NETWORK AWARE COMPUTE RESOURCE MANAGEMENT USE CASE FOR ORAN NON-RT RIC

BACKGROUND

[0001] A radio access network (RAN) is an important component in a telecommunications system, as it connects end-user devices (or user equipment) to other parts of the network. The RAN includes a combination of various network elements (NEs) that connect end-users to a core network. Traditionally, hardware and/or software of a particular RAN is vendor specific.

[0002] Open RAN (O-RAN) technology has emerged to enable multiple vendors to provide hardware and/or software to a telecommunications system. Since different vendors are involved, the type of hardware and/or software provided may also be different. That is, different types of NEs may be provided by different vendors, and depending on the specific service, the NE could be virtualized in software form (e.g., virtual machine (VM)-based), or could be in physical hardware form (e.g., non-VM based).

[0003] RAN functions in the O-RAN architecture are controlled and optimized by a RAN Intelligent Controller (RIC). The RIC is a software-defined component that implements modular applications to facilitate the multivendor operability required in the O-RAN system, as well as to automate and optimize RAN operations.

[0004] The RIC is divided into two types: a non-real-time RIC (Non-RT RIC) and a near-real-time RIC (Near-RT RIC). The Non-RT RIC operates on a timescale greater than 1 second within the Service Management and Orchestration (SMO) framework. Its functionalities are implemented through modular applications called rApps, and include: providing policy based guidance and enrichment across the A1 interface, which is the interface that enables

communication between the Non-RT RIC and the Near-RT RIC; performing data analytics; Artificial Intelligence/Machine Learning (AI/ML) training and inference for RAN optimization; and/or recommending configuration management actions over the O1 interface, which is the interface that connects the SMO to RAN managed elements (e.g., Near-RT RIC, O-RAN Centralized Unit (O-CU), O-RAN Distributed Unit (O-DU), etc.).

[0005] The Near-RT RIC operates on a timescale between 10 milliseconds and 1 second and is coupled with the O-CU control plane (O-CU-CP) and the O-CU user plane (O-CU-UP) via the E2 interface. It hosts xApps to implement functions such as interference mitigation, load balancing, security, etc. The two types of RICs work together to optimize the O-RAN. For example, the Non-RT RIC provides the policies, data, and AI/ML models enforced and used by the Near-RT RIC for RAN optimization.

[0006] As mentioned above, the Non-RT RIC is located within the SMO framework, which manages and orchestrates RAN elements. Specifically, the SMO manages and orchestrates what is referred to as the O-Ran Cloud (O-Cloud). The O-Cloud is a collection of physical RAN nodes that host the RICs, O-CUs, and O-DUs, the supporting software components (e.g., the operating systems and runtime environments), and the SMO itself. In other words, the SMO manages the O-Cloud from within.

[0007] In further detail, the O-RAN E2 nodes (i.e., the O-CU, O-DU, etc., that are connected to the Near-RT RIC via the E2 interface) are orchestrated on the O-Cloud as Virtualized Network Functions (VNFs). SMO anchored functionalities (Network Function Orchestrator (NFO), Federated O-Cloud Orchestration and Management (FOCOM), etc.) handle management and orchestration of VNFs and the O-Cloud. In the related art, however, SMO anchored functionalities are limited to utilizing intelligence by correlation of observability from E2 nodes,

infrastructure, and network function (NF) for influencing management/orchestration. O-Cloud resources cannot be optimized based on intent (NFO/FOCOM) or through configuration changes over the O2 interface (i.e., the interface between the SMO and the O-Cloud) from Non-RT RIC rApps. That is, in the related art, rApps of the Non-RT RIC do not have a scope for optimizing or influencing orchestration and management of the O-Cloud or operation and maintenance of O-RAN elements. Nor has any use case been defined that utilizes intelligence through rApps with the help of AI/ML for optimizing or influencing orchestration and management of the O-Cloud or operation and maintenance of O-RAN elements.

SUMMARY

[0008] According to example embodiments, systems and methods are provided that utilize one or more rApps of the Non-RT RIC to optimize O-Cloud resources.

[0009] According to example embodiments, systems and methods are provided that enable the Non-RT RIC to provide intelligence to O-Cloud operations and functionalities through optimizing O-Cloud resources.

