



**NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS**

**SCHOOL OF SCIENCE  
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATION**

**MSc THESIS**

**5G/6G Radio Access:  
Study, Deployment, and Testing Using O-RAN**

**Mahmoud – E – ESAWI BABIKER OSMAN**

**Supervisors: Dr. Nikos Passas  
Dr. Dimitrios Tsolkas**

**ATHENS**

**September 2021**

## **MSc THESIS**

5G/6G Radio Access: Study, Deployment, and Testing Using O-RAN

**Mahmoud – E – Esawi Babiker Osman**

**S.N.: 1200004**

**SUPERVISORS:**   **Dr. Nikos Passas**  
                          **Dr. Dimitrios Tsolkas**

## **ABSTRACT**

A run through 5G, O-RAN architecture, and AI/ML enabled O-RAN use cases are presented in this work. Due to 5G denser site deployments and the advanced features it had incorporated (e.g., mMIMO, Beamforming, mmWave, ...) which demanded stricter requirements on the radio network optimization tasks, intelligence was introduced into RAN. For that reason and more, open RAN movement emerged, and O-RAN alliance specified an O-RAN architecture that can enable AI/ML into RAN. As a result of this, numerous use cases and ways to improve network functionality were unlocked. The thesis presented, a novel AI/ML enabled ANR system architecture, that would benefit from the central intelligence characteristic of O-RAN architecture to achieve an ANR system that can outperform current systems. The novel design is vendor and RAT agnostic, is network aware, and reacts to network changes in real-time. It can also help detect abnormalities in the network performance that can't be detected by normal monitoring systems. According to our best knowledge, the presented approach is the first one that aims to achieve these characteristics.

**SUBJECT AREA:** Mobile Wireless Networks

**KEYWORDS:** RAN, O-RAN, 5G, ANR

*I dedicate my dissertation work to my family. To my parents, Esawi, and Huda, and to my amazing brother Ahmed. Their words of encouragement and support pushed me to do better and be the person I am today.*

## **AKNOWLEDGMENTS**

This Master Thesis has been accomplished in the framework of the European Funded Project: SMART Telecom and Sensing Networks (SMARTNET) - Erasmus+ Programme Key Action 1: Erasmus Mundus Joint Master Degrees – Ref. Number 2017 – 2734/001 – 001, Project number - 586686-EPP-1-2017-1-UK-EPPKA1-JMD-MOB, coordinated by Aston University, and with the participation of Télécom SudParis, member of IP Paris and National and Kapodistrian University of Athens.

# CONTENTS

<b>1. INTRODUCTION .....</b>	<b>11</b>
<b>2. VISION FOR 5G AND BEYOND .....</b>	<b>13</b>
<b>2.1 5G Network Architecture and Characteristics .....</b>	<b>13</b>
2.1.1 5G Spectrum .....	14
2.1.2 New Air Interface Technologies .....	15
<b>3. INTRODUCTION TO OPEN RAN .....</b>	<b>16</b>
<b>3.1 Open RAN Movement .....</b>	<b>16</b>
<b>3.2 O-RAN &amp; 3GPP .....</b>	<b>16</b>
<b>3.3 O-RAN, Software and Hardware Disaggregation.....</b>	<b>17</b>
<b>3.4 O-RAN Architecture .....</b>	<b>17</b>
3.4.1 O-RAN elements .....	17
3.4.2 Control Loops .....	18
3.4.3 NR Functional Split .....	19
3.4.4 Implementation Options of O-RAN functions .....	22
3.4.5 RAN Intelligent Controller (RIC) .....	22
<b>3.5 Virtualization, CI/CD, O-Cloud .....</b>	<b>23</b>
<b>3.6 O-RAN Benefits and Use Cases .....</b>	<b>25</b>
3.6.1 O-RAN benefits .....	25
3.6.2 O-RAN use cases .....	26
<b>3.7 Global O-RAN Adoption and Challenges .....</b>	<b>26</b>
3.7.1 Market Potential and Current O-RAN Deployments .....	27
3.7.2 Main Challenges Facing O-RAN .....	28
<b>4. O-RAN AND AI/ML .....</b>	<b>29</b>
<b>4.1 Review of ML .....</b>	<b>29</b>
4.1.1 Supervised Learning .....	29
4.1.2 Unsupervised Learning .....	29
4.1.3 Reinforcement Learning .....	30
4.1.4 Federated Learning .....	31

<b>4.2</b>	<b>Mapping AI/ML functionality in O-RAN.....</b>	<b>31</b>
<b>4.3</b>	<b>AI/ML General Procedure and Interface Framework .....</b>	<b>31</b>
4.3.1	Data, Model Training and Model Evaluation pipeline .....	33
4.3.2	ML Model Lifecycle Implementation.....	34
4.3.3	Deployment Scenarios .....	34
<b>4.4</b>	<b>Adoption of AI/ML in O-RAN.....</b>	<b>36</b>
<b>5.</b>	<b>AUTOMATIC NEIGHBOR RELATIONS (ANR) OPTIMIZATION .....</b>	<b>37</b>
<b>5.1</b>	<b>Introduction to Neighbor Relation Optimization .....</b>	<b>37</b>
5.1.1	Importance of NR .....	37
<b>5.2</b>	<b>Handover and NR Measurements .....</b>	<b>37</b>
5.2.1	NR Process .....	38
<b>5.3</b>	<b>From Manual to Automatic NR Optimization .....</b>	<b>39</b>
5.3.1	SON.....	40
<b>5.4</b>	<b>O-RAN ANR Optimization .....</b>	<b>41</b>
<b>6.</b>	<b>PROPOSED NOVEL AI/ML ENABLED ANR SYSTEM ARCHITECTURE .....</b>	<b>43</b>
<b>6.1</b>	<b>Motivation .....</b>	<b>43</b>
<b>6.2</b>	<b>Proposed Design .....</b>	<b>43</b>
6.2.1	Initial Operation Phase .....	43
6.2.2	Main Operation Phase .....	44
<b>6.3</b>	<b>Advantages of the Novel Design.....</b>	<b>45</b>
<b>6.4</b>	<b>Challenges and Future Work .....</b>	<b>47</b>
<b>7.</b>	<b>CONCLUSION .....</b>	<b>48</b>
	<b>ABBREVIATIONS - ACRONYMS.....</b>	<b>49</b>
	<b>ANNEX I.....</b>	<b>50</b>
	<b>REFERENCES .....</b>	<b>51</b>

## LIST OF FIGURES

Figure 1: 5G and beyond usage scenarios [1] .....	13
Figure 2: 5G system architecture overview [2].....	13
Figure 3: NR interfaces overview [2].....	14
Figure 4: 5G Spectrum ranges .....	14
Figure 5: Beam scanning and broadcast beamforming .....	15
Figure 6: Combining 3.5GHz in DL with 1.8GHz in UL .....	15
Figure 7: "Some" of the O-RAN Alliance members [4].....	16
Figure 8: Open RAN software and hardware disaggregation .....	17
Figure 9: Logical architecture of O-RAN [6].....	18
Figure 10: Control loops in O-RAN [6] .....	19
Figure 11: A downlink view for the functional splits [8] .....	20
Figure 12: F1/W1/E1/X2/Xn Interfaces .....	20
Figure 13: FHM mode usage for shared cell deployment [6] .....	22
Figure 14: Cascade mode usage for shared cell deployment [6].....	22
Figure 15: Centralized Near-RT RIC Serving 4G and 5G Simultaneously [6].....	23
Figure 16: Distributed Near-RT RIC [6] .....	23
Figure 17: High-level NFV architecture [12].....	24
Figure 18: Cloud Deployment Scenarios [13] .....	25
Figure 19: O-RAN market growth projections 2020-2026 [17].....	27
Figure 20: Supervised learning/unsupervised model training and actor locations [21] ..	30
Figure 21: Reinforcement learning framework.....	30
Figure 22: Reinforcement learning model training and actor locations [21] .....	30
Figure 23: Learning-based closed-control loops in an O-RAN architecture [24].....	31
Figure 24: ML training host and inference locations [21] .....	32
Figure 25: O-RAN Spec. - Chained modular models with common inputs [21] .....	33
Figure 26: O-RAN Spec. - Data pipeline [21].....	33



Figure 27: O-RAN Spec. - Model training and evaluation pipelines [21].....	34
Figure 28: O-RAN Spec. - ML model lifecycle (an implementation example) [21] .....	34
Figure 29: O-RAN Spec. - Deployment scenario 1.1 [21] .....	35
Figure 30: O-RAN Spec. - Deployment scenario 1.2 [21] .....	35
Figure 31: Examples of image based and file-based ML model deployment [21].....	36
Figure 32: NR ranking based on MR data (manual NCL update) .....	38
Figure 33: NR ranking based on PM data (manual NCL update) .....	39
Figure 34: SON location .....	40
Figure 35: The ANR System during Initial Operation Phase.....	43
Figure 36: Basic ANR Operation Flow.....	44
Figure 37: The ANR System during Main Operation Phase .....	45
Figure 38: Example of NCL .....	46
Figure 39: Example of Relation level issue.....	46

## LIST OF TABLES

Table 1: AI/ML deployment scenarios [21] .....	32
Table 2: Summary on characteristics of different CU-DU split option [11] .....	50

## 1. Introduction

The world demand of data and connectivity is exponentially increasing and this led to researchers to go from 1G (First Generation networks) all the way to 5G trying to find better and more efficient networks. With each of these new techniques and generations, new challenges arise and more research is needed, which led to the emergence of open RAN (Radio Access Network) as a way to tackle some of these challenges and achieve a better performance. In this research, a study of how this style of network came to be, their deployment, and use cases.

The first section of the thesis includes a brief description of 5G networks and their characteristics. It will show the 5G architecture, the air interface techniques used, and all the new challenges that have risen from that.

In the following section, the open RAN movement and the emergence of O-RAN alliance will be introduced. It will cover the need for O-RAN specifications and the O-RAN architecture in detail, by going through the O-RAN elements, the control loops in the architecture, how O-RAN can split the RAN functionality between the different elements, and how virtualization techniques are used in O-RAN. The section will also introduce RICs (Radio Intelligent Controllers) and the use cases and benefits achieved by using intelligence in O-RAN networks. Lastly, the section will show the market potential for O-RAN, the current O-RAN deployments, and the challenges facing O-RAN today.

The section after that will dig deeper in how intelligence has been added to the RAN. It will first give a brief review of machine learning (ML) in general and then describe how these procedures will be mapped into O-RAN architecture and their different deployments scenarios. It will also show few of the use cases that benefited from AI/ML adoption in O-RAN.

The last two sections will focus on automatic neighbor relations (ANR) optimization, which is one of the main use cases of applying AI/ML into RAN networks. One of the sections will cover a literature review on ANR, by first describing the need for neighbor relation lists and how they are used within the handover procedure. Also the section will describe how ANR was achieved in traditional RAN networks in both manual and automatic systems. Then, it will show how ANR can benefit from ANR architecture and current systems and their limitations.

Lastly, in the final section, a novel AI/ML enabled ANR system architecture will be introduced. The section will cover why do we need such a novel approach, a description of how the system works, and the expected gains. It will show how the system will take advantage of the intelligence in O-RAN architecture to bring out the most value of it. In the end a brief run through what kind of challenges that this novel architecture deployment might face, and some suggestions on future work to tackle these challenges.

## 2. Vision for 5G and Beyond

In Sep-2015 ITU-R published its vision for 5G and beyond, which defined three main application scenarios: Enhanced Mobile Broadband (eMBB), Ultra-reliable and low latency communications (uRLLC), and Massive Machine Type Communications (mMTC) (seen in Figure 1).

This is achieved by introducing a new RAN (Radio Access Network) and a new network architecture and applying new techniques into the mobile networks, such as: software defined networking (SDN), network function virtualization (NFV), and slicing.

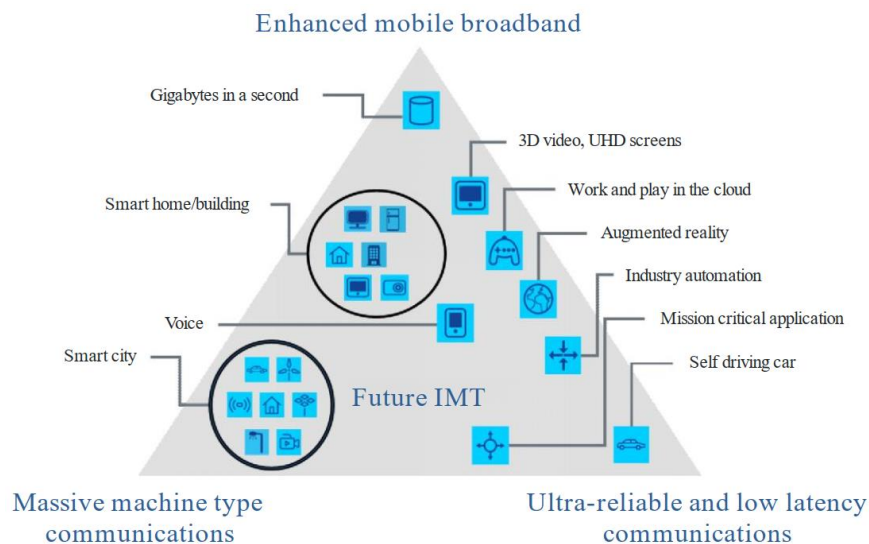


Figure 1: 5G and beyond usage scenarios [1]

### 2.1 5G Network Architecture and Characteristics

5G network consists of the following components:

- Core network: Next Generation Core (NGC) / 5G Core (5GC)
- Wireless network: New RAN (NR)

Service-Based Architecture (SBA) framework is applied in 5GC architecture, which resulted in approach with modularity and reusability. This is done by defining network functions (NFs) to replace the traditional network nodes. As seen in Figure 2, By interfacing these NFs into a common framework, they can offer services to the other NFs or any consumer [2].

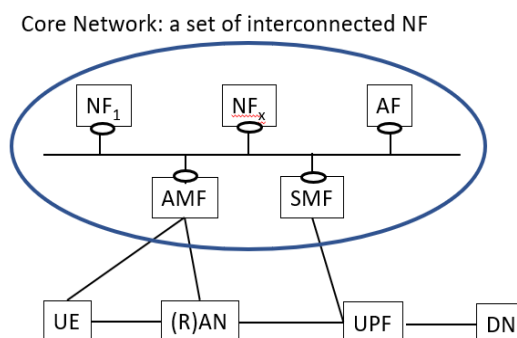
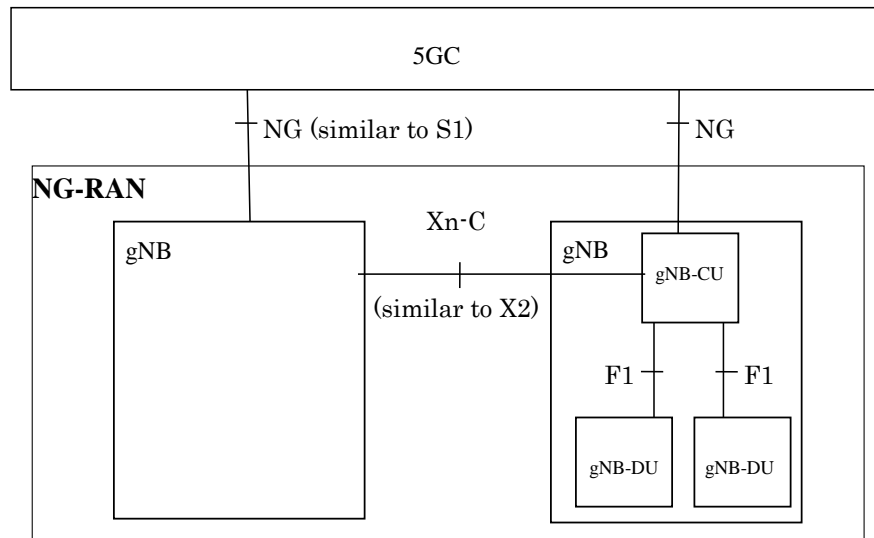


Figure 2: 5G system architecture overview [2]

In the other hand the architecture of the NR is just a simple single entity -the gNodeB- which is connected to the 5GC via the NG interface as seen in Figure 3.



**Figure 3: NR interfaces overview [2]**

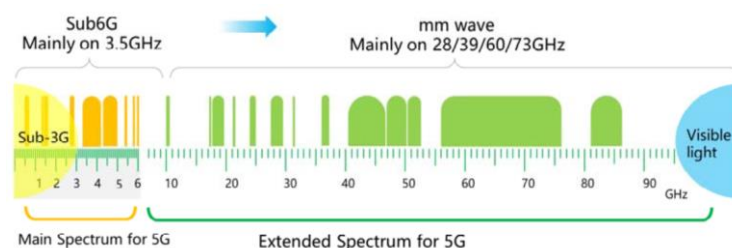
Taking advantage of the split architecture of the gNB, the network can utilize different distribution of protocol stacks between the Central Unit (CU) and the Distributed Units (DU) depending on the mid-haul availability and the use case (Further details on this in Section 3.4.3).

### 2.1.1 5G Spectrum

To achieve the high performance of 5G needed for the application scenarios mentioned earlier, new spectrum was needed.

3GPP divided 5G spectrum into two main frequency ranges (FR) [2]:

- FR1: This sub 6GHz frequency band is the main 5G band for coverage.
- FR2: Millimeter-wave (mm-Wave) band, is the extended band for 5G and would provide the capacity.



**Figure 4: 5G Spectrum ranges**

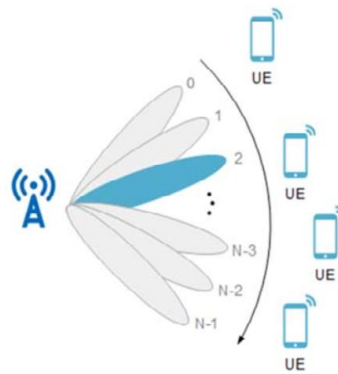
These higher frequencies will require denser base station deployments which will make the performance of the network in terms of mobility management, user equipment (UE) camping band selection and load balancing between the different cells is much more critical. Furthermore, the handover between the different radio access technologies (RAT) (Inter-RAT handovers) to 4G and 2G when it is necessary will be more challenging to optimize.

Also, because of the crowded spectrum -specially for the sub6G band- spectrum sharing and management would be of the most importance. Operators will need to prioritize defragmenting and clearing prime bands of the spectrum.

### 2.1.2 New Air Interface Technologies

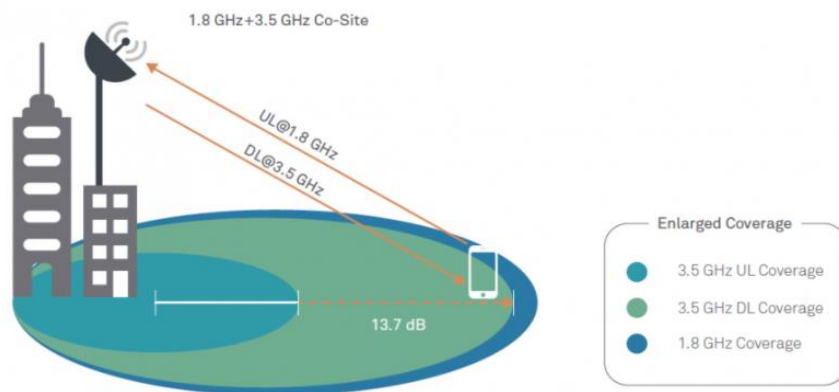
The NR adopted a number of technologies to achieve the flexibility needed to facilitate the diverse new applications of 5G, but because these features need to be optimized per slice/user to achieve the required service-level agreements (SLAs) their optimization is becoming even more critical.

For example, 5G mMIMO expands the legacy systems by increasing the number of antennas to 64T64R. It unlocked advanced applications like 3D beamforming -which provided better services for top floors of high building or for UAVs) and broadcast beamforming (Figure 5).



**Figure 5: Beam scanning and broadcast beamforming**

Another example, is the introduction of uplink (UL) and downlink (DL) decoupling to mitigate the UL & DL coverage imbalance and the UL coverage limitations when using high frequencies (mm-wave). The idea of switching the UL transmission to a different band (e.g., 1.8 GHz) will improve the coverage, but only if it was optimized properly.



**Figure 6: Combining 3.5GHz in DL with 1.8GHz in UL**

Because of the spectrum management, denser site deployments, and the stricter requirements to optimize the radio network features with all these new techniques, introducing intelligence in RAN was proposed as a good method to mitigate some of these challenges.

### 3. Introduction to Open RAN

#### 3.1 Open RAN Movement

A need for a fresh look into the RAN ecosystem is becoming evident as a way to meet the expectations of 5G with its diverse applications and challenging latencies, speeds, number of users, and lack of spectrum. Hence the push towards the idea of deploying intelligent networks that have flexibility in its different logical functions and can be placed at different locations and be coordinated to power 5G applications.

Open RAN is a movement to disaggregate hardware and software in mobile radio access networks and create open interfaces between the different entities of the network. But for this to come into reality and achieve inter-operability standardizing the open RAN operations is crucial [3].

There are many groups and bodies who are involved in different aspects of Open RAN and trying to promote the new approach in next generation RANs, in this thesis work the focus will be on what's recommended and specified by O-RAN alliance.

In February 2018, O-RAN Alliance was founded, which is a group of vendors, operators and academic contributors that is leading the Open RAN movement and is producing standards to allow an inter-operable Open RAN deployment. It was formed by a merger between two different organizations, the C-RAN Alliance and X-RAN forum.



Figure 7: "Some" of the O-RAN Alliance members [4]

#### 3.2 O-RAN & 3GPP

One of the most popular and valid questions is "why do we need O-RAN when we already have 3GPP?". The main thing to point out, is that O-RAN specifications are complementary to the 3GPP Standards.

The 3GPP standard based interfaces are supposed to be open in traditional RAN, but actually the software and interfaces are either proprietary or closed by tying it to an underlying hardware by the same vendor. And the important role of bringing the industry to work together and implement a more flexible, virtualized, and intelligent approach is being championed by organizations like the O-RAN Alliance.

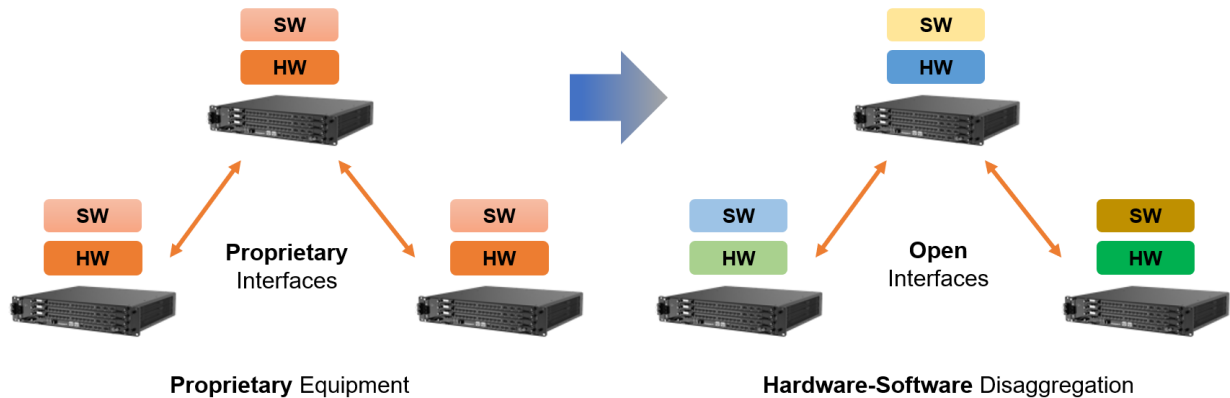
Specifically, in traditional RAN, the front-haul interface uses the CPRI protocol and even though it is standardized by 3GPP different deployments generally has vendor-specific implementation that is not necessarily open [5]. Furthermore, there is a lock-in in 4G



because each vendor creates their own version of X2 interfaces which will force the lock-in to continue with 5G.

### 3.3 O-RAN, Software and Hardware Disaggregation

O-RAN networks are specially characterized by having a disaggregated software and hardware. As seen in Figure 8, Open RAN will move the networks from being a single-vendor fully integrated network into multi-vendor disaggregated network.



**Figure 8: Open RAN software and hardware disaggregation**

It should be noted that this software and hardware disaggregation is the main difference between O-RAN and other RAN architectures, such as C-RAN (Centralized RAN) and vRAN (Virtualized RAN), where in C-RAN the RAN is centralized but with proprietary interfaces and no disaggregation, and in vRAN the functions have been virtualized but it no full disaggregation and the interfaces are proprietary.

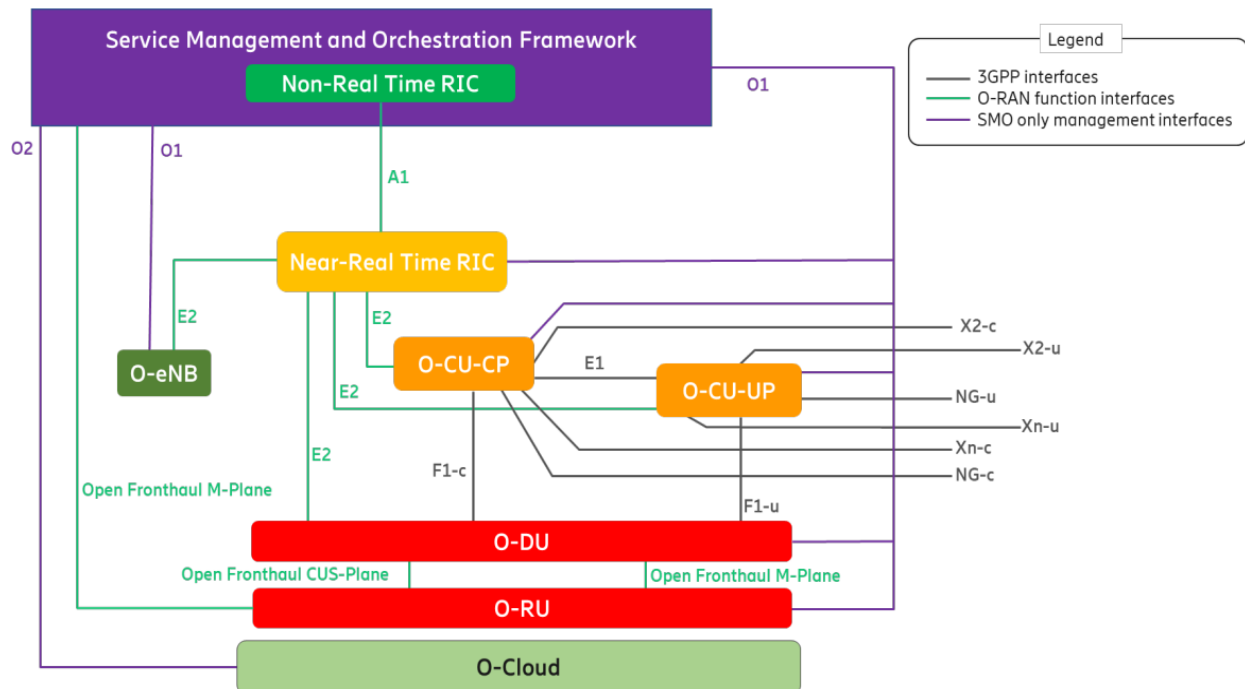
### 3.4 O-RAN Architecture

#### 3.4.1 O-RAN elements

In [6], O-RAN Alliance specifies O-RAN architecture as seen in Figure 9. A Service Management and Orchestration Framework that contains a Non real time radio intelligent controller (RIC) responsible of the management side. The radio side include a Near-RT RIC, O-CU-CP, O-CU-UP, and O-RU.

Below is a brief description of these entities:

- near-RT RIC: O-RAN near-real-time RAN Intelligent Controller
  - A logical function that enables near-real-time control and optimization of O-RAN elements and resources via fine-grained data collection and actions over E2 interface.
- non-RT RIC: O-RAN non-real-time RAN Intelligent Controller
  - A logical function that enables non-real-time control and optimization of RAN elements and resources, Artificial Intelligence and machine learning (AI/ML) workflows including model training and updates, and policy-based guidance of applications/features in near-RT RIC.

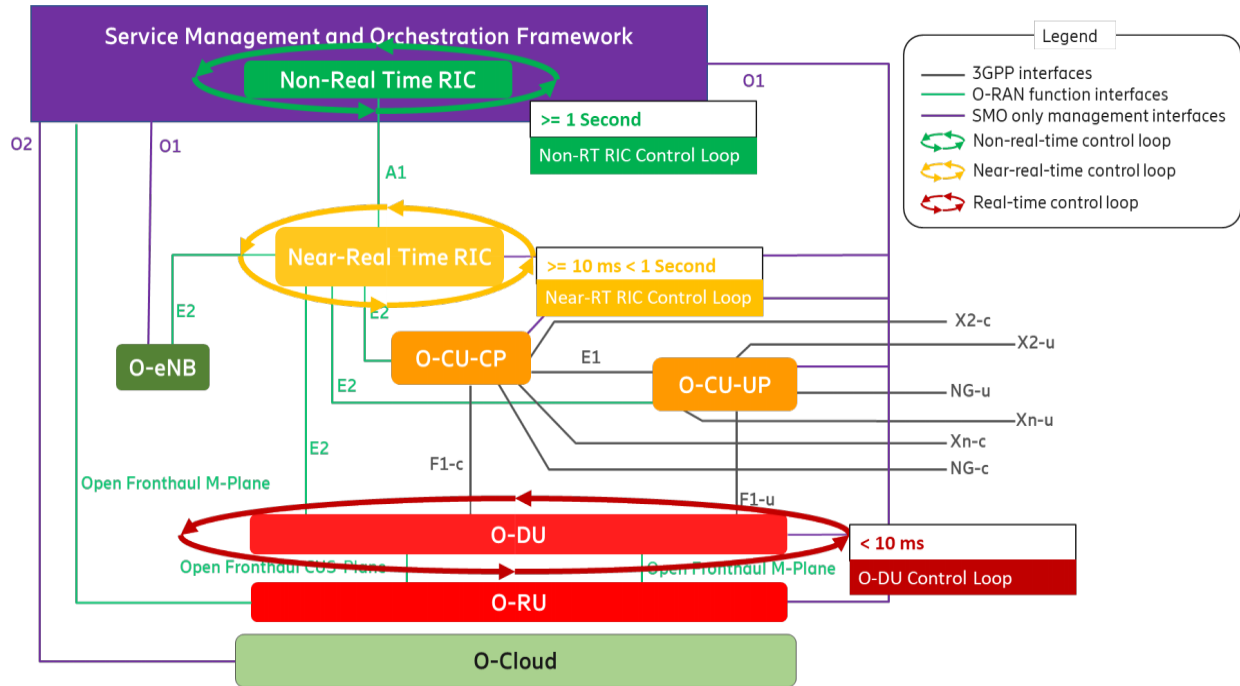


**Figure 9: Logical architecture of O-RAN [6]**

- O-CU: O-RAN Central Unit
  - A logical node hosting RRC, SDAP and PDCP protocols
  - O-CU-CP: O-RAN Central Unit – Control Plane
  - O-CU-UP: O-RAN Central Unit – User Plane
- O-DU: O-RAN Distributed Unit
  - A logical node hosting RLC/MAC/High-PHY layers based on a lower layer functional split. (More Details on the functional split is on Section 3.4.3)
- O-RU: O-RAN Radio Unit:
  - A logical node hosting Low-PHY layer and RF processing based on a lower layer functional split.
- O-Cloud: O-RAN Cloud:
  - A cloud computing platform comprising a collection of physical infrastructure nodes that meet O-RAN requirements to host the relevant O-RAN functions (such as Near-RT RIC, O-CU-CP, O-CU-UP and O-DU etc.) (The virtualization of the O-RU is for future study)
- SMO: Service Management and Orchestration
  - Which is responsible for RAN domain management and within it the Non-RT RIC. It should be noted that there is no formal interface between the Non-RT RIC and the SMO [6].

### 3.4.2 Control Loops

Three main control loops are present in O-RAN architecture as seen in Figure 10. All of these runs in parallel and they interact with each other depending on the use case.



**Figure 10: Control loops in O-RAN [6]**

- 1- Non-RT RIC control loop have a timing of 1 second or more. rApps are used in this control loop as an Independent software plug-in to provide extra functionality.
- 2- Near-RT RIC control loop have executing time between 10ms to 1 second. xApps are used in this control loop to provide extra functionality.
- 3- O-DU Control loop have a timing of less than 10ms, but it should be noted it is still not standardized as of yet by O-RAN and the main focus on the first two control loops.

rApps and xApps provide great flexibility and innovation opportunities to allow for better performing networks and unlock the O-RAN use cases. (More of those use cases will be discussed in Section 3.6).

### 3.4.3 NR Functional Split

Depending on the use case scenario, network design, and mid-haul availability, Network functions (NFs) can be flexibly distributed between the CUs and DUs [7].

3GPP proposed eight splits going from Option 1 (higher layer split - HLS) to Option 8 (lower layer split – LLS) (Figure 11). And by having this HLS and LLS, 3GPP also introduced in Rel15 the F1, E1, and W1 interfaces as seen in Figure 12.

Taking a step back, it should be noted that separating the RU from the DU was proposed because it can reduce the cost by having a less intelligent RU (The complexity of the RU is intrinsically related to the way of implementing Low-PHY functions). Also having the processing on the DU that control multiple RUs will result in pooling gains and enable easier implementations of features like CoMP (Coordinated Multi-Point).

Following that logic, even though O-RAN specify the nodes (Near-RT RIC, O-CU-CP, O-CU-UP, O-DU, and O-RU) as separate entities, it is possible to group some or all of them together, and thus collapsing internal interfaces like the F1-c, F1-u, E1, and E2 [6].

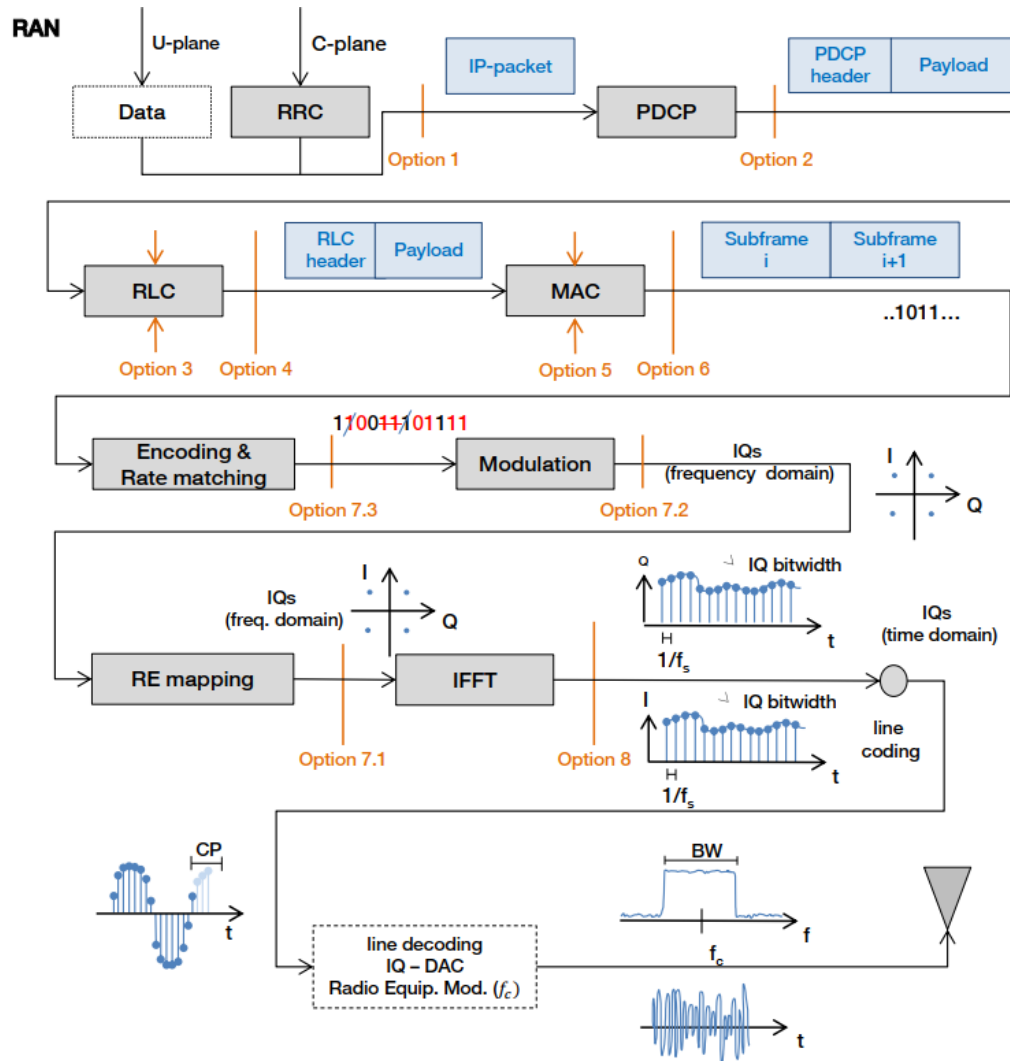


Figure 11: A downlink view for the functional splits [8]

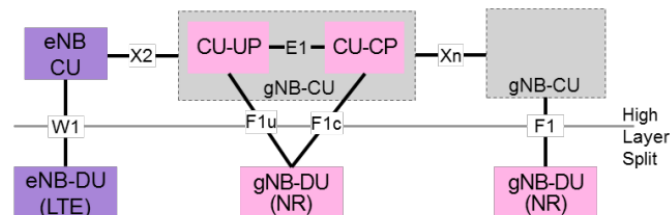


Figure 12: F1/W1/E1/X2/Xn Interfaces

Below is a description of the 8 split options:

**RAN Functional Split 1 (1A-like split):** In this split, the entire UP is located in the DU. Many of the functions will be handled locally with the centralized RRC, but at the same time faster mobility management will be achieved without the need for the operator to manage the X2 interface.

**RAN Functional Split 2 (PDCP/RLC split):** With the RRC & PDCP are in the central unit and RLC, MAC, physical layer and RF are in the distributed unit, this already standardized interface makes the inter-operation between the elements simpler. This split has the most relaxed link requirements as because all the real-time aspects are located in the DU [9].

**RAN Functional Split 3 (intra RLC split):** Low RLC, MAC, Physical layer and RF are in DU, which allows a single RLC entity to be shared with multiple MAC entities.

**RAN Functional Split 4 (RLC-MAC split):** MAC, Physical Layer and RF are in DU, but as subframes would be shorted in 5G, it will require shorter scheduling time intervals. For that, a data buffer and a flow control scheme will be necessary to implement at the DU.

**RAN Functional Split 5 (intra MAC split):** In this split scheduling is centralized in the CU, with the parts of the MAC layer, physical layer and the RF in the DU.

**RAN Functional Split 6 (MAC-PHY Split):** In this split, only the physical and RF layers in the DU. Which results in all the physical processing being handled locally, and having a centralized MAC scheduling.

**RAN Functional Split 7 (Intra PHY Split):** This is the most popular split option, because it gives a good solution for distributed RAN deployments by its pooling gains while keeping the processing load at minimum [5]. Only parts of the physical layer and RF in the DU, and depending on which parts are kept in the DU, 3 flavors of split 7 came to light.

In [8], 7.x family splits were said to offer the best balance between the inter-cell cooperation and the complexity of the RU, even though 7.1 and 7.2 splits were bandwidth hungry. On that regard, 7.x splits are the ideal deployment choice to support traffic in dense urban areas for 4G and 5G [5].

- **7.1 Split:** This split consists of transmitting I/Q symbols in the frequency domain, which saves the frequency to time conversion overhead [8]. One of the main characteristics of this split is that the fronthaul capacity is independent of the traffic generated by the UE [9]. Furthermore, the only other split option with fixed bitrate is split 8.
- **7.2 Split:** Similar to 7.1 split, I/Q signals will be sent in time domain, but multiple signals from multiple antenna ports will be combined [8] and within the DU precoding and resource element (RE) mapping will be done.

Because of its less complex approach, and its support for various fronthaul requirements, and virtualization benefits, this split option and its variations (named 7-2x), is what has been adopted by O-RAN in their fronthaul specifications [10].

- **7.3 Split:** By having the scrambling, modulation and layer mapper in the DU, a lower bitrate will be needed on the fronthaul link. Nevertheless, 3GPP only considers this split in the downlink direction [11]. The authors in [8] proposes a bi-directional 7.3 split, as it will have the least bandwidth consumption.

**RAN Functional Split 8 (PHY-RF Split):** In this split, the DU contains the RF functionality and all upper layers in the CU. In 2G and 3G systems this split is effective because the traffic rates are lower, and cost optimized RUs with minimal logic processing can be used [5]. On the other hand, this split is not scalable for mMIMO deployments, as it will require a constant high bitrate that scales with the number of antennas [9].

The legacy split 8 used the industry standard CPRI interface, but newer deployments utilize the eCPRI instead.

The table in ANNEX I provides a high-level summary on the characteristics of the different CU-DU split options [11].

### 3.4.4 Implementation Options of O-RAN functions

O-RAN gives the flexibility of connecting multiple O-RUs with one or multiple component carriers by either Fronthaul Multiplexer (FHM) or by Cascading O-RUs [6].

In Figure 13, FHM is placed between an O-DU and several O-RUs to realize a shared cell. While in Figure 14, several O-RUs are placed in chain to realize the shared cell.

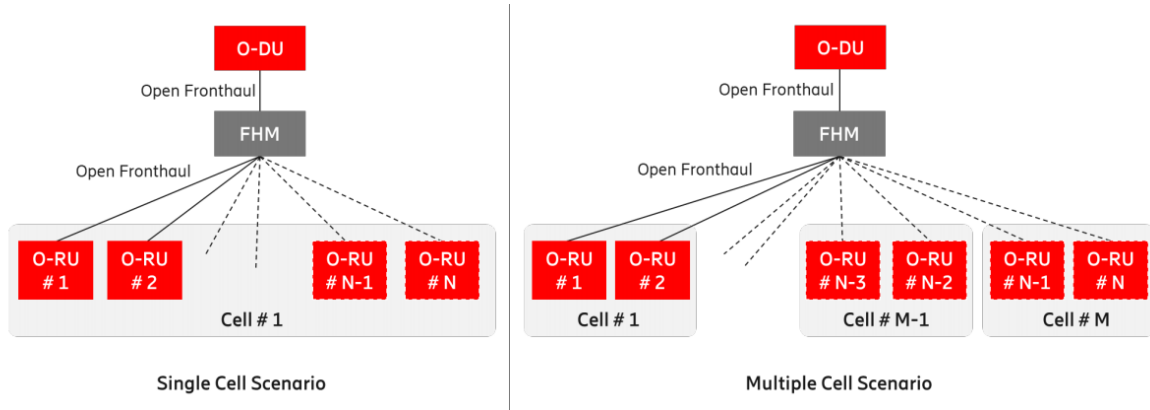


Figure 13: FHM mode usage for shared cell deployment [6]

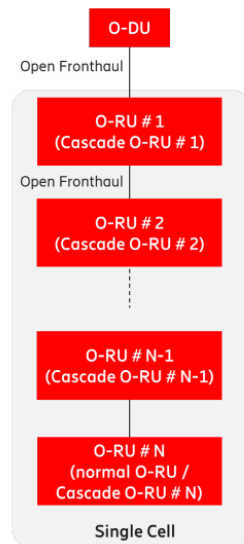


Figure 14: Cascade mode usage for shared cell deployment [6]

### 3.4.5 RAN Intelligent Controller (RIC)

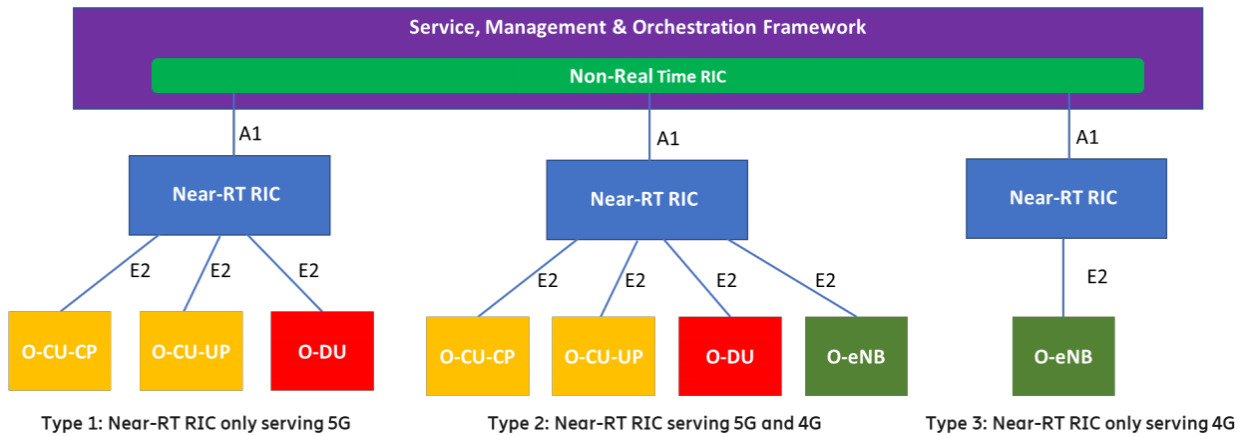
By leveraging analytics and AI/ML tools, RIC provides advanced control functionality and increase the efficiency of radio management algorithms. Four functional software elements seen in the overall RIC architecture, which are the DU software function, multi-RAT (Radio Access Technology) CU protocol stack, the Near-RT RIC, and the SMO with the Non-RT RIC within it.

#### 3.4.5.1 Near Real-Time RIC

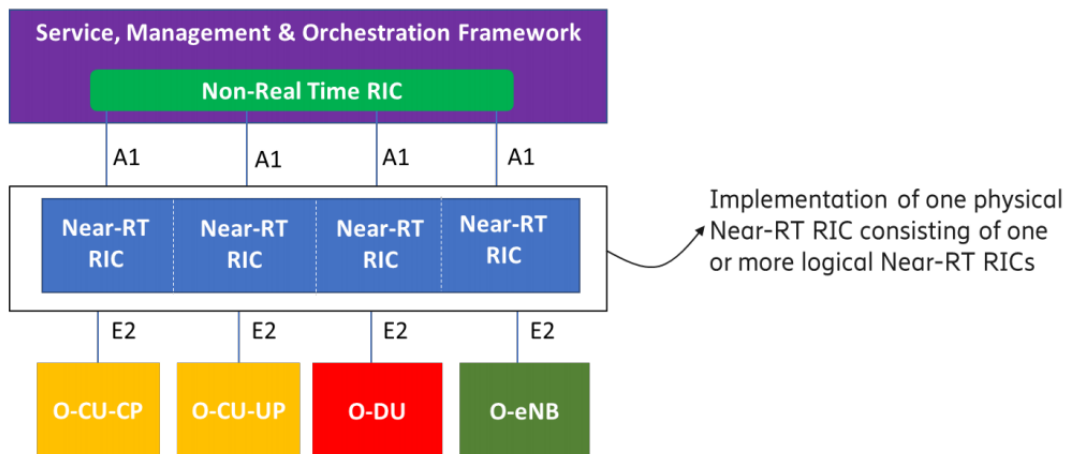
As mentioned previously in Section 3.4.1, the Near-RT RIC is a software platform for hosting xApps (micro-service-based applications). It provides the xApps with an infrastructure for controlling the different RAN entities (eNodeB, gNodeB, CU, DU) via the E2 interface. Furthermore, it receives data and policies via the A1 interface from the Non-

RT RIC. These policies the near-RT RIC can monitor, suspend, or override with its own policies.

A single or multiple E2 nodes can be controlled via a single Near-RT RIC, by either having a centralized deployment model as in Figure 15 or a distributed deployment model as in Figure 16.



**Figure 15: Centralized Near-RT RIC Serving 4G and 5G Simultaneously [6]**



**Figure 16: Distributed Near-RT RIC [6]**

### 3.4.5.2 Non-Real-Time RIC

The Non-RT RIC framework provides added services to RAN operation by hosting rApps. Using these rApps, advanced and intelligent. The functionality of the Non-RT RIC is internal to the SMO and it interfaces with the near-RT RIC via the A1 interface.

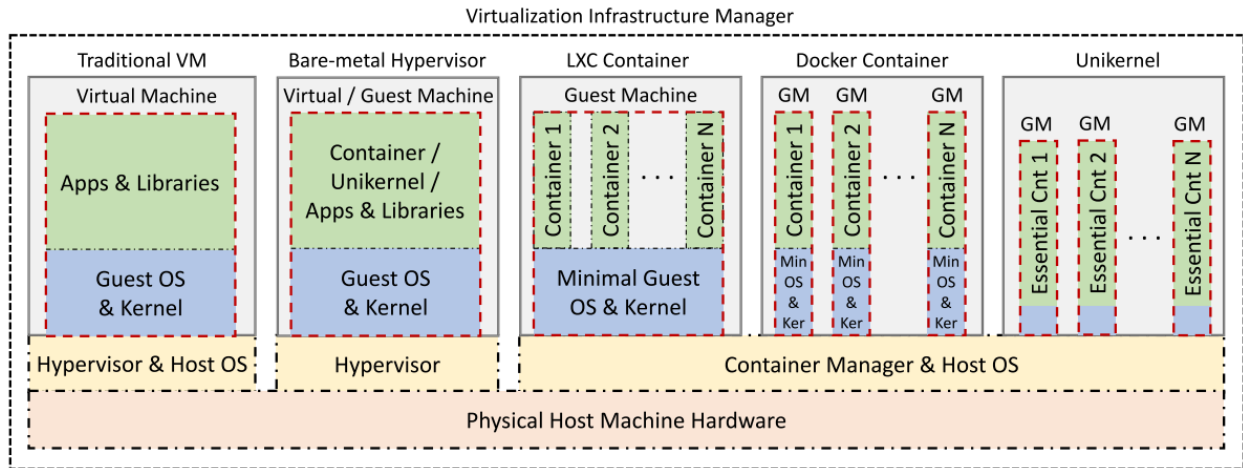
Using data analysis techniques and AI/ML, the non-RT RIC produces optimization actions and advanced functionality for configuration management, device management, fault management, performance management, and lifecycle management for all network elements in the network.

## 3.5 Virtualization, CI/CD, O-Cloud

Advanced RAN virtualization combined with SDN and functional splits and slicing is becoming a requirement to achieve the 5G expected applications. Furthermore, by delivering network automation the deployment of the network would be easier, and this

can only be achieved by taking a cloud-native environment approach for the RIC and virtualize (or containerize) the network functions.

There are many ways to decouple applications and services, from the hardware they run on top, Figure 17 shows some of those different ways.



**Figure 17: High-level NFV architecture [12]**

In O-RAN specifications, network functions can be VNFs (Virtualized Network Function). As in Figure 9, the O-RAN NFs will be containers sitting above the O-cloud and will support the O1 interface [6].

- The virtual machine will emulate a computer operating system while a container is the environment that a specific service will run in.

Going further, the continuous Integration/Continuous Delivery (CI/CD) DevOps methodology will be adopted for Open RAN. This will allow development processes to support short software development cycles and rolling out features and releases more quickly than legacy networks.

It is suggested to follow four principles in Open RAN that were currently used for data centers [5]:

- 1- RAN functions should be deployed as a microservice
- 2- Rather than legacy-based infrastructure, a container-based environment should be adopted.
- 3- Promote DevOps approaches for frequent and fast software delivery.
- 4- CI/CD methodology must be used, to better enable automation.

The final thing to discuss in this section is the O-Cloud. Which as mentioned in Section 3.4.1 refers to the collection of elements that has a decoupled software and hardware. Several deployment scenarios are specified by O-RAN.

As seen in Figure 18, these different scenarios depend on either the network element is proprietary or O-RAN compliant O-cloud. Currently Scenario B is the first priority of focus by O-RAN Alliance. Furthermore, O-RAN has a vision to have an entire metropolitan area BBU processing needs located in a single location [13].



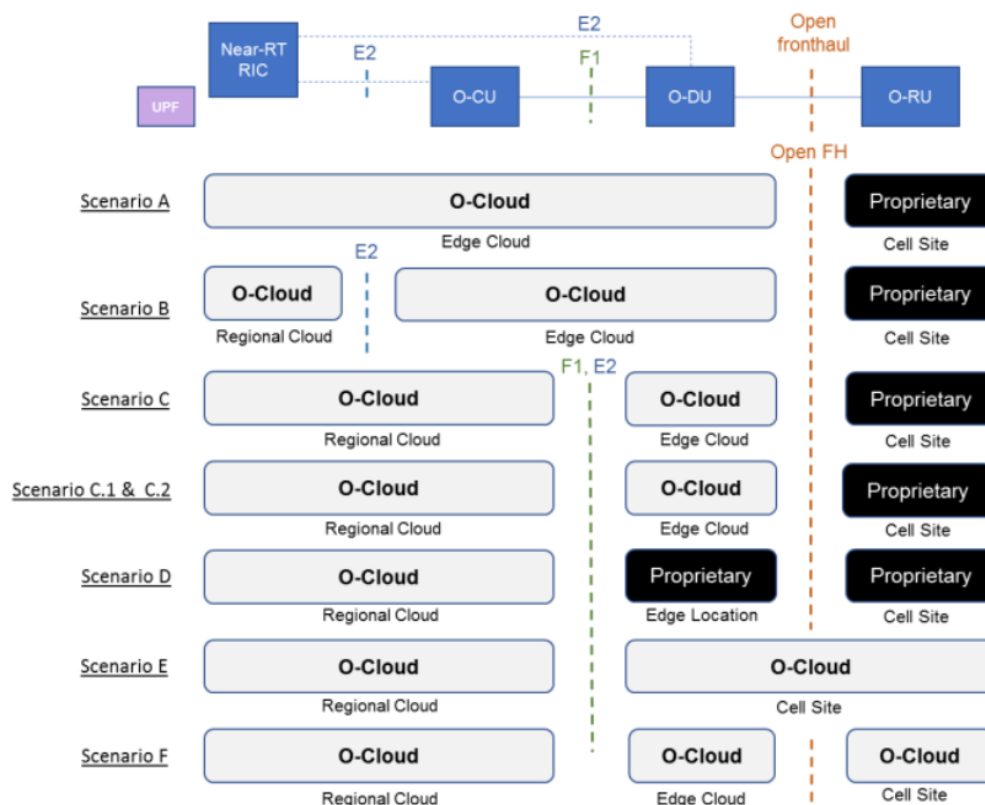


Figure 18: Cloud Deployment Scenarios [13]

### 3.6 O-RAN Benefits and Use Cases

#### 3.6.1 O-RAN benefits

When compared to traditional RAN, the main value added to O-RAN (Other than the openness and the flexibility from it) is the intelligence. The xApps hosted on the Near-RT RIC through the E2 interface and the policy changes from the Non-RT RIC through the A1 interface allow for added functionality and more efficient network performance.

O-RAN adoption will reduce network CAPEX and OPEX. And this is achieved by:

- 1- The open interfaces, which will allow more companies to join in and increase the competition in the RAN supply market.
- 2- Adopting the open-source software will result in faster innovation.
- 3- Using cloud native O-RAN concepts enables a cheaper scale out designs rather than the scale up designs required in traditional RAN.

As mentioned before, due to intelligence introduction in the RAN, O-RAN will also improve network efficiency and performance. Automating the RAN, will bring forth a better fine tuning for the different radio control algorithms (e.g., mobility management, load balancing, QoS assurance, energy saving functions, ... etc.).

Furthermore, O-RAN networks will have a significant agility edge comparing it to traditional RAN. Cloud native O-RAN networks would be capable of importing/rolling-out new features and functionality by easy software upgrades.

### 3.6.2 O-RAN use cases

In this section, a summary of the O-RAN use-cases listed in O-RAN specifications [14].

**Traffic Steering:** The intelligent RAN will allow for a more UE-centric tuning, while also reducing the manual intervention and increasing the response time.

**QoE optimization:** O-RAN will unlock vertical applications that require QoE (quality of experience) prediction and support for the dynamic traffic volume and fast fluctuations in radio transmission.

**QoS based resource optimization:** As mentioned, O-RAN will enable a more UE centric view of the network, which will allow for user priorities and per user resource allocation.

**Massive MIMO optimization:** with the mMIMO increase of number of antennas, O-RAN AI/ML adoption will offer a better coverage and capacity control to achieve the intended SLAs.

**RAN slice SLA assurance:** The O-RAN intelligence will allow for a better per slice fine tuning to ensure slice SLAs dynamically.

**Context based dynamic handover management for V2X:** An AI/ML functionality will customize handover (HO) sequences to avoid abnormalities by analyzing past navigation and radio statistics.

**Flight path based dynamic UAV resource allocation:** Flight path, climate, and payload information can be analyzed and with-it relevant ML models can be trained to better fine tune the coverage and HO during the flight.

**Radio resource allocation for UAV applications:** O-RAN enables edge and local processing of the video and control information.

**RAN sharing:** sharing scenarios can be deployed where a sperate carrier can be utilized by each operator allowing for more freedom and independent control of the radio resources. This is done by taking advantage of “remote” E2/O1/O2 interfaces that will enable coordination between the operators.

**Multi-vendor slices:** O-RAN will enable a slice to be composed of functions provided by different providers.

**Dynamic spectrum sharing (DSS):** Having the RIC control the resources and interference mitigation functions between nodes with LTE/5G shared frequency bands will unlock an efficient DSS.

**NSSI resource allocation optimization:** AI/ML models can be trained to predict the traffic and performance in different times and locations, which will bring forth a more capable NSSI (Network Slice Subnet Instance) functionality to realize the service requirements.

### 3.7 Global O-RAN Adoption and Challenges

O-RAN is getting a lot of traction, and it is touted to be the way future mobile network is going to be deployed. An important indication of that, preliminary thoughts and visions of 6G is agreeing on that 6G is going to have open interfaces and intelligence built into its architecture [15].

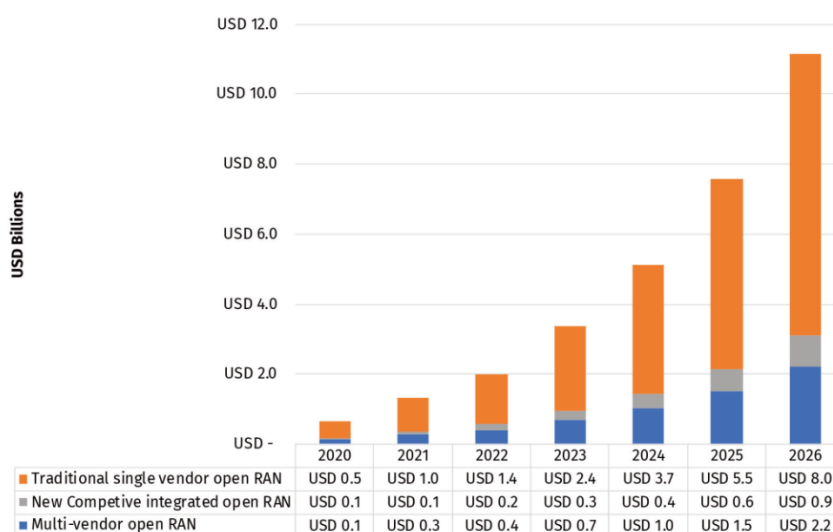
At the same time, fully adopting O-RAN into 6G is still facing some challenges. As this will require a merger of O-RAN and 3GPP specifications into one, which is a challenge because not all contributors of these standards are members in both bodies [16].

### 3.7.1 Market Potential and Current O-RAN Deployments

Historically radio networks were not the most agile systems, but adopting open RAN will unlock data center economics and cloud agility to it. It is predicted that O-RAN components and services to grow to 11.2b USD in 2026 [17].

This appeal is mainly due to the CAPEX and OPEX reduction potential. Any reduction in the RAN segment will significantly impact the cost as it believed that 80% of the total network cost could be from the RAN [5].

Going towards Open RAN will mitigate two of the most pressing challenges facing networks today, the cost to deploy and maintaining the networks that are getting more and more complex within a vendor lock-in environment.



**Figure 19: O-RAN market growth projections 2020-2026 [17]**

Trails and limited rural deployments have already been tested for a while now. Open RAN first was deployed in rural areas as a low-cost solution that is simplified and can economically support the very limited number of users. From there, sites can be upgraded remotely with a software push to the edge; hence the maintenance and operation cost will be reduced.

For example, Vodafone has a live open RAN deployment in Turkey since 2019, and according to Vodafone's Andy Dunkin (Open RAN RF & Digital Platform Development manager) their open RAN deployments have network KPIs and performance matched with the other parts of the network. With that success Vodafone is already planning to deploy open RNA in UK in 2022 [18].

Currently, the world largest Open RAN architecture with end-to-end cloud native deployment is Rakuten Mobile in Japan. As of April 2021, they have more than 15000 cells on air, already covering Tokyo and planning to cover entirety of Japan by end of 2021. According to Tariq Ameen (Rakuten Group Chief Technology Officer), the entire operations team comprises about 175 employees and it is never expected to have more than 350 [19]. This, and the automation brought by the intelligent RAN, resulted in a reduction in operation cost allowed Rakuten to offer cheaper subscriptions that made Japan from the most expensive mobile data rates in the world, to one of the cheapest.

### 3.7.2 Main Challenges Facing O-RAN

**Integration:** Having different parts of the network from different vendors increases the complexity of deploying the network, hence the need for system integrators. They will act as a buffer between the operators and the vendors.

Furthermore, operators will need to improve the skill set of their employees as it will be more about software integration than hardware.

**Testing:** There will be an increase in testing cost, because of different parts of the network are coming from different vendors. Software and hardware Badging and certifications are already proposed to mitigate this cost increase.

**Chipsets:** The new players coming to RAN parts manufacturing will need more time to catch up to the big tech giants now dominating the market (Ericsson, Huawei, Nokia, ...) in term of producing modules that have competitive performance and energy consumption levels.

## 4. O-RAN and AI/ML

The breakthroughs and improvements on how much computing capacity can be achieved, and the exponential increase in data have made it feasible for techniques like machine learning (ML) and artificial intelligence (AI) to be used in RAN [20]. For O-RAN, applying AI/ML on the RIC is one of the top main enablers of O-RAN use cases.

O-RAN specifies four main principles, when applying ML to O-RAN [21]:

- 1- Offline learning is suggested as a best practice, even in scenarios where reinforcement learning is used.
- 2- Before deploying the models on the network, they need to be trained and tested.
- 3- Modular approach in designing the ML applications is suggested.
- 4- When possible, the decision of whether the ML application would be deployed in the Non-RT RIC or the near-RT RIC should be left to the service provider.

### 4.1 Review of ML

The field of study that looks into how to give computers the ability to learn without being explicitly programmed is called machine learning (ML). Depending on the application needed, different types of machine learning algorithms will be used.

#### 4.1.1 Supervised Learning

Using this type of algorithms, the aim is to learn a mapping function from the input to the output, given a labelled data set.

Such algorithms:

- 1- Regression: Linear Regression, Logistic Regression
- 2- Instance-based Algorithms: k-Nearest Neighbor (KNN)
- 3- Decision Tree Algorithms: CART
- 4- Support Vector Machines: SVM
- 5- Bayesian Algorithms: Naive Bayes
- 6- Ensemble Algorithms: Extreme Gradient Boosting, Bagging: Random Forest

A common issue such algorithms is the dependency on the availability of a training set. In many cases data will not be available or it would be inaccurate.

#### 4.1.2 Unsupervised Learning

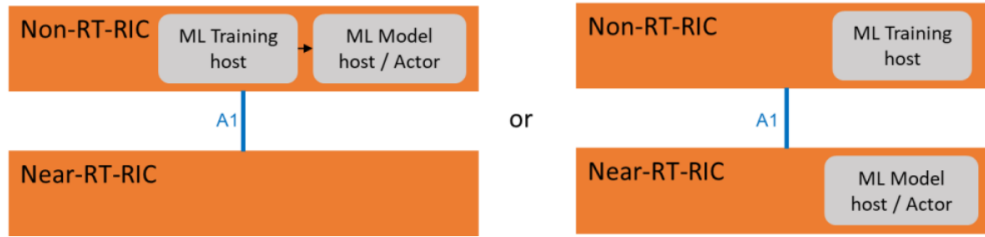
Unsupervised learning, is a machine learning task that aims to learn a function to describe a hidden structure from unlabeled data.

Such algorithms:

- 1- K-means clustering
- 2- Principle component analysis (PCA).
- 3- Isolation Forest (iForest).

Because in unsupervised learning, patterns in data are inferred without a known reference, a major challenge would be requiring a subject-matter expert intervention.

In both cases of supervised or unsupervised learning, O-RAN specify the locations of the model training and the actor (the entity that hosts the ML assisted solution) as seen in Figure 20.

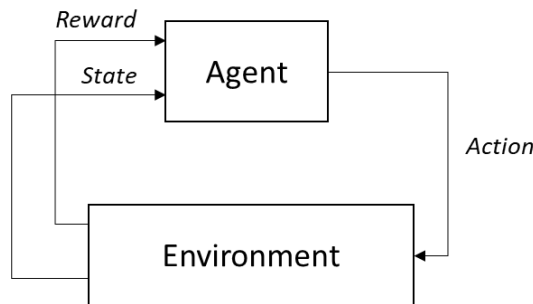


**Figure 20: Supervised learning/unsupervised model training and actor locations [21]**

#### 4.1.3 Reinforcement Learning

Reinforcement learning (RL) is where the machine learning is based on the interaction with the environment. This is done by trying to find a policy that connects a state of the world to an action that the RL agent could take.

The RL agent based on trial and error will look for actions that will produce the highest rewards or actions that might result in better rewards in long term [22].

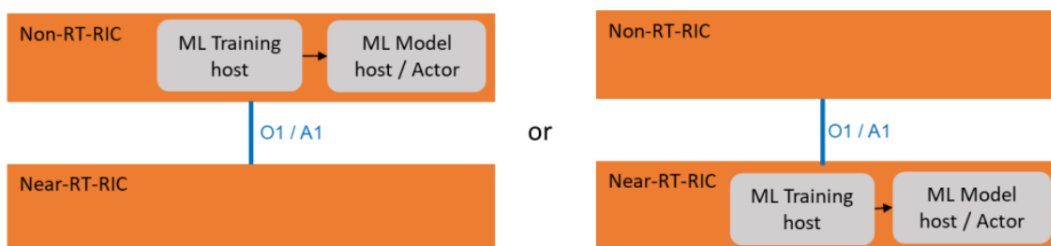


**Figure 21: Reinforcement learning framework**

Such algorithms:

- 1- Q-learning
- 2- Multi-armed bandit learning
- 3- Deep RL.

In this type of algorithms, O-RAN specify the locations of the model training and the as seen in Figure 22.



**Figure 22: Reinforcement learning model training and actor locations [21]**

Applying RL modules immediately to real networks is risky, as there is no guarantee to how the algorithm would behave on uncommon events, hence the need for rule-based system that would run alongside the RL-based system until it stabilizes [23].

#### 4.1.4 Federated Learning

Federated Learning (FL) is where the model learns in a distributed manner, which will save bandwidth needed for model updates by exchanging model parameters instead of raw data.

In Section 3.5, deployment scenario B is suggested where FL can be applied in multiple ways (Non-RT RIC in central cloud, Near-RT RIC in regional cloud, and O-CU/O-DU to edge cloud) [23]

#### 4.2 Mapping AI/ML functionality in O-RAN

As discussed in Section 3.4.2, there are three control loops within O-RAN architecture. The ML assisted solutions will fall into one or more of those loops depending on the use case.

An overview of how the intelligence will be added within the network architecture is seen in Figure 23. It should be noted that the additional inference timescale below 1ms shown in Figure 23 is not part of the O-RAN as it will require a device and RU level standardization that is not available yet [24].

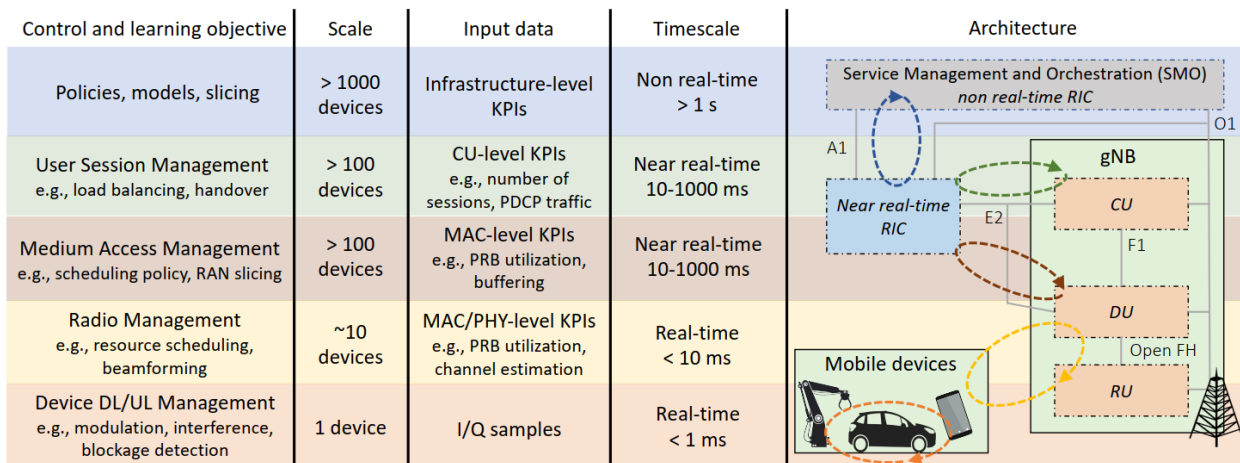


Figure 23: Learning-based closed-control loops in an O-RAN architecture [24]

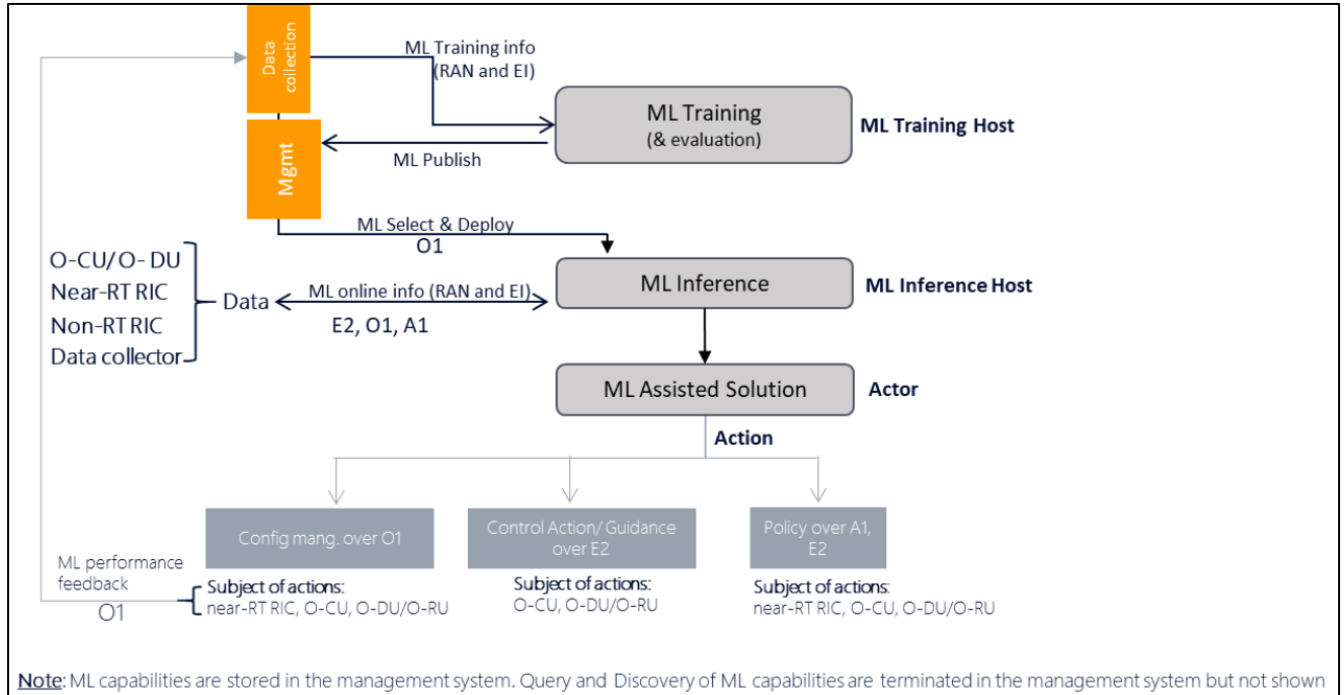
#### 4.3 AI/ML General Procedure and Interface Framework

The potential mapping relationship between the ML components and the NFs interfaces were described by O-RAN [21] as in Figure 24.

Different scenarios for the ML framework deployment were described by O-RAN in [21], shown in Table 1. O-RAN also specifies some expected key phases to any ML-assisted solution in O-RAN architecture.

- 1- ML model capability query/discovery: where the SMO will discover various capabilities and properties of the ML inference host (such as: processing capability, supported ML model formats, and data sources available)
- 2- ML model selection and training.
- 3- ML model deployment and inference: Over O1/E2/A1 interfaces, events and counters will be used for inference and policy changes or actions will be exchanged.

- 4- ML model performance monitoring
- 5- ML model redeploy/update



**Figure 24: ML training host and inference locations [21]**

**Table 1: AI/ML deployment scenarios [21]**

Deployment Scenario	ML Training Host	ML Inference Host	Interface for ML model deployment / update	Subject of Action	Action from inference host to subject		Enrichment data for inference
					Config Mgmt. (CM)	Policy / Control	
Scenario 1.1	SMO/Non-RT RIC	Non-RT RIC	SMO internal	Near-RT RIC	O1	A1 (policy)	SMO internal
				O-CU, O-DU, O-RU	O1	N/A	SMO internal
Scenario 1.2	SMO/Non-RT RIC	Near-RT RIC	O1, O2	Near-RT RIC	near-RT RIC internal	near-RT RIC internal	A1
				O-CU, O-DU, O-RU	N/A	E2 (control/policy)	E2 (if applicable)
Scenario 1.3 (FFS*)	SMO/Non-RT RIC	O-CU / O-DU	O1, O2	O-CU, O-DU, O-RU	FFS*	FFS*	FFS*

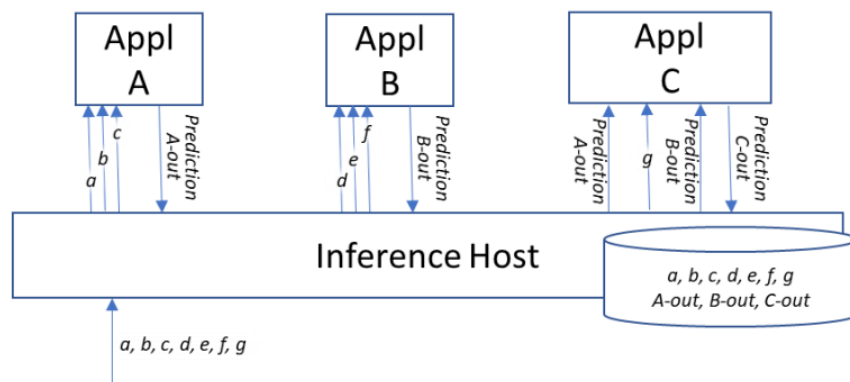
\* FFS (For Future Study)



### 4.3.1 Data, Model Training and Model Evaluation pipeline

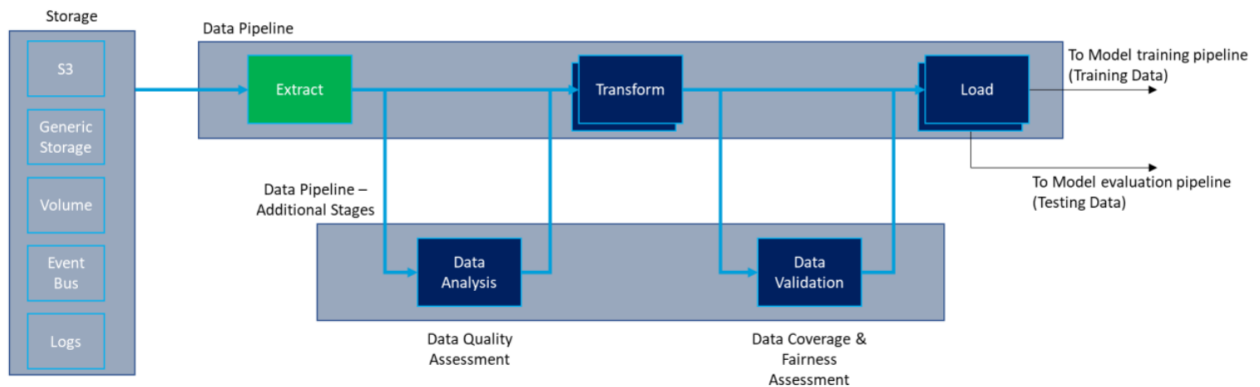
Regarding the design and composition of the ML model, a number of approaches were specified by O-RAN: whether a single model with many inputs or a modular chained model with common inputs would be used [21]. And As mentioned before on the principles of ML adoption in O-RAN, the modular approach is the best practice suggested. Furthermore, this modular approach would be extended to applications, as vendors/operators would want to provide/acquire a complete solution and not a simple ML model.

As for accessing this data (with the common inputs), it is crucial to avoid having a single application “owning” data. A better approach would be to have the data commonly held by all applications as seen in Figure 25. With that, a newly loaded application would register with the platform and declare what types of data it would consume and what data it would produce.



**Figure 25: O-RAN Spec. - Chained modular models with common inputs [21]**

How data is being handled greatly impact the effectiveness of data-driven applications. Figure 26 shows how “extract-transform-load” (ETL) process handle the data, where more stages for data quality check and validation can be inserted into this pipeline as well.



**Figure 26: O-RAN Spec. - Data pipeline [21]**

Going further, depending on the model type the model training pipeline may change, but a generic approach is usually present for the evaluation pipeline. An extension for the evaluation process might be used to select the best model from a range of models. (Figure 27 shows an overview of the model training and evaluation pipelines)

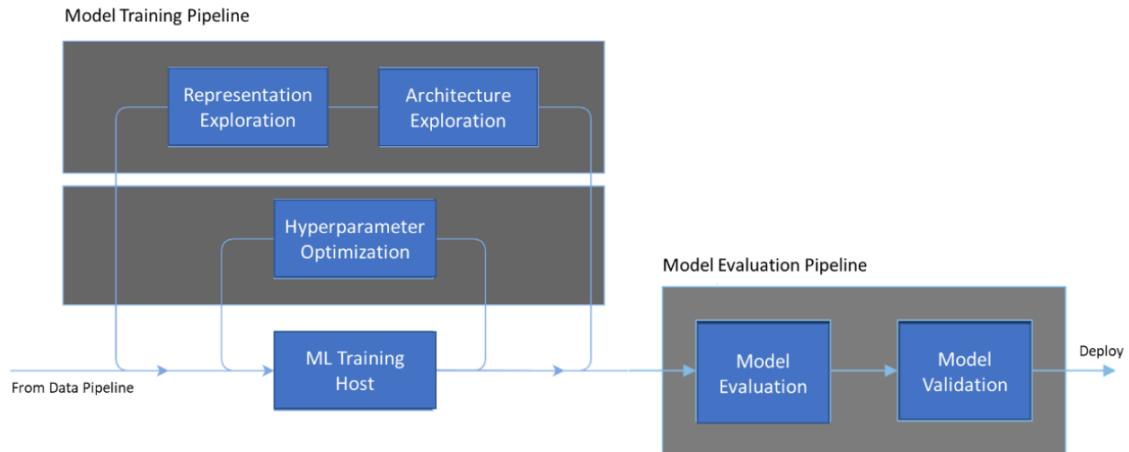


Figure 27: O-RAN Spec. - Model training and evaluation pipelines [21]

#### 4.3.2 ML Model Lifecycle Implementation

Figure 28 shows the typical steps involved in ML assisted solution in O-RAN, which takes into account a supervised or unsupervised learning models. For reinforcement models, these steps could vary depending on the use case.

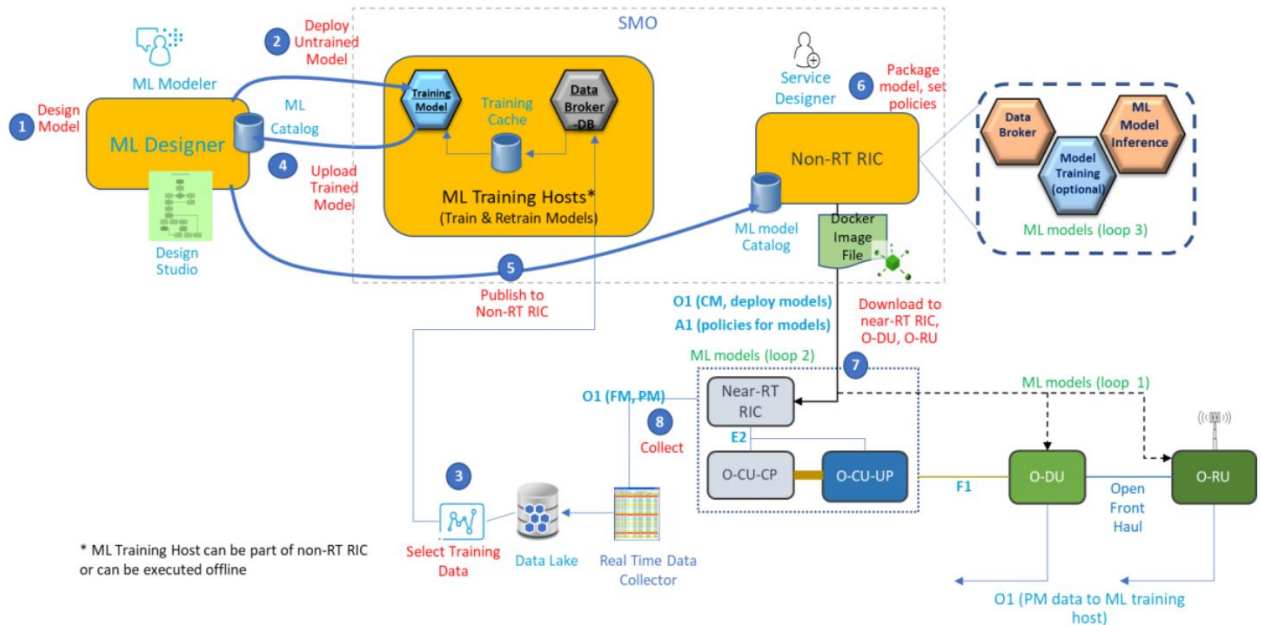


Figure 28: O-RAN Spec. - ML model lifecycle (an implementation example) [21]

#### 4.3.3 Deployment Scenarios

Depending on the amount of data needed by the model and the latency requirements, the decision of the deployment scenario would be made. (From Table 1, scenarios 1.1 and 1.2 deployments are shown in Figure 29 and Figure 30 respectively.)

Considerations which must be taken into account include: The data availability, cost of moving the data, latency, and the significance/benefits of centralizing vs localizing the data.

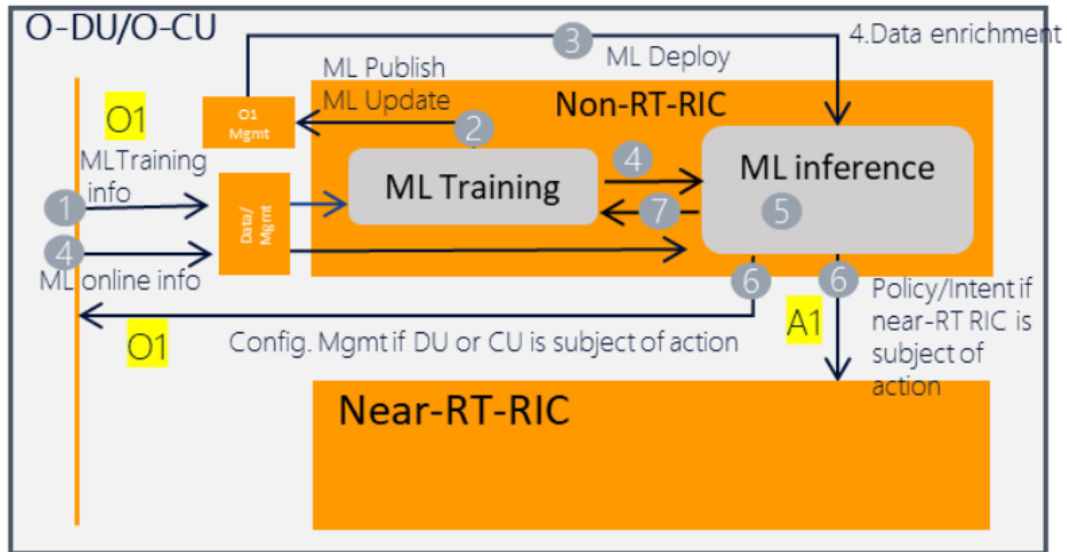


Figure 29: O-RAN Spec. - Deployment scenario 1.1 [21]

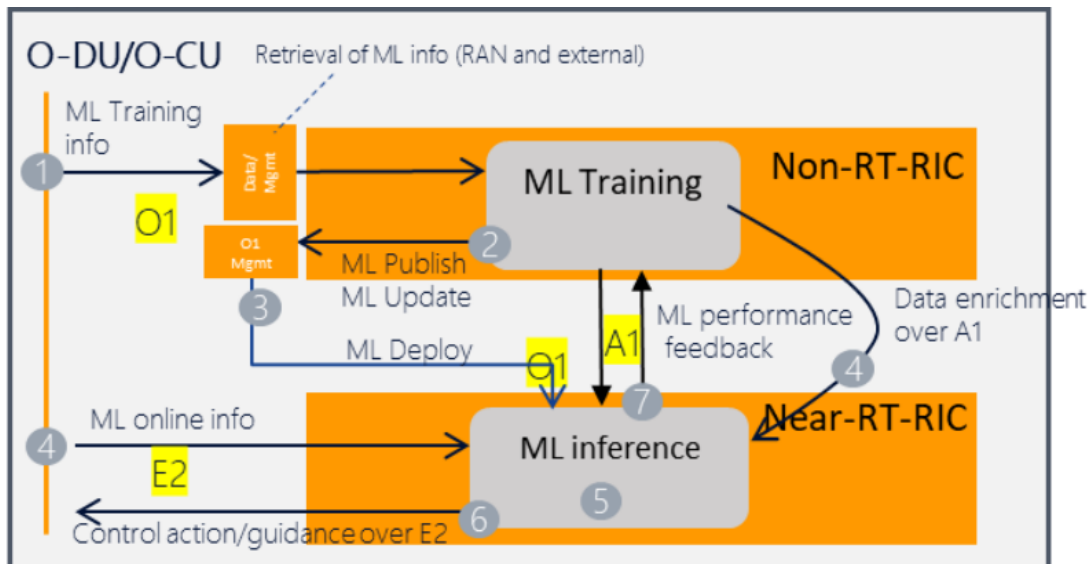


Figure 30: O-RAN Spec. - Deployment scenario 1.2 [21]

Furthermore, different deployment options can be used to enable a ML model in the Near-RT RIC, as shown in Figure 31:

- 1- Image based deployments: Where the AI/ML model will be deployed as an xApp. This option will have a more flexible deployment but its efficiency will depend on the container capability.
- 2- File based deployment: Where the AI/ML model will be deployed based on a model file, decouple with the xApp. This will give better customization by exploiting the on-device model optimization and the potential use of standard file formats for ML models, but it will require the matching of the model format and the inference engine.

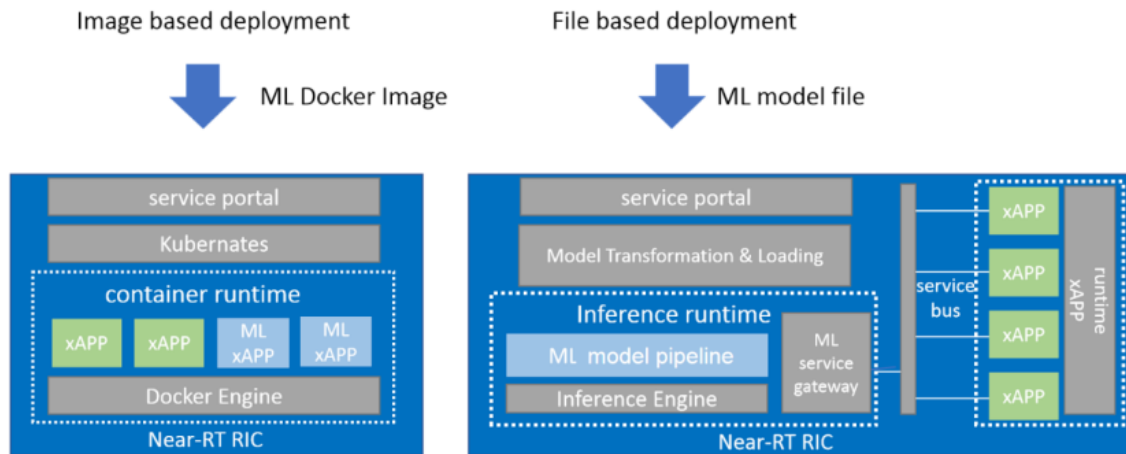


Figure 31: Examples of image based and file-based ML model deployment [21]

#### 4.4 Adoption of AI/ML in O-RAN

As mentioned in previous section, O-RAN has specified a number of use cases. There are a number of research work on such use cases and more.

For example, in [24] the authors applied deep reinforcement learning (DRL) models to schedule and control slices in 5G network. They proved the feasibility of having a closed-control loop (with the DRL enabled xApp) to select the best scheduling policy for each RAN slice. Their experimental results showed an improved network performance with 20% gain in spectral efficiency and buffer occupancy.

In [25] a performance self-diagnosis system was applied into a cellular network. Two unsupervised ML applications were used on the RIC to enable real-time performance self-diagnosis. The first one would detect the anomalous performance, and finds the root cause behind it. While the second application would learn the relationship between two performance measures.

Another example is the use of long short-term memory (LSTM) recurrent neural network (RNN) in [26] to predict traffic patterns and identify potential congested cells of a real-world network in a densely populated area. After the identification of such cells, cell splitting is applied to improve and maintain the quality of experience.

## 5. Automatic Neighbor Relations (ANR) Optimization

### 5.1 Introduction to Neighbor Relation Optimization

Since the early days of digital mobile communications, neighboring cell list planning along with frequency planning and power planning has been considered the main aspects determining the quality and capacity of the mobile network [27]. This is because handover enables the mobility of network users while maintaining continuity of the connection, and a proper neighbor relation planning will allow efficient use of network resources.

The premise is that, a UE camping in a cell with poor radio conditions instead of a better cell nearby will result in the UE suffering from poor QoE. Also, it will waste resources, as the user will spend more time reserving resources (time, RBs, power, transmission bandwidth, etc.) trying to finish its service -e.g., downloading a file- and causing interference with the multiple retransmissions and high power needed to keep the poor connection. Furthermore, this UE will share resources from users that have better radio conditions that could have used the same resources more efficiently.

The neighbor relations -most often grouped in a neighboring cell list (NCL)-, is used to make the process of reporting the potential options for a target cell much faster. Instead of the user scanning the whole allowed spectrum and camping on all cells, decoding their identity (Cell ID, BSIC, CGI, or PCI depending on the RAT), and measuring the signal strength and quality, the UE will have a smaller list of cells. Which will make the measurement time much quicker. This is even more important in case of Inter-RAT handovers, as in some cases the UE will interrupt the ongoing service to make the measurement on a different frequency band. It will also allow for a more frequent measurements, resulting in better handover decisions.

#### 5.1.1 Importance of NR

As mentioned, neighbor relations planning and optimization is an important ongoing activity when operating a RAN. It makes sure the NCL is updated and making sure it is large enough to contain all necessary neighbors to ensure optimum handover but also small enough to make the measurement time as short as possible to allow for frequent measurements resulting in a more accurate estimate.

### 5.2 Handover and NR Measurements

The radio conditions can be measured by checking the signal quality of the serving and neighboring cells, for example: Depending on the RAT, there are diverse measurements can be collected; RSRP (Reference Signal Received Power), RSRQ (Reference Signal Received Quality), and SINR (Signal to Interference & Noise Ratio) [28].

The UE will be reporting these measurements to the cell and based on a predefined mechanism events would be triggered and a RRC messages regarding these events would be sent to the base station. The following events specified in 3GPP 38.311:

- A1 event: indicate that a serving cell becomes better than a predefined threshold.
- A2 event: indicate that a serving cell becomes worse than a predefined threshold.
- A3 event: indicate that a neighbor cell becomes better than the serving cell.
- A4 event: indicate that a neighbor cell becomes better than a predefined threshold.
- A5 event: indicate that a serving cell becomes worse than threshold1 and neighbor cell becomes better than threshold2.

- B1 event: indicate that an Inter RAT neighbor cell becomes better than a predefined threshold.
- B2 event: indicate that a primary cell becomes worse than threshold 1 and inter RAT neighbor becomes better than threshold 2.

By using these events as an input to the handover algorithm, quality of service will be maintained by handing the UE over to a better serving cell during mobility.

Some work has been done in recent years to apply machine learning techniques into handover decision making. [22] describes a method where a centralized RL agent will take all the measurement reports and accordingly decide on the handover actions that will yield the best long-term network performance.

### 5.2.1 NR Process

Except of few minor improvements, the same process has been used for adding and deleting relations of the NCL. A static policy would be used to monitor the performance of the NR cells, and the handover failures and based on that policy NRs will be added or deleted. The exact way of implementing this method and the steps for achieving it was improved over the years but it is mainly the same concept.

A fully manual way of doing NR planning and optimization in the older days, was by having an initial configuration (using the geographical locations of the base stations and theoretical coverage estimations) for a single RAT and then by analyzing drive tests, missing NRs will be identified and added. These steps are mostly followed in network planning phase, where no traffic and the network are still in the configuration and deployment stage. Because the real radio coverage is unknown during those stages, the resulting NCL will not be accurate [29].

After the network is deployed and there is live traffic on the network, another phase of the NR audit can be done by monitoring the handover counters. Neighbor relations would be deleted in case no attempts were made for that relation or only failed attempts were observed for a predefined time.

Below is a summary of the manual process:

#### Adding missing neighbor relations:

- Inputs:
  - o Basic Configuration File, that contain the cell identities.
  - o Measurement Report Files, where the radio conditions measurements would be obtained for each relation (RSRP, RSRQ, ...).
- Process:
  - o Using the configuration file and the measurement report files, a table similar to the one below (Figure 32) can be extracted.

Pcell	Ncell	No. MR	Avg. RSRP	Avg. RSRQ	Distance
A	B				
A	C				
A	D				
A	?	...	...	...	...

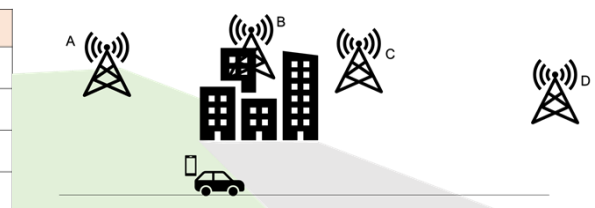


Figure 32: NR ranking based on MR data (manual NCL update)

- By ranking the relations in Figure 32, the ones with high number of measurement reports and with a good average radio condition must be added.
- Optionally, a geographical location of the cells can be used to add a distance column. Which can be used to make sure neighboring cells were added.

Deleting redundant neighbor relations:

- Inputs:
  - Performance data, that contains the KPIs and the performance counters of the cells.
- Process:
  - The same ranking seen in Figure 32 can be used, to delete those relations with low number of reports or bad radio condition measured.
  - Furthermore, using the performance data, relations that are performing badly, or not used (low HO attempts) can be deleted.

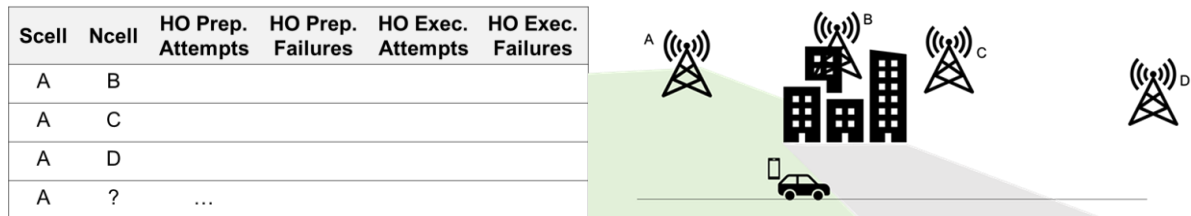


Figure 33: NR ranking based on PM data (manual NCL update)

Three main draw backs for this approach:

- 1- Delay: To delete any relation, a predefined time must pass with that relation being unused, which means the optimization would not reflect real time need/behavior of the network.
- 2- Inaccuracy: for rural areas low number of attempts for rural areas might cause the deletion of a necessary relation. On the other hand, for metropolitan areas due to imperfections in propagation models a lot of important relations would be missed.
- 3- Trial and Error: A relation would be added and then observed to know if it is needed or would be deleted.
- 4- Static policy: The optimization/planning policy of adding/deleting the relations is static. (Predefined time with handover attempts detected below a predefined value)

As the number of users stabilize and enough drive tests have been done for the network, the NCL would be more and more accurate. But it should be noted that, its accuracy in a specific area will drop again whenever a new site/sector has been added, changes in the antenna directions/tilts were made or even just new buildings constructed/demolished which will impact the coverage.

### 5.3 From Manual to Automatic NR Optimization

As the complexities of mobile networks risen with the stacking of different RAT on top of each other, having multiple diverse frequency bands, denser deployments, and beam-forming capable antennas, it became more critical to have an efficient mobility functions



in the networks. And with this growth of complexity and all this unpredictability of the coverage a need to reduce the manual radio planning has been apparent and intelligent tools must be envisioned for it.

In [27], an open loop algorithm (a system operator will have to accept/reject modifications) and a closed loop algorithm (no need for a system operator) that takes into consideration long-term measurements and statistics is proposed to suggest neighbor relations to be added/deleted for early 2G mobile networks.

In [29] a self-optimization approach was proposed for NCL optimization. It was done by having a self-configuration phase, where each base station collects measurements by scanning neighboring cells. Later on, after network operation, based on measurements from the UEs, adjustment on the resulted NCL would be done as part of the self-optimization phase.

As mentioned, improvements were made over the years by using programs to automatically do the analysis instead of it being manually done by an engineer. Some vendors provided tools, that would connect automatically to the network, where it would record/collect the needed PM, CM, and site locations and produce a list of relations to be deleted and others to be added.

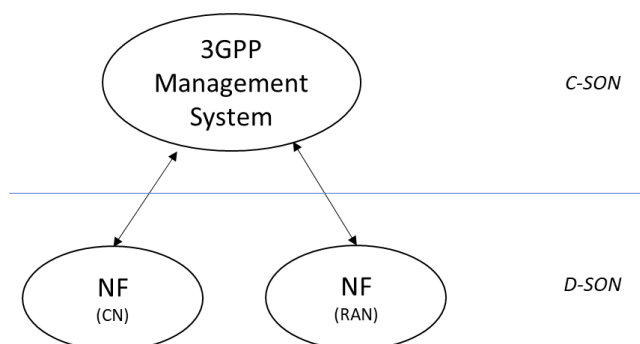
Also, some tools will add configuration check to avoid PCI collisions and PCI confusions that might severely impact the handover performance or the general KPIs of the cells by doing PCI check in 4G, CGI check in 3G, or BCCH and BSIC check in 2G networks.

### 5.3.1 SON

Self-Organizing Network (SON) is a group of algorithms that are used in mobile networks to automate network planning, configuration and management, optimization, and healing. In [30] 3GPP categorized SON algorithms based on their location into centralized SON (C-SON), distributed SON (D-SON), and hybrid SON (H-SON).

- C-SON: The algorithm is located in the management system, where the data will be analyzed and actions will be executed towards the network.
- D-SON: The algorithm is located in the NFs, where decisions will be made in the NFs, and the 3GPP management system may include some control by switching on/off a D-SON function.
- H-SON: The algorithm will be located partially in the 3GPP management system and the NFs.

A key requirement for SON is scalability, stability and convergence.



**Figure 34: SON location**



The core functions of SON -which are vendor specific- were introduced in 2007, and to correct that, some recommendations were introduced in 2010 [31].

SON systems can be used for PCI configuration, load balancing optimization, coverage optimization, energy saving management and for automatic neighbor relation (ANR) optimization.

It should be mentioned that, even though SON functions are already defined for LTE, they are still not fully defined for 5G RAN, especially resources allocation optimization and inter-slice issues [32] .

#### **5.3.1.1 ANR SON**

The first automation efforts that has used SON systems for ANR were following the same approach explained earlier in section 5.2.1. Where that same process will be initiated and executed regularly by the OSS (Operation and Support System) system of the network.

- 1- The ANR algorithm will use the geographical data and create an NCL.
- 2- The ANR algorithm will optimize the NCL by monitoring the performance (handover failures, RRC drop rate, RLF, ... etc.)

In [25] 3GPP specified SON functionality for ANR management where intra-5G relations and inter RAT relations are automatically added by the gNB.

A major issue with those is that they are vendor specific. The OSS of a specific vendor will be capable of collecting the data and apply the needed actions for that vendor base stations, and because of proprietary and closed systems, it won't be able to interact with other vendors base stations.

Furthermore, almost all of these OSS solutions work on single RAT, and with all network having multi-RAT deployments these solutions were generally used as a tool that an expert engineer will check their results and modify them before applying them to the commercial live network.

#### **5.3.1.2 SON and O-RAN**

Great benefits will be achieved by properly integrate SON and O-RAN, specially by removing redundant functional blocks and exposing SON functions to O-RAN RIC and allow SON decisions to be re-used by xApps.

Also, extra attention must be paid during implementation to make sure if any conflicting decisions were made (SON and O-RAN) a proper way of handling them is ready to ensure a stable performance.

It should be noted that, some researchers claim that to achieve optimum performance with SON & O-RAN by reducing the complexity and enabling cross-layer operations a new functional decomposition of the network will be required [32].

### **5.4 O-RAN ANR Optimization**

O-RAN will provide the base that will unlock a truly fully automated neighbor relations planning and optimization. This is mainly because O-RAN has intelligence capabilities, that will allow for AI/ML functions to be used for ANR, which will give rise to better performance. Furthermore, with O-RAN open interfaces, it will make sure the ANR system can function without being to a specific vendor or RAT.

In [33] a methodology to avoid handover failure inside the O-RAN architecture framework via ANR optimization and policy enforcement was proposed. By using the MR data to

train a ML model to predict HO failures and prohibit handovers to that target cell by modifying the NCL.

The drawbacks of the method described in [33], is that it requires a relation level performance prediction. This would be impractical for continuous Realtime changes specially for a fully developed network that has dense site deployment with multiple RATs stacked on top of each other. Due to the number of relations (Intra-Freq, Inter-Freq, and Inter-RAT) the amount of data and processing needed would be huge. Furthermore, the ANR system has no knowledge of the Target Cell Status and not benefiting from O-RAN RIC's global view of the network.

Taking all of what was discussed in mind, a novel AI/ML enabled ANR System architecture is proposed in this thesis.

## 6. Proposed Novel AI/ML Enabled ANR System Architecture

### 6.1 Motivation

As mentioned in the previous sections, the continuous complexity increases of mobile networks with their multiple RAT, multiple bands, HCS planning, denser site deployments, and beam forming capabilities, have made the ANR process more critical and more expensive. The design of a novel AI/ML enabled ANR system taking advantage of the O-RAN architecture was proposed to meet those challenges.

### 6.2 Proposed Design

The main premise that would be used is to rank neighbor relations by a given score that would be calculated using the different rApps available. Based on this score relations would be added and deleted.

Two main phases of operation.

- 1- Initial operation phase.
- 2- Main operation phase.

#### 6.2.1 Initial Operation Phase

Here, the system has just started with a fresh new network, or there are not enough measurements from handsets to be used for analysis.

The relation score would be calculated from only the rApps that don't require live network measurements. The score would be calculated using mainly inter-cell distance. Using this would provide an inaccurate list, but it would be updated as soon as the measurements begin to accumulate. Also, when the other rApps for predicting Radio conditions and QoE starts producing outputs, the system would switch to main operation mode, which will give a more accurate and responsive changes.

Figure 35 shows the rApps involved during this initial phase.

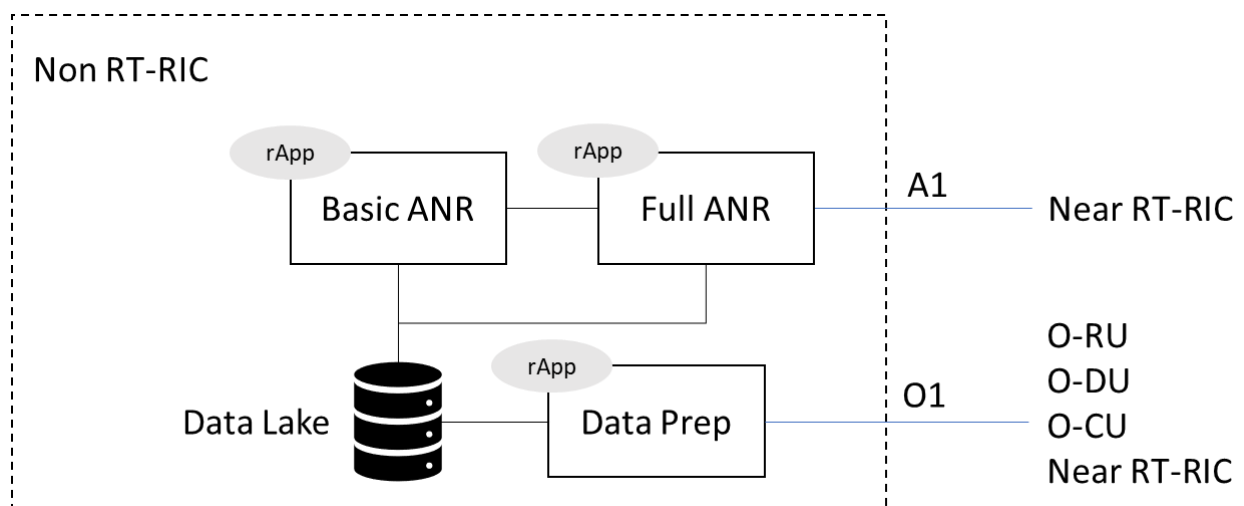


Figure 35: The ANR System during Initial Operation Phase

- **The “Data Prep” rApp:** It would be connected to the other entities of the network through the O1 interface and would format and store the data needed by the ANR system in a way that would be easier to access. Data would include, network

configuration (mainly Cell ID, BSIC, CGI, or PCI depending on the RAT, and the current NCL), also the measurement reports collected from the handsets.

The main job of this rApp that it would aggregate the data from the original resolution (15min, 30mins, or 60 mins) to daily values.

- **The “Basic ANR” rApp:** Figure 36 shows the process flow of the “Basic ANR” operation. The rApp will first check if there are enough measurement reports collected in the data lake. If there is, for each cell, all its neighbors (both configured already or missing) would be ranked based on the signal strength and quality. A score would be given for each relation based on it.

Furthermore, the rApp would use geographical location of the sites and the antenna specifications to calculate inter-cell distances and coverage overlap and would rank all neighboring cells (both configured already or missing neighbors) based on distance and a score would be given for each.

The rApp output would be like a table (expected to be a Pandas Dataframe) with all the neighbor relations and their scores from distance calculation and measurements as columns.

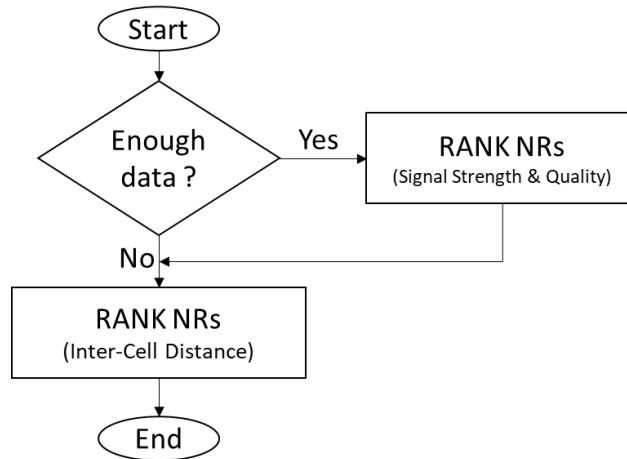


Figure 36: Basic ANR Operation Flow

- **The “Full ANR” rApp:** Will calculate a final weighted score, and produce a new NCL. Modifications requests with Add and delete NRs will be created and sent to the near RT-RIC via the A1 interface.

## 6.2.2 Main Operation Phase

In this operation phase all rApps would be involved and this is where this system would differ from traditional ones and show its advantages. Figure 37 shows the rApps involved during this operation phase.

The “Full ANR” rApp will make use of data provided by these other rApps to generate a faster response and more accurate changes. Such rApps:

- **KPI predictors / QoS Predictors:** rApps that would use AI/ML to predict the incoming handover success rate of a target cell. The “Full ANR” rApp would use that information to not add relations that might be more prone for failures. This way the system will not wait for a day/week to determine a relation unnecessary. It will also give it the power to add a different relation to another cell.

This would be achieved by giving a score for each relation by checking the predicted performance of its target cell.

Most importantly, it will predict main KPIs of a target cell (Throughput and Drop Rate) and the “Full ANR” will score NRs based on it. Which will also make sure that the UE has a less chance attempting HO to a poor performing cell.

