O-RAN.WG1.NESUC-R003-v02.00

*Technical Report*

**O-RAN Work Group 1 (Use Cases and Overall Architecture) Network Energy Saving Use Cases Technical Report**

Copyright © 2023 by the O-RAN ALLIANCE e.V.

The copying or incorporation into any other work of part or all of the material available in this document in any form without the prior written permission of O-RAN ALLIANCE e.V. is prohibited, save that you may print or download extracts of the material of this document for your personal use, or copy the material of this document for the purpose of sending to individual third parties for their information provided that you acknowledge O-RAN ALLIANCE as the source of the material and that you inform the third party that these conditions apply to them and that they must comply with them.

O-RAN ALLIANCE e.V., Buschkauler Weg 27, 53347 Alfter, Germany

Contents

[Foreword 5](#_TOC_250102)

[Modal verbs terminology 5](#_TOC_250101)

1. [Scope 6](#_TOC_250100)
2. [References 6](#_TOC_250099)
   1. [Informative references 6](#_TOC_250098)
3. [Definition of terms, symbols and abbreviations 7](#_TOC_250097)
   1. [Terms 7](#_TOC_250096)
   2. [Symbols 7](#_TOC_250095)
   3. [Abbreviations 7](#_TOC_250094)
4. [Objectives and Requirements 8](#_TOC_250093)
   1. [Objectives 8](#_TOC_250092)
   2. [Requirements 9](#_TOC_250091)
5. [Carrier and Cell Switch Off/On 10](#_TOC_250090)
   1. [Problem Statement, Solution and Value Proposition 10](#_TOC_250089)
   2. [Architecture/Deployment Options 11](#_TOC_250088)
      1. [Option 1: Non-RT RIC Deployment 11](#_TOC_250087)
         1. [Description and UML Diagram 11](#_TOC_250086)
         2. [O-RAN Entity Roles 14](#_TOC_250085)
         3. [Input/Output Data Requirements 14](#_TOC_250084)
            1. Summary 14
            2. Detailed Input Requirements 15
            3. Detailed Output Requirements 18
      2. [Option 2: Near-RT RIC Deployment 19](#_TOC_250083)
         1. [Description and UML Diagram 19](#_TOC_250082)
         2. [O-RAN Entity Roles 22](#_TOC_250081)
         3. [Input/Output Data Requirements 22](#_TOC_250080)
            1. Summary 22
            2. Detailed Input Requirements 24
            3. Detailed Output Requirements 27
   3. [Impact Analysis on O-RAN Work Groups 27](#_TOC_250079)
   4. [Relation and Impact on 3GPP Specifications 29](#_TOC_250078)
      1. [Relation to 3GPP RAN Specifications 29](#_TOC_250077)
      2. [Ongoing Work in 3GPP Rel.18 RAN 29](#_TOC_250076)
      3. [Impact on 3GPP RAN Specifications 30](#_TOC_250075)
      4. [Relation to 3GPP System Architecture Specifications 30](#_TOC_250074)
      5. [Ongoing Work in 3GPP Rel.18 System Architecture 31](#_TOC_250073)
      6. [Impact on 3GPP System Architecture Specifications 31](#_TOC_250072)
   5. [Gain Analysis 31](#_TOC_250071)
      1. [Cell and Carrier Switch Off/On Energy Saving for 4T4R O-RU 33](#_TOC_250070)
      2. [Cell and Carrier Switch Off/On Energy Saving for 64T64R O-RU 34](#_TOC_250069)
   6. [Feasibility Analysis 35](#_TOC_250068)
6. [RF Channel Reconfiguration 36](#_TOC_250067)
   1. [Problem Statement, Solution and Value Proposition 36](#_TOC_250066)
   2. [Architecture/Deployment Options 37](#_TOC_250065)
      1. [Option 1: Non-RT RIC Deployment 37](#_TOC_250064)
         1. [Description and UML Diagram 38](#_TOC_250063)
         2. [O-RAN Entity Roles 41](#_TOC_250062)
         3. [Input/Output Data Requirements 41](#_TOC_250061)
            1. Summary 41
            2. Detailed Input Requirements 43
            3. Detailed Output Requirements 47
      2. [Option 2: Near-RT RIC Deployment 47](#_TOC_250060)
         1. [Description and UML Diagram 48](#_TOC_250059)
         2. [O-RAN Entity Roles 51](#_TOC_250058)
         3. [Input/Output Data Requirements 52](#_TOC_250057)
            1. Summary 52
            2. Detailed Input Requirements 53
            3. Detailed Output Requirements 56
   3. [Impact Analysis on O-RAN Work Groups 56](#_TOC_250056)
   4. [Relation and Impact on 3GPP Specifications 58](#_TOC_250055)
   5. [Gain Analysis 58](#_TOC_250054)
      1. [RF Channel Reconfiguration ES Gain Analysis for 4T4R O-RU 58](#_TOC_250053)
      2. [RF Channel Reconfiguration ES Gain Analysis for 64T64R O-RU 60](#_TOC_250052)
   6. [Feasibility Analysis 61](#_TOC_250051)
      1. [Continuous operation during RF Channel Reconfiguration 61](#_TOC_250050)
      2. [Impact on Coverage 61](#_TOC_250049)
      3. [Impact and Relation to UE specific Base Station Algorithms 61](#_TOC_250048)
      4. [Limited O-RU / O-DU Capabilities 61](#_TOC_250047)
7. [Advanced Sleep Mode Selection 62](#_TOC_250046)
   1. [Problem Statement, Solution and Value Proposition 62](#_TOC_250045)
   2. [Architecture/Deployment Options 64](#_TOC_250044)
      1. [Option 1: Training and Inference in Non-RT RIC 64](#_TOC_250043)
         1. [Description and UML Diagram 65](#_TOC_250042)
         2. [O-RAN Entity Roles 69](#_TOC_250041)
         3. [Void 69](#_TOC_250040)
      2. [Option 2: Training in Non-RT RIC and Inference in Near-RT RIC 69](#_TOC_250039)
         1. [Description and UML Diagram 70](#_TOC_250038)
         2. [O-RAN Entity Roles 74](#_TOC_250037)
         3. [Void 74](#_TOC_250036)
   3. [Impact Analysis on O-RAN Work Groups 75](#_TOC_250035)
   4. [Relation and Impact on 3GPP Specifications 76](#_TOC_250034)
   5. [Void 76](#_TOC_250033)
   6. [Feasibility Analysis 76](#_TOC_250032)
      1. [Impact to Continuous Operation during Advance Sleep Modes 76](#_TOC_250031)
      2. [Impact to Coverage 76](#_TOC_250030)
      3. [Impact and Relation to Vendor Specific Scheduling Algorithms 77](#_TOC_250029)
      4. [Limited O-RU/O-DU Capabilities 77](#_TOC_250028)
8. [O-Cloud Resource Energy Saving Mode 78](#_TOC_250027)
   1. [Sub Use Case 1: O-Cloud Node Shutdown 78](#_TOC_250026)
      1. [Problem Statement, Solution and Value Proposition 78](#_TOC_250025)
      2. [Architecture/Deployment Option 79](#_TOC_250024)
         1. [Description and UML Diagram 81](#_TOC_250023)
         2. [O-RAN Entity Roles 85](#_TOC_250022)
         3. [Void 85](#_TOC_250021)
      3. [Impact Analysis on O-RAN Work Groups 86](#_TOC_250020)
      4. [Relation and Impact on 3GPP Specifications 86](#_TOC_250019)
      5. [Void 86](#_TOC_250018)
      6. [Feasibility Analysis 86](#_TOC_250017)
         1. [Service Continuity during NF relocation 86](#_TOC_250016)
         2. [Pooling vs. Scaling Gains 86](#_TOC_250015)
         3. [Start-up Time for Scale Out Operation 86](#_TOC_250014)
   2. [Sub Use Case 2: O-Cloud CPU Energy Saving Mode 87](#_TOC_250013)
      1. [Problem Statement, Solution and Value Proposition 87](#_TOC_250012)
      2. [Architecture/Deployment Option 88](#_TOC_250011)
         1. [Description and UML Diagram 89](#_TOC_250010)
         2. [O-RAN Entity Roles 92](#_TOC_250009)
         3. [Void 92](#_TOC_250008)
      3. [Impact Analysis on O-RAN Work Groups 92](#_TOC_250007)
      4. [Relation and Impact on 3GPP Specifications 92](#_TOC_250006)
      5. [Void 93](#_TOC_250005)
      6. [Feasibility Analysis 93](#_TOC_250004)
         1. [Not to Restrict Fast CPU Energy Saving Mode Switching 93](#_TOC_250003)
9. [Summary and Conclusion 94](#_TOC_250002)

Annex A (Informative): Design Principles for NES Features 96

Annex B (Informative): Load profile and O-RU functional blocks 98

Annex C (Informative): O1 interface principles 100

Annex D (Informative): Examples of Advanced Sleep Modes 104

[Revision history 105](#_TOC_250001)

[History 105](#_TOC_250000)

# Foreword

This Technical Report (TR) has been produced by O-RAN Alliance.

# Modal verbs terminology

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

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

# Scope

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

Version x.y.z where:

1. the first digit is incremented for all changes of substance, i.e. technical enhancements, corrections, updates, etc. (the initial approved document will have x=01).
2. the second digit is incremented when editorial only changes have been incorporated in the document.
3. the third digit included only in working versions of the document indicating incremental changes during the editing process.

The present document provides a technical report on Network Energy Saving use cases.

# References

## 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, O-RAN 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.

1. 3GPP TR 21.905, Vocabulary for 3GPP Specifications
2. ETSI ES 203 228, Environmental Engineering (EE); Assessment of mobile network energy efficiency
3. O-RAN Technical Specification, O-RAN Architecture Description
4. O-RAN Technical Specification, Non-RT RIC Architecture
5. O-RAN Technical Report, Cloud Architecture and Deployment Scenarios for O-RAN Virtualized RAN
6. O-RAN Technical Specification, Use Cases Detailed Specification
7. 3GPP TS 28.554, Management and orchestration; 5G end to end Key Performance Indicators (KPI)
8. 3GPP TS 28.310, Management and orchestration; Energy efficiency of 5G
9. 3GPP TS 38.300, NR; NR and NG-RAN Overall Description, Stage-2
10. ETSI ES 202 706-1, Metrics and measurement method for energy efficiency of wireless access network equipment; Part 1: Power consumption - static measurement method
11. 3GPP TS 28.541, Management and orchestration; 5G Network Resource Model (NRM); Stage 2 and stage 3
12. 3GPP TS 28.552, Management and orchestration; 5G performance measurements
13. 3GPP TS 32.425, Telecommunication management; Performance Management (PM); Performance measurements Evolved Universal Terrestrial Radio Access Network (E-UTRAN)
14. 3GPP TS 32.451, Telecommunication management; Key Performance Indicators (KPI) for Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Requirements
15. 3GPP TS 32.551, Telecommunication management; Energy Saving Management (ESM); Concepts and requirements
16. 3GPP TS 38.331, NR; Radio Resource Control (RRC); Protocol specification
17. O-RAN Technical Specification, O-RAN Outdoor Macrocell Hardware Architecture and Requirements (FR1)
18. O-RAN Technical Specification, O-RAN Operations and Maintenance Interface Specification
19. 3GPP TR 38.864, Study on network energy savings for NR
20. O-RAN Technical Specification, O2 Interface General Aspects and Principles

# Definition of terms, symbols and abbreviations

## Terms

For the purposes of the present document, the terms and definitions given in 3GPP TR 21.905 [1] and the following apply. A term defined in the present document takes precedence over the definition of the same term, if any, in 3GPP TR 21.905 [1].

**Energy Efficiency:** relation between the useful output and energy/power consumption as defined in ETSI ES 203 228 [2]

**Energy Consumption:** integral of power consumption over time as defined in ETSI ES 202 706-1 [10]

## Symbols

No symbol is defined in this TR.

## Abbreviations

For the purposes of this document, the abbreviations given in 3GPP TR 21.905 [1], 3GPP TS 38.300 [9], ETSI ES 203 228 [2], O-RAN.WG1.O-RAN-Architecture-Description [3], O-RAN.WG6.CADS [5] apply.

# Objectives and Requirements

## Objectives

This Technical Report captures the outcome of the WG1 UCTG Network Energy Saving pre-normative phase. The objectives of the pre-normative phase are as follows:

* study and investigate all required counters/KPI’s for monitoring and reporting of Energy Consumption and Energy Efficiency in real time for all O-RAN defined nodes such as O-RU, O-DU, O-CU as well as 3GPP defined logical components of network such as Cell, Carrier, gNB,
* study requirements, key issues, proposed solutions, benefits of the Energy Saving proposals,
* study potential impact and required enhancements to O-RAN interfaces such as E2, O1, A1, FH M- plane, FH CUS-Plane, R1, and Near-RT RIC API,
* study potential impact and required enhancements on data models of all O-RAN entities,
* identify the any possible impact on Non-RT RIC architecture, Near-RT RIC architecture and AI/ML workflow.

The use cases studied in the NES pre-normative phase are:

* Carrier and Cell Switch Off/On
* RF Channel Reconfiguration Off/On
* Advanced Sleep Mode Selection
* O-Cloud Resource Energy Saving Mode

Algorithms discussed and analyzed as part of the Network ES pre-normative phase will be examples only and will not be part of any specification as outcome of this pre-normative phase or subsequent work items.

For each use case solution proposal, the detailed objectives are:

* Review, evaluate applicability of, and select from existing deployment alternatives (Non-RT and/or Near-RT RIC) and AI/ML deployment scenarios and document respective findings:
  + Evaluate energy savings gains based on a E2 Node/O-RU/PNF/VNF energy consumption model using evaluation metrics/KPIs and including assessment of impact on network and user performance.
  + Study on a per sub-use case basis potential impact and enhancements on O-RAN interfaces (e.g. O1, O2, A1, E2, O-FH) in terms of required input data and output configuration data.
  + Review existing counters, KPIs, and data models as specified/studied in 3GPP.
  + Study potential enhancements on existing counters, KPIs, and data models or define new counters, KPIs, and data models of all involved O-RAN entities.

NES use cases should reuse the existing 3GPP measurements and measurement reporting for input data as well as the existing 3GPP configuration and provisioning management for the output data as much as possible.

If new measurements or configuration parameters are essential to support new NES use cases, then these should preferably be based on parameters, variables, definitions and procedures that are already used in the 3GPP/O- RAN specifications. The current standardization approach (specifying the new measurement in 3GPP specifications and referring to it in O-RAN specifications) should be prioritized over inventing new standardization approaches.

## Requirements

This TR shall capture Energy consumption and Energy Efficiency related counters and KPIs including the following:

##### O-RU Specific KPIs

- Energy Efficiency and power consumption KPIs provided by real-time metering.

##### O-CU/O-DU Hardware & Software / O-Cloud Software & Platform KPIs

* O-CU/O-DU hardware (e.g., CPU, accelerators, NIC cards, fans and power supply, etc.) shall have the capability to measure and report power consumption values to the O-Cloud. The O-Cloud software shall be able to collect measurement data at the hardware component level (e.g., CPU, NIC, accelerator card, fans and power supply, etc.) and provide power, energy and environmental (PEE) parameters/KPIs. O-Cloud shall be able to report EE through O2 interface to the SMO or through North-Bound Interfaces (NBI) to external tools.
* The O-Cloud software shall be able to provide PEE parameters/KPIs at the workload level (e.g., pod, O-CU/O-DU CNF, etc.), as well as for the O-Cloud platform software components themselves. O- Cloud shall be able to report energy efficiency through O2 interface to the SMO or through NBI to external tools.
* O-CU and O-DU shall provide CNF level energy efficiency counters/KPIs (e.g., power consumption, Traffic load, data volume, throughput), which shall be reported through O1 interface to the SMO or through NBI to external tools.

# Carrier and Cell Switch Off/On

## Problem Statement, Solution and Value Proposition

Mobile networks often utilize multiple frequency layers (carriers) to cover the same service area. At low load, i.e. when the expected traffic volume is lower than a fixed threshold, ES can be achieved by switching off one or more carriers or entire cells without impairing the user experience. UEs previously served by the carrier or cell will be offloaded by the E2 Node to a new target carrier or cell prior to the switch off.

However, the switch off/on decisions are not a trivial task. There are conflicting targets between system performance and energy savings. Other carriers and/or cells will have to serve the additional traffic and traffic is changing over time. E2 Nodes support a number of techniques effecting energy consumption which might also be load dependent. While energy savings for the switched off carrier and/or cell is maximized, the overall energy consumption of the network might even increase.

Carrier and cell switch off/on control by the Non-RT or Near-RT RIC can consider overall network energy efficiency instead of local optimization. The switch off/on decision can optionally be made by an AI/ML model within the inference host, deployed at the Non-RT RIC (deployment option 1) or at the Near-RT RIC (deployment option 2) to further improve decision making. Among others, the AI/ML models' functionality may include prediction of future traffic, user mobility, and resource usage and may also predict expected energy efficiency enhancements, resource usage, and network performance for different ES optimization states.

Before switching off/on carrier(s) and/or cell(s), the E2 Node may need to perform some preparation actions for off switching (e.g. check ongoing emergency calls and warning messages, to enable, disable, modify Carrier Aggregation and/or Dual Connectivity, to trigger HO traffic and UEs from cells/carriers to other cells or carriers, informing neighbour nodes via X2/Xn interface etc.) as well as for on switching (e.g., cell probing, informing neighbour nodes via X2/Xn interface etc.).

## Architecture/Deployment Options

### Option 1: Non-RT RIC Deployment

In option 1, decision making, potentially including AI/ML Model Training and Inference, is done at the Non- RT RIC.

#### Description and UML Diagram

##### Table 5.2.1.1-1: Carrier and Cell Switch Off/On: AI/ML inference via Non-RT RIC

|  |  |  |
| --- | --- | --- |
| **Use Case Stage** | **Evolution / Specification** | **<<Uses>> Related use** |
| Goal | Enable Carrier and Cell switch off/on Energy Saving functions in the Network by means of configuration parameter change and Actions controlled by Non-RT RIC and allow for AI/ML-based solutions. |  |
| Actors and Roles | Non-RT RIC acting as inference host for Energy Savings decision making.  E2 Node and O-RU are the subject of action for configuration enforcement. |  |
| Assumptions | O1 interface connectivity is established.  Open FH M-Plane interface is established between E2 Node and O-RU and/or SMO and O-RU directly.  Network is operational.  Non-RT RIC has knowledge about overlapping carriers/cells and the  coverage of those carriers/cells (e.g., which carrier/cell is a coverage layer and which is a capacity layer). |  |
| Pre-conditions | The operator has set the targets for Energy Saving functions in the Non- RT RIC. |  |
| Begins when | Operator enables the optimization functions for carrier and cell switch off/on Energy Saving functions and E2 Node and O-RU become  operational. |  |
| Step 1 (M) | SMO initiates specific measurement data collection request towards E2 Node and O-RU (via E2 Node and O-FH) for AI/ML model training and inference. |  |
| Step 2 (M) | E2 Node and O-RU send the configured measurement data to SMO periodically or event based. |  |
| Step 3 (M) | Non-RT RIC retrieves the collected measurement data for processing. |  |
| Step 4 (M) | Non-RT RIC trains the AI/ML models with the collected data. Trained AI/ML models are deployed, configured, and activated.  Non-RT RIC constantly monitors   1. performance and energy consumption of the E2 Node(s) 2. energy consumption of O-RU(s) |  |
| Step 5 (M) | Based on the AI/ML inference the Non-RT RIC may request the SMO to configure E2 Node to prepare and execute cell or carrier switch off/on. |  |
| Step 6 (M) | SMO instructs E2 Node via O1 interface to perform the received request(s) from the Non-RT RIC. O-RU is informed about the updated O-RU configuration via Open FH M-Plane interface by E2 Node. E2  Node will notify SMO once cell or carrier switch off/on is completed. |  |
| Step 7 (M) | Non-RT RIC continuously analyzes performance of AI/ML model. If  energy saving objectives are not achieved, it may decide to initiate fallback mechanism, and/or AI/ML model update or retraining. |  |
| Ends when | E2 Node becomes non-operational or when the operator disables the optimization functions for Energy Saving. |  |
| Exceptions | None. |  |
| Post Conditions | Non-RT RIC continues closed-loop monitoring of Energy Saving function at E2 Node and O-RU.  E2 Node(s) and O-RU(s) operate using the newly deployed parameters/models and state (off/on). |  |

@startuml Skin rose

skinparam defaultFontSize 12 autonumber

Box "Service Management & \n Orchestration Framework" #gold

Participant "Collection & Control" as SMO Participant "Non-RT RIC" as NRTRIC

End box

Box "O-RAN Nodes" #lightpink Participant "Near-RT RIC" as RTRIC Participant "E2-Nodes" as E2NODES Participant "O-RUs" as ORUs

End box

group Data Collection autonumber 1.1

SMO -> E2NODES : <<O1>> Data collection request for Energy Saving E2NODES -> ORUs : <<FH>> Data collection request for Energy Saving

autonumber 2.1

ORUs -> E2NODES : <<FH>> Measurement Data Collection for Energy Saving E2NODES -> SMO : <<O1>> Measurement Data Collection for Energy Saving autonumber 3

SMO -> NRTRIC : Data retrieval

end

group Data Analysis Training and Inference autonumber 4.1

NRTRIC -> NRTRIC : Performance and Energy Consumption Monitoring \n(E2 Node(s) & O- RU(s))

NRTRIC -> NRTRIC : AI/ML Model training and inference

end

autonumber 5

group Actor Decision Making

NRTRIC -> SMO : Request to prepare and execute \nfor carrier(s) and cell(s) switch off/on

autonumber 6.1

SMO <-> E2NODES : <<O1>> E2-Node parameter configurations and recommendation for carrier(s) and cell(s) switch off/on, \nmessage response and/or notification of execution.

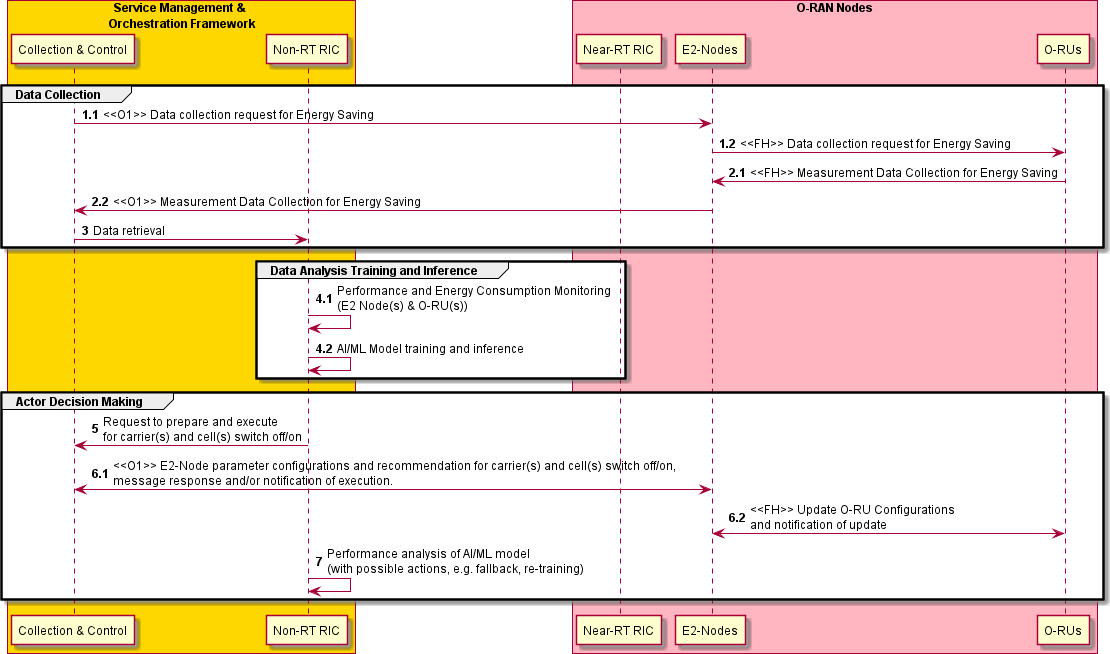
E2NODES <-> ORUs : <<FH>> Update O-RU Configurations \nand notification of update

autonumber 7

NRTRIC -> NRTRIC : Performance analysis of AI/ML model \n(with possible actions, e.g. fallback, re-training)

end

@enduml



##### Figure 5.2.1.1-1: Carrier and Cell Switch Off/On flow diagram: AI/ML inference via Non-RT RIC

#### O-RAN Entity Roles

1. Non-RT RIC
   1. Collect configurations, performance indicators and measurement reports (e.g., cell load related information and traffic information, EE/EC measurement reports, geolocation information) from SMO, E2 Nodes and O-RUs (forwarded by SMO), for the purpose of decision making, optionally using training and inference of AI/ML models that assist such EE/ES functions.
   2. (Optionally) Trigger EE/ES AI/ML model training/retraining.
   3. (Optionally) Deploy, update, configure EE/ES AI/ML models in Non-RT RIC.
   4. Analyze the received data from SMO, E2 Nodes, and O-RUs to determine EE/ES optimization (i.e. if carriers or cells are recommended to be switched off/on), optionally using AI/ML models
   5. Signal updated configurations for EE/ES optimization to E2 Nodes (O-CU) via R1/O1.
2. E2 Node
   1. Report cell configuration, performance indicators and measurement reports (e.g., cell load related information and traffic information, EE/EC measurement reports) to SMO via O1 interface.
   2. Perform actions required for EE/ES optimization
      * e.g. check ongoing emergency calls and warning messages, perform some preparation actions for Off Switching (e.g., to enable, disable, modify Carrier Aggregation and/or Dual Connectivity, to trigger HO traffic and UEs from cells/carriers to other cells or carriers, informing neighbour nodes via X2/Xn interface etc.) as well as for On Switching (e.g., cell probing, informing neighbour nodes via X2/Xn interface etc.) and make final decision on switch off/on and notify SMO via O1 about performed actions
3. O-RU
   1. Report EC and EE related information via Open FH M-Plane interface to O-DU or alternatively to SMO directly.
   2. Support actions required to perform EE/ES optimization.
      * updated carrier configuration (i.e. activation, deactivation or sleep)

#### Input/Output Data Requirements

* + - * 1. Summary

Input Data

1. E2 Node to SMO/Non-RT RIC
   * Carrier/cell characteristics
   * EE/EC measurement reports
   * Load statistics per cell and per carrier
   * UE mobility information including cell or beam level measurements (e.g. RSRP, RSRQ, SINR)
   * Energy consumption
2. O-RU to E2 Node
   * Power consumption metrics: Mean total/per carrier power consumption, mean total/per carrier transmit power

Output Data

1. SMO to E2 Node
   * Carrier(s) and/or cell(s) recommended to be switched off/on
     + - 1. Detailed Input Requirements

Initialization:

##### Table 5.2.1.3-1: Initialization

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Input Data** | | | | | |
| **Interface** | **Source / Target** | **Name/Description** | **Units** | **Reporting Period** | **Existing /New Definitions** |
| R1 | SMO / rApp | Optimization target for Carrier/Cell Switch Off/On |  | manual or event triggered | New |
| E.g., (average or max) NG-RAN data Energy Efficiency | bit/J |  | 3GPP TS 28.554 [7] (Cl. 6.7.1) |
| O1, R1 | E2 Node (O- DU) / SMO /  rApp | Carrier/cell characteristics | - | manual or event triggered (e.g., SMO  feature activation, E2 Node startup / failure / reconfig.) | New |
| E.g., physical location, transmit direction, carrier frequency, coverage parameters, configured transmit power, beam width, coverage shape, tilt, azimuth, carrier-cell mapping, carrier-HW mapping |  |  | 3GPP TS 28.541 [11] (Cl.  4.3.6, 4.3.38, 4.3.39, 4.3.40,  4.3.74)  O-RAN.WG5.O-DU-O1 (Sec.  10), O-RAN-WG5.O-CU-O1  (Sec. 9) |

AI/ML Model Training:

##### Table 5.2.1.3-2: AI/ML Model Training

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Input Data** | | | | | |
| **Interface** | **Source / Target** | **Name/Description** | **Units** | **Reporting Period** | **Existing Definitions** |
| O1, R1 | E2 Node (O- CU) / SMO /  rApp | DL PDCP SDU Data Volume per interface (Data Volume in DL delivered from O-CU-UP to O- DU, per PLMN, per QoS level, per slice, per Interface (F1-U, Xn- U, X2-U)) | Mbit | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.3.6.2.3) |
| O1, R1 | E2 Node (O- CU) / SMO /  rApp | UL PDCP SDU Data Volume per interface (Data Volume in UL delivered to O-CU-UP from O- DU, per PLMN, per QoS level, per slice, per Interface (F1-U, Xn- U, X2-U)) | Mbit | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.3.6.2.4) |
| O1, R1 | E2 Node (O- CU) / SMO /  rApp | RSRQ measurement per SSB per cell | dB | (non-real time for training) | Measurement: 3GPP TS 28.552 [12] (Cl.  5.1.1.31) |
| O1, R1 | E2 Node (O- CU) / SMO /  rApp | RSRP measurement per SSB per cell | dBm | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.22) |
| O1, R1 | E2 Node (O- CU) / SMO /  rApp | SINR measurement per SSB per cell | dB | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.32) |
| O1, R1 | E2 Node (O- CU) / SMO /  rApp | Energy consumption | kWh | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.19.3) |
| O-FH (M-  Plane) or  O-FH (M-  Plane), O1, R1 | O-RU / O-DU  or  O-RU / O-DU /  SMO / rApp | Power consumed by hardware component | W | (non-real time for training) | Measurement: 3GPP TS 28.552 [12] (Cl.  5.1.1.19.2)  Reporting:  O-RAN.WG4.MP (Sec. B.1, B.5) |
| O-FH (M-  Plane) or  O-FH (M-  Plane), O1, R1 | O-RU / O-DU  or  O-RU / O-DU /  SMO / rApp | Transmit power | mW | (non-real time for training) | Measurement and reporting: O-RAN.WG4.MP (Sec. B.1, B.2.1) |

Input Decision Making / AI/ML Inference:

##### Table 5.2.1.3-3: Input Decision Making / AI/ML Inference

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Input Data** | | | | | |
| **Interface** | **Source / Target** | **Name/Description** | **Units** | **Reporting Period** | **Existing Definitions** |
| O1, R1 | E2 Node (O- CU) / SMO /  rApp | Data Volume in DL delivered from O- CU-UP to O-DU, per PLMN, per QoS level, per slice, per Interface (F1-U, Xn-U, X2-U) | Mbit | > min | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.3.6.2.3) |
| O1, R1 | E2 Node (O- CU) / SMO /  rApp | Data Volume in UL delivered to O- CU-UP from O-DU, per PLMN, per QoS level, per slice, per Interface (F1- U, Xn-U, X2-U) | Mbit | > min | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.3.6.2.4) |
| O1, R1 | E2 Node (O- CU) / SMO /  rApp | RSRQ measurement per SSB per cell | dB | > min | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.31) |
| O1, R1 | E2 Node (O- CU) / SMO /  rApp | RSRP measurement per SSB per cell | dBm | > min | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.22) |
| O1, R1 | E2 Node (O- CU) / SMO /  rApp | SINR measurement per SSB per cell | dB | > min | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.32) |
| O1, R1 | E2 Node (O- CU) / SMO /  rApp | Energy consumption | kWh | > min | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.19.3) |
| O-FH (M-  Plane) or  O-FH (M-  Plane), O1, R1 | O-RU / O-DU  or  O-RU / O-DU /  SMO / rApp | Power consumed by hardware component | W | > min | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.19.2)  Reporting:  O-RAN.WG4.MP (Sec. B.1, B.5) |
| O-FH (M-  Plane) or  O-FH (M-  Plane), O1, R1 | O-RU / O-DU  or  O-RU / O-DU /  SMO / rApp | Transmit power | mW | > min | Measurement and reporting:  O-RAN.WG4.MP (Sec. B.1, B.2.1) |

* + - * 1. Detailed Output Requirements

Output Decision Making / AI/ML Inference:

##### Table 5.2.1.3-4: Output Decision Making / AI/ML Inference

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Output Data** | | | | | |
| **Interface** | **Source / Target** | **Name/Description** | **Units** | **Config. Period** | **Existing Definitions** |
| O1, R1 | rApp / SMO / E2 Node (O- CU) | Candidate carrier(s) / cells (s) recommended for energy saving to enter energySaving state (e.g., 3GPP TS 28.310 [8] Sec. 6.2.2) | - | > min | 3GPP TS 28.541 [11]  (Cl. 4.3.63)  O-RAN-WG5.O-CU-O1 (Sec. 9) |
| Candidate carrier(s) / cell (s) for compensation (e.g., 3GPP TS 32.551  [15] compensatingForEnergySaving, optional) | - | > min | - |
| O1, R1 | E2 Node (O- CU) / SMO /  rApp | Confirmation (Success/Failure) | - | event triggered | 3GPP TS 28.310 [8] (Cl.  6.2.2) |
| O-FH (M-  Plane) | E2 Node (O- DU) / O-RU | Updated carrier configuration (i.e., activation, deactivation, or sleep) | - | > min | O-RAN.WG4.MP (Sec. 15.3.2) |

### Option 2: Near-RT RIC Deployment

In option 2, decision making, potentially using AI/ML Model Inference, is done at Near-RT RIC. While AI/ML Model Training might be hosted in Non-RT or Near-RT RIC, the description below is based on AI/ML Model Training in the Non-RT RIC.