[0010] According to example embodiments, a system for optimizing O-Cloud resources, includes: an O-RAN Cloud platform (O-Cloud) including a plurality of physical nodes configured to host virtualized network functions (VNFs), the VNFs including O-RAN centralized units (O-CUs) and O-RAN distributed units (O-DUs); a Service Management and Orchestration (SMO) framework configured to manage and orchestrate the O-Cloud, the SMO including anchored functionalities, a non-real-time radio access network (RAN) intelligent controller (Non-RT RIC), an O1 interface for communication with the VNFs and an O2 interface for communication with Infrastructure Management Services (IMS) and Deployment Management Services (DMS) of the O-Cloud; and a near-real-time RAN intelligent controller (Near-RT RIC) configured to perform

RAN optimization using policies and data provided by the Non-RT RIC, wherein the Non-RT RIC is configured to implement one or more applications (rApps) to: collect O-Cloud inventory and telemetry data on node level or O-Cloud level over the O2 interface, collect cellwise traffic data, including mobility and neighbor cell load data, over the O1 interface, train and deploy an artificial intelligence/machine learning (AI/ML) model to generate policies based on collected O1 and O2 data, and generate policies for the O-Cloud or VNFs based on at least one of priority, load, and energy consumption, and wherein the anchored functionalities of the SMO framework are configured to receive and manage the policies for the O-Cloud from the Non-RT RIC, and generate optimization controls for transmission to the O-Cloud via the O2 interface and/or transmit the policies to the IMS and DMS via the O2 interface for implementation.

[0011] The anchored functionalities of the SMO framework may include a Network Function Orchestrator (NFO) and a Federated O-Cloud Orchestration and Management (FOCOM).

BRIEF DESCRIPTION OF THE DRAWINGS

[0012] Features, advantages, and significance of exemplary embodiments of the disclosure will be described below with reference to the accompanying drawing:

[0013] FIG. 1 is a diagram of an O-RAN system architecture according to example embodiments.

DETAILED DESCRIPTION

[0014] The following detailed description of example embodiments refers to the accompanying drawings. The same reference numbers in different drawings may identify the same or similar elements.

[0015] The foregoing disclosure provides illustration and description, but is not intended to be exhaustive or to limit the implementations to the precise form disclosed. Modifications and variations are possible in light of the above disclosure or may be acquired from practice of the implementations. Further, one or more features or components of one embodiment may be incorporated into or combined with another embodiment (or one or more features of another embodiment). Additionally, in the flowcharts and descriptions of operations provided below, it is understood that one or more operations may be omitted, one or more operations may be added, one or more operations may be performed simultaneously (at least in part), and the order of one or more operations may be switched.

[0016] It will be apparent that systems and/or methods, described herein, may be implemented in different forms of hardware, firmware, or a combination of hardware and software. The actual specialized control hardware or software code used to implement these systems and/or methods is not limiting of the implementations. Thus, the operation and behavior of the systems and/or methods were described herein without reference to specific software code.—It is being understood that software and hardware may be designed to implement the systems and/or methods based on the description herein.

[0017] Even though particular combinations of features are recited in the claims and/or disclosed in the specification, these combinations are not intended to limit the disclosure of possible implementations. In fact, many of these features may be combined in ways not specifically recited in the claims and/or disclosed in the specification. Although each dependent claim listed below may directly depend on only one claim, the disclosure of possible implementations includes each dependent claim in combination with every other claim in the claim set.

[0018] No element, act, or instruction used herein should be construed as critical or essential unless explicitly described as such. Also, as used herein, the articles “a” and “an” are intended to include one or more items, and may be used interchangeably with “one or more.” Where only one item is intended, the term “one” or similar language is used. Also, as used herein, the terms “has,” “have,” “having,” “include,” “including,” or the like are intended to be open-ended terms. Further, the phrase “based on” is intended to mean “based, at least in part, on” unless explicitly stated otherwise. Furthermore, expressions such as “at least one of [A] and [B]” or “at least one of [A] or [B]” are to be understood as including only A, only B, or both A and B.

[0019] Example embodiments of the present disclosure provide systems and methods that utilize one or more rApps of the Non-RT RIC to optimize O-Cloud resources. Example embodiments of the present disclosure provide systems and methods that enable the Non-RT RIC to provide intelligence to O-Cloud operations and functionalities through optimizing O-Cloud resources.

[0020] FIG. 1 is a diagram of an O-RAN system architecture according to example embodiments.

[0021] Referring to FIG. 1, the Non-RT RIC hosts third party applications as rApps in the SMO which can read various observability through O1 and O2 related services. These third party rApps can be leveraged to provide policy-based guidance via NFO/FOCOM or Configuration Management (CM) changes via direct O2 interface for optimization of O-Cloud resources.

[0022] According to example embodiments, the third party rApps can influence:

- Automation / Optimization aspect of O-Cloud Infrastructure Life Cycle Management
    - Energy Saving (With the use of rApp /AI/ML, correlation between real time Load Prediction, NF performance and O-cloud performance data)

    - Hardware (HW) Healing (e.g., restart of HW) and automated maintenance
    - Security (e.g., password change)
    - Root Cause Analysis
- O-Cloud Deployment Life Cycle Management
    - Healing of NFs (e.g., restart, relocate, re-instantiate)
    - Scaling In/Out in Horizontal Scaling (Scaling In/Out of NFs based on configuration and loading related data from O-Cloud and E2 Nodes to improve efficiency in terms of load balancing and energy saving)
    - Scale Up/Down in Vertical Scaling
    - Health / State Check (e.g., Auditing, Reconciliation of real time state, health of NFs)
    - Diagnostic
- O-Cloud Provisioning in terms of capacity, availability, etc.
- O-Cloud Inventory in terms of the mapping of resources to their use.
- Assist Acceleration Abstraction Layer Interface (AALI) and Acceleration Abstraction Layer (AAL)-Logical Profile Unit (LPU) management aspects. LPU is a logical representation of resources within an instance of a HW Accelerator (e.g., there can be multiple processing units or subsystems on a hardware accelerator, or resource partitioning (hard – dedicated resources, soft – soft resources) and these can be logically represented as a AAL Logical Processing Unit).

[0023] According to example embodiments, the third party rApps can influence the above functions with the help of data over O2 and O1 such as:

- O2-IMS Related

- O2ims_InfrastructureInventory Services
- O2ims_InfrastructureMonitoring Services
- O2ims_InfrastructureProvisioning Services
- O2ims_InfrastructureLifecycleManagement Services

- O2-DMS Related
  - Deployment Lifecycle Management Services
  - Deployment Monitoring Services
- O1 FCAPS Data

[0024] Using this data, correlation between the deployed NF events and underneath O-Cloud events can be performed to find out the root cause and take wise actions towards NF and O-Cloud.

[0025] Further, the rApps according to example embodiments can make the O-Cloud more operator driven.

[0026] Action / Intent Management via:

- NFO / FOCOM:
  - NFO takes care of life cycle management (LCM) action for NFs, rApp can pass intent to NFO which shall guide to perform action toward deployment management services (DMS) for NFs For example, in case of NF healing scenarios, rApp can assist NFO to restart, relocate, re-instantiate NFs
  - FOCOM takes care of LCM of O-Cloud, rApp can pass intent to FOCOM to guide to perform action towards IMS. For example, rApp can assist in case of root cause analysis of failure in O-Cloud related to Hardware or Software based on observability correlation.

- Configuration changes over O2:
    - There are few use cases where fast and continuous actions would be required towards O-Cloud, like changing the P-state, C-state of the vCPU for energy saving purposes and direct actions via O2 would be ideal scenarios for less latency.

[0027] Aspects of example embodiments make O-Cloud orchestration and management more interoperable with the use of third party applications (rApps) and AI/ML as an intelligent input to NFO/FOCOM or direct input via O2 Interface.

Use Cases for O-Cloud Resource Optimization

[0028] Hereinbelow, uses cases for O-Cloud resource optimization according to various example embodiments are described.

1.0 Use Case 1: O-Cloud Energy Saving

[0029] This use case provides the motivation, description, and requirements for O-Cloud energy saving use case, allowing operators to define different objectives for energy saving of various components of O-RAN elements such as NFs deployed over O-Cloud, O-Cloud Infrastructure, O-Cloud underneath hardware, etc.

1.1 Background and Goal of this Use Case

[0030] Energy consumption of mobile networks is a key topic for mobile operators. Energy Saving (ES) features involve different network mechanisms, network nodes and layers, operating at different time scales. Certain ES features are already deployed in network equipment and are vendor specific.

[0031] This use case shