (*i.e.,* mobile phones and businesses) to the mobile core network using radio waves. They are made of both hardware and software components. Recently, a new RAN architecture, called Open-RAN, has been gaining popularity among researchers and industrial players, given its ability to increase network competitiveness, flexibility, and affordability. Its purpose is to create a multi-supplier RAN solution that enables hardware and software disaggregation, open interfaces and virtualization, and network control and upgrades software housed in the cloud. Open-RAN promotes three key aspects, summarised as follows:

- Cloudification: advocates the conversion of RAN applications to cloud-native functionalities.
- Automation and Intelligence: promotes open management and orchestration through RAN automation interfaces coupled with AI/ML capabilities.
- 3GPP-compliancy: supports 3GPP-defined interfaces (*e.g.,* higher layer split and inter-mode communication) along with new internal interfaces such as O-RAN lower split and near-RT RIC [9].

Fig. 2 depicts the Open-RAN architecture. It has several components, described as follows:

1) ***Service and Management Orchestration (SMO):*** reduces capacity costs by allowing users to scale compute resources up and down based on the infrastructure availability. With continuous integration and development, SMO performs fast and automated life cycle management of virtualized network services and cloud-native capabilities. It also takes care of applications' performance awareness to distribute applications' workload across many distributed cloud resources. Finally, it introduces a closed loop of management, control, and KPI reporting to increase network performance.

*2) **RAN Intelligent Controller (RIC):*** The O-RAN Alliance [2] defines RIC as a logical function in the RAN that controls and delivers intelligence to optimize radio resource allocation, implement handovers, manage interference, and balance load between cells. It consists of a real-time (RT) controller for tasks that require a latency of less than 1 second and a near-RT controller for tasks that require a latency of 1 second or more. Mobile operators can use RIC to install and manage their Open-RAN to ensure: interoperability and vendor variety, predictive and intelligence resource management, and subscriber QoS.

*3) **Non-Real Time RIC:*** is a logical function that allows latency higher than 1 second. It is a micro-service-based software platform that hosts remote applications (rApp). It comes in two flavours: virtual native functions and cloud-native functions.Configuration management, device management, fault management, performance management, and life-cycle management for all network pieces are among the significant aspects of Non-RT RIC.

*4) **Near-Real-Time RIC:*** is a logical function that ensures a one second latency. It is a micro-service-based software platform that hosts extensible applications (xApps), defined as micro-service-based applications that uses standardized interfaces and service models to perform radio resource management by taking data from the RAN (*e.g.,* UEs, MBSs, and FBSs), computing the control actions, and delivering back to UEs, MBSs, and FBSs. Near-RT RIC can be used as a Virtual Native Functions or a Cloud Native Function, and its core roles are: handover management, traffic and radio monitoring in real-time, QoS control, collection and maintenance of historical traffic data, and interaction with Non-RT RIC.

*5) **Distributed Unit (DU):*** introduced by the Third Generation Partnership Project (3GPP) as part of the evolution path toward dis-aggregated RAN. It is software installed on a Customized-off-the-shelf server on-site. The Radio Link Control, Media Access Control, and sections of the Physical layer are all run by DU, usually installed near the Radio Unit on-site. Depending on the functional split choice, this logical node includes a subset of the eNB/gNB functions.

*6) **Centralised Unit (CU):*** runs the Radio Resource Control and Packet Data Convergence Protocol levels along with controlling DU operations. The gNodeB (gNB) is composed of a CU and a DU coupled via Fs-C and Fs-U interfaces for control and user planes. Multiple gNBs can be supported by a single CU with multiple DUs. Due to the split architecture, a 5G network can use varying distributions of protocol stacks between CU and DUs depending on mid-haul availability and network design. CU runs functions such as user data transfer, mobility control, RAN sharing, location, and session management.