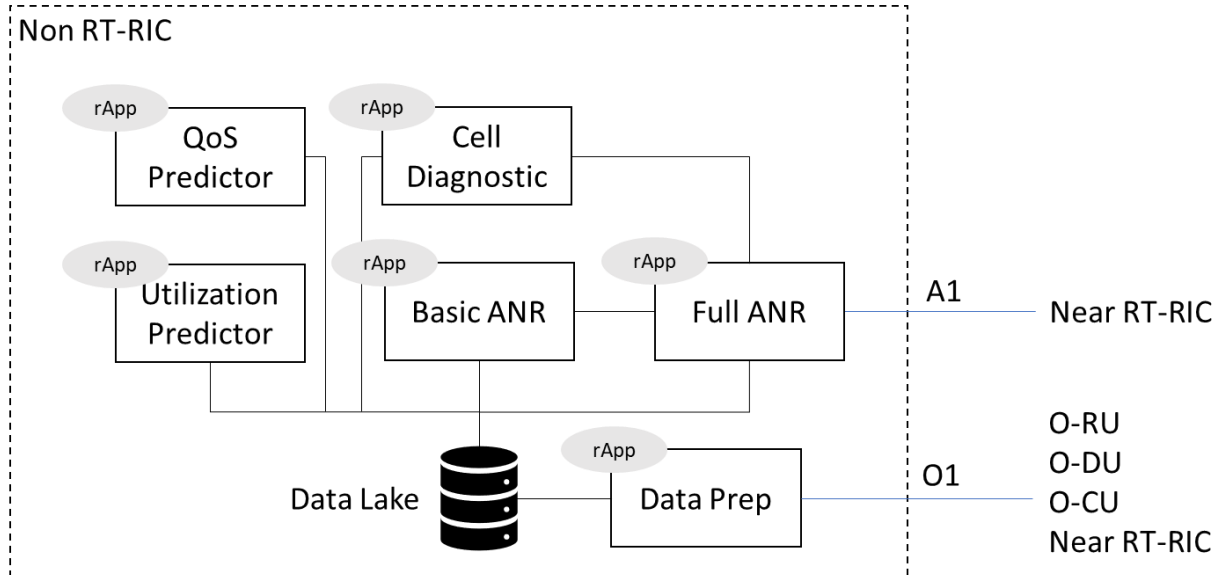


Figure 37: The ANR System during Main Operation Phase

- **Resources' utilization predictors:** rApps that would use AI/ML to predict the radio resources utilization of a cell. Similar to the previous case with the KPI predictors, the “Full ANR” rApp would use that data to score each relation target and allow for a more responsive system to utilization and give a more efficient NCL to achieve maximum performance.
- **Cells Health Checkers / diagnostic rApps:** rApps that would be used to diagnose/solve/report issues with a cell. The “Full ANR” rApp would use them in cases where a relation have a high score but bad performance, and instead of deleting the relation right away, it will try reporting the cell to the diagnostic rApp as an abnormality.

The different rApps would be used as mentioned above with the output provided by the “Basic ANR” rApp to add more details (columns to the dataframe) and then have a weighted score calculated. After that create the new NCL and send the modification requests.

### 6.3 Advantages of the Novel Design

Applying such novel ANR system will have many advantages over the other approaches.

#### 1- RAT Agnostic.

This design will take advantage of O-RAN architecture and the centralized intelligence in it; hence it will not care about the RAT.

#### 2- Vendor Agnostic.

As previously mentioned, by taking advantage of O-RAN architecture and the open interfaces it provides, this novel ANR system will support multiple vendors.

### 3- Network Awareness.

A huge advantage of this design, is that it takes advantage of O-RAN architecture and its RIC global view to ensure QoS for UEs and provide a more efficient network aware NCL updates.

For example, assuming a network as shown in Figure 38, Cell A has an NCL configured as shown and cells B and C are the main serving cells for the university campus in the image.

During school days the utilization of cells B and C will be very high. With the help of the “QoS Predictor” rApp the relation A-B and A-C will have lower scores. Which would result in the “Full ANR” updating the NCL to delete these relations.

The opposite would happen during weekends when the utilization on B and C drops. And the NCL configured in cell A would return to the original state as seen in Figure 38.

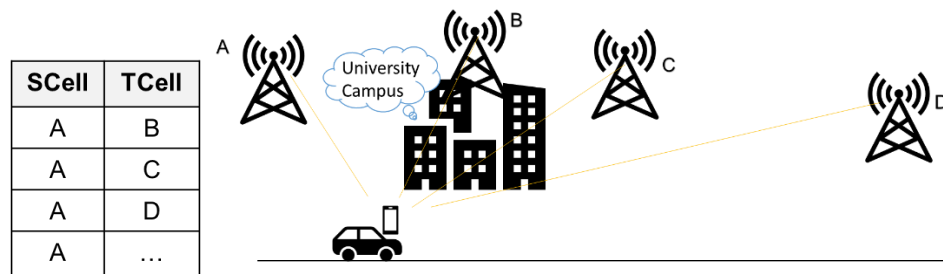


Figure 38: Example of NCL

### 4- Fast response time.

As explained in the previous example, this system will be able to predict the changes in network performance and update the NCL accordingly, while in old systems a predefined must pass before deleting a relation for example.

### 5- Report hidden performance issues.

This is one of the main advantages of such novel design. Some performance issues in the network can't be detected when monitoring cell level KPIs. Hence the need for expert engineers monitoring the network relation level performance and analyzing drive tests.

For example, assuming Figure 39 is showing part of a network in rural area with very low traffic. It will be hard to catch the KPI abnormality by the interference on cell B, but this novel architecture will be able to report it, as users moving along the road will show good radio conditions score for relation A-B but at the same time poor HO performance score.

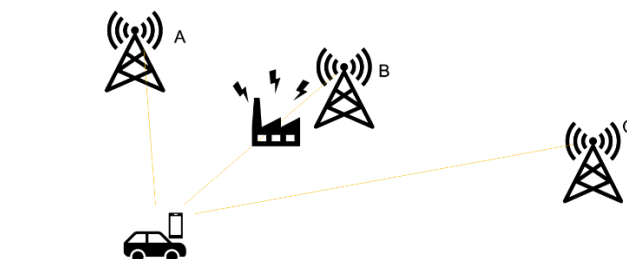


Figure 39: Example of Relation level issue

- 6- This ANR system would be acceptable by operators.

This novel AI/ML enabled ANR is addressing all the challenges and holes that other ANR systems might miss, that previously required an engineer intervention/confirmation.

It should be mentioned that the modularity approach taken while designing this ANR system will ensure the system operation even with no extra rApps added, but it would have an increased accuracy and response time with each addition.

## 6.4 Challenges and Future Work

There are a few challenges that are facing such a novel approach:

- 1- Recording/Collecting measurement reports for inter-RAT neighbors.

Because this is usually requiring the UE to go into compressed mode to make these measurements. Basically, the UE will interrupt/limit the current service by allowing for short monitoring gaps where the UE receiver reconfigure for the different frequency, collect the measurements, and reconfigure itself back to the original band. This process has the great advantage of allowing a single receiver to be used for both the original service and perform the inter-frequency measurements, but it is taxing on the overall network performance.

In addition to the inherent drawback of this procedure of interrupting/limiting the service, the UE does not have information about the radio conditions of the other band, hence high power signaling will be needed. For that reason, in live networks a full network action to record measurements that include inter-frequency measurements are only allowed during the least active period of the network (usually 3AM-5AM), because it will significantly raise the interference levels in the network.

Further research on how to take advantage of O-RAN centralized view of the network. Maybe by figuring out a way to use the intra-frequency measurements recorded by the UEs already camped in the target frequency band as an inter-frequency measurement can be achieved instead of depending on compressed mode.

- 2- The amount of MR data for a multi-RAT network.

Because of the amount of neighbor relations, the amount of data to handle by the system will be big. More research on how to apply big data management techniques might allow for using hourly aggregated MR data which will give a more responsive system.

- 3- Implementing and testing the system.

Small networks or simulations can't be used to test the performance of this system. Real network data will be needed as all the gains of using such AI/ML enabled architecture is geared towards achieving the ANR automation with the best performance possible in complex fully developed networks.

Another big challenge in testing this system is the rApps needed. To showcase the advantages of this approach, these Predictors rApps must be either acquired or built.

Addressing these challenges for future work will ensure optimum performance for such a novel architecture.

## 7. Conclusion

This thesis covered a broad view of 5G and the main characteristics of the new deployments that required stricter optimization with its denser site deployments and advanced features (mMIMO, Beamforming, mmWave, ...) and how intelligence was proposed as a way to tackle these challenges.

Then a detailed description of O-RAN architecture was covered and how AI/ML applied in it. It also showcased some of the different use cases and the benefits of applying intelligence in the RAN.

For the later part of the work, the research focused on ANR systems. Starting by how ANR was done in traditional RAN manually or with automatic systems. And later by how ANR is achieved as a use case in O-RAN.

Finally, a novel AI/ML enabled ANR system architecture for O-RAN was proposed. The system addressed many of the short comings of the previous systems. The design ensured the system would work regardless of the vendor or the RAT, and by taking advantage of AI/ML enabled O-RAN architecture it would produce a network aware ANR functionality. This network awareness will allow the system to not require an engineer confirmation before applying the changes on the network, which is something the operators won't accept for current systems. Another advantage of using this novel approach, is that it can detect performance abnormalities that can't be caught by regular monitoring systems.

A few of the challenges facing this novel approach were discussed, and suggestions on how to address them were proposed.

In conclusion, this research tried to introduce O-RAN network architecture, the AI/ML application in it and the use cases that can be unlocked. An example of those was ANR and a novel system for achieving it was proposed.



## ABBREVIATIONS - ACRONYMS

LLS	lower layer split
FHM	Fronthaul Multiplexer
RL	Reinforcement learning
FL	Federated Learning
ETL	extract-transform-load
QoE	quality of experience
RB	Resource Block
NCL	neighboring cell list
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
SINR	Signal to Interference & Noise Ratio
KPI	Key Performance Indicator
Pcell	Primary Cell
Ncell	Neighboring Cell
Scell	Serving Cell
MR	Measurement Report
PM	Performance Monitoring
CM	Configuration Management
PCI	Physical Cell Identifier
CGI	Cell Global Identity
BCCH	Broadcast Control CHannel
BSIC	Base Station Identity Code
SON	Self-Organizing Networks
OSS	Operation and Support System

## ANNEX I

**Table 2: Summary on characteristics of different CU-DU split option [11]**

	Opt. 1	Opt. 2	Opt. 3-2	Opt. 3-1	Opt. 5	Opt. 6	Opt. 7-3 (only for DL)	Opt. 7-2	Opt. 7-1	Opt. 8
Baseline available	No	Yes (LTE DC)	No						Yes (CPRI)	
Traffic aggregation	No	Yes								
ARQ location	DU			CU May be more robust under non-ideal transport conditions						
Resource pooling in CU	Lowest	in between (higher on the right)							Highest	
	RRC only	RRC + L2 (partial)				RRC + L2	RRC + L2 + PHY (partial)			RRC + L2 + PHY
Transport NW latency requirement	Loose				Not clarified during study phase	Tight				
Transport NW Peak BW requirement	N/A	Lowest	in between (higher on the right)						Highest	
	No UP req.	baseband bits						Quantized IQ (f)		Quant. IQ (t)
	-	Scales with MIMO layers							Scales with antenna ports	
Multi-cell/freq. coordination	multiple schedulers (independent per DU)				centralized scheduler (can be common per CU)					

## REFERENCES

- [1] ITU, "IMT Vision - "Framework and overall objectives of the future development of IMT for 2020 and beyond", ITU, 2015.
- [2] 3GPP, "Release 15 Description - 3GPP TR 21.915 V15.0.0," 2019.
- [3] S. K. Singh, R. Singh and B. Kumbhani, "The Evolution of Radio Access Network Towards Open-RAN: Challenges and Opportunities," in 2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), Seoul, 2020.
- [4] O-RAN, "O-RAN Alliance Membership," [Online]. Available: <https://www.o-ran.org/membership>. [Accessed 4 June 2021].
- [5] Parallel Wireless, "Everything You Need to Know About Open RAN," 2020.
- [6] O-RAN, "O-RAN Architecture Description 3.0," O-RAN WG1, 2020.
- [7] P. Arnold, N. Bayer, J. Belschner and G. Zimmermann, "5G radio access network architecture based on flexible functional control / user plane splits," in 2017 European Conference on Networks and Communications (EuCNC), Oulu, Finland, 2017.
- [8] V. Q. Rodriguez, F. Guillemin, A. Ferrieux and L. Thomas, "Cloud-RAN functional split for an efficient fronthaul network," in 2020 International Wireless Communications and Mobile Computing (IWCMC), Limassol, Cyprus, 2020.
- [9] L. M. P. Larsen, A. Checko and H. L. Christiansen, "A Survey of the Functional Splits Proposed for 5G Mobile Crosshaul Networks," IEEE Communications Surveys & Tutorials, vol. 21, no. 1, 2018.
- [10] O-RAN, "O-RAN Fronthaul - Control, User and Synchronization Plane Specification v5.0," O-RAN WG4, 2020.
- [11] 3GPP, "Study on new radio access technology, Radio access architecture and interfaces - 3GPP TR 38.801 V14.0," 2017.
- [12] L. Bonati, M. Polese, S. D'Oro, S. Basagni and T. Melodia, "Open, Programmable, and Virtualized 5G Networks: State-of-the-Art and the Road Ahead," Computer Networks, vol. 182, 2020.
- [13] O-RAN, "O-RAN Use Cases and Deployment Scenarios," 2020.
- [14] O-RAN, "O-RAN Use Cases Detailed Specification v04.00," O-RAN WG1, 2020.
- [15] The 5G Infrastructure Association, "European Vision for the 6G Network Ecosystem," 2021.
- [16] I. Morris, "Open RAN adds to 6G dilemma," 22 July 2021. [Online]. Available: <https://www.lightreading.com/6g/open-ran-adds-to-6g-dilemma/d/d-id/771026>. [Accessed 2 August 2021].
- [17] VMWARE, "O-RAN Analyst Report - WHY THE FUTURE IS OPEN," 2021.
- [18] A. Dunkin, Interviewee, Open RAN World Digital Conference. [Interview]. 6 April 2021.
- [19] T. Ameen, Interviewee, Open RAN World Digital Conference. [Interview]. 6 April 2021.
- [20] Ericsson, "Enhancing RAN Performance with AI," 2019.
- [21] O-RAN, "O-RAN AI-ML Workflow Description and Requirements 1.01," O-RAN WG2, 2020.
- [22] V. Yajnanarayana, H. Rydén and L. Hévizi, "5G Handover using Reinforcement Learning," in 2020 IEEE 3rd 5G World Forum (5GWF), Bangalore, India, 2020.
- [23] H. Lee, J. Cha, D. Kwon, M. Jeong and I. Park, "Hosting AI/ML Workflows on O-RAN RIC Platform," in 2020 IEEE Globecom Workshops (GC Wkshps), Taipei, Taiwan, 2020.
- [24] L. Bonati, S. D'Oro, M. Polese, S. Basagni and T. Melodia, "Intelligence and Learning in O-RAN for Data-driven NextG Cellular Networks," arXiv:2012.01263 [cs.NI], 2020.
- [25] F. B. Mismar and J. Hoydis, "Unsupervised Learning in Next-Generation Networks: Real-Time Performance Self-Diagnosis," arXiv:2104.06993 [cs.NI], 2021.
- [26] S. Niknam, A. Roy, H. S. Dhillon, S. Singh, R. Banerji, J. H. Reed, N. Saxena and S. Yoon, "Intelligent O-RAN for Beyond 5G and 6G Wireless Networks," arXiv:2005.08374, 2020.
- [27] S. Magnusson and H. Olofsson, "Dynamic neighbor cell list planning in a microcellular network," in Proceedings of ICUPC 97 - 6th International Conference on Universal Personal Communications, San Diego, CA, USA, 1997.
- [28] Techplayon.com, "5G NR Measurement Events," 26 February 2020. [Online]. Available: <https://www.techplayon.com/5g-nr-measurement-events/>. [Accessed 28 June 2021].

- [29] V. M. Nguyen and H. Claussen, "Efficient self-optimization of neighbour cell lists in macrocellular networks," in 21st Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Istanbul, Turkey, 2010.
- [30] 3GPP, "Study on the Self-Organizing Networks (SON) for 5G networks - 3GPP TR 28.861 V16.0," 2019.
- [31] H. Claussen, D. Lopez-Perez, L. Ho, R. Razavi and S. Kucera, "Automation of Cellular Networks," in Small Cell Networks: Deployment, Management, and Optimization, Wiley-IEEE Press, 2018, pp. 55 - 90.
- [32] S. Kukliński, L. Tomaszewski and R. Kołakowski, "On O-RAN, MEC, SON and Network Slicing," in 2020 IEEE Globecom Workshops (GC Wkshps, Taipei, Taiwan, 2020.
- [33] H. Kumar, V. Sapru and S. K. Jaisawal, "O-RAN based proactive ANR optimization," in 2020 IEEE Globecom Workshops (GC Wkshps, Taipei, Taiwan, 2020.
- [34] VIAVI Solutions, "O-RAN Open Ecosystem Power 5G Applications".
- [35] P. H. Masur and J. H. Reed, "Artificial Intelligence in Open Radio Access Network," arXiv:2104.09445 [cs.NI], 19 April 2021.