



## **Publicly Available Specification (PAS); O-RAN Use Cases Analysis Report (O-RAN.WG1.Use-Cases-Analysis-Report-R003-v12.00)**

### **CAUTION**

*The present document has been submitted to ETSI as a PAS produced by O-RAN Alliance and approved by the ETSI Technical Committee Mobile Standards Group (MSG).*

*ETSI had been assigned all the relevant copyrights related to the document O-RAN.WG1. Use Cases Analysis Report-R003-v12.00 on an "as is basis". Consequently, to the fullest extent permitted by law, ETSI disclaims all warranties whether express, implied, statutory or otherwise including but not limited to merchantability, non-infringement of any intellectual property rights of third parties. No warranty is given about the accuracy and the completeness of the content of the present document.*

---

Reference

DTR/MSG-001145

---

Keywords

PAS, use case

***ETSI***

650 Route des Lucioles  
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - APE 7112B  
Association à but non lucratif enregistrée à la  
Sous-Préfecture de Grasse (06) N° w061004871

***Important notice***

The present document can be downloaded from the  
[ETSI Search & Browse Standards](#) application.

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the prevailing version of an ETSI deliverable is the one made publicly available in PDF format on [ETSI deliver](#) repository.

Users should be aware that the present document may be revised or have its status changed,  
this information is available in the [Milestones listing](#).

If you find errors in the present document, please send your comments to  
the relevant service listed under [Committee Support Staff](#).

If you find a security vulnerability in the present document, please report it through our  
[Coordinated Vulnerability Disclosure \(CVD\)](#) program.

***Notice of disclaimer & limitation of liability***

The information provided in the present deliverable is directed solely to professionals who have the appropriate degree of experience to understand and interpret its content in accordance with generally accepted engineering or other professional standard and applicable regulations.

No recommendation as to products and services or vendors is made or should be implied.

No representation or warranty is made that this deliverable is technically accurate or sufficient or conforms to any law and/or governmental rule and/or regulation and further, no representation or warranty is made of merchantability or fitness for any particular purpose or against infringement of intellectual property rights.

In no event shall ETSI be held liable for loss of profits or any other incidental or consequential damages.

Any software contained in this deliverable is provided "AS IS" with no warranties, express or implied, including but not limited to, the warranties of merchantability, fitness for a particular purpose and non-infringement of intellectual property rights and ETSI shall not be held liable in any event for any damages whatsoever (including, without limitation, damages for loss of profits, business interruption, loss of information, or any other pecuniary loss) arising out of or related to the use of or inability to use the software.

***Copyright Notification***

No part may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm except as authorized by written permission of ETSI.

The content of the PDF version shall not be modified without the written authorization of ETSI.

The copyright and the foregoing restriction extend to reproduction in all media.

© ETSI 2025.  
All rights reserved.

---

# Contents

Intellectual Property Rights .....	7
Foreword.....	7
Modal verbs terminology.....	7
Introduction .....	8
1 Scope .....	9
2 References .....	9
2.1 Normative references .....	9
2.2 Informative references.....	9
3 Definition of terms, symbols and abbreviations.....	11
3.1 Terms.....	11
3.2 Symbols.....	12
3.3 Abbreviations .....	12
4 Use Cases .....	12
4.1 Use case 1: Context-Based Dynamic HO Management for V2X.....	12
4.1.1 Background Information.....	12
4.1.2 Motivation.....	13
4.1.3 Proposed Solution .....	13
4.1.4 Benefits of O-RAN Architecture .....	14
4.1.5 Notes .....	14
4.2 Use case 2: Flight Path Based Dynamic UAV Radio Resource Allocation .....	14
4.2.1 Background Information.....	14
4.2.2 Motivation.....	14
4.2.3 Proposed Solution .....	14
4.2.4 Benefits of O-RAN Architecture .....	15
4.3 Use case 3: Radio Resource Allocation for UAV Application Scenario.....	15
4.3.1 Background Information.....	15
4.3.2 Motivation.....	16
4.3.3 Proposed Solution .....	16
4.3.4 Benefits of O-RAN Architecture .....	17
4.4 Use case 4: QoE Optimization .....	17
4.4.1 Background Information.....	17
4.4.2 Motivation.....	18
4.4.3 Proposed Solution .....	18
4.4.3.0 Non-RT RIC and Near-RT RIC Proposed Solutions .....	18
4.4.3.1 User-Centric QoE Connection Policy .....	19
4.4.3.2 User-Centric QoE Performance Policy, Non-RT RIC .....	20
4.4.3.3 User-Centric QoE Performance Policy, Near-RT RIC.....	21
4.4.3.4 AI/ML QoE Enhancements .....	21
4.4.3.5 Radio Performance Analytics.....	22
4.4.4 Benefits of O-RAN Architecture .....	23
4.5 Use case 5: Traffic Steering .....	23
4.5.1 Background Information.....	23
4.5.2 Motivation.....	23
4.5.3 Proposed Solution .....	23
4.5.3.1 Policy based traffic management optimization .....	23
4.5.3.2 Enrichment information based traffic steering optimization .....	24
4.5.4 Benefits of O-RAN Architecture .....	24
4.6 Use case 6: Massive MIMO Beamforming Optimization .....	24
4.6.1 Background Information.....	24
4.6.2 Motivation.....	25
4.6.3 Proposed Solution .....	25
4.6.3.0 Optimization Loops for mMIMO Beamforming.....	25
4.6.3.1 Non-RT Massive MIMO GoB Beam Forming Optimization .....	26
4.6.3.2 Near-RT Massive MIMO BeamBased Mobility Robustness Optimization (bMRO) .....	26

4.6.3.3	Near-RT Massive MIMO Beam Selection Optimization (BSO).....	27
4.6.4	Benefits of O-RAN Architecture .....	28
4.7	Use case 7: RAN Sharing .....	29
4.7.1	Background Information.....	29
4.7.2	Motivation.....	29
4.7.3	Proposed Solution.....	29
4.7.4	Benefits of O-RAN Architecture .....	31
4.8	Use case 8: QoS Based Resource Optimization .....	31
4.8.1	Background Information.....	31
4.8.2	Motivation.....	31
4.8.3	Proposed Solution.....	32
4.8.4	Benefits of O-RAN Architecture .....	32
4.9	Use case 9: RAN Slice SLA Assurance .....	33
4.9.1	Background Information.....	33
4.9.2	Motivation.....	33
4.9.3	Proposed Solution.....	33
4.9.4	Benefits of O-RAN Architecture .....	34
4.10	Use case 10: Multi-vendor Slices .....	34
4.10.1	Background Information.....	34
4.10.2	Motivation.....	34
4.10.3	Proposed Solution.....	35
4.10.3.1	Multi-vendor Slices.....	35
4.10.3.2	RAN Sharing.....	35
4.10.4	Benefits of O-RAN Architecture .....	36
4.11	Use case 11: Dynamic Spectrum Sharing (DSS).....	36
4.11.1	Background Information.....	36
4.11.2	Motivation.....	36
4.11.3	Proposed Solution.....	37
4.11.4	Benefits of O-RAN Architecture .....	38
4.12	Use case 12: NSSI Resource Allocation Optimization.....	38
4.12.1	Background Information.....	38
4.12.2	Motivation.....	38
4.12.3	Proposed Solution.....	39
4.12.4	Benefits of O-RAN Architecture .....	40
4.13	Use case 13: Local Indoor Positioning in RAN.....	40
4.13.1	Background Information.....	40
4.13.2	Motivation.....	40
4.13.3	Proposed Solution.....	40
4.13.4	Benefits of O-RAN Architecture .....	41
4.14	Use case 14: Massive MIMO SU/MU-MIMO Grouping Optimization.....	41
4.14.1	Background Information.....	41
4.14.2	Motivation.....	41
4.14.3	Proposed Solution.....	41
4.14.3.1	Solution 1 .....	41
4.14.3.2	Solution 2 .....	42
4.14.4	Benefits of O-RAN Architecture .....	43
4.15	Use case 15: O-RAN Signalling Storm Protection.....	43
4.15.1	Background Information.....	43
4.15.2	Motivation.....	44
4.15.3	Proposed Solution.....	44
4.15.4	Benefits of O-RAN Architecture .....	45
4.16	Use case 16: Congestion Prediction & Management.....	45
4.16.1	Background Information.....	45
4.16.2	Motivation.....	46
4.16.3	Proposed Solution.....	46
4.16.4	Benefits of O-RAN Architecture .....	47
4.17	Use case 17: Industrial IoT Optimization.....	47
4.17.1	Background Information.....	47
4.17.2	Motivation.....	48
4.17.3	Proposed Solution.....	48
4.17.3.0	Non-RT RIC and Near-RT RIC Optimization Solutions .....	48
4.17.3.1	Non-RT RIC solution of IIoT optimization .....	48

4.17.3.2	Near-RT RIC solution of IIoT optimization.....	49
4.17.4	Benefits of O-RAN Architecture .....	50
4.18	Use case 18: BBU Pooling to achieve RAN Elasticity.....	50
4.18.1	Background Information.....	50
4.18.2	Motivation.....	50
4.18.3	Proposed Solution.....	50
4.18.4	Benefits of O-RAN Architecture .....	51
4.19	Use case 19: Integrated SON Function.....	52
4.19.1	Background Information.....	52
4.19.2	Motivation.....	52
4.19.3	Proposed Solution.....	53
4.19.4	Benefits of O-RAN Architecture .....	54
4.20	Use case 20: Shared O-RU .....	54
4.20.1	Background Information.....	54
4.20.2	Motivation.....	54
4.20.3	Proposed Solution.....	55
4.20.3.1	Static allocation of resources between O-DU nodes .....	55
4.20.3.1.1	Introduction .....	55
4.20.3.1.2	Single Operator Use Case.....	56
4.20.3.1.3	Multi-Operator Use Case.....	56
4.20.3.2	Dynamic allocation of resources between O-DU nodes.....	57
4.20.3.2.1	Introduction .....	57
4.20.3.2.2	Single Operator Use Case.....	58
4.20.4	Benefits of O-RAN Architecture .....	58
4.21	Use case 21: Energy Saving .....	58
4.21.1	Background Information.....	58
4.21.2	Motivation.....	59
4.21.3	Proposed Solution.....	59
4.21.3.0	ES sub-use cases .....	59
4.21.3.1	ES sub-use case 1: Carrier and cell switch off/on .....	59
4.21.3.2	EE/ES sub-use case 2: RF channel switch off/on.....	60
4.21.3.3	ES sub-use case 3: Advanced Sleep Modes .....	61
4.21.4	Benefits of O-RAN Architecture .....	62
4.22	Use case 22: MU-MIMO Optimization.....	62
4.22.1	Background Information.....	62
4.22.2	Motivation.....	62
4.22.3	Proposed Solution.....	62
4.22.4	Benefits of O-RAN Architecture .....	63
4.23	Use case 23: Sharing Non-RT RIC Data with the Core .....	63
4.23.1	Background Information.....	63
4.23.2	Motivation.....	64
4.23.3	Proposed Solution.....	64
4.23.3.0	Proposed Options .....	64
4.23.3.1	SMO/Non-RT RIC Direct to 5GC Consumer with Façade of an NWDAF .....	64
4.23.3.2	SMO/Non-RT RIC to DCCF/MFAF with Façade of an NWDAF.....	66
4.23.3.3	SMO/Non-RT RIC to Separate NWDAF with Façade of OA&M.....	68
4.23.3.4	SMO/Non-RT RIC to Separate NWDAF with Façade of RAN NF.....	71
4.23.3.5	SMO/Non-RT RIC to External NWDAF with Façade of 3GPP MDA.....	72
4.23.4	Benefits of O-RAN Architecture .....	74
4.24	Use case 24: Industrial vision SLA Assurance.....	74
4.24.1	Background Information.....	74
4.24.2	Motivation.....	74
4.24.3	Proposed Solution.....	75
4.24.4	Benefits of O-RAN Architecture .....	76
4.25	Use case 25: Non-Public Network (NPN) RAN-Sharing via Midhaul for Multi-Operator Coverage.....	76
4.25.1	Background Information.....	76
4.25.2	Motivation.....	77
4.25.3	Proposed Solution .....	78
4.25.3.0	NPN RAN-Sharing solution.....	78
4.25.3.1	RAN Sharing to support multiple operators.....	78
4.25.3.1.0	NPN RAN-Sharing solution using O-RAN architecture .....	78
4.25.3.1.1	Spectrum Partitioning to support multiple operators .....	79

4.25.3.2	RAN-Sharing via Midhaul Architecture .....	79
4.25.3.2.0	NPN RAN-sharing via Midhaul .....	79
4.25.3.2.1	O-RAN impacts of Proposed Solution .....	80
4.25.3.2.2	Midhaul Connection .....	81
4.25.3.2.3	E2-Interface changes .....	82
4.25.3.2.4	O1-Interface changes .....	82
4.25.3.2.5	Mobility .....	82
4.25.4	Benefits of O-RAN Architecture .....	83
<b>Annex A (informative): Change history .....</b>		<b>84</b>
History .....		85

---

# Intellectual Property Rights

## Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the [ETSI IPR online database](#).

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

## Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

**DECT™, PLUGTESTS™, UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™, LTE™** and **5G™** logo are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM®** and the GSM logo are trademarks registered and owned by the GSM Association.

---

# Foreword

This Technical Report (TR) has been produced by ETSI Technical Committee Mobile Standards Group (MSG).

The content of the present document is subject to continuing work within O-RAN and may change following formal O-RAN approval. Should the O-RAN ALLIANCE modify the contents of the present document, it will be re-released by O-RAN with an identifying change of version date and an increase in version number as follows:

version xx.yy.zz

where:

- xx: the first digit-group is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc. (the initial approved document will have xx=01). Always 2 digits with leading zero if needed.
- yy: the second digit-group is incremented when editorial only changes have been incorporated in the document. Always 2 digits with leading zero if needed.
- zz: the third digit-group included only in working versions of the document indicating incremental changes during the editing process. External versions never include the third digit-group. Always 2 digits with leading zero if needed.

---

# Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the [ETSI Drafting Rules](#) (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

---

## Introduction

The present document provides O-RAN WG1 high level use case descriptions. Any multi-WG use case defined in O-RAN is expected to be documented in the present document. If the use case is to be studied further, it will be covered in O-RAN WG1 detailed use case specification next, and then in relevant WGs. It should be noted that not all of the use cases presented here are currently supported by O-RAN specifications and these use cases will be addressed in future O-RAN work.

---

# 1 Scope

The present document specifies potential O-RAN use cases as defined by O-RAN WG1 UCTG (Use Case Task Group). The use cases are described at a very high level, emphasizing how the use is enabled by O-RAN architecture along with basic input data expectations and resulting actions. These high level use cases are prioritized within O-RAN, and selected use cases are further detailed in O-RAN WG1 UCTG and relevant O-RAN WGs to define the requirements for O-RAN components and their interfaces.

---

## 2 References

### 2.1 Normative references

Normative references are not applicable in the present document.

### 2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

- [i.1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".
- [i.2] 3GPP TR 38.802: "Study on new radio access technology Physical layer aspects", Release 14, September 2017.
- [i.3] 3GPP TR 38.889: "Study on NR-based access to unlicensed spectrum", Release 16, December 2018.
- [i.4] 3GPP TSG-RAN WG3 Meeting #104.
- [i.5] ETSI EN 302 637-2: "Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 2: Specification of Cooperative Awareness Basic Service".
- [i.6] GSMA Future Networks: "Infrastructure Sharing: An Overview", June 2019.
- [i.7] GSMA NG.116: "Generic Network Slice Template", Version 2.0, October 2019.
- [i.8] López-Pérez David et al.: "A Survey on 5G Radio Access Network Energy Efficiency: Massive MIMO, Lean Carrier Design, Sleep Modes, and Machine Learning." IEEE Communications Surveys & Tutorials (2022).
- [i.9] 3GPP TS 22.261: "Service requirements for the 5G system; Stage 1", Release 16, December 2019.
- [i.10] 3GPP TS 23.285: "Architecture enhancements for V2X services", Release 16, June 2019.
- [i.11] 3GPP TS 23.501: "System Architecture for the 5G System (5GS); Stage 2", Release 16, December 2019.
- [i.12] 3GPP TS 26.247 "Transparent end-to-end Packet-switched Streaming Service (PSS); Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH)", Release 16, October 2020.
- [i.13] 3GPP TS 28.310: "Management and orchestration; Energy efficiency of 5G", Release 17, December 2021.

- [i.14] 3GPP TS 28.313: "Management and orchestration; Self-Organizing Networks (SON) for 5G networks", Release 16, December 2020.
- [i.15] 3GPP TS 28.530: "Management and orchestration; Concepts, use cases and requirements", Release 16, September 2019.
- [i.16] 3GPP TS 28.541: "5G Network Resource Model (NRM); Stage 2 and stage 3", Release 16, January 2020.
- [i.17] 3GPP TS 28.552: "Management and orchestration; 5G performance measurements", Release 16, September 2019.
- [i.18] 3GPP TS 36.300: "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2 (Release 16)", Release 16, January 2021.
- [i.19] 3GPP TS 36.902: "Evolved Universal Terrestrial Radio Access Network E-UTRAN); Self-configuring and self-optimizing network (SON) use cases and solutions (Release 9)", Release 9, April 2011.
- [i.20] 3GPP TS 37.340: "NR; Multi-connectivity; Overall description; Stage-2", Release 16, April 2020.
- [i.21] 3GPP TS 38.331: "NR; Radio Resource Control (RRC) protocol specification (Release 16)", Release 15, January 2021.
- [i.22] O-RAN WG6: "Orchestration Use Cases and Requirements for O-RAN Virtualized RAN".
- [i.23] O-RAN TS 29.520: "5G System; Network Data Analytics Services".
- [i.24] 3GPP TS 23.288: "Architecture enhancements for 5G System (5GS) to support network data analytics services".
- [i.25] 3GPP TS 29.510: "Network Function Repository Services, Stage 3".
- [i.26] 3GPP TS 23.502: "Procedures for the 5G System (5GS); Stage 2".
- [i.27] 3GPP TS 28.550: "Management and orchestration; Performance assurance".
- [i.28] 3GPP TS 28.532: "Management and orchestration; Generic management services".
- [i.29] O-RAN R1 interface: "General Aspects and Principles".
- [i.30] O-RAN.WG1.OAD: "O-RAN Architecture Description".
- [i.31] O-RAN.WG2.Non-RT-RIC-ARCH-TS: "Non-RT RIC: Functional Architecture".
- [i.32] Void.
- [i.33] 3GPP TS 38.174: "Technical Specification Group Radio Access Network; NR; Integrated access and backhaul radio transmission and reception", Release 16, v16.3.0.
- [i.34] 3GPP TS 32.130: "Network sharing; Concepts and requirements".
- [i.35] 3GPP TS 23.251: "Network sharing; Architecture and functional description".
- [i.36] 3GPP TS 22.101: "Technical Specification Group Services and System Aspects; Service aspects; Service principles", v16.3.0.
- [i.37] 3GPP TS 38.300: "Technical Specification Group Radio Access Network; NR; NR and NG-RAN Overall Description; Stage 2", v16.7.0.
- [i.38] 3GPP TS 38.401: "Technical Specification Group Radio Access Network; NG-RAN; Architecture description", Release 16, v16.6.0.
- [i.39] 3GPP TS 38.340: "Technical Specification Group Radio Access Network; NR; Backhaul Adaptation Protocol (BAP) specification", Release 16, v16.5.0.

- [i.40] 3GPP TS 38.473: "Technical Specification Group Radio Access Network; NG-RAN; F1 application protocol (F1AP)", Release 16, v16.6.0.
- [i.41] 3GPP TS 28.104: "3<sup>rd</sup> Generation Partnership Project; Technical Specification Group Services and System Aspects; Management and orchestration; Management Data Analytics (MDA)".

## 3 Definition of terms, symbols and abbreviations

### 3.1 Terms

For the purposes of the present document, the terms and definitions given in [i.1] and the following apply.

**NOTE:** A term defined in the present document takes precedence over the definition of the same term, if any, in [i.1].

**A1:** interface between non-RT RIC and Near-RT RIC to enable policy-driven guidance of Near-RT RIC applications/functions, and support AI/ML workflow

**A1 policy:** type of declarative policies expressed using formal statements that enable the non-RT RIC function in the SMO to guide the near-RT RIC function, and hence the RAN, towards better fulfilment of the RAN intent

**A1 Enrichment information:** information utilized by near-RT RIC that is collected or derived at SMO/non-RT RIC either from non-network data sources or from network functions themselves

**E2:** interface connecting the Near-RT RIC and one or more O-CU-CPs, one or more O-CU-UPs, and one or more O-DUs

**E2 Node:** Logical node terminating E2 interface. In the present document, O-RAN nodes terminating E2 interface are:

- for NR access: O-CU-CP, O-CU-UP, O-DU or any combination;
- for E-UTRA access: O-eNB.

**Intents:** declarative policy to steer or guide the behavior of RAN functions, allowing the RAN function to calculate the optimal result to achieve stated objective

**O-RAN near-Real-Time RAN Intelligent Controller (Near-RT RIC):** logical function that enables real-time control and optimization of RAN elements and resources via fine-grained data collection and actions over E2 interface

**O-RAN non-Real-Time RAN Intelligent Controller (Non-RT RIC):** logical function that enables non-real-time control and optimization of RAN elements and resources, AI/ML workflow including model training and updates, and policy-based guidance of applications/features in Near-RT RIC

**O-RAN Central Unit (O-CU):** logical node hosting O-CU-CP and O-CU-UP

**O-RAN Central Unit – Control Plane (O-CU-CP):** logical node hosting the RRC and the control plane part of the PDCP protocol

**O-RAN Central Unit – User Plane (O-CU-UP):** logical node hosting the user plane part of the PDCP protocol and the SDAP protocol

**O-RAN Distributed Unit (O-DU):** logical node hosting RLC/MAC/High-PHY layers based on a lower layer functional split

**O-RAN Radio Unit (O-RU):** logical node hosting Low-PHY layer and RF processing based on a lower layer functional split

**O1:** interface between management entities (NMS/EMS/MANO) and O-RAN managed elements, for operation and management, by which FCAPS management, Software management, File management can be achieved

**Radio Access Network (RAN):** in terms of the present document, any component below Near-RT RIC per O-RAN architecture, including O-CU/O-DU/O-RU

## 3.2 Symbols

Void.

## 3.3 Abbreviations

For the purposes of the present document, the abbreviations [given in [i.1] and the following] apply.

NOTE: An abbreviation defined in the present document takes precedence over the definition of the same abbreviation, if any, in [i.1].

AI/ML	Artificial Intelligence/Machine Learning
ASM	Advanced Sleep Mode
DCCF	Data Collection Coordination and Delivery Function
EC	Energy Consumption
EE	Energy Efficiency
eNB	eNodeB (applies to LTE)
ES	Energy Saving
FCAPS	Fault Configuration Accounting Performance, Security
gNB	gNodeB (applies to NR)
KPI	Key Performance Indicator
MFAF	Messaging Framework Adaptor Function
MIMO	Multiple Input, Multiple Output
NRF	Network Repository Function
NWDAF	NetWork Data Analytics Function
PRB	Physical Resource Block
QoE	Quality of Experience
RAT	Radio Access Technology
RIC	O-RAN RAN Intelligent Controller
SINR	Signal-to-Interference-plus-Noise Ratio
SSB	Synchronization Signal Block
UAV	Unmanned Aerial Vehicle
V2X	Vehicle to Everything
VM	Virtual Machine
SMO	Service Management and Orchestration
UDM	Unified Data Management

## 4 Use Cases

### 4.1 Use case 1: Context-Based Dynamic HO Management for V2X

#### 4.1.1 Background Information

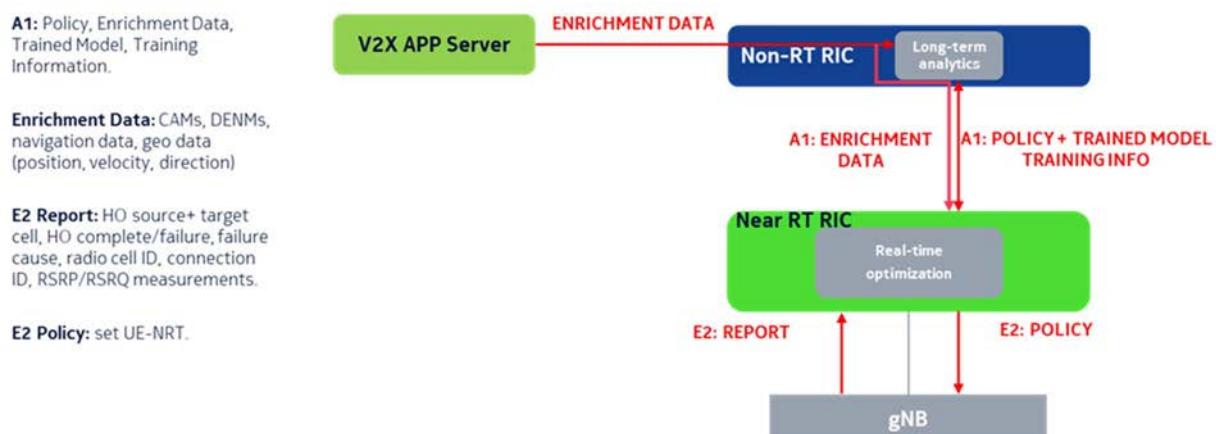
V2X communication allows for numerous potential benefits such as increasing the overall road safety, reducing emissions, and saving time. Part of the V2X architecture is the V2X UE (SIM + device attached to vehicle) which communicates with the V2X Application Server (V2X AS). The exchanged information comprises Cooperative Awareness Messages (CAMs) from UE to V2X AS [i.5], radio cell IDs, connection IDs, and basic radio measurements (RSRP, RSRQ, etc.).

## 4.1.2 Motivation

As vehicles traverse along a highway, due to their high speed and the heterogeneous natural environment V2X UE-s are handed over frequently, at times in a suboptimal way, which can cause handover (HO) anomalies: e.g. short stay, ping-pong, and remote cell. Such suboptimal HO sequences substantially impair the functionality of V2X applications. Since HO sequences are mainly determined by the Neighbour Relation Tables (NRTs), maintained by the xNBs, there is hardly room for UE-level customization.

This UC aims to present a method to avoid and/or resolve problematic HO scenarios by using past navigation and radio statistics in order to customize HO sequences on a UE level. To this end, the AI/ML functionality that is enabled by the Near-RT RIC is employed.

## 4.1.3 Proposed Solution



**Figure 4.1.3-1: Dynamic Handover Management for V2X use case**

In order to prevent anomalous HO sequences causing performance degradation for the V2Xs, an xApp can be built with two main functionalities as shown in figure 4.1.3-1. Applying long-term analytics by using AI/ML algorithms is the first function expected from the xApp. As it is suggested by O-RAN framework, non-real-time intelligent management of RAN functions are deployed in the Non-RT RIC. V2X AS maintains the UE-based HO events and mobility data which are shared with Non-RT RIC over O1 interface. Hence, Non-RT RIC can identify causes of HO anomalies and discover optimal HO sequences. Finally, a data base is maintained to keep track of resolutions to anomalies based on these feedbacks. The other function of the xApp is performing real-time optimization which is offered by trained ML model on the Near-RT RIC. Near-RT RIC will monitor UE specific real-time mobility context based on V2X data. Deployed ML model is used to detect/predict unexpected HO events and generate desired HO sequence which can eliminate/avoid HO anomaly. Finally, Near-RT RIC is expected to create and update UE-specific NRTs to improve performance of the V2X users.

The xNB is assumed to host, besides the default NRT, also UE-specific NRTs (UE-NRTs) for V2X (and potentially other types of) users. If the UE-NRT for a specific V2X user exists, it is used. If not, the default NRT remains valid.

The input samples for handover profiling can come from the V2X AS and the xNB:

- CAMs, navigation data, geo data (position, velocity, direction);
- Radio cell ID, connection ID, basic radio measurements (RSRP, RSRQ, etc.);
- Handover event (source and target cell IDs);
- Handover success / error codes.

The long-term analytics function and the real-time optimization function can be thought of as ML training and ML inference, respectively.

#### 4.1.4 Benefits of O-RAN Architecture

The long-term optimization function can jointly process UE level radio and navigation data. Both the analysis and the prediction functions can employ a range of AI/ML methods, whose availability is facilitated by the O-RAN architecture.

#### 4.1.5 Notes

An option for associating the data coming from the gNB and that coming from the V2X AS can be achieved by the ECGI+C-RNTI [i.10] and [i.18]. The ECGI uniquely identifies the cell, while the C-RNTI uniquely identifies a UE within a gNB. The ECGI+C-RNTI pair constitutes a globally unique UE identifier.

### 4.2 Use case 2: Flight Path Based Dynamic UAV Radio Resource Allocation

#### 4.2.1 Background Information

This use case provides the background, motivation, and requirements for the support the use case of flight path based dynamic UAV Radio Resource Allocation, allowing operators to adjust radio resource allocation policies through the O-RAN architecture, reducing unnecessary handover and improving radio resource utilization.

#### 4.2.2 Motivation

The field trials' results show that the coverage for low altitude is good and can provide various services for terrestrial UEs with good performance. However, since the site along the flight is mainly for terrestrial UEs, the altitude of the UAV is always not within the main lobe of the ground station antenna. And the side lobes give rise to the phenomenon of scattered cell associations particularly noticeable in the sky. The cell association pattern on the ground is ideally contiguous area where the best cell is most often the one closest to the UE. As the UE move up in height, the antenna side lobes start to be visible, and there is a possibility of the best cell no longer being the closest one. The cell association pattern in this particular scenario becomes fragmented especially at the height of 300 m and above. Hence, at higher altitudes, several challenges that lead to a different radio environment are:

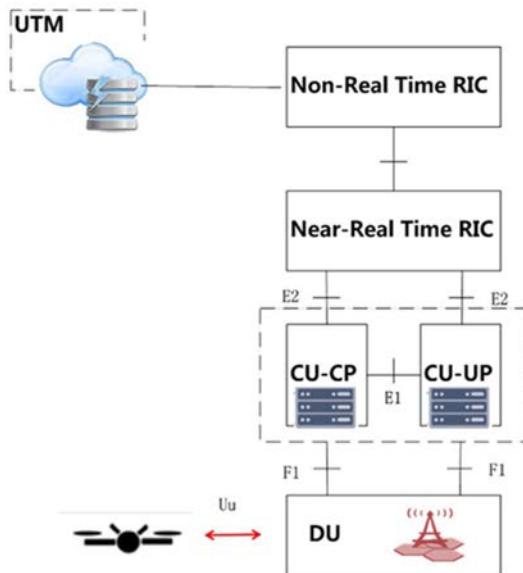
- LOS propagation/uplink interference.
- Poor KPI caused by antenna side lobes for base stations.
- Sudden drop in signal strength.

These challenges directly impact on the mobility performance of the drone and the service experience of the user. Hence, O-RAN would like to support the use case of flight path based dynamic UAV Radio Resource Allocation to resolve the above issues.

#### 4.2.3 Proposed Solution

To manage resource allocation of UAVs based on Flight Path, Non-RT RIC retrieves necessary Aerial Vehicles related metrics from network level measurement reports. Moreover, flight path information of Aerial Vehicle, climate information, flight forbidden/limitation area information and Space Load information are received from an application, e.g. UTM (Unmanned Traffic Management) for constructing/training relevant AI/ML model. For example, this could be UL/DL interference from/to Aerial vehicles, the detection of Aerial Vehicle UEs, and available radio resource (e.g. frequency, cell, beam, BWP, numerology) prediction. The Near-RT RIC supports deployment and execution of AI/ML model from Non-RT RIC. Based on this, Near-RT RIC can perform the radio resource allocation for UAVs considering the radio channel condition, flight path information and other application information with on-demand coverage. Moreover, Aerial Vehicles performance reports should be sent to Non-RT RIC for evaluation and optimization of the ML model.

The architectural context of the flight path based dynamic UAV Radio Resource Allocation use case is shown in figure 4.2.3-1.



**Figure 4.2.3-1: Use case of flight path based dynamic UAV Radio Resource Allocation**

#### 4.2.4 Benefits of O-RAN Architecture

An effective functional module, retrieving the application information, performing machine learning and training is not provided by current gNB/eNB architecture. Though, the O-RAN architecture components can collect both the acquired application information and radio environment information, and execute AI/ML models based on received information. The flight path based dynamic UAV Radio Resource Allocation mechanism can be supported by the RIC function modules, i.e. Non-RT RIC and Near-RT RIC. Therefore, the description of O-RAN support use case for flight path based dynamic UAV Radio Resource Allocation is provided.

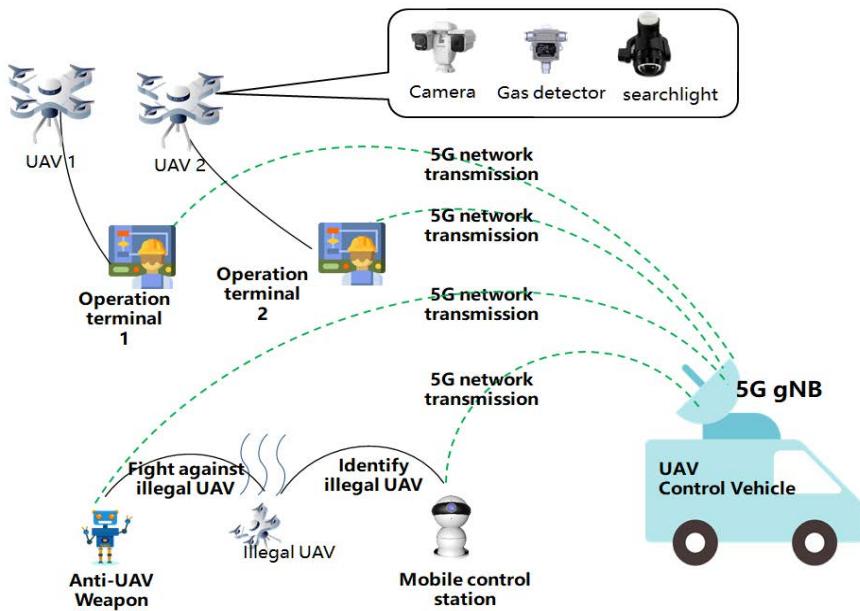
### 4.3 Use case 3: Radio Resource Allocation for UAV Application Scenario

#### 4.3.1 Background Information

As shown in figure 4.3.1-1, this scenario refers to a Rotor UAV flying at low altitude and low speed and carrying cameras, sensors and other devices mounted. The Operation Terminal uses the 5.8 GHz frequency band remote control the UAV for border/forest inspection, high voltage/base station inspection, field mapping, pollution sampling, and HD live broadcast.

At the same time, the UAV control mobile station and the UAV anti-weapon combination provide the UAV control, fight against illegal UAV and other services to ensure low-altitude safety in special areas.

The UAV Operation terminal, the anti-UAV weapon, and the UAV control mobile station are connected with the UAV Control Vehicle using 5G network. UAV Control Vehicle deploys network equipment, including O-CU,O-DU, the Non-RT RIC, the Near RT RIC function modules and Application Server (In this use case, it is an Edge computing Service Platform) to provide reliable network services through 5G networks.



**Figure 4.3.1-1: UAV Control Vehicle Application scenario**

### 4.3.2 Motivation

UAV terminal access is a new scenario of 5G networks. It has a large amount of high real-time data that is more suitable for local processing. For the existing radio resource allocation strategy, there is no solution for such uplink high-rate data transmission.

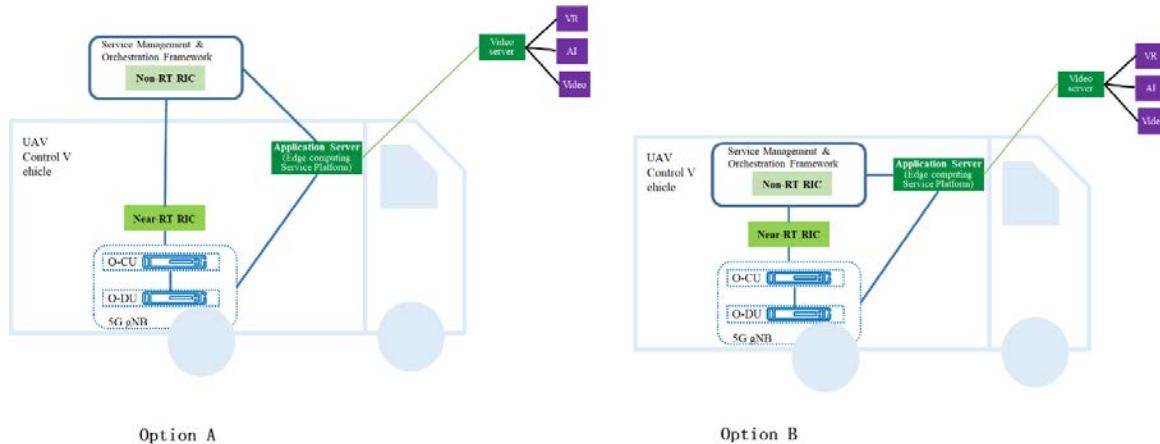
In this use case, UAV Control Vehicle is required to provide reliable network services through 5G networks. The data of the UAV terminal interacting with the network includes control data and application data, and the control data includes navigation commands, configuration changes, flight status data reporting, etc. Control data requires low latency and low bandwidth requirements. The application data includes 4K high-definition video data, which has obvious uplink and downlink service asymmetry, and the uplink has high requirements on network bandwidth.

The UAV Control Vehicle deploys edge computing services that application services and content are placed on the edge instead of core network so that local processing of video and control information can be managed. At the same time, real-time data services can be provided to third-party applications through a video server.

### 4.3.3 Proposed Solution

As shown in figure 4.3.3-1, this use case involves two options of network architecture. Option A is that O-CU, O-DU and Near-RT RIC are deployed on the Control Vehicle, Non-RT RIC and core network are deployed on the central cloud. The Control Vehicle is connected to core network and Non-RT RIC via fiber optics. Option B is a private network, all modules, including the gNB, Near-RT RIC, Non-RT RIC and the necessary core network function modules, are deployed in the Control Vehicle.

In both deployment options, radio resource allocation for UAV application use case requires some of the basic functionalities of the O-RAN components. Non-RT RIC should retrieve UE-level radio resource requirements from Application Server to generate UE based radio resource allocation policies. In this scenario, Near-RT RIC should support execution and interpretation of retrieved policies to create configuration parameters to be applied on the E2 nodes. In addition to UE-specific radio resource adjustment with respect to received parameters, E2 nodes should provide information about UE registration or UE status change to Near-RT RIC so that it can be transferred to Application Server. As it is stated in clause 4.3.2, UAV terminals require low latency and high throughput in the uplink direction. For this reason, Application Server is needed to receive user plane data from UAV terminals.



**Figure 4.3.3-1: Network architecture for UAV Control Vehicle Application scenario**

#### 4.3.4 Benefits of O-RAN Architecture

The 5G network supports real-time high-definition video transmission and remote low-latency control of UAV, and finally provides various industry services such as inspection, security, surveying and mapping.

In the UAV Control Vehicle Application scenario, there are a small amount of control data interaction requirements between the terminal and the network interaction, as well as the large bandwidth requirements for uploading HD video.

The service asymmetry raises new requirements for resource allocation of the gNB. At the same time, the existing network operation and maintenance management platform (OSS system) can only optimize the parameters of a specific group of UEs, but not for individual users. Therefore, under the O-RAN architecture, the radio resource requirements for different terminals are sent to the gNB for execution by means of the Near-RT RIC function module.

The UAV control vehicle has flexible layout features. In this use case, the application service and content are deployed on the edge computing platform instead of the core network; the Non-RT RIC function module is used to schedule radio resources instead of the core network's QoS mechanism. In this way, the load and overhead of the core network can be reduced, and the forwarding and processing time in data transmission can be reduced, and the delay can be reduced.

### 4.4 Use case 4: QoE Optimization

#### 4.4.1 Background Information

Highly demanding 5G native applications like Cloud VR and industrial automation are both bandwidth consuming and latency sensitive. Demanding contemporary 4G and 5G applications like online multiplayer gaming and connected vehicles are today often handled in a best effort way with low or no application specific optimization. These traffic-intensive and highly interactive applications are not well served by current semi-static QoS framework which does not efficiently satisfy diversified QoE requirements. These requirements can vary during an application lifetime, especially taking into account potentially significant fluctuation of radio transmission capability and applications with dynamic performance requirements. It is expected that QoE estimation/prediction from application level can help deal with such uncertainty and improve the efficiency of radio resources, and eventually improve user experience and yield a more efficient use of RAN resources. Furthermore, an increased set of mobile applications with varying QoE demands will increasingly become unmanageable if semi-static profiles are "preloaded" into the relevant RAN nodes without a more automated closed-loop approach. RAN performance exposure to an external application are also envisioned to be helpful for the application to improve the user experience.

## 4.4.2 Motivation

The main objective is to ensure QoE optimization be supported within the O-RAN architecture and its open interfaces in a way that allows per-user, slice or 5QI flow modification of RAN behavior, features, scheduling procedures and other configuration based on user application requirements or other input. This includes:

- 1) A user-centric QoE policy approach. Input from external systems, such as user applications, can be used to set or request specific RAN behavior automatically and without preloading of static configuration data and QoS profiles into E2 Nodes. Feedback is provided to these systems on the acceptance of the request.
- 2) A user-centric QoE update and feedback approach. During a user or application session lifetime, a closed-loop including the user application or other external input can provide enriched performance targets to the RAN, varying the targeted performance in response to these external factors. Feedback is provided to the requesting source on the RAN performance for measurement against an SLA or for the application itself to take action to improve the user QoE.
- 3) Using AI / ML approaches embedded in the RAN. Multi-dimensional data, e.g. user traffic data, QoE measurements, network measurement report, can be acquired and processed via ML algorithms to support traffic recognition, QoE prediction, QoS enforcement decisions. ML models can be trained offline and model inference will be executed in a real-time manner.
- 4) Radio performance analytics and exposure. Based on the data analysis and ML algorithms, Near-RT RIC provides either statistics or predictions on the cell level or UE level, e.g. traffic rate, latency, packet loss rate. Furthermore, Near-RT RIC can expose the real time RAN analytics information to the external applications, and helps external applications to execute logic control, e.g. TCP sending window adjustment, video coding rate selection to improve the user QoE.

The use case focus is a general solution that supports any specific QoE use case (e.g. Cloud VR, video, gaming, connected vehicles, etc.).

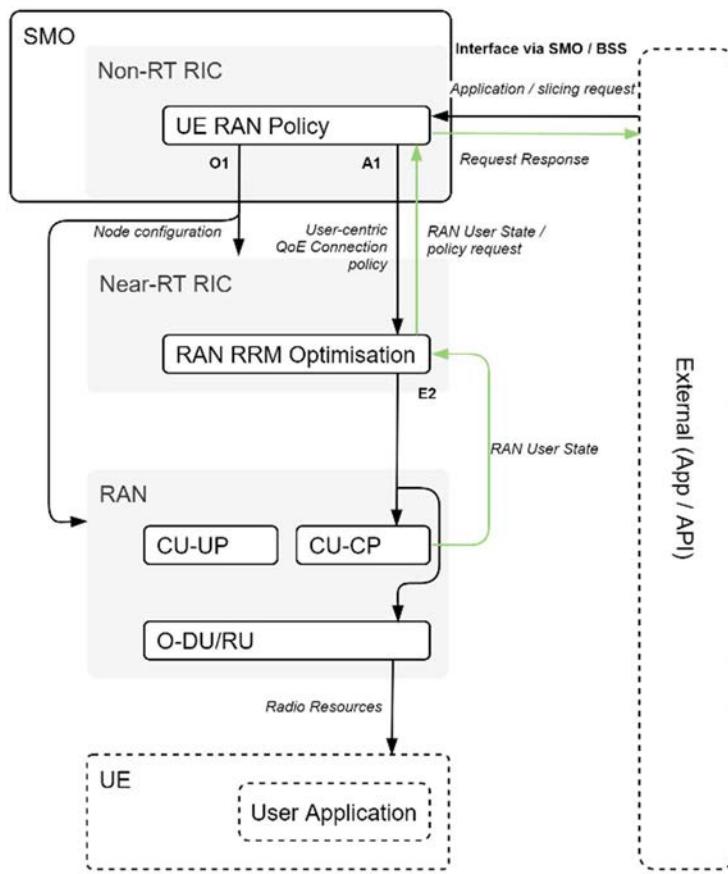
## 4.4.3 Proposed Solution

### 4.4.3.0 Non-RT RIC and Near-RT RIC Proposed Solutions

Traditional technologies involve manual configuration of RAN parameters for the congested cells to improve QoE of the network users. This can include per-site or per-cell parameter modification to target congested cells, or deployment of network-wide configurations to create application or user specific handling policies. The O-RAN architecture facilitates QoE optimization in a real-time way with proactive closed-loop network optimization, both within the RAN and including input from external applications. This will improve the way congested cells are detected and automatically allocate resources based on the end user experience, whose demands can change over time.

The proposed solution consists of O-RAN components, Non-RT RIC, Near-RT RIC and E2 Nodes, empowered with Machine Learning algorithms and user, slice or 5QI flow-centric feature and QoS modification interfaces. The solution introduces a "User RAN Policy", hosted at the Non-RT RIC (clause 4.4.3.2) or Near-RT RIC (clause 4.4.3.3). The User RAN Policy will be instantiated as an rApp or xApp which will apply the operator's desired QoE configuration for specific user, slice or 5QI flow types on the network in response to requests from external systems or in response to UE mobility, and provide feedback and reporting mechanisms, including SLA information, to external systems which are hosting the user application(s). The solution also includes exposing per-user or per-cell radio performance analytics information to external applications, and the application can optimize user QoE based on the RAN analytics information.

#### 4.4.3.1 User-Centric QoE Connection Policy



**Figure 4.4.3.1-1: QoE Connection Policy application concept**

This solution is expected to be invoked in response to changes in the applications being delivered to a UE, updates in the intended use-case handling behavior by an operator, implementation of a new network slice within the network, or mobility of a UE to a node which requires an update of the relevant handling policy.

The concept of QoE Connection Policy application is shown in figure 4.4.3.1-1.

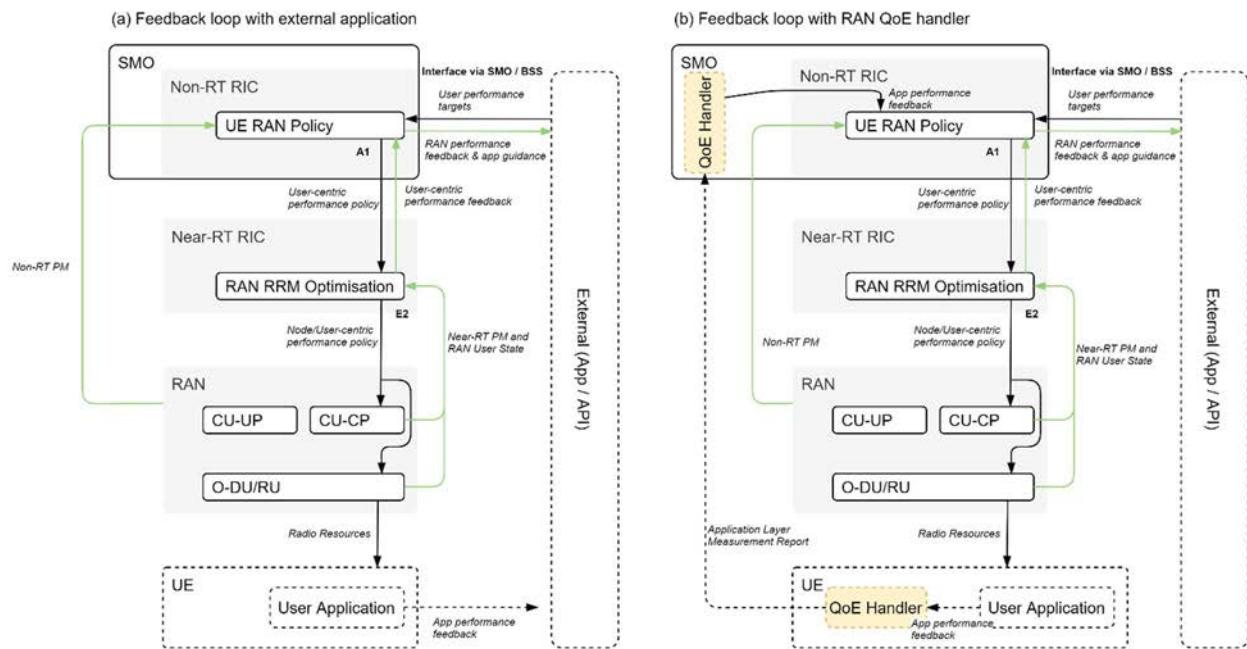
The Non-RT RIC hosts a RAN QoE connection policy configuration which can be modified by operators to apply RAN connection level behaviors to a network slice, 5QI flow per device, specific user types or specific user ID(s).

The Non-RT RIC will use the O1 interface to activate and configure the features of a particular QoE Connection Policy as they are needed across the network, while the A1 interface is used to request policy updates from the Non-RT RIC by specific network nodes (Policy Request) and by the Non-RT RIC to communicate the performance handling features that apply to a specific user, slice or 5QI flow (User-centric QoE Connection Policy).

For example, in response to an A1 interface Policy Request: The O1 interface is first used to implement or update a constellation of carrier aggregation configuration, traffic steering, mobility, power control or scheduling priority or other behaviors (e.g. DRX) as a constellation of features which can be referred to by a feature activation policy index. The A1 interface is then used to indicate the users or flows this set of features is to be applied to.

E2 nodes provide updated RAN user state information with sufficient granularity to the Near-RT RIC to trigger Policy Requests as needed.

#### 4.4.3.2 User-Centric QoE Performance Policy, Non-RT RIC



**Figure 4.4.3.2-1: QoE Performance Policy application concept, Non-RT RIC version**

This solution is expected to be invoked in response to changes in the user experience which could happen at any time during the lifetime of an application or session. This will typically operate as a continuous closed-loop including feedback from the user application(s).

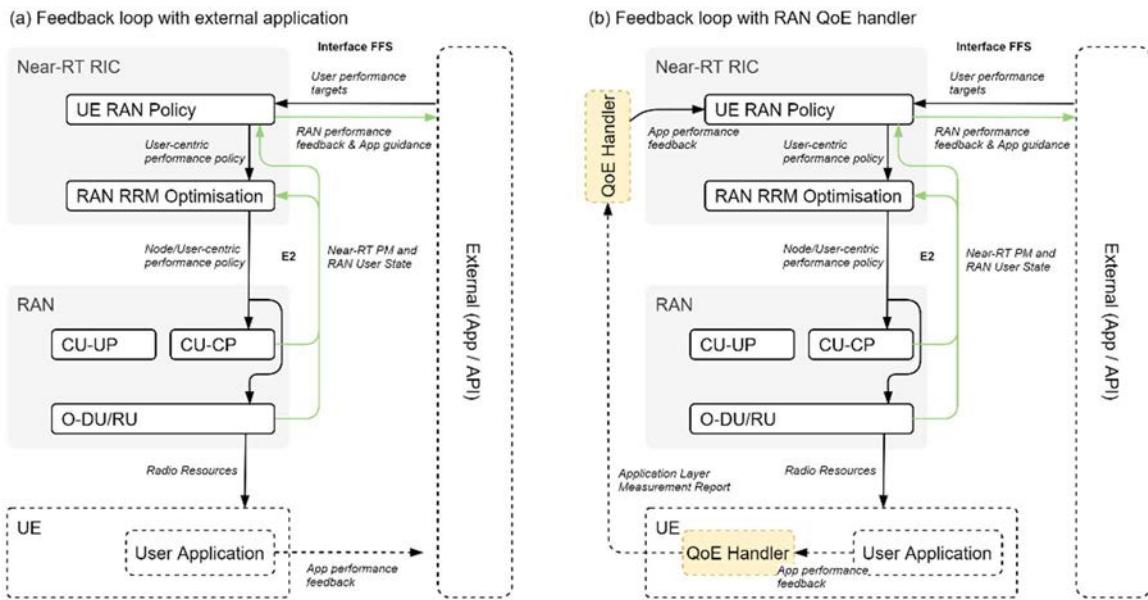
The concept of QoE Performance Policy application, Non-RT RIC version is shown in figure 4.4.3.2-1.

In (a), the Non-RT RIC hosts a RAN QoE performance policy configuration which accepts input from external systems (user application or via other SMO/BSS functions) that provide closed-loop feedback on the user experience. This will be translated into a User-Centric performance policy which is communicated via the A1 interface towards the Near-RT RIC which in turn, enforces these policies on the set of E2 nodes which currently host UEs of the type configured in the performance policy. For example, this policy would include specific latency, bitrate, and jitter targets which are used in conjunction with the connection policy described at 4.4.3.1 to determine how a specific user is handled in real time.

An alternative, optional approach is shown at (b), which instead uses Application Layer Measurement Reporting from the UE to the RAN as defined in [i.21] (*MeasReportAppLayer* RRC Container) and [i.12] for LTE, or equivalent when standardized for NR. This approach is complementary to (a) and will be limited to those user applications with access to UE AT interfaces or appropriate middleware, and so is not universally applicable. The *MeasReportAppLayer* is forwarded from the RAN to a QoE handling node (outside of the O-RAN architecture) and provides input to the Non-RT RIC analogous to the external system but with reduced latency.

E2 nodes provide updated User State information with sufficient granularity to allow for the distribution of connection policy, and PMs with required granularity to SMO to report on performance KPIs given the performance policy targets. This is then used to determine if an agreed performance target or SLA has been met, or alternatively, to indicate the RAN performance to an external application which can itself take action to improve the user QoE.

#### 4.4.3.3 User-Centric QoE Performance Policy, Near-RT RIC



**Figure 4.4.3.3-1: QoE Performance Policy application concept, Near-RT RIC version**

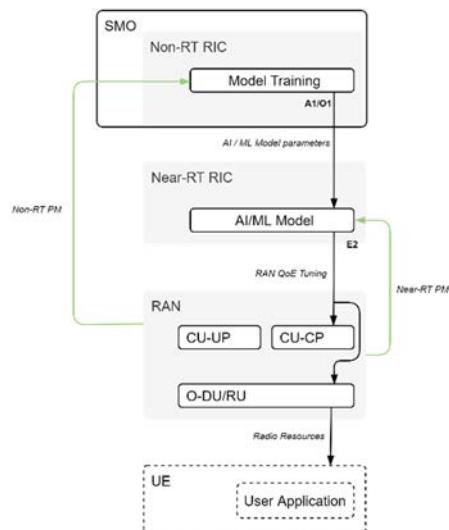
This solution has the same set of usage patterns as described at 4.4.3.2, except the RAN QoE performance policy configuration is moved to the Near-RT RIC. This solution is only expected to be necessary to support QoE input from edge-hosted applications co-located with the Near-RT RIC or CU-UP nodes.

The concept of QoE Performance Policy application, Near-RT RIC version is shown in figure 4.4.3.3-1.

In (a), the edge hosted external app provides input to the Near-RT RIC. The specific interface used is not defined here, and can be vendor specific or part of a future O-RAN release. In the (b) variant, which instead uses Application Layer Measurement Reporting [i.21] and [i.12], the QoE handler is also implemented at the edge location, with *MeasReportAppLayer* RRC containers forwarded for handling and then providing input to the Near-RT RIC rather than relying on the external app.

#### 4.4.3.4 AI/ML QoE Enhancements

The concept of AI/ML QoE enhancements is shown in figure 4.4.3.4-1.

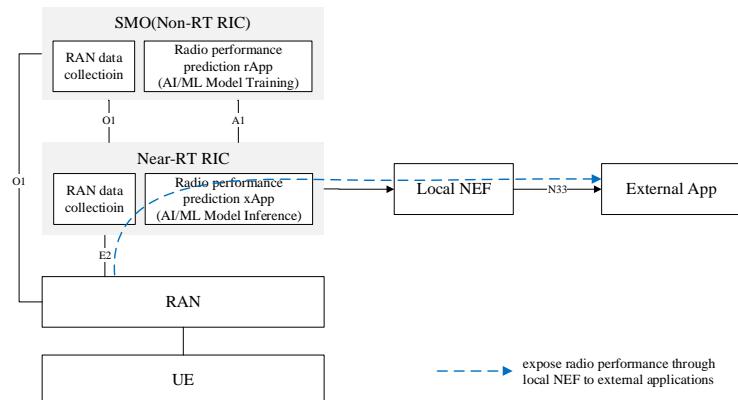


**Figure 4.4.3.4-1: AI/ML QoE enhancements concept**

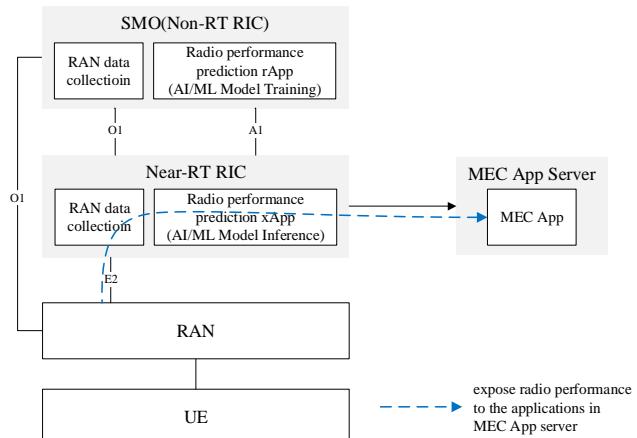
The Non-RT RIC will construct AI/ML models trained with data retrieved from SMO, network level measurements and policies to be sent Near-RT RIC for managing RAN parameters. The ML model will be deployed in the Near-RT RIC to assist QoE optimization such as making predictions on application/traffic types, QoE and available bandwidth.

To achieve all these functions, E2 Nodes should provide the PMs with required granularity to SMO over O1 and to the Near-RT RIC over E2. Also, RRM behaviour updates should be allowed by E2 Nodes through E2 to support QoS enforcement.

#### 4.4.3.5 Radio Performance Analytics



**Figure 4.4.3.5-1: Radio Performance Analytics Information Exposure, through Local NEF**



**Figure 4.4.3.5-2: Radio Performance Analytics Information Exposure, through MEC**

This solution is expected to be invoked when the external application requests for RAN performance from near-RT RIC. In response to application requests, near-RT RIC subscribes measurements through E2 interface and runs radio performance prediction model to generate predicted performance for a specific UE or cell.

As shown in figure 4.4.3.5-1, Near-RT RIC exposes radio performance prediction information through Local NEF (defined in 3GPP) to External Apps. As shown in figure 4.4.3.5-2, Near-RT RIC exposes radio performance prediction information directly to MEC App deployed in MEC App server (defined in ETSI MEC).

External application receives radio performance prediction information and executes logic control based on the service and user requirements, e.g. for 4K/8K/VR videos, application could execute TCP sending window adjustment, video coding rate selection to improve user experience.

#### 4.4.4 Benefits of O-RAN Architecture

The proposed solution to support a real-time QoE optimization for any specific use e.g. Cloud VR, video. It requires features provided by O-RAN architecture to orchestrate these requirements via interoperable A1/O1/E2 interfaces, and including input from external systems and user applications. Retrieving measurement metrics, AI/ML training and executing the AI/ML model in near-real time are also offered by the O-RAN architecture.

### 4.5 Use case 5: Traffic Steering

#### 4.5.1 Background Information

5G systems will support many different combinations of access technologies namely; LTE (licensed band), NR (licensed band), NR-U (unlicensed band), Wi-Fi (unlicensed band), (see TR 38.889 [i.3]). Several different multi-access deployment scenarios are possible with 5GC to support wide variety of applications and satisfy the spectrum requirements of different service providers:

- Carrier aggregation between licensed band NR (Primary Cell) and NR-U (Secondary Cell)
- Dual connectivity between licensed band NR (Primary Cell) and NR-U (Secondary Cell)
- Dual connectivity between licensed band LTE (Primary Cell) and NR-U (Secondary Cell)
- Dual connectivity between licensed band NR (Primary Cell) and Wi-Fi (Secondary Cell)

NOTE: The scenario of dual connectivity between NR and Wi-Fi is for future study.

The traffic steering use case allows, using the A1 interface, flexibly configure the desired optimization policies and utilize the appropriate performance criteria to proactively manage user traffic across different access technologies. A1 interface can also provide the enrichment information, e.g. radio fingerprint information based on the data analytics of the historical RAN data

#### 4.5.2 Motivation

Imbalances in the traffic load across cells of different access technologies, or variances in their available bandwidth and QoS attributes, can give rise to situations with suboptimal spectrum utilization and negative impact on the user experience. The 3GPP Self-Organizing Network (SON) function Mobility Load Balancing (MLB) is the legacy approach to improve the user experience by balancing the load through optimization of the handover triggers and handover decisions using load information shared between neighbouring cells. Normally all users are treated equally in this respect. In addition, the statistical characteristics of the radio network information have not been fully exploited and utilized to enhance the network and user experience performance.

#### 4.5.3 Proposed Solution

##### 4.5.3.1 Policy based traffic management optimization

Impairments of the user experience due to local load imbalances across cells, or due to variances in their available bandwidth offered, are example scenarios addressed by traffic management policies over the A1 interface. The traffic management policy allows allocating cells in order of priority to individual users, for the control plane and user plane respectively. A traffic management policy could be issued for any reason, e.g. reasons not known to the RAN.

The Non-RT RIC monitors the user experience by UE level performance measurements and the resource utilization on cell level. The Non-RT RIC assesses the observed performance vs. the expected service level requirements. If the requirements are breached the Non-RT RIC locates the cell where the breach is detected and assesses the local load or bandwidth conditions in the associated cells. It can desire to relocate one or more users to other cells e.g. in order to increase their available radio resources, offer increased bandwidth, desire to off-load a certain cell to improve conditions for the remaining users, or for any other reason.

Further in a multi-access system, apart from selecting the appropriate access technology to steer the traffic, there is a need to switch the traffic across different access technologies based on changes in radio environment and application requirements and even split the traffic across multiple access technologies to satisfy performance requirements. The different types of traffic and frequency bands in a commercial network can require appropriate bearer selection (Master Cell Group (MCG) bearer, Secondary Cell Group (SCG) bearer, Split bearer) and bearer type change for achieving load balancing, low latency and best in class throughput in a multi-access scenario with 5GC networks., (see 3GPP TS 37.340 [i.20]).

The Non-RT RIC creates corresponding traffic management policies directed to identify UEs expressing the priority order of cells to be allocated to each one for their control planes and user planes respectively. The policies are sent over the A1 interface to the Near-RT RIC, who uses the policies when enforcing the radio resource control.

#### 4.5.3.2 Enrichment information based traffic steering optimization

The current traffic management solutions are usually implemented by relocating users among cells, which highly depends on the measurement report (MR) from the UE feedback. The statistical characteristics of the radio network and the UE behaviors information have not been fully exploited and utilized to enhance the network and user experience performance.

The enrichment information can be utilized to optimize traffic management performance. The enrichment data could include Network Radio fingerprint information, UE trajectory information (e.g. way points in cell-level or beam-level), UE mobility profile (e.g. stationary, horizontal, vertical speed), UE service type (e.g. delay sensitivity or bit error rate sensitive), traffic pattern (e.g. average UL/DL packet size, periodicity). With the assistance of the enrichment information provided by Non-RT RIC, Near-RT RIC can efficiently derive optimized solutions.

For example Non-RT RIC can derive the radio fingerprint enrichment information via the data statistical analysis of historical measurement results. The radio fingerprint information captures the mapping relationship of the intra-frequency measurement results and the inter-frequency measurements. Then Non-RT RIC can deliver it to Near-RT RIC to help it predict the inter-frequency measurement, reduce the unnecessary inter-frequency measurement, accelerate the process of traffic optimization and improve the network system performance and user experience.

#### 4.5.4 Benefits of O-RAN Architecture

The use case explores the opportunity with the policy based A1 interface allowing to address specific users to obtain an intent driven by e.g. UE specific service level requirements.

### 4.6 Use case 6: Massive MIMO Beamforming Optimization

#### 4.6.1 Background Information

Massive MIMO is seen as one of the key technologies for 5G. Due to the multi-antenna transmission and reception, this technology can inherently provide diversity and improve capacity by targeting high gain antenna beams towards one or multiple subscribers, thus improving the receive power levels and spatially filtering the interference from neighboring subscribers and transmission points. In addition, a spatial multiplexing operations regime can improve the network capacity by transferring multiple data streams towards/from one or different subscribers utilizing a spatial re-use of the scarce time/frequency resource blocks. Further advantages include the controlling of electromagnetic (EM) emissions or advanced network management technologies like beam shaping, beam-based load balancing, optimized beam mobility, adaptive cell coverage areas, especially in highly dense urban 3D environments and mobile subscribers. In order to further optimize networks, e.g. maximize spectral efficiency, optimize coverage, or maximize cell capacity, fully digital beamforming (BF) methods are to be employed for below 6 GHz frequency wireless telecommunications. Grid of Beams (GoB) is a BF method which aims at selectively covering regions of interest with a suitable subset of radio beams (chosen from a dictionary of possible beams). Beam-based Mobility Robustness Optimization is a BF method enhancing beam specific mobility performance e.g. by adding beam specific individual offsets. Mobility Robustness Optimization is a well-known SON concept [i.19] and [i.21]. First supporting measurements have been specified in LTE in Release 9 [i.18].

## 4.6.2 Motivation

The massive MIMO BF optimization use case aims at proactively and continuously improving cell and/or user-centric coverage, capacity and mobility performance in a massive MIMO deployment area by adapting beam configurations (e.g. number of beams, beam boresights, vertical beam widths, horizontal beam widths, beam black lists/white lists, beam mobility thresholds) in a multi-cell, possibly multi-vendor, deployment scenario to non-real time (e.g. 3D construction, 3D terrain topology, network, weather seasons, intra-day cell splitting/merging/shaping, traffic distribution, beam conflicts) and near-real time (e.g. moving users/hotspots, changing traffic distribution, crowd source data) changes. The high number of configuration parameters per array antenna, the amount of available measurement input data, the complexity, pro-activeness as well as non- and near-real time requirements suggest the application of machine learning techniques of input data analytics as well as use case decision generation.

The objective of this use case is to allow the operator to flexibly configure Massive MIMO system parameters by means of policies, configuration or machine learning techniques, according to objectives defined by the operator.

## 4.6.3 Proposed Solution

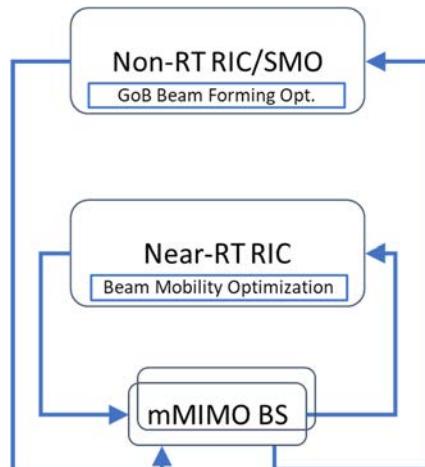
### 4.6.3.0 Optimization Loops for mMIMO Beamforming

- 1) The performance of BF is strongly dependent on the choice of the beam pattern  
Traditional GoB solutions do not take contextual/per-site information into account and operate with a narrow set of manually predefined beam patterns. Contextual/per-site information is, among others, comprised of cell geometry and other cell parameters, antenna array parameters, real-time or long-term (e.g. seasonal) traffic patterns, real-time or long-term user mobility patterns etc. One goal of this use case is to demonstrate that by taking contextual/per-site information into account, GoB/mMIMO BF can enhance the network performance by allowing the operator to perform extensive BF customization and optimization.
- 2) A key factor in multiuser (MU) mMIMO is the management of Beam Failures (BFAs)  
In order to prevent connectivity and user experience degradation, two AI/ML based Beam Mobility Optimization solutions are proposed.

In this use case three optimization loops for mMIMO BF are proposed (see figure 4.6.3.0-1):

- 1) An outer, Non-RT loop Massive MIMO GoB Beam Forming Optimization
- 2) An inner, Near-RT loop Massive MIMO Beam-based Mobility Robustness Optimization
- 3) An inner, Near-RT loop Massive MIMO Beam Selection Optimization

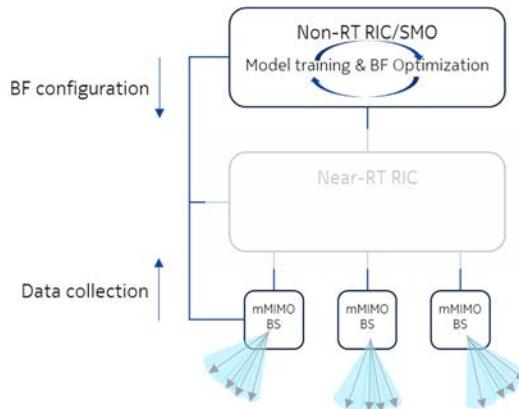
The two optimization loops can be implemented and deployed either individually or jointly in a nested fashion.



**Figure 4.6.3.0-1: Non-RT and Near-RT Optimization loops**

#### 4.6.3.1 Non-RT Massive MIMO GoB Beam Forming Optimization

The concept of Non-RT Beam Forming (BF) Optimization is shown in figure 4.6.3.1-1.



**Figure 4.6.3.1-1: Non-RT Beam Forming (BF) Optimization**

Non-RT RIC hosts an application with long-term analytics function (=ML training, Non-RT RIC), whose task is to collect, process and analyze antenna array parameters, cell performance KPIs, UE mobility/spatial density data, traffic density data, interference data and BF gain/beam RSRP and MDT measurement data.

The input data for BF optimization training and inference can be comprised of [i.2] antenna timing advance and Angle-of-Arrival measurements (for positioning estimation unless derived from another entity), aggregated & preprocessed beam-based reference signal measurements (average) traffic density measurements (across the respective mMIMO spatial grid) with associated positioning information, CSI measurements or, power headroom reports (PHRs), neighboring cells' beams/interference information and beam RSRPs/gains as well as cell performance measurements such as handover and beam failure statistics.

The output of the BF optimization inference can be optimized BF configuration, number of beams, beam elevation, beam horizontal & vertical widths and power allocation of beams.

The long-term analytics function and the BF optimization function can be thought of as ML training and ML inference, respectively.

Operator objectives can include desired coverage, defined in terms of the cell geometry (SSB beams), cell capacity requirements (CSI-RS beams), per UE cell edge throughput and traffic density weighted average RSRP and BF gain requirements. (E.g. the operator might wish to implement the strategy to cover low-density areas with wide beams and high-density areas with narrow beams.)

#### 4.6.3.2 Near-RT Massive MIMO BeamBased Mobility Robustness Optimization (bMRO)

The concept of Near-RT Beam-based Mobility Robustness Optimization is shown in figure 4.6.3.2-1.

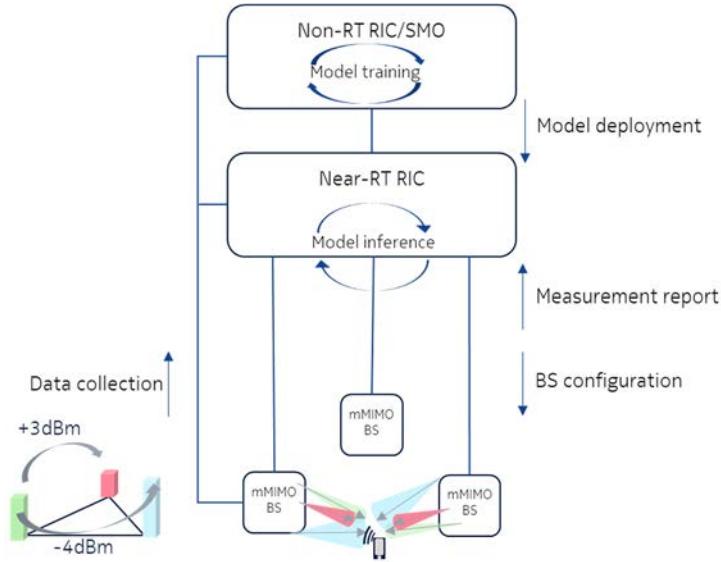
Near-RT RIC hosts an xApp with BF optimization function (=ML inference, Near-RT RIC), whose task is to monitor UE/traffic density measurements, UE positioning information, beam based mobility and beam failure KPIs and select/deploy optimized mMIMO BF parameters to E2 nodes.

The near-RT RIC can host an xApp to optimize inter-cell beam mobility such as a beam-based Mobility Robustness Optimization. In this case the Near-RT RIC might for instance configure beam individual offsets for inter-cell mobility decisions.

The Near-RT bMRO function can run individually, without the Non-RT BF Optimization. However, it can be deployed in a nested fashion within a Non-RT BF Optimization loop. In that case, upon change of the GoB configuration, the Near-RT bMRO function is reset or reconfigured. There are several options for coordinating the outer and the inner loops such as:

- 1) The outer loop's output comes from a finite set of configurations, and to each configuration the inner loop employs a trained model.

- 2) The inner loop employs a reinforcement or adaptive learning technique which is reset upon change of the GoB configuration, or, depending on implementation, adapted to it.



**Figure 4.6.3.2-1: Near-RT Beam-based Mobility Robustness Optimization**

Non-RT RIC hosts an application with long-term analytics function (=ML training, Non-RT RIC), whose task is to collect and analyze underlying GoB configuration, if GoB configuration exists, beam mobility and failure statistics, L1/L2 RSRP values, potential source-target beam pairs.

Near-RT RIC hosts an xApp with bMRO Optimization function (=ML inference, near-RT RIC), whose task is to monitor potential source-target beam pairs and optimize beam mobility for scheduling by managing user-beam pairing.

The input data for Beam Mobility Robustness Optimization training and inference can be comprised of [i.2] per-user measurements (e.g. RSRP, SINR, etc.), handover failure statistics (e.g. overall HO failures, too early or to late or HO to wrong cell), such as number of BFAs, times of BFAs, BFA rate etc., per-user potential source-target beam pairs and neighbouring cells' beams/interference information.

The output of the bMRO optimization function can be [i.19] adjusted offsets for candidate source-target beam pairs for beam mobility.

The long-term analytics function and the bMRO function can be thought of as ML training and ML inference, respectively.

Operator objectives can include minimization of BFA rate for a group of users (e.g. high-mobility users).

#### 4.6.3.3 Near-RT Massive MIMO Beam Selection Optimization (BSO)

The concept of Near-RT Beam Selection Optimized (BSO) function is shown in figure 4.6.3.3-1.

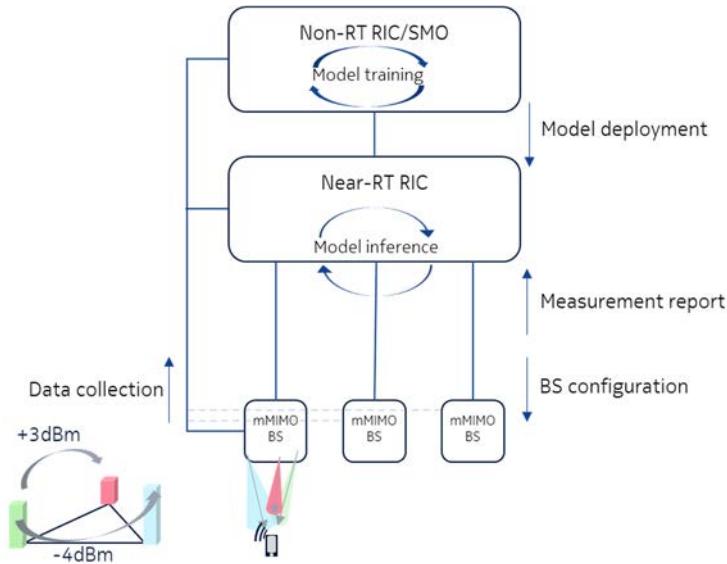
Near-RT RIC hosts an xApp with BF optimization function (=ML inference, Near-RT RIC), whose task is to monitor UE/traffic density measurements, UE positioning information, beam based mobility and beam failure KPIs and select/deploy optimized mMIMO BF parameters to E2 nodes.

The Near-RT RIC can host an xApp to optimize intra-cell beam mobility. In this case the Near-RT RIC might for instance configure parameters for intra-cell beam switching decisions.

The Near-RT BSO function can run individually, without the Non-RT BF Optimization. However, it can be deployed in a nested fashion within a Non-RT BF Optimization loop. In that case, upon change of the GoB configuration, the Near-RT BSO function is reset or reconfigured. There are several options for coordinating the outer and the inner loops such as:

- 1) The outer loop's output comes from a finite set of configurations, and to each configuration the inner loop employs a trained model.

- 2) The inner loop employs a reinforcement or adaptive learning technique which is reset upon change of the GoB configuration, or, depending on implementation, adapted to it.



**Figure 4.6.3.3-1: Near-RT Beam Selection Optimized (BSO) function**

Non-RT RIC hosts an application with long-term analytics function (=ML training, Non-RT RIC), whose task is to collect and analyze underlying GoB configuration, if GoB configuration exists, beam mobility and failure statistics, L1/L2 RSRP values, potential source-target beam pairs.

Near-RT RIC hosts an xApp with BSO function (=ML inference, near-RT RIC), whose task is to monitor potential source-target beam pairs, and to optimize beam mobility for scheduling by managing user-beam pairing.

The input data for BSO training and inference can be comprised of per-user measurements (e.g. RSRP, SINR etc.), beam failure statistics, per-user potential source-target beam pairs and neighbouring cells' beams/interference information.

The output of the BSO optimization function can be adjusted offsets for candidate source-target beam pairs for beam mobility.

The long-term analytics function and the BSO function can be thought of as ML training and ML inference, respectively.

Operator objectives can include minimization of BFA rate for a group of users (e.g. high-mobility users).

#### 4.6.4 Benefits of O-RAN Architecture

The advantages the O-RAN architecture provides to this use case include the chance to apply and combine both, non- and near real-time analytics, machine-learning, and decision making for various sub-tasks of this use case for a cell-centric and /or user (group) centric point of view. The Non-RT RIC oversees by definition the long-term traffic-, coverage-, and interference situation etc. in a whole cluster of cells. O-RAN interfaces such as O1, A1, and E2 will support necessary data, policy, and configuration exchanges between the architectural elements. By taking contextual information and past failure statistics into account, mMIMO BF can be much more customized and thus achieve purposeful operator desires/requirements. Optimized mMIMO BF depends on the collection, processing, and the analysis of both non-real-time and real-time cell-/UE-level data, which is facilitated by the O-RAN architecture.

## 4.7 Use case 7: RAN Sharing

### 4.7.1 Background Information

One of the main challenges for network operators is to deploy a massive number of services, while providing different Quality of Service (QoS) requirements and keeping reasonable the level of investment in the network. Network sharing is envisioned as an efficient and sustainable way to accelerate the deployment of 5G, while taking advantage of a common pool of physical infrastructures and resources, made available for two or more partner operators.

Besides, regulatory requirements often force operators to provide coverage in not business attractive areas, causing profitability issues. To this end, RAN sharing is seen as a promising solution that can reduce network costs, increase network capacity and coverage, while enhancing customer satisfaction. Accordingly, the open and multivendor nature of the O-RAN architecture can accelerate the introduction and development of RAN sharing solutions, by enhancing the deployment of virtual network functions (VNF) on commodity shared hardware, while taking into account diverse QoS requirements.

Among the different RAN-sharing models that have been experimented so far, a special focus is put here on the evaluation of the compatibility of the "Geographical Split" RAN sharing model with the O-RAN architecture. In such a model, a coverage area is split between two or more operators; each operator manages the RAN in a specific area, while offering access to its RAN resources to its operator partners. Two main configurations have been used worldwide [i.6] Multi Operator Core Network (MOCN) and Multi Operator RAN (MORAN). In MOCN, both the RAN infrastructure and carriers are pooled. Even though this model enables further cost saving, especially in rural areas due to a lower number of carriers, it requires the presence of a regulatory entity that takes care of the allocation of different parts of the shared spectrum between operators. Conversely, in MORAN, each operator utilizes a separate carrier, while getting more freedom and independency on the control of the radio resources. MORAN is the most widely used sharing configuration as it can provide appropriate independency to each sharing partner operator, while maximizing the benefits of sharing in terms of CAPEX and OPEX.

Besides, the O-RAN architecture can provide new opportunities for implementing this RAN sharing model in a more efficient way, thanks to its multi-vendor interfaces and abstraction control provided by the RIC (RAN Intelligent Controller). Accordingly, O-RAN can accelerate the deployment of 3GPP-based solutions by providing more flexibility in the management of the shared resources.

### 4.7.2 Motivation

Currently, in 3GPP there are ongoing discussions for identifying which RAN functions should be included in the RAN sharing procedures in 5G [i.11] and [i.2]. Specifically, it has been analysed the feasibility to share the whole or a part of the DU or CU functional blocks while assuming the sharing of a common physical layer (PHY). In [i.4], a first agreement has been achieved stating that multiple logical CU-CP/ DU can control the same PHY radio resources but coordination between logical CU-CP needs to be ensured by an appropriate implementation (not standardized).

In this context, O-RAN can be seen as the ideally enabler of the 3GPP RAN sharing model by providing the required coordination between the shared network nodes. To this end, the RIC can enable the coordination of multiple CU-CP/ DU via the E2 interface, while opening the road for diverse RAN sharing scenarios. Moreover, O-RAN can accelerate the deployment of compliant 5G RAN sharing solutions by taking advantage of the multi-vendor nature of the F1, X2 interfaces.

### 4.7.3 Proposed Solution

The logic architecture of the proposed MORAN use-case is given in figure 4.7.3-1. It is assumed that two operators, denoted as Operator A and Operator B, share the same RAN infrastructure, while keeping the core network independent.

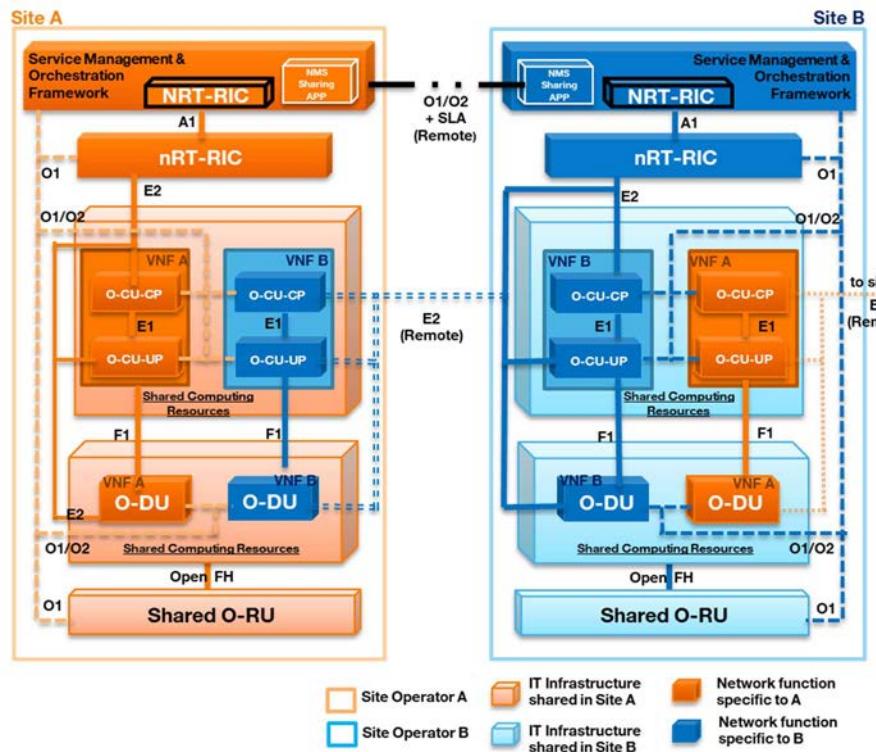
Specifically, Operator A owns the site A and shares the PHY Layer (LOW) with Operator B (Shared O-RU shown in figure 4.7.3-1). Indeed, multiple PLMN IDs are broadcasted [i.6], while each operator operates in a different carrier.

Moreover, site A hosts VNF instances of Operator B in a shared O-DU and O-CU site. Specifically, the computing resources of the site A are shared among multiple VNFs, belonging to the operator A and B respectively. Each VNF represents a logic implementation of the O-DU and O-CU functionalities and should be controlled by each partner operator in an independent manner.

While Operator A can directly control its VNFs, Operator B needs to control its VNFs in a remote manner. The challenge here is to enable Operator B to control resources in an infrastructure that is owned by another operator. Accordingly, it is assumed that Operator B can monitor and control the remoted resources via the RIC node of site B. Note that in the proposed architecture, the RIC are not shared and kept independent at the site A and B respectively.

Such a scenario presents a set of implementation challenges:

- A common interface is needed to control and coordinate the usage of shared resources.
- An orchestrator has to be able to communicate effectively with the shared nodes regardless of the manufacturer or vendor of the used hardware devices.



**Figure 4.7.3-1: MORAN Use-Case in O-RAN**

To this end, this use-case proposes to enable Operator B to remote control the O-CU and O-DU VNFs via a "remote E2 interface", giving it the freedom to implement and configure its RIC policies in an autonomous manner.

Besides, it is assumed that Operator A instantiates all the network nodes in site A, including the VNFs of Operator B, via the O2 interface, while the management and orchestration is provided by Service & Orchestration Framework. To better orchestrate the shared resources, the Service & Orchestration Framework of Site A should interact with the one of Operator B. This interaction can be enabled as follows:

- by designing a new interface between the Service & Orchestration Framework of the two sites (RAN-Sharing Orchestration Interface in figure 4.7.3-1);
- by enabling the Operator B to directly orchestrate its VNFs deployed in site A via "remote O1 & O2" interfaces.

Regarding the control plane, it is assumed that Operator A can control only the shared physical resources, while Operator B can handle only the virtual resources that belong to it.

The "remote E2" interface should provide secure access to Op. A site, while from the Operator B RIC perspective, the O-DU and O-CU VNFs hosted at site A should be controlled in a transparent manner as in non-shared scenario.

Finally, this use-case proposes an extension of the E2 interface in order to support the remote control of shared resources, while taking account security aspects related to the risk of:

- i) offering to an external actor the possibility to have access to the resources of the hosting RAN infrastructure; and
- ii) enabling an external actor to orchestrate the resources of the partner operator.

#### 4.7.4 Benefits of O-RAN Architecture

The proposed MORAN sharing architecture shown in figure 4.7.3-1 lets operators to configure shared network resources independently from configuration and operating strategies of the other sharing operators. Accordingly, this architecture enables the RIC of the Operator B to monitor the radio state of the customers served by the partner operator's site, facilitating the optimization of the radio allocation process and the remote configuration of QoS parameters. Moreover, this approach can favour the deployment of slicing scenarios facilitating the differentiation of services running on the shared RAN.

### 4.8 Use case 8: QoS Based Resource Optimization

#### 4.8.1 Background Information

QoS based resource optimization can be used when the network has been configured to provide some kind of preferential QoS for certain users. One such scenario can be related to when the network has been configured to support e2e slices. In this case, the network has functionality that ensures resource isolation between slices as well as functionality to monitor that slice Service Level Specifications (SLS) are fulfilled.

In RAN, it is the scheduler that ensures that Physical Resource Block (PRB) resources are isolated between slices in the best possible way and also that the PRB resources are used in an optimal way to best fulfill the SLS for different slices. The desired default RAN behavior for slices is configured over O1. For example, the ratio of physical resources (PRBs) reserved for a slice is configured at slice creation (instantiation) over O1. Also, QoS can be configured to guide the RAN scheduler how to (in real-time) allocate PRB resources to different users to best fulfill the SLS of a specific slice. In the NR NRM this is described by the resource partition attribute.

Instantiation of a RAN sub-slice will be prepared by rigorous planning to understand to what extent deployed RAN resources will be able to support RAN sub-slice SLS. Part of this procedure is to configure RAN functionality according to above. With this, a default behavior of RAN is obtained that will be able to fulfill slice SLSs for most situations. However, even through rigorous planning, there will be times and places where the RAN resources are not enough to fulfill SLS given the default configuration. To understand how often (and where) this happens, the performance of a RAN slice will continuously be monitored by SMO. When SMO detects a situation when RAN SLS cannot be fulfilled, Non-RT RIC can use A1 policies to improve the situation. To understand how to utilize A1 policies and how to resolve the situation, the non RT-RIC will use additional information available in SMO.

#### 4.8.2 Motivation

To motivate the use case an example with an emergency service as a slice tenant is used. For this example, it is understood (at slice instantiation) that 50 % of the PRBs in an area can be enough to support the emergency traffic under normal circumstances. Therefore, the ratio of PRBs for the emergency users is configured to 50 % as default behavior for the pre-defined group of users belonging to the emergency slice. Also, QoS is also configured in CN and RAN so that video cameras of emergency users get a minimum bitrate of 500 kbps.

Now, suppose a large fire is ongoing and emergency users are on duty. Some of the personnel capture the fire on video on site. The video streams are available to the Emergency Control Command. Because of the high traffic demand in the area from several emergency users (belonging to the same slice), the resources available for the Emergency slice is not enough to support all the traffic. In this situation, the operator has several possibilities to mitigate the situation. Depending on SLAs towards the Emergency slice compared to SLAs for other slices, the operator could reconfigure the amount of PRB reserved to Emergency slice at the expense of other slices. However, there is always a risk that Emergency video quality is not good enough irrespective if all resources are used for Emergency users. It might be that no video shows sufficient resolution due to resource limitations around the emergency site.

In this situation, the Emergency Control Command decides, based on the video content, to focus on a selected video stream to improve the resolution. The Emergency Control System gives the information about which users to up- and down-prioritized to the e2e slice assurance function (through e.g. an Edge API) of the mobile network to increase bandwidth for selected video stream(s). Given this additional information, the Non-RT RIC can influence how RAN resources are allocated to different users through a QoS target statement in an A1 policy. By good usage of the A1 policy, the Emergency Control Command can ensure that dynamically defined group of UEs provides the video resolution that is needed.

The use case can be summarized as per below:

- A fire draws a lot of emergency personnel to an area.
- Because of this RAN resources becomes congested which affects the video quality for all video feeds in the area.
- The Emergency Control Command have 5 active video feeds and selects one video feed which is of specific interest.
- The Emergency Control Command requests higher resolution of a selected feed, while demoting the other.
- With this information, the Non-RT RIC will evaluate how to ensure higher bandwidth for the feed selected by Emergency Control Command (and lower for other feeds).
- The Non-RT RIC updates the policy for the associated UEs in the associated Near-RT RIC over the A1 interface.
- Near-RT RIC enforce the modified QoS target for the associated UEs over the E2 interface to fulfill the request.
- The Emergency Control Command experiences a higher resolution of the selected video feed.

#### 4.8.3 Proposed Solution

The main functions of the O-RAN components are utilized to support an improved QoS based resource optimization. QoS based resource optimization use case deploys O-CU, O-DU, the Non-RT RIC and the Near RT RIC function modules. To achieve intelligent resource optimization, Non-RT RIC should provide policies to Near-RT RIC which are used to drive QoS based resource optimization at the RAN level. Non-RT RIC should monitor QoS related metrics from network and SMO functions. O-CU and O-DU components should provide UE performance metrics with the configured granularity to SMO via O1. In addition to performance metrics retrieved from network elements, external information sources might also be utilized to solve the problem of allocation limited RAN resources. For example, external server could provide Non-RAN data about priorities of the UEs to SMO. Finally, the E2 Nodes should execute QoS enforcement decisions received from Near-RT RIC which are expected to influence RRM behaviour.

#### 4.8.4 Benefits of O-RAN Architecture

The main features of O-RAN architecture are pointed by the proposed solution which aims to offer more advanced QoS based resource optimization.

## 4.9 Use case 9: RAN Slice SLA Assurance

### 4.9.1 Background Information

The 3GPP standards architected a sliceable 5G infrastructure which allows creation and management of customized networks to meet specific service requirements that can be demanded by future applications, services and business verticals. Such a flexible architecture needs different requirements to be specified in terms of functionality, performance and group of users which can greatly vary from one service to the other. The 5G standardization efforts have gone into defining specific slices and their Service Level Agreements (SLAs) based on application/service type [i.11]. Since network slicing is conceived to be an end-to-end feature that includes the core network, the transport network and the Radio Access Network (RAN), these requirements should be met at any slice subnet during the life-time of a network slice [i.15], especially in RAN side. Exemplary slice performance requirements are defined in terms of throughput, energy efficiency, latency and reliability at a high level in SDOs such as 3GPP [i.9] and GSMA [i.7]. These requirements are defined as a reference for SLA/contractual agreements for each slice, which individually need proper handling in NG-RAN.

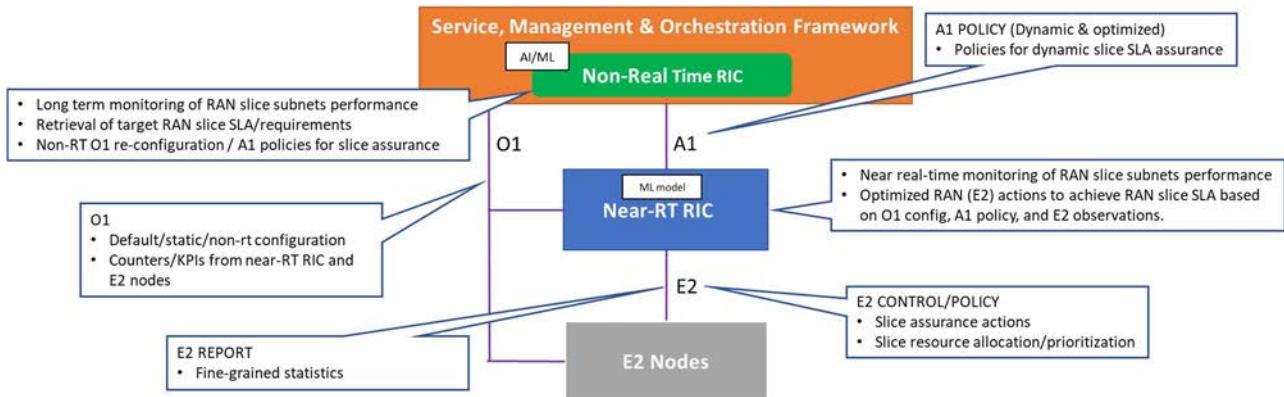
Although network slicing support is started to be defined with 3GPP Release 15, slice assurance mechanisms in RAN needs to be further addressed to achieve deployable network slicing in an open RAN environment. It is necessary to assure the SLAs by dynamically controlling slice configurations based on slice specific performance information. Existing RAN performance measurements [i.17] and information model definitions [i.16] are not enough to support RAN slice SLA assurance use cases. This use case is intended to clarify necessary mechanisms and parameters for RAN slice SLA assurance.

### 4.9.2 Motivation

In the 5G era, network slicing is a prominent feature which provides end-to-end connectivity and data processing tailored to specific business requirements. These requirements include customizable network capabilities such as the support of very high data rates, traffic densities, service availability and very low latency. According to 5G standardization efforts, the 5G system can support the needs of the business through the specification of several service needs such as data rate, traffic capacity, user density, latency, reliability, and availability. These capabilities are always provided based on a Service Level Agreement (SLA) between the mobile operator and the business customer, which brought up interest for mechanisms to ensure slice SLAs and prevent its possible violations. O-RAN's open interfaces and AI/ML based architecture will enable such challenging mechanisms to be implemented and help pave the way for operators to realize the opportunities of network slicing in an efficient manner.

### 4.9.3 Proposed Solution

As shown in figure 4.9.3-1, RAN slice SLA assurance scenario involves Non-RT RIC, Near-RT RIC, E2 Nodes and SMO interaction. The scenario starts with the retrieval of RAN specific slice SLA/requirements (possibly within SMO or from NSSMF depending on Operator deployment options). Based on slice specific performance measurements from E2 Nodes, Non-RT RIC and Near-RT RIC can fine-tune RAN behavior aligned with O-RAN architectural roles to assure RAN slice SLAs dynamically. Non-RT RIC monitors long-term trends and patterns for RAN slice subnets' performance, and employs AI/ML methods to perform corrective actions through SMO (e.g. reconfiguration via O1) or via creation of A1 policies. Non-RT RIC can also construct/train relevant AI/ML models that will be deployed at Near-RT RIC. A1 policies possibly include scope identifiers (e.g. S-NSSAI) and statements such as KPI targets. On the other hand, Near-RT RIC enables optimized RAN actions through execution of deployed AI/ML models in near-real-time by considering both O1 configuration (e.g. static RRM policies) and received A1 policies, as well as received slice specific E2 measurements.



**Figure 4.9.3-1: Slice SLA Assurance**

#### 4.9.4 Benefits of O-RAN Architecture

Current standardization efforts focus on defining business and system requirements, slice management functionalities and procedures, and possible RRM policies on shared resources without specifying how such a flexible envisioned system can be addressed within 5G RAN. Considering the dynamic nature of RAN, providing desired levels of service quality for each RAN slice is a challenging topic that requires further investigation and standardization efforts for a multi-vendor open RAN environment. O-RAN's open interfaces combined with its AI/ML based innovative architecture can enable such RAN SLA assurance mechanisms, which could potentially change the way network operators do their business and also enable new business models. For example, O-RAN architecture and interfaces can enable operators to manage spectrum resource allocation across slices more efficiently and dynamically in response to usage patterns, thereby allowing more efficient use of spectrum resources.

### 4.10 Use case 10: Multi-vendor Slices

#### 4.10.1 Background Information

This use case enables multiple slices with functions provided by multi-vendors, such as slice #1, composed of DU(s) and CU(s), provided by vendor A and slice #2, composed of DU(s) and CU(s), provided by vendor B.

#### 4.10.2 Motivation

When providing multiple slices, it is assumed that suitable vO-DU/scheduler and vO-CU treat each slice respectively. A vendor who provides vO-DU and vO-CU function can have a strength of a customized scheduler for a certain service. With accomplishment of multi-vendor circumstances, following benefits can be expected:

1) More flexible and time to market deployment

Operators can maximize options to choose suitable vO-DU/scheduler and vO-CU to offer various slices. For example, some vendors can have a strength of a scheduler for eMBB service and the other can have a strength of scheduler for URLLC service. Or, vendor A can provide vO-DU/scheduler and vO-CU suitable for URLLC earlier than vendor B, therefore operators can choose vO-DU and vO-CU functions from vendor A to meet their service requirements.

Also, when an operator wants to add a new service/slice, new functions from a new vendor can be introduced with less consideration for existing vendors if multi-vendor circumstance was realized. This can help expand vendor's business opportunities rapidly.

2) Flexible deployment when sharing RAN equipment among operators

When operators want to share RAN equipment and resources, RAN vendors and their placement of each RAN functions can be different. If multi-vendor circumstance was introduced, then it can relax restrictions among operators to share RAN equipment and resources. This can help expanded opportunities for reaching agreements of RAN sharing among operators. With expansion of RAN sharing, operators CAPEX and OPEX can be optimized, helping with additional investment opportunities.

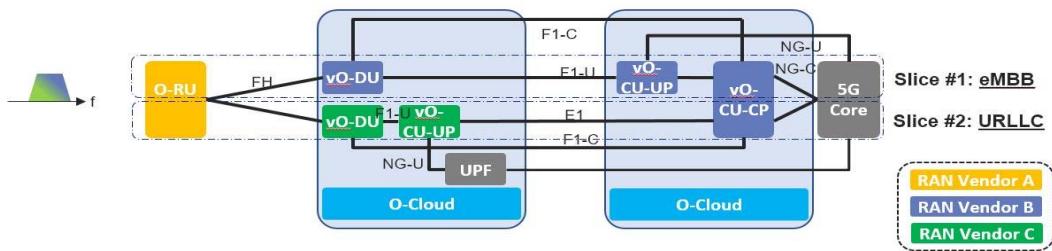
3) Reducing supply chain risk

If an existing vendor providing a certain pair of vO-DU and vO-CU functions withdraws from the market due to business reasons, operators can deploy new vO-DU and vO-CU functions alternatively from other vendors under this multi-vendor circumstance. This can reduce a risk for operators' business continuity.

### 4.10.3 Proposed Solution

#### 4.10.3.1 Multi-vendor Slices

An architecture for multi-vendor slices use case is given in figure 4.10.3.1-1. There are multiple slices with which have vO-DU and vO-CU functions associating respectively. As shown in figure 4.10.3.1-1, slice-1 is composed of vO-DU(s) and vO-CU(s) provided by vendor B, and slice-2 is composed of vO-DU(s) and vO-CU-UP(s) provided by vendor C. Each vO-DU/scheduler and vO-CU functions treat one slice as an example. O-RU provided by vendor A is shared between two vO-DU(s) supplied by two different vendors, vendor B and C. The case of vO-DU and vO-CU from different vendors in a slice is for further study.



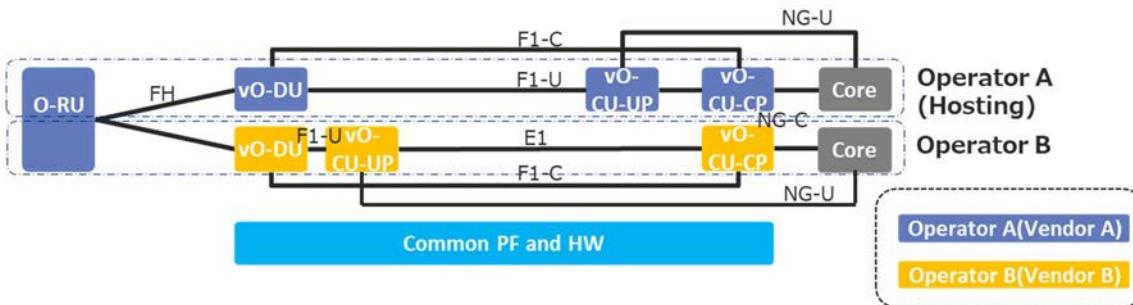
**Figure 4.10.3.1-1: Multi-vendor Slices use case**

#### 4.10.3.2 RAN Sharing

An additional application of multi-vendor slices use case is RAN sharing where, operator A has a pair of vO-DU and vO-CU from vendor A, and operator B has a different pair of vO-DU and vO-CU from vendor B, and O-RU is shared among these two operators (see figure 4.10.3.2-1).

As mentioned in clause 4.10.2, through RAN sharing, CAPEX and OPEX are expected to be reduced. Savings achieved by RAN sharing can then be invested again for additional expansions of RAN sharing area.

This use case considers single slice in each operator and multiple slices in each operator is for further study.

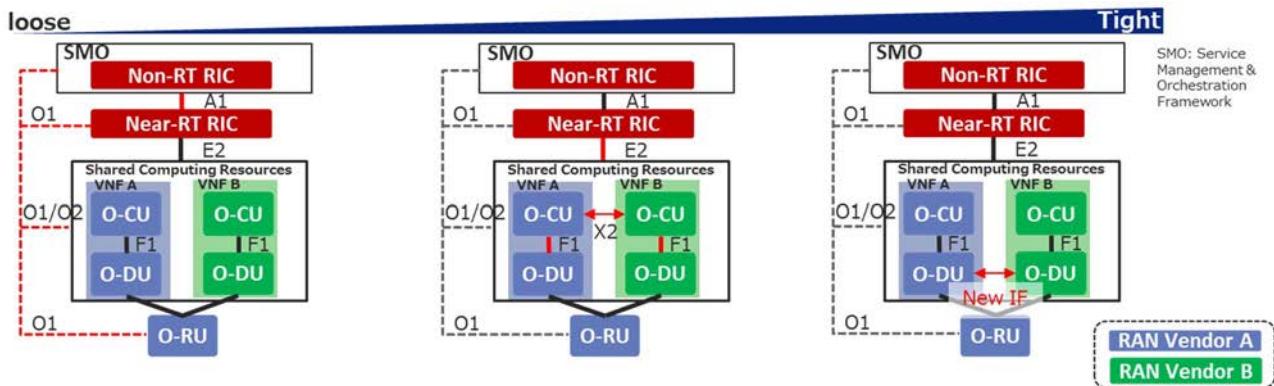


**Figure 4.10.3.2-1: RAN Sharing use case**

To realize multi-vendor slices, some coordination between vO-DU/vO-CUs will be required since radio resource should be assigned properly and without any conflicts. Depending on different service goals and the potential impact on O-RAN architecture, a required coordination scheme needs to be determined. The possible cases are:

- 1) Loose coordination through O1//A1 interface (Case 1 shown in figure 4.10.3.2-2).
- 2) Moderate coordination through E2/X2/F1 interface (Case 2 shown in figure 4.10.3.2-2).

- 3) Tight coordination through a new interface between vO-DUs (Case 3 shown in figure 4.10.3.2-2).



**Figure 4.10.3.2-2: Multi-vendor Slices Coordination Scheme Options**

In case 1, a resource allocation between slices or vO-DU/vO-CUs is provisioned through O1/A1 interface and each pair of vO-DU and vO-CU will allocate radio resources to each customer within radio resources allocated by Near-RT RIC and/or Non-RT RIC.

In case 2, a resource allocation can be negotiated between slices or vO-DU/vO-CUs through E2, X2 and F1 after provisioned through O1/A1 interface.

If a more adaptive radio resource allocation is needed (case 3), a more frequent negotiation would be required. This can potentially be achieved via an interface or API extension between vO-DU(s) which is not defined in the present document.

#### 4.10.4 Benefits of O-RAN Architecture

The proposed multi-vendor slices approach will not only enable operators to utilize different vendor's strengths for particular service/slice type but also provide opportunities for additional RAN Sharing use cases. The coordination needed between different vendors CU(s)/DU(s) can be realized through O-RAN architecture where Near-RT RIC and Non-RT RIC can coordinate CU(s)/DU(s) to efficiently allocate and use resources across these multi-vendor slices.

### 4.11 Use case 11: Dynamic Spectrum Sharing (DSS)

#### 4.11.1 Background Information

As transition from 4G to 5G is being made, the spectral resources used for 5G deployment is a key consideration and this situation varies from one operator to another. Though, new C-band resources between 3-6 GHz and mmWave bands have been acquired by operators, these bands suffer from great propagation and penetration loss, limiting their coverage to users in proximity of the cell site, this is compelling particularly on the UL where the UE device is power constrained. This mandates the requirement of 5G deployment on lower bands (i.e. below 2GHz), which are widely used in existing 4G LTE deployments. Operating on lower bands along with non-standalone mode of 5G deployment helps to cover large geography, enables seamless mobility between 4G and 5G while being sensitive to overall cost of deployment. In addition, DSS offers the advantage of dynamically sharing the available spectrum adapting to the changing work loads of 4G and 5G network.

#### 4.11.2 Motivation

DSS is compelling considering the need for operators to dynamically share already deployed spectral resources between LTE and NR devices without degrading the QoE of the current 4G subscribers while offering the same level of coverage to support NR devices, with the consideration that they will be rolled out in an incremental manner. The objective of this use case is to propose DSS in the context of the O-RAN architecture, specifically to realize it as an application in the RIC framework. This would particularly benefit vRAN implementations when the 4G/5G CU/DU are from different vendors and one could leverage RAN data over ORAN's framework for traffic prediction, resource management and control functions. Towards this, the intelligent control functions, which can be realized as a DSS application to augment the L2/L1 control functions defined as part of LTE-NR coexistence in Rel-15/16, are identified.

### 4.11.3 Proposed Solution

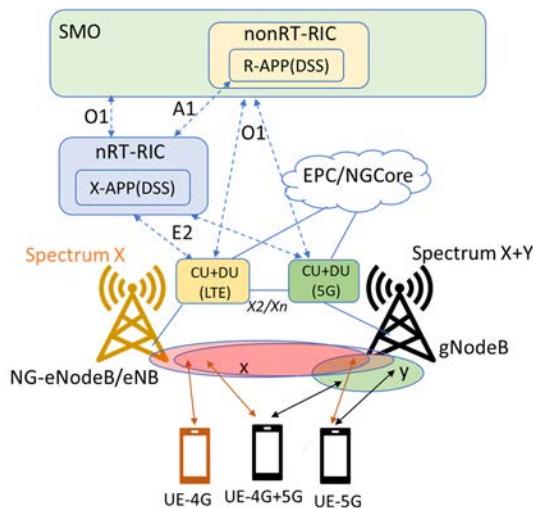
The architectural context for the proposed solution is shown in figure 4.11.3-1, where DSS enables 4G and 5G UEs to operate over the same spectrum identified as X (typically low band), while 5G itself could operate on new bands Y (typically high band) not used by current 4G deployment. In a typical setting, Y would offer higher capacity, low latency and smaller coverage, while X would be used to offer reasonable capacity along with larger coverage. 3GPP specifications offers DSS support over X2/Xn interface to enable dynamic sharing of spectrum resource, in addition to L2/L1 adaptation for 5G-NR to co-exist with LTE subscriber.

When DSS is enabled in the SA mode, 5G UE would be capable of operating on lower LTE bands (below 2 GHz), C and mmWave bands and requires connectivity only to the gNBs. The sharing of the LTE bands between LTE and 5G data channels are achieved by both 4G scheduler and 5G scheduler, assisted by the coordination function, complimented by the RIC, between them.

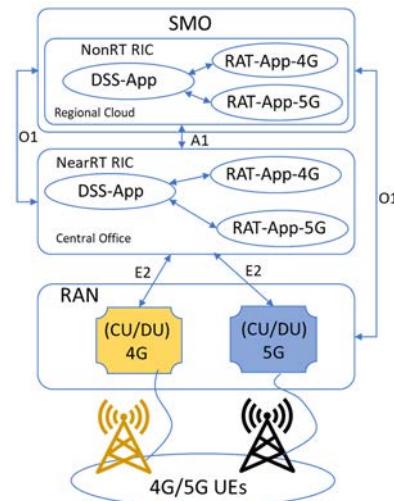
For 5G NSA mode, the 5G UE is required to have dual connectivity capability and be able to connect to eNBs on LTE bands for control plane requirements and user plane connectivity towards the LTE and/or 5G depending on deployment requirements. In the scenario where gNB only operates on 5G C or mmWave bands, the sharing of the LTE frequency band between 4G and 5G UEs can be solely fulfilled by eNB MAC scheduler, as the two devices are indistinguishable from L2/L1 perspective. While, if the gNB is required to operate on lower LTE bands as well, then spectral sharing needs to be coordinated between the LTE and 5G schedulers.

The use case proposes to conduct DSS related policy, configuration, resource management and control functions using the Non-RT and Near-RT functions over open interfaces proposed by ORAN.

An abstracted view of how DSS application can be realized using the Non-RT and Near-RT RIC components is shown in figure 4.11.3-2. The DSS over RIC can be realized as multiple applications considering its multiple optimization and operational objectives. One possible logical breakdown is as a resource management application (DSS-App) managing the shared spectrum resource adapting to dynamic 4G and 5G specific workload requirements in various local contexts, and another application (RAT-App) to configure, control and monitor DSS rules in the CU/DU corresponding to the LTE and 5G cells. The DSS-App engineers to translate the global DSS objectives such as workload requirements for a region and time-of-day to spectrum sharing policies such as max/min bandwidth threshold at a local level (e.g. central office). The RAT-App then translates the DSS-App's resource policies to RAT specific configuration and control actions communicating with the respective CU/DU instances.



**Figure 4.11.3-1: RIC Based DSS Architecture**



**Figure 4.11.3-2: RIC Based DSS Realization**

The main goal of the Non-RT DSS-App is to provide long-term policy or intent as a scheduling guidance to 4G and 5G scheduler considering business, user, spatial and temporal workload factors and the main functionality of Non-RT RAT-App is to translate the global DSS policies from Non-RT DSS-App to RAT specific policies to the RAT-App in the Near-RT RIC over A1.

The main functionalities of the Near-RT DSS-App include policy translation between Non-RT DSS-App to RAT specific configuration to the Near-RT RAT-App. Furthermore, it is actively involved in closed-loop decision using the KPIs from the RAN adapting to the needs of the 4G and 5G cells. The main functionality of Near-RT RAT-App is to perform RAT specific configuration, control and data subscription over E2 interface with RAN (CU/DU components).

#### 4.11.4 Benefits of O-RAN Architecture

Using RIC based implementation of DSS benefits from using the non-real-time and near-real-time data from the RAN to influence the 4G/5G MAC scheduler's near and long-term resource management policies and scheduler configuration. This is in addition to the value driven from control over open interfaces and using third party DSS applications. This is in contrast to synchronizing the 4G and 5G schedulers using closed resource management functions. Some scenarios, for which the advantage of a RIC based DSS is foreseen, include:

- 1) DSS over RIC allows policy driven resource management between the 4G and 5G schedulers, aided by predictive intelligence. Furthermore, DSS requires synchronizing the MAC schedulers to avoid scheduling interference. However, the granularity of synchronization depends on the nature of workload. While workloads that are bursty in nature can require close to per-TTI synchronization, predictable workload could be handled without trading-off significant efficiency with coarser TTI synchronization. The latter would be workloads DSS over RIC would be suitable for, which would also be in line with Near-RT RIC's operational requirement of 10ms-1s control loop latency.
- 2) DSS over RIC can also improve resource management efficiency over multiple 4G/5G spectrum sharing cells. In this scenario, the RIC could use the global data spanning distributed cell sites, aided with predictive models to share the spectral resources dynamically. For example, data on predictive LTE or 5G user mobility can be used to predict the bandwidth requirement in adjacent 4G and 5G cells. This bandwidth decision can be then coordinated in advance across multiple distributed 4G/5G cells.
- 3) DSS over RIC can also be useful when the 4G/5G cells can only overlap over the cell edge, in which case UE related information such as RSRP/RSRQ, PDCP throughput can be used by the Near-RT RIC to identify the set of overlapping cells along with the projected workload. This information is then used to slice the shared resource among the cell edge users to avoid interference by the respective schedulers while optimizing resources for the cell center users.

### 4.12 Use case 12: NSSI Resource Allocation Optimization

#### 4.12.1 Background Information

5G networks are becoming increasingly complex with the densification of millimeter wave small cells, and various new services, such as enhanced Mobile Broadband (eMBB), Ultra Reliable Low Latency Communications (URLLC), and massive Machine Type Communications (mMTC) that are characterized by high speed high data volume, low speed ultra-low latency, and infrequent transmitting low data volume from huge number of emerging smart devices, respectively. It is a challenging task for 5G networks to allocate resources dynamically and efficiently among multiple network nodes to support various services.

#### 4.12.2 Motivation

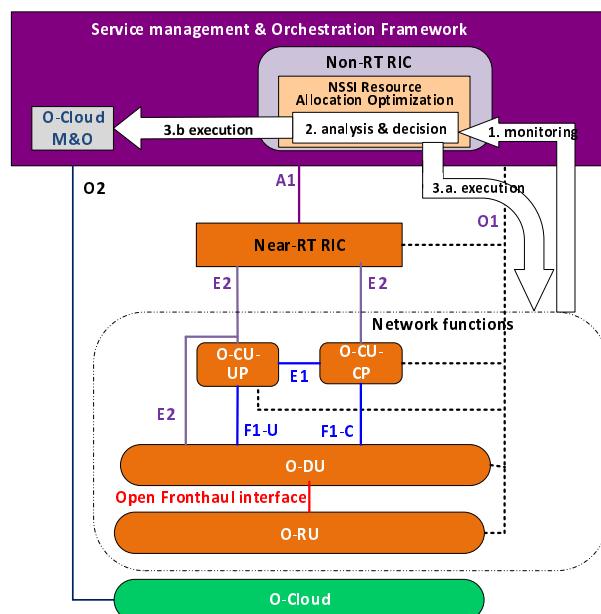
eMBB, URLLC, and mMTC services in 5G are typically realized as Network Slice Instance(s) (NSI(s)) and the resources allocated to Network Slice Subnet Instance (NSSI) to support the O-RAN nodes can be optimized according to the service requirements. As the new 5G services have different characteristics, the network traffic tends to be sporadic, where there can be different usage pattern in terms of time, location, UE distribution, and types of applications. For example, most IoT sensor applications can run during off-peak hours or weekends. Special events, such as sport games, concerts, can cause traffic demand to shoot up at certain times and locations. Therefore, NSSI resource allocation optimization function trains the AI/ML model, based on the huge volume of performance data collected over days, weeks, months from O-RAN nodes. It then uses the AI/ML model to predict the traffic demand patterns of 5G networks in different times and locations for each network slice subnet, and automatically re-allocates the network resources ahead of the network issues surfaced.

The resource quota policies associated with RAN NFs (E2 Nodes) included in the respective NSSIs enable 5G network providers to optimize or prioritize the utilization of the RAN resources across slices and supports the flexibility to share resources optimally across critical service slices during resource surplus or scarcity. For example, an NSSI allocated for premium service can receive a major share of the resources compared to a slice allocated for a standard/best-effort service. Another such example is the scenario of additional resource allocation for emergency services. An important consideration here is that the NSSI resource quota policies focuses on maximization of resource utilization across the NSSIs. The resource quota policies can be used as a constraint for resource allocation that defines the range of resources that can be allocated per slice. One use case for applying such a constraint is the analysis and decision based on history of resource allocation failure that can be reflected in the RAN Node measurements. Here resource quota policy can be provisioned to control the minimum, maximum and dedicated resources that need to be allocated based on the historical pattern.

#### 4.12.3 Proposed Solution

The NSSI resource allocation optimization on the Non-RT RIC is shown in figure 4.12.3-1, and can consist of the following steps:

- 1) Monitoring: monitor the radio network(s) by collecting data via the O1 interface, including performance measurements (that are measured per S-NSSAI (3GPP TS 28.552 [i.17])) such as DL PRB used for data traffic, UL PRB used for data traffic, Average DL UE throughput in gNB, Average UL UE throughput in gNB, Number of PDU Sessions requested to setup, Number of PDU Sessions successfully setup, Distribution of UL UE throughput in gNB, etc.
- 2) Analysis & Decision:
  - 2a) Utilize AI/ML models to analyze the measurements and predict the future traffic demand, including the RRMPolicyRatio IOC limits, for each NSSI for a given time and location.
  - 2b) Determine the actions needed to add or reduce the resources (e.g. capacity, VNF resources, slice subnet attributes (3GPP TS 28.541 [i.16]), etc.) for the RAN NFs (E2 Nodes) included in the respective NSSI at the given time, and location.
- 3) Execution: execute the following actions to reallocate the NSSI resources:
  - 3a) Re-configure the NSSI attributes, including RRMPolicyRatio IOCs (3GPP TS 28.541 [i.16]) via the OAM Functions in SMO which uses O1 interface to configure the E2 Nodes.
  - 3b) Update the cloud resources via the O2 interface.



**Figure 4.12.3-1: The realization of NSSI resource allocation optimization over Non-Real Time RIC**

#### 4.12.4 Benefits of O-RAN Architecture

This use case is based on the retrieval and analysis of long term performance data captured from RAN nodes and the generation of AI/ML models for re-allocation of network resources based on the prediction of traffic demand patterns of 5G networks, which is made possible by O-RAN's AI/ML based architecture.

### 4.13 Use case 13: Local Indoor Positioning in RAN

#### 4.13.1 Background Information

Cellular network based positioning is an important technology for 5G vertical industries, individuals and operators, especially in local indoor scenarios. For instance, Mall can provide value-added services, such as local indoor navigation and shop recommendation by leverage real-time indoor positioning. Industrial manufacturing field should send real-time safe warning to remind operators keep away from the dangerous area which would also need real-time indoor positioning.

NR positioning is introduced by 3GPP Rel.16. The location management function (LMF) resides in core network acts as a location server. LTE positioning protocol (LPP) is reused for the UE measurements and NR positioning Protocol "a" (NRPPa) based on LPPa is used for the gNB measurements [i.18]. The NR Positioning Protocol A (NRPPa) PDUs carry E-CID, OTDOA are routed between eNB/gNB and the LMF via AMF. The AMF routes the NRPPa PDUs transparently based on a Routing ID corresponding to the involved LMF over NG-C interface without knowledge of the involved NRPPa transaction. This long route messages between eNB/gNB and centralized LMF can suffer network jitters and leads to un-real-time UE location results. For some local indoor scenario, e.g. mall and industrial manufacturing, it would be better to directly deploy the positioning function in the RAN side and expose UE location to local applications.

#### 4.13.2 Motivation

For local indoor scenarios, the positioning function inside RAN is envisioned to be a promising solution. It not only reduces the latency of positioning but also can reuse the edge cloud infrastructure in RAN. In the context of O-RAN architecture, the positioning function can be deployed as a positioning xApp in the Near-RT RIC. The positioning xApp computes the UE location and optional velocity based on the positioning measurement obtained via the E2 interface.

One example is that distributed indoor small base station in the operator's domain can provide UE positioning service by adding Near-RT RIC with positioning xApp. The positioning related information report from small base station to the positioning xApp inside Near-RT RIC can be measured based on the uplink SRS information. By leveraging multi-point positioning, field strength positioning and other positioning algorithms, positioning xApp inside Near-RT RIC can conduct a real-time position of UE.

#### 4.13.3 Proposed Solution

Cellular network based positioning procedure mainly includes two main steps: 1) positioning measurements reports; and 2) position computation and optional velocity estimation based on the measurements. The local indoor positioning in RAN use case mainly involves the Near-RT RIC, E2 nodes and potentially Non-RT RIC in the O-RAN architecture. To enable the local indoor positioning in RAN, Near-RT RIC should be able to discover the location measurement capability of the E2 nodes and subscribe the positioning measurements based on the chosen positioning algorithm and positioning QoS requirements.

Non-RT RIC can also be leveraged to provide the AI/ML model if machine learning based algorithm is selected for the positioning. In this case, Non-RT RIC trains the AI/ML position model based on the historical positioning measurements (e.g. RSSI) and the labeled user location (e.g. by manual or by minimal drive test (MDT)). Then, the trained positioning AI/ML model can be deployed to Near-RT RIC for real-time positioning inference.

The E2 nodes are expected to provide positioning measurements to Near-RT RIC as required. The measurements report can be periodical or event driven based on Near-RT RIC subscription. Near-RT RIC can also adjust the corresponding positioning measurement configurations in E2 nodes to assure the measure requirement. What kind of positioning measurements need to be reported from E2 Node depend on the subscription request from positioning xApp in Near-RT RIC. For instance, the positioning measurements can include E-CID, OTDOA, UTDOA, TOA, RSSI, RSRQ and RRU antenna ID if it is the indoor system. The report granularity is expected in the order of 10-100 ms.

The positioning xApp in Near-RT RIC can pass the positioning results to the SMO for further exposure. Furthermore, Near-RT RIC can also expose the positioning results to edge application nearby in a secure manner (via. API gateway with firewall).

#### 4.13.4 Benefits of O-RAN Architecture

By enabling the local indoor positioning xApp in Near-RT RIC, it helps to reduce the latency of positioning in local indoor scenarios. Moreover, Near-RT RIC covers multiple E2 nodes, it can track UE location in case of mobility within certain areas.

### 4.14 Use case 14: Massive MIMO SU/MU-MIMO Grouping Optimization

#### 4.14.1 Background Information

The Massive MIMO is one of the key technologies for 4G and 5G. Due to the multi-antenna transmission and reception, this technology can inherently provide diversity and improve capacity by targeting high gain antenna beams towards one or multiple subscribers, thus improving the receive power levels and spatially filtering the interference from neighboring subscribers and transmission points. The case that only one subscriber is served by a given physical resource block of a massive MIMO cell in a transmission interval is called single-user MIMO (SU-MIMO). While the case that multiple subscribers are served by a given physical resource block of a massive MIMO cell in a transmission interval is called multi-user MIMO (MU-MIMO).

In the commercial deployment, the massive MIMO eNB/gNB support both SU-MIMO and MU-MIMO transmission. The SU-MIMO and MU-MIMO are targeting for different scenarios, and can achieve different spectral efficiency. Some subscribers can be stationary, some subscribers can be pedestrian, and some subscribers can be moving at a high speed. The performance of MU-MIMO is very susceptible to subscriber's moving speed, while the performance of SU-MIMO is relatively robust to subscriber's moving speed. Additionally, some subscribers can have VoLTE or VoNR traffic, some subscribers can have download/FTP traffic, and some subscribers can have video streaming traffic. The performance of MU-MIMO considerably reduces in the low traffic volume scenario, while the performance of SU-MIMO is relatively robust to low traffic scenarios.

The spectral efficiency of SU-MIMO and MU-MIMO differ from each other. The spectral efficiency of MU-MIMO can be several times of spectral efficiency of SU-MIMO. Furthermore, SU-MIMO and MU-MIMO have different requirement for C-DRX, SRS, and other RRC configurations. These RRC configurations can be optimized based on scenario information.

#### 4.14.2 Motivation

The massive SU/MU-MIMO grouping optimization use case aims at proactively and continuously improving cell and/or user-centric transmission efficiency and reliability in a massive MIMO deployment area by adapting appropriate transmission methods (e.g. SU-MIMO, MU-MIMO) for each user. The massiveness of available measurement input data, the complexity, pro-activeness as well as near-real time requirement suggest the application of machine learning techniques of input data analytics as well as use case decision generation.

The objective of this use case is to allow the operator and vendor to optimize SU/MU-MIMO transmission by means of policies, configuration, or machine learning techniques, according to objectives defined by the operator.

#### 4.14.3 Proposed Solution

##### 4.14.3.1 Solution 1

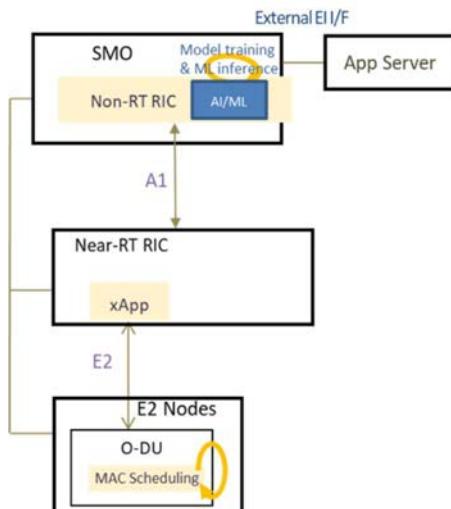
Non-RT RIC Inference solution is shown in figure 4.14.3.1-1. In this solution, Non-RT RIC is for SU/MU-MIMO grouping decision.

The SMO should collect the necessary configurations, performance indicators, measurement reports data from RAN nodes triggered by Non-RT RIC if required, and Non-RT RIC should retrieve user enrichment information (e.g. GPS Information, traffic information) from the application server. When the optimization objective fails, it triggers the AI/ML model re-training, data analytics and optimization in Non-RT RIC.

The Non-RT RIC should enable to retrieve necessary configurations, performance indicators, measurement reports and other information (e.g. user GPS Information, traffic information) for the purpose of constructing/training relevant AI/ML models. The Non-RT RIC should use the trained AI/ML model to decide the UE list for SU-MIMO group and MU-MIMO group by inferring the mobility, traffic model of each user. Additionally, Non-RT RIC should also decide RRC configuration for SU-MIMO group and MU-MIMO group, such as SRS and C-DRX configuration etc.

The Near-RT RIC should be able to retrieve SU/MU-MIMO grouping, and related RRC configurations from non-RT RIC (if it configures over A1). Moreover, it can send the configurations to E2 nodes by policy.

The RAN Nodes should send proper RRC configuration accordingly for UE in both SU-MIMO and MU-MIMO groups, do SU-MIMO scheduling for UE in SU-MIMO group, and do SU-MIMO or MU-MIMO scheduling for UE in MU-MIMO group dynamically. The RAN nodes should collect and report the performance measurement to SMO related to SU-MIMO and MU-MIMO spectral efficiency. For example, average layer, rank, and throughput for SU-MIMO and MU-MIMO.



**Figure 4.14.3.1-1: Solution 1: Non-RT RIC centered SU/MU-MIMO grouping**

#### 4.14.3.2 Solution 2

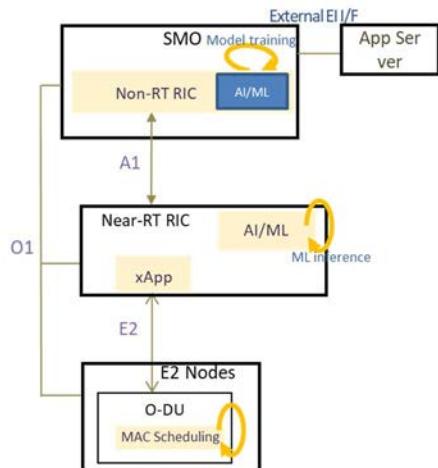
Near-RT RIC Inference solution is shown in figure 4.14.3.2-1. This solution is Near-RT RIC centered SU/MU-MIMO grouping decision. Another alternative is that the Near-RT RIC just provides the mobility and the traffic model prediction result over E2 interface. The E2 node makes final decision on SU/MU-MIMO grouping.

The SMO should collect the necessary configurations, performance indicators, measurement reports data from RAN nodes triggered by Non-RT RIC if required, and Non-RT RIC should retrieve user Enrichment Information (e.g. GPS Information, traffic information) from the application server when the optimization objective fails, and it triggers the AI/ML model re-training, data analytics and optimization in Non-RT RIC.

The Non-RT RIC should enable the operators to retrieve necessary configurations, performance indicators, measurement reports and other information (e.g. user GPS Information, traffic information) for the purpose of constructing/training relevant AI/ML models. The Non-RT RIC should sends Enrichment Information to Near-RT RIC by A1-EI I/F. The Near-RT RIC supports deployment and execution of AI/ML model from Non-RT RIC. The Near-RT RIC does ML inference to decide the UE list for SU-MIMO group and MU-MIMO group by inferring the mobility, the traffic model of each user. Additionally, Near-RT RIC should decide RRC configuration for SU-MIMO group and MU-MIMO group, such as SRS and C-DRX configuration, etc.

The Near-RT RIC should be able to send the configurations to E2 nodes by policy.

E2 Nodes should send proper RRC configuration accordingly for UE in both SU-MIMO and MU-MIMO groups, do SU-MIMO scheduling for UE in SU-MIMO group, and do SU-MIMO or MU-MIMO scheduling for UE in MU-MIMO group dynamically. E2 nodes should collect and report performance measurement to SMO related to SU-MIMO and MU-MIMO spectral efficiency. For example, average layer, rank, and throughput for SU-MIMO and MU-MIMO. An E2 node can also decide on the SU/MU-MIMO grouping; the E2 nodes should retrieve the mobility and the traffic model prediction result over the E2. E2 nodes should support the advanced MAC scheduling algorithms that decide to do SU-MIMO or MU-MIMO transmission for each user considering the user mobility and traffic model prediction.



**Figure 4.14.3.2-1: Solution 2: Near-RT RIC centered SU/MU-MIMO grouping**

#### 4.14.4 Benefits of O-RAN Architecture

The advantages the O-RAN architecture provides to massive MIMO SU/MU-MIMO grouping optimization use case include the ability to apply and combine both, non- and near real-time analytics, machine-learning, and decision making for various sub-tasks of this use case for a cell-centric and /or user (group) centric point of view. The Non-RT RIC supports long-term AI/ML training based on enrichment information. O-RAN interfaces such as O1, A1, and E2 has support for necessary data, policy, and configuration exchanges between the architectural elements. These advantages motivate massive MIMO SU/MU-MIMO grouping optimization to run on O-RAN based RAN implementations.

### 4.15 Use case 15: O-RAN Signalling Storm Protection

#### 4.15.1 Background Information

Society is increasingly dependent on network connectivity at any time and in any place and increasing diversity of device types ranging from complex devices such as smart phone to very simple and low-cost IoT devices are connecting to the network. The sheer number of connected devices, as well as the wide range of device types, makes the mobility network subject to accidental or intentional attacks that can disrupt the regular usage of the network. Given that life-critical applications are moving to wireless networks, such network disruptions are not only an inconvenience but can have impact on life and health of individuals. The O-RAN architecture offers an opportunity to address such security challenges in customizable and creative ways by utilizing the near-real time RIC xApps and non-real time RIC rApps.

## 4.15.2 Motivation

The main defense mechanism against attacks coming from the devices toward the network is based on configuration of the devices themselves and trust that the devices will indeed comply to restrictions defined by mobility standards. One such defense mechanism is the back-off timer that restricts the number of repeated device registrations, thus preventing devices from overloading the network with attaches. If this trust is breached there are no other options for defending the network rather than rejecting (denying service) randomly to both benign and malicious devices, a state which is equivalent to DDoS. Unfortunately, even today the network has few hundreds of device types that accidentally breach this trust and allow devices to aggressively attach to the network in a rate of few thousand times per hour (the maximum allowed number by standard is less than 20 attaches per hour). An attacker that finds a way to manipulate a large set of these vulnerable devices remotely can cause an attach storm that would lead to a long outage of large parts of the network. Furthermore, this attacker can continue this attack over many hours, each time picking few thousand of devices from a large pool of millions of devices available; the network carrier will not be able to stop this attack without intelligent and fine-grained controls to act against a certain patterns of behaviour. The good news is that detecting these aggressive devices is possible as their behaviour is very different from the other devices in the network. What the network really needs is to apply dynamic restriction over these devices to prevent them from overloading the control plane of the network. This restriction should be smart enough to still allow benign devices to register to the network without interruption. Having smart security control at the RAN can stop such attack and without overloading deeper parts of the network in the core.

## 4.15.3 Proposed Solution

In order to protect the network from such UE originated signaling storms, an xApp can be built with two main functionalities: a DDoS detection capability and a DDoS mitigation capability. The DDoS detection capability has two parts: the near real time detection, which takes place in a RIC xApp and a non-real time detection, which takes place at the SMO and relies on enrichment data originated in external system (e.g. 5G Core or OAM system) (see figure 4.15.3-1). The reason for this functionality split is that some detection logic relies on information that is available only at the core (such as IMSI, IMEI, device types, PLMN etc.), while for other detection logic, the information available at the RAN will suffice. It is assumed that the Near Real-Time RIC DDoS Detection xApp provides a faster response with coarse grained aggressive device detection while the 5G Core detection results computed by the SMO rApp provide slower response with finer grained aggressive device detection. It is noted that for some attack scenarios the xApp can perform the detection without the help of external information coming from the Non Real-Time RIC. The DDoS Mitigation xApp can implement an E2 INSERT-CONTROL control loop where the xApp can decide for each attach request if it should be accepted or rejected, or it can update an appropriate E2 POLICY when a UE is determined to be suspect. An exemplary E2 POLICY in this use case can allow suspected UEs to be blocked completely or throttled at a given rate at the E2 Node.

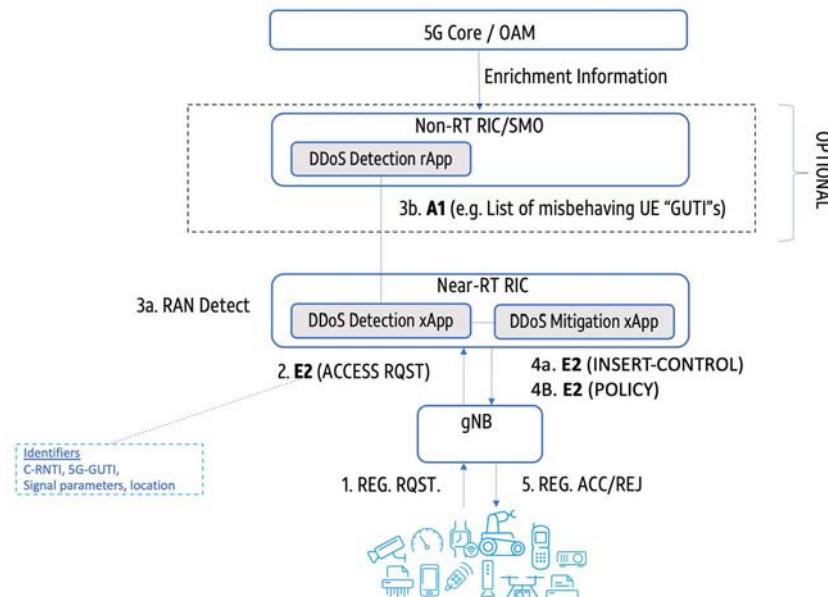


Figure 4.15.3-1: RIC DDoS Capability for 5G Stand Alone Setup

The detection algorithms should detect abnormal activity of high volume of control plane messages originating from a set of devices. The detection algorithm can change based on the network environment (metro, rural, enterprise, etc.) as well as dynamic usage trends and available devices. The output of the detection algorithm can also be of different levels of quality, depending on the attack type, in some cases the detection algorithm will produce list of specific devices while in other cases, it would only point to problematic geographic locations, specific gNBs or cell towers where the attack seems to be originated. Generally speaking, a detection algorithm should track registration behavior of individual devices and alert in the case that a threshold is exceeded for a given time threshold (e.g. one minute). The thresholds per tracking unit (i.e. cell site, gNB, AMF group etc.) should be calculated dynamically based on normal behaviour of the tracking unit. The learning of the normal behavior can be achieved using ML algorithm over time.

An important concern regarding the detection report quality is the identifiers lifetime. Identifiers of network elements such as a gNB or cell tower are considered persistent over long periods of time. When considering device or subscriber related identifiers their life span depends on where they are collected. Specifically, equipment identifier (such as IMEI) or subscriber identifier (such as IMSI) are only visible at the network core and are not available at the RAN. Some RAN identifiers change upon every network attach and are allocated randomly (for example C-RNTI), while others persist longer such as the 5G TMSI. The available identifiers in the detection report depends on the nature of the attack: location, number of devices used, exploited vulnerability type and more. For example, if an aggressive device uses an RRC Connection Re-Establishment procedure, the C-RNTI remains the same over the numerous connection attempts, while in the case of RRC Connection Establishment procedure there is a fresh randomly selected C-RNTI for each attempt. There are detection techniques to identify an aggressive UE that does not rely solely on its identifiers, rather on signal parameters such as signal quality or timing advance, which can in turn, help build a unique fingerprint of the aggressive device. An important disclaimer is that not all attack scenarios can be prevented. Specifically, in the case where the modem or chipset itself are compromised and communication parameters can be faked by the attacker, the detection methods mentioned above could be avoided.

The mitigation functionality should support a set of actions, depending on the policy applied. These actions can be applied to a single UE, or a set of UEs that apply to a certain criterion. The criteria can be a location, a cell tower or a gNB. A UE set can be defined in a custom manner to include a list of UEs, where the criteria is based on considerations external to RIC, such as UE reputation, vendor, or vulnerability type. The actions applied to a single UE or set of UE should include (but not be limited to) rejecting attach requests, throttling attach requests (attach rate should not exceed a certain threshold), handover a set of devices to a different radio or applying a different slice to set.

#### 4.15.4 Benefits of O-RAN Architecture

The O-RAN architecture allows applying protection to the network at the edge, thus minimizing the effect of these signaling storm DDoS attacks on network resources.

The combination of near real time logic at the Near Real-Time RIC for fast detection, with slower scale analysis and input from the Non Real-Time RIC and SMO provides a good mixture between quick reaction and advanced detection schemes. Lighter detection processes can be applied as xApps while more advanced heavy processing ML analytics can run externally and send input to RIC over the A1 interface.

### 4.16 Use case 16: Congestion Prediction & Management

#### 4.16.1 Background Information

This use case provides a proactive approach to congestion handling in the base station by analyzing the radio resource utilization and taking timely corrective action so as to mitigate any potential congestion in the system.

## 4.16.2 Motivation

Large-scale commercial cellular networks have many problems - like cell congestions leading to RLF (Radio Link Failure etc.), Handover Failure, poor data rates etc. Network congestion is a crucial problem for the telecom operators as it affects the Quality of Service (QoS) of the users directly. An operator has many solutions like Offload (across carriers, Wi-Fi, etc.) or Antenna techniques (Cell Split, Higher order MIMO, etc.) to handle congestion. However, the congestion patterns in the network are not fully understood and mitigation is done post facto at the expense of the prolonged user experience degradation. The congestion mitigation is critical for operators to retain their subscribers for operator and individual user experience. With 5G, the congestion needs to be handled to as to best utilize the radio resources. Today operators do not have a well-defined mechanism to predict congestion. The main objective of this use case is to use the embedded intelligence of O-RAN to predict the congestion ahead of time, so that operators can keep the cell congestion mitigation solution in place before the congestion is predicted to happen.

## 4.16.3 Proposed Solution

CPM (Congestion Prediction & Management) architecture is proposed to detect and mitigate congestion proactively. In the CPM architecture, E2 node statistics [counters] are collected by the data collector of SMO. This is done over the O1 interface. The pre-processing of data is also done in the same place. Pre-processing includes adding VNF/cell names/numbers and ids to the data and converting counters into KPI using KPI logic. After preprocessing data is shared with Non-RT RIC deployed in the SMO using a data sharing entity. Non-RT RIC will invoke the corresponding training model/application in an AI server inside SMO (it can be placed outside SMO also). The data cleaning and training will happen and the predicted KPIs will be sent back to CPM rApp in Non-RT RIC. Machine learning models, can be used to learn and predict the future traffic for the next hour/day/month. The prediction window can be configured by operators as per the available data and its periodicity. CPM rApp in Non-RT RIC will form the inference. Inference logic to define cell congestion can be like:

- 1) The average user-perceived IP throughput < P Mbps.
- 2) DL PRB utilization > Q%.
- 3) Average RRC User > R.

The inference will contain the cell ids, information about whether the cell is congested or not, time stamp of cell congestion and predicted KPI value (to decide the congestion intensity). As per the CPM rApp information in Non-RT RIC, there are two options to mitigate cell congestion as follows:

**Option a:** CPM rApp in Non-RT RIC transfers the inference to the CPM xApp in Near-RT RIC through A1 interface. Near-RT RIC can decide the mitigation solutions as per the inference. Mitigation solutions can include:

- 1) Switching to dual connectivity mode.
- 2) Debarring of user access.
- 3) Load sharing.

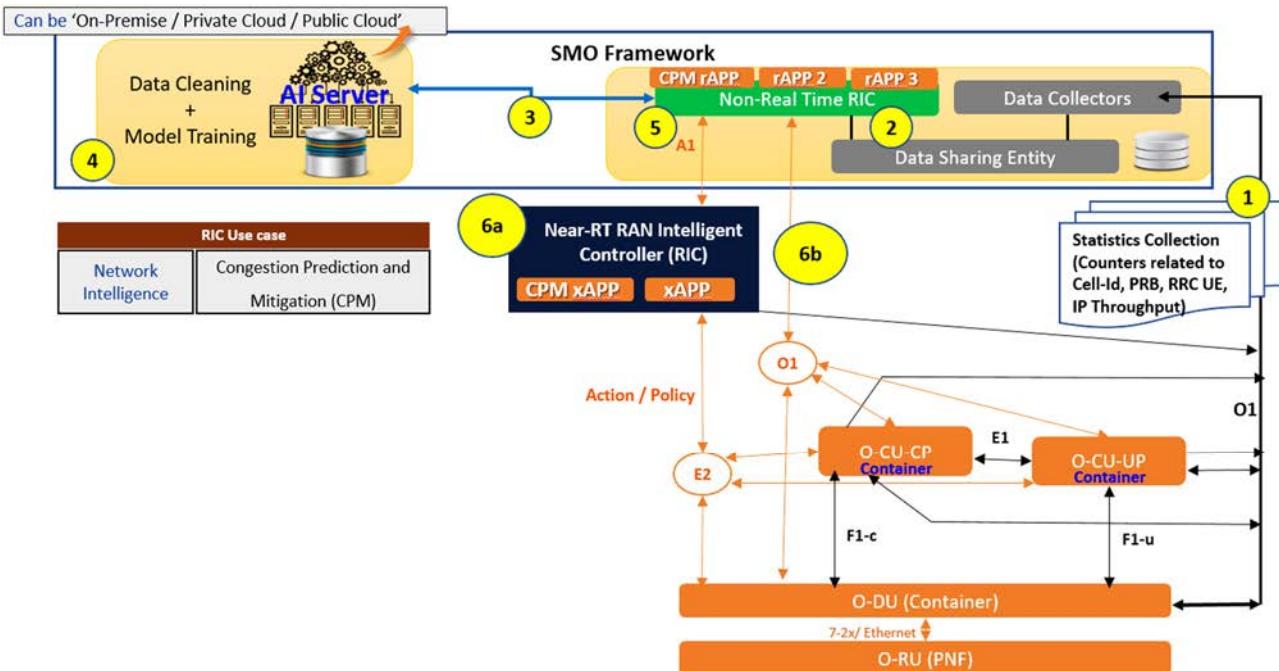
These solutions can be controlled over E2 interface.

**Option b:** Non-RT RIC can also directly help to mitigate the congestion with the help of O1 interface. Some of the mitigation solutions can be:

- 1) Splitting a cell (assuming hardware support available).
- 2) Add more carriers.
- 3) Switching to higher order MIMO.
- 4) Switching some of the users to Wi-Fi.

Finally, the E2 nodes feedback can be sent to Non-RT RIC through O1 interface. The proposed CPM solution is shown in figure 4.16.3-1 in detail.

NOTE: There can be more mitigation solutions apart from the one provided in steps 6a and 6b (see figure 4.16.3-1). Choosing the mitigation solution is up to the operator to decide as per the congestion intensity (specified in the inference).



**Figure 4.16.3-1: Proposed Congestion Prediction Management Solution**

#### 4.16.4 Benefits of O-RAN Architecture

All the main O-RAN components can leverage from proposed CPM use case. It predicts congestion before its occurrence so that an operator can configure the cell and prevent the degradation of QoS of a cell. AI server in SMO can be utilized for model training of KPIs. Similar models can be used to predict KPIs for other use cases as well. Non-RT RIC can be utilized to prepare inference based on predicted KPIs and congestion logic. These inference can be helpful for operators to choose on what solutions can be applied for a give congestion severity. CPM utilize Near-RT RIC for configuring a cell well before the congestion based on the inference and infrastructure dynamics.

### 4.17 Use case 17: Industrial IoT Optimization

#### 4.17.1 Background Information

In 3GPP Industrial Internet of Thing (IIoT) item, new scenarios with high reliability (i.e.  $10^{-5} \sim 10^{-6}$ ) have been considered including factory automation, transport industry. To satisfy these scenarios requirements, several key features as below have been supported for IIoT in 5G system, such as data duplication and multi-connectivity enhancement (PDCP duplication, PDU session duplication, QoS flow duplication), Time Sensitive Networking (Ethernet Header Compression (EHC), Accurate reference timing provisioning, QoS and scheduling enhancements) and different prioritized transmission multiplexing.

Based on O-RAN architecture some of these features can be optimized, e.g. PDCP duplication, EHC, and different prioritized transmission multiplexing. For PDCP duplication, up to 4 RLC entities/legs can be supported. RRC signaling is used for initial configurations. And MAC CE is used for dynamic control. According to 3GPP specifications definition, PDCP duplication can only be applicable for NR. For EHC, PDCP entities associated with DRBs can be configured by RRC to use EHC. Each PDCP entity carrying user plane data can be configured to use EHC. Every PDCP entity uses at most one EHC compressor instance and at most one EHC decompressor instance. For different prioritized transmission multiplexing, the lower priority transmission can be cancelled and preempted by the higher priority transmissions. RRC signaling is used for the related configurations.

Network deployment of IIoT use case includes both public network and non-public network (NPN). For public network, network slicing can be used to implement the resource isolation to guarantee the performance of IIoT use case. For NPN or private network, it can have different ways to implement the network deployment. For example, one way is to deploy private network equipment completely, such as gNB, UPF and 5GC. The network equipment can only provide services for the private use. The other way is to use network slicing to implement a virtual NPN, which means only some of the network equipment are deployed for private use, such as gNB. Meanwhile, the others, such as UPF, 5GC, are shared with the public network. The network resource isolation can be implemented by network slicing. If network slicing is used, IIoT use case only focuses on optimization for the specific piece of slicing bearing IIoT traffic.

## 4.17.2 Motivation

The objective of the IIoT optimization use case is to guarantee reliability, and resource utilization efficiency in the above scenarios by optimizing the enhancement mechanisms of related technique features, e.g. PDCP duplication, EHC, different prioritized transmission multiplexing and etc.

PDCP duplication and EHC is applied to the corresponding DRB. Improper configuration of PDCP duplication and EHC will leads to the waste of HW/SW and transmission resources. What service or data radio bearer (DRB) needs to use PDCP duplication is not only depends on the reliability requirement, but also related to the network environment, such as network load and channel quality/interference condition. Besides, what service or DRB needs to use EHC is depends on the service characteristics, such as service type, traffic period, duration, and packet size. EHC is especially beneficial when payload sizes are relatively small compared to the overall size of the frame, which is typically the case in IIoT networks based on Ethernet. Additionally, for different prioritized transmission multiplexing, which and where users are configured to listen to the Pre-emption Indication (PI) or Cancellation Indication (CI) depends on the service characteristics and network conditions. Improper configuration of PI or CI related configurations will cause invalid monitoring and power consumptions.

Therefore, the enhancement mechanisms for IIoT should be semi-static or dynamic configured for some specific users and some specific services based on the service prediction and performance prediction, considering the variant network load and channel quality/interference environment. AI/ML can be used for service prediction and KPI prediction to infer and derive the strategies.

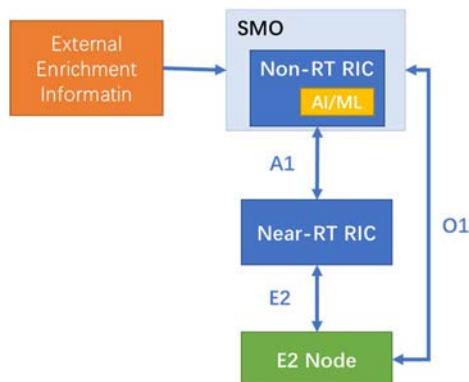
## 4.17.3 Proposed Solution

### 4.17.3.0 Non-RT RIC and Near-RT RIC Optimization Solutions

The proposed solutions include Non-RT RIC optimization solution and Near-RT RIC optimization solution. The Non-RT RIC optimization is a slower loop while Near-RT RIC solution has faster loop.

#### 4.17.3.1 Non-RT RIC solution of IIoT optimization

The architecture of Non-RT RIC solution of IIoT optimization use case is given in figure 4.17.3.1-1.



**Figure 4.17.3.1-1: Architecture of Non-RT RIC solution of IIoT optimization use case**

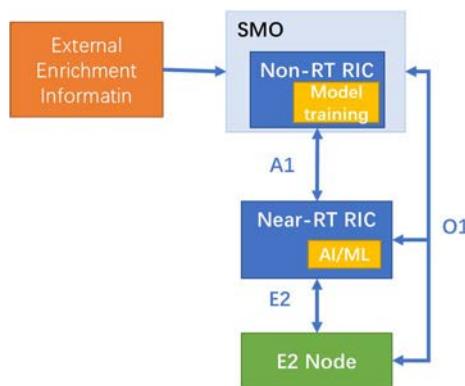
Non-RT RIC should be able to retrieve network data (network load, performance metrics, QoS flow 5QI and etc.) and external service information from SMO. The external enrichment information such as service type, traffic period, duration, packet size, KPI requirements can be provided by MEC server, APP server or industrial control platform and etc., to SMO. Non-RT RIC should be able to retrieve configured policy from SMO, e.g. cell edge user reliability KPI target, Ethernet header compression efficiency. Additionally, Non-RT RIC should be able to deploy and update AI/ML model from SMO. Non-RT RIC should use AI/ML model to infer and decide related IIoT key techniques configurations, such as whether the QoS flow should map to DRB supporting duplication and EHC, whether high priority traffic can preempt transmission resource by cancelling the low priority traffic transmission.

Near-RT RIC should be able to retrieve IIoT related configurations from Non-RT RIC (if it configures over A1). Near-RT RIC should be able to send the configurations to E2 nodes by policy.

E2 Node (O-CU, O-DU) should provide UE measurement report, performance metrics, and etc. to SMO through O1 interface. E2 Node should enforce E2 control policy or O1 configuration, if O1 configuration per UE can be supported.

#### 4.17.3.2 Near-RT RIC solution of IIoT optimization

The architecture of Near-RT RIC solution of IIoT optimization use case is given in figure 4.17.3.2-1.



**Figure 4.17.3.2-1: Architecture of Near-RT RIC solution of IIoT optimization use case**

Non-RT RIC should be able to retrieve network data (network load, performance metrics, QoS flow 5QI, and etc.) and external service information from SMO. The external enrichment information such as service type, traffic period, duration, packet size, KPI requirements can be provided by MEC server, APP server or industrial control platform and etc., to SMO. Non-RT RIC should be able to train AI/ML model for Near-RT RIC. Additionally, Non-RT RIC should be able to evaluate the collected data and A1 policy feedback, if required, and generate or update the appropriate optimization policy, such as reliability targets, and sends it to the Near-RT RIC via A1 interface.

Near-RT RIC should be able to deploy or update AI/ML model via O1 interface. Near-RT RIC should be able to receive an A1 policy from the Non-RT RIC, and initiate the corresponding optimization procedure. Additionally, Near-RT RIC should be able to evaluate the performance data from the E2 Nodes and monitor whether the performance is out of KPI targets which are indicated in the A1 policy. E.g. cell edge user performance cannot meet the A1 policy. Based on service estimation, i.e. service characteristics, and channel quality estimation, and A1 policy, the Near-RT RIC should infer and make decision of PDCP duplication on/off, RLC entity selection and high priority traffic preemption. Then, Near-RT RIC can generate new E2 policies or modify the existing ones to send them to the E2 Nodes. If required, Near-RT RIC should send report to Non-RT RIC for evaluation and optimization.

E2 Node (O-CU, O-DU) should provide network metrics and etc. to SMO through O1 interface. E2 Node should provide UE measurement report, performance metrics, and cell resource utility to Near-RT RIC through E2 interface. Additionally, E2 Node should enforce E2 control policy.

#### 4.17.4 Benefits of O-RAN Architecture

The benefits of O-RAN architecture for IIoT use case include the ability to use external enrichment information provided by MEC server, APPs server, or industrial control platform to SMO. Based on the external service information and RAN information, AI/ML method is introduced to help generating IIoT key techniques strategies to improve performance of system and users. For SDAP sub-layer, it can provide semi-static QoS flow to DRB mapping rules. For PDCP and RLC sub-layer, semi-static EHC on/off configuration, semi-static or dynamic PDCP duplication activate/deactivate decision and RLC entities dynamic control/selection decision can be provided. For MAC and PHY, it can provide semi-static and dynamic PI and CI configurations when different prioritized transmissions are multiplexed. Besides, multi-E2 Node coordination will be beneficial based on O-RAN architecture.

### 4.18 Use case 18: BBU Pooling to achieve RAN Elasticity

#### 4.18.1 Background Information

Driving cost efficiency by improving the elasticity of workloads is one of the principal goals of cloudification. This use case presents BBU pooling as an opportunity to achieve RAN elasticity. O-RAN CADS describes multiple deployment options for the O-RAN cloudified NFs, which allows for different variants of centralization of these resources in the edge and regional clouds. In particular, cloudified BBUs can be deployed on a common hardware pool, centralized at the same cloud location. This centralization opens up the opportunity for O-RUs to be flexibly mapped to BBUs, thereby enabling RAN elasticity, while potentially lowering the costs in the long term.

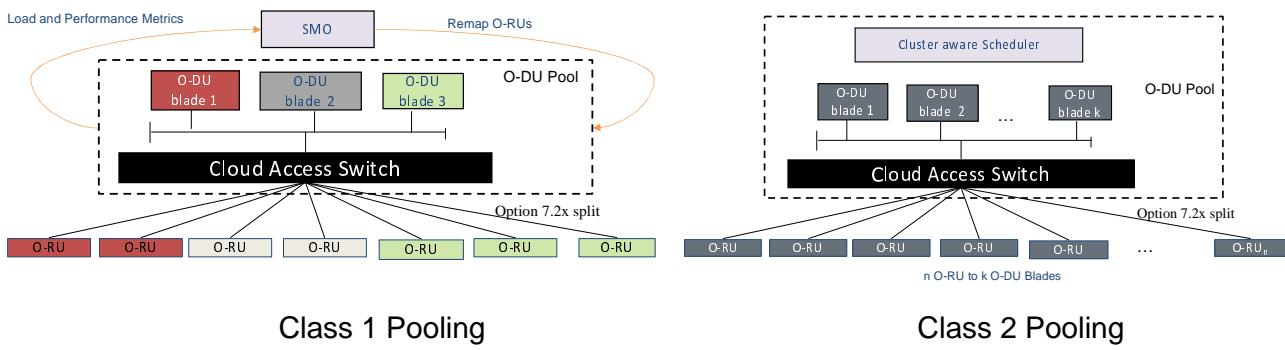
#### 4.18.2 Motivation

Pooling of BBU resources provides many benefits like having to maintain only a single large, controlled environment for many cell sites, better resiliency and agility to deal with unexpected failures and increases in demands, and allowing capacity to be added incrementally and assigned to cell sites as needed resulting in better overall utilization. More importantly, BBU Pooling provides significant statistical multiplexing benefits due to better resource consolidation. The total resources required from the shared pool will be less than total resources required across distributed locations, because the peak of the sum of the traffic will generally be lower than the sum of the individual cell site traffic peaks. BBU Pooling enables more aggressive capacity dimensioning based on different percentiles of resource consumption rather than for the peak consumption. This proposal addresses the scenarios where both O-CUs and O-DUs can pooled, but for simplicity the following sections focuses on the description of O-DU pooling.

#### 4.18.3 Proposed Solution

A high-level depiction of approaches to O-DU Pooling is shown in figure 4.18.3-1. The O-DU resources for multiple cell sites are pooled at a single centralized location or the *O-DU Pool*. The Cloud Access Switch (CAS) at the O-DU Pool aggregates traffic from the multiple Front Haul Gateways and routes it to the appropriate O-DU blade based on the load balancing requirement. For the O-DU pooling use case, it is assumed that an open Front Haul exists with O-RAN defined 7-2x split. Based on the association between O-RUs and O-DUs and the granularity at which the traffic is assigned from the O-RUs to the O-DUs, the following classes of O-DU pooling are defined.

- 1) **Class 0 pooling:** A scenario where an O-RU is assigned to a single specific O-DU statically, and the traffic from the O-RU is not split into subsets that could be assigned to different O-DUs. This is the same as Simple Centralization as defined in CADS V2.0. With Class 0 pooling, re-assigning an O-RU to connect to a different O-DU would require significant "hands on" activity, and is performed very infrequently during specific maintenance windows.
- 2) **Class 1 pooling:** A scenario where  $n$  O-RUs are initially assigned to a single O-DU during a specific period of time, but the O-RUs can be re-assigned to different O-DU at any point of time via orchestration procedures with the help of SMO. This automated re-assignment can be triggered outside maintenance windows, based on various performance criterion or for load balancing. Like with Class 0 pooling, the traffic from the O-RU is not split into subsets that could be assigned to different O-DUs.
- 3) **Class 2 pooling:** A scenario where  $n$  O-RUs are assigned to  $m$  O-DUs, and subsets of traffic from one O-RU are dynamically distributed (and load-balanced) across the O-DU resources within the O-DU pool using a cluster aware scheduler.



**Figure 4.18.3-1: O-DU Pooling**

With all classes of O-DU pooling, the total resources required from the shared pool will be less than total resources required across distributed locations, because the peak of the sum of the traffic will generally be lower than the sum of the individual cell site traffic peaks. In particular, Class 1 and Class 2 O-DU pooling enables more aggressive capacity dimensioning based on different percentiles of resource consumption rather than for the peak consumption. However, the key distinction between the classes of pooling is the granularity at which traffic can be load balanced across the O-DUs, which provides different statistical multiplexing benefits. While Class 0 pooling provides a one-time statistical multiplexing gains at the time of initial deployment due to the centralization, Class 1 pooling allows further load balancing benefits due to the automated and flexible re-mapping of O-RUs to the O-DUs periodically. However, the re-assignment of O-RUs needs to be performed during maintenance windows (or via hitless migration) to minimize service interruptions. Finally, Class 1 Pooling enables better handling of hot spots, and more efficient management of traffic loads and maintenance/upgrades.

On the other hand, Class 2 pooling dynamically assigns subsets of traffic to the O-DU resources in the shared O-DU pool. Hence, if some cell sites experience light traffic demand, while others experience high traffic demand, then subsets of traffic can be routed to O-DU resources in the shared pool such that the traffic is optimally consolidated among the O-DU resources needed to meet the performance requirements, leading to significant statistical multiplexing benefits. At one extreme, Class 2 pooling allows each O-DU in the O-DU pool to have similar load (once the optimal number of blades have been predicted for a given time period), if each UE flow or each slot processing can be mapped to the least-busy blade for the duration of that flow. In addition, there is no service interruption as with Class 1 pooling. Class 2 pooling could also be performed with more coarse-grained distribution of traffic subsets to the O-DUs. For example, traffic could be mapped to the least-utilized blade on a per UE basis or on a per-carrier basis within a TRP.

#### 4.18.4 Benefits of O-RAN Architecture

As described in clause 4.18.3, both Class 1 and Class 2 Pooling are enabled by the O-RAN Open Fronthaul with LLS (option 7.2x split). Further, for Class 1 Pooling, O-RAN Architecture standardizes the SMO functions needed to enable pooled BBUs including:

- Management of O-DU & O-RU (Registration/De-registration of O-RU to O-DU, Registration of RU locations to SMO).
- Orchestration features via O1/O2 interfaces and procedures for remapping O-RUs from one O-DU to another O-DU, implementing the Orchestration control loops provided by the non-RT RIC, hitless draining of cell traffic from one O-DU to another O-DU.
- Monitoring Performance/Fault metrics to determine DU load and KPIs.

Class 2 pooling is relatively less explored in the industry, aiming to provide fine-grained load scheduling across the O-DUs in the pool. For Class 2 Pooling, O-RAN Architecture helps with:

- Affinity of control and downlink data, supported by cluster aware scheduler.
- Affinity between control and uplink data, supported by the Cloud Access Switch or in the O-RU.
- Configuration of the O-DU to support a wider range of sectors through the cluster aware scheduler.
- Monitoring Performance/Fault metrics to determine DU load and KPIs.

- Triggers for the scale-in/scale-out workflows in the SMO that manages the O-DU Pool.
- Fine-grained load balancing functions through the RIC that standardizes the role of CU-CP High.

## 4.19 Use case 19: Integrated SON Function

### 4.19.1 Background Information

Self-Organizing Network (SON) functionalities for 5G (3GPP TS 28.313 [i.14]) reduces the cost of running a mobile network by providing automated configuration, optimization, protection and healing functions and thus eliminating manual configuration of network elements right from initial deployment through the network operation. SON also helps better network performance and customer experience and can significantly improve Opex-to-revenue ratio and help in realizing avoidable Capex.

SON is an automation technology that enables the network to set itself up and self-manage resources and configuration to achieve optimal performance. The SON functions are handled by SON algorithms either individually or in groups. SON algorithms monitor the network(s) by collecting management data including MDAS (Management Data Analytics Service) data and subsequently analyze the data to determine if there are issues in the network(s) and what needs to be resolved.

3GPP SON specification 3GPP TS 28.313 [i.14] specifies the concepts, use cases, requirements, and procedures for the SON functionality. Based on the location of the SON algorithm, the implementation of SON solution is termed as:

- Centralized SON (C-SON - where the SON algorithms are executed in the 3GPP management system);
- Distributed SON (D-SON - where the SON algorithms are executed in the Network Function layer); and
- Hybrid SON (where the SON algorithm execution is spread across the network function layers and the management layers).

Since the interfaces between various entities hosting the SON algorithms are not standardized, there can be interoperability issues in multi-vendor network deployments. This is further explained in detail in clause 4.19.2.

### 4.19.2 Motivation

The objective of this use case is to enable the realization of SON functions in the O-RAN architecture framework i.e. as rApps, xApps or as management entity functions through open interfaces in a way that inter vendor interoperability issues can be addressed.

The following SON functions are of interest for operators for this use case definition:

- Initial PCI allocation, conflict resolution and ANR (Automatic neighbor relations)
- Mobility load balancing (MLB) and Mobility robustness optimization (MRO)
- RACH optimization
- Energy savings
- Coverage and Capacity Optimization

When the SON algorithms are realized in a centralized location, as a C-SON variant, the interface between C-SON and the RAN nodes, being proprietary, demands the operator to choose the same vendor for both SON solution and the RAN nodes. When the SON algorithms are realized in distributed fashion, as a D-SON variant, the interface between D-SON and the RAN nodes could be proprietary and internal to vendor's design. The performance of such vendor specific SON solutions affect KPIs as typical 5G deployments are expected to be multi-vendor based.

When operators use management entities like NMS, EMS from a different vendor and RAN nodes (gNB-CUs, gNB-DUs) from different vendors, the following issues are observed:

- D-SON implementations in different gNB-CU vendors cannot interoperate though they communicate each other via open Xn interface.

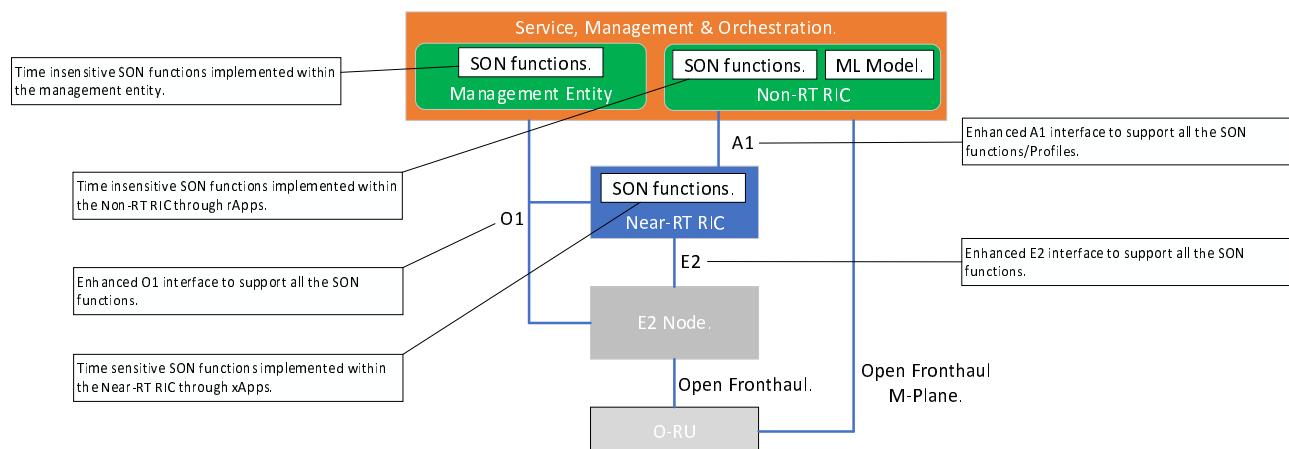
- C-SON can be realized as either collocated with management entities (e.g. EMS/NMS) or as a standalone entity. Integrating the C-SON as a standalone entity with RAN nodes could lead to interoperability issues due to vendor specific interfaces.

Integrating third party SON solutions partially leads to degradation of the overall operator KPIs. Non interoperable SON / RRM implementations have been seen to significantly impact the overall network performance. Even with operator assisted vendor and integrated third-party SON Solutions, it is difficult to deterministically quantify/confirm the output performance mainly due to the nature of multi-vendor implementations.

The nature of band allocations in different geographic regions have led to some operators to consider using these new bands along with LTE bands for 5G network deployment. Large-scale deployment of small cells in urban areas is required to accommodate large traffic especially with smaller millimeter wave band cells (i.e. cell ISD of 100~200 meters). Therefore, the acquisition, deployment, and operation of ultra-dense cells become more important. This will force operators to explore small-sized and low weight base stations where multi-vendor interoperability of end-to-end system and realization of SON technology becomes necessary.

#### 4.19.3 Proposed Solution

The general SON functional solution in the O-RAN architecture is shown in figure 4.19.3-1. The SON algorithms are realized as rApps or as xApps or as a combination thereof through open interfaces in a way that inter vendor interoperability issues can be addressed. The A1 and E2 and optionally O1 interfaces will be used for the realization of the SON functions.



**Figure 4.19.3-1: SON Functions**

As shown in figure 4.19.3-1, time insensitive SON functions like Initial PCI, Initial ANR, and Energy Saving can be implemented either at Non-RT RIC using separate rApp or in the management entities like EMS, NMS or at the SMO. A group of such functionalities can also be combined and executed in an rApp. The SON function related information can be exchanged either directly via O1 interface or via A1 & E2 Interfaces. The dynamic or time sensitive SON functions will be realized in the xApps of RT-RIC. The actual SON algorithm implementation within xApps/rApps/Management entities are implementation specific and it is out of scope of the current proposal, and the SON functions provide the derived SON configurations to the associated E2 Nodes via E2 interface (in case of xApps).

This approach helps provide a framework for the realization of the following SON functions within the existing O-RAN architecture. As an illustrative example, consider the PCI allocation SON function:

When an operator wants to deploy a cell, the location for the E2 Node/O-RU deployment is identified. As part of the plug-and-play functionality, the E2 Node first establishes the link with the Management entity via the O1 interface, to get the FCAPS configurations. Before the E2 node becomes operational, it needs the initial PCI allocated. At this stage as the E2 node will not be able to establish any link with the Near-RT RIC, there should be some SON functions within the SMO that can derive the initial SON configurations. Within the SMO, the SON functions can be realized either as rApps or through SON functions within the management entities. The Initial PCI SON function should be equipped to get the location information of the E2 Node being deployed. This can be provided by the E2 Node itself across the O1 interface or an application can directly provide the location information to the management entity. The management entity can internally share it with the Initial PCI function rApp. The initial PCI rApp can use this location information [Azimuth, elevation, Altitude, etc.,] to derive the possible neighbor nodes already deployed in and around the current node's location and the associated PCIs of these neighbors. The rApp can then derive a mutually exclusive PCI for the current node being deployed. This derived PCI acts as an initial PCI for the E2 Node being deployed. This allocated PCI will be shared to the management entity internally and which in turn forwards to the node via the O1 interface. The E2 Node will be operational with this initial PCI. Similarly, the Initial PCI SON function can also be implemented within the Management entity itself. In this case it can access the Node's location information within the Management entity database and the rest of the functionalities defined above for rApp case follows.

When an E2 Node becomes operational, there can be race conditions with other E2 nodes becoming operational resulting in a PCI conflict. To detect and correct such PCI conflicts during the lifetime of a deployed node, there can be a dynamic PCI SON functions, which could reside at the Near-RT RIC level as such PCI conflicts that needs to be detected and corrected at the earliest.

#### 4.19.4 Benefits of O-RAN Architecture

The functional entities of the current O-RAN Architecture can be reused to realize SON functions to operators. The O-RAN specified interfaces such as O1, E2 and A1 interfaces will also be reused with appropriate modifications to enable the SON functions.

By using the O-RAN framework, third-party SON algorithms (realized as an rApp, xApp or as a management entity function) can be easily integrated to achieve optimal performance, across a multi-vendor deployment scenario.

### 4.20 Use case 20: Shared O-RU

#### 4.20.1 Background Information

The existing fronthaul specifications include role-based access control capabilities to restrict permissions to certain sections of an O-RU's configuration based on privileges as well as supporting parallel management interfaces for realizing hybrid fronthaul deployments. However, the current lower layer split architecture is constrained by having an O-RU node operate with a single O-DU node. This limits the ability of fronthaul system to be used to deliver enhanced use cases where multiple O-DU nodes can be used to enhance the capabilities offered using O-RAN's open fronthaul specifications.

#### 4.20.2 Motivation

Already defined use cases, including RAN Sharing as defined in clause 4.7 and Multi-Vendor Slices as defined in clause 4.10 describe use cases where a common O-RU is operated with a plurality of O-DUs. More generically, deployments of a common O-RU node with multiple O-DU nodes offers the ability to support enhanced capabilities when using O-RAN's lower layer split, including:

- Enhanced scalability and/or enhanced availability and/or enable access-specific node deployments in fronthaul systems, where multiple O-DU nodes are deployed by a single operator and connected to a Shared O-RU.
- Enhanced sharing capabilities where multiple O-DU nodes are deployed by different operators and connected to a common Shared O-RU node.

For single operator deployments, the decomposition of the RAN into O-RU and O-DU functional components based on interoperable O-RAN defined interfaces enables new approaches to the delivery of enhanced systems whereby multiple O-DU nodes can be deployed in different configurations, including:

- "active/standby configuration" to support increased availability;
- "horizontal configuration" by adding more O-DU nodes to (or removing nodes from) an O-RAN system to enable scale in/out;
- "access-centric configuration" to support combined LTE and NR deployments, where one or more O-DU nodes are used to deliver LTE functionality and one or more other O-DU nodes are used to deliver NR functionality;
- configuration supporting a case to add new O-DU provided from a different vendor, migrating to O-DU with providing slices. Radio resource is allocated among O-DUs/slices dynamically. Single O-CU will cover multiple O-DUs per UE.

As described elsewhere in the present document, RAN sharing is envisioned as an efficient and sustainable way to reduce the network deployment costs, while increasing network capacity and coverage. Accordingly, the open and multivendor nature of the O-RAN architecture can accelerate the introduction and development of RAN sharing solutions.

For multi-operator deployments, the decomposition of the RAN into O-RU and O-DU functional components based on interoperable O-RAN defined interfaces leads to new options for RAN sharing:

- Shared O-RU, where a common O-RU node is shared and dedicated O-RU resources (e.g. spectrum, transmit power, etc.) and dedicated O-DU node is used by each sharing operator.
- Shared O-RU, where a common O-RU is shared and dedicated O-DU node is used by each sharing operator and O-RU resources are dynamically shared between sharing operators.

Whilst there are a variety of RAN sharing deployment models, a key motivation of this use case is to enable O-RAN based architectures to be able to support Shared O-RU based RAN sharing. In the conventional radio node based sharing, a "host" operator makes available a shared radio node as part of distributed antenna system (DAS) that can be partitioned between one or more mobile operators. The host operator then integrates with the RAN equipment of the one or more mobile operators (MNOs), where the interface between DAS radio node and RAN equipment can re-use an attenuated Uu interface. Translating this into the O-RAN architecture, the host operator is responsible for operating the Shared O-RU that can integrate with the O-DUs of one or more mobile operators. Hereafter such a host operator is referred to as a Shared Resource Operator – Host (SRO-H) and the mobile operators operating the O-DUs are referred to as Shared Resource Operator - Tenants (SRO-T). The SRO-H can operate its own O-DU and O-CU functionalities, in which case the O-RU is shared between the SRO-H and SRO-T(s). Alternatively, the host can operate the O-RU as a neutral-host, in which case the O-RU is shared between the SRO-Ts. Hereafter such a host operator is referred to as a Shared Resource Operator - Neutral Host (SRO-NH).

When such RAN sharing is coupled with established open fronthaul specified shared cell functionality, where multiple O-RU nodes are used to support a common cell, then it is evident that there is an opportunity for the O-RAN ecosystem to be re-applied to address the distributed antenna system market; a market that is estimated to have a global revenue of over USD 19 Bn by 2025 according to Market Data Forecast (<https://www.marketdataforecast.com>).

## 4.20.3 Proposed Solution

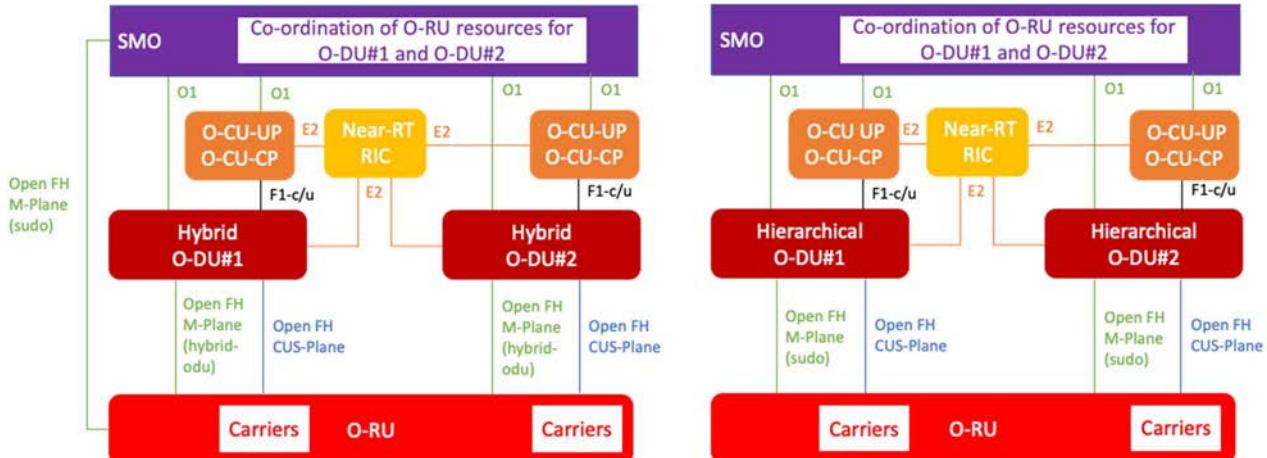
### 4.20.3.1 Static allocation of resources between O-DU nodes

#### 4.20.3.1.1 Introduction

The initial phase of the proposed solution is to enhance the current open fronthaul architecture to enable a common O-RU node to operate with multiple O-DU nodes with the limitation that resources are statically allocated to separate O-DU nodes.

#### 4.20.3.1.2 Single Operator Use Case

The single operator use cases is shown in figure 4.20.3.1.2-1, for either hybrid or hierarchical open fronthaul deployments. Because resources are statically allocated, co-ordination of O-RU resources is performed by the operator's SMO. In this figure, two set of O-CU and O-DU are described but single O-CU can also cover multiple O-DUs per one UE.



**Figure 4.20.3.1.2-1: Single Operator Solution for Hybrid and Hierarchical Open Fronthaul Deployments**

In the case of Active/Standby operation, the SMO is responsible for configuring identical carrier configurations on separate O-DU nodes. The data model of the O-RU is enhanced with capabilities to associate the fronthaul transport flows associated with an active O-DU node, which necessarily includes the O-DU node's CU-Plane transport endpoint, with the corresponding fronthaul transport flows associated with a standby O-DU node. The operation of the stand-by O-DU node is enhanced such that carrier configuration of the O-RU is omitted, which is instead performed only by the active O-DU node. The stand-by O-DU node subscribes to receive supervision notifications from the O-RU. The switchover from active to stand-by O-DU node can be initiated by the SMO, e.g. enabling operation of a planned procedure to de-commission the active O-DU node. Alternatively, the switchover from active to stand-by O-DU node can be initiated automatically whereby the stand-by O-DU node becomes aware that the Active O-DU has become non-operational which triggers the switchover procedure.

In the case of Scale in/Scale out operation, the SMO statically partitions O-RU resources between a set of candidate O-DU nodes. The SMO oversees the definition and reporting of long-term traffic metrics in a whole cluster of cells that can be served by the candidate O-DUs, e.g. using the Non-RT RIC. The SMO uses the O1 interface to either:

- i) activate and configure; or
- ii) de-activate and de-commission candidate O-DU nodes in response to varying traffic load.

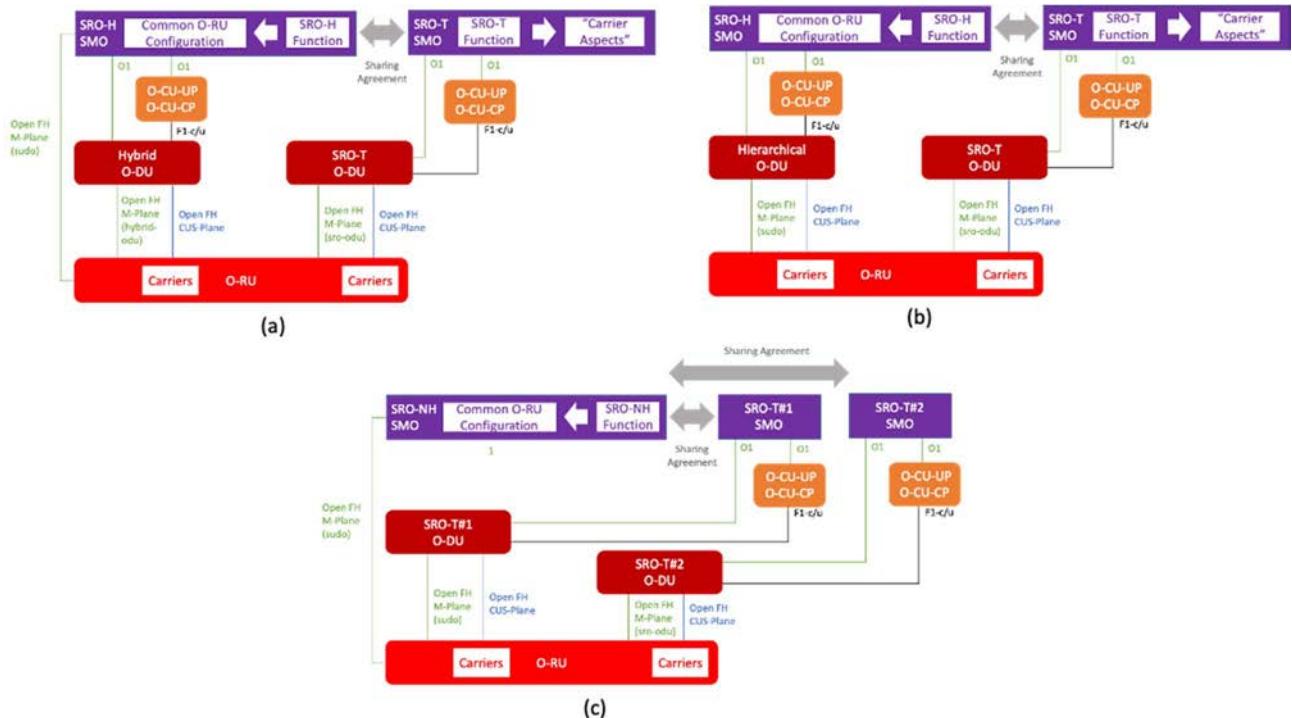
In the case of "access-centric configuration" to support combined LTE and NR deployments, the SMO statically partitions O-RU resources between LTE and NR carriers. The SMO uses the O1 interface to configure one or more O-DU nodes for each of the respective radio technologies, e.g. including associated DSS aspects such as MBSFN configuration, subframe masks and rate matching configurations.

#### 4.20.3.1.3 Multi-Operator Use Case

The multi-operator use cases are shown in figure 4.20.3.1.3-1, where:

- a) shows the deployment with hierarchical open fronthaul;
- b) shows the deployment with hybrid open fronthaul; and
- c) shows the deployment with neutral host.

In all three cases a SRO-H enters into an agreement with one or more SRO-Ts to offer access to the resources of a Shared O-RU. The sharing agreement covers off those aspects of the O-RU's resources which are available to the SRO-T and which can be used in the SRO-T's configuration. In all cases, the SRO-H is responsible for configuring the common aspects of the Shared O-RU.



**Figure 4.20.3.1.3-1: Multi Operator Solution for Open Fronthaul Deployments**

The Shared O-RU is enhanced with additional role-based access control capabilities to restrict the ability of an SRO-T from accessing the configuration of the Shared O-RU associated with resources to which it has not been authorized, e.g. those resources that have been allocated to a second SRO-T. Furthermore, the Shared O-RU's enhanced role-based access control capabilities restrict the ability of an SRO-T from receiving performance management information to that which is associated with its authorized set of Shared O-RU resources. Additional capabilities are defined to support supervision of the communications link between the Shared O-RU and an SRO-T, with an alarm indication able to indicate to the SRO-H that the communications to the SRO-T have been interrupted. Unlike in conventional O-RU operation, the loss of supervision to an SRO-T O-DU does not result in the O-RU ceasing all radio transmissions and performing an autonomous reset. Instead, only those carriers associated with the SRO-T O-DU are disabled.

The SRO-H is able to validate that the SRO configuration adheres to the specifics of the sharing agreement between the parties. The SMO of the SRO-H can configure to be notified of modifications to the O-RU's configuration datastore. These notifications can be used to confirm that the sharing agreement is being adhered to. Any remedial action required, e.g. if the SRO-T configuration is not in compliance with the sharing agreement, is handled between the SRO-H and the SRO-T.

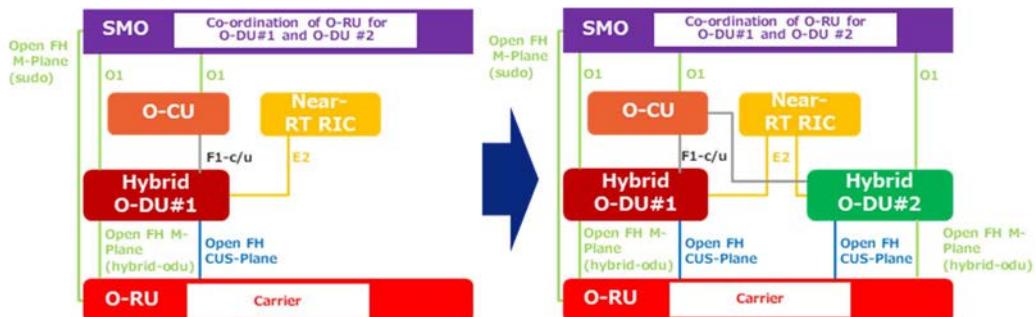
## 4.20.3.2 Dynamic allocation of resources between O-DU nodes

### 4.20.3.2.1 Introduction

This phase of the proposed solution is to enhance the current open fronthaul architecture to enable a common O-RU node to operate with multiple O-DU nodes with the limitation that resources are dynamically allocated to separate O-DU nodes.

#### 4.20.3.2.2 Single Operator Use Case

The single operator use cases is shown in figure 4.20.3.2.2-1 (shows only hybrid model), for either hybrid or hierarchical open fronthaul deployments. Because resources are dynamically allocated, co-ordination of O-RU resources is performed by the Near RT-RIC or M-plane based on demand and performance for slices. To keep improving frequency efficiency with providing slices concurrently, radio resource allocation will be expected frame by frame or in seconds. A single O-CU will cover multiple O-DUs per UE.



**Figure 4.20.3.2.2-1: Single Operator Solution for Hybrid Open Fronthaul Deployments migrating to vO-DU**

#### 4.20.4 Benefits of O-RAN Architecture

The proposed single operator Shared O-RU architectures shown in figure 4.20.3.1.2-1 let operators use the open fronthaul to enable a common O-RU node to operate with a plurality of O-DU nodes. Such a capability can be used to deliver enhanced deployment scenarios where multiple O-DU nodes are used to increase the resilience or scalability of the O-RAN fronthaul architecture. The proposed access-centric approach allows operators to utilize different O-DU vendor's strengths for a particular radio technology, enabling a Shared O-RU node configured for DSS to operate with an LTE O-DU node from vendor#1 and an NR O-DU node from vendor#2.

The proposed multi-operator Shared O-RU architectures shown in figure 4.20.3.1.3-1 let operators use the open fronthaul to enable a single O-RU node to be shared between multiple different operators. The architecture allows shared resource operators to configure Shared O-RU resources independently from configuration and operating strategies of the other shared resource operators while preventing one shared resource operator from gaining insight into the configuration and operational status of Shared O-RU resources allocated to a second shared resource operator. Moreover, the architecture enables the SRO-H to operate as an independent neutral host provider.

**NOTE:** Use cases where resources are dynamically allocated to separate O-DU nodes for Multi Operator use case are for further study.

## 4.21 Use case 21: Energy Saving

#### 4.21.1 Background Information

Energy Consumption (EC) of the RAN is an important topic for network operators, especially for 5G network. RAN Energy Saving (ES) depends on rigorous planning and configuration. Due to the varying nature of traffic load and to user mobility, the optimization of EC of the RAN is complex and can be applied to different network layers and in different time scales. There is a risk that RAN equipment consume much energy while serving low traffic, or even no traffic at all.

Different ES features are investigated in industry, some of which are deployed in operational networks. Examples are deep sleep mode (e.g. shut down of a base station of a given technology); carrier shut down; and RF channels' switch off/on). More recently, short time scales ES mechanisms have been proposed, at symbol-, subframe- and frame-levels, known as Advanced Sleep Modes (ASM).

O-RU is responsible for the major part of the mobile network EC. EC of VNFs have not yet been standardized in Release 17 of 3GPP and can be only estimated, e.g. based on mean vCPU usage of the underlying VMs. EC measurements will be further addressed in release 18. In O-RAN O-Cloud, the only feature with relation to ES is scale in/out that is mentioned in the use case document [i.22]. Sub-use cases involving scale-in / scale-out processes, workload placement or HW processors' sleep modes are not defined in the present document.

## 4.21.2 Motivation

ES for legacy and 5G networks can be carried out using manual configuration or SON functions. 3GPP defines both centralized and distributed ES features [i.13], which are mainly targeting intra- or inter-RAT cell on/off switching. The motivation of the ES use case is to leverage on ORAN AI/ML services and open interfaces in order to introduce optimized ES and EE solutions involving switching off/on of different network components at different time scale. Will be considered off/on switching solutions supported by 3GPP configurations or supported by propriety implementations that require additional standardization.

## 4.21.3 Proposed Solution

### 4.21.3.0 ES sub-use cases

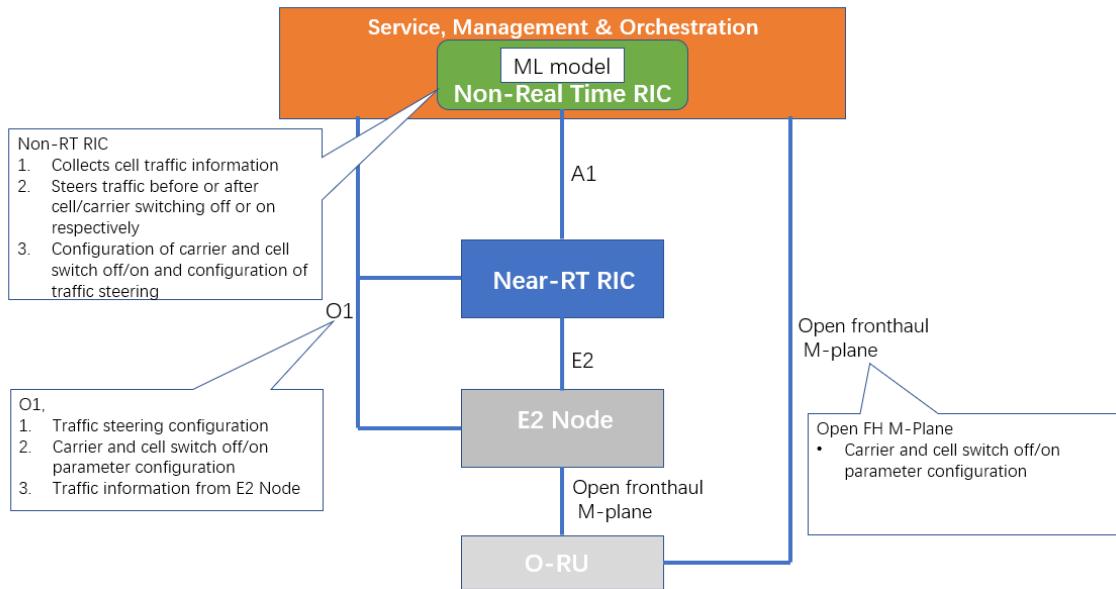
The ES use cases is divided into three sub-use cases, according to the time scale of the control and the controlled system involved:

- 1) Carrier and cell switch off/on ES. Time scale: non-real time for both control and controlled system. The feature aims at reducing O-CU/DU/RU power consumption by switching off/on one or more carriers or a cell of a given technology. AI/ML-assisted solutions in the Non-RT RIC can be used to control the traffic load of the carriers and the cell, and to automatically decide when to switch off/on one or more carriers or a cell using O1 and/or Open fronthaul M-plane parameter configurations. Off/on switching is accompanied with adequate traffic steering, guided by policies, to ensure service continuity and quality of service. Carrier switch off/on has two modes of operation:
  - i) hibernate mode in which the radio's power amplifiers remain with minimum current draw; and
  - ii) complete switch off.
- 2) RF channel switch off/on ES. Time scale: non- or near-real-time are possible for both control and controlled system. This feature aims at reducing power consumption of O-RU with massive MIMO deployment by switching off/on certain RF channels [i.8]. Using AI/ML assisted solutions, rApp or xApp will trigger switching off/on certain RF channels, based on traffic information such as load, user location and mobility. As example, one can switch off 32 out of 64 RF channels in a digital M-MIMO architecture or reduce the number of layers and/or number of Multi-User scheduled UEs in a hybrid architecture. The O-RU reconfiguration can be performed using the Open fronthaul M-plane from E2 node or SMO.
- 3) Advanced Sleep Mode ES. Time scale for control: near-real-time. Time scale for the controlled system: real-time and near-real-time. This feature is expected to reduce power consumption by partially switching off O-RU components. The impact of ASMs on O-DU and O-CU EC is not supported in the present document. Using multi-dimensional data, e.g. traffic load, user service type, energy efficiency measurements, etc., the Near-RT RIC can configure cell parameters, such as the SSB periodicity needed for the operation of ASMs.

### 4.21.3.1 ES sub-use case 1: Carrier and cell switch off/on

The carrier and cell switch off/on is a non-real-time sub-use case where the Non-RT RIC is in charge of configuring E2 nodes. This ES feature can be invoked in response to changes in traffic load or change in the cell/carrier off/on switching policy. The Non-RT RIC will use the O1 interface to configure the parameters of the O-CU/O-DU and will use the Open fronthaul M-plane from the SMO to configure O-RU switch off/on, after triggering handovers to move users to other cells / carriers. The O1 interface is also used to collect traffic information per cell and per carrier to the SMO as input to the Non-RT RIC. rApps hosting AI/ML assisted solutions can be used to achieve optimized ES configurations, such as the switching time that can provide optimal desired tradeoff between ES and users' QoS.

The concept of carrier and cell switch off/on ES sub-use case is given in figure 4.21.3.1-1.

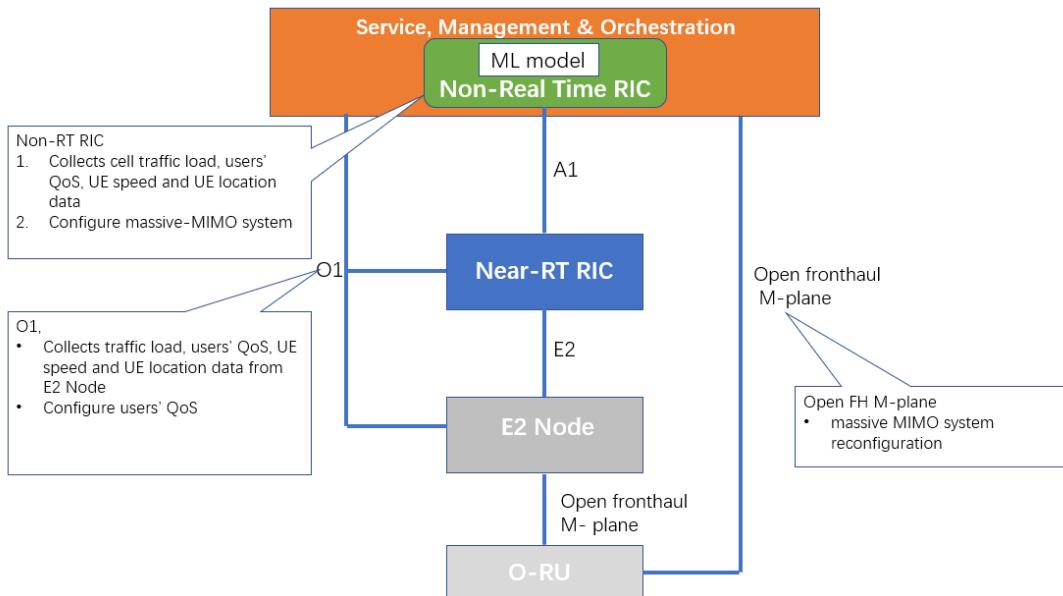


**Figure 4.21.3.1-1: Carrier and cell switch off/on ES sub-use case**

#### 4.21.3.2 EE/ES sub-use case 2: RF channel switch off/on

The first implementation of this sub-use case is non-real-time. This solution is expected to reduce power consumption by switching off certain RF channels of O-RU with massive MIMO deployment. Based on data collected from E2 nodes and O-RU, the solution will automatically switch off/on certain RF channels and configure the massive MIMO system to minimize the impact on users' QoS.

The concept of RF channel switch off/on ES sub-use case: Non-real time implementation is given in figure 4.21.3.2-1.

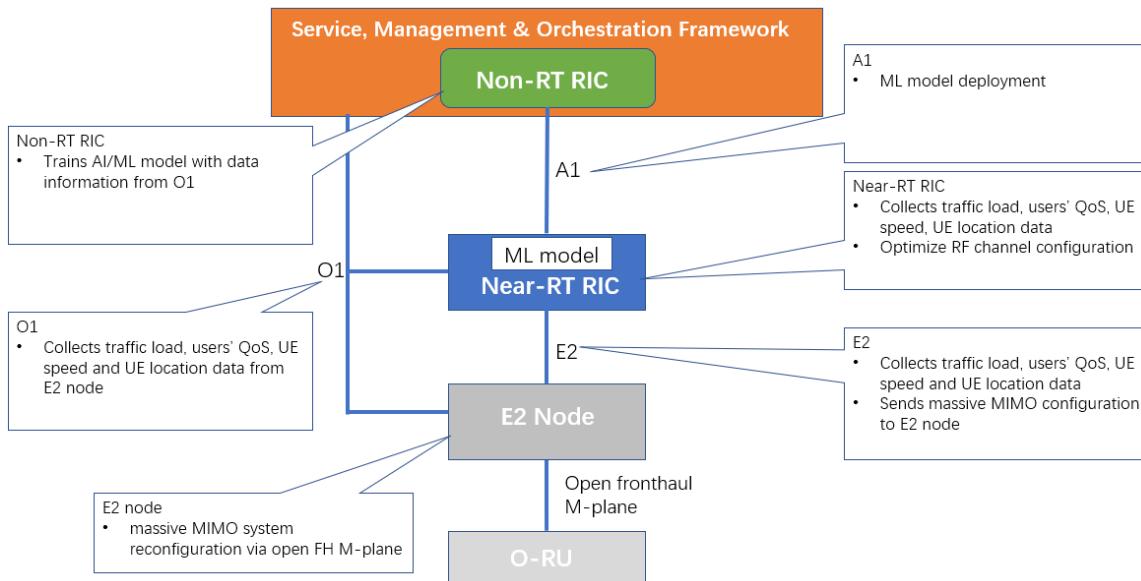


**Figure 4.21.3.2-1: RF channel switch off/on ES sub-use case. Non-real time implementation**

The second implementation of this sub-use case is near-real-time. As in the non-real-time implementation, this solution is expected to reduce EC by switching off certain RF channels of O-RU with massive MIMO.

The AI/ML training host in the SMO/Non-RT RIC will train AI/ML model based on data collected from applications and/or O1 interface such as cell traffic, users' QoS, users' speed or geolocation data. The Near-RT RIC uses the trained model to select the off/on switching time and to infer the configuration of the massive MIMO system based on information from E2 and Open fronthaul M-plane interfaces.

The concept of RF channel switch off/on ES sub-use case: Near-real time implementation is given in figure 4.21.3.2-2.



**Figure 4.21.3.2-2: RF channel switch off/on ES sub-use case. Near-real-time implementation**

#### 4.21.3.3 ES sub-use case 3: Advanced Sleep Modes

ASMs are used to reduce EC by partially and automatically switching off O-RU components, at a duration of a symbol (ASM1), a slot (ASM2) and a frame (ASM3). The deeper (longer) is the sleep mode level, the longer is its wake-up time. The impact of ASMs on O-DU and O-CU is not supported in the present document.

The aim of this sub-use case is to maximize the ASMs duration (and the corresponding ES) in which O-RU components are switched-off, taking into account user QoS constraints in terms of latency. To this end, ASM activation policy can be set (e.g. the order and number of ASMs in each level), and system configurations can be applied (such as setting the SSB periodicity; compress data transmissions in particular slots and empty the remaining ones that are put "into sleep state"; reduce the synchronization signals' transmissions and the opportunities for Random Access and paging).

The ASM activation policy and configurations can be derived using AI/ML models, aiming at achieving optimal tradeoffs between ES and QoS. The AI/ML training host in the SMO/Non-RT RIC will train an AI/ML model based on measurements such as traffic load and latency, on service information and will derive optimal system configuration (e.g. SSB periodicity, ASM sleep level and sleep duration). The Near-RT RIC will request the AI/ML model from A1 interface that, once deployed and activated, will optimize ES by setting ASM policies and system configurations via E2 and Open Fronthaul M-plane interfaces.

The concept of Advance Sleep Mode ES sub-use case is given in figure 4.21.3.3-1.

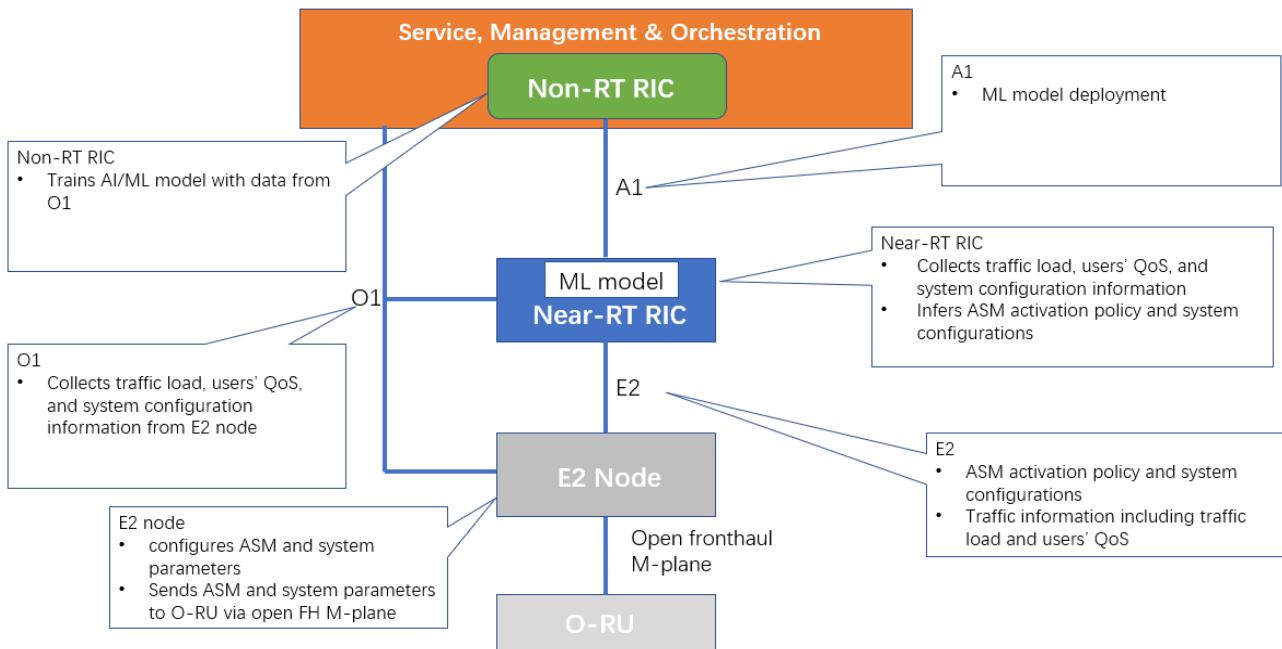


Figure 4.21.3.3-1: Advance Sleep Mode ES sub-use case

#### 4.21.4 Benefits of O-RAN Architecture

The proposed solutions will support the AI/ML assisted solutions at both Non-RT RIC and Near-RT RIC as inference hosts, to efficiently optimize ES and EE at different time scales. The ES solutions will leverage on O-RAN open interfaces to support the use case in a multi-vendor disaggregated RAN.

### 4.22 Use case 22: MU-MIMO Optimization

#### 4.22.1 Background Information

MU-MIMO is one of the key technologies available for increasing UE and cell capacities using existing time/frequency resources. The use of multiple antennas enables the pointing of beams to multiple UEs with each beam spatially filtering the interference from the other beams. This has the potential to provide higher total cell capacity when there are N eNB/gNB antennas.

In a commercial deployment, some subscribers can be stationary, some can be pedestrian moving slowly, and some can be moving at high speed. Traditional MU-MIMO solutions are very sensitive to subscriber mobility and as a result the capacity gains achieved with multiple antennas is limited.

New beamforming solutions are emerging that support MU-MIMO with less time sensitivity allowing them to be implemented in the Near-RT RIC. These solutions are applicable to both downlink and uplink data channels and both TDD and FDD, and can provide high user and cell performance for subscribers moving wide range of speeds.

#### 4.22.2 Motivation

The MU-MIMO optimization use case aims at enabling new spatial multiplexing and precoding solutions that have the potential for increasing user and overall cell capacity in a massive MIMO deployment area by selecting appropriate users over time and frequency resources, and for each selection of users recommending the applicable precoding coefficients and modulation and coding schemes (MCS) for most efficient resource usage.

#### 4.22.3 Proposed Solution

The use case proposes to have the MU-MIMO optimization function as an application in the Near-RT RIC. The application collects information from the E2 nodes and sends control messages to the E2 nodes using the E2 interface.

The main functionalities of the Near-RT RIC application include collecting UE state from the O-CU, configuring the O-CU / O-DU to report RRC parameters and UL/DL traffic and channel information, using the reported channel and traffic information to select the best UEs to be spatially multiplexed, calculating for each selection the applicable frequency/time resources, the MCS, and the beamforming coefficients for best MU-MIMO performance, and sending the recommended UEs selections, applicable resources, MCS, and beamforming coefficients to the O-DU over E2 interface.

The E2 nodes (O-CU/ O-DU) provide RRC parameters and UL/DL traffic and channel information to the Near-RT RIC and the O-DU can use the information received from the Near-RT RIC to schedule MU-MIMO transmissions, while still handling time critical events separately.

#### 4.22.4 Benefits of O-RAN Architecture

Using a RIC-based implementation of MU-MIMO can benefit from using the near-real-time data from the E2 nodes to influence how the 4G/5G MAC scheduler assigns the near-term resources and how the PHY forms the MU-MIMO beams for achieving improved per subscriber and overall cell throughput performance in certain deployment scenarios. Using the RIC-based implementation to support this use case also opens the door for future expansion of the MU-MIMO to supporting CoMP covering both ICIC and joint multi sites MU-MIMO.

### 4.23 Use case 23: Sharing Non-RT RIC Data with the Core

#### 4.23.1 Background Information

As networks have become more and more complicated, the need has arisen for complex analytics, being available in close to real time, so as to facilitate automated network "control loops" through entities such as the Non-RT RIC and the Near-RT RIC. The need for such analytics is not, however, restricted to the RAN but also exists in the 5G Core. Thus, 3GPP has defined the NWDAF function in the 3GPP 5GC architecture [i.23].

The NWDAF uses standard interfaces to deliver analytics to other 5G Core network functions, which use those analytics to optimize their network processing. This is a similar pattern that is seen with the Non-RT RIC, which uses a standard interface (i.e. A1-EI) to deliver analytics to 5G RAN network functions (i.e. Near-RT RIC instances), which use those analytics to optimize network processing.

3GPP defines a plethora of optional functionality that an NWDAF can provide, but fundamental to the NWDAF is a functionality 3GPP defines as the Analytics Logical Function (AnLF). AnLF is the basic ability, perhaps using AIML, to produce analytics information content useful to 5GC NFs, which includes determining what source data content is necessary to do so. It is this AnLF functionality that the NWDAF function holds in common with the Non-RT RIC.

So that any 5G Core NF can semantically interpret the analytic content produced by any NWDAF, 3GPP has defined the specific "Analytics Ids" associated with each NWDAF-producible analytic type, as well as the data structures associated with each Analytic Id. It is worth noting that O-RAN has resisted similarly defining the data structures into which all rApp analytics should be carried so as to allow for innovation.

rApps can produce analytics for multiple data consumers, such as other rApps, Near-RT RIC/xApps, general external entities (e.g. an Application Function, AF). This WID seeks to add to this list the 5G Core NFs. In such a multi-consumer environment, it is a good architectural practice to avoid coupling of the data producer with the data consumer, in fact even hiding from the producer the identity of its data consumers or even whether the data consumers are single or multiple. This is, in fact, how the R1 DME is defined. In keeping with this principle, rApps will likely structure the information content of instance data they publish over R1 in a manner that is "generally useful" and not tailored to any particular data consumer. Some data consumers will be able to accept these "generally useful" data structures as is. Such might be the case, for example, for an external Application Function (AF). However, as described above, this is not the case with the 5G Core NFs.

One common pattern that will be used in various solution proposals builds on an internal functionality of the Non-RT RIC, background on which will be described here. As per [i.29], an rApp data producer should "one-time" register its output via DME services. If that rApp also intends to make its data also available via A1-EI, that rApp should also "one-time" register that same analytic output as an A1-EI Type (refer to [i.29], clause 3.3.3.2). Because A1-EI defines some standard data structures, presumably this A1-EI registration would also include the information necessary for the A1-Related Function to determine how to map any "generally useful" information content into these A1-EI data structures.

3GPP has also defined that the 5G Core NFs would use the NRF discover which specific analytics types (Analytics Ids) are available and the service endpoint through which to obtain instances of that analytic. In addition, 3GPP has defined an optional capability through which the 5G Core NF can leverage the UDM to discover available UE-level analytics and the service endpoint through which to obtain instances of such analytics for a given UE.

## 4.23.2 Motivation

There exist use cases such that the Non-RT RIC will produce, as a by-product of its normal function, analytics that would be useful to the Core. For these purposes an analytic would be considered to be of 'useful interest' to a 5G Core NF if that analytic corresponds in its semantic content to an 'Output Analytic' described in 3GPP 23.288.

For example, it would be within the scope of the Non-RT RIC to produce analytics related to UE mobility and QoS sustainability (e.g. Traffic Steering and Congestion Prediction and Management). The semantic content of these could reasonably correspond to that described in the "Output Analytic" structures of the UE Mobility and QoS Sustainability analytics types appropriate to an NWDAF (see clause 4.2.1.1 of [i.23]). The same could be said of other NWDAF analytics types as well (e.g. Slice Load Level, Network Performance, etc.)

Given that some Non-RT RIC generated analytics would be directly useful to 5GC network functions, rather than implementing such an analytic twice, assuming that such an analytic is implemented via an rApp in the Non-RT RIC, there should be a mechanism whereby such an analytic can be shared with the Core.

## 4.23.3 Proposed Solution

### 4.23.3.0 Proposed Options

Several possible solutions can be envisioned, which are treated below. A "fundamental challenge" that each solution should address is where the mapping takes place of the analytic information content from the "generally useful" structure that the rApp publishes across R1 for multiple consumers to the very specific data structures required by 5G Core. The various Options explored differ not only on the answer to this fundamental challenge, but also on whether the AnLF functionality is replicated in the Functional Architecture (i.e. in both the Non-RT RIC Function and the NWDAF Function) or shared, in the mechanism for whether and how to update the NRF/UDM, etc.

Two options, whereby the SMO/Non-RT RIC can perform the role of a 5GC NWDAF without a separate intervening NWDAF function, will be explored, and then three options, whereby the SMO/Non-RT RIC interact with a separate intervening NWDAF function, will be explored.

#### 4.23.3.1 SMO/Non-RT RIC Direct to 5GC Consumer with Façade of an NWDAF

This Option 1 approach follows the pattern defined in [i.24], clauses 6.1.1 and 6.1.2. In this approach, the SMO/Non-RT RIC exposes the analytic directly to the consuming 5GC NF, performing the role of and presenting the façade of an NWDAF relative to that network function. Thus, in this option the AnLF functionality is not duplicated but exists only in the Non-RT RIC.

Regarding the "fundamental challenge", in this Option 1 the SMO/Non-RT RIC is responsible for mapping the analytic information content from the "generally useful" data structures published across R1 into the specific data structures expected by [i.24]. The mechanism for how to do this mapping extends the established pattern defined in [i.28] for rApp registration of A1-EI types. Specifically, Option 1 would require an rApp that intends to make its data also available to 5G Core NFs to also "one-time" register its R1 data type as a [i.24] Analytic Id. This subsequent registration would also include the information necessary to map the "generally useful" information content into the data structures expected by [i.24]. The entity that would perform this mapping during subsequent runtime analytics exposure process will be described generically in the present document as the "NWDAF façade" of the SMO/Non-RT RIC.

In this option the 5GC NF discovers the correct "NWDAF" instance of interest through the UDM (for UE-related analytics) and the NRF (for non-UE related analytics). Accordingly, the SMO/Non-RT RIC, in the role of an NWDAF:

- 1) Should register with the NRF for the capabilities that it supports. (See [i.26], clause 5.2.7.1 and [i.25], clause 6.1.6.2.45.)
- 2) Can optionally register with the UDM for the UEs that it is serving or collecting data for, and for the related Analytics ID(s). (See [i.24], clause 6.1C.2)

### **Observations:**

- 1) The process through which a new analytics type would be introduced would likely entail:
  - a) One or more rApp being loaded onto the Non-RT RIC capable of generating the desired analytic content in the "generally useful" structures, communicating to the NWDAF façade how to subsequently render that content into the NWDAF structures (see [i.24], clauses 6.6 and following).
  - b) The NWDAF façade of the SMO/Non-RT RIC registering that new Analytics Id with the NRF and optionally UDM (per UE) as being available for consumption.
  - c) A new feature being loaded on a 5GC NF type capable of utilizing a particular NWDAF Analytics Id, and which uses the NRF and optionally UDM (per UE) to determine that the SMO/Non-RT RIC is the entity from which to consume that specific Analytics Id (perhaps with further discrimination based on TAC or NF mapping; refer to [i.11], clause 6.3.19).
- 2) The Non-RT RIC is within the SMO, where "M" is "management". If the SMO/Non-RT RIC presents itself to the 5GC NFs as an NWDAF, it is taking the role of a Network Function. The implications of this will be explored in the Detailed Specification.
- 3) Similar to Option 2, a DCCF/MFAF could be among the analytics consumer "5GC NFs". Because of that, the real differentiator between Option 1 and Option 2 is in which entity, if any, registers new Analytics Ids with the NRF and optionally UDM.
- 4) If there is a desire for 5GC NFs to also be able to access historical SMO/Non-RT RIC analytics, then the NWDAF façade would also need to support the optional Nnwdfaf\_DataManagement Subscribe/Notify services.
- 5) The NWDAF specification (see [i.24]) lists many other functionalities that an NWDAF can optionally support, such as supporting an external ADRF, providing training services to other 5GC NFs, Aggregation and Transfer of analytics, etc. Thus, other optional interfaces of the NWDAF (e.g. Nnwdfaf\_DataManagement, Nnwdfaf\_MLModelProvision and Nnwdfaf\_MLModelInfo) are not considered. Because the SMO/Non-RT RIC does not support such functionality at this time nor are they critical to the topic of sharing data between RAN and Core, this assumes that the NWDAF façade to the 5GC would be that of an NWDAF instance that does not support this optional functionality.

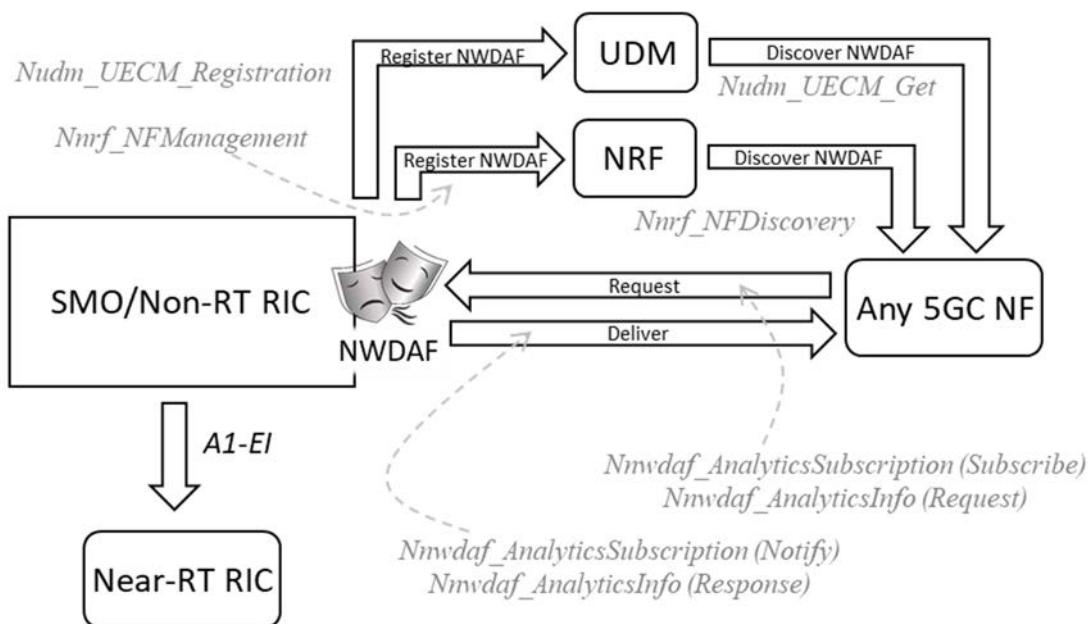
### **Benefits of this approach:**

- 1) No duplication of AnLF functionality between the SMO/Non-RT RIC and a separate NWDAF, resulting in less cost and management complexity for a carrier/operator.
- 2) No changes to the 3GPP defined interfaces, and no new interfaces need to be defined.
- 3) This Option addresses the "fundamental challenge" by following an established pattern used for delivery of rApp analytics to the Near-RT RIC via A1-EI.
- 4) No configuration changes needed (outside of SMO/Non-RT RIC) with each new analytics type being exposed to the 5GC NFs. Rather, the 5GC knows that the source of that analytic type is the SMO/Non-RT RIC through dynamic NRF updates.

### **Negatives of this approach:**

While it is not truly a negative, it is worth pointing out that this option does require support for new interfaces that are unlike any functionality that is currently associated with the SMO/Non-RT RIC, specifically the interfaces for SMO/Non-RT RIC to interact with the 5GC NRF and UDM to register its analytics types. However, such functionality can be generically useful for support of other external data consumers.

The approach taken by Option 1 is shown in figure 4.23.3.1-1 below.



**Figure 4.23.3.1-1: SMO/Non-RT RIC Direct to 5GC Consumer with Façade of an NWDAF**

#### 4.23.3.2 SMO/Non-RT RIC to DCCF/MFAF with Façade of an NWDAF

This Option 2 approach follows the pattern defined in [i.24], clause 6.1.4.2. In this approach, the SMO/Non-RT RIC shares its analytics with a DCCF/MFAF, which in turn makes those analytics available to requesting 5GC NFs. Thus, in this option the AnLF functionality is not duplicated but exists only in the Non-RT RIC.

Regarding the "fundamental challenge", this Option 2 approach also gives the SMO/Non-RT RIC responsibility for mapping the analytic information content from the "generally useful" data structures published across R1 into the specific data structures expected by [i.24]. The mechanism for how to do this is that described in Option 1, with the separate rApp data registration event for 5G Core NF data sharing.

Because in this option the SMO/Non-RT RIC performs the AnLF and data mapping function, the SMO/Non-RT RIC will present the façade of an NWDAF relative to the DCCF/MFAF, leveraging the [i.24] NWDAF services for data subscription and delivery. The DCCF/MFAF is responsible for receiving analytics requests from 5GC NFs, forwarding the corresponding request to the NWDAF façade of the SMO/Non-RT RIC, receiving from the SMO/Non-RT RIC the associated analytics information, and returning the same to the requesting 5GC NFs.

On the surface this Option 2 can appear similar, from an SMO/Non-RT RIC perspective, to Option 1 in that, in both cases, the SMO/Non-RT RIC presents the façade of an NWDAF towards the 5G Core, in this case a 5G Core DCCF/MFAF. However, in this Option 2 approach a simplification of the SMO/Non-RT RIC functionality is achieved by eliminating the ability of 5GC NFs to discover NWDAF instances via the NRF and UDM. Rather, in this Option 2 approach, the 5GC NFs can only discover via the NRF the DCCF/MFAF instances, which they would do through standard means (refer to [i.11], clause 6.3.19). Neither does the DCCF/MFAF discover NWDAF façade instances, but rather Option 2 assumes that the DCCF/MFAF knows how to reach the various NWDAF façade instances through local configuration.

**NOTE:** See [i.11], clause 6.3.13: "The NF consumers should utilize the NRF to discover NWDAF instance(s) unless NWDAF information is available by other means, e.g., locally configured on NF consumers." In this case, the DCCF/MFAF is the sole "NF consumer".

#### Observations:

- 1) The process through which a new analytics type would be introduced would likely entail:
  - a) One or more rApp being loaded onto the Non-RT RIC capable of generating the desired analytic content in the "generally useful" structures, communicating to the NWDAF façade how to subsequently render that content into the NWDAF structures (see [i.24], clauses 6.6 and following).
  - b) Configure the DCCF with a mapping from Analytics Id to the source NWDAF instance.

- c) A new feature being loaded on a 5GC NF type capable of utilizing a particular NWDAF Analytics Id, and which uses the NRF and optionally UDM (per UE) to determine that the SMO/Non-RT RIC is the entity from which to consume that specific Analytics Id (perhaps with further discrimination based on TAC or NF mapping; refer to [i.11], clause 6.3.19).
- 2) The Non-RT RIC is within the SMO, where "M" is "management". If the SMO/Non-RT RIC presents itself to the 5GC NFs as an NWDAF, it is taking the role of a Network Function. The implications of this will be explored in the Detailed Specification.
- 3) If there is a desire for the DCCF/MFAF to also be able to access and share historical SMO/Non-RT RIC analytics, then the NWDAF façade would also need to support the optional Nnwdaf\_DataManagement Subscribe/Notify services.
- 4) The NWDAF specification (see [i.24]) lists many other functionalities that an NWDAF can optionally support, such as supporting an external ADRF, providing training services to other 5GC NFs, Aggregation and Transfer of analytics, etc. Thus, other optional interfaces of the NWDAF (e.g. Nnwdaf\_DataManagement, Nnwdaf\_MLModelProvision and Nnwdaf\_MLModelInfo) are not considered. Because the SMO/Non-RT RIC does not support such functionality at this time nor are they critical to the topic of sharing data between RAN and Core, this assumes that the NWDAF façade to the 5GC would be that of an NWDAF instance that does not support this optional functionality.
- 5) Because in both Option 2 and Option 1 the SMO/Non-RT RIC assumes the façade of an NWDAF instance relative to the 5GC, Option 2 can be viewed as an interim deployment step that can be seamlessly evolved to an Option 1 approach.

#### **Benefits of this approach:**

- 1) No duplication of AnLF functionality between the SMO/Non-RT RIC and a separate NWDAF, resulting in less cost and management complexity for a carrier/operator.
- 2) No changes to the 3GPP defined interfaces, and no new interfaces need to be defined.
- 3) This Option addresses the "fundamental challenge" by following an established pattern used for delivery of rApp analytics to the Near-RT RIC via A1-EI.
- 4) Simplification, over Option 1, of the functionality of the NWDAF façade of the SMO/Non-RT RIC, eliminating the NRF and UDM registration interfaces assumed in Option 1.

#### **Negatives of this approach:**

- 1) The practical cost of this simplification over that of Option 1 is perhaps the elimination of some deployment options for SMO/Non-RT RIC instances that present an NWDAF façade, in order to avoid adding practical complexity to the DCCF/MFAF configuration. For example:
  - a) In this Option approach it is unlikely that a carrier would want to deploy many, many instances of SMO/Non-RT RIC, distinguished both by Analytics type and by geographical area of service (e.g. TAC and/or NF mapping), because that would significantly increase the complexity of the DCCF/MFAF configuration for distinguishing between these instances. Rather, because the 5GC NF's ability to discover a DCCF/MFAF instance would likely be limited to a TAC or NF mapping (refer to [i.11], clause 6.3.19), perhaps a simple one-to-many mapping between DCCF/MFAF instance and NWDAF façade would be more manageable through DCCF/MFAF configuration, the granularity of this mapping also likely being at the TAC or NF level.
  - b) In this Option approach there is no update of the UDM to indicate when UE-level analytics are being collected, so there is no ability for the 5GC NFs to leverage the UDM to determine the source for UE-level analytics. (However, as per Observation #5, Option 2 does provide a seamless path to overcome this negative over time).
- 2) Configuration changes need to be made in the DCCF/MFAF with each new analytic type being exposed to the 5GC NFs, so that the DCCF/MFAF knows that the source of that analytic is the SMO/Non-RT RIC.

The approach of Option 2 is shown in figure 4.23.3.2-1 below.

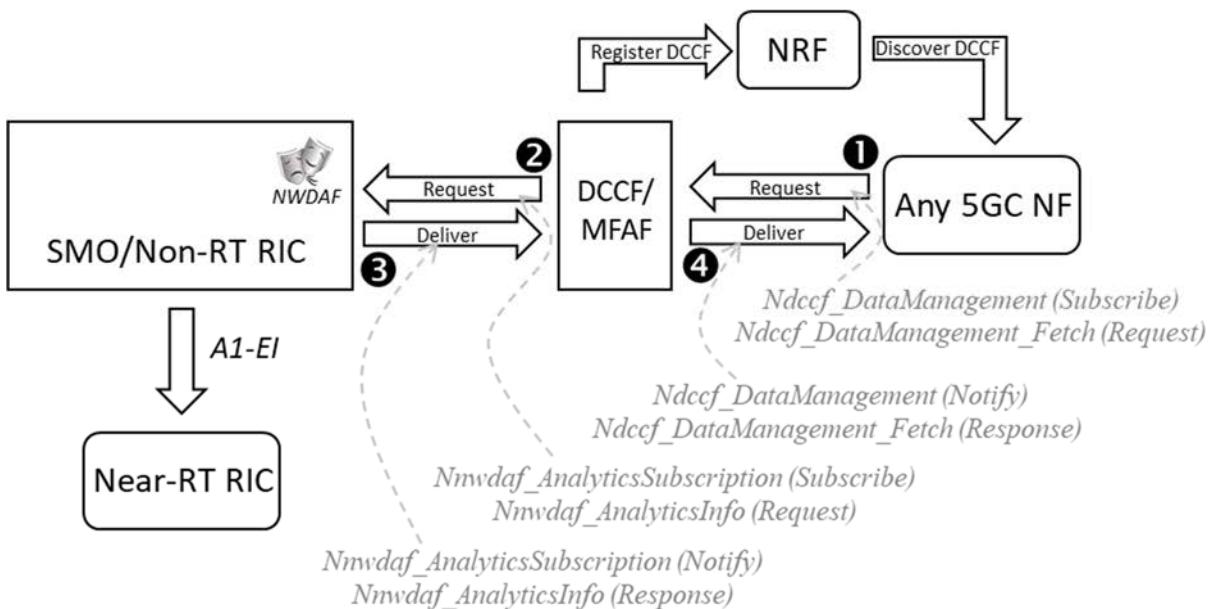


Figure 4.23.3.2-1: SMO/Non-RT RIC to DCCF/MFAF

#### 4.23.3.3 SMO/Non-RT RIC to Separate NWDAF with Façade of OA&M

This Option 3 approach follows the pattern defined in [i.24], clause 6.2.3.2. In this approach, the SMO/Non-RT RIC shares its analytics with a separate NWDAF, which in turn makes those analytics available to requesting 5GC NFs. Thus, in this option the AnLF functionality is duplicated in both the Non-RT RIC Function and the NWDAF Function.

There are multiple sub-options regarding how to approach the "fundamental challenge". These are described in "Observations" below.

In this Option 3 approach, the SMO/Non-RT RIC performs the role and presents the façade of an OAM system relative to the NWDAF (see i.28). In this option the separate NWDAF performs its normal functionality: receiving analytics requests from 5GC NFs, determining the source data that should be collected in order to fulfill that analytics request, using that source data to construct the analytics response, and forwarding the resultant analytics response to the requesting 5GC NF.

##### Observations:

- 1) The OAM interface presented as a façade towards the NWDAF (refer to [i.28]) follows the SA5 approach whereby content is specified via Network Resource Models (NRMs) which define certain standardized attributes (e.g. PLMN ID, S-NSSAI, etc.). There is also the provision for vendor-specific extensions which, of course, would not be standardized. In order for the separate NWDAF to understand the semantic content of the information being sent from the OAM façade, the SMO/Non-RT RIC will need to either map its analytics content into these standard NRM structures or map that content into the extensions. Approaches for addressing the "fundamental challenge" include:
  - a) SMO/Non-RT RIC maps the "generally useful" data structures into the existing [i.28] standardized NRM attributes, with each rApp using data type registration to communicate to the OAM façade its own mapping into these standardized NRM attributes. However, given that the standard NRM attributes were defined by 3GPP independent of the NWDAF data structures, finding such a mapping is expected to be challenging at best, and hence is not considered a viable path forward.
  - b) SMO/Non-RT RIC maps the "generally useful" data structures into a set of new [i.28] standardized NRM attributes, with each rApp using data type registration to communicate to the OAM façade its own mapping into these standardized NRM attributes. This would require O-RAN to negotiate with 3GPP changes to [i.28] to define the new NRMs, and changes to [i.24] for the NWDAF to know how to generically map these NRM structures into the structures expected by [i.24]. For a variation on this approach, see item "e" below.

- c) SMO/Non-RT RIC maps the "generally useful" data structures into a set of rApp-by-rApp defined [i.28] vendor extensions, with each rApp using data type registration to communicate to the OAM façade its own mapping into these one-off vendor extensions. With this approach, for each new analytic produced by an rApp, a negotiation would take place between the rApp vendor and the separate NWDAF vendor on how to encode its specific analytic information content into the extensions supported by [i.28]. This approach results in additional complexity which can stifle innovation and is thus not considered a viable path forward.
  - d) SMO/Non-RT RIC maps the "generally useful" data structures into a set of O-RAN defined [i.28] vendor extensions, with each rApp using data type registration to communicate to the OAM façade its own mapping into this set of O-RAN defined extensions. This would require O-RAN to negotiate with 3GPP changes to [i.28] and to [i.24] to define a standard set of [i.28] data structures into which the SMO/Non-RT RIC can map "generally useful" data structures, and the [i.24] mapping required for the NWDAF to derive the [i.24] data structures. This would, in effect, be negotiating an entirely new interface between O-RAN and 3GPP, a significant effort. A more practical variation of this approach can be found in item "e" below.
  - e) SMO/Non-RT RIC maps the "generally useful" data structures into the structures defined in [i.24], with each rApp using data type registration to communicate to the OAM façade its own mapping into this set of O-RAN defined extensions. The OAM façade would wrap those structures into the extensions supported by [i.28]. This would require modifications to [i.28] to define the standard method for wrapping these data structures, as well as a change to [i.24] such that an NWDAF would expect to find [i.24] data structures embedded in a [i.28] interface. Note that this approach has the SMO/Non-RT RIC performing some limited NWDAF functionality in that it knows how to map its analytics output into [i.24] data structures. Of the available approaches listed, this seems to be the more viable path forward.
- 2) The analysis above related to the runtime exchange of instance data between the SMO/Non-RT RIC, however the NWDAF is also responsible for registering with the NRF the analytics types that it supports. (See [i.26], clause 5.2.7.1 and [i.25] 6.1.6.2.45.) It can also optionally register with the UDM its capability to produce an Analytics ID for a specific UE. To do this, the separate NWDAF would need to discover the capabilities of the SMO/Non-RT RIC. Options include:
- a) The SMO/Non-RT RIC could present the façade of an NWDAF towards the separate NWDAF, that separate NWDAF acting as an Aggregation NWDAF. (See [i.24], clause 6.1A.). However, this approach conflicts with the fundamental premise of this Option: that SMO/Non-RT RIC presents an OAM façade to the separate NWDAF. In addition, from an SMO/Non-RT RIC perspective this approach would be no different than Option 1 and so will not be further considered.
  - b) A new interface being negotiated between O-RAN and 3GPP for automatically communicating to an NWDAF the capabilities of an OAM entity. This would be a significant effort.
  - c) Similar to Option 2, leveraging the capability of [i.11], clause 6.3.13, such that the separate NWDAF knows how to reach the various NWDAF façade instances through local configuration. This seems like the only viable path forward.
- 3) Given the above, the process through which a new analytics type would be introduced assuming the most viable paths forward as described above:
- a) One or more rApp being loaded onto the Non-RT RIC that are capable of generating the desired analytic content.
  - b) Configure the separate NWDAF with a mapping from Analytics Id to the source OAM instance so that it can register that Analytics Id with the NRF (and optionally, UDM).
  - c) A new feature being loaded on a 5GC NF type that requires access to a particular NWDAF Analytics Id, and which uses the NRF and optionally UDM (per UE) to determine that the separate NWDAF entity is that from which to consume that specific Analytics Id (perhaps with further discrimination based on TAC or NF mapping; refer to [i.11], clause 6.3.19).

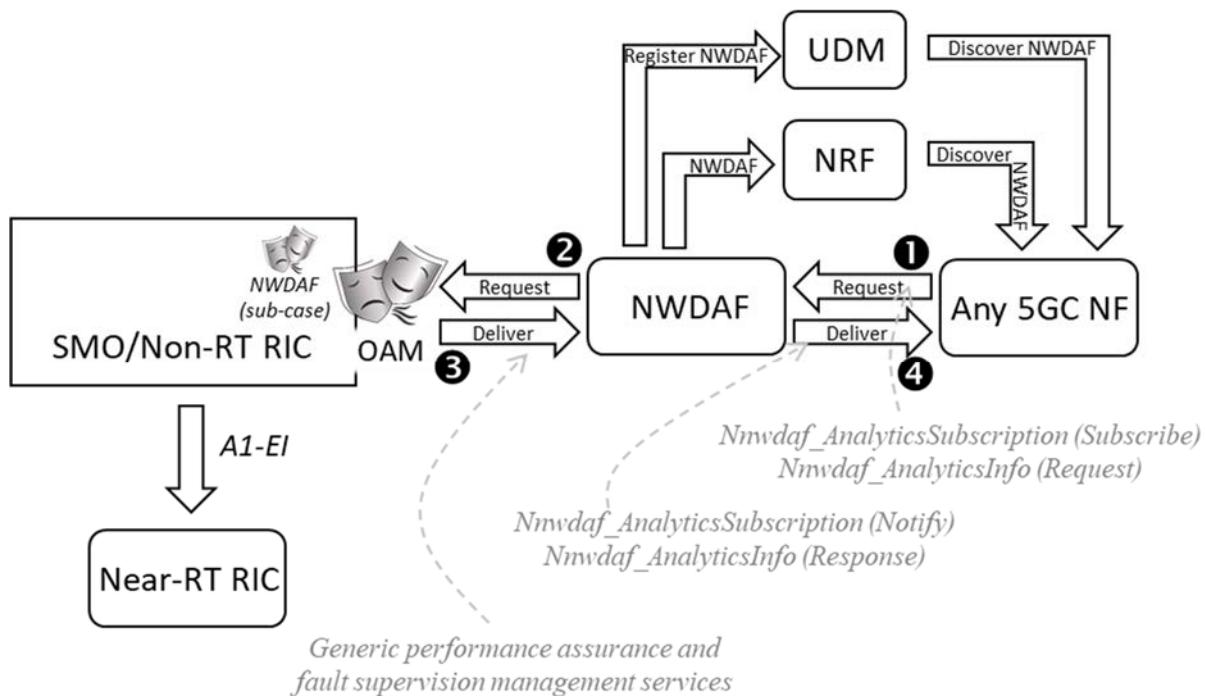
**Benefits of this approach (assuming the most viable path forward):**

- 1) Separate NWDAF determines SMO/Non-RT RIC instance for its source analytics, avoiding the need for new interfaces for SMO/Non-RT RIC to interact with the 5GC NRF and UDM. (Of course, see negative below as to how the NWDAF would discover new analytics types supported in the SMO/Non-RT RIC so as to register them.)
- 2) This Option addresses the "fundamental challenge" by following an established pattern used for delivery of rApp analytics to the Near-RT RIC via A1-EI.

**Negatives of this approach (assuming the most viable path forward):**

- 1) Duplication of AnLF functionality in both the SMO/Non-RT RIC and the separate NWDAF, resulting in less cost and management complexity for a carrier/operator.
- 2) In addition, the NWDAF AnLF function as envisioned by 3GPP is capable of performing complex trending and AI/ML functionalities to process raw source data into the desired analytics output. In this Option 3 approach, however, the NWDAF's role is only of adaptation and reformatting of the analytics from the SMO/Non-RT RIC for presentation to the 5GC NF. With this option, a service provider would be using its likely complex NWDAF function to perform what is essentially a data adaptation function.
- 3) As per observation #1e above, this Option requires changes to two separate 3GPP interface specifications: [i.25] and [i.24]. Note that other alternatives were also explored in #1 above, however each of those alternatives also involved either changes to existing or defining entirely new 3GPP interfaces entailing even more effort, or else requiring per analytic custom coding in the NWDAF stifling innovation in the Non-RT RIC, both alternatives seeming less palatable than this documented negative.
- 4) Configuration changes need to be made in the separate NWDAF with each new analytic type being exposed to the 5GC NFs, so that the NWDAF knows that the source of that analytic is the SMO/Non-RT RIC.
- 5) The practical cost of the simplification taken in Observation "2c" is perhaps the elimination of some deployment options for SMO/Non-RT RIC instances in order to avoid adding practical complexity to the DCCF/MFAF configuration. For example:
  - a) In this Option approach it is unlikely that a carrier would want to deploy many, many instances of SMO/Non-RT RIC, distinguished both by Analytics type and by geographical area of service (e.g. TAC and/or NF mapping), because that would significantly increase the complexity of the DCCF/MFAF configuration for distinguishing between these instances. Rather, because the 5GC NF's ability to discover a DCCF/MFAF instance would likely be limited to a TAC or NF mapping (refer to [i.11], clause 6.3.19), perhaps a simple one-to-many mapping between DCCF/MFAF instance and NWDAF façade would be more manageable through DCCF/MFAF configuration, the granularity of this mapping also likely being at the TAC or NF level.
  - b) In this Option approach there is no ability for the 5GC NFs to leverage the UDM to determine the source for UE-level analytics. In addition, unlike Option 2, because the SMO/Non-RT RIC presents the façade of an OAM endpoint, there is no seamless path to overcome this negative.

This approach is shown in figure 4.23.3.3-1 below.



**Figure 4.23.3.3-1: SMO/Non-RT RIC to Separate NWDAF with Façade of OA&M**

#### 4.23.3.4 SMO/Non-RT RIC to Separate NWDAF with Façade of RAN NF

This Option 4 approach also follows the pattern defined in [i.24], clause 6.2.3.2. In this approach, the SMO/Non-RT RIC shares its analytics with a separate NWDAF, which in turn makes those analytics available to requesting 5GC NFs. Thus, in this option the AnLF functionality is duplicated in both the Non-RT RIC Function and the NWDAF Function.

There are multiple sub-options regarding how to approach the "fundamental challenge" (in fact the same that were presented in Option 4). These are addressed below.

In the Option 4 approach, the SMO/Non-RT RIC performs the role and presents the façade of a RAN NF providing performance management OAM relative to the NWDAF (refer to [i.27], clause 5.1.1). In this option the separate NWDAF performs its normal functionality: receiving analytics requests from 5GC NFs, determining the source data that should be collected in order to fulfill that analytics request, using that source data to construct the analytics response, and forwarding the resultant analytics response to the requesting 5GC NF.

##### Observations:

- 1) Similar to Option 3, the RAN NF interface presented as a façade towards the NWDAF (refer to [i.27]) follows the SA5 approach whereby content is specified via Network Resource Models (NRMs) which define certain standardized attributes (e.g. PLMN ID, S-NSSAI, etc.). There is also the provision for vendor-specific extensions which, of course, would not be standardized. In order for the separate NWDAF to understand the semantic content of the information being sent from the RAN NF façade, either the SMO/Non-RT RIC will need to either map its analytics content into these standard NRM structures (see [i.24], clauses 6.2 and following), or map that content into vendor-specific extensions. The options and the viability assessment for addressing the "fundamental challenge" would be basically the same as that of Option 3.
- 2) The options and viability assessment thereof for the separate NWDAF discovering the analytics capabilities of the SMO/Non-RT RIC would be basically the same as that of Option 3.
- 3) The process through which a new analytics type would be introduced would be the same as that described in Option 3.

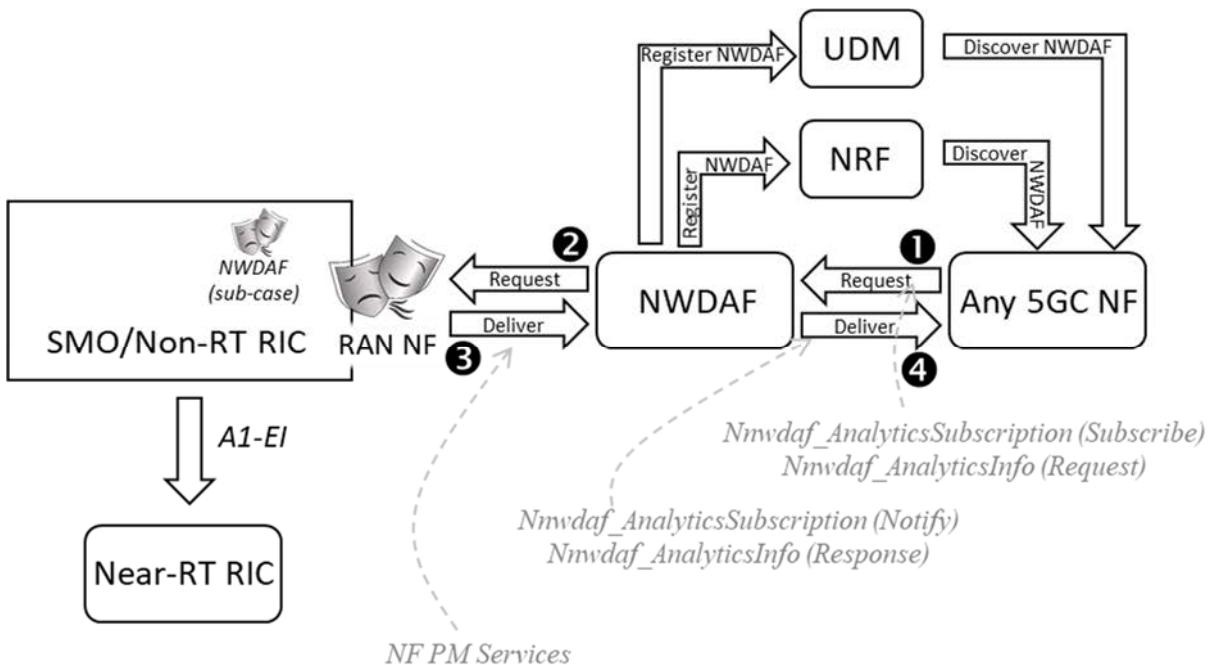
### Benefits of this approach:

- 1) Separate NWDAF determines SMO/Non-RT RIC instance for its source analytics, avoiding the need for new interfaces for SMO/Non-RT RIC to interact with the 5GC NRF and UDM. (Of course, the NWDAF should support interfaces to the NRF and UDM to register new analytics types that are supported in the SMO/Non-RT RIC.)
- 2) This Option follows the same basic rApp-to-Framework pattern already established for delivery of rApp analytics to the Near-RT RIC via A1-EI.
- 3) The same interface to the NWDAF can be leveraged by both the Non-RT and Near-RT RICs.

### Negatives of this approach:

- 1) This approach shares the same negative as Option 3 list item #1.
- 2) This approach shares the same negative as Option 3 list item #2.
- 3) This approach shares the same negative as Option 3 list item #3, substituting occurrences of reference [i.28] with reference [i.27] as appropriate.
- 4) This approach shares the same negative as Option 3 list item #4.
- 5) This approach shares the same negative as Option 3 list item #5.

This approach is shown in figure 4.23.3.4-1 below.



**Figure 4.23.3.4-1: SMO/Non-RT RIC to Separate NWDAF with Façade of RAN NF**

### 4.23.3.5 SMO/Non-RT RIC to External NWDAF with Façade of 3GPP MDA

This Option 5 is a variation of Option 3 hence it follows the pattern defined in [i.24], clause 6.2.3.2 but leverages the 3GPP MDA specification in [i.41], instead of OAM as in Option 3. Similar to Option 3, the SMO/Non-RT RIC performs the role, and presents the façade of, a 3GPP management service towards the NWDAF except that in Option 5 this is an MDA service (see [i.24] and [i.41]). Same as in Option 3, the external NWDAF performs its normal functionality: receiving analytics requests from 5GC NFs, determining the source data that should be collected in order to fulfill that analytics request, requesting the analytics from the target MDA service source data to construct the analytics response, and forwarding the resultant analytics response to the requesting 5GC NF.

**Observations:**

- 1) The main difference with Option 3 is that Option 5 uses 3GPP MDA façade instead of OAM:
  - a) MDA service exposed to the NWDAF (refer to [i.28]) follows the 3GPP SA5 approach whereby content is specified in the MDA specification (refer to [i.41]). There is also the provision for vendor-specific extensions which, of course, would not be standardized. As per reference [i.28], the NWDAF understands the 3GPP MDA semantic content (refer to [i.41]), so the SMO/Non-RT RIC can reuse the existing 3GPP MDA analytics data types. In addition, O-RAN can define new RAN analytics data types as needed, as extensions to the 3GPP MDA data types.
- 2) Approach for addressing the "fundamental challenge" include:
  - a) SMO/Non-RT RIC can reuse the existing [i.41] standardized MDA data types, with each rApp using data type registration to communicate to the MDA façade its own mapping into these standardized 3GPP MDA attributes. There are two possible alternatives:
    - i) O-RAN defines its own RAN analytics data types as per the vendor-extension mechanism enabled in 3GPP MDA, where O-RAN defines its extensions. The NWDAF [i.24] will need to understand the O-RAN standardized extensions; or
    - ii) O-RAN to request 3GPP SA5 to define a dedicated O-RAN data type extension in [i.41], to enable O-RAN to define its contents/structure.
  - b) SMO/Non-RT RIC maps the data structures into a set of O-RAN defined vendor extensions [i.28] and [i.41]], with each rApp using data type registration with DME to communicate the supported set of O-RAN defined RAN analytics extensions.
  - c) In case of rApps which do not produce RAN analytics types in a standardized format (3GPP MDA, O-RAN), then the rApps are expected to provide their own mappings into the standardized set of RAN analytics data types. The rApps use data type registration with DME to communicate this mapping so that they can be discoverable for consumption. The MDA façade would wrap those structures into the extensions supported by [i.24] and [i.41], or specified in O-RAN.
- 3) Same as for Option 3 - the NWDAF is responsible for registering with the NRF the analytics types that it supports, to enable discovery by 5GC NFs (see [i.26], clause 5.2.7.1 and [i.25] 6.1.6.2.45.). It can also optionally register with the UDM its capability to produce an Analytics ID for a specific UE. To do this, the external NWDAF would need to discover the RAN analytics capabilities of the MDA service produced by the SMO/Non-RT RIC. Options include:
  - a) Same as in Option 3, leveraging the capability of [i.11], clause 6.3.13, such that the external NWDAF knows how to reach the various MDA façade instances through local configuration.
  - b) NWDAF can discover the RAN analytics service/s of the SMO/Non-RT RIC via SME service discovery mechanisms, and it can discover the supported RAN analytics data types via DME. That implies that NWDAF has the ability to consume the SME and DME services, and this is an impact to [i.24].
- 4) Given the above, the process through which a new analytics type would be introduced assuming the most viable paths forward as described above:
  - a) One or more rApp being loaded onto the non-RT RIC that are capable of generating the desired RAN analytic content.
  - b) Configure the external NWDAF with a mapping from Analytics Id to the source MDA instance so that it can register that Analytics Id with the NRF (and optionally, UDM) on behalf of the SMO/non-RT RIC.
  - c) A new feature being loaded on a 5GC NF type that requires access to a particular NWDAF Analytics Id, and which uses the NRF and optionally UDM (per UE) to determine that the external NWDAF entity is that from which to consume that specific Analytics Id (perhaps with further discrimination based on TAC or NF mapping; refer to [i.11], clause 6.3.19).

**Benefits of this approach (assuming the most viable path forward):**

- 1) External NWDAF determines SMO/Non-RT RIC instance for its source analytics using the NWDAF consumption of MDA services as described in [28], avoiding the need for new interfaces for SMO/Non-RT RIC to interact with the 5GC NRF and UDM.
- 2) It offers a single/common solution to provide both the RAN analytics data, such as the UE location predictions described in this use case, as well as RAN management data analytics.
- 3) New RAN analytics can be flexibly added in non-RT RIC as needed, via e.g. rApps.
- 4) Dynamic service and data discovery by NWDAF: the MDA service and the data types produced by the SMO/non-RT RIC can be discovered using the service discovery via SME, and respectively the data discovery via DME.

**Negatives of this approach (assuming the most viable path forward):**

- 1) If a dynamic discovery of the MDA service and RAN analytics types supported by the SMO/non-RT RIC is desired, then NWDAF needs to be capable to consume the SME and DME services. This is an impact to [i.24].
- 2) If the static method is used, then configuration changes need to be made in the external NWDAF with each new analytic type being exposed to the 5GC NFs, so that the NWDAF knows that the MDA source of that analytic is the SMO/Non-RT RIC.

NOTE: As the static/configuration approach is the one provided in 3GPP [i.24], this is a common negative in most options, as in all cases there is some level of configuration needed in NWDAF to recognize different sources of data.

#### 4.23.4 Benefits of O-RAN Architecture

The Non-RT RIC will be capable of generating analytics that have general usefulness to the Core as well. Thus, it is beneficial to overall efficiency that such analytics produced by the Non-RT RIC can be shared with the Core, rather than requiring a separate NWDAF to also generate them. Even more so, there are some analytics that the RAN is in a better position to generate than the Core, such as UE Mobility predictions. This is true irrespective of which specific solution approach is used to enable such sharing.

### 4.24 Use case 24: Industrial vision SLA Assurance

#### 4.24.1 Background Information

Industrial vision is an image recognition technology used for automatic inspection, work piece processing and assembly automation, as well as the monitoring and control of production process by using machines to replace human eyes for various measurements and judgments. The main feature of industrial vision system is that it needs to collect the images of the products tested or processed in the area with dense production lines, and then transmit them to the vision server for detection and feedback the results.

#### 4.24.2 Motivation

5G pre-scheduling technology can greatly reduce the transmission delay and improve the transmission reliability. However, the traditional static pre-scheduling mechanism does not fully consider the actual industrial production line uplink transmission characteristics. Regardless of whether the terminal has uplink transmission data, it always allocates air interface resources according to the existing configuration, resulting in a large waste of resources, and a significant decline in the actual bearable traffic of the cell, which is difficult to deploy commercially.

Therefore, the commercial pre-scheduling mechanism generally adopts a dynamic pre-scheduling fashion, in which scheduling start time can be dynamically adjusted according to the actual business characteristics. In order to realize dynamic pre-scheduling, the network side needs to perceive the arrival of uplink transmission data on the terminal side, prepare uplink air interface resources for this terminal in advance, and configure the pre-scheduling adaptively.

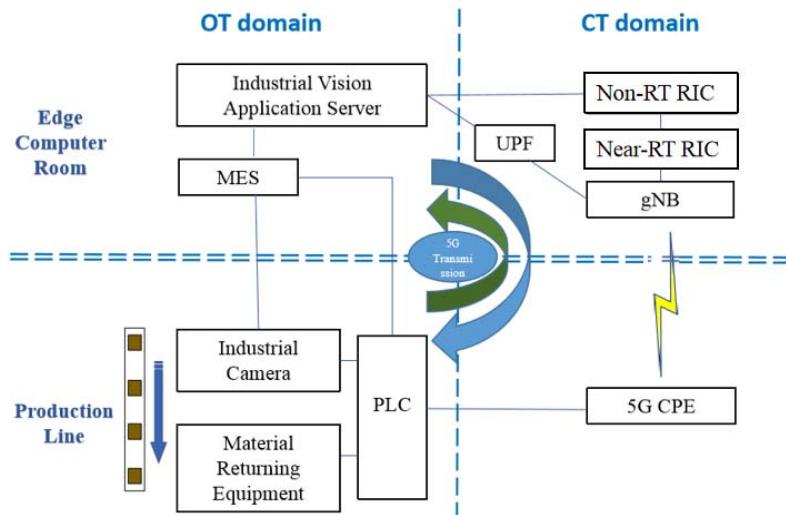
O-RAN based architecture will enable such adaptive pre-scheduling parameters configuration to be implemented and help to realize the high efficiency of industrial vision.

### 4.24.3 Proposed Solution

Industrial vision use case mainly involves the Non-RT RIC, Near-RT RIC, E2 nodes in the O-RAN architecture, and industrial vision server (MES), industrial camera in the production system. The data transmission to Non-RT RIC approach includes, but not limited to, below method: Non-RAN data about production line/camera information and data transmission information is transmitted to Non-RT RIC as Enrichment Information [i.30] and [i.31].

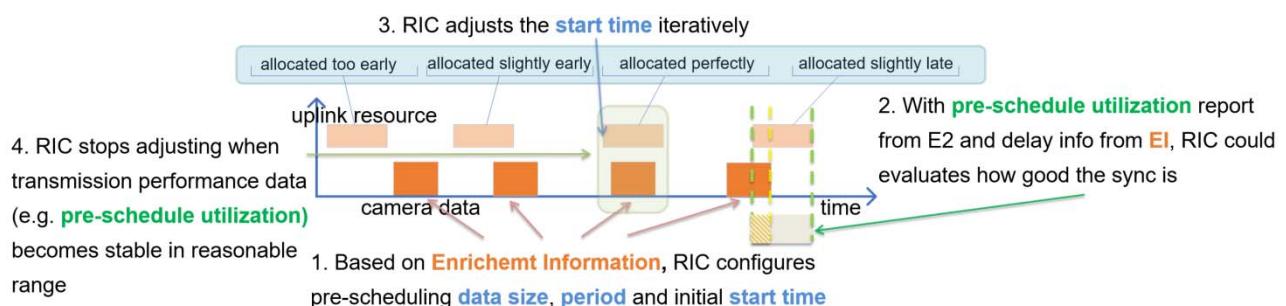
NOTE: The Enrichment Information mentioned above can be also obtained from the 5G core network.

With Enrichment Information from Industry server application server, base station can sync the camera data uplink resource allocation with work piece arrival time through dynamic pre-scheduling, as shown in figure 4.24.3-1.



**Figure 4.24.3-1: Industrial Vision SLA Assurance overview**

In order to realize the adaptive configuration of pre-scheduling process, the Near-RT RIC needs to adaptively calculate three pre-scheduling parameters matching with the industrial vision periodical business characteristics, including pre-scheduling data size, pre-scheduling period and pre-scheduling start time, so as to configure the E2 node via E2 control/policy service with those pre-scheduling parameters.



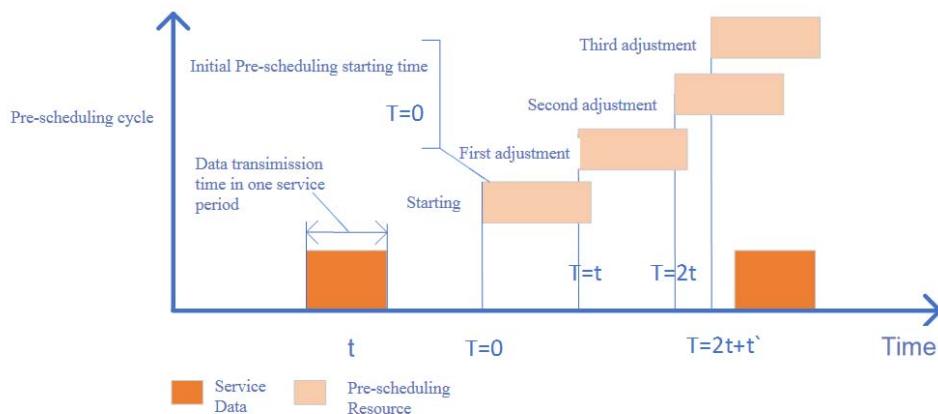
**Figure 4.24.3-2: Industrial Vision Dynamic pre-scheduling**

The dynamic pre-scheduling workflow is shown in figure 4.24.3-2. At the start-up stage of production line, pre-scheduling data size and pre-scheduling period are calculated by the Near-RT RIC based on production line speed and configured industrial camera attributes, such as image color, pixel, etc. These production line data can be imported as enrichment information to the Non-RT RIC, and then delivered to the Near-RT RIC. Then, the Near-RT RIC decides an initial pre-scheduling start time, which can be absolute time or relative time depending on the Near-RT RIC's capability to synchronize time with E2 nodes. The pre-scheduling start time is the time at which the first pre-scheduling of the whole iterative process should start.

The Near-RT RIC then configures the E2 node via E2 control/policy service with pre-scheduling data size, pre-scheduling period and pre-scheduling start time parameters. Note that the configured parameters mentioned above only serve as scheduling recommendations to E2 Nodes. Actual PRB scheduling depends on many other factors not captured by this use case. E2 Node might for instance supersede Near-RT RIC's recommendation in case high priority delay critical data needs to be scheduled.

After initial configuration, the Near-RT RIC adaptively adjusts the pre-scheduling start time based on the service data transmission information, including pre-scheduling time-domain resource utilization and image data transmission delay related enrichment information. The pre-scheduling time-domain resource utilization is to be provided by E2 node, but it is currently not defined in the E2 interface specification, and it can be defined as the ratio of the time-domain resource of the data to be scheduled to the total pre-scheduled time-domain resource. The image data transmission delay related enrichment information is provided to the Non-RT RIC by industrial vision server (MES) which hosts industrial vision application, and is then delivered to the Near-RT RIC.

The initial pre-scheduling starting time can be set to  $T=0$ , and then iteratively adjusted based on the service data transmission information, as shown in figure 4.24.3-3, until the service data transmission information converges to a reasonable range.



**Figure 4.24.3-3: Pre-scheduling starting time adjustment illustration**

#### 4.24.4 Benefits of O-RAN Architecture

Near-RT RIC based recommendation of pre-scheduling configuration towards the E2 Node might reduce the latency and improve the timing accuracy of image data transmission and might reduce scheduling related uplink signal overhead. The proposed solution to support industrial vision business assurance requires features provided by O-RAN architecture to satisfy the requirements via interoperable E2 interfaces and including input from external systems and external applications.

### 4.25 Use case 25: Non-Public Network (NPN) RAN-Sharing via Midhaul for Multi-Operator Coverage

#### 4.25.1 Background Information

This clause describes use-case analysis for a new O-RAN use-case to support Non-Public Network (NPN) RAN-Sharing via Midhaul for Multi-Operator Coverage. This NPN RAN-sharing solution is different from traditional Multi-operator RAN (MORAN) and Multi-operator Core (MOCN) RAN-sharing models used in the past. In this use-case, Non-Public Network (NPN) RAN is shared to extend mobile-operator coverage inside private-spaces where Mobile-operators have poor coverage and need partnership– like buildings or Airports, Ports, hospitals, hotels etc. in which better mobile-operator network coverage is desired.

This use-case ensures security & privacy for NPN-Network traffic by keeping private-traffic separate & private from public-traffic. This use-case enables NPN-RAN sharing to extend MNO-network coverage for multiple MNOs in a way that solves MNO security concern of connecting to un-trusted NPN-network.

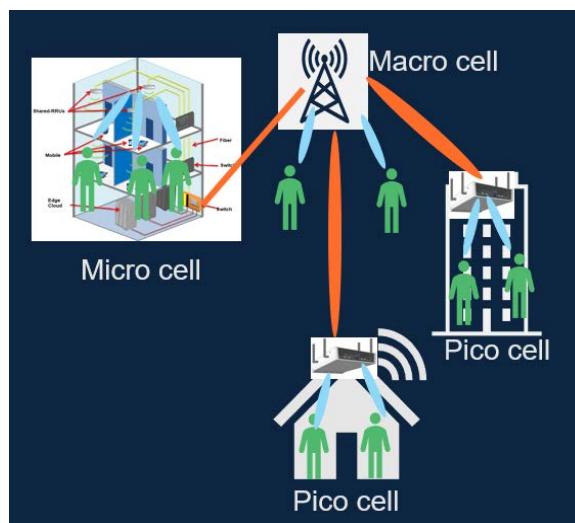
## 4.25.2 Motivation

The motivation is to create new shared 5G-RAN use-case in which Non-Public Network (NPN) 5G-RAN is shared to extend wireless coverage for multiple partner operators. O-RAN can accelerate the deployment of compliant 5G RAN sharing solutions by taking advantage of O-RAN efforts to create multi-vendor interoperability profiles for 3GPP-defined F1-interface. Refer to 3GPP specifications for RAN-sharing, F1-interface and IAB-interface [i.33], [i.34], [i.35], [i.36], [i.37], [i.38], [i.39], [i.40].

Main motivation of this use-case is to enable RAN-sharing of NPN Micro-cell or Pico-cells deployments within buildings or private-spaces to extend wireless Macro-cell coverage of multiple mobile operators within private campuses. NPN owner gets additional benefit of getting better MNO-coverage inside campus to achieve better ROI on NPN-investment. MNO gets benefit of extending MNO-coverage without spending CAPEX for it.

This use-case solves in-building MNO coverage problem in a cost-effective & secure manner to keep Private traffic private for NPN 5G-RAN deployment and enable secure NPN-RAN sharing for MNO-network traffic. This use-case supports midhaul connection across different security domains i.e. NPN RAN managed by NPN-RAN operator & partner RAN managed by a different mobile operator. RAN-Sharing via Midhaul enables MNO's macro-cell coverage extension without affecting back-haul traffic impacting mobile-operator 5G-core workload. In a way, it creates hierarchical cell structure to enhance capacity and coverage of Mobile-operator macro cell (see figure 4.25.2-1). In this use-case, 3GPP based Midhaul (F1) and Integrated Access Backhaul (IAB) interface are used to seamlessly integrate NPN-owned Micro cells & Pico-cells with multiple MNO's Macro-cells to extend coverage and create Macro-cell hierarchical cell structure with heterogenous ownership.

NPN RAN-sharing via Midhaul avoids use of roaming model for NPN network integration with public networks. This avoids complex network roaming based integration of MNO-network with several NPN-networks partners.



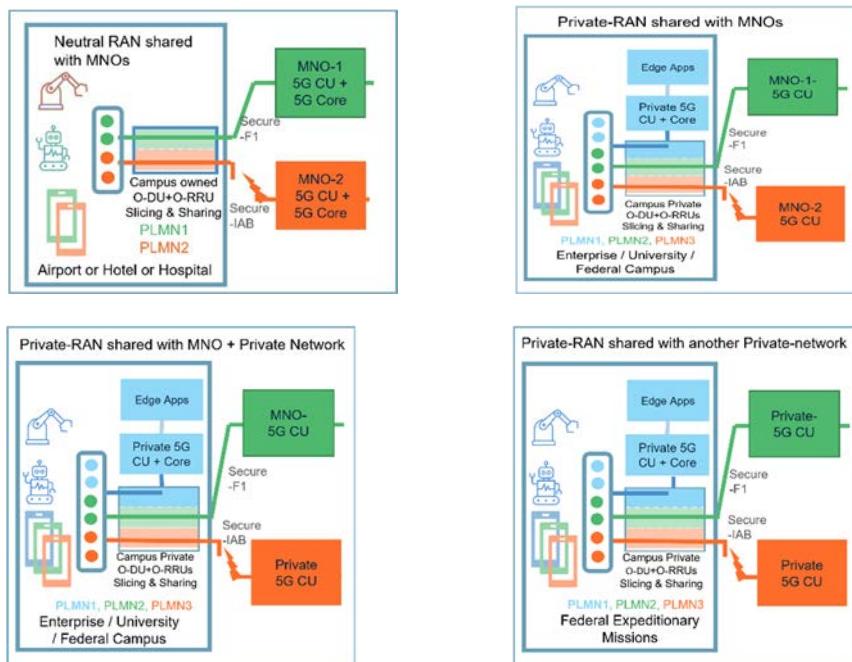
**Figure 4.25.2-1: Hierarchical Cell Structure to extend macro-cell coverage**

In this use-case, NPN-RAN O-DU is shared with multiple external partners via Midhaul connection. NPN O-DU connects to multiple PLMN Partner O-CU-CPs with each PLMN representing traffic from one Mobile Operator. So, NPN O-DU is connected to only one Partner O-CU-CP for one PLMN as defined in 3GPP TS 38.401 [i.38]. This use-case enabled several NPN RAN-sharing partnership models to extend wireless coverage of multiple operators at lower CAPEX and OPEX:-

- 1) Neutral-RAN sharing via Midhaul with Multiple MNOs (see case 1 in figure 4.25.2-2). In this case, NPN RAN is installed by campus owner (Hotel, Airport or Hospital). One NPN-RAN O-DU is installed and shared with multiple Mobile operators to extend wireless coverage of Mobile operators within Campus at lower cost instead of each Mobile operator deploying their own Network Extension solution (example- DAS/Relay)
- 2) NPN-RAN shared with Multiple MNOs (see case 2 in figure 4.25.2-2). In this deployment option, NPN-RAN is used for private CAMPUS -traffic and multiple Mobile-Operator's traffic. NPN O-DU RAN-sharing partitions based on agreements with Mobile Operators to allow offload of Mobile-operator traffic to NPN 5G-Network and extend Mobile-Operator cell-coverage inside Campus.

- 3) NPN-RAN shared with MNO & another NPN-network (see case 3 in figure 4.25.2-2). In this deployment option, NPN-RAN is shared with a mix of Mobile-Operator (public 5G-network) and another Private 5G-Network. This option can be used to connect multiple private 5G-Network partners together with seamless mobility as well as extend public 5G-network coverage inside campus.
- 4) NPN-RAN sharing with multiple NPN-Networks (see case 4 in figure 4.25.2-2). In this deployment model, several NPN 5G-networks are connected to each other for seamless mobility via RAN-Sharing. Each NPN 5G-network is represented by their own separate PLMN with multiple campus locations support PLMNs from all partners sharing the RAN.

Proposed solution for this use-cases should work for all partnership models mentioned above.



**Figure 4.25.2-2: Different NPN RAN-Sharing partnership scenarios**

## 4.25.3 Proposed Solution

### 4.25.3.0 NPN RAN-Sharing solution

Since NPN RAN-sharing is proposed for Micro-cell / Pico-cells deployed in private spaces, NPN RAN-sharing solution for this use-cases needs to be simple to manage & deploy to enable adoption. Hence, NPN RAN-sharing solution is developed using simple business agreement templates built using RAN-sharing policy & frameworks that are created via regulatory or open MNO-partnership process.

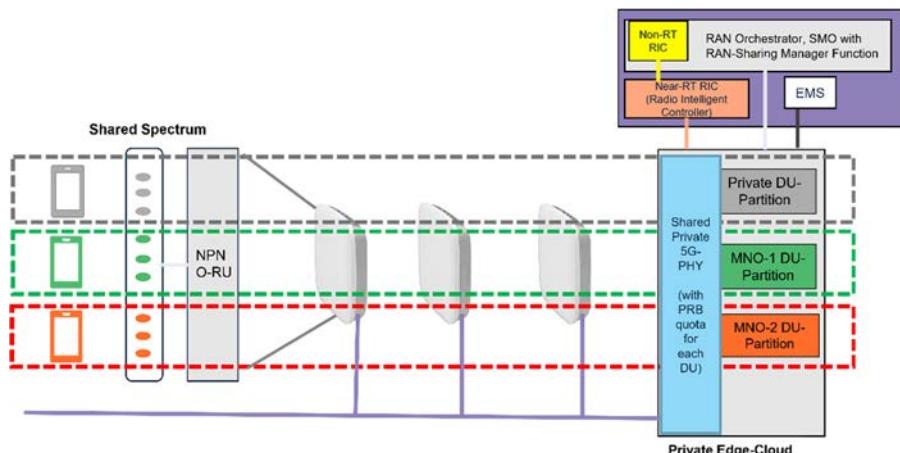
Though proposed solution can be used for NPN RAN-sharing using multi-band O-RUs to maintain different spectrum for different operator traffic (similar to MORAN), proposed solution for this use-case can also be used for NPN-RAN sharing using one spectrum band license for NPN-RAN to simplify NPN-RAN deployment and share it to support multiple Mobile-Operators.

### 4.25.3.1 RAN Sharing to support multiple operators

#### 4.25.3.1.0 NPN RAN-Sharing solution using O-RAN architecture

See figure 4.25.3.1.0-1 for NPN RAN-sharing solution using O-RAN architecture. Within O-RAN SMO & RAN-orchestrator, NPN-RAN sharing Manager function is defined to divide NPN-RAN into RAN-partitions with quota of NPN-RAN Physical Resource Blocks (PRB) assigned to each Partner as per business agreement with partner operator.

Typically, shared NPN-RAN should consist of shared NPN-RAN O-DU connected to network of O-RU & antennas covering entire campus or NPN-owner coverage area. This solution does not propose any change in NPN O-RU. NPN-RAN sharing is enabled using COTS / existing O-RU. O-DU creates RAN-sharing partitions by allocation of PHY Resource Blocks (PRBs) to each partner PLMN and connects to multiple partner's O-CUs to extend wireless coverage for multiple partners. To ensure security of partner's remote O-CU, multi-vendor O-RAN F1 interface is used to connect to local O-CU Proxy of remote partner O-CU. To support NPN RAN-sharing, NPN O-DU is aware of multiple PLMNs and maps PHY-resources to different MAC-layers associated with different PLMN in uplink as well as downlink.



**Figure 4.25.3.1.0-1: NPN RAN Sharing model to support multiple mobile operators**

#### 4.25.3.1.1 Spectrum Partitioning to support multiple operators

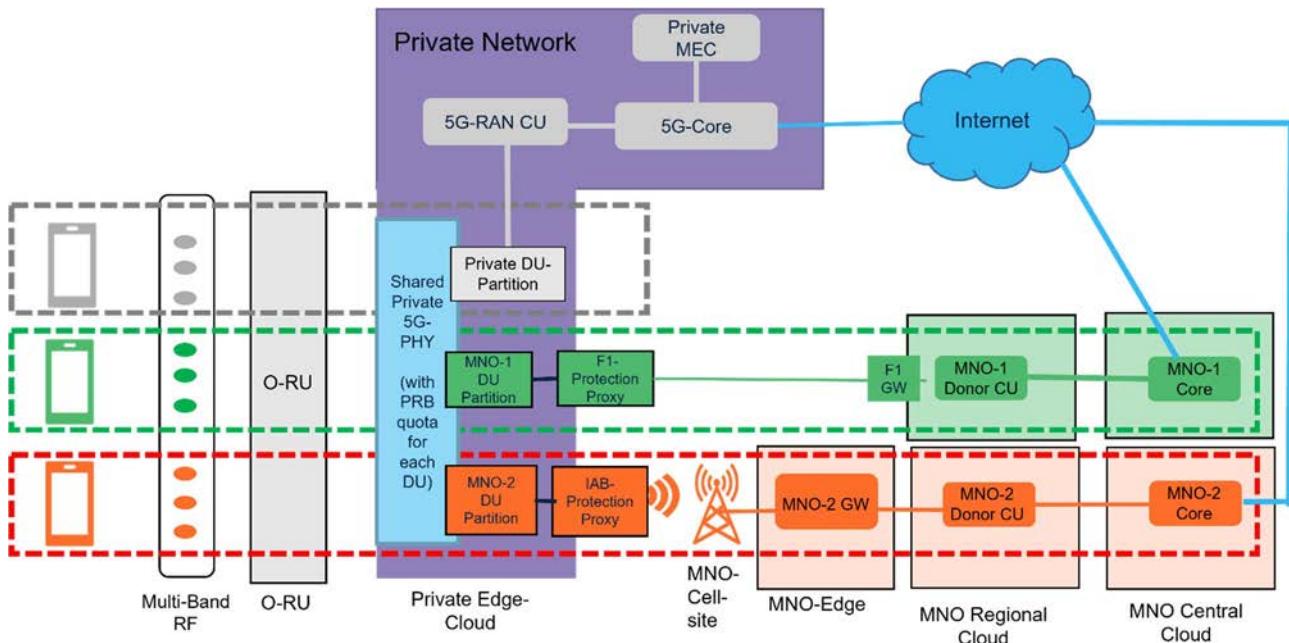
NPN-RAN spectrum license is partitioned (divided) to support multiple Mobile-Operators as shown in figure 4.25.3.1.0-1. NPN O-RUs deploy licensed spectrum through-out NPN-coverage area for NPN 5G-Network deployment. All NPN O-RUs are connected to a common NPN O-DU. NPN O-DU creates RAN-partition subnets by allocating a quota of PHY-Layer resource blocks of 5G PHY-layer to enable RAN-sharing.

Spectrum licensing for NPN network deployment is beyond the scope of O-RAN standard and not covered here. Different countries should provide spectrum licensing for NPN Network deployment via different policy & business agreements frameworks. To be cost-effective, NPN-owner is expected to deploy 5G-Network using one spectrum license and divide available spectrum as partitions to support different PLMNs of multiple partner mobile-operators. Hence, NPN-spectrum licensing options needs to be defined in a way that all operators within a country support NPN-spectrum for their devices.

#### 4.25.3.2 RAN-Sharing via Midhaul Architecture

##### 4.25.3.2.0 NPN RAN-sharing via Midhaul

See figure 4.25.3.2.0-1 below for RAN-sharing via Midhaul architecture in which NPN-RAN is shared with multiple mobile operators via Midhaul connection from NPN Shared O-DU to Partner-Operator's O-CU.



**Figure 4.25.3.2.0-1: NPN RAN-sharing Via Midhaul**

See figure 4.25.3.2.0-1 for NPN 5G RAN-sharing architecture to support multiple operator's traffic. As shown in NPN RAN-sharing diagram in figure 4.25.3.2.0-1, NPN RAN is shared and connected to multiple partner operator's O-CU via Midhaul. This use-case supports midhaul connection across different security domains i.e. NPN RAN managed by NPN-RAN operator & partner O-CU managed by a different mobile operator. This is an untrusted Midhaul connection in which MNO's O-CU connects to shared NPN RAN-partition. To resolve security concerns, Partner Operator creates proprietary wired F1-interface or wireless IAB-interface to remote gNB-CU connection to Partner-owned O-CU proxy which supports 3GPP-defined F1-interface locally in NPN-network and connects to NPN O-CU to extend Partner operator's wireless coverage. For NPN-RAN sharing, NPN O-DU connects to partner's local O-CU Proxy using O-RAN defined F1-iteroperability profiles and does not change 3GPP defined F1-interface.

#### 4.25.3.2.1 O-RAN impacts of Proposed Solution

The proposed solution is designed in a way that existing Private 5G-RAN deployments can be upgraded for RAN-sharing leveraging existing investments as much as possible. Proposed solution impacts to different O-RAN components are as follows:

- 1) **O-RU + Antennas** - Proposed solution does not propose any changes to COTS O-RU that is used to deploy Private-RAN or NPN-RAN solution. This enables the upgrade path of existing 5G-network to enable RAN-sharing without changing O-RUs which represent major component of RAN-deployment investment.
- 2) **Fronthaul Transport Network** - Higher Transport capacity will be needed if a higher spectrum bandwidth is used at O-RU. However, if deployed NPN-RAN does not require higher spectrum resources then fronthaul transport will not have any impacts due to proposed solution.
- 3) **Fronthaul Interface** - Fronthaul interface is not impacted by this use-case if Open fronthaul mandated by O-RAN is used. If Split 7.2 (adopted by O-RAN) is used then PHY common channels can support multiple RAN-sharing PLMNs and user-data resources blocks are allocated to RAN-sharing partners as PRB resource-block quota as shown in figure 4.25.3.2.0-1 above. NPN-RAN fronthaul interface is shared for NPN-RAN sharing and support traffic from all partners using common fronthaul interface from NPN-RAN O-DU to NPN-RAN O-RUs as shown in figure 4.25.3.1.0-1.

- 4) **O-DU impact** - O-DU architecture has impacts as identified in figure 4.25.3.2.0-1. In this use-case, NPN-RAN O-DU is shared with multiple external operators. Different partnership options for NPN RAN-Sharing are shown in figure 4.25.2-2. To support partner traffic, NPN O-DU needs to be modified to support F1-traffic coming from multiple PLMNs. To support NPN RAN-sharing, NPN O-DU needs to support multiple PLMNs and maps PHY-resources to different MAC-layers associated with different PLMN in uplink as well as downlink. This requires separate MAC-layer processing for different PLMN traffic and mapping of PHY-layer PRBs correctly to different PLMNs in uplink and downlink traffic. Legacy NPN-RANs can support NPN-RAN sharing and add wireless coverage of partner 5G-networks within campus using existing NPN-RAN deployments.
- 5) **Midhaul Connection** - This use-case supports connection between NPN-network and MNO-network with different security and ownership domains. This untrusted midhaul connection across different security domains is solved in a way that resolves security concerns of NPN-network and MNO-network. To resolve this security concern, this use-case supports F1-interoperability between NPN-Network and MNO-network via a proxy O-CU as a local midhaul-connection to NPN O-DU. This proxy O-CU connects to remove MNO O-CU as a proprietary wired-F1 or wireless-IAB interface. In a way, this security issue is solved by MNO offering a midhaul-proxy which MNO connects locally to shared NPN O-DU to ensure security for MNO-network. NPN O-DU creates RAN-sharing partitions and creates midhaul-interface connection for partner networks to connect and use shared O-DU resources.
- 6) **O-CU impacts** - NPN O-CU connects locally to shared NPN O-DU using RAN-transport endpoint at NPN O-DU created for partner midhaul-connection. Partner MNO-networks enables connection to remote external untrusted NPN-RAN via MNO-owned O-CU proxy interface. MNO's O-CU connection to remote O-CU proxy is developed as a proprietary interface by MNO to resolve MNO's security concerns of connecting to external O-DU. Partner O-CU Proxy is installed locally in NPN-campus and connects locally to NPN 5G-Network.

#### 4.25.3.2.2 Midhaul Connection

##### 4.25.3.2.2.0 Impact on Interfaces

NPN-RAN O-DU connects locally with partner O-CU proxy which supports midhaul-interface inter-operability between O-DU and O-CU as defined by O-RAN alliance F1 interoperability profile. O-RAN F1-interoperability profile can have some impacts which should be worked / added as part of O-RAN WG-5 work-group effort.

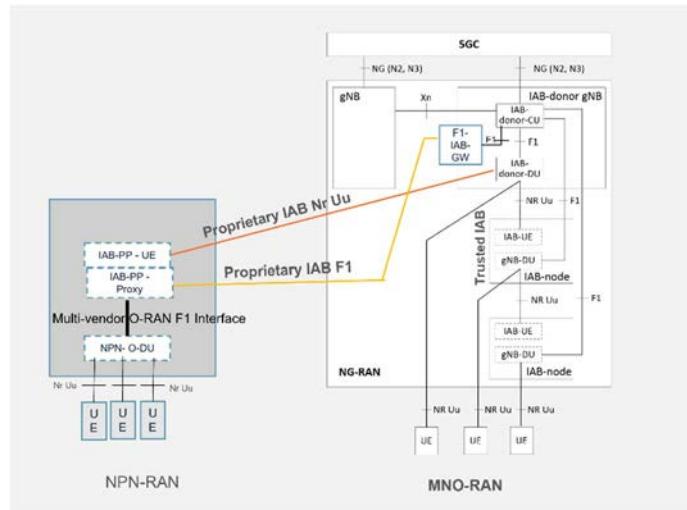
##### 4.25.3.2.2.1 Local Proxy based Wired F1 Interface

Local proxy-based wired F1 interface is defined as a wired connection between O-CU F1-GW and partner's local F1-protection proxy (F1-PP) installed in NPN campus. This is a MNO-proprietary wired midhaul interface between F1-GW of the MNO's donor CU and MNO-provided F1-protection-proxy (F1-PP) installed in NPN Network.

#### 4.25.3.2.2.2

#### Local Proxy based Wireless IAB Interface

Local Proxy-based wireless Integrated Access Backhaul (IAB) interface is defined as a wireless connection between MNO's donor IAB-GW and MNO's local O-DU IAB-protection proxy (F1-PP) installed in NPN campus. This is a MNO-proprietary wireless midhaul connection of the NPN-hosted IAB-Node and MNO's IAB donor. See figure 4.25.3.2.2.2-1 for untrusted IAB Interface diagram.



**Figure 4.25.3.2.2.2-1: Proprietary IAB Interface**

#### 4.25.3.2.3

#### E2-Interface changes

O-RAN defines a common E2 interface between co-hosted / trusted O-DU + O-CU and Near-RT RIC. Since untrusted Midhaul interface between partner O-CU and NPN O-DU is proposed as part of this use-case, E2-Interface changes should be proposed to meet Near-RT RIC requirements for all mobile operators involved in NPN RAN-sharing. Proposed changes should be contributed to O-RAN WG3 workgroup for adoption.

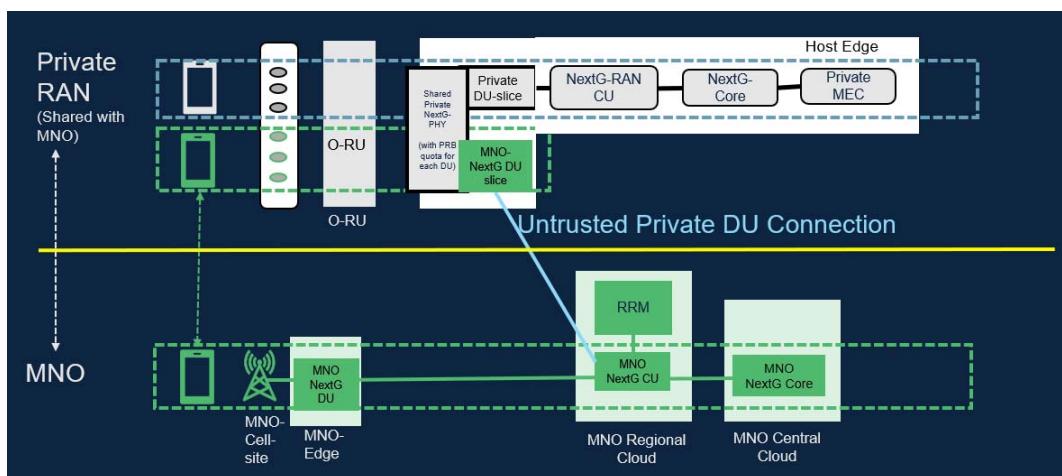
#### 4.25.3.2.4

#### O1-Interface changes

O-RAN defines a common O1 interface between co-hosted / trusted O-DU + O-CU and Service Management Orchestrator (SMO). Since untrusted Midhaul interface between partner O-CU and NPN O-DU is proposed as part of this use-case, O1-Interface changes should be proposed to meet SMO-requirements of NPN RAN-sharing independently from MNO's SMO & RIC-orchestrator. Proposed changes should be contributed to O-RAN WG2 workgroup for adoption.

#### 4.25.3.2.5

#### Mobility

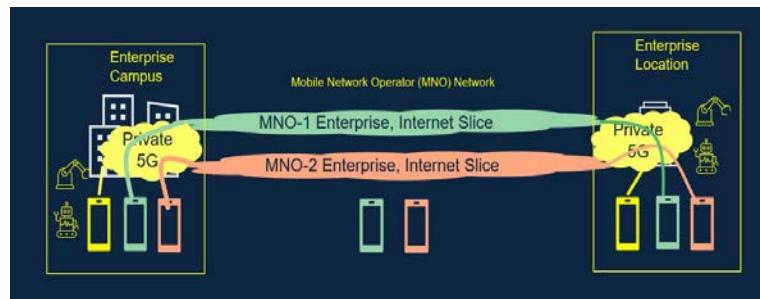


**Figure 4.25.3.2.5-1: Seamless UE Mobility**

Radio Resource Control (RRC) and Radio Resource Management (RRM) functions are supported in donor/partner CU-processing. So, all Radio Measurement reports from UE are processed by the MNO-CU to enable smooth mobility between the shared NPN-RAN and MNO's 5G-RAN (see figure 4.25.3.2.5-1). Since the shared NPN RAN is part of the MNO's hierarchical-cell connected to the MNO's macro-cell, NPN O-DU Cell-ID integration within the MNO's O-CU can be supported via an untrusted NPN O-DU-registration process.

#### 4.25.4 Benefits of O-RAN Architecture

One major benefit of this NPN RAN-sharing use-case is to enable Enterprise NPN-owner and MNO partnership to offer end-to-end QoS for MNO's Enterprise users/services inside Enterprise campuses where MNO wireless coverage is poor. See figure 4.25.4-1 below for Enterprise RAN-sharing scenario in Private 5G-RAN is shared to support multiple mobile-operators traffic.



**Figure 4.25.4-1: End-to-end QoS mobile service for Enterprise users**

## Annex A (informative): Change history

Date	Version	Information about changes
2019.10.17	01.00	<ul style="list-style-type: none"> <li>First version with initial set of use cases. Describes the use cases at a very high level, emphasizing how the use case is enabled by O-RAN architecture along with basic input data expectations and resulting actions.</li> </ul>
2020.03.11	02.00	<ul style="list-style-type: none"> <li>Updates to the existing use cases.</li> <li>Addition of QoS based resource optimization use case.</li> <li>Addition of RAN slice SLA assurance use case.</li> </ul>
2020.07.16	03.00	<ul style="list-style-type: none"> <li>Addition of multi-vendor slices use case.</li> <li>Addition of dynamic spectrum sharing use case.</li> <li>Addition of long-term NSSI optimization use case.</li> <li>Addition of local positioning use case.</li> <li>Enhancements to massive MIMO use case.</li> <li>Enhancements to traffic steering use case.</li> </ul>
2020.11.02	04.00	<ul style="list-style-type: none"> <li>Addition of SU-MU MIMO TM optimization use case.</li> <li>Addition of signaling storm protection use case.</li> <li>Addition of cell congestion prediction use case.</li> <li>Addition of L2 IIoT optimization use case.</li> </ul>
2021.03.13	05.00	<ul style="list-style-type: none"> <li>Addition of BBU pooling to achieve RAN elasticity use case.</li> <li>Updates to massive MIMO optimization use case.</li> </ul>
2021.07.19	06.00	<ul style="list-style-type: none"> <li>Additions to QoE optimization use case.</li> </ul>
2021.11.23	07.00	<ul style="list-style-type: none"> <li>Addition of integrated SON function use case.</li> <li>Updates to NSSI resource optimization use case.</li> </ul>
2022.04.04	08.00	<ul style="list-style-type: none"> <li>Addition of Shared O-RU use case.</li> <li>Addition of energy saving use case.</li> <li>Updates to QoE optimization use case based on RAN analytics information exposure feature.</li> </ul>
2022.08.01	09.00	<ul style="list-style-type: none"> <li>Addition of MU-MIMO optimization use case.</li> <li>Additions to Shared O-RU use case for multi-vendor slices use case impact.</li> </ul>
2022.11.18	10.00	<ul style="list-style-type: none"> <li>Addition of RAN-core data sharing use case.</li> <li>Minor editorial updates.</li> </ul>
2023.03.24	11.00	<ul style="list-style-type: none"> <li>Addition of industrial vision SLA assurance use case.</li> <li>Minor updates to some of the existing use cases.</li> </ul>
2023.07.27	12.00	<ul style="list-style-type: none"> <li>Updates for O-RAN Drafting Rules (ODR) compliancy.</li> <li>Addition of RAN sharing via midhaul use case.</li> <li>Addition of an option for Non-RT RIC data sharing with core use case.</li> </ul>

---

## History

<b>Document history</b>		
V12.0.0	April 2025	Publication