#### Description and UML Diagram

##### Table 5.2.2.1-1: Carrier and Cell Switch Off/On: AI/ML inference via Near-RT RIC

|  |  |  |
| --- | --- | --- |
| **Use Case Stage** | **Evolution / Specification** | **<<Uses>> Related use** |
| Goal | Enable Carrier and Cell switch off/on Energy Saving functions in the Network by means of configuration parameter change and Actions controlled by Near-RT RIC and allow for AI/ML-based solutions. |  |
| Actors and Roles | Non-RT RIC acting as AI/ML model management entity.  Near-RT RIC acting as inference host for Energy Savings decision making.  E2 Node and O-RU are the subject of action for configuration enforcement. |  |
| Assumptions | O1 connectivity is established between the SMO with E2 Node, and Near-RT RIC and Non-RT RIC.  E2 interface connectivity is established between E2 Node and Near-RT RIC.  A1 interface is established between Non-RT RIC and Near-RT RIC. Open FH M-Plane interface is established between E2 Node and O-RU. Network is operational.  Near-RT RIC has knowledge about overlapping carriers/cells and the coverage of those carriers/cells (e.g., which carrier/cell is a coverage layer and which is a capacity layer). |  |
| Pre-conditions | The operator has set the targets for Energy Saving function in the Non- RT RIC. |  |
| Begins when | The operator enables the optimization functions for carrier and cell switch off/on Energy Saving functions and E2 Node and O-RU become  operational. |  |
| Step 1 (M) | SMO initiates specific measurement data collection request towards E2 Node and O-RU for AI/ML model training. |  |
| Step 2 (M) | E2 Node and O-RU send the configured measurement data to SMO periodically or event based. |  |
| Step 3 (M) | Non-RT RIC retrieves the collected measurement data for processing |  |
| Step 4 (M) | Non-RT RIC trains the AI/ML models with the collected data. Trained AI/ML models are deployed, configured, and activated in the Near-RT RIC. |  |
| Step 5 (M) | SMO may trigger EE/ES optimization and might provide policies guiding the Near-RT RIC EE/ES function via O1 and/or via Non-RT RIC and A1 interface. |  |
| Step 6 (M) | Near-RT RIC constantly monitors   1. performance and energy consumption of the E2 Node(s) 2. energy consumption of O-RU(s)   Based on the AI/ML inference, considering optimization policies, the Near-RT RIC may request the E2 Node to prepare and execute cell or carrier switch off/on. E2 Node may request O-RU Node to prepare and execute cell or carrier switch off/on. E2 Node will notify Near-RT RIC  once cell or carrier switch off/on is completed. |  |
| Ends when | E2 Node becomes non-Operational or when the operator disables the optimization functions for Energy Saving. |  |
| Exceptions | None. |  |
| Post Conditions | Near-RT RIC continues closed-loop monitoring of Energy Saving function at E2 Node and O-RU.  E2 Node and O-RU operate using the newly deployed parameters/models and state (off/on). |  |

@startuml skin rose

skinparam defaultFontSize 12 autonumber

Box "Service Management & \n Orchestration Framework" #gold

Participant "Collection & Control" as SMO Participant "Non-RT RIC" as NRTRIC

End box

Box "O-RAN Nodes" #lightpink Participant "Near-RT RIC" as RTRIC Participant "E2-Nodes" as E2NODES Participant "O-RUs" as ORUs

End box

group Data Collection autonumber 1.1

SMO -> E2NODES : <<O1>> Data collection request for Energy Saving E2NODES -> ORUs : <<FH>> Data collection request for Energy Saving

autonumber 2.1

ORUs -> E2NODES : <<FH>> Measurement Data Collection for Energy Saving E2NODES -> SMO : <<O1>> Measurement Data Collection for Energy Saving autonumber 3

SMO -> NRTRIC : Data retrieval

end

autonumber 4.1

group AI/ML workflow

NRTRIC -> NRTRIC : AI/ML Model training

NRTRIC -> RTRIC : <<O1>> or <<O2>> Deploy AI/ML model

end

autonumber 5.1

group Optimization Trigger and Policy alt

SMO -> RTRIC : <<O1>> Optimization Trigger/Target else

NRTRIC -> RTRIC : <<A1>> A1 Policy

end

end

autonumber 6.1

group Actor Data Collection & Decision Making

RTRIC -> E2NODES : <<E2>> Data collection request for Energy Saving E2NODES -> ORUs : <<FH>> Data collection request for Energy Saving ORUs -> E2NODES : <<FH>> Measurement Data Collection for Energy Saving

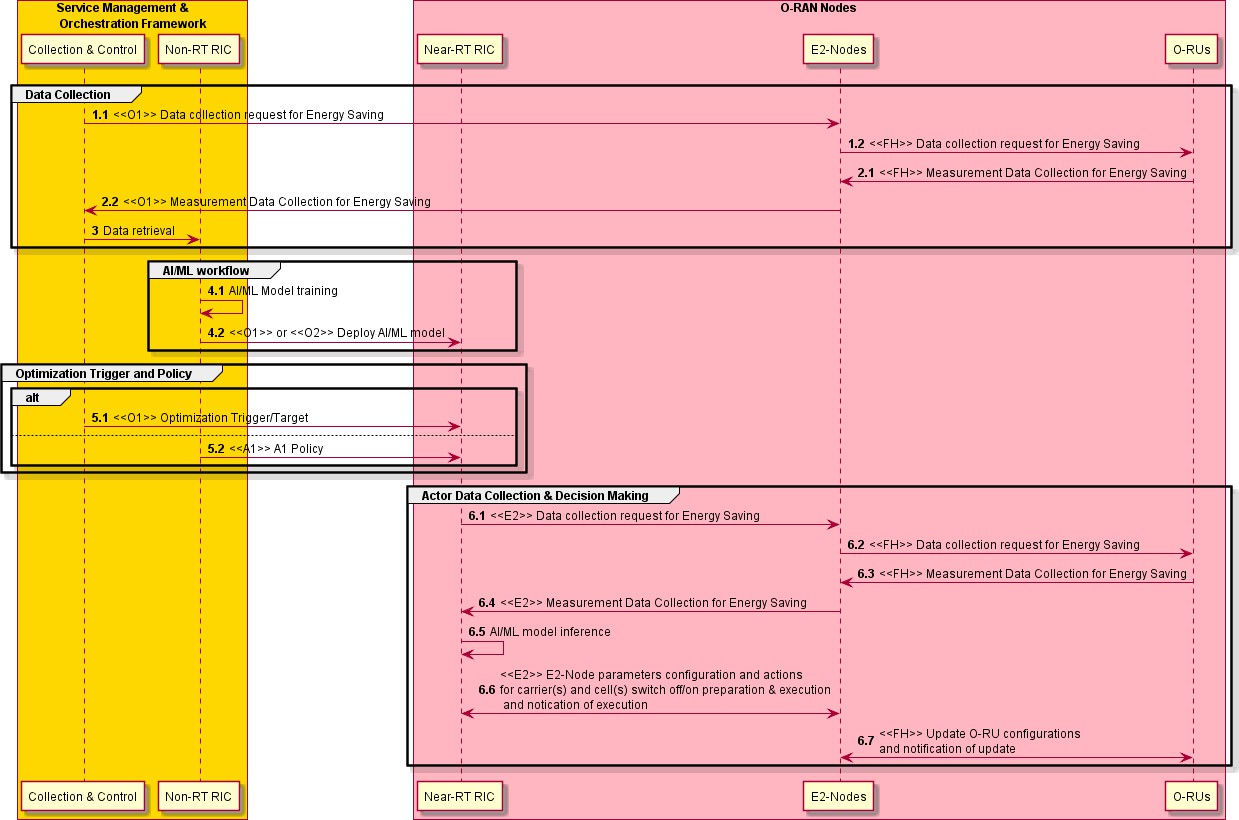
E2NODES -> RTRIC : <<E2>> Measurement Data Collection for Energy Saving

RTRIC -> RTRIC: AI/ML model inference

RTRIC <-> E2NODES : <<E2>> E2-Node parameters configuration and actions \nfor carrier(s) and cell(s) switch off/on preparation & execution \n and notication of execution

E2NODES <-> ORUs: <<FH>> Update O-RU configurations \nand notification of update

end @enduml



##### Figure 5.2.2.1-1: Carrier and Cell Switch Off/On flow diagram: AI/ML inference via Near-RT RIC

#### O-RAN Entity Roles

1. Non-RT RIC
   1. (Optionally) Collect configurations, performance indicators and measurement reports (e.g., cell load related information and traffic information, EE/EC measurement reports, geolocation information) from SMO, E2 Nodes and O-RUs (forwarded by SMO), for the purpose of training AI/ML models that assist such EE/ES functions.
   2. (Optionally) Trigger EE/ES AI/ML model training/retraining.
   3. (Optionally) Deploy, update, configure EE/ES AI/ML models, in Near-RT RIC via O1/O2 interface.
   4. Provide optimization trigger, optimization targets and intent-based policies (e.g., set energy target to 50% of peak power consumption) to Near-RT RIC via R1/O1 or A1 interface.
2. Near-RT RIC
   1. Collect configurations, performance indicators and measurement reports (e.g., cell load related information and traffic information, EE/EC measurement reports) from E2 Nodes.
   2. (Optionally) Receive EE/ES AI/ML model for deployment via O1 or O2 interface.
   3. Receive EE/ES related configuration management via O1 interface and/or policies via A1 interface for consideration during optimization.
   4. Analyze the received data from E2 Nodes and perform AI/ML model inference to determine EE/ES. optimization (i.e. if carriers or cells are recommended to be switched off/on) considering the optimization targets/policies.
   5. Provide policies or required information to E2 Node (O-CU) via E2 to trigger actions for EE/ES optimization.
3. E2 Node
   1. Report cell configuration, performance indicators and measurements reports (e.g., cell load related information and traffic information, EE/EC measurement reports) per cell/carrier via O1 interface to SMO and via E2 interface to Near-RT RIC.
   2. Perform actions required for EE/ES optimization
      * e.g. check ongoing emergency calls and warning messages, perform some preparation actions for Off Switching (e.g., to enable, disable, modify Carrier Aggregation and/or Dual Connectivity, to trigger HO traffic and UEs from cells/carriers to other cells or carriers, informing neighbour nodes via X2/Xn interface etc.) as well as for On Switching (e.g., cell probing, informing neighbour nodes via X2/Xn interface etc.) and make final decision on switch off/on and notify Near-RT RIC via E2 about performed actions
4. O-RU
   1. Report EC and EE related information via Open FH M-Plane interface to O-DU or alternatively to SMO directly.
   2. Support actions required to perform EE/ES optimization.
      * updated carrier configuration (i.e. activation, deactivation or sleep)

#### Input/Output Data Requirements

* + - * 1. Summary

Input Data

1. E2 Node to SMO/Non-RT RIC (training) and Near-RT RIC (decision making/inference)
   * Carrier/cell characteristics
   * EE/EC measurement reports
   * Load statistics per cell and per carrier
   * UE mobility information including cell or beam level measurements (e.g., RSRP, RSRQ, SINR)
   * Energy Consumption
2. O-RU to E2 Node
   * Power consumption metrics: Mean total/per carrier power consumption, mean total/per carrier transmit power

Output Data

1. Near-RT RIC to E2 Node
   * Carrier(s) and/or cell(s) recommended to be switched off/on
     + - 1. Detailed Input Requirements

Initialization:

##### Table 5.2.2.3-1: Initialization

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Input Data** | | | | | |
| **Interface** | **Source / Target** | **Name/Description** | **Units** | **Reporting Period** | **Existing / New Definitions** |
| O1 | SMO / Near-RT RIC | Optimization target for Carrier/Cell Switch Off/On |  | manual or event triggered | New |
| E.g., (average or max) NG-RAN data Energy Efficiency | bit/J |  | 3GPP TS 28.554 [7] (Cl.  6.7.1) |
| E2, O1, R1 | E2 Node (O- DU) / Near-RT RIC / SMO /  rApp | Carrier/cell characteristics | - | manual or event triggered (e.g., SMO  feature activation, E2 Node startup  / failure / reconfig.) | New |
| E.g., physical location, transmit direction, carrier frequency, coverage parameters, configured transmit power, beam width, coverage shape, tilt, azimuth, carrier-cell mapping, carrier-HW mapping |  |  | 3GPP TS 28.541 [11] (Cl.  4.3.6, 4.3.38, 4.3.39,  4.3.40, 4.3.74)  O-RAN.WG5.O-DU-O1 (Sec. 10), O-RAN-  WG5.O-CU-O1 (Sec. 9) |

AI/ML Model Training:

##### Table 5.2.2.3-2: AI/ML Model Training

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Input Data** | | | | | |
| **Interface** | **Source / Target** | **Name/Description** | **Units** | **Reporting Period** | **Existing Definitions** |
| O1, R1 | E2 Node (O- DU) / SMO /  rApp | DL PDCP SDU Data Volume per interface (Data Volume in DL delivered from O-CU-UP to O-DU, per PLMN, per QoS level, per slice, per Interface (F1-U, Xn-U, X2-U)) | Mbit | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.3.6.2.3) |
| O1, R1 | E2 Node (O- DU) / SMO /  rApp | UL PDCP SDU Data Volume per interface (Data Volume in UL delivered to O-CU-UP from O-DU, per PLMN, per QoS level, per slice, per Interface (F1-U, Xn-U, X2-U)) | Mbit | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.3.6.2.4) |
| O1, R1 | E2 Node (O- DU) / SMO /  rApp | RSRQ measurement per SSB per cell | dB | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.31) |
| O1, R1 | E2 Node (O- DU) / SMO /  rApp | RSRP measurement per SSB per cell | dBm | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.22) |
| O1, R1 | E2 Node (O- DU) / SMO /  rApp | SINR measurement per SSB per cell | dB | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.32) |
| O1, R1 | E2 Node (O- DU) / SMO /  rApp | Energy consumption | kWh | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.19.3) |
| O-FH (M-  Plane) or  O-FH (M-  Plane), O1, R1 | O-RU / O-DU  or  O-RU / O-DU /  SMO / rApp | Power consumed by hardware component | W | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.19.2)  Reporting:  O-RAN.WG4.MP (Sec. B.1, B.5) |
| O-FH (M-  Plane) or  O-FH (M-  Plane), O1, R1 | O-RU / O-DU  or  O-RU / O-DU /  SMO / rApp | Transmit power | mW | (non-real time for training) | Measurement and reporting:  O-RAN.WG4.MP (Sec. 10.2, B.1, B.2.1) |

Input Decision Making / AI/ML Inference:

##### Table 5.2.2.3-3: Input Decision Making / AI/ML Inference

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Input Data** | | | | | |
| **Interface** | **Source / Target** | **Name/Description** | **Units** | **Reporting Period** | **Existing / New Definitions** |
| E2 | E2 Node (O- DU) / Near-RT RIC | Data Volume in DL delivered from O- CU-UP to O-DU, per PLMN, per QoS level, per slice, per Interface (F1-U, Xn-U, X2-U). In case of split gNB architecture. | Mbit | ~ per N x 100ms | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.3.6.2.3)  Reporting:  O-RAN.WG3.E2SM- KPM |
| E2 | E2 Node (O- DU) / Near-RT RIC | Data Volume in UL delivered to O- CU-UP from O-DU, per PLMN, per QoS level, per slice, per Interface (F1- U, Xn-U, X2-U). In case of split gNB architecture. | Mbit | ~ per N x 100ms | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.3.6.2.4)  Reporting:  O-RAN.WG3.E2SM- KPM |
| E2 | E2 Node (O- DU) / Near-RT RIC | RSRQ measurement per SSB per cell | dB | ~ per N x 100ms | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.31)  Reporting:  O-RAN.WG3.E2SM- KPM |
| E2 | E2 Node (O- DU) / Near-RT RIC | RSRP measurement based on SSB per UE | dBm | ~ per N x 100ms | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.22)  Reporting:  O-RAN.WG3.E2SM- KPM |
| E2 | E2 Node (O- DU) / Near-RT RIC | SINR measurement based on SSB per UE | dB | ~ per N x 100ms | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.32)  Reporting:  O-RAN.WG3.E2SM- KPM |
| E2 | E2 Node (O- DU) / Near-RT RIC | Energy consumption | kWh | ~ per N x 100ms | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.19.3)  New reporting: E2 |
| O-FH (M-  Plane) | O-RU / O-DU | Power consumed by hardware component | W | ~ per N x 100ms | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.19.2)  Reporting:  O-RAN.WG4.MP (Sec. B.1, B.5) |
| O-FH (M-  Plane) | O-RU / O-DU | Transmit power | mW | ~ per N x 100ms | Measurement and reporting:  O-RAN.WG4.MP (Sec. 10.2, B.1, B.2.1) |

* + - * 1. Detailed Output Requirements

Output Decision Making / AI/ML Inference:

##### Table 5.2.2.3-4: Output Decision Making / AI/ML Inference

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Output Data** | | | | | |
| **Interface** | **Source / Target** | **Name/Description** | **Units** | **Config. Period** | **Existing / New Definitions** |
| E2 | Near-RT RIC / E2 Node (O- CU) | Candidate carrier(s) / cells (s) recommended for energy saving to enter energySaving state (e.g., 3GPP TS 28.310 [8] Sec. 6.2.2) | - | > min | 3GPP TS 28.541 [11]  (Cl. 4.3.63)  New configuration: E2 |
| Candidate carrier(s) / cell (s) for compensation (e.g., 3GPP TS 32.551  [15] compensatingForEnergySaving, optional) | - | > min | - |
| E2 | E2 Node (O- CU) / Near-RT RIC | Confirmation (Success/Failure) | - | event triggered | 3GPP TS 28.310 [8] (Cl.  6.2.2) |
| O-FH (M-  Plane) | E2 Node (O- DU) / O-RU | Updated carrier configuration (i.e. activation, deactivation, or sleep) | - | > min | O-RAN.WG4.MP (Sec. 15.3.2) |

## Impact Analysis on O-RAN Work Groups

This is an initial impact analysis as part of the WG1 UCTG Network Energy Saving work on Carrier and Cell Switch Off/On use case. The intention is to estimate the expected standardization effort within the O- RAN work groups. It is up to the WGs to decide how Carrier and Cell Switch Off/On use case functionality should be specified in specifications of each WG.

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **WGs/FGs** | **Spec. No** | **Objective description** |
| **1** | WG1 (Use  case) | O-RAN.WG1.NES-USE-CASES-TR  O-RAN.WG1.Use-Cases-Detailed- Specification | Update WG1 NES use case analysis report and use- case detailed specification with Carrier and Cell Switch Off/On use case. No impact to existing  architecture. |
| **2** | WG2 (Non-RT RIC, A1, R1) | O-RAN.WG2.R1GAP O-RAN.WG2.R1UCR  O-RAN.WG2.R1TD (TBD) | Updates to R1 services and procedures for Carrier and Cell Switch Off/On use case |
| O-RAN.WG2.Non-RT-RIC-ARCH | No impact identified |
| O-RAN.WG2.Use-Case-Requirements | Specifying Carrier and Cell Switch Off/On use case and its requirements in WG2 UCR specification |
| O-RAN.WG2.A1GAP O-RAN.WG2.A1TD | Review and implement potential requirements for policy driven implementation 2nd deployment option of Carrier and Cell Switch Off/On use case captured  in WG2 UCR specification. |
| 3 | WG3  (Near-RT RIC, E2) | O-RAN.WG3.UCR | Specifying 2nd deployment option of Carrier and Cell Switch Off/On use case and its requirements in WG3 UCR specification |
| O-RAN.WG3.RICARCH | No impact identified |
| O-RAN.WG3.E2GAP | No impact identified |
| O-RAN.WG3.E2AP | No impact identified |
| O-RAN.WG3.E2SM-RC or NEW: O-RAN.WG3.E2SM-CC | Identify and specify RAN E2 actions necessary for 2nd deployment option of Carrier and Cell Switch Off/On use case through E2 Node |
| O-RAN.WG3.E2SM-KPM | Identify and specify RAN E2 measurement required analysis of ES and EC for 2nd deployment option of  Carrier and Cell Switch Off/On use case through Near-RT RIC |
| 4 | WG4 (O-FH) | O-RAN.WG4.MP | Review and specify requirements for M-Plane. Identify and specify relevant impacts on M-Plane for both Hierarchical and hybrid model to accommodate management features requirements towards O-RU.  Define O-RU Energy efficiency KPIs and counters. |
| O-RAN-WG4.CUS | Review and specify requirements for CUS-Plane. Identify and specify relevant impacts on CUS Plane and data model to support various Tx/Rx Array  Carrier Off/On. |
| 5 | WG5 (O1) | O-RAN.WG5.O-DU-O1 O-RAN.WG5.O-CU-O1 O-RAN.WG5.MP | Identify specific O-DU operational and data model aspects of the feature content including the interface between SMO and O-DU, and the one between SMO and O-CU. Make appropriate changes to the O-DU data model and other WG5 specifications as  needed. |
| 6 | WG10 | O-RAN.WG10.O1-Interface  O-RAN.WG10.OAM-Architecture  O-RAN.WG10.Information Model and Data Models | Review, identify enhancements and update the relevant impacts on O1 interface and IM/DM to support Carrier and Cell Switch Off/On use case. |

## Relation and Impact on 3GPP Specifications

### Relation to 3GPP RAN Specifications

|  |  |  |  |
| --- | --- | --- | --- |
| **Specs** | **Release** | **Title** | **Description** |
| **TR 36.927** | LTE  Rel.10 | Study on Potential solutions for energy saving for E-UTRAN | Study on energy saving for E-UTRAN. |
| **TS 36.300**  **and other specs** | LTE  Rel.10 | E-UTRAN; Overall description; Stage 2 | Support of LTE energy saving by cell/carrier off/on switching. The switch-off decision might be taken by the capacity booster cell autonomously or may be centrally controlled by OAM. The eNB may initiate handover with appropriate root cause values and thereby inform its neighbour cells about the future switch off. The load of the neighbour cells may also be monitored by load information exchange over X2. Before switching off, all peer eNBs can be informed about the switch off over the X2 interface. For the switch off a Deactivation Indication was added to the ENB CONFIGURATION UPDATE procedure.  For switch on a CELL ACTIVATION REQUEST/RESPONSE  procedure was introduced in LTE Rel.10. The autonomous switch on decision is more difficult to realize particularly for small cells. Even with high load on the coverage cell, it is not obvious if there are UEs in the vicinity of such small cells. |
| **TS 36.300**  **and other specs** | LTE  Rel.11 | E-UTRAN; Overall description; Stage 2 | LTE Rel.11 defined respective S1 messages for Inter-RAT (i.e. GERAN, UMTS) support. |
| **TR 36.887** | LTE  Rel.12 | Study on energy saving enhancement for E- UTRAN | Study on energy saving enhancements for E-UTRAN. |
| **TS 36.300**  **and other specs** | LTE  Rel.15 | E-UTRAN; Overall description; Stage 2 | EN-DC CONFIGURATION UPDATE procedure was specified to support LTE-NR Dual Connectivity in the Non-Standalone Configurations. |
| **TS 38.300**  **and other specs** | 5G  Rel.15 | NR and NG-RAN  Overall description; Stage-2 | Support of 5G energy saving by cell/carrier off/on switching. Similar functionality as defined for LTE. E-UTRA or NR cell providing additional capacity via single or dual connectivity. |

### Ongoing Work in 3GPP Rel.18 RAN

The following study is conducted in 3GPP Rel.18 in RAN. This includes:

* Definition of a base station energy consumption model
* Definition of an evaluation methodology and KPIs
* Study and identify techniques on the gNB and UE side to improve network energy savings in terms of both BS transmission and reception

|  |  |  |  |
| --- | --- | --- | --- |
| **Reference** | **Release** | **Title** | **Documentation and impact on Specifications** |
| **FS\_Netw\_Energy\_NR** | 5G Rel.18 | Study on network energy savings for NR | Results will be captured in 3GPP TR 38.864 [19]. |

While 3GPP RAN WG1 is leading this study item, RAN WG3 will study aspects of information exchange/coordination over network interfaces such as X2/Xn. For this purpose, RAN WG3 also considers Enhanced Carrier/Cell Switch Off/On mechanisms.

### Impact on 3GPP RAN Specifications

Non-RT or Near-RT RIC controlled Cell Switch Off/On is not considered to have any impact on 3GPP RAN specifications.

### Relation to 3GPP System Architecture Specifications

|  |  |  |  |
| --- | --- | --- | --- |
| **Specs** | **Release** | **Title** | **Description** |
| **TR 36.927** | LTE  Rel.10 | Study on Energy Savings Management | Study on Energy Savings Management |
| **TS 32.551**  **and other specs** | LTE  Rel.10 | Energy Saving Management (ESM); Concepts and requirements | Concepts and requirements for ESM. According to 3GPP TS 32.551, the energy saving procedure initiating capacity booster cell, to enter or exit “energySaving” state, should be able to initiate energy saving compensation activation and/or deactivation on one or multiple cells or network elements. 3GPP TS 32.425 [13] defines the measurements and 3GPP TS 32.451  [14] defines the KPIs. |
| **TR 32.834** | LTE  Rel.11 | Study on OAM aspects of inter-RAT Energy Saving | Inter-RAT aspects of Energy Saving. The following RATs are considered in this study: GSM, UMTS, LTE, CDMA. |
| **TR 23.866** | LTE  Rel.12 | Study on System Enhancements for Energy Efficiency | Study on System Enhancements for Energy Efficiency |
| **TR 32.856** | 5G  Rel.15 | Study on OAM support for assessment of energy efficiency in mobile access networks | Support for ETSI metrics and methods for energy efficiency assessment and related measurements. Principles for monitoring equipment power, energy and environmental parameters. 2G, 3G and 4G have been covered. |
| **TS 28.310**  **and other specs** | 5G  Rel.16 | Energy efficiency of 5G | This TS defines solutions and requirements for   * Data Volume collection Power, * Power, Energy and Environmental (PEE) measurement collection * Energy Saving use cases   The Energy Saving use case defines the scenario of energy saving by switching off capacity booster cells when the traffic demand is low and re-activate them on a need basis. The energy saving consists of two scenarios where the capacity booster cell gNB is fully or partially overlaid by the candidate cell(s).  The configuration of energy saving policies for these solutions is defined in 3GPP TS 28.541 [11]. The measurements related to data volume and PEE measurements are defined in 3GPP TS  28.552 [12]. The Energy Efficiency related KPIs are defined in TS 28.554 [7]. |

### Ongoing Work in 3GPP Rel.18 System Architecture

Energy saving compensation activation and deactivation procedures for LTE are defined in 3GPP TS 32.551

[15] as:

* Energy saving compensation activation: procedure to increase the coverage area for the candidate cell(s).
* Energy saving compensation deactivation: procedure to decrease a previously increased coverage area.