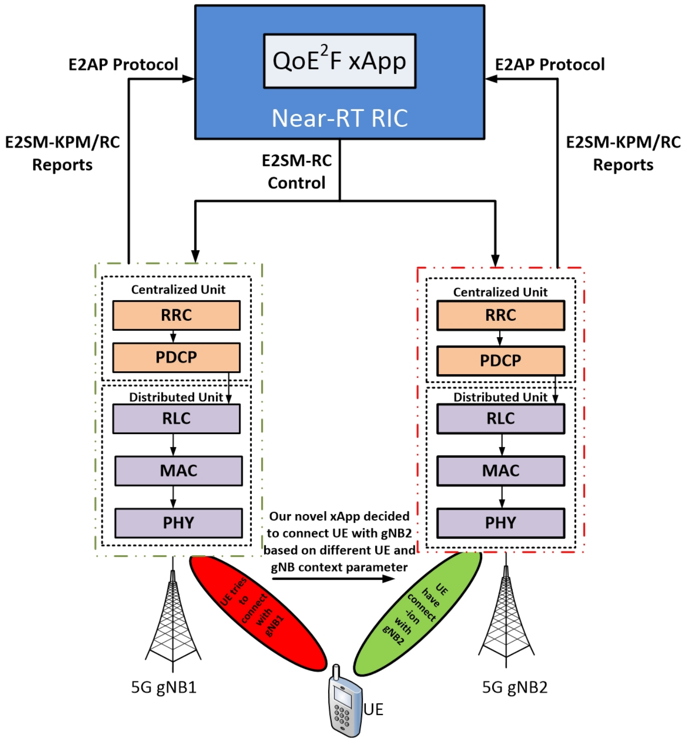


Figure 3. QoE^2F xApp based on E2AP.

To improve the functionality of near-RT RIC for optimal resource allocation, we propose the QoE Enhancement Function (QoE^2F), a xApp that collects information on resource demand and prioritizes users for resource allocation for improved QoE. QoE^2F ensures: 1) optimal access and subscriber session structure to save resources and supply services based on demands and 2) optimal connection between service providers and customers. As illustrated in Fig. 2, the QoE^2F xApp is hosted in Near RT-RIC to control RRM decisions for individual RAN functionalities at near-RT granularities (10ms-1sec).

IV. THE PROPOSED SOLUTION

This section first describes the AGA algorithm then introduces the workflow of the QoE^2F xApp.

A. AGA

QoE^2F has at its core the novel AGA, a decentralised solution to the UA-RA-PA problem, as illustrated in Fig. 3. Our objective is to find a solution that associates each UE with the BS that offers the highest throughput (*i.e.,* high channel quality index, good channel conditions). For that, we decompose the original problem into two sub-problems and solve them using AGA. Both problems are formulated as a dual '0/1' MKP, considering the maximum BS capacity, transport block index, and total BS power capacity restrictions. MKP is broadly described as follows. Given a set of items, each of which has a weight and a value, we need to identify which items to be included in the knapsack so that the total weight is smaller than the capacity of the knapsack while the total value is maximum. A collection of *m* knapsacks with various capacities is given in the case of MKP. The knapsacks considered in this work are represented by BSs (MBS and FBS), and the items to fit into the knapsacks denote the UEs. The item weights are user demands in the UA-RA sub-problem and power consumption in the PA sub-problem. Item values are the available throughput for each knapsack in both sub-problems. The knapsack capacity for the UA-RA sub-problem is the maximum BS capacity and is designated as the total power capacity for the PA sub-problem.

Compared to standard algorithms, the Default Genetic Algorithm (DGA) has intelligence, parallelism, self-organization, expansibility, quick application, and high robustness. As a result, it may be used to solve problems involving combinatorial optimization, such as MKP [10]. However, DGA has limitations such as poor local searchability, high convergence rate, difficulty avoiding local optima, and quick diversification loss [11]. Also, the evolution process with constant crossover and mutation probability makes DGA fall in a local optimum. Therefore, we propose AGA, which extends DGA to include Simulated Annealing, a search algorithm that avoids the local optimum problem when looking for a global solution. Because of its simplicity and convergence, it has become a popular approach for addressing combinatorial optimization problems like MKP. Furthermore, while DGA is prone to premature convergence due to its weak climbing capability, simulated annealing has good asymptotic convergence and speed [12]. As a result, when combined, the resulting AGA algorithm leaps out of its local optimum, speeding up its convergence. AGA involves adaptive crossover probability, adaptive mutation probability, metropolis acceptance criteria and simulated annealing termination condition.

B. *QoE^2F xApp Workflow*

The E2 interface connects two endpoints: the near-RT RIC and E2 nodes (*i.e.,* DUs, CUs, and O-RAN-compliant BSs) [13]. E2 allows RIC to control E2 nodes' procedures and functions, collecting measurements from RAN and feeding them to the near-RT RIC regularly or in response to predefined trigger events. Procedures for control and data gathering can be applied to one or more cells, slices, QoS classes, or particular UEs. E2 Application Protocol (E2AP) and E2 Service Model (E2SM) are the protocols used at the E2 interface [13]. The E2AP is a simple procedural protocol that governs how the near-RT RIC and E2 nodes connect to provide essential services. Different E2SMs, which implement specific capabilities, can be embedded in E2AP messages (*e.g.,* reporting of RAN metrics or controlling RAN parameters).

Using E2AP and E2SM, the QoE improvement use-case is created in the Near-RT RIC via a QoE^2F xApp to control RRM for optimizing UEs' UA-RA-PA decisions. To develop a xApp, two E2SMs have been standardized in O-RAN: E2SM-Key Performance Measurement (E2SM-KPM) and E2SM-RAN Control (E2SM-RC) specifications. The former allows performance metrics at the UE and BS to be computed at near-RT granularities and reported from CU and DU at near-RT intervals. The latter provide UEs with identification and information reporting capabilities. In addition, the E2AP is in charge of interface administration (including E2 interface setup, reset, and problem reporting) and near-real-time RIC service updates (*i.e.,* the exchange of the list of the RAN functions supported by the E2 node). Fig. 4 represents the message flow between the Near RT-RIC and E2 Nodes.

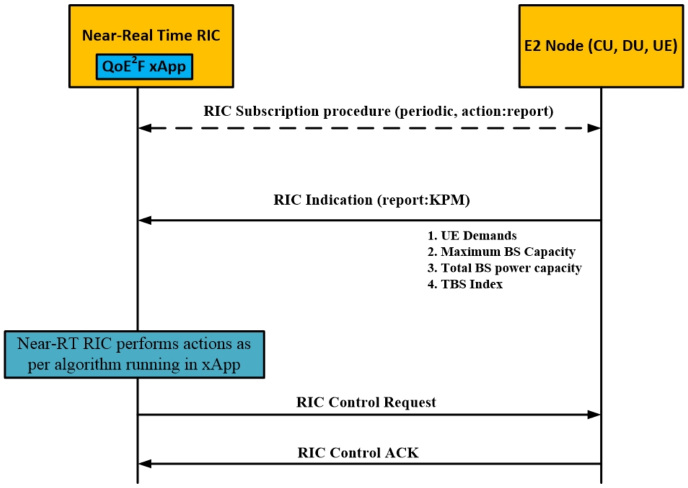


Figure 4. Message Exchange between Near RT-RIC and E2 Nodes.

1) **RIC Subscription Procedure (Periodic, action:report)**: The E2 node provides the metrics it may expose during the E2 setup activities. After then, a xApp in the near-RT RIC can send a subscription message specifying which KPMs are of interest, and frequent reporting will be carried out.
2) **RIC Indication (report:KPM):** To stream the selected KPMs, the E2 node uses E2 Indication messages of type report. Different E2 nodes, such as UEs, CUs, and DUs, create different KPM messages. For example, TBS Index, maximum BS capacity, total BS power capacity, and UE requests are the KPMs that were exchanged.
3) **RIC Control Request:** One of the major goals of our xApp is to optimize the RRM in E2 nodes. Once the xApp receives all the defined KPMs, AGA will be executed, and the desired output will be reported to E2 nodes in a RIC control request message.
4) **RIC Control ACK:** Once the message is received, E2 nodes reply with ACK to the xApp.

V. CASE STUDY: ENTERTAINMENT SERVICES

This section presents a simulation-based performance evaluation of QoE^2F. The case study considered in this paper replicates a real-life scenario where a HetNet consists of 19 fixed BSs (one MBS and 18 FBSs) and randomly placed UEs within the coverage area of these BSs, as depicted in Fig. 1. All BSs are connected to a high-speed backhaul with minimal delay (*e.g.,* fiber optic). We simulate two services with increasing numbers of UEs: video streaming (*i.e.,* a YouTube video with resolution 720p) as it requires high throughput and VoIP, used as background traffic. We denote the number of UEs that request the video service by v and the number of UEs requesting VoIP as o. Given that the coverage area of BSs may

Table I
SIMULATION PARAMETERS

Parameter	Value
Simulation Area	500m x 500m
Number of UEs	40 (v=30, o=10) 70 (v=50, o=20) 100 (v=70, o=30)
UE maximum data rate demand (video)	2.5 Mbps
Total transmit power (MBSs and FBS)	$\{46, 26\}$ dBm
Number of sub-channels (MBSs and FBS)	12 x 100
System Bandwidth	20 MHz
Power per RB (W/RB) (MBS and FBS)	0.39 and 0.0039

overlap, we ensure that each UE can only be associated with at most one BS at any time instance. We use the log-distance path loss model to represent the channel fading between BSs and UEs. We use Network Simulator NS3 and Evalvid framework to run extensive simulations using the parameters listed in Table. I.

We examine video multi-method assessment fusion (VMAF) and mean opinion score (MoS) for the QoE assessment. We compare AGA with two other schemes: default single carrier (DSC) and DGA. In DSC, all UEs connect to a single BS (i.e., macro-cell BS). We have selected DSC as a baseline because we wanted to evaluate the added benefit of the proposed solution in the context of HetNets. DGA was selected for comparison because it is a population-based algorithm that finds a sub-optimal solution. DGA has also limitations that include poor local searchability, high convergence rate, difficulty avoiding local optima, and the proposed QoE2F addresses these shortcomings.

Fig. 5 illustrates the QoE evaluation based on VMAF, denoted as V, and MoS, denoted as M (The mapping of VMAF to MOS is done based on: $MOS\_Score = (1 + (4 * VMAF\_Score)/100)$. Thus, the range of VMAF score is 0-100, and the range of MOS is 1-5.). Two 720p resolution videos categorized as High Spatial Low Temporal (HSLT) (*e.g.,* Cricket Match Video)), Low Spatial High Temporal (LSHT) (*e.g.,* Moving Bus Video) have been considered for the evaluation. They have bitrates of 3261 and 2716 Kbps with a frame rate of 25 and 50 fps, respectively. Fig. 5(a) represents the VMAF Score and Mean Opinion Score (MoS) for different UE densities scored when HSLT video was streamed under three different schemes. Under QoE^2F, for 30 users, the average VMAF score was 99.52 mapped to 4.98 MoS, suggesting that the video quality perceived by users was excellent. The same trend was observed for the remaining UE densities. Under DGA, the video quality remains excellent for 30 and 50 UEs, but drops to fair for 70 UEs as the VMAF score was around 60.6, which is 35% less than what was achieved by QoE^2F. Note that for DSC, the video quality perceived by UEs was always poor, and users could not receive the video when the number of UEs was set to 70. The poor performance of the DSC scheme shows the importance of HetNets in 5G and how QoE^2F outperformed the other schemes even at high UE densities.

Fig. 5(b) depicts VMAF and MoS scores for different UE densities when LHST video was streamed under the three different schemes. Using QoE^2F, the average VMAF score for 30 users was 60.61 mapped to 3.42 MoS, suggesting that the video quality perceived by users was fair. The same trend was observed for the remaining UE densities. Under DGA, the video quality was poor for all the UE densities. For DSC, the perceived video quality was poor for 30 and 50 UE densities, and users were not able to receive the video when the UE density was set at 70. These results show how QoE^2F outperforms the other two schemes considered in terms of various QoE metrics and considering different UE densities, hence enhancing the users' visual experience. QoE^2F performs much better than DGA due to the selection of UEs with high channel quality indices while incurring low execution time. A marginal drop was observed in the perceived video quality at higher UE density because of an increase in load on BSs. In addition, QoE^2F seems to be more suitable for streaming LHST videos in comparison to HSLT videos because of the high frame rate. Note that when streaming high frame rate videos, selecting the best network is not enough to ensure excellent perceived video quality from these UEs. Though integrating Dynamic Adaptive Streaming Over HTTP (DASH) with our approach could increase the quality even further, this is outside the scope of this paper.

VI. CONCLUSIONS

This article discusses optimal user association (UA) to base stations, resource allocation (RA) during service distribution, and power allocation (PA) in the context of 5G HetNets. It introduces a novel QoE^2F xApp, which enhances the Near RT-RIC functionality of the 5G O-RAN system architecture. The QoE^2F xApp caters to optimal UA-RA-PA to enhance the user QoE for two different categories of videos, *i.e.,* HSLT and LHST. This is achieved through the use of an innovative adaptive genetic algorithm (AGA) that runs within QoE^2F App. AGA breaks the UA-RA-PA problem into two sub-problems and solves them as a multi-knapsack problem by putting constraints on total BS capacity, TBS index, and BS power capacity.

We believe that QoE^2F xApp offers some gains over the state-of-the-art as it enables both UEs and BSs to achieve their specific needs. From a user perspective, this means getting services at high data rates and quality levels.From the mobile network operators perspective, this implies meeting the performance targets indicated in service level agreements while also operating efficiently. Simulation results show that QoE^2F provides excellent quality for HSLT videos while maintaining good quality levels for LHST videos.

The following potential research directions can be pursued in relation to our proposed QoE^2F xApp.

- **Real-time control:** The optimization and control of O-RAN is performed at a coarse timescale (i.e. greater than 10 milliseconds). Future evolutions of O-RAN will need to integrate specific real-time components, such as for example, dApps [14], enable dynamic real-time data-driven control and optimization of O-RAN. These elements can collaborate with xApps to use data that

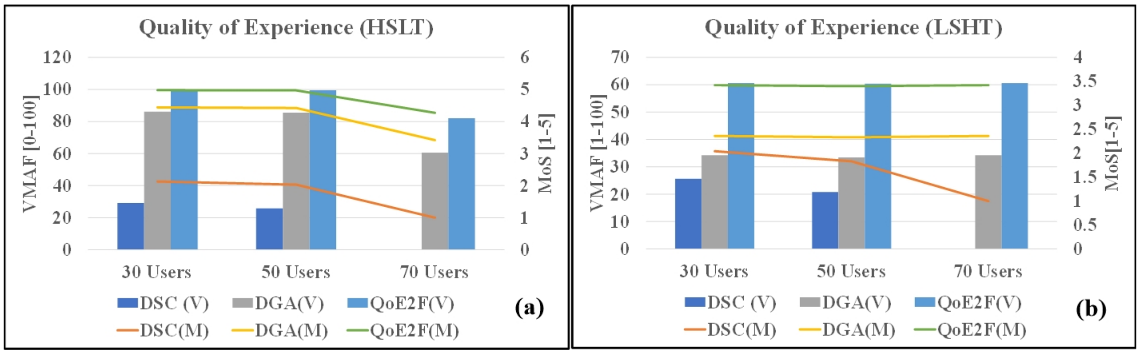


Figure 5. Evaluation of QoE (a) for different delivery schemes employed for 30, 50 and 70 HSLT video users, (b) for different delivery schemes employed for 30, 50 and 70 LSHT video users

cannot be transported from the RAN to RIC for analysis (e.g., (In-phase and Quadrature Phase (I/Q) samples or fine-grained channel data).

- **Effective AI/ML Algorithms design, testing, and deployment:** The AI/ML workflow puts O-RAN in a position to serve as a framework for using ML solutions in the RAN. While this process is being standardized, there are still a number of difficulties. These difficulties stem from (i) the need to gather, train, and test heterogeneous datasets that are indicative of large-scale deployments; and (ii) test and improve data-driven solutions through online training.

VII. ACKNOWLEDGEMENTS

This work received financial support from the Science Foundation Ireland (SFI) via grants 18/CRT/6224 (SFI Centre for Research Training in Digitally-Enhanced Reality D-REAL), 16/SP/3804 (Enable) and 12/RC/2289_P2 (Insight).

REFERENCES

[1] “Open RAN explained: Openness, innovation and flexibility” Ericsson [Online]. Available: https://www.ericsson.com/en/openness-innovation/open-ran-explained. [Accessed: 26-Jan-2022].

[2] S. K. Singh, R. Singh and B. Kumbhani, “The Evolution of Radio Access Network Towards Open-RAN: Challenges and Opportunities,” *2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW),* 2020, pp. 1-6.

[3] A. Marotta, D. Cassioli, K. Kondepu, C. Antonelli and L. Valcarenghi, “Exploiting flexible functional split in converged software defined access networks,” *in IEEE/OSA Journal of Optical Communications and Networking,* vol. 11, no. 11, pp. 536-546, November 2019.

[4] C. Zou, H. Lu and J. Meng, “QoE-Driven Optimization for Video Services in Cache-Enabled Software-Defined HetNets,” *2019 11th International Conference on Wireless Communications and Signal Processing (WCSP)*, 2019, pp. 1-7.

[5] J. Chen, Y. Deng, J. Jia, M. Dohler and A. Nallanathan, “Cross-Layer QoE Optimization for D2D Communication in CR-Enabled Heterogeneous Cellular Networks,” *in IEEE Transactions on Cognitive Communications and Networking,* vol. 4, no. 4, pp. 719-734, Dec. 2018.

[6] P. Trakas, F. Adelantado and C. Verikoukis, “QoE-Aware Resource Allocation for Profit Maximization Under User Satisfaction Guarantees in HetNets With Differentiated Services,” *in IEEE Systems Journal,* vol. 13, no. 3, pp. 2664-2675, Sept. 2019.

[7] P. Trakas, F. Adelantado, N. Zorba and C. Verikoukis, “A quality of experience-aware association algorithm for 5G heterogeneous networks,” *IEEE Int. Conf. on Communications (ICC),* 2017, pp. 1-6.

[8] “Ran Software Community,” *O*, 31-Aug-2021. [Online]. Available: https://o-ran-sc.org/. [Accessed: 29-Dec-2021].

[9] C.-L. I and S. Katti, “O-RAN alliance introduces 48 new specifications released since July 2021” - O-RAN Alliance, 22-Oct-2021. [Online]. Available: https://www.o-ran.org/blog/2021/6/23/o-ran-alliance-introduces-33-new-specifications-released-since-march-2021-fdxtb. [Accessed: 06-Feb-2022].

[10] B. Agarwal, M. Ruffini and G. -M. Muntean, “Reduced Complexity Optimal Resource Allocation for Enhanced Video Quality in a Heterogeneous Network Environment,” in *IEEE Transactions on Wireless Communications*, vol. 21, no. 5, pp. 2892-2908, May 2022.

[11] B. Agarwal, M. A. Togou, M. Ruffini, and G.-M. Muntean, “Joint Performance-Resource Optimization for Improved Video Quality in Fairness Enhanced HetNets,” *IEEE International Conference on Communications (ICC),* Seoul, South Korea, 2022 pp. 1-6.

[12] B. Agarwal, M. A. Togou, M. Ruffini, and G.-M. Muntean, “PIRS3A: A Low Complexity Multi-knapsack-based Approach for User Association and Resource Allocation in HetNets,” *IEEE Intl. Conf. on Communications Workshops (ICC Workshops),* Montreal, Canada, 2021 pp. 1-6.

[13] O-RAN Working Group 3, “O-RAN Near-Real-time RAN Intelligent Controller Architecture & E2 General Aspects and Principles 2.00,” O-RAN.WG3.E2GAP-v02.01 Technical Specification, July 2021.

[14] S. D’Oro, L. Bonati, M. Polese, and T. Melodia. ”OrchestRAN: Network Automation through Orchestrated Intelligence in the Open RAN.” *arXiv preprint arXiv:2201.05632* (2022).

Bharat Agarwal received the B.Eng. degree in Electrical and Electronics Engineering from Galgotias University, India, in 2016. He was a Project Scientist at the Department of Telecommunications, India Institute of Technology, Delhi. He is pursuing the PhD degree with the School of Electronics Engineering, Dublin City University (DCU), Ireland. His research interests include heterogeneous networks, traffic optimization, 5G, LTE, convex optimization. He received a SFI PhD scholarship from the DCU CRT D-REAL Centre.

Mohammed Amine Togou is a postdoctoral researcher at Dublin City University’s SFI Insight Centre for Data Analytics. He holds a B.S. and M.S. in computer science and computer networks from Al Akhawayn University in Ifrane, Morocco, as well as a Ph.D. in computer science from the University of Montreal in Canada. 5G/6G networks, blockchain, machine learning, and IoT are some of his current research interests.

Marco Ruffini is a Trinity College Dublin Associate Professor and Fellow. He is a pioneer in the convergence of fixed-mobile and access-metro networks, and virtualization of next-generation networks, in his core area of research. He gives keynotes and lectures at a number of important international conferences throughout the world.

Gabriel-Miro Muntean is a Professor with Dublin City University’s School of Electronic Engineering in Ireland. Quality, energy, and performance-related challenges of rich media content delivery are among his research interests. Prof. Muntean is an IEEE Transactions on Broadcasting Associate Editor and the IEEE Communication Surveys and Tutorials’ Multimedia Communications Area Editor. He coordinated the EU Horizon2020 project NEWTON and leads the DCU team in the EU Horizon2020 project TRACTION. (Contact: gabriel.muntean@dcu.ie).