This procedure is not defined yet for energy saving procedures for 5G, and hence 3GPP TS 28.310 [8] needs to be enhanced to introduce this procedure for 5G networks. Related study and work items are existing.

|  |  |  |  |
| --- | --- | --- | --- |
| **Reference** | **Release** | **Title** | **Documentation and impact on Specifications** |
| **FS\_EE5G\_Ph2** | 5G Rel.18 | Study on new aspects of EE for 5G networks Phase 2 | Results will be captured in 3GPP TR 28.913 Study on new aspects of EE for 5G networks phase 2. |
| **EE5GPLUS\_Ph2** | 5G Rel.18 | Enhancements of EE for 5G Phase 2 | Specification enhancements are expected for the following 3GPP specifications: TS 28.310 [8], TS  28.552 [12], TS 28.554 [7] and TS  28.541 [11]. |

### Impact on 3GPP System Architecture Specifications

Non-RT or Near-RT RIC controlled Cell Switch Off/On is not considered to have any impact on 3GPP System Architecture specifications.

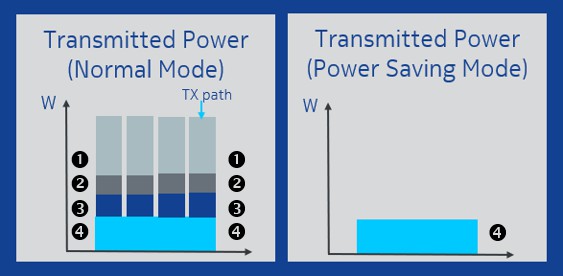
## Gain Analysis

An operator can expect strong energy savings, and thus OPEX savings, by using Carrier/Cell Switching in a network with multiple frequency layers. The exact power saving strongly depends on network deployment details, cell configuration (e.g., configured bandwidth, common channels etc.), used hardware, and network load. Multiple frequency layers may or may not use the same power amplifier in the RF module. Major power savings are naturally obtained when some power amplifiers can be switched off completely.

Numerical analysis based on an estimated power saving per carrier / cell to be switched off might provide a good estimate of the expected savings. O-RU power consumption might be categorized into power consumption related to:

1. user channels' load (e.g., traffic on uplink/downlink shared channels),
2. common signals and channels' load (e.g., SSB transmissions, Reference Signals, broadcast channel, paging channel etc.),
3. regular PA operational power (i.e., the required energy to power up the power amplifier),
4. regular operational power (i.e., the required energy to power up the radio module).

Power consumption related to (1), (2) and (3) scales with the number of transmission paths as illustrated in Figure 5.5-1.



##### Figure 5.5-1: O-RU power saving for a 4TX4RX sector configuration

Considering that the requested traffic (1) needs to be transmitted in one of the remaining carriers/cells and that regular operational power (4) will be required after carrier/cell switch off, the power savings of the Carrier/Cell Switch algorithm is mainly related to the power consumption of (2) and (3).

The yearly energy savings might be calculated as follows:

𝑌𝑒𝑎𝑟𝑙𝑦𝑃𝑜𝑤𝑒𝑟𝑆𝑎𝑣𝑖𝑛𝑔[𝑀𝑊ℎ] =

𝑂-𝑅𝑈 𝑃𝑜𝑤𝑒𝑟𝑆𝑎𝑣𝑖𝑛𝑔[𝑊] 1.000.000

𝑃𝑜𝑤𝑒𝑟𝑆𝑎𝑣𝑖𝑛𝑔𝐻𝑜𝑢𝑟𝑠[ℎ]

∗

𝑑𝑎𝑦

∗ #𝑑𝑎𝑦𝑠 ∗ #𝑂-𝑅𝑈𝑠

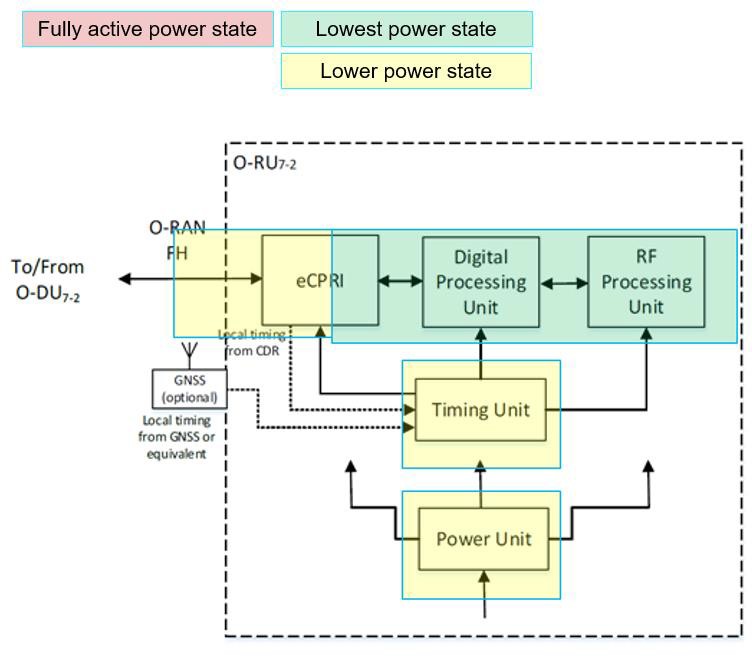
Two example calculations, simplifying the factors as cited above and ignoring deployment, cell configuration, hardware specifics and load dependencies, are provided in the table below:

##### Table 5.5-1: Simplified power savings calculation examples

|  |  |  |
| --- | --- | --- |
| **Factor** | **Value** | **Value** |
| O-RU power savings during switch off | 100 W | 200 W |
| Number of hours of switch off per day | 2 | 4 |
| Number of O-RUs | 10000 | 20000 |
| Yearly energy saving [MWh] | 730 | 5840 |

To be more specific considering deployment, cell configuration, antenna type, hardware and load dependencies of the O-RU power consumption, computational analysis results for two example scenarios are provided. The calculations are based on the load profile and the power consumption categorization of O-RU functional blocks as outlined in Annex B.

Energy savings gain from Carrier and Cell Switch Off/On would be equivalent to power consumption of the O-RU hardware components that can be shut down or put into energy savings mode during low traffic load. As illustrated in Figure 5.5-2, major energy savings gain would be derived from shutting down RF Processing Unit and Digital Processing Unit (depending on the antenna configuration), while FH Processing Unit, Power Unit and other components would still be fully or partially functioning, and the power consumption would stay more or less consistent during the switching off period.



##### Figure 5.5-2: Energy savings for O-RU

### Cell and Carrier Switch Off/On Energy Saving for 4T4R O-RU

The energy saving gain from Carrier and Cell Switch Off/On is analyzed based on the system parameters in Table 5.5.1-1 and the example power profile considering power consumption from operational experience provided in Table 5.5.1-2.

##### Table 5.5.1-1: O-RU Configuration for ES gain analysis

|  |  |
| --- | --- |
| **No. of antennas** | 4T4R |
| **No. of layers** | 4 |
| **Bandwidth** | 100 MHz |
| **Carrier frequency** | 3.5 GHz |
| **Tx power per antenna** | 30 W |
| **Technology** | 5G NR |

##### Table 5.5.1-2: Example power profile for 4T4R O-RU

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Operating Load (Traffic)** | **Total O- RU (W)** | **ORAN**  **Fronthaul Processing Unit (W)** | **Digital Processing Unit**  **(W)** | **RF**  **Processing Unit (W)** | **Power Unit & other components (W)** |
| Busy hour load | 550 | 10 | 25 | 495 | 20 |
| Low load | 200 | 10 | 25 | 145 | 20 |
| Sleep (Energy Savings State) | 50 | 5~10 | 0~5 | 0~15 | 15~20 |
| Energy saving gain | 150~180 | 0~5 | 20~25 | 130~145 | 0~5 |

Compared to the low load scenario, energy savings of up to 150~180 W per O-RU can be achieved in the energy saving state. An example calculation for yearly energy saving potential considering a shutdown of 10000 O-RUs during 3 hours per day is provided in Table 5.5.1-3.

##### Table 5.5.1-3: Power saving calculation example for 4T4R O-RU

|  |  |
| --- | --- |
| **Factor** | **Value** |
| O-RU power savings during switch off (low load) [W] | 150~180 |
| Number of hours of switch off per day (50% of low load period) | 3 |
| Number of O-RUs | 10000 |
| Yearly energy saving [MWh] | 1643~1971 |

### Cell and Carrier Switch Off/On Energy Saving for 64T64R O-RU

The energy saving gain from Carrier and Cell Switch Off/On is analyzed based on the system parameters in Table 5.5.2-1 and the example power profile considering power consumption from operational experience provided in Table 5.5.2-2.

##### Table 5.5.2-1: O-RU Configuration for ES gain analysis

|  |  |
| --- | --- |
| **No. of antennas** | 64T64R |
| **No. of layers** | 16 |
| **Bandwidth** | 100 MHz |
| **Carrier frequency** | 3.5 GHz |
| **Tx power per antenna** | 30 W |
| **Technology** | 5G NR |

##### Table 5.5.2-2: Example power profile for 64T64R O-RU

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Operating Load (Traffic)** | **Total O- RU (W)** | **ORAN**  **Fronthaul Processing Unit**  **(W)** | **Digital Processing Unit (W)** | **RF**  **Processing Unit (W)** | **Power Supply & other components (W)** |
| Busy hour load | 1100 | 50 | 200 | 800 | 50 |
| Low load | 400 | 50 | 200 | 100 | 50 |
| Sleep (Energy Savings State) | 60~140 | 30~50 | 0~20 | 0~20 | 30~50 |
| Energy saving gain | 260~340 | 0~20 | 180~200 | 80~100 | 0~20 |

Compared to the low load scenario, energy savings of 260-340 W per O-RU can be achieved in the energy saving state. Example calculations for yearly energy saving potential considering a shutdown of 10000 O- RUs during 3 hours per day are provided in Table 5.5.2-3.

##### Table 5.5.2-3: Power saving calculation examples for 64T64R O-RU

|  |  |
| --- | --- |
| **Factor** | **Value** |
| O-RU power savings during switch off (low load) [W] | 260~340 |
| Number of hours of switch off per day (50% of low load period) | 3 |
| Number of O-Rus | 10000 |
| Yearly energy saving [MWh] | 2847~3723 |

## Feasibility Analysis

No challenges concerning feasibility have been identified during the pre-normative phase.

# RF Channel Reconfiguration

## Problem Statement, Solution and Value Proposition

In mobile networks mMIMO antennas are used for beamforming techniques to enhance cell capacity and throughput. In order to achieve beamforming, O-RUs need to concentrate the power amplifiers at the radome by combining radiating elements. At low load, i.e., when the expected traffic volume or number of connected users are lower than the configured threshold, ES can be achieved by reducing the power consumption of O- RUs by switching off certain Tx/Rx arrays. For example, 32 out of 64 Tx/Rx Arrays of an O-RU can be switched off in a digital mMIMO architecture and correspondingly the number of spatial layers and SSBs can be reduced. The procedure (involvement of respective O-RAN interfaces) of the RF Channel Reconfiguration depends on the management architecture model (hybrid or hierarchical) and the deployment option. The reconfiguration decision can be made by an AI/ML model within the inference host deployed at the Non-RT RIC (denoted as deployment Option 1), or at the Near-RT RIC (denoted as deployment option 2). Among others the AI/ML models may include prediction of future traffic, user mobility, and resource usage and may also predict expected energy efficiency enhancements, resource usage, and network performance for different ES optimization states.

The main aim of RF Channel Reconfiguration is to perform O-RU Tx/Rx Array selection. However, Tx/Rx Array reselection may require modifying the maximum number of spatial streams, the number of SSB Beams or the O-RU Antenna transmit power. Hence the overall scope of RF Channel Reconfiguration includes the actions listed in Table 6.1-1.

##### Table 6.1-1: Actions in the context of RF Channel Reconfiguration

|  |  |  |
| --- | --- | --- |
| Action Name | Explanation | Possible Implementation Method |
| O-RU Tx/Rx Array selection | O-RU Tx/Rx Array selection means switching off certain Tx/Rx Arrays or Array elements to reduce power consumption of O-RU.  Reselecting Rx/Tx Arrays may impact cell coverage. | * O-RU reports all supported Tx/Rx Array selections to O-DU or to SMO via Open FH M-Plane. * Based on traffic load and user distribution the Non-RT/Near RT RIC will optimize Tx/Rx Array selection. |
| Modification of number of SU/MU MIMO spatial streams or data layers | The maximum number of spatial streams may depend on the O-RU Tx/Rx Array selection, and it may be necessary to be modified along with a new Tx/Rx Array selection.  A reduction of the number of spatial streams without Tx/Rx Array selection is not excluded as in certain scenarios the energy consumption of O-RU and/or O-DU may be reduced by turning off certain processing units of FPGA/GPUs, by running at a lower clock speed or turning off CPU/GPU cores along with reduced processing requirements. | * O-RU may report the maximum number of supported spatial streams for each Tx/Rx Array selection. For example, 16 spatial streams for 64 Tx/Rx and 8 spatial Layers for 32 Tx/Rx. * Based on the traffic load and the user distribution the O-RU will be instructed about the maximum number of spatial streams to be used. Hence, the O-RU and/or O-DU may be able turn off certain processing units that would be required to processing a higher number of spatial streams. |
| Modification of the number of SSB beams | The number of SSB beams may depend on the O-RU Tx/Rx Array selection and the number of spatial streams.  Hence it may be necessary to adapt the number of SSB beams. | - The number of SSB beams is controlled by the O-DU. Hence the O-DU can be instructed via Non-RT/Near-RT RIC to set the number of SSB beams based on Tx/Rx Array selection. |
| Modify the O-RU Antenna Transmit power | The maximum O-RU total transmit power may depend upon the O-RU Tx/Rx Array selection; hence it may need to be modified along with Tx/Rx Array selection to compensate for reduced coverage in the downlink. | - The O-RU power is configured per Tx Array. Hence, the modification may not be required when O-RU Tx/Rx Array selection from 64 Tx/Rx to 32 Tx/Rx. However, if possible, O-RU Tx power can be increased to compensate for reduced coverage. |

## Architecture/Deployment Options

### Option 1: Non-RT RIC Deployment

In option 1, decision making, potentially including AI/ML Model Training and Inference, is done at the Non- RT RIC.

#### Description and UML Diagram

##### Table 6.2.1.1-1: RF Channel Reconfiguration: AI/ML inference via Non-RT RIC

|  |  |  |
| --- | --- | --- |
| **Use Case Stage** | **Evolution / Specification** | **<<Uses>>**  **Related use** |
| Goal | Enable RF Channel Reconfiguration Energy Saving functions in the Network by means of configuration parameter change and actions  controlled by Non-RT RIC to enable AI/ML-based solutions. |  |
| Actors and Roles | Non-RT RIC acting as inference host for Energy Savings decision making.  E2 Node and O-RU are the subject of action for configuration  enforcement. |  |
| Assumptions | O1 interface connectivity is established between E2 Nodes and SMO. Open FH M-Plane interface is established between E2 Node and O- RU or between O-RU and SMO.  Network is operational. |  |
| Pre-conditions | The operator has set the targets for Energy Saving functions in the Non-RT RIC (rApp).  Capability of O-RU to support various RF Channel Reconfiguration  actions. |  |
| Begins when | The operator enables the optimization functions for RF Channel  Reconfiguration Energy Saving and E2 Node and O-RU become operational |  |
| Steps 1.1, 1.2, 1.3,  1.4 (M) | Non-RT RIC initiates specific measurement data collection request through SMO via O1 towards E2 Node or via Open FH M-Plane directly  towards O-RU for AI/ML model training for Energy Saving optimization |  |
| Steps 2.1, 2.2, 2.3,  2.4 (M) | E2 Node and O-RU send the configured measurement data to SMO periodically or event based for Non-RT RIC processing. Non-RT RIC  retrieves data through SMO from E2 Node and from O-RU. |  |
| Step 3 (M) | Non-RT RIC retrieves the collected measurement data for processing. |  |
| Steps 4.1, 4.2, 4.3 (M) | Non-RT RIC trains the AI/ML models with the collected data. Trained AI/ML models are deployed and activated in the Non-RT RIC.  Non-RT RIC constantly monitors performance and energy consumption of the E2 Node and the O-RU for inference, such as cell load related and traffic information, EE/EC measurement reports, geolocation information etc. |  |
| Step 5 (M) | Based on the AI/ML inference, the Non-RT RIC may request the SMO to configure E2 Node (O-DU) and execute RF Channel  Reconfiguration for ES/EE optimization. |  |
| Steps 6.1, 6.2, 6.3,  6.4 (M) | SMO instructs E2 Node via O1 to perform the received request(s) from Non-RT RIC   1. O-RU Tx/Rx Array selection. 2. Modify the number of SU/MU MIMO spatial streams or data layers. 3. Modify the number of SSB beams. 4. Modify O-RU Antenna Transmit power. |  |
| Step 7 (M) | Non-RT RIC continuously monitors the performance of AI/ML model. If energy saving objectives are not achieved, it may decide to initiate  fallback mechanism, and/or AI/ML model update or retraining. |  |
| Ends when | If E2 Node becomes non-operational or when the operator disables the  optimization functions for Energy Saving. |  |
| Exceptions | None |  |
| Post Conditions | Non-RT RIC continues closed loop monitoring of Energy Saving function for E2 Node and O-RU.  E2 Node and O-RU operate using the newly deployed parameters and  state. |  |

@startuml

skinparam defaultFontSize 12 autonumber

Box "Service Management & \n Orchestration Framework" #gold Participant “Collection & Control” as smo

Participant " Non-RT RIC" as NRTRIC End box

Box "O-RAN Nodes" #lightpink Participant "Near-RT RIC" as RTRIC Participant "E2-Node" as E2NODES Participant "O-RU" as ORUs

End box

autonumber 1.1

group Data Collection alt via O1

smo -> E2NODES : <<O1>> Data collection request for Energy Saving E2NODES -> ORUs : <<FH>> Data collection request

Else via OFH-MP

smo -> E2NODES : <<O1>> Data collection request for Energy Saving smo -> ORUs : <<OFH-MP>> Data collection request

end

autonumber 2.1 alt Via O1

ORUs -> E2NODES : <<FH>> Measurement Data Collection

E2NODES -> smo : <<O1>> Measurement Collection for Energy Saving Else via OFH-MP

ORUs -> smo : <<FH>> Measurement Data Collection

E2NODES -> smo : <<O1>> Measurement Collection for Energy Saving End

autonumber 3

smo -> NRTRIC : <<O1>> Data Retrieval end

group Data Analysis Training and Inference autonumber 4.1

NRTRIC -> NRTRIC : AI/ML Model training

NRTRIC -> NRTRIC : Deploy & activate trained AI/ML Model for inferencing

NRTRIC -> NRTRIC : Monitoring & Analysis of Energy Efficiency \n & Consumption (E2 Node & O-RU)

autonumber 5

group Actor Decision Making

NRTRIC -> smo : Request to prepare and execute RF Channel Reconfiguration

\nenforecement for Energy saving autonumber 6.1

smo -> E2NODES : <<O1>> RF Channel Reconfiguration enforecement for Energy saving E2NODES -> ORUs : <<FH>> Updated O-RU Configurations

end autonumber 7

NRTRIC -> NRTRIC : Performance analysis of AI/ML model \n(with possible actions, e.g. fallback, re-training)

end @enduml



**1.2** FH Data collection request

**2.1** FH Measurement Data Collection

**2.3** FH Measurement Data Collection

**2.4** O1 Measurement Collection for Energy Saving

**3** O1 Data Retrieval

* 1. AI/ML Model training
  2. Deploy activate trained AI/ML Model for inferencing

**4.3** Monitoring Analysis of Energy Efficiency

Consumption (E2 Node O-RU)

**6.1** O1 RF Channel Reconfiguration enforecement for Energy

**6.2** FH Updated O-RU Configurations

O-RU

E2-Node

Near-RT RIC

Non-RT RIC

Collection Control

)

rmance analysis of AI/ML model

possible actions, e.g. fallback, re-training

(with

Perfo

saving

enforecem ent for Energy saving

**5** Request to prepare and execute RF Channel Reconfiguration

**ision Making**

**Actor Dec**

**is Training and Inference**

**Data Analys**

**via OF -MP**

**2.2** O1 Measurement Collection for Energy Saving

**ia O1**

**alt**

**1.4** OFH-MP Data collection request

**1.3** O1 Data collection request for Energy Saving

**via OF -MP**

**via O1**

**1.1** O1 Data collection request for Energy Saving

**alt**

**tion**

**Data Collec**

O-RU

E2-Node

Near-RT RIC

Non-RT RIC

Collection Control

**O-RAN Nodes**

**Service Management**

**Orchestration Framework**

##### Figure 6.2.1.1-1: RF Channel Reconfiguration flow diagram: AI/ML inference in Non-RT RIC

#### O-RAN Entity Roles

1. SMO (including Non-RT RIC)
   1. Collect necessary cell configurations, performance indicators and measurement reports (e.g., cell load related and traffic information, EE/EC measurement reports, geolocation information) from E2 Node and O-RU, for the purpose of training and inference of AI/ML models that assist in the EE/ES functions.
   2. Trigger EE/ES AI/ML model training/retraining.
   3. Deploy, update, and configure EE/ES AI/ML models
   4. Analyze the data received from E2 Nodes and O-RU to determine RF Channel Reconfiguration actions for EE/ES, i.e., O-RU Tx/Rx Array selection, modification of the number of SSB beams, modification of the O-RU Antenna Transmit power, modification of the number of SU/MU MIMO spatial streams or data layers using AI/ML models.
   5. Signal updated RF Channel Reconfiguration and execution of optimization actions to E2 Node via O1 Interface.
2. E2 Node
   1. Report necessary cell configurations, performance indicators, and measurement reports (e.g., cell load related and traffic information, EE/EC measurement reports) to SMO via O1 interface.
   2. Perform actions required to perform RF Channel Reconfiguration (i.e., O-RU Tx/Rx Array selection, modification of the number of SSB beams, modification of the O-RU Antenna Transmit power, modification of the number of SU/MU MIMO data layers or spatial streams) as part of EE/ES optimization.
3. O-RU
   1. Report EC and EE related information over Open FH M-Plane to O-DU or alternatively to SMO directly.
   2. Perform actions required to be performed due to RF Channel Reconfiguration (i.e., O-RU Tx/Rx Array selection, modification of the number of SSB beams, modification of the O-RU Antenna Transmit power, modification of the number of SU/MU MIMO spatial streams or data layers) as part of EE/ES optimization.

#### Input/Output Data Requirements

* + - * 1. Summary

Input Data

1. SMO and E2 Node
   * Load statistics per cell and per carrier, such as number of active users, average number of RRC connections, average number of scheduled active users per TTI, PRB utilization, DL/UL Cell/User throughput, PMI/CSI reports.
   * Latency statistics per cell (if URLLC slices are involved, latency is used in the EE definition, 3GPP TS 28.554 [7]).
2. O-RU
   * Power consumption metrics: Mean total/per carrier power consumption, mean total/per carrier transmit power.
   * Information on supported Tx/Rx Array selections along with power consumption (site/O-RU input power needed for certain EE KPIs)

Output Data

1. SMO to E2 Node
   * RF Channel Reconfiguration actions which can include,
     + O-RU Tx/Rx Array selection
     + Modify the number of SU/MU MIMO spatial streams or data layers
     + Modify the number of SSB beams
     + Modify O-RU Antenna Transmit power
       - 1. Detailed Input Requirements

Initialization:

##### Table 6.2.1.3-1: Initialization

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Input Data** | | | | | |
| **Interface** | **Source / Target** | **Name/Description** | **Units** | **Reporting Period** | **Existing / New Definitions** |
| R1 | SMO / rApp | Optimization target for RF Channel Reconfiguration | - | manual or event triggered | New |
| E.g., (average or max) NG-RAN data Energy Efficiency | bit/J |  | 3GPP TS 28.554 [7]  (Cl. 6.7.1) |
| R1 | SMO / rApp | Maximum Initial Access Latency for given SSB Beam Configuration | ms | Initialization | New: KPI metric input by the operator |
| O1, R1 | O-DU / SMO  / rApp | Carrier/cell characteristics | - | manual or event triggered (e.g., SMO  feature activation, E2 Node startup/failure/ reconfig.) | New |
| E.g., physical location, transmit direction, carrier frequency, coverage parameters, configured transmit power, beam width, coverage shape, tilt, azimuth, carrier-cell mapping, carrier-HW mapping | - |  | 3GPP TS 28.541 [11]  (Cl. 4.3.6, 4.3.38,  4.3.39, 4.3.40,  4.3.74)  O-RAN.WG5.O- DU-O1 (Sec. 10), O- RAN-WG5.O-CU-  O1 (Sec. 9) |
| O-FH (M-  Plane) or  O-FH (M-  Plane), O1, R1 | O-RU / O-DU  or  O-RU / O-DU  / SMO / rApp | Supported common beam configuration from O-RU per cell | file | Initialization | O-RAN.WG5.MP (Sec. 8.1) |
| O-FH (M-  Plane) | O-RU / O-DU | Beamforming weights or attributes via YANG module per cell | values in IE | Initialization | O-RAN.WG4.CUS (Sec. 12.4.2) |
| R1, O1, O- FH (M-  Plane) or  O-FH (M-  Plane) | rApp / SMO / O-DU / O-RU  or  O-DU / O-RU | Inferred SSB beam configuration in specified file per cell | file | Initialization | O-RAN.WG4.CUS (Sec. 12.4.2) |
| O1, R1 | O-DU / SMO  / rApp | Supported SSB and CSI-RS TRS configurations per cell | - | Initialization | 3GPP TS 38.331 [16]  (Sec. 6.3.2) |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| O-FH (M-  Plane), O1, R1 | O-RU / O-DU  / SMO / rApp | O-RU reports all offered Tx/Rx Arrays | file | Initialization | New Definition required in O- RAN.WG5.MP  o-ran-uplane- conf.yang module  o-ran- beamforming.yang module |
| O-FH (M-  Plane) | O-RU / O-DU | O-RU reports all offered Tx/Rx Arrays | values in IE | Initialization | New Definition required in O- RAN.WG4.MP  o-ran-uplane- conf.yang module  o-ran- beamforming.yang module |

AI/ML Model Training:

##### Table 6.2.1.3-2: AI/ML Model Training

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Input Data** | | | | | |
| **Interface** | **Source / Target** | **Name/Description** | **Units** | **Reporting Period** | **Existing / New Definitions** |
| O1, R1 | E2 Node (O- CU) / SMO /  rApp | DL PDCP SDU Data Volume per interface (Data Volume in DL delivered from O-CU-UP to O-DU, per PLMN, per QoS level, per slice, per Interface (F1-U, Xn-U, X2-U)) | Mbit | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.3.6.2.3) |
| O1, R1 | E2 Node (O- CU) / SMO /  rApp | UL PDCP SDU Data Volume per interface (Data Volume in UL delivered to O-CU-UP from O-DU, per PLMN, per QoS level, per slice, per Interface (F1-U, Xn-U, X2-U)) | Mbit | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.3.6.2.4) |
| O1, R1 | E2 Node (O- CU) / SMO  / rApp | Number of active UEs in NG-RAN (Number of RRC\_CONECTED UEs) per cell | Integer | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Sec. 5.1.1.23, A7, A.60) |
| O1, R1 | E2 Node (O- CU) / SMO /  rApp | RSRQ measurement per SSB per cell | dB | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.31) |
| O1, R1 | E2 Node (O- CU) / SMO /  rApp | RSRP measurement per SSB per cell | dBm | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.22) |
| O1, R1 | E2 Node (O- CU) / SMO /  rApp | SINR measurement per SSB per cell | dB | > min | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.32) |
| O1, R1 | E2 Node (O- CU) / SMO /  rApp | Energy consumption | kWh | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.19.3) |
| O-FH (M-  Plane) or  O-FH (M-  Plane), O1, R1 | O-RU / O-DU  or  O-RU / O-DU /  SMO / rApp | Power consumed by O-RU or its hardware components | W | (non-real time for training) | Measurement: 3GPP TS28.552 [12]  (Cl. 5.1.1.19.2)  Reporting:  O-RAN.WG4.MP (Sec. B.1, B.5) |
| O1, R1 | E2 Node (O- DU) / SMO /  rApp | Over the air transmit power by O-RU as calculated in O-DU  Note: This measurement is required to calculate energy efficiency of O- RU. | W | (non-real time for training) | New Measurement and New Reporting required in O- RAN.WG5.MP |

Input Decision Making / AI/ML Inference:

##### Table 6.2.1.3-3: Input Decision Making / AI/ML Inference

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Input Data** | | | | | |
| **Interface** | **Source / Target** | **Name/Description** | **Units** | **Reporting Period** | **Existing / New Definitions** |
| O1, R1 | E2 Node (O-CU)  / SMO / rApp | DL PDCP SDU Data Volume per interface (Data Volume in DL delivered from O-CU-UP to O-DU, per PLMN, per QoS level, per slice, per Interface (F1-U, Xn-U, X2-U)) | Mbit | > min | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.3.6.2.3) |
| O1, R1 | E2 Node (O-CU)  / SMO / rApp | UL PDCP SDU Data Volume per interface (Data Volume in UL delivered to O-CU-UP from O-DU, per PLMN, per QoS level, per slice, per Interface (F1-U, Xn-U, X2-U)) | Mbit | > min | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.3.6.2.4) |
| O1, R1 | E2 Node (O-CU)  / SMO / rApp | Number of active UEs in NG-RAN (Number of RRC\_CONECTED UEs) per cell | Integer | > min | Measurement: 3GPP TS 28.552 [12]  (Sec. 5.1.1.23, A.7, A.60) |
| O1, R1 | E2 Node (O-CU)  / SMO / rApp | RSRQ measurement per SSB per cell | dB | > min | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.31) |
| O1, R1 | E2 Node (O-CU)  / SMO / rApp | RSRP measurement per SSB per cell | dBm | > min | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.22) |
| O1, R1 | E2 Node (O-CU)  / SMO / rApp | SINR measurement per SSB per cell | dB | > min | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.32) |
| O1, R1 | E2 Node (O-CU)  / SMO / rApp | Energy consumption | kWh | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.19.3) |
| O-FH (M-  Plane) or  O-FH (M-  Plane), O1, R1 | O-RU / O-DU  or  O-RU / O-DU /  SMO / rApp | Power consumed by O-RU or its hardware components | W | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.19.2)  Reporting:  O-RAN.WG4.MP (Sec. B.1, B.5) |
| O1, R1 | E2 Node (O-DU)  / SMO / rApp | Over the air transmit power by O-RU as calculated in O-DU  Note: This measurement is required to calculate energy efficiency of O- RU. | W | (non-real time for training) | New Measurement and New Reporting required in O- RAN.WG5.MP |

* + - * 1. Detailed Output Requirements

Output Decision Making / AI/ML Inference:

##### Table 6.2.1.3-4: Output Decision Making / AI/ML Inference

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Output Data** | | | | | |
| **Interface** | **Source / Target** | **Name/Description** | **Units** | **Reporting Period** | **Existing / New Definitions** |
| O-FH (M-  Plane) | O-DU / O-RU | Candidate O-RU’s appropriate array selection | - | > min | O-RAN.WG4.MP  o-ran-uplane-conf.yang |
| Maximum number of Max SU/MU MIMO spatial streams or data layers | - | > min | New Definition required in O-RAN.WG4.MP |
| Recommended O-RU Antenna Transmit power | W | > min | O-RAN.WG4.MP  urn:o-ran:uplane-conf Module  gain |
| R1, O1 | rApp / SMO  / O-DU | Candidate O-RU’s appropriate array selection | - | > min | New Definition required in WG2 R1 and O- RAN.WG5.O-DU-O1 |
| Maximum number of Max SU/MU MIMO spatial streams or data layers | - | > min | New Definition required in O-RAN.WG4.MP |
| Inferred O-RU SS Burst Set (SS Block Number and SS Burst Periodicity) and CSI-RS TRS Configuration | - | Initialization | 3GPP TS 38.331 [16] IE:  ServingCellConfigCom mon |
| Recommended O-RU Antenna Transmit power | W | > min | O-RAN.WG4.MP  urn:o-ran:uplane-conf Module  gain |
| O1, R1 | E2 Node (O- DU) / SMO  / rApp | Confirmation (Success/Failure) | - | event triggered | New |

### Option 2: Near-RT RIC Deployment

In option 2, decision making, potentially using AI/ML Model Inference, is done at Near-RT RIC. While AI/ML Model training might be hosted in the Non-RT or in the Near-RT RIC, the description below is based on AI/ML Model Training in the Non-RT RIC.

#### Description and UML Diagram

##### Table 6.2.2.1-1: RF Channel Reconfiguration: AI/ML inference via Near-RT RIC

|  |  |  |
| --- | --- | --- |
| **Use Case Stage** | **Evolution / Specification** | **<<Uses>> Related**  **use** |
| Goal | Enable RF Channel Reconfiguration Energy Saving functions in the Network  by means of configuration parameter change and actions controlled by Near- RT RIC and allow for AI/ML-based solutions. |  |
| Actors and Roles | Near-RT RIC acting as inference host for Energy Savings decision making. E2 Node and O-RU are the subject of action for configuration enforcement. SMO/Non-RT RIC acting as policy maker or trigger for ES optimization in  Near-RT RIC. |  |
| Assumptions | O1 interface connectivity is established between E2 Node, Near-RT RIC, and SMO.  E2 interface connectivity is established between E2 Node and Near-RT RIC. A1 interface is established between Non-RT RIC and Near-RT RIC.  Open FH M-Plane interface is established between E2 Node and O-RU or between O-RU and SMO.  Network is operational. |  |
| Pre-conditions | The operator has set the targets for the Energy Saving function in the Non- RT RIC (rApp).  Capability of O-RU for various RF Channel Reconfiguration actions. |  |
| Begins when | Operator enables the optimization functions for RF Channel Reconfiguration  Energy Saving and E2 Node and O-RU become operational. |  |
| Steps 1.1,1.2, 1.3  (M) | SMO initiates specific measurement data collection request towards E2  Node and O-RU for AI/ML model training for Energy Saving optimization. |  |
| Steps 2.1, 2.2, 2.3,  2.4 (M) | E2 Node and O-RU send the configured measurement data to SMO  periodically or event based for Non-RT RIC processing. |  |
| Step 3 (M) | Non-RT RIC retrieves the collected measurement data for processing. |  |
| Steps 4.1, 4.2, 4.3 (M) | Non-RT RIC trains the AI/ML models with the collected data. Trained AI/ML models are deployed and activated in Near-RT RIC through SMO over O1 or O2 Interface.  Non-RT RIC constantly monitors performance and energy consumption of the E2 Node and the O-RU for inference, such as cell load related and traffic  information, EE/EC measurement reports, geolocation information etc. |  |
| Steps 5.1, 5.2 (M) | SMO may trigger RF Channel Reconfiguration in the Near-RT RIC via O1 interface or optionally Non-RT RIC may provide policies via A1 interface  guiding the Near-RT RIC RF Channel Reconfiguration EE/ES function. |  |
| Steps 6.1, 6.2, 6.3,  6.4, 6.5, 6.6 (M) | Based on the AI/ML inference in Near-RT RIC, considering optimization policies, the Near-RT RIC may request the E2 Node to prepare and execute RF Channel Reconfiguration.  E2 Node may request O-RU to prepare and execute RF Channel Reconfiguration changes such as,   1. O-RU Tx/Rx Array selection 2. Modification of the number of SU/MU MIMO spatial streams or data layers. 3. Modification of the number of SSB beams. 4. Modification of the O-RU Antenna Transmit power. |  |
| Steps 7.1, 7.2 (M) | Based on performance analysis of AI/ML model, Non-RT RIC may update  AIML model in Near-RT RIC. |  |
| Ends when | E2 Node becomes non-operational or when the operator disables the  optimization functions for Energy Saving. |  |
| Exceptions | None |  |
| Post Conditions | Non-RT RIC continues closed loop monitoring of Energy Saving function for E2 Node and O-RU.  E2 Node and O-RU operate using the newly deployed parameters/models  and state. |  |

@startuml skin rose

skinparam defaultFontSize 15 autonumber

Box "Service Management & \n Orchestration Framework" #gold Participant “Collection & Control” as smo Participant " Non-RT RIC" as NRTRIC

End box

Box "O-RAN Nodes" #lightpink Participant "Near-RT RIC" as RTRIC Participant "E2-Nodes" as E2NODES Participant "O-RUs" as ORUs

End box autonumber 1.1

group Data Collection alt via O1

smo -> E2NODES : <<O1>> Data collection request for Energy Saving E2NODES -> ORUs : <<FH>> Data collection request

Else via OFH-MP

smo -> ORUs : <<OFH-MP>> Data collection request end

autonumber 2.1 alt Via O1

ORUs -> E2NODES : <<FH>> Measurement Data Collection

E2NODES -> smo : <<O1>> Measurement Collection for Energy Saving Else via OFH-MP

ORUs -> smo : <<FH>> Measurement Data Collection

E2NODES -> smo : <<O1>> Measurement Collection for Energy Saving End

autonumber 3

smo -> NRTRIC : <<O1>> Data Retrieval end

autonumber 4.1

group AI/ML workflow

NRTRIC -> NRTRIC : AI/ML Model training

NRTRIC -> NRTRIC : Monitoring & Analysis of Energy Efficicincy \n & Consumption (E2 Nodes & O-RU)(s

NRTRIC -> RTRIC : <<O1>> or <<O2>> Deploy AI/ML model

end

autonumber 5.1

group Optimization Trigger and Policy alt via O1

smo -> RTRIC : <<O1>> Optimization Trigger/Target else via A1

NRTRIC --> RTRIC : <<A1>> Intent based Policy

end

autonumber 6.1

group Actor Data Collection & Decision Making

RTRIC -> E2NODES : <<E2>> Data collection request for Energy Saving E2NODES -> ORUs : <<FH>> Data collection request for Energy Saving ORUs -> E2NODES : <<FH>> Measurement Data Collection for Energy Saving E2NODES -> RTRIC : <<E2>> Measurement Data Collection for Energy Saving RTRIC -> RTRIC: AI/ML model inference

RTRIC -> E2NODES: <<E2>> RF Channel Reconfiguration \nenforecement for Energy saving E2NODES -> ORUs: <<FH>> Updated O-RU Configurations

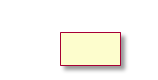
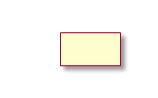
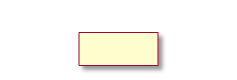
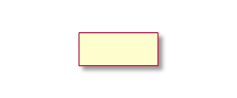
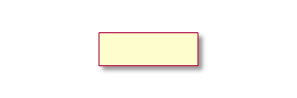
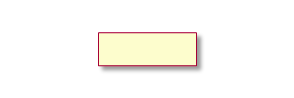
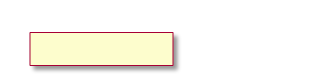
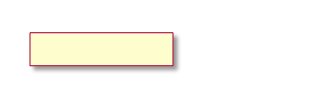
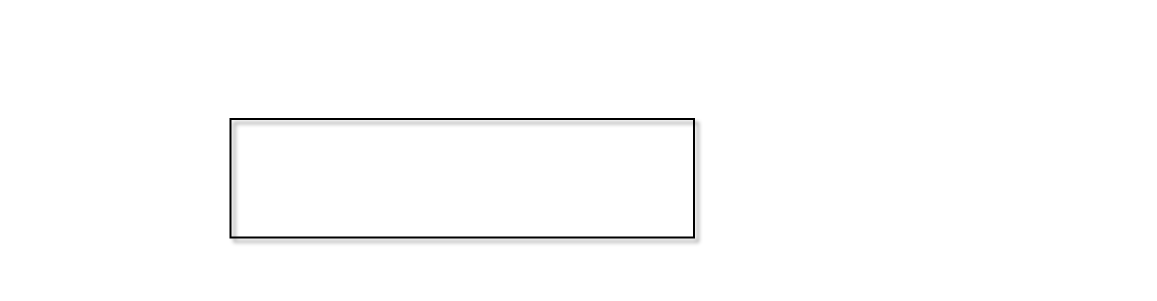
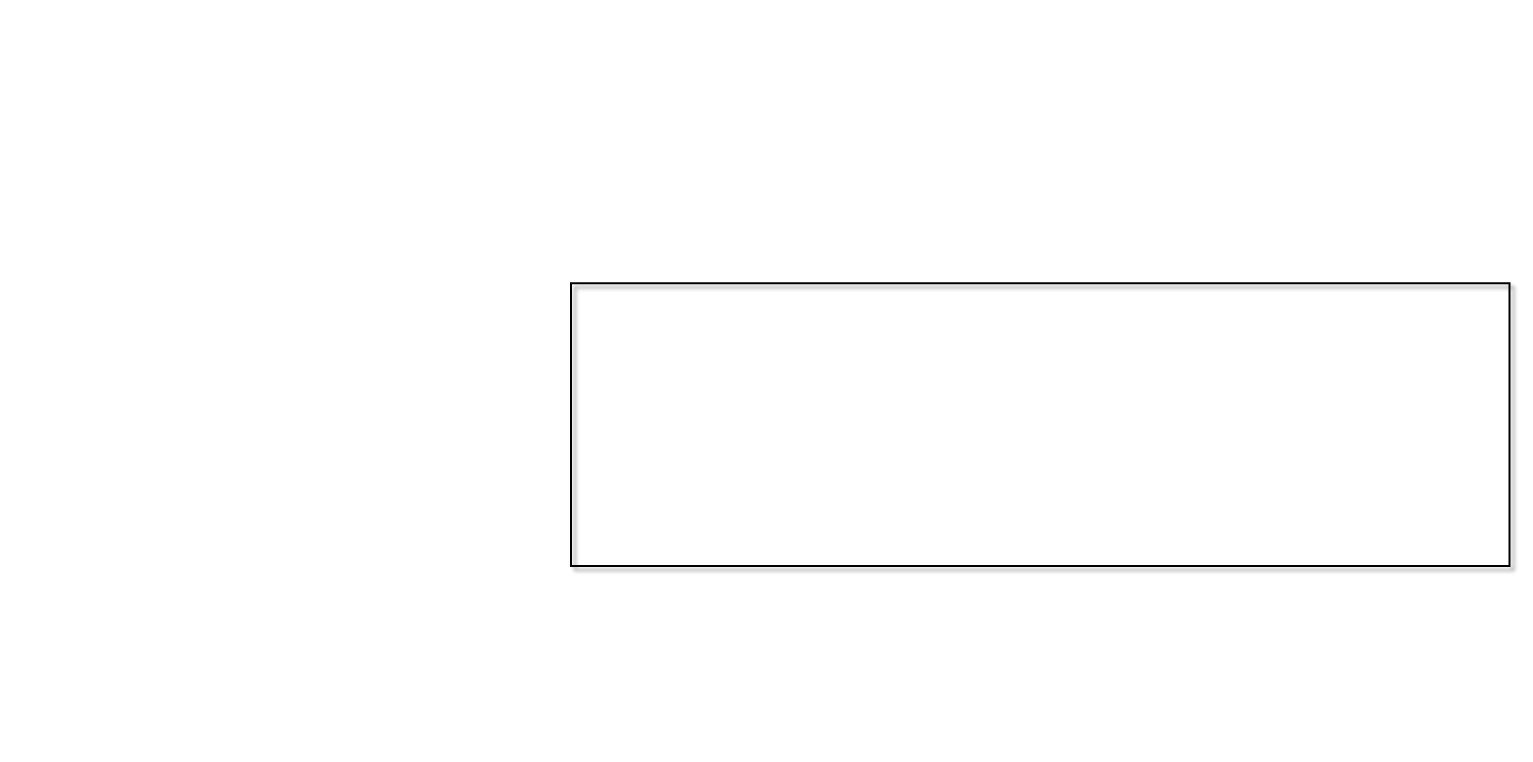
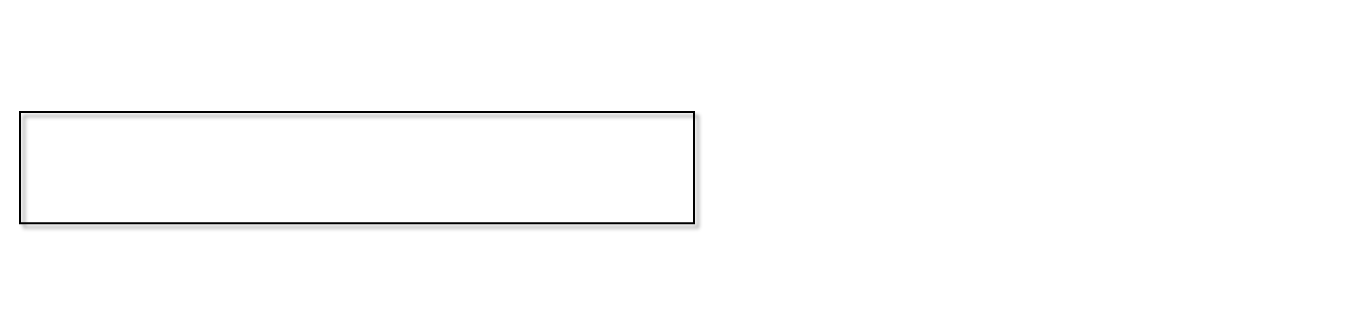
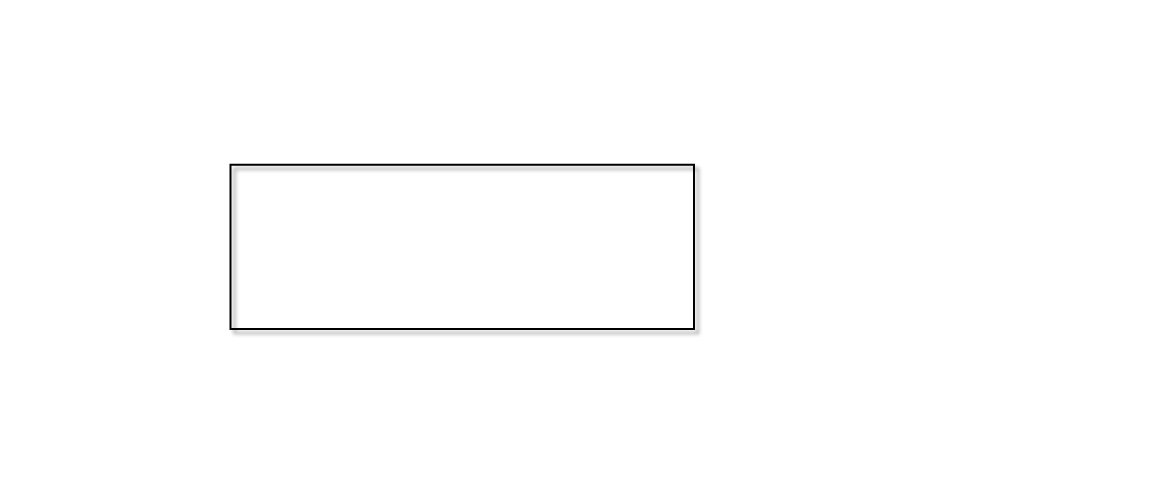
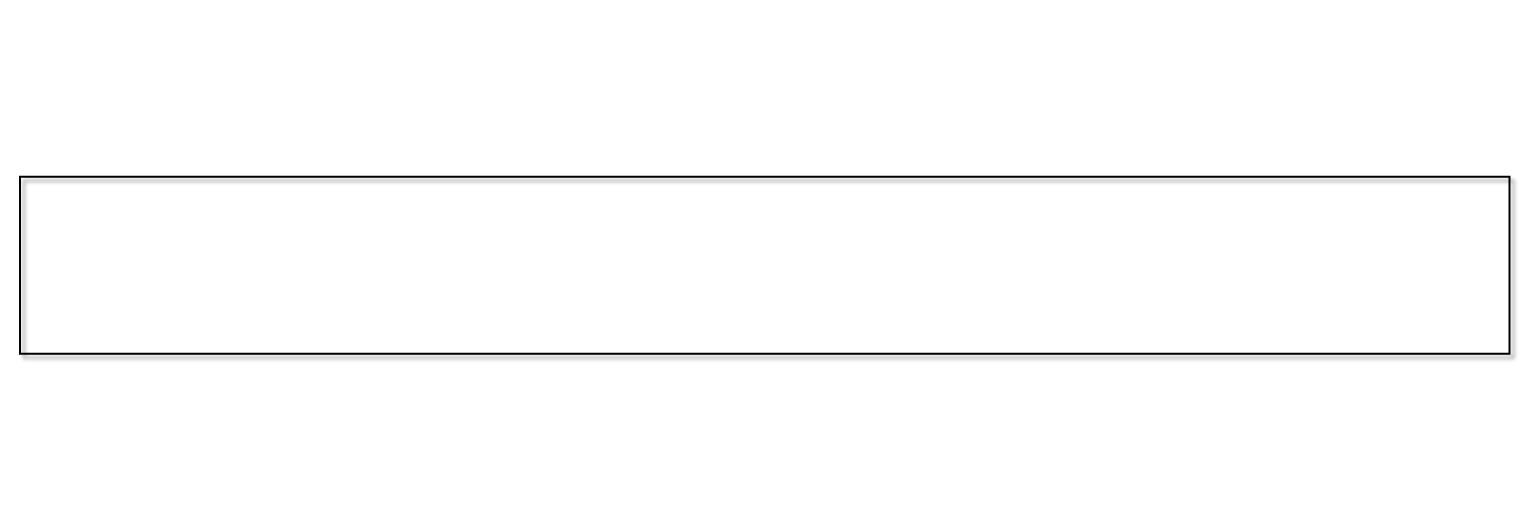
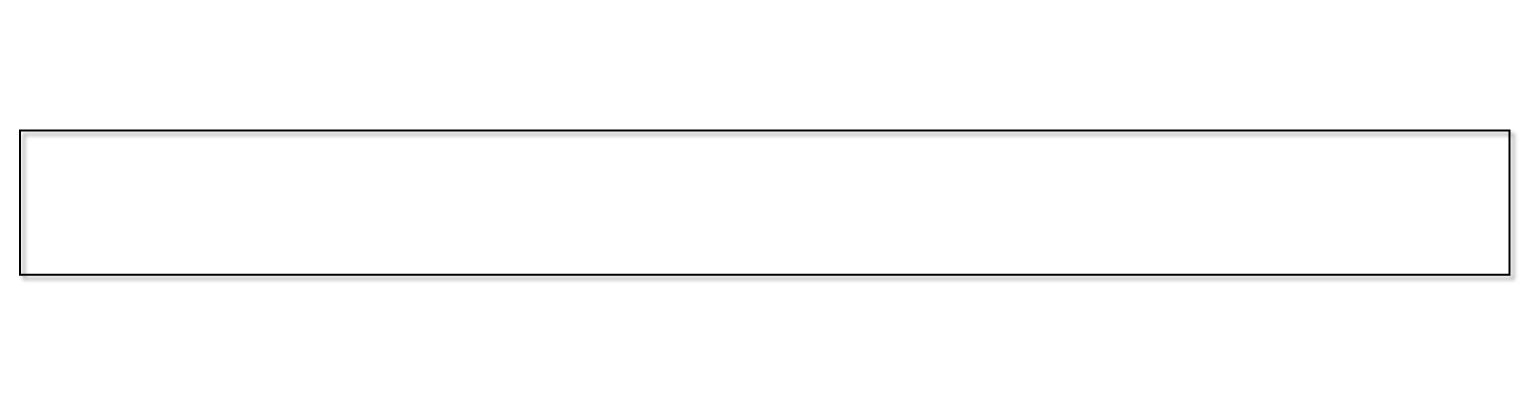
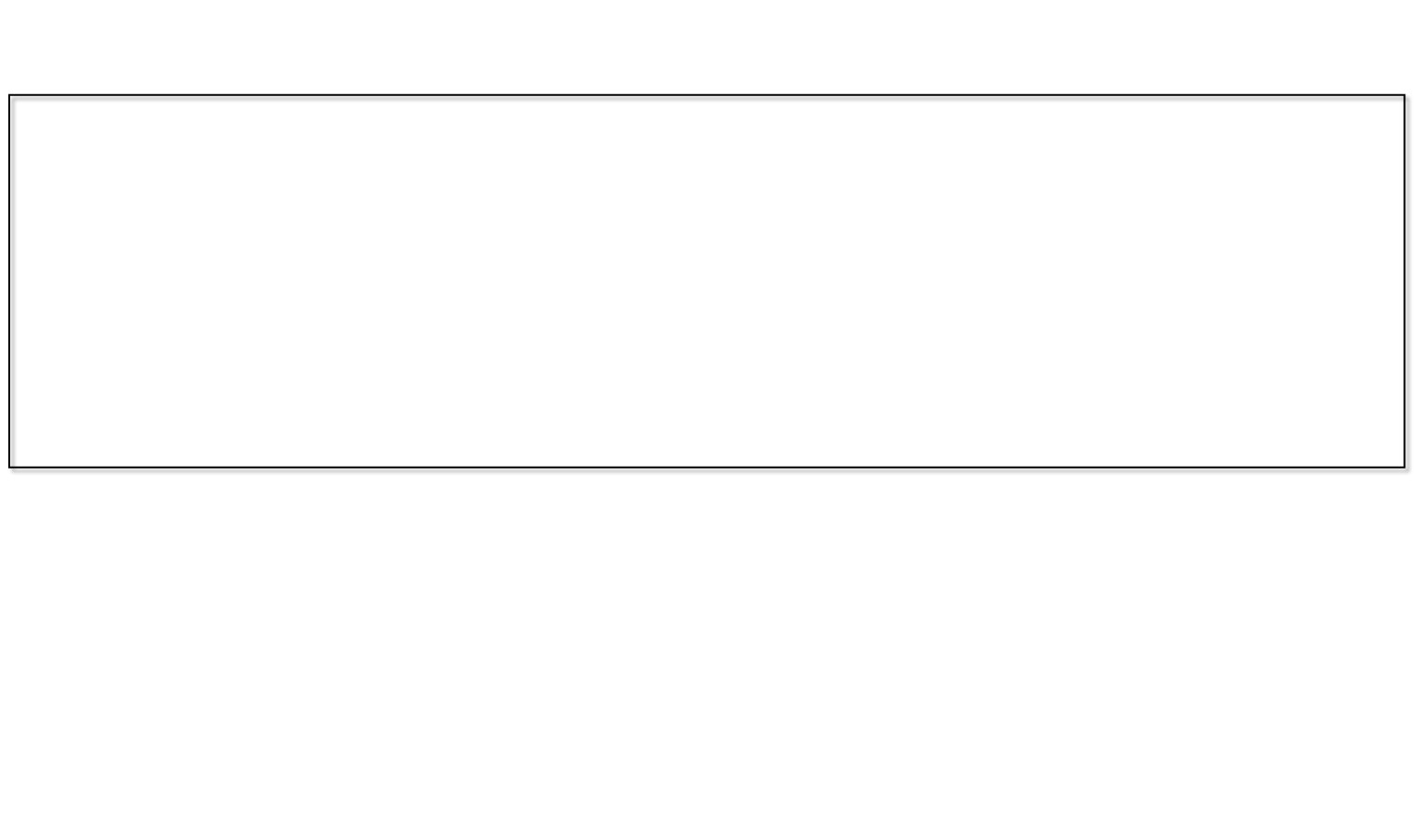
autonumber 6.1

group AI/ML workflow

NRTRIC -> NRTRIC : Performance analysis of AI/ML model \n(with possible actions, e.g. fallback, re-training)

NRTRIC -> RTRIC : <<O1>> or <<O2>> Update AI/ML model

end @enduml



**Service Management Orchestration Framework**

**O-RAN Nodes**

Collection Control Non-RT RIC Near-RT RIC E2-Nodes O-RUs

**Data Collection alt via O1**

* 1. O1 Data collection request for Energy Saving
  2. FH Data collection request

**via OF -MP**

* 1. OFH-MP Data collection request

**alt**

**ia O1**

**2.1** FH Measurement Data Collection

* 1. O1 Measurement Collection for Energy Saving

**via OF -MP**

* 1. FH Measurement Data Collection
  2. O1 Measurement Collection for Energy Saving

**3** O1 Data Retrieval

**AI/ML workflow**

**4.1** AI/ML Model training

**4.2** Monitoring Analysis of Energy Efficicincy

Consumption (E2 Nodes O-RU)(s

**4.3** O1 or O2 Deploy AI/ML model

**alt**

**via O1**

**5.1** O1 Optimization Trigger/Target

**via A1**

**5.2** A1 Intent based Policy

**Actor Data Collection Decision Making**

**6.1** E2 Data collection request for Energy Saving

**6.2** FH Data collection request for Energy Saving

**6.3** FH Measurement Data Collection for Energy Saving

**6.4** E2 Measurement Data Collection for Energy Saving

**6.5** AI/ML model inference

**6.6** E2 RF Channel Reconfiguration

enforecement for Energy saving

**6.** FH Updated O-RU Configurations

**AI/ML workflow**

**6.1** Performance analysis of AI/ML model

(with possible actions, e.g. fallback, re-training)

**6.2** O1 or O2 Update AI/ML model

Collection Control

Non-RT RIC

Near-RT RIC

E2-Nodes

O-RUs

##### Figure 6.2.2.1-1: RF Channel Reconfiguration flow diagram: AI/ML inference in Near-RT RIC

#### O-RAN Entity Roles

1. SMO (including Non-RT RIC)
   1. Collect necessary cell configurations, performance indicators and measurement reports (e.g., cell load related and traffic information, EE/EC measurement reports, geolocation information) from E2 Node and O-RU, for the purpose of training AI/ML models that assist in the EE/ES functions.
   2. Trigger and perform EE/ES AI/ML model training/retraining.
   3. Analyze the data received from E2 Node and O-RU to trigger optimization for RF Channel Reconfiguration.
   4. Provide optimization trigger, optimization targets, and A1 policies (e.g., change of Tx/Rx Array Selection based on 50 % peak power consumptions) to Near-RT RIC via O1 or A1 interface.
2. Near-RT RIC
   1. Collect necessary cell configurations, performance indicators and measurement reports (e.g., cell load related and traffic information, EE/EC measurement reports) from E2 Nodes and O-RU.
   2. Receive EE/ES AI/ML model for deployment via O1 or O2.
   3. Receive EE/ES related configuration management via O1 interface and/or policies via A1 interface for consideration during optimization.
   4. Analyze the data received from E2 Node and perform AI/ML model inference to determine RF Channel Reconfiguration actions for EE/ES (e.g., O-RU Tx/Rx Array Selection, modification of the number of SSB beams, modification of the O-RU Antenna Transmit power, modification of the number of SU/MU MIMO spatial streams or data layers) to be performed considering the optimization targets/policies.
   5. Provide policies and/or required information over E2 interface to trigger actions for EE/ES optimization.
3. E2 Node
   1. Report necessary cell configurations, performance indicators and measurement reports (e.g., cell load related and traffic information, EE/EC measurement reports) to SMO via O1 interface and to Near-RT RIC via E2 or O1 Interface.
   2. Perform O-RU Tx/Rx Array Selection, modification of the number of SSB beams, modification of the O-RU Antenna Transmit power, modification of the number of SU/MU MIMO spatial streams or data layers as part of RF Channel Reconfiguration actions for EE/ES optimization.
4. O-RU
   1. Report EC and EE related information over Open FH M-Plane interface to O-DU or alternatively to SMO directly.
   2. Perform actions required to perform O-RU RF Channel Reconfiguration as part of EE/ES optimization.
   3. Perform actions such as O-RU Tx/Rx Array Selection, modification of the number of SSB beams, modification of the O-RU Antenna Transmit power, modification of the number of SU/MU MIMO spatial streams or data layers as part of RF Channel Reconfiguration actions for EE/ES optimization.

#### Input/Output Data Requirements

* + - * 1. Summary

Input Data

1. SMO and E2 Node
   * Load statistics per cell and per carrier, such as number of active users, average number of RRC connections, average number of scheduled active users per TTI, PRB utilization, DL/UL Cell/User throughput, PMI/CSI reports.
   * Latency statistics per cell (if URLLC slices are involved, latency is used in the EE definition, 3GPP TS 28.554 [7]).
2. O-RU
   * Power consumption metrics: Mean total/per carrier power consumption, mean total/per carrier transmit power.
   * Information of O-RU for supported Tx/Rx Array selection along with power consumption (site/O-RU input power needed for certain EE KPIs).

Output Data

1. Non-RT RIC to Near RT RIC
   * O1 Configuration (i.e., ES optimization trigger/target) OR
   * A1 Policy for ES optimization
2. Near-RT RIC to E2 Node to O-RU
   * RF Channel Reconfiguration actions which can include,
     + O-RU Tx/Rx Array selection
     + Modification of the number of SU/MU MIMO spatial streams or data layers
     + Modification of the number of SSB beams
     + Modification of the O-RU Antenna Transmit power
       - 1. Detailed Input Requirements

Initialization:

##### Table 6.2.2.3-1: Initialization

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Input Data** | | | | | |
| **Interface** | **Source / Target** | **Name/Description** | **Units** | **Reporting Period** | **Existing / New Definitions** |
| O1 | SMO / Near-RT RIC | EE Optimization target for RF Channel Re-configuration |  | manual or event triggered | New |
| E.g., (average or max) NG-RAN data Energy Efficiency | bit/J |  | 3GPP TS 28.554 [7]  (Cl. 6.7.1) |
| O1 | SMO / Near-RT RIC | Maximum Initial Access Latency for given SSB Beam Configuration | ms | Initialization | New: KPI metric input by the operator |
| E2, O1, | E2 Node (O- | Carrier/cell characteristics | - | manual or | New |
| R1 | DU) / Near-RT |  |  | event triggered |  |
|  | RIC / SMO |  |  | (e.g., SMO |  |
|  | / rApp |  |  | feature |  |
|  |  |  |  | activation, E2 |  |
|  |  |  |  | Node |  |
|  |  |  |  | startup/failure/ |  |
|  |  |  |  | reconfig.) |  |
|  |  | E.g., physical location, transmit | - |  | 3GPP TS 28.541 [11] |
|  |  | direction, carrier frequency, coverage |  | (Cl. 4.3.6, 4.3.38, |
|  |  | parameters, configured transmit |  | 4.3.39, 4.3.40, |
|  |  | power, beam width, coverage shape, |  | 4.3.74) |
|  |  | tilt, azimuth, carrier-cell mapping, |  |  |
|  |  | carrier-HW mapping |  |  |
| O-FH (M-  Plane) | O-RU / O-DU | Supported common beam configuration from O-RU per cell | file | Initialization | O-RAN.WG5.MP  (Chapter 8) |
| E2 | O-DU / Near- RT RIC | Supported SSB and CSI-RS TRS configurations per cell | - | Initialization | 3GPP TS 38.331 [16]  (Sec. 6.3.2) |
| O-FH (M-  Plane), O1, R1 | O-RU / O-DU /  SMO / rApp | O-RU reports all offered Tx/Rx Arrays | file | Initialization | New Definition required in O- RAN.WG5.MP |
|  |  |  |  |  | o-ran-uplane- conf.yang module |
|  |  |  |  |  | o-ran- beamforming.yang module |
| O-FH (M-  Plane) | O-RU / O-DU | O-RU reports all offered Tx/Rx Arrays | values in IE | Initialization | New Definition required in O- RAN.WG4.MP |
|  |  |  |  |  | o-ran-uplane- conf.yang module |
|  |  |  |  |  | o-ran- beamforming.yang module |

AI/ML Model Training:

##### Table 6.2.2.3-2: AI/ML Model Training

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Input Data** | | | | | |
| **Interface** | **Source / Target** | **Name/Description** | **Units** | **Reporting Period** | **Existing/New Definitions** |
| O1, R1 | E2 Node (O- CU) / SMO  / rApp | DL PDCP SDU Data Volume per interface (Data Volume in DL delivered from O-CU-UP to O-DU, per PLMN, per QoS level, per slice, per Interface (F1-U, Xn-U, X2-U)) | Mbit | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.3.6.2.3) |
| O1, R1 | E2 Node (O- CU) / SMO  / rApp | UL PDCP SDU Data Volume per interface (Data Volume in UL delivered to O-CU-UP from O-DU, per PLMN, per QoS level, per slice, per Interface (F1-U, Xn-U, X2-U)) | Mbit | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.3.6.2.4) |
| O1, R1 | E2 Node (O- CU) / SMO  / rApp | Number of active UEs in NG-RAN (Number of RRC\_CONECTED UEs) per cell | Integer | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Sec. 5.1.1.23, A.7, A.60) |
| O1, R1 | E2 Node (O- CU) / SMO  / rApp | RSRQ measurement per SSB per cell | dB | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.31) |
| O1, R1 | E2 Node (O- CU) / SMO  / rApp | RSRP measurement per SSB per cell | dBm | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.22) |
| O1, R1 | E2 Node (O- CU) / SMO  / rApp | SINR measurement per SSB per cell | dB | > min | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.32) |
| O1, R1 | E2 Node (O- DU) / SMO  / rApp | Energy consumption | kWh | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.19.3) |
| O-FH (M-  Plane) or  O-FH (M-  Plane), O1, R1 | O-RU / O-DU  or  O-RU / O-DU  / SMO / rApp | Power consumed by O-RU or its hardware components | W | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.19.2)  Reporting:  O-RAN.WG4.MP (Sec. B.1, B.5) |
| O1, R1 | E2 Node (O- DU) / SMO  / rApp | Over the air Transmit power by O-RU  Note: This measurement is required to calculate energy efficiency of O-RU. | W | (non-real time for training) | New Measurement and New Reporting required in O- RAN.WG5.MP |
| O1, R1 | E2 Node (O- CU) / SMO  / rApp | Number of active UEs in NG-RAN (Number of RRC\_CONECTED UEs) per cell | Integer | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Sec. 5.1.1.23, A.7, A.60) |

Input Decision Making / AI/ML Inference:

##### Table 6.2.2.3-3: Input Decision Making / AI/ML Inference

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Input Data** | | | | | |
| **Interface** | **Source / Target** | **Name/Description** | **Units** | **Reporting Period** | **Existing / New Definitions** |
| E2 | E2 Node (O- CU) / Near-RT RIC | DL PDCP SDU Data Volume per interface (Data Volume in DL delivered from O-CU-UP to O-DU, per PLMN, per QoS level, per slice, per Interface (F1-U, Xn-U, X2-U)) | Mbit | per N x 100ms | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.3.6.2.3) |
| E2 | E2 Node (O- CU) / Near-RT RIC | UL PDCP SDU Data Volume per interface (Data Volume in UL delivered to O-CU-UP from O-DU, per PLMN, per QoS level, per slice, per Interface (F1-U, Xn-U, X2-U)) | Mbit | per N x 100ms | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.3.6.2.4) |
| E2 | E2 Node (O- CU) / Near-RT RIC | Number of active UEs in NG-RAN (Number of RRC\_CONECTED UEs) per cell | Integer | per N x 100ms | Measurement: 3GPP TS 28.552 [12]  (Sec. 5.1.1.23, A.7, A.60) |
| E2 | E2 Node (O- DU) / Near-RT RIC | PRACH correlation power for every received PRACH corresponding to each active SSB Beam Index | dBm | per N x 100ms | New Measurement and New Reporting: The measurement at O-DU may be delivered |
| E2 | E2 Node (O- CU) / Near-RT RIC | RSRQ measurement per SSB per cell | dB | per N x 100ms | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.31) |
| E2 | E2 Node (O- CU) / Near-RT RIC | RSRP measurement per SSB per cell | dBm | per N x 100ms | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.22) |
| E2 | E2 Node (O- CU) / Near-RT RIC | SINR measurement per SSB per cell | dB | per N x 100ms | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.32) |
| E2 | E2 Node (O- CU) / Near-RT RIC | Energy consumption | kWh | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.19.3) |
| O-FH (M-  Plane) or  O-FH (M-  Plane), E2 | O-RU / O-DU  or  O-RU / O-DU / Near-RT RIC | Power consumed by O-RU or its hardware components | W | (non-real time for training) | Measurement: 3GPP TS 28.552 [12]  (Cl. 5.1.1.19.2)  Reporting:  O-RAN.WG4.MP (Sec. B.1, B.5) |
| E2 | E2 Node (O- DU) / Near-RT RIC | Over the air Transmit power by O-RU  Note: This measurement is required to calculate energy efficiency of O-RU. | W | (non-real time for training) | New Measurement and New Reporting required in O- RAN.WG5.MP |

* + - * 1. Detailed Output Requirements

Output Decision Making / AI/ML Inference:

##### Table 6.2.2.3-4: Output Decision Making / AI/ML Inference

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Output Data** | | | | | |
| **Interface** | **Source / Target** | **Name/Description** | **Units** | **Reporting Period** | **Existing / New Definitions** |
| O-FH (M-  Plane) | O-DU / O-RU | Candidate O-RU’s appropriate Array selection | - | > min | O-RAN.WG4.MP  o-ran-uplane-conf.yang |
| Maximum number of Max SU/MU MIMO spatial streams or data layers | - | > min | New Definition required in O-RAN.WG4.MP |
| Recommended O-RU Antenna Transmit power | - | > min | O-RAN.WG4.MP  urn:o-ran:uplane-conf Module  gain |
| E2 | Near-RT RIC / O-DU | Candidate O-RU’s appropriate Array selection | - | > min | New Definition required in WG2 R1 and O- RAN.WG5.O-DU-O1 |
| Maximum number of Max SU/MU MIMO spatial streams or data layers | - | > min | New Definition required in O-RAN.WG4.MP |
| Inferred O-RU SS Burst Set (SS Block Number and SS Burst Periodicity) and CSI-RS TRS Configuration | - | Initialization | 3GPP TS 38.331 [16] IE:  ServingCellConfigCom mon |
| Recommended O-RU Antenna Transmit power | - | > min | O-RAN.WG4.MP  urn:o-ran:uplane-conf Module  gain |
| E2 | E2 Node (O- DU) / Near-RT RIC | Confirmation (Success/Failure) | - | event triggered | New |

## Impact Analysis on O-RAN Work Groups

This is an initial impact analysis as part of the WG1 UCTG Network Energy Saving work on RF Channel Reconfiguration use case. The intention is to estimate the expected standardization effort within the O-RAN work groups. It is up to the WGs to decide how RF Channel Reconfiguration functionality should be specified in specifications of each WG.

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **WGs/FGs** | **Spec. No** | **Objective description** |
| **1** | WG1 (Use  case) | O-RAN.WG1.NES-USE-CASES-TR  O-RAN.WG1.Use-Cases-Detailed- Specification | Update WG1 NES use case analysis report and use- case detailed specification with RF Channel Reconfiguration use case. No impact to existing architecture |
| **2** | WG2 (Non-RT RIC, A1, R1) | O-RAN.WG2.R1GAP O-RAN.WG2.R1UCR  O-RAN.WG2.R1TD (TBD) | Updates to R1 services and procedures for RF Channel Reconfiguration use case |
| O-RAN.WG2.Non-RT-RIC-ARCH | Reviewing procedures to implement RF Channel Reconfiguration use cases and generate corresponding A1 policies/updates of O1 and O2-  related services via R1 interface, if any. |
| O-RAN.WG2.Use-Case-Requirements | Specifying RF Channel Reconfiguration use case and its requirements in WG2 UCR specification |
| O-RAN.WG2.A1GAP O-RAN.WG2.A1TD | Reviewing requirement for policy driven implementation 2nd deployment option of RF Channel Reconfiguration use case captured in document. |
| 3 | WG3  (Near-RT RIC, E2) | O-RAN.WG3.UCR | Specifying RF Channel Reconfiguration use case and its requirements in WG3 UCR specification |
| O-RAN.WG3.RICARCH | No impact identified |
| O-RAN.WG3.E2GAP | No impact identified |
| O-RAN.WG3.E2AP | No impact identified |
| O-RAN.WG3.E2SM-RC or NEW: O-RAN.WG3.E2SM-CC | Identify and specify RAN E2 actions necessary for RF Channel Reconfiguration use case specially modifying mMIMO layer, Power, Tx/Rx Array selections, number of SSB Beams through E2 Node  towards O-RU. |
| O-RAN.WG3.E2SM-KPM | Identify and specify RAN E2 measurement required analysis of ES and EC for RF Channel Reconfiguration optimization through Near-RT RIC |
| 4 | WG4 (O-FH)  Impact | O-RAN.WG4.MP | Identify the relevant impacts on M-Plane for both Hierarchical and hybrid model to accommodate management features requirements towards O-RU. Define O-RU Energy efficiency KPIs and counters. |
| O-RAN-WG4.CUS | Identify the relevant impacts on CUS-Plane and data model to support various Tx/Rx Array selections. |
| 5 | WG5 (O1)  Impact | O-RAN.WG5.O-DU-O1 O-RAN.WG5.O-CU-O1 O-RAN.WG5.MP | Identify specific O-DU operational and data model aspects of the feature content including the interface between SMO and -O-DU, and the one between SMO and O-CU. Make appropriate changes to the  O-DU data model and other WG5 specifications as needed. |
| 6 | WG7 | - | WG7 needs to provide reference architecture of O- RU which supports Tx/Rx Array selection.  Study energy savings gains from Tx/Rx array selection including limiting maximum number of layers. |
| 7 | WG10 | O-RAN.WG10.O1-Interface  O-RAN.WG10.OAM-Architecture  O-RAN.WG10.Information Model and Data Models | Identify the relevant impacts on O1 interface to support RF Channel Reconfiguration use case and IM/DM to capture requirements for use cases. |

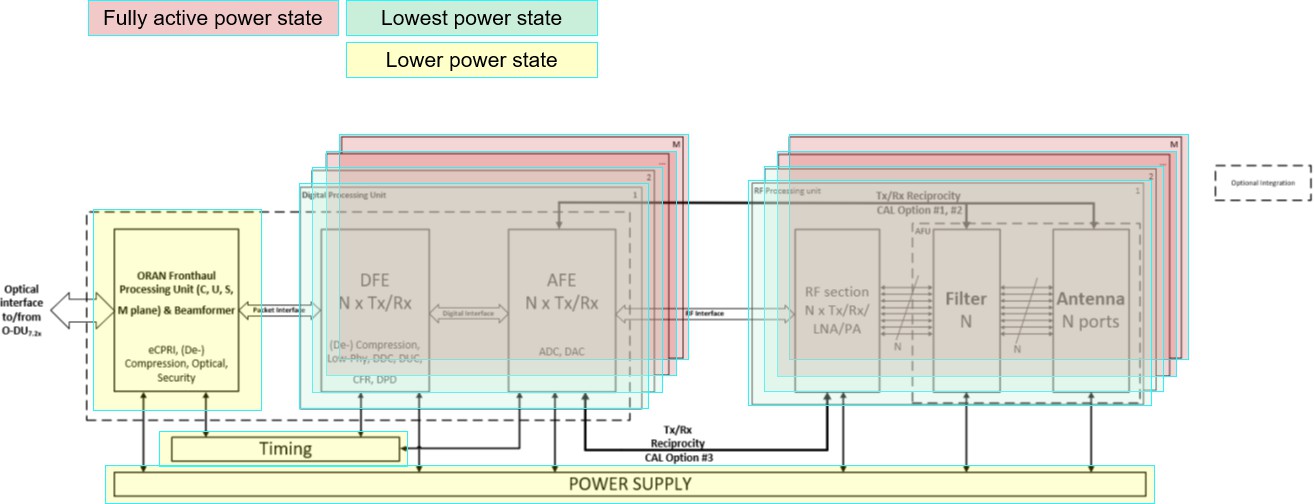
## Relation and Impact on 3GPP Specifications

3GPP RAN does a Rel.18 study on network energy savings for NR (i.e. FS\_Netw\_Energy\_NR, see Sec. 5.4.2 for details). Within this study 3GPP RAN WGs consider related techniques such as RU Tx/Rx array selection, modification of number of SSB beams, adaptation of RU transmit power etc.

## Gain Analysis

The potential of energy savings from RF Channel Reconfiguration is dependent upon several factors such as network deployment, traffic pattern/load, user distribution, antenna type and specific O-RU architecture etc. Computational analysis results for two example scenarios, one for 4T4R O-RU and one for 64T64R O-RU, are provided below. The calculations are based on the load profile and the power consumption categorization of O-RU functional blocks as outlined in Annex B.

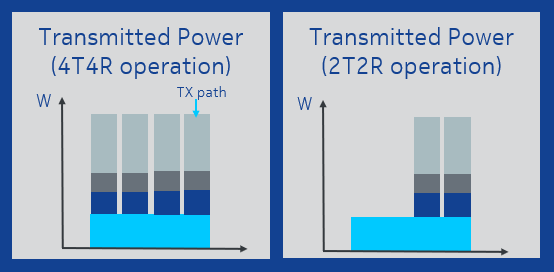
Energy savings gain from RF Channel Reconfiguration would be equivalent to power consumption of the O- RU hardware components that can be shut down or put into energy savings mode when the number of active antennas is reduced. As illustrated in Figure 6.5-1, energy savings gain would mostly be derived from partial shutdown of the RF Processing Unit and Digital Processing Unit (depending on the antenna configuration), while other functional modules such as O-RAN Fronthaul Processing Unit, Power Supply Unit and some other components would still be fully or partially functioning, and the power consumptions would be more or less consistent irrespective of the number of active antennas.



##### Figure 6.5-1: Energy savings for RF Channels

### RF Channel Reconfiguration ES Gain Analysis for 4T4R O-RU

When the traffic load is low, a 4T4R O-RU might be scaled down to 2T2R such that some part of the RF components in the O-RU might be switched off to save energy. When scaling down from 4T4R to 2T2R, RF Processing Unit for the two RF channels to be switched off might be shut down as illustrated in Figure 6.5.1- 1.



##### Figure 6.5.1-1: O-RU power saving for RF Channel Reconfiguration

The energy saving gain from reducing the number of RF channels from 4T4R to 2T2R is analyzed based on the system parameters in Table 6.5.1-1 and the example O-RU power profile considering power consumption from operational experience provided in Table 6.5.1-2.

##### Table 6.5.1-1: O-RU Configuration for ES gain analysis

|  |  |
| --- | --- |
| **No. of antennas** | 4T4R |
| **No. of layers** | 4 |
| **Bandwidth** | 100 MHz |
| **Carrier frequency** | 3.5 GHz |
| **Tx power per array element** | 30 W |
| **Technology** | 5G NR |

##### Table 6.5.1-2: Example power profile for 4T4R O-RU

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Operating Load (Traffic)** | **Configuration** | **Total O- RU (W)** | **ORAN**  **Fronthaul**  **Processing Unit (W)** | **Digital Processing Unit (W)** | **RF**  **Processing Unit (W)** | **Power Unit & other components**  **(W)** |
| Busy hour Load | 4T4R | 550 | 10 | 25 | 495 | 20 |
| Low load | 4T4R | 200 | 10 | 25 | 145 | 20 |
| Low load | 2T2R  RF Channel Reconfiguration ES Mode | 120 | 8 | 20 | 75 | 17 |
| Maximum energy saving gain | 2T2R vs. 4T4R  (low load) | 80 | 2 | 5 | 70 | 3 |

Compared to the low load scenario, a maximum energy savings of up to 80 W per O-RU can be achieved. Achievable energy savings might be lower and might vary in practice. Besides an impact on receive signal quality and spectral efficiency, also coverage and beam characteristics will be affected. A lower spectral efficiency may for instance increase overall transmission time of the active UEs when transmitting a packet.

An example calculation for yearly energy saving potential considering a shutdown of 10000 O-Rus during 3 hours per day is provided in Table 6.5.1-3.

##### Table 6.5.1-3: Power saving calculation example for 4T4R O-RU

|  |  |
| --- | --- |
| **Factor** | **Value** |
| RU power savings during RF channel switch off (low load) [W] | 80 |
| Number of hours of switch off per day (50% of low load period) | 3 |
| Number of O-RUs | 10000 |
| Yearly energy saving [MWh] | 876 |

### RF Channel Reconfiguration ES Gain Analysis for 64T64R O-RU

The energy saving gain from RF Channel Reconfiguration of a 64T64R O-RU to 32T32R is analyzed based on the system parameters in Table 6.5.2-1 and the example O-RU power profile considering power consumption from operational experience provided in Table 6.5.2-2.

##### Table 6.5.2-1: O-RU Configuration for ES gain analysis

|  |  |  |
| --- | --- | --- |
|  | Normal Mode | RF Channel Reconfiguration ES |
| **Configuration** | 64T64R | 32T32R |
| **No. of layers** | 16 | 8 |
| **Bandwidth** | 100 MHz | 100 MHz |
| **Carrier frequency** | 3.5 GHz | 3.5 GHz |
| **Tx power per array element** | 3 W | 3 W |

##### Table 6.5.2-2: Example power profile for 64T64R O-RU

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Operating Load (Traffic)** | **O-RU RF**  **Channel Configuration** | **Total O-RU Power Consumption (W)** | **ORAN**  **Fronthaul Processing Unit (W)** | **Digital Processing Unit (W)** | **RF**  **Processing Unit (W)** | **Power Supply & other components (W)** |
| Busy hour load | 64T64R | 1120 | 50 | 200 | 800 | 70 |
| Low load | 64T64R | 520 | 50 | 200 | 200 | 70 |
| Low load | 32T32R  RF Channel Reconfiguration ES Mode | 290 | 35 | 100 | 100 | 55 |
| Maximum energy saving gain | 64T64R vs.  32T32R  (low load) | 230 | 15 | 100 | 100 | 15 |

Compared to the low load scenario, a maximum energy saving gain of up to 230 W per O-RU can be achieved. As stated before, the achievable energy savings might be lower and might vary in practice.

An example calculation for yearly energy saving potential considering a shutdown of 10000 O-RUs during 3 hours per day is provided in Table 6.5.2-3.

##### Table 6.5.2-3: Power saving calculation example for 64T64R O-RU

|  |  |
| --- | --- |
| **Factor** | **Value** |
| RU power savings during RF channel switch off (low load) [W] | 230 |
| Number of hours of switch off per day (50% of low load period) | 3 |
| Number of O-RUs | 10000 |
| Yearly energy saving [MWh] | 2519 |

## Feasibility Analysis

### Continuous operation during RF Channel Reconfiguration

Changing configuration of O-RU such as shutting down 32 TRX of the 64TRX for the purpose of energy savings may impact the existing users if it requires full or partial reset of O-RU, hence care should be exercised on how the reconfiguration of antenna array is performed.

### Impact on Coverage

The coverage during TRX switch should be maintained as much as possible. A reduction of transceiver chains may impact the number of transmitted SSB beams, the beam shapes, as well as the overall transmit power, affecting the various downlink physical channels. A limited coverage might result in reduced user throughput, radio link failures, beam failures, and call drops.

### Impact and Relation to UE specific Base Station Algorithms

A TRX switch might impact the cell coverage and the beam characteristics. This sudden change in radio channel characteristics might affect UE specific base station algorithms (e.g., inner and outer loop link adaptation, power control, beam selection, adaptive MIMO, channel quality reporting etc.) and may result in a transitory phase until transmission can be continued or is stabilized.

It is up to the proprietary scheduler algorithm to handle such events most efficiently. Scheduling (e.g. user selection, resource allocation), adaptive MIMO (e.g. MIMO mode, spatial streams and layers) and link adaptation (e.g. Coding/Modulation Scheme, transmit power) will be adjusted instantaneously every scheduling instance and transmission time interval by the base station without time constrained external control.

### Limited O-RU / O-DU Capabilities

The capabilities of the O-RU in terms of configurable antenna arrays, transition times, as well as the capabilities of the O-DU scheduling respective antenna arrays with the various MIMO transmission modes may be limited. Hence, respective capabilities are required to be known. The capabilities between the different array configurations may differ significantly in terms of supported SU/MU MIMO modes as well as in terms of supported MIMO layers, streams etc.

# Advanced Sleep Mode Selection

## Problem Statement, Solution and Value Proposition

This Use Case describes a method to achieve intelligent energy saving by optimizing the sleep mode via RIC-based guidance.

##### Background

O-RU and E2 Nodes (O-CU, O-DU) may implement various Sleep Modes (SMs). The SMs are enabled by Non-RT RIC/SMO and/or Near-RT RIC. When enabled, the E2 Nodes select among the SMs considering their capabilities, the actual traffic situation, and the network conditions. Different SM operate at different time scales (e.g., symbol, slot, frame).

##### Problem Statement

In a single vendor network where components are not disaggregated, the O-RU's capabilities are well-known within the E2 Node, allowing for autonomous execution of sleep modes. However, implementing sleep modes in O-RAN requires the O-DU and O-CU to understand the sleep mode capabilities of the O-RU, including the number and depth of sleep modes. Also, Near/Non-Real-Time RIC need to be aware of the O- DU and O-CU capabilities to drive policies related to sleep modes. For optimization, O-RU and O-DU need to report energy efficiency related measurements and KPIs from which the power saving can be observed. To achieve this, effective communication and coordination between all components is necessary.

##### Solution

In this solution the O-RU will expose its SM capabilities to the O-DU in the initial configuration phase. Next, the O-DU will expose a set of commonly supported SMs to the SMO (over O1) and Non-RT RIC (via SMO and R1) and/or the Near-RT RIC (over E2) which is available for utilization during operation.

Information exposed towards the Non-RT and/or Near-RT RIC will include a unique identifier of each SM, and additional operational parameters, e.g., minimum duration of activation, transition times between SMs, whether to be applied on UL/DL, etc.

Moreover, the Non-RT RIC and/or Near-RT RIC will collect for the different SMs data related to network load and performance, cell configuration, and energy/power consumption.

Based on the above two sets of information the Non-RT RIC and/or Near-RT RIC train an ML model, which will infer optimized SM utilization range (i.e., a set of SMs the O-DU may dynamically select from at a given or pre-defined time), which is then transmitted to the O-DU as a guidance. Based on the optimized (allowed) SM utilization range, O-DU will select specific SMs.

Depending on the SMs, O-RU and O-DU may apply such SM selection internally or O-RU may apply a SM based on O-DU request. For instance, O-RU may (or may not) apply a micro sleep internally, while O-DU may (or may not) internally adjust its scheduling strategy to maximize possible O-RU micro sleep periods considering O RU SM capabilities. Energy saving is thus obtained during regular operation when O-DU schedules data in a more optimized way. O-DU may for example prioritize time-domain scheduling over frequency domain scheduling or may compress data transmission to increase the number of symbols or slots without data. O-DU cannot enforce the O-RU to apply its micro sleep due to the fast internal operation and Near-RT RIC may not be able to enforce O-DU to adjust its internal scheduling strategy. The energy saving gains can be observed by monitoring respective Energy Efficiency KPIs.

In the alternative on request operation, O-DU may for instance configure the O-RU to switch off a certain part of the O-RU to support a deep sleep operation considering the timing constraints as communicated

during the capability exchange. Energy saving is thus obtained when O-RU applies a new configuration as requested by O-DU.

##### Value Proposition

Firstly, the Non-RT RIC and/or Near-RT RIC based AI/ML algorithm may extend the time period or increase the granularity during which the E2 Nodes are allowed to select from their implemented SMs.

Secondly, this solution allows for an optimized SMs selection in a multi-vendor O-RU/O-DU deployment, by the O-RU/O-DU exposing the required information as part of their SM capabilities.

## Architecture/Deployment Options

### Option 1: Training and Inference in Non-RT RIC

In this deployment option, the Non-RT RIC retrieves the E2 Node SM capabilities over O1 and R1 via SMO, collects the necessary data from the E2 Nodes and trains an ML model, which is then deployed in the rApp. Inference of the optimized SM utilization range is provided as guidance by the rApp. Additionally, external parties (such as prediction rApps or external application servers) may provide additional information for optimization.

#### Description and UML Diagram

##### Table .2.1.1-1: Advanced Sleep Mode Selection: Training and Inference in Non-RT RIC

|  |  |  |
| --- | --- | --- |
| **Use Case Stage** | **Evolution / Specification** | **<<Uses>>**  **Related use** |
| Goal | Provide optimized SM utilization range as guidance to the O-DU and allow  for optimized SM selection in O-DU / O-RU. |  |
| Actors and Roles | * SMO/Non-RT RIC: Retrieval of E2 Node SM capabilities. * SMO/Non-RT RIC: Data collection for ML model training and ML model deployment. * Non-RT RIC: Network data collection and analysis, and inference of SM utilization range guidance provided to O-DU. * O-DU: Expose common O-DU/O-RU SM capabilities over O1. * O-RU: Expose SM related capabilities to O-DU. * (Optional) Additional parties: Providing additional data for optimization. |  |
| Assumptions | - O-RU and O-DU expose the necessary information about their available SMs capabilities. |  |
| Pre- conditions | - All relevant functions and components are instantiated and available. |  |
| Begins when | The operator has enabled or has set targets for SM utilization guidance  rApp. |  |
| Step 1.1, 1.2 (M) | O-RU transmits to O-DU its SM capabilities. O-DU decides on a common  set of O-DU/O-RU SM capabilities to utilize during operation. |  |
| Step 1.3 (M) | O-DU transmits information about the available O-DU/O-RU SM capabilities  to Non-RT RIC/SMO, including minimal operational information. |  |
| Step 2.1, 2.2 (M), 2.3 (O)  Step 2.4, 2.5 (M), 2.6 (O)  Step 2.7 (M) | Data collection request from SMO to E2 Nodes (M) and external entities (O) for training and inference.  Data collection from E2 Nodes (M) and external entities (O) for training in SMO/Non-RT RIC.  Non-RT RIC retrieves data for training and inference. |  |
| Step 3.1, 3.2, 3.3 (M) | Based on continuous/periodically collected data, Non-RT RIC trains the AI/ML model and deploys it. Non-RT RIC performs inference using the  AI/ML model. |  |
| Step 4.1 (M)  Step 4.2, 4.3 (O),  Step 4.4, 4.5, 4.6 (O) | Non-RT RIC provides SM utilization guidance via SMO and O1 to O-DU.  O-DU considers Non-RT RIC’s SM utilization guidance in its updated scheduling strategy and O-RU applies SM selection internally.  Based on Non-RT RIC’s SM utilization guidance, O-DU alternatively selects a SM, O-DU requests a new SM configuration by O-RU over O-FH and O- RU applies the new SM configuration. |  |
| Ends when | Operator disables or changes targets for SM utilization guidance rApp. |  |
| Exceptions | None identified. |  |
| Post-conditions | Non-RT RIC continues to monitor the energy consumption and efficiency and the RAN performance by collecting and monitoring the relevant data over O1. |  |

@startuml

skinparam defaultFontSize 12 autonumber

Box "SMO" #gold

Participant “Collection & Control” as COLL Participant “Non-RT RIC” as NON

End box

Box "O-RAN" #lightpink Participant "O-CU" as OCU Participant "O-DU" as ODU

End box

Box "O-RU" #turquoise Participant “O-RU” as ORU

End box

Box "EXT" #lightcyan Participant “EXT” as EXT

End box

== Initialization == autonumber 1.1

ORU -> ODU : <<O-FH>> SM capabilities ODU -> ODU : SM capability decision

ODU -> COLL : <<O1>> SM capability information exposure ODU -> NON : <<E2>> SM capability information exposure

== Data Collection == autonumber 2.1

Group Data for training and for inference Loop

COLL -> OCU : <<O1>> Data collection request COLL -> ODU : <<O1>> Data collection request COLL -> EXT : Data collection request

OCU -> COLL : <<O1>> Data retrieval ODU -> COLL : <<O1>> Data retrieval EXT -> COLL : Data retrieval

COLL -> NON : Data retrieval End loop

End group

== AI/ML Flow == autonumber 3.1

NON -> NON : AI/ML model training NON -> NON : AI/ML model deployment

Loop

NON -> NON : AI/ML model inference

End Loop

== SM Guidance (Loop) == autonumber 4.1

NON -> COLL: Guidance to utilize SM

COLL -> ODU: <<O1>> Guidance to utilize SM

group Configuration Update group alt1

ODU -> ODU: Updated scheduling strategy ORU -> ORU: Internal SM selection

end

group alt2

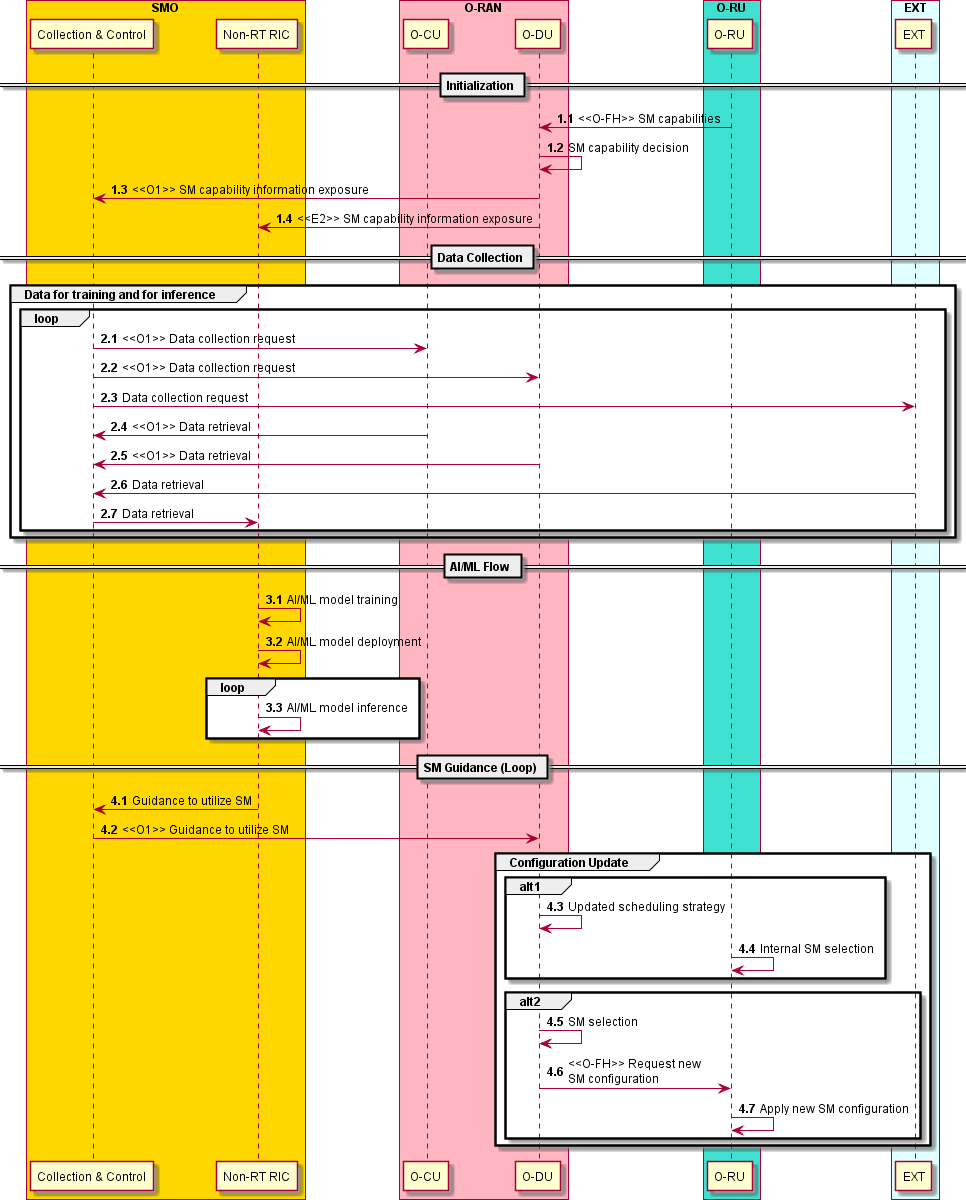
ODU -> ODU: SM selection

ODU -> ORU: <<O-FH>> Request new \nSM configuration ORU -> ORU: Apply new SM configuration

end

end

@enduml



##### Figure .2.1.1-1: Flow diagram for SM Selection, ML model training in the Non-RT RIC

#### O-RAN Entity Roles

1. SMO (including Non-RT RIC)
   1. Receive common SM capability information and additional operational parameters from O-DU.
   2. Subscribe to and retrieve necessary performance indicators (incl. Energy Efficiency KPIs), measurement reports, UE context information, RAN configurations, and SM usage data from E2 Nodes via the O1 interface for the purpose of AI/ML model training, inference, and performance monitoring.
   3. Optionally, i) collect enrichment information from Application servers and associate enrichment information with collected measurements and configurations, ii) collect prediction or optimization related information from other rApps.
   4. Perform AI/ML model training and deployment.
   5. Send SM utilization guidance for Sleep Mode optimization to E2 Nodes over O1.
2. E2 Nodes (O-DU in disaggregated architecture)
   1. Support retrieving the O-RU SM capabilities and additional operational parameters from O-RU via O-FH.
   2. Support reporting the common O-DU/O-RU SM capabilities and additional operational parameters to SMO over O1.
   3. Support reporting of necessary performance indicators (incl. Energy Efficiency KPIs), measurement reports, UE context information, RAN configurations, and SM usage data with required granularity to SMO via the O1 interface.
   4. Receive SM utilization guidance from the SMO via the O1interface.
   5. Adjust scheduling strategy (to allow O-RU to internally update its used SMs configuration) or alternatively perform SM selection based on SM utilization guidance received from the SMO.
   6. Optionally, request O-RU over O-FH to update its used SM configuration (e.g., switch off a certain O-RU functionality).
3. O-RU
   1. Support reporting the O-RU SMs capabilities and additional operational parameters to O-DU via O- FH.
   2. Internally apply SM selection or alternatively receive over O-FH and apply O-DUs request for updated SM configuration.

#### Void

### Option 2: Training in Non-RT RIC and Inference in Near-RT RIC

In this deployment option, the Non-RT RIC retrieves the E2 Node SM capabilities over O1 and R1 via SMO, collects the necessary data from the E2 Nodes and trains an ML model, which is then deployed in the rApp. Inference of the optimized SM utilization range is provided as guidance by the rApp. Additionally, external parties (such as prediction rApps or external application servers) may provide additional information for optimization.

#### Description and UML Diagram

##### Table .2.2.1-1: Advanced Sleep Mode Selection: Training in Non-RT RIC and Inference in Near-RT RIC

|  |  |  |
| --- | --- | --- |
| **Use Case Stage** | **Evolution / Specification** | **<<Uses>>**  **Related use** |
| Goal | Provide optimized SM utilization range as guidance to the O-DU and allow  for optimized SM selection in O-DU/O-RU. |  |
| Actors and Roles | * SMO/Non-RT RIC: Retrieval of E2 Node SM capabilities. * SMO/Non-RT RIC: Data collection for ML model training and ML model deployment. * Near-RT RIC (xApp): Network data collection and analysis, and inference of SM utilization range guidance provided to O-DU. * O-DU: Expose common O-DU/O-RU SM capabilities over O1/E2. * O-RU: Expose SM related capabilities to O-DU. * (Optional) Additional parties: Providing additional data for optimization. |  |
| Assumptions | - O-RU and O-DU expose the necessary information about their available SMs capabilities. |  |
| Pre- conditions | - All relevant functions and components are instantiated and available. |  |
| Begins when | The operator has enabled or has set targets for SM utilization guidance  xApp. |  |
| Step 1.1, 1.2 (M) | O-RU transmits to O-DU its SM capabilities. O-DU decides on a common  set of O-DU/O-RU SM capabilities to utilize during operation. |  |
| Step 1.3, 1.4 (M) | O-DU transmits information about the available O-DU/O-RU SM capabilities to Non-RT RIC/SMO and Near-RT RIC, including minimal operational  information. |  |
| Step 2.1, 2.2 (M), 2.3 (O)  Step 2.4, 2.5 (M), 2.6 (O)  Step 2.7 (M)  Step 2.8, 2.9 (M), 2.10 (O)  Step 2.11, 2.12 (M),  2.13 (O) | Data collection request from SMO to E2 Nodes (M) and external entities (O) for training.  Data collection from E2 Nodes (M) and external entities (O) for training in SMO.  Non-RT RIC retrieves data for training.  Data collection request from Near-RT RIC to E2 Nodes (M) and external entities (O) for inference.  Data collection from E2 Nodes (M) and external entities (O) to Near-RT RIC for inference. |  |
| Step 3.1, 3.2, 3.3 (M) | Based on continuous/periodically collected data, Non-RT RIC trains the AI/ML model and deploys it to Near-RT RIC via O1/O2. Near-RT RIC  performs inference using the AI/ML model. |  |
| Step 4.1 (M)  Step 4.2, 4.3 (O),  Step 4.4, 4.5, 4.6 (O) | Near-RT RIC provides SM utilization guidance via E2 to O-DU.  O-DU considers Near-RT RIC’s SM utilization guidance in its updated scheduling strategy and O-RU applies SM selection internally.  Based on Near-RT RIC’s SM utilization guidance, O-DU alternatively selects a SM, O-DU requests a new SM configuration by O-RU over O-FH and O- RU applies the new SM configuration. |  |
| Ends when | Operator disables or changes targets for SM utilization guidance xApp. |  |
| Exceptions | None identified. |  |
| Post-conditions | Near-RT RIC/Non-RT RIC continues to monitor the energy consumption and efficiency and the RAN performance by collecting and monitoring the relevant data over E2/O1. |  |

@startuml

skinparam defaultFontSize 12 autonumber

Box "SMO" #gold

Participant “Collection & Control” as COLL Participant “Non-RT RIC” as NON

End box

Box "O-RAN" #lightpink

Participant “Near-RT RIC” as NEAR Participant "O-CU" as OCU Participant "O-DU" as ODU

End box

Box "O-RU" #turquoise Participant “O-RU” as ORU

End box

Box "EXT" #lightcyan Participant “EXT” as EXT

End box

== Initialization == autonumber 1.1

ORU -> ODU : <<O-FH>> SM capabilities ODU -> ODU : SM capability decision

ODU -> COLL : <<O1>> SM capability information exposure ODU -> NEAR : <<E2>> SM capability information exposure

== Data Collection == autonumber 2.1

Group Data for training

COLL -> OCU : <<O1>> Data collection request COLL -> ODU : <<O1>> Data collection request COLL -> EXT : Data collection request

OCU -> COLL : <<O1>> Data retrieval ODU -> COLL : <<O1>> Data retrieval EXT -> COLL : Data retrieval

COLL -> NON : Data retrieval End group

Group Data for inference

NEAR -> OCU : <<E2>> Data collection request NEAR -> ODU : <<E2>> Data collection request NEAR -> EXT : Data collection request

Loop

OCU -> NEAR : <<E2>> Data retrieval ODU -> NEAR : <<E2>> Data retrieval EXT -> NEAR : Data retrieval

End Group End Loop

== AI/ML Flow == autonumber 3.1

NON -> NON : AI/ML model training

NON -> NEAR : <<O1/O2>> AI/ML model deployment

Loop

NEAR -> NEAR: AI/ML model inference

End Loop

== SM Guidance (Loop) == autonumber 4.1

NEAR -> ODU: <<E2>> Guidance to utilize SM

group Configuration Update group alt1

ODU -> ODU: Updated scheduling strategy

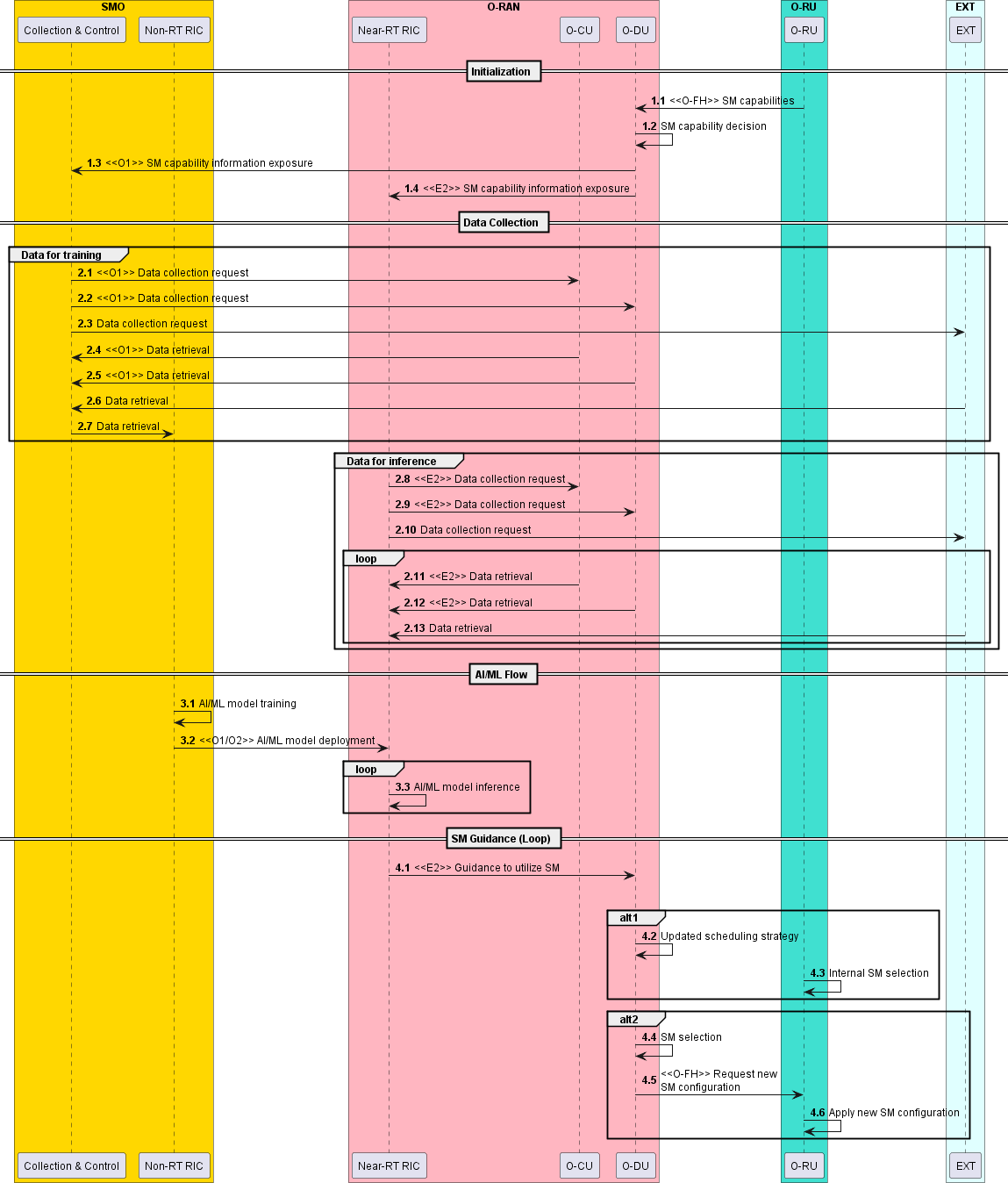
ORU -> ORU: Internal SM selection end

group alt2

ODU -> ODU: SM selection

ODU -> ORU: <<O-FH>> Request new \nSM configuration ORU -> ORU: Apply new SM configuration

end @enduml



##### Figure .2.2.1-1: Flow diagram for SM Selection in the Near-RT RIC, ML model training in the Non-RT RIC

#### O-RAN Entity Roles

1. SMO (including Non-RT RIC)
   1. Receive common SM capability information and additional operational parameters from O-DU.
   2. Subscribe to and retrieve necessary performance indicators (incl. Energy Efficiency KPIs), measurement reports and RAN configurations from E2 Nodes via the O1 interface for the purpose of AI/ML model training and performance monitoring.
   3. Perform AI/ML model training and deployment.
   4. Optionally, i) collect enrichment information from Application servers and associate enrichment information with collected measurements and configurations, ii) collect prediction or optimization related information from other rApps.
2. Near-RT RIC (xApp)
   1. Receive common SM capability information and additional operational parameters from O-DU.
   2. Subscribe to and retrieve necessary performance indicators (incl. Energy Efficiency KPIs), measurement reports, UE context information, RAN configurations, and SM usage data from E2 Nodes via the E2 interface for the purpose of AI/ML model inference, and performance monitoring.
   3. Optionally, i) collect enrichment information from Application servers and associate enrichment information with collected measurements and configurations, ii) collect prediction or optimization related information from other xApps (Near-RT RIC internally) or rApps.
   4. Send SM utilization guidance via a policy message for Sleep Mode optimization to E2 Nodes via the E2 interface.
3. E2 Nodes (O-DU in disaggregated architecture)
   1. Support retrieving the O-RU SM capabilities and additional operational parameters from O-RU via O-FH.
   2. Support reporting the common O-DU/O-RU SM capabilities and additional operational parameters to Near-RT RIC over E2 and to SMO over O1.
   3. Support reporting of necessary performance indicators (incl. Energy Efficiency KPIs), measurement reports, UE context information, RAN configurations, and SM usage data with required granularity to Near-RT RIC via the E2 interface and SMO via the O1 interface.
   4. Receive SM utilization guidance via policy message from the Near-RT RIC via the E2 interface.
   5. Adjust scheduling strategy (to allow O-RU to internally update its used SMs configuration) or alternatively perform SM selection based on SM utilization guidance received from the Near-RT RIC.
   6. Optionally, request O-RU over O-FH to update its used SM configuration (e.g., switch off a certain O-RU functionality).
4. O-RU
   1. Support reporting the O-RU SMs capabilities and additional operational parameters to O-DU via O- FH.
   2. Internally apply SM selection or alternatively receive over O-FH and apply O-DUs request for updated SM configuration (e.g., switch off a certain O-RU functionality).

#### Void

## Impact Analysis on O-RAN Work Groups

This is an initial impact analysis as part of the WG1 UCTG Network Energy Saving work on Advance Sleep Mode Selection use case. The intention is to estimate the expected standardization effort within the O-RAN working groups. It is up to the WGs to decide how Advance Sleep Mode Selection use case functionality should be specified in specifications of each WG.

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **WGs/FGs** | **Spec. No** | **Objective description** |
| **1** | WG1 (Use  case) | O-RAN.WG1.NES-USE-CASES-TR  O-RAN.WG1.Use-Cases-Detailed- Specification | Update WG1 NES use case analysis report and use- case detailed specification with Advance Sleep Mode Selection use case. No impact to existing  architecture |
| **2** | WG2 (Non-RT RIC, A1, R1) | O-RAN.WG2.R1GAP O-RAN.WG2.R1UCR  O-RAN.WG2.R1TD (TBD) | Updates to R1 services and procedures for Advanced Sleep Mode Selection use case |
| O-RAN.WG2.Non-RT-RIC-ARCH | Reviewing procedures to implement Advanced Sleep Mode Selection use case and generate corresponding A1 policies/updates of O1 and O2-related services  via R1 interface, if any. |
| O-RAN.WG2.Use-Case-Requirements | Specifying Advanced Sleep Mode Selection use case and its requirements in WG2 UCR specification |
| O-RAN.WG2.A1GAP O-RAN.WG2.A1TD | Reviewing requirement for policy driven implementation of Advance Sleep Mode Selection |
| 3 | WG3  (Near-RT RIC, E2) | O-RAN.WG3.UCR | Specifying Advanced Sleep Mode Selection use case and its requirements in WG3 UCR specification |
| O-RAN.WG3.RICARCH | No impact identified |
| O-RAN.WG3.E2GAP | No impact identified |
| O-RAN.WG3.E2AP | No impact identified |
| O-RAN.WG3.E2SM-RC or NEW: O-RAN.WG3.E2SM-CC | Identify and specify RAN E2 actions necessary for  implementation of Advance Sleep Mode Selection from E2 Node towards O-RU. |
| O-RAN.WG3.E2SM-KPM | Identify and specify RAN E2 measurement required analysis of ES and EC for Advance Sleep Mode Selection through Near-RT RIC |
| 4 | WG4 (O-FH)  Impact | O-RAN.WG4.MP | Identify the relevant impacts on M-Plane for both hierarchical and hybrid model to accommodate management features requirements towards O-RU. Define O-RU Energy efficiency KPIs and counters. |
| O-RAN-WG4.CUS | Identify the relevant impacts on CUS-Plane and data model to support various Advance Sleep Mode  Selection use case. |
| 5 | WG5 (O1)  Impact | O-RAN.WG5.O-DU-O1 O-RAN.WG5.O-CU-O1 O-RAN.WG5.MP | Identify specific O-DU operational and data model aspects of the feature content including the interface between SMO and O-DU, and the one between SMO and O-CU. Make appropriate changes to the O-DU data model and other WG5 specifications as  needed. |
| 7 | WG7 | WG7 Energy Savings TR | WG7 needs to provide reference architecture of O- RU which supports Advance Sleep Mode Selection use case. |

|  |  |  |  |
| --- | --- | --- | --- |
| 8 | WG10 | O-RAN.WG10.O1-Interface  O-RAN.WG10.OAM-Architecture  O-RAN.WG10.Information Model and Data Models | Identify the relevant impacts on O1 interface to support Advanced Sleep Mode Selection use cases and IM/DM to capture requirements. |

## Relation and Impact on 3GPP Specifications

Within the 3GPP Rel.18 RAN WG1 led study item on network energy saving 3GPP agreed on three Power States for evaluation purpose Micro, Light and Deep Sleep as documented in TR 38.864 [19]. The states represent a simplified model for simulation and evaluation purpose only and have not been part of the normative work in Rel.18. The examples of Advanced Sleep Modes are shown in Annex D.

|  |  |  |  |
| --- | --- | --- | --- |
| **Reference** | **Release** | **Title** | **Documentation and impact on Specifications** |
| **FS\_Netw\_Energy\_NR** | 5G Rel.18 | Study on network energy savings for NR | Results will be captured in 3GPP TR 38.864 [19]. |

3GPP TSG RAN 98-e approved a new Rel.18 WID: Network energy savings for NR (RP-223540) which includes following objective.

1. *Specify enhancement on cell DTX/DRX mechanism including the alignment of cell DTX/DRX and UE DRX in RRC\_CONNECTED mode, and inter-node information exchange on cell DTX/DRX*

*[RAN2, RAN1, RAN3]*

* + *Note: No change for SSB transmission due to cell DTX/DRX.*
  + *Note: The impact to IDLE/INACTIVE UEs due to the above enhancement should be avoided.*

While 3GPP standards concentrate on enhancing the co-ordination between UE and gNB for cell DTX/DRX operation, O-RAN Advanced Sleep Mode Selection focuses on co-ordinating O-RU and O-DU based on Non-RT/Near-RT RIC policy.

## Void

## Feasibility Analysis

### Impact to Continuous Operation during Advance Sleep Modes

Advanced Sleep Modes that are limited to symbol level are not expected to have a significant impact on user performance. However, longer sleep modes that involve shutting down more components or reducing their activity for longer periods of time may impact user performance such as increased latency and/or reduced data transfer rates.

### Impact to Coverage

If an O-RU transmission is suspended temporarily (e.g., some kind of deep sleep modes), UEs may be unable to connect to gNB due to missing synchronization, measurements and/or system information, hence the coverage in the area may be impacted. The impact might be mitigated by intelligent configuration and/or using multiple carriers.

### Impact and Relation to Vendor Specific Scheduling Algorithms

O-RU and O-DU may implement their own sleep modes. However, adjustments may be needed to align with ES/EE policies from Non-RT/Near-RT RIC. Operators need to balance user/cell performance against energy savings, and E2 Nodes (O-DU/O-CU) may require additional functionality to implement these policies.

### Limited O-RU/O-DU Capabilities

The O-RU and O-DU may have limitations in terms of implementing sleep modes capabilities. It is up to implementation how many Advanced Sleep Modes are supported in O-RU or O-DU. Furthermore, the generic solution should support the transmission of operational parameters from O-RU to O-DU, e.g., minimum duration of activation of SMs, transition times between sleep modes, whether to be applied on UL/DL, etc.

# O-Cloud Resource Energy Saving Mode

In O-RAN split option 7.2x major signal processing is performed by O-DU and O-CU. Hence, O-DU and O- CU hardware components consume energy to maintain a certain level of system availability even in case of low or no traffic load on VNFs or CNFs. As more and more operators adopt virtualized or containerized architecture, there is a strong need to have solution(s) for energy saving in the O-Cloud. The O-Cloud Resource Energy Saving Mode use case should cover various components of the O-Cloud such as CPU/GPU, accelerators, NIC cards and other components at node level.

The aim of this use case is to enable energy savings in the O-Cloud by reducing the power consumption of various O-Cloud components without impairing the network performance. Given the network status, the O- Cloud components' power consumption can be optimized through actions such as adaptive shutdown of hardware, scaling out Network Functions, and optimization of CPU/FPGA power, memory usage, CPU frequency, etc. By using multi-dimensional data (e.g., traffic load data at E2 Nodes, load over O-Cloud in terms of compute/storage) the Non-RT RIC can configure changes towards the O-Cloud.

The time scale for the control of O-Cloud Resource Energy Saving Mode solutions is Non-Real Time.

## Sub Use Case 1: O-Cloud Node Shutdown

### Problem Statement, Solution and Value Proposition

This use case describes a method to perform O-Cloud Energy Saving by shutting down physical O-Cloud Node(s) in idle times through Non-RT RIC.

When O-Cloud Node is operating at low load, then the deployed Network Functions or its microservices can be relocated or shut down or blocked from starting on the node in order to free up the node. Moving VNFs or CNFs within O-Cloud and/or evacuating the O-Cloud Nodes are O-Cloud internal functionalities which might be triggered by the SMO.

Idle O-Cloud Nodes can be shut down to reduce energy consumption during low-load times. Non-RT RIC subscribes to O2 data via SMO, which, among others, includes configuration of O-Cloud Nodes (e.g., K8s cluster, resource pools, NF-pod-node associations). Non-RT RIC provides guidance to the SMO. The SMO monitors O-Cloud and E2 Node resources based on O1/O2 data and requests shut down (scale in) or add O- Cloud resources via O2ims (scale out) based on Non-RT RIC guidance. O-Cloud (IMS) will estimate if sufficient O-Cloud resources are remaining or are available to serve respective requests. After execution or rejection of respective requests O-Cloud (IMS) will communicate its actions via O2ims.

Evacuation of Node 1

**NF**

**NF**

**NF**

O-Cloud Node 1 O-Cloud Node O-Cloud Node

Shutdown of Node 1

**NF**

**NF**

**NF**

O-Cloud Node Shutdown

Network Function relocation

**NF**

**NF**

**NF**

**NF**

**NF**

**NF**

O-Cloud Node O-Cloud Node 2 O-Cloud Node

##### Figure 8.1.1-1: O-Cloud energy saving by O-Cloud pod relocation and node shutdown

A Node is the smallest unit of a computing hardware, such as a physical or virtual machine. Programs running on Nodes are packaged as containers. One or more containers form a structure are called a pod. A pod might comprise functionalities supporting one or multiple cells or parts of one or multiple cells. As shown in Figure 8.1.1-1, O-Cloud Node 1 has one VNF pod deployed which can be relocated to the O-Cloud Node 2, which would make O-Cloud Node 1 idle. Post relocation of the VNF pod, the idle O-Cloud Node 1 can be shut down to reduce energy consumption.

Non-RT RIC will use O1 and O2 data to provide guidance for shutdown decision. For example, O-Cloud resource utilization matrices such as CPU, memory, and storage utilization can be analyzed by using O- Cloud Monitoring Service telemetry. Such data can be correlated with RAN related load and energy consumption information obtained per Network Function via the O1 interfaces.

### Architecture/Deployment Option

Background information on O-Cloud architecture and management: Federated O-Cloud Orchestration and Management (FOCOM), Network Function Orchestrator (NFO), Infrastructure Management Services (IMS) and Deployment Management Services (DMS) are defined in O-RAN.WG6.O2-GA&P [20]. The FOCOM is responsible for accounting and asset management of the resources in the cloud. The NFO is responsible for orchestrating the assembly of the network functions as a composition of NF deployments in the O-Cloud. The IMS is responsible for management of the O-Cloud resources and the software which is used to manage those resources, and the DMS is responsible for management of NF deployments into the O-Cloud.

In this deployment option, decision making for O-Cloud Node Shutdown configuration and guidance, potentially including AI/ML model training and inference, is done at the SMO/Non-RT RIC. The overall mechanism is shown in Figure 8.1.2-1.

**SMO**

**O-RU**

Open FH

M-Plane

**E2 Node**

DMS

E2

IMS

A1

O1

O2-DMS

Non-Real Time RIC

& rApp

FOCOM NFO

1. Receive guidance from Non-RT RIC for optimizing energy consumption for O- Cloud resource evacuation and shutdown.
2. Provide optimized configuration for O- Cloud resource evacuation and shutdown towards O-Cloud.

Service, Management & Orchestration

O-Cloud ES Guidance

**Non-RT RIC (rApp)**

1. Collect O-Cloud FCAPS data via SMO over

O2/R1.

1. Collect E2 Node data via SMO over O1/R1.
2. Train Deploy AI/ML Model to generate guidance based on O1 and O2 data.
3. Generate guidance for O-Cloud configuration based on priority, load and energy consumption.

O2-IMS

Near-RT RIC

DMS

DMS

**O-Cloud**

* 1. Expose O2 (IMS/DMS) data towards SMO.
  2. Apply configuration provided by SMO, e.g., execute Node evacuation/shutdown.
  3. Communicate its action towards

SMO/Non-RT RIC.

**O-Cloud**

Open FH M-Plane

##### Figure 8.1.2-1: Cloud Resource Energy Saving via SMO/Non-RT RIC

#### Description and UML Diagram

##### Table 8.1.2.1-1: O-Cloud Node Shutdown: AI/ML inference via Non-RT RIC

|  |  |  |
| --- | --- | --- |
| **Use Case Stage** | **Evolution / Specification** | **<<Uses>>**  **Related use** |
| Goal | Shutting down O-Cloud Nodes during idle times or low load scenario for  energy saving purpose. |  |
| Actors and Roles | * SMO (FOCOM): Receives guidance from Non-RT RIC and generates optimized configurations towards O-Cloud. * SMO (NFO): Receives guidance from Non-RT RIC and generates optimized configurations towards O-Cloud. * Non-RT-RIC (rApp): Trains AI/ML models and generates guidance. * O-Cloud (IMS): Executes changes recommended by SMO/FOCOM. * O-Cloud (DMS): Executes changes recommended by SMO/NFO. |  |
| Assumptions | * The “Service Request” from Non-RT RIC to the FOCOM includes the identifiers of the O-Cloud Node(s) to be evacuated or shut down. * The “Service Request” from Non-RT RIC to the NFO includes the identifiers of the NFs to be modified. * Non-RT RIC (rApp) subscribed to receive notifications from SMO. * SMO (FOCOM and NFO) gets guidance from Non-RT RIC (rApp). |  |
| Pre-conditions | * All relevant functions and components are instantiated and available. * Non-RT RIC can receive configurations, load and performance measurements from E2 Nodes via O1 and telemetry data from O-Cloud via O2. * O-Cloud ensures that relocation of NF pod from one Node to another Node does not impair network services. * O-Cloud ensures that NF Pod shutdown does not impair network services related to that NF. |  |
| Begins when | The operator has enabled or has set targets for O-Cloud Node Shutdown. |  |
| Step 1.1, 1.2, 1.3,  1.4 (M) | Network Function (applications) related data is subscribed to and received over O1 interface, for example traffic load of NF, NF utilization, NF energy  consumption etc. |  |
| Step 2.1, 2.2, 2.3,  2.4, 2.5, 2.6, 2.7, 2.8 (M) | O-Cloud utilization matrices (such as CPU, Memory, Storage utilization) and NF placement are acquired by Non-RT RIC from O-Cloud via SMO framework  over O2 interface. |  |
| Step 3.1, 3.2 (O) | Non-RT RIC trains the AI/ML models with the collected data. Trained AI/ML models are deployed and activated in the Non-RT RIC. Non-RT RIC constantly monitors performance and energy consumption of the O-Cloud Nodes, NF deployments, as well as cell load related and traffic data, EE/EC  measurement reports etc. |  |
| Step 4.1, 4.2, 4.3, 4.4,  4.5, 4.6 (M) | Non-RT RIC provides guidance to FOCOM and NFO to modify NF deployments and to evacuate the specific O-Cloud Node(s).  NFO requests the DMS over O2dms to modify NF deployment.  FOCOM requests the IMS using O2ims interface to evacuate one or more O- Cloud Nodes after modification (shutdown, relocation etc.) of NF deployments. |  |
| Step 5.1, 5.2, 5.3 (O) | IMS notifies SMO that the O-Cloud Node(s) evacuation is completed (or  rejected). |  |
| Step 6.1, 6.2, 6.3, (M) | Non-RT RIC provides guidance to FOCOM, then FOCOM configures IMS to  shutdown Node(s). |  |
| Step 7 (M) | IMS shuts down the requested O-Cloud Node(s). |  |
| Step 8.1, 8.2, 8.3 (M) | IMS notifies SMO that O-Cloud Node(s) shut down is completed. |  |
| Ends when | O-Cloud Node(s) have been shut down. |  |
| Exceptions | None identified. |  |
| Post-conditions | O-Cloud Node(s) may be shut down for energy saving.  SMO/Non-RT RIC continues to monitor the load on the O-Cloud and RAN data by collecting and monitoring the relevant performance and matrices using O1 and O2. |  |

@startuml

skinparam defaultFontSize 12 autonumber

Box “O-Cloud Platform” #lightseagreen participant “IMS” as IMS participant “DMS” as DMS

End box

Box "Service Management & \n Orchestration Framework" #gold Participant “FOCOM” as FOCOM

Participant “NFO” as NFO Participant “OAM Functions” as OAM

Participant " Non-RT RIC & rApp" as nRT End box

Box "O-RAN" #lightpink Participant "E2-Nodes" as E2N

End box

== O1 & O2 Data Collection == autonumber 1.1

note over nRT,OAM

User traffic load on NF(O-DU/O-CU), NF(O-DU/O-CU)Resource utilization etc. data to retreieved

End note

nRT -> OAM : Data subscription request

Loop

OAM -> E2N :<<O1>>Data subscription request E2N -> OAM :<<O1>>Data Retrieval

OAM -> nRT : Data Retrieval

End Loop

Note over nRT,IMS

O2ims & O2dms telemetry, configurations such as CPU,Memory, Storage utilization,O-Cloud Node configurations etc

End note autonumber 2.1

nRT -> FOCOM : Data collection request nRT -> NFO : Data collection request

FOCOM -> IMS : <<O2ims>> Data collection request NFO -> DMS : <<O2dms>> Data collection request

Loop

IMS -> FOCOM : <<O2ims>> Data retrieval DMS -> NFO : <<O2dms>> Data retrieval FOCOM -> nRT : Data retrieval

NFO -> nRT : Data retrieval

End Loop

== AI/ML Flow == autonumber 3.1

nRT -> nRT: AI/ML model training nRT -> nRT: AI/ML model inference

== O-Cloud Node Evacuation (Optional) == Note over nRT,FOCOM

Draning request to be included with necessary O-Cloud Node identifiers End note

autonumber 4.1

nRT -> FOCOM: Guidance to evacuate O-Cloud Node nRT -> NFO: Guidance to modify NF deployment NFO -> DMS: <<O2dms>> Modify NF deployment FOCOM -> IMS:<<O2ims>> Evacuate O-Cloud Node

Loop for each \nNF Deployment DMS -> DMS: NF Modification

End Loop

Loop for each \nO-Cloud Node

IMS -> IMS: Evacuate O-Cloud Node End Loop

autonumber 5.1

DMS -> NFO :<<O2dms>> Completion of NF \n deployment modification IMS -> FOCOM :<<O2ims>> Completion \nof O-Cloud Node Evacuation

NFO -> nRT : Notify completion of NF \ndeployment modification FOCOM -> nRT : Notify completion \nof O-Cloud Node Evacuation

== O-Cloud Node Shutdown == Note over nRT , FOCOM

Shutdown request to be included with necessary O-Cloud Node identifiers End note

autonumber 6.1

nRT -> FOCOM: Guidance to shutdown O-Cloud node FOCOM -> IMS: <<O2ims>> Shutdown O-Cloud Node

autonumber 7

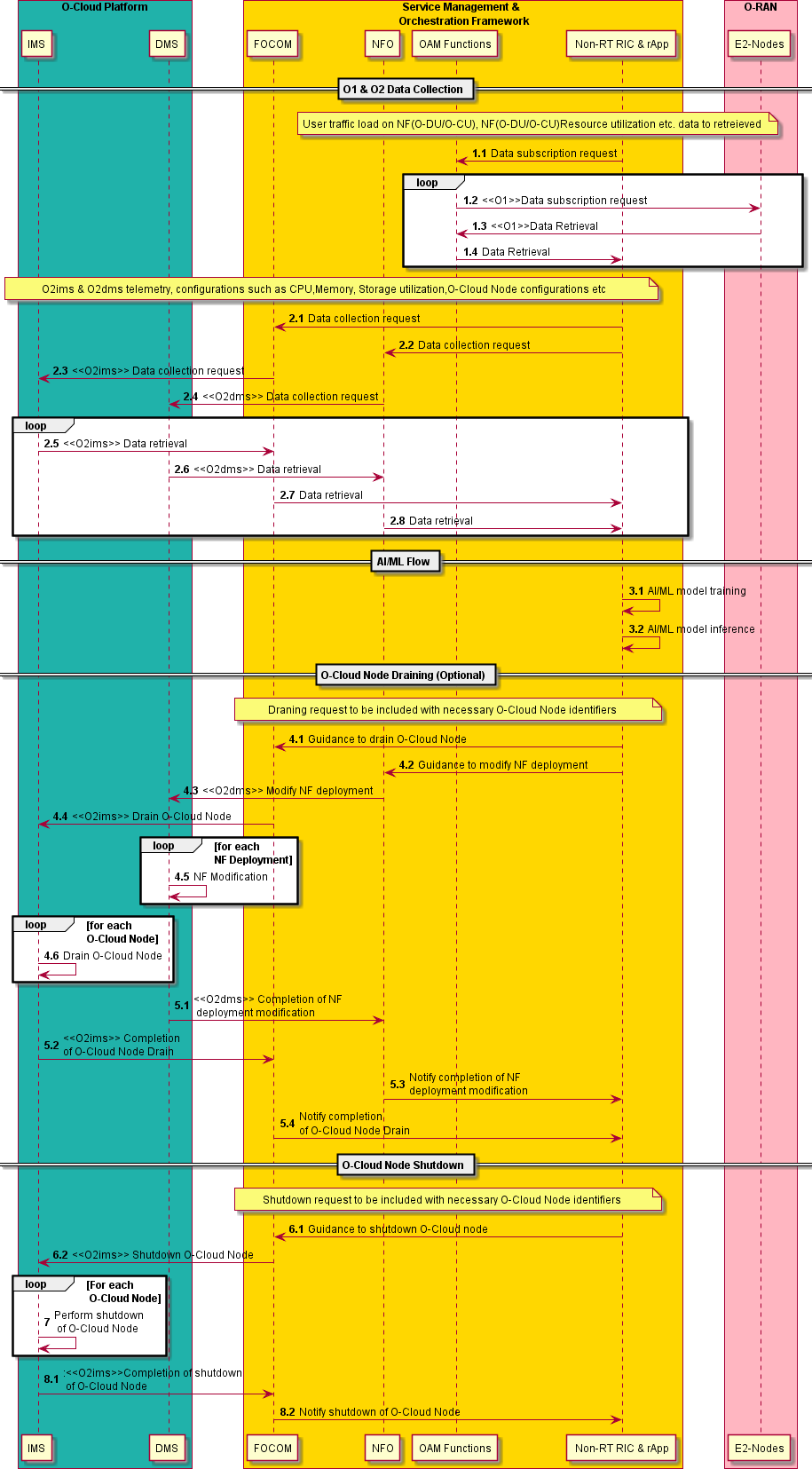
Loop For each\n O-Cloud Node

IMS -> IMS: Perform shutdown\n of O-Cloud Node End Loop

autonumber 8.1

IMS -> FOCOM: :<<O2ims>>Completion of shutdown\n of O-Cloud Node FOCOM -> nRT: Notify shutdown of O-Cloud Node

@enduml



##### Figure 8.1.2.1-1: Flow diagram for O-Cloud Node shutdown

#### O-RAN Entity Roles

1. Non-RT RIC (rApp)
   1. Collect configurations, performance indicators and power consumption information from E2 Nodes and O-Cloud Nodes via SMO, for the purpose of decision making, optionally using training and inference of AI/ML models that assist such EE/ES functions.
   2. Collect O-Cloud configuration data from O2 via SMO / from FOCOM and from NFO including policies applied to O-Cloud.
   3. Provide guidance towards FOCOM to evacuate or shut down O-Cloud Nodes.
   4. (Optionally) Deploy, update, configure and trigger execution of EE/ES AI/ML models.
2. E2 Node
   1. Report KPIs and measurements related to load and power consumption information to SMO via O1 interface.
3. O-Cloud (DMS)
   1. Report NF deployment information to SMO via O2dms interface.
   2. Apply NF deployment configurations received over O2dms.
   3. Provide feedback on NF deployment modification actions.
4. O-Cloud (IMS)
   1. Report infrastructure and deployment telemetry data to SMO via O2 interface. For example, O- Cloud Node configuration, health status, server memory utilization and I/O load overtime, CPU, network, and memory usage.
   2. Apply configurations received over O2ims related to O-Cloud Node draining and shutdown as part of O-Cloud energy saving optimization.
   3. Provide feedback post completion or non-completion of actions to SMO.
5. SMO (NFO)
   1. Collect the necessary configurations and NF deployment data from O-Cloud triggered by Non-RT RIC and forward information towards Non-RT RIC.
   2. Decide about optimized O-Cloud deployment configuration based on guidance received from Non- RT RIC to modify NF deployments. Send optimized O-Cloud deployment configuration to O-Cloud via O2dms interface.
   3. Monitor the NF deployment of O-Cloud NF via O2dms and report to Non-RT RIC.
6. SMO (FOCOM)
   1. Collect the necessary configurations, performance indicators, and measurement reports data from O- Cloud triggered by Non-RT RIC and forward information towards Non-RT RIC.
   2. Decide about optimized O-Cloud configuration based on guidance received from Non-RT RIC to evacuate or shut down O-Cloud Nodes. Send optimized O-Cloud Node configuration to O-Cloud via O2ims interface.
   3. Monitor the performance of O-Cloud Nodes via O2ims and report to Non-RT RIC.

#### Void

### Impact Analysis on O-RAN Work Groups

No detailed impact analysis has been conducted during this study. NOTE 8.1-1 should be resolved by WG6 during normative phase.

### Relation and Impact on 3GPP Specifications

No specific relation and impact on 3GPP specifications have been identified during the study.

### Void

### Feasibility Analysis

#### Service Continuity during NF relocation

NF pod relocation between cluster nodes involves shutting down the NF pod on its original node and starting a new NF pod on another node. In order to avoid impairment of the network performance and service continuity, it has to be guaranteed that all functionalities and traffic allocated to the original NF pod are reallocated before shutdown.

#### Pooling vs. Scaling Gains

Nodes can be shut down only if all pods running on them are shut down. O-Cloud Node Shutdown will result in concentration of pods on a fewer number of nodes, which also decreases resiliency and robustness.

#### Start-up Time for Scale Out Operation

Scale out operation after node shutdown involves startup of several pieces of hardware, initialization of its operating system and virtualization layers, configuring it as a cluster node, and adding it to the cluster as a node. Overall, this sequence of procedures can take several minutes. Therefore, Energy Saving procedures and resiliency/robustness procedures are to be optimized jointly.

**NOTE 8.1-1:** To be assessed by WG6, how fast reconfigurations of the O-Cloud Node Shutdown by SMO / Non-RT RIC can be applied in the O-Cloud and if there is any specific requirement.

## Sub Use Case 2: O-Cloud CPU Energy Saving Mode

### Problem Statement, Solution and Value Proposition

This use case describes a method of O-Cloud Energy Saving in which preferred CPU Energy Saving modes can be configured by SMO/Non-RT RIC. CPU Energy Saving modes, implemented by the vendor or standardized, may correspond to different CPU energy saving states (related to, e.g., CPU frequency, voltage, certain sleep modes) that can be controlled externally for CPU Power Management. O-Cloud might for instance be configured with a range (or utilization factor) of allowed or suggested CPU Energy Saving modes. O-Cloud is allowed to do fast adaptations of the CPU Energy Saving modes autonomously (e.g., based on instantaneous load of one or multiple CPUs) within that range. Alternatively, O-Cloud might be configured with a maximum CPU Energy Saving mode. O-Cloud is allowed to select among Energy Saving modes up to the maximum O-Cloud Energy Saving mode autonomously. By this, energy savings can be maximized, while still limiting the impact on QoS/user experience (e.g., potential latency impact on user plane traffic). The operator is allowed to control and tune O-Cloud energy saving gains versus O-Cloud performance.

Examples for CPU Energy Saving modes are the control of CPU frequency/voltage (P-state) and/or certain C-state of the O-Cloud Node CPUs as explained in the following.

##### CPU frequency (P-State):

P-States provide a way to scale the CPU frequency and voltage to reduce the power consumption of the CPU. The number of available P-States can vary with the type of CPU, even those from the same family, or one can change the corresponding frequency of the CPU. P-states can typically be limited or disabled in a system’s firmware such as UEFI/BIOS.

Frequency of the CPU (P-State) can be dynamically changed based on the load on the O-Cloud Node and thus O-Cloud Node energy consumption can be modified. The decision to select a desired CPU frequency (P- State) can be done by examining the current operational CPU frequency (P-State), CPU utilization and other related O2 and O1 (Traffic) telemetry from the O-Cloud instance.

P-State or CPU Frequency can be changed for the whole CPU (all cores) as well as for individual cores of a CPU.

##### C-State:

C-States are usually starting in C0, which is the normal CPU operating mode, i.e., the CPU is 100% turned on. With increasing CNumber, the CPU sleep mode is deeper, i.e., more circuits are turned off and more time is required to return the CPU back to C0 mode, i.e., to wake-up. Each mode is named individually and several of them have sub-modes with different power saving and thus wake-up time levels.

Example C-states:

* C0 is the operational state, meaning that the CPU is 100% turned on.
* C1 is the first idle state.
* C2 is the second idle state.

At the O-Cloud Node CPU level, power usage can be controlled in various ways. One way is by controlling the C-State of the CPU in CPE idle condition. C-States reflect the capability of an idle processor to turn off unused components to save power.

The decision of desired C-State must consider the NF deployment distribution on the O-Cloud instance and might involve consolidating the deployments on the limited set of O-Cloud Nodes based on O2 telemetry analysis such as CPU utilization, memory utilization, etc.

C-State can be changed at the CPU level, or at the individual core level of that CPU.

### Architecture/Deployment Option

In this deployment option, decision making for CPU Energy Saving Mode utilization control, potentially including AI/ML Model Training and Inference, is done at the Non-RT RIC. The overall mechanism is shown in Figure 8.2.2-1.

**SMO**

**O-RU**

Open FH

M-Plane

**E2 Node**

DMS

E2

IMS

A1

O1

O2-DMS

Non-Real Time RIC

& rApp

NFO

FOCOM

1. Receive guidance from NonRT RIC for optimizing CPU Energy Saving Mode.
2. Provide optimized configuration forCPU Energy Saving Mode towards O-Cloud.

Service, Management & Orchestration

O-Cloud ES Guidance

**Non-RT RIC (rApp)**

1. Collect O-Cloud FCAPS data via SMO over O2/R1.
2. Collect E2 Node data via SMO over O1/R1.
3. Generate guidance/configuration [(NOTE 0-1,](#_bookmark0) NOTE 8.2-2) based on O1 and O2 data for CPU Energy Saving Mode control based on priority, load and energy consumption.

O2-IMS

Near-RT RIC

DMS

DMS

**O-Cloud**

* 1. Expose O2 (IMS/DMS) data toward SMO/Non-RT RIC.
  2. Apply configuration provided by FOCOM in its decision making.
  3. Communicate its actions towards

SMO/Non-RT RIC.

**O-Cloud**

Open FH M-Plane

##### Figure 8.2.2-1: O-Cloud CPU Energy Saving Mode via SMO/Non-RT RIC

**NOTE 0-1:** The interface between Non-RT RIC and FOCOM is for illustrative purposes, subject to further modification by WG1 ATG architecture specifications.

**NOTE 0-2**: To be studied by WG2 if rApp has all information from O-Cloud to generate the updated guidance towards FOCOM.

#### Description and UML Diagram

##### Table 8.2.2.1-1: O-Cloud CPU Energy Saving Mode: AI/ML inference via Non-RT RIC

|  |  |  |
| --- | --- | --- |
| **Use Case Stage** | **Evolution / Specification** | **<<Uses>> Related use** |
| Goal | Optimizing CPU Energy Saving Mode for power management of the O- Cloud Node CPUs (e.g., to preferably use low power mode when CPU is  idle or operating at low load). |  |
| Actors and Roles | * SMO/FOCOM: Receives guidance from Non-RT RIC and generates optimized configurations towards O-Cloud. * Non-RT RIC (rApp): Trains AI/ML models and generates guidance. * O-Cloud (IMS): Executes changes recommended by SMO/Non-RT RIC. |  |
| Assumptions | * The “Service Request” to the FOCOM includes the identifiers of the O- Cloud Node(s) to be controlled. * Non-RT RIC (rApp) has subscribed to receive notifications from SMO. * SMO (FOCOM) gets guidance from Non-RT RIC (rApp), where rApp may host the algorithm(s) that determine to control the CPU Energy Saving Mode. |  |
| Pre- conditions | * All relevant functions and components are instantiated and available. * Non-RT RIC can receive configurations, load, performance measurements from RAN nodes via the O1 and telemetry data from O- Cloud via O2. * Changes of the CPU Energy Saving Mode(s) within the O-Cloud are enabled from O-Cloud Node BIOS/UEFI. |  |
| Begins when | The operator has enabled or has set targets for this CPU Energy Saving Mode control. |  |
| Step 1.1, 1.2, 1.3 (M) | Collection of traffic load and utilization and RAN configurations related Network function (O-DU and O-CU) data over O1 interface. |  |
| Step 2.1, 2.2, 2.3, 2.4,  2.5, 2.6, 2.7, 2.8 (M) | O-Cloud Node telemetry and inventory, for example CPU utilization, supported and currently used CPU Energy Saving Mode etc. over O2 interface to be acquired from SMO. |  |
| Step 3.1, 3.2 (O) | Non-RT RIC optionally trains the AI/ML models with the collected data. Trained AI/ML models are deployed and activated in the Non-RT RIC. Non- RT RIC constantly monitors RAN performance and energy consumption of the O-Cloud Nodes for inference; for example, load related and traffic information, EE/EC measurement reports, CPU, Memory, Storage utilization,  etc. |  |
| Step 4.1, 4.2 (M) | Non-RT RIC provides guidance to FOCOM, then FOCOM configures IMS with preferred CPU Energy Saving Mode(s) for the CPU of O-Cloud Node(s).  NOTE: CPU cores of specific O-Cloud Node might be identified in guidance or O2 configuration. |  |
| Step 5 (M) | IMS applies optimized CPU Energy Saving Mode of each O-Cloud Node to the configured CPU Energy Saving Mode. |  |
| Step 6.1, 6.2, 6.3 (M) | IMS notifies SMO and SMO notifies Non-RT RIC that change of Energy Saving Mode utilization is completed. |  |
| Ends when | CPU Energy Saving Mode has been changed. |  |
| Exceptions | None identified. |  |
| Post-conditions | Based on the load on the O-Cloud Nodes, the desired CPU Energy Saving mode is changed. Thus, CPU Energy Saving mode (e.g., CPU frequency (P- State) and/or C-State of CPU) is optimized for or energy saving purposes. |  |

@startuml autonumber

Box “O-Cloud Platform” #lightseagreen participant “IMS” as IMS participant “DMS” as DMS

End box

Box "Service Management & \n Orchestration Framework" #gold Participant “FOCOM” as FOCOM

Participant “NFO” as NFO Participant “OAM ” as OAM Participant “NonRT RIC” as nRT

End box

Box "O-RAN" #lightpink Participant "E2-Nodes" as E2N

End box

== O1 & O2 Data Collection == autonumber 1.1

note over nRT,OAM

User Traffic Load, RAN configurations related to Network functions (O-DU & O-CU) data to be retrieved

End note

Loop

nRT -> OAM : Data subscription request

OAM -> E2N :<<O1>> Data subscription request E2N -> OAM :<<O1>> Data retrieval

OAM -> nRT : Data retrieval

End Loop

Note over nRT,IMS

O2ims & O2dms telemetry & inventory such as supported CPU Energy Saving Modes and status, CPU utilization, current operational CPU Power, Frequency, Voltage etc of O-Cloud node to be retrieved

End note autonumber 2.1

Loop

nRT -> FOCOM : Data collection request

FOCOM -> IMS : <<O2ims>> Data collection request nRT -> NFO : Data collection request

NFO -> DMS : <<O2dms>> Data collection request

IMS -> FOCOM : <<O2ims>> Data retrieval DMS -> NFO : <<O2dms>> Data retrieval FOCOM -> nRT : Data retrieval

NFO -> nRT : Data retrieval

End Loop

== AI/ML Flow == autonumber 3.1

nRT -> nRT: AI/ML model training nRT -> nRT: AI/ML model inference

== CPU Energy Saving Mode Optimization == autonumber 4.1

nRT -> FOCOM: Guidance of CPU Energy Saving Mode

FOCOM -> IMS: <<O2ims>> Configure CPU Energy Saving Mode

autonumber 5

Loop For each O-Cloud Node

IMS -> IMS: Perform Change of CPU Energy Saving Mode End Loop

autonumber 6.1

IMS -> FOCOM: <<O2ims>> Notify change of CPU Energy Saving Mode of O-Cloud Node FOCOM -> nRT: Notify change of CPU Energy Saving Mode of O-Cloud Node

@enduml



##### Figure 8.2.2.1-1: Flow diagram for O-Cloud CPU Energy Saving Modes

#### O-RAN Entity Roles

1. Non-RT RIC (rApp)
   1. Collect configurations, performance indicators and power consumption information from SMO, E2 Nodes and O-Cloud Nodes, for the purpose of decision making, optionally using training and inference of AI/ML models that assist such EE/ES functions.
   2. Collect O-Cloud configuration data from FOCOM including supported CPU Energy Saving Modes and planned policies or configuration changes towards O-Cloud.
   3. Collect O-Cloud configuration data from NFO including NF deployment configuration and planned policies and configuration changes towards O-Cloud.
   4. Provide guidance to FOCOM to change CPU Energy Saving Mode.
   5. (Optional) Deploy, update, configure and trigger execution of EE/ES AI/ML models.

**NOTE 0-3**: Framework and procedure for guidance from Non-RT RIC to NFO/FOCOM within SMO to be studied in WG1 ATG.

1. E2 Node
   1. Report KPIs and measurements related to load and power consumption information to SMO via O1 interface.
2. O-Cloud (including IMS and DMS)
   1. Report infrastructure telemetry data to SMO via O2 interface, for example O-Cloud Node health status, server memory utilization and I/O load over time and deployment telemetry data like CPU, network, and memory usage etc.
   2. Apply configurations received over O2ims related to CPU Energy Saving Mode.
   3. Provide feedback post completion or non-completion of actions to SMO and/or to Non-RT RIC through SMO.
3. SMO (FOCOM/NFO)
   1. Collect the necessary configurations, performance indicators, and measurement reports data from RAN nodes and O-Cloud triggered by Non-RT RIC and forward information towards Non-RT RIC.
   2. Decide about optimized O-Cloud configuration based on guidance received by Non-RT RIC to optimize CPU Energy Saving Mode utilization. Send optimized O-Cloud Node configuration to O- Cloud via O2ims interface.
   3. Monitor the performance of O-Cloud Nodes via O2ims and RAN Network functions via O1; when the optimization objective fails, initiate fall-back procedure; meanwhile, trigger the AI/ML model re- training, data analytics and optimization in Non-RT RIC.

#### Void

### Impact Analysis on O-RAN Work Groups

No detailed impact analysis has been conducted during this study.

NOTE 8.2-1 through NOTE 8.2-4 shall be resolved by WG1 ATG, WG2 and WG6 during normative phase.

### Relation and Impact on 3GPP Specifications

No specific relation and impact on 3GPP specifications have been identified during the study.

### Void

### Feasibility Analysis

#### Not to Restrict Fast CPU Energy Saving Mode Switching

O-Cloud has instantaneous knowledge about O-Cloud resource and energy consumption and is therefore capable of fast adaptions. Moreover, different energy saving features might be available and run on different time scales, with different latency requirements. SMO/Non-RT RIC configurations should not restrict the O- Cloud from utilizing other (faster) energy saving features in case those are deemed optimal in the given context and decrease the energy consumption.

**NOTE 8.2-4**: To be assessed by WG6, how fast reconfigurations of the CPU Energy Saving Mode by SMO/FOCOM can be applied in the O-Cloud, and if there is any specific requirement.

# Summary and Conclusion

In the O-RAN WG1 UCTG Network Energy Saving pre-normative phase, Network Energy Saving use cases have been analyzed. This Technical Report presents the results of the pre-normative phase Network Energy Saving work item.

Carrier and Cell Switch Off/On is a technique which enables turning off cells/carriers in case there is no load/UEs in the respective cell/carrier, and the neighboring cells/carriers can take over the expected additional load. The algorithm to trigger off/on switching may be hosted in the Non-RT RIC or in the Near- RT RIC. Performance, load, resource utilization, and energy/power consumption related KPIs are forwarded to the RIC from the E2 Nodes and O-RUs, and the RIC-based applications determine recommendations for cells/carriers to be switched off (energySavingState). Ultimately, the switch-off operations are performed by the E2 Node (O-CU) since additional critical operations are required which are handled by the O-CU. While for the Non-RT RIC deployment all input/output parameters are existing, enhancements to the E2 interface are required when the algorithm is hosted by the Near-RT RIC. Numerical analysis shows a high potential of this energy saving technique for overall Network Energy Saving. No challenges on feasibility have been identified. Carrier and Cell Switch Off/On is suggested for normative standardization with the foreseen impact on O-RAN WGs as outlined in Section 5.3.

RF Channel Reconfiguration is a technique which enables O-RU to be requested to perform Tx/Rx Array selection. At low load, i.e., when the expected traffic volume or number of connected users are lower than the configured threshold, the power consumption of O-RUs can be reduced by switching off certain Tx/Rx arrays. For example, 32 out of 64 Tx/Rx Arrays of an O-RU can be switched off in a digital mMIMO architecture. The algorithm to trigger RF Channel Reconfiguration may be hosted in the Non-RT RIC or in the Near-RT RIC. Respective KPIs are collected by the Near-RT/non-RT RIC from the E2 Nodes and O- RUs, to be used by the xApps/rApps to determine recommendations for RF Channel Reconfiguration to be executed by the E2 Nodes. Numerical analysis shows a large potential of this energy saving technique for overall Network Energy Saving. Some feasibility aspects are listed in Section 6.6 which should be considered during normative phases in relevant WGs. RF Channel Reconfiguration is suggested for normative standardization with the foreseen impact on O-RAN WGs as outlined in Section 6.3.

Advanced Sleep Modes typically refer to standardized or implementation specific intelligent energy saving techniques in the RAN. The Advanced Sleep Mode Selection use case is a technique which achieves more granular energy saving by optimizing the range of allowed sleep modes in the O-DU/O-RU via RIC-based guidance. Two architecture/deployment options have been identified, 1) Training and Inference in Non-RT RIC, and 2) Training in Non-RT RIC and Inference in Near-RT RIC. In both options, the Non-RT RIC retrieves the E2 Node Sleep Mode capabilities over O1 and R1 via SMO, collects the necessary data from the E2 Nodes and trains an ML model, which is then deployed: for option 1) in the rApp, and for option 2) in the Near-RT RIC (xApp). In option 1) inference of the optimized Sleep Mode utilization range is provided as guidance by the rApp, while in option 2) by the Near-RT RIC (xApp). Some feasibility aspects are listed in Section 7.6 which should be considered during normative phases in relevant WGs. Advanced Sleep Mode Selection is suggested for normative standardization with the foreseen impact on O-RAN WGs as outlined in Section 7.3.

O-Cloud Resource Energy Saving Mode comprises a set of techniques which enable energy savings in the O- Cloud by reducing the power consumption of various O-Cloud components without impairing the network performance. Two sub use cases have been identified, O-Cloud Node Shutdown and O-Cloud CPU Energy Saving Mode. The O-Cloud Node Shutdown use case is a method for reducing O-Cloud energy consumption by focusing NF deployments on a fewer number of (physical) O-Cloud Nodes and thereby potentially freeing up physical resources. Thereafter, shutting down the idle physical nodes is achieved through guidance from Non-RT RIC to FOCOM. The O-Cloud CPU Energy Saving Mode use case is a method for reducing the O- Cloud energy consumption by guiding the available CPU-internal energy saving states via guidance provided

by Non-RT RIC to FOCOM. Some feasibility aspects are listed in Section 8.1.6 and Section 8.2.6, respectively, which should be considered during normative phases in relevant WGs. O-Cloud Resource Energy Saving Mode is suggested for normative standardization with the foreseen impact on O-RAN WGs as outlined in Section 8.1.3 and Section 8.2.3.

Implementation of NES features should ensure stable and conflict-free network operation. Ideas on design principles for NES applications facilitating appropriate conflict mitigation are outlined in Annex A.

O-RAN O1 interfaces are used for exchange of data for analysis and provisioning of configuration changes in the O-RAN Nodes. How Network Energy Saving use cases utilize the O1 interface for such purpose is described in Annex C in a generic way.

Annex A (Informative):

Design Principles for NES Features

Efficient and stable utilization of NES features requires an intelligent automation and optimization framework, leveraging on collected KPIs and data analytics with AI/ML algorithms. The SMO, the Non-RT RIC and the Near-RT RIC have an essential role in orchestrating energy saving mechanisms across the overall RAN infrastructure. Network operators should be able to control each NES feature at SMO level, for instance by scheduling the activation and deactivation of such features over pre-defined time periods.

Intelligent management and orchestration of O-CU, O-DU and O-Cloud resources are required. For example, optimized traffic steering will allow for a dynamic adaptation of active hardware resources.

NES features will differ in their algorithmic solutions with respect to the considered scales, both temporal and spatial. In order to harmonize different features, first a concise scale evaluation is required.

The higher the number of different NES features being deployed in a radio network, the higher the probability of conflicts between actions initiated by individual NES solutions. These kinds of conflicts can be minimized or avoided by either by a respective conflict coordination entity controlling the actions of each NES function during activation, or by an appropriate design of the NES functions facilitating conflict free coexistence, e.g., by deploying NES solutions with diverging time scales.

Currently rApp conflict management is not specified in O-RAN. Therefore, it may be investigated whether the xApp conflict mitigation approach could be expanded to apply for rApps as well. Furthermore, definition of rules for conflict mitigation – to be derived from the design guidelines – is essential.

Scales of Energy Saving Algorithms

Three different scales of energy saving algorithms are identified:

1. Decoupled case

In this case locality in time and (network) space is assumed which leads to local decision-making. Stability of operation is ensured by appropriately chosen hysteresis values. Decoupled NES features are investigated with respect to their potential of being overruled by non-local optimization algorithms. Example use cases are carrier and cell switch on/off and RF channel re-configuration.

1. Non-locality in time

Such an algorithm requires a robust forecast. Traffic load is considered being the principal steering basis for this kind of NES features. A generic traffic forecast with statistical guarantees as a generally available feature is essential. This information should be available in the SMO, and not in a dedicated rApp/xApp. An example use case would be switching off network entities during the night.

1. Non-locality in space

Earlier research showed the energy saving potential of nearest-neighbour handover algorithms in order to shut off specific cells. Coverage and capacity distinction will assure MNO’s legal obligations and product considerations.

Solution approach

Temporal and spatial scales of NES solutions need to be specified by App vendors. Concurrent update intervals in capturing the respective data lead to placing the respective Apps in the Non-RT RIC or the Near- RT RIC. App “twins” are possible.

Detecting and Gauging Potential Conflicts

In order to achieve scale-separation and conflict-free optimization hierarchies, NES features have to be charted and given a measure for their potential to figure in non-local optimization strategies. This measure derives from nearest-neighbour interaction of the feature and HW-induced hysteresis requirements.

Decoupled NES features are investigated with respect to their potential of being overruled by non-local optimization algorithms (see above).

Potential conflict and instability cases are for instance switching oscillations, SON-like conflicts between different AI/ML procedures for NES features, or SON-like conflicts between NES features and other use cases (e.g., QoE Optimization, Traffic Steering, Massive MIMO Beamforming Optimization, …).

Design Guidelines

Stability of network operation is key and to be considered first in place. Thus, non-local NES features with inherent instability risks according to the above analysis are to be deprioritized. To that end, proposed Apps are to be evaluated with regard to operational destabilization, and appropriate measures are to be derived.

Secondly the outcome, i.e. the potential for optimization, is to be taken into account. Off-line studies equip MNOs with aggregate estimates of NES effects within the representative time and space intervals. An NES feature with a higher potential of energy savings should overrule those with lower energy saving ability.

Appropriate harmonization between different Apps is essential. MNOs may derive criteria for rApps/xApps in order to ensure conflict-free energy saving hierarchies which App vendors are encouraged to comply to. Prior study work regarding conflict resolution in SON can be used to harmonize AI/ML Apps with different cost functions (use cases) but overlapping or identical action spaces.

Annex B (Informative):

Load profile and O-RU functional blocks

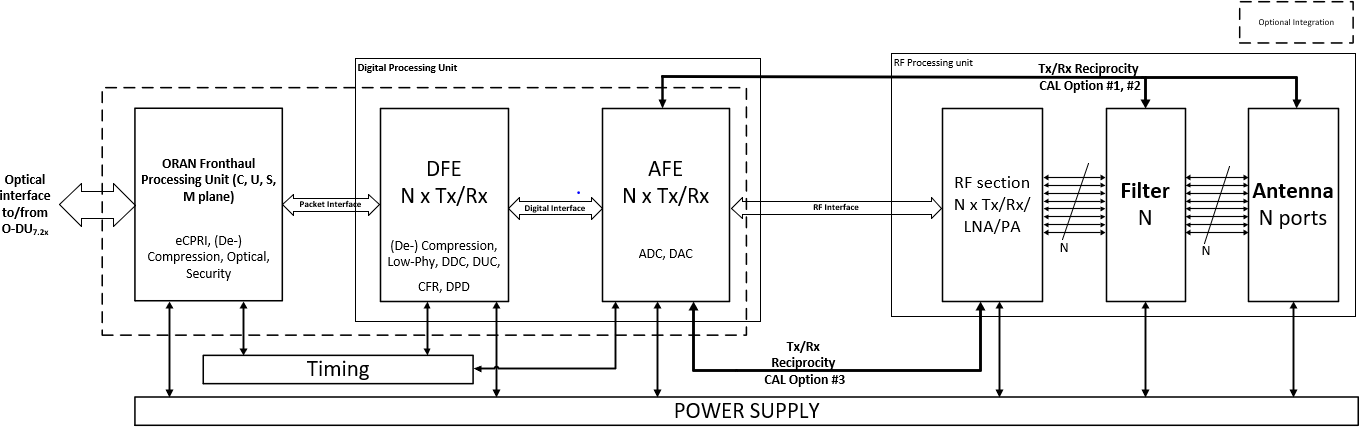
Generally, it is up to the operator to define a load distribution profile reflecting the situation in the network. For the calculation of the average power/energy consumption, weighting factors based on a daily (24 hours) load distribution profile given in Table C-1 as described in ETSI ES 202 706-1 [10] has been used.

##### Table B-1: Load level durations for daily average calculation

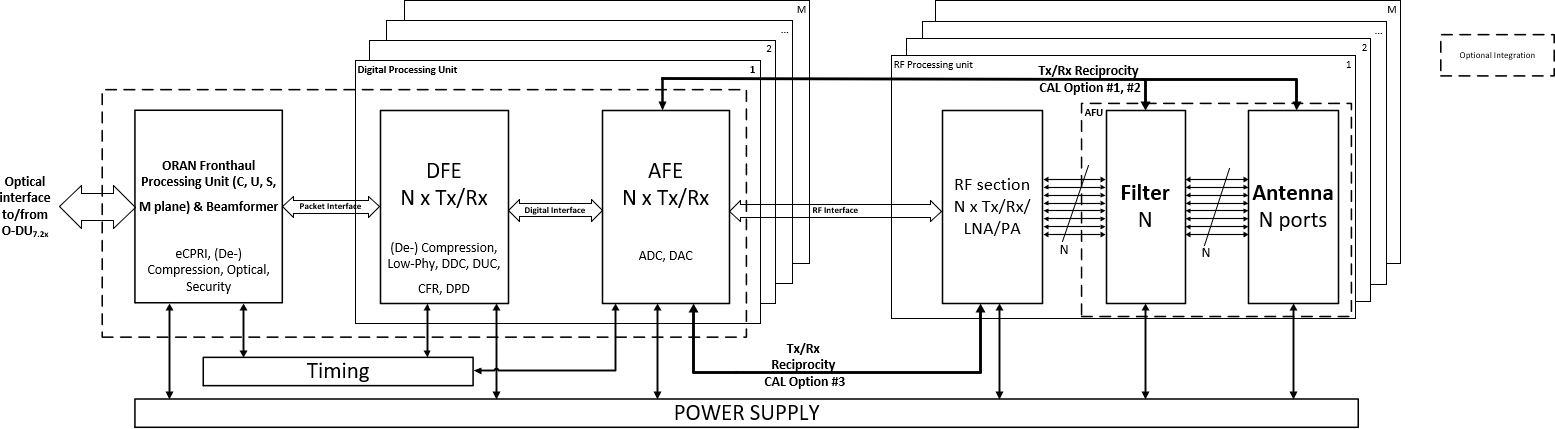
|  |  |  |  |
| --- | --- | --- | --- |
|  | **Low load** | **Medium load** | **Busy hour load** |
| **Duration/day** | 6 hours | 10 hours | 8 hours |

In Figure B-1 through B-4 as described in ORAN.WG7.OMAC-HAR [17], high-level functional block diagrams for an 4T4R/8T8R and mMIMO O-RUs are shown, depicting the major HW/SW components. Accordingly, the power consumption can be roughly categorized into the following parts:

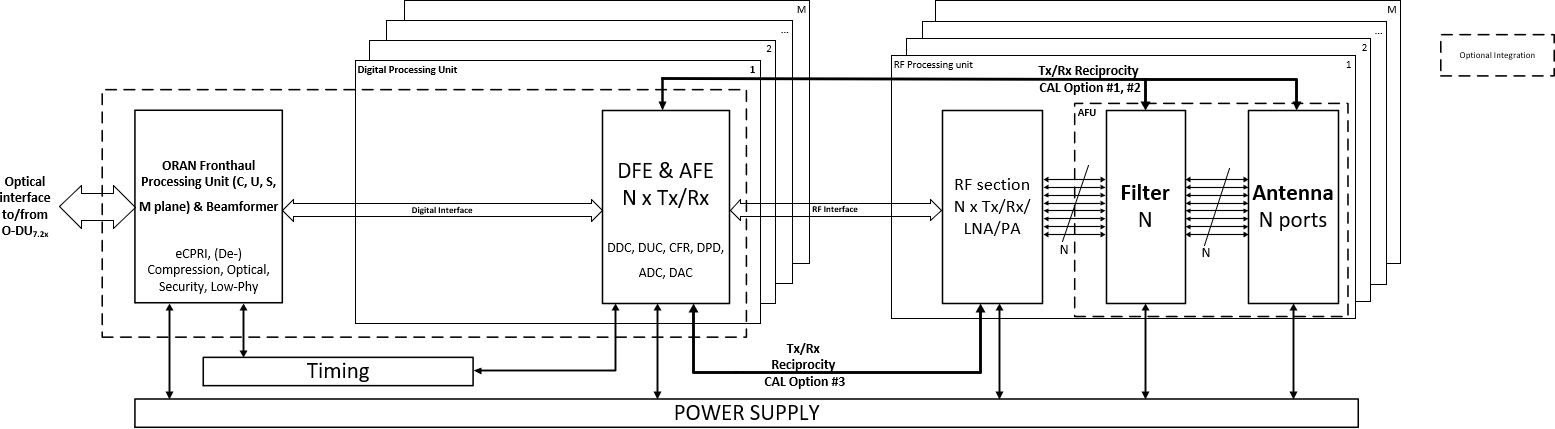
1. Digital Processing Unit
2. RF Processing Unit
3. FH Processing Unit
4. Power Unit & Other Components



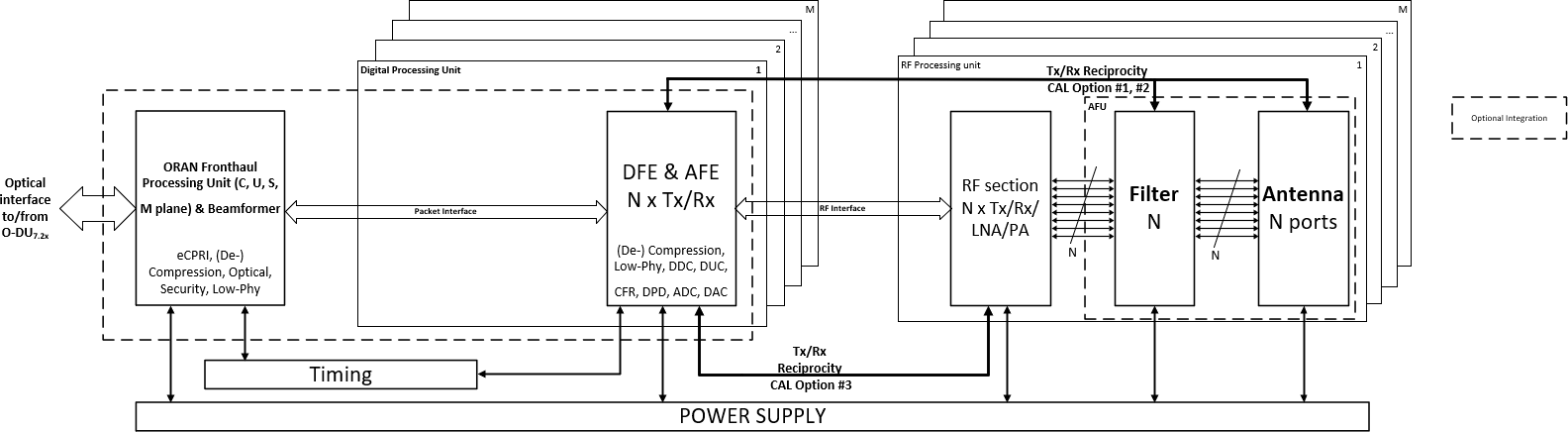
##### Figure B-1: Generic O-RU -2x Functional Module Diagram of 4T4R/8T8R



##### Figure B-2: mMIMO NxM T NxM R O-RU igh Level System Architecture #1



##### Figure B-3: mMIMO NxM T NxM R O-RU igh Level System Architecture #2



##### Figure B-4: mMIMO NxM T NxM R O-RU igh Level System Architecture #3

Annex C (Informative):

O1 interface principles

C.1: Overview

O-RAN O1 interfaces are used for exchange of data for analysis and provisioning of configuration changes in the O-RAN Nodes (Near-RT RIC, E2 Nodes, O-RUs) required for O-RAN use cases. This informative annex describes the general principles how Network Energy Saving use cases utilize the O1 interface for such purpose in a generic way.

C.2: O1 interface usage for data collection

The following message flow describes the usage of O1 interface of the Performance Assurance Management Services (see in Section 2.3 **Error! Reference source not found.**) for data collection.

@startuml Skin rose

skinparam defaultFontSize 12 autonumber

Box "Service Management & \n Orchestration Framework" #gold Participant "SMO" as SMO

End box

Box "O-RAN Nodes" #lightpink

Participant "O-RAN Node" as ORANNODES Participant "PerfMetricJob" as JOB

End box

== Data Collection ==

ref over SMO : Decision making on data collection \t

autonumber 1.1

SMO -> ORANNODES : <<O1>> Data collection job creation\

\n\t\tNETCONF edit-config create

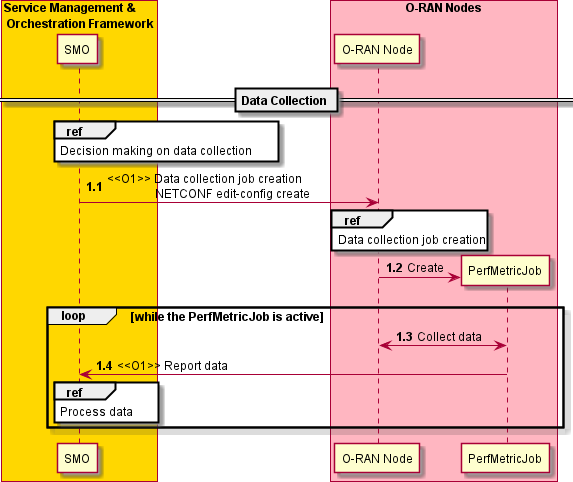
ref over ORANNODES : Data collection job creation ORANNODES -> JOB \*\*: Create

loop while the PerfMetricJob is active JOB <-> ORANNODES: Collect data

JOB -> SMO : <<O1>> Report data

ref over SMO: Process data \t end

@enduml



##### Figure C-1: O1 interfaces usage for data collection

O1 Interfaces are used in the data collection flow as described below:

1. In Step 1.1, the Performance Assurance Management Services (see Section 2.3 **Error! Reference source not found.**) are used to initiate performance measurement data collection. SMO, in the role of Performance Assurance MnS Consumer, sends request to O-RAN Nodes, in the role of Performance Assurance MnS Producer, for collection and retrieval of measurement data. The O1 interface for this procedure is described in Section 2.3.3 **Error! Reference source not found.**. The

Performance Assurance MnS Consumer may specify the reporting mechanism as data file reporting or data streaming.

1. In Step 1.2, O-RAN Nodes internally perform actions to create data collection job, to collect the requested data.
2. While the data collection job is active, the O-RAN Nodes collect the requested data (in Step 1.3). When the request data is ready, O-RAN Nodes send the collected measurement data to the SMO (in Step 1.4). Depending on the reporting mechanism specified by the SMO in Step 1.1, the O-RAN Nodes shall send the measurement data as data file or data streaming. The O1 interface for this procedure is described in Section 2.3.1 **Error! Reference source not found.** and 2.3.2 **Error! Reference source not found.** respectively.
3. SMO, after the retrieval of the measurement data, may internally perform further processing of the data and decision making required for use case.

C.3: O1 interface usage for configuration changes

The following message flow describes the usage of O1 interface of the Provisioning Management Services (see Section 2.1 **Error! Reference source not found.**) for configuration changes.

@startuml Skin rose

skinparam defaultFontSize 12 autonumber

Box "Service Management & \n Orchestration Framework" #gold Participant "SMO" as SMO

End box

Box "O-RAN Nodes" #lightpink

Participant "O-RAN Node" as ORANNODES End box

== Configuration changes ==

|||

ref over SMO: Decision making on necessary configuration changes \t autonumber 1.1

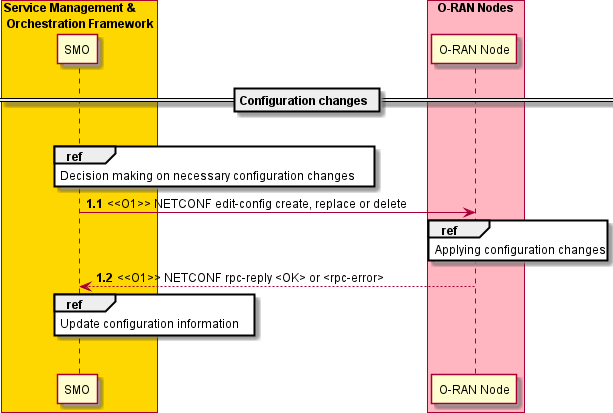
SMO -> ORANNODES: <<O1>> NETCONF edit-config create, replace or delete ref over ORANNODES: Applying configuration changes

SMO <-- ORANNODES: <<O1>> NETCONF rpc-reply <OK> or <rpc-error>

ref over SMO: Update configuration information \t

|||

@enduml



##### Figure C-2: O1 interfaces usage for configuration changes

O1 Interfaces are used in the configuration changes as described below:

1. In Step 1.1, the Provisioning Management Services (see Section 2.1 **Error! Reference source not found.**) are used to modify the parameters. SMO, in the role of Provisioning MnS Consumer, sends the request to configure parameters to the O-RAN Nodes, in the role of Provisioning MnS Producer, using the O1 interface described in Section 2.1.3 **Error! Reference source not found.**.
2. The changes are applied to the O-RAN Nodes. O-RAN Nodes may internally update the configurations across internal nodes.
3. In Step 1.2, O-RAN Nodes send the response to the SMO using the O1 interface described in Section

2.1.3 **Error! Reference source not found.** to indicate the result of the operation in Step 1.1.

1. SMO may read the values of the configuration parameters at any time using the O1 interface described in Section 2.1.5 **Error! Reference source not found.** and may subscribe to notification using the O1 interface described in Section 2.1.10 **Error! Reference source not found.** to receive notifications related to any changes to the configuration parameters with the O1 interface described in Section 2.1.9 **Error! Reference source not found.**.
2. SMO may internally update the configuration information across internal nodes.

Annex D (Informative):

Examples of Advanced Sleep Modes

The O-RAN specifications may be able to offer the flexibility to support different types of O-RUs with varying sleep mode capabilities, including the number of sleep modes available, their depth, and wake-up times. These sleep modes may range from shallow modes such as Symbol and Slot to deeper modes like Radio Frame or even longer durations in seconds. Depending on the sleep mode, different entities might be in control of selection. For instance, O-RU may internally execute Symbol to Slot level sleep modes, while O-DU might be responsible for controlling sleep modes from Slot to Radio Frame. Furthermore, Non-RT or Near-RT RIC may provide the operational range of Sleep Mode selection or may provide control for longer sleep modes.

Within the 3GPP Rel.18 study item on Network Energy Savings for NR, the following power states have been captured in Section 5.1 of 3GPP TR 38.864 [19] for evaluation purpose. Energy consumption mode for BS:

1. Micro sleep power state with instantaneous (i.e. per symbol basis) transition time
2. Light sleep power state with transition times between 6 ms and 640 ms
3. Deep sleep power state with transition times between 50 ms to 10 sec

# Revision history

|  |  |  |
| --- | --- | --- |
| **Date** | **Revision** | **Description** |
| 2023.02.07 | 01.00.01 | Removed v01.00 history Initial version towards v02.00 Agreed CR captured:  - NOK-2023.01.24-ORAN-CR-0037-  UC4\_CPUPowerModes\_WG6\_WG2\_ATG\_Feedback\_v02.docx  - NOK-2023.01.26-ORAN-CR-0038-  UC4\_OCloud\_NodeShutdown\_WG2\_ATG\_WG6\_Feedback\_v02.docx  NOTE: This version contains the following agreed CRs which have not been reflected in v01.00 publication.   * NOK-2022.09.19-ORAN-CR-0009-UC4\_O-Cloud Energy Saving Modes v07.docx * NOK-2022.09.19-ORAN-CR-0008-UC4\_O-Cloud Node Shutdown v02.docx |
| 2023.03.07 | 01.00.02 | Agreed CR captured:   * VIA.AO-2023.02.28-ORAN-CR-0040-Conclusions for Phase 2\_v01.docx * VIA.AO-2023.02.28-ORAN-CR-0039-Editorial Corrections for Phase 2\_v02.docx * VIA-2023.03.02-ORAN-CR-0041-Editorial Corrections for UC1-2\_v01.docx * RMI.AO-2023.03.02-ORAN-CR-0038-UseCase3\_Advance Sleep Modes\_v03.docx   - NOK-2023.02.23-ORAN-CR-0036-  UC3\_ASMOptimization\_DeploymentOption1\_v03.docx  - NOK-2023.02.23-ORAN-CR-0035-  UC3\_ASMOptimization\_DeploymentOption2\_v03.docx Editorial corrections |
| 2023.03.10 | 01.00.03 | Voided Section 7.2.1.3, 7.2.2.3, 7.2.5, 8.1.2.3, 8.1.5, 8.2.2.3, 8.2.5  Editorial corrections  Clean version for WG1 approval |
| 2023.03.22 | 01.00.04 | Captured WG1 review comments   * Table 7.2.1.1-1: Corrected extra node “Non” * “Energy Savings decision making function” and “configuration enforcement function” corrected as “inference host for Energy Savings decision making” and “the subject of action for configuration enforcement” * Editorial corrections |
| 2023.03.23 | 02.00 | Clean version for TSC approval and publication |

# History

|  |  |  |
| --- | --- | --- |
| **Date** | **Revision** | **Description** |
| 2022.11.19 | 01.00 | v01.00 publication |
|  |  |  |
|  |  |  |
|  |  |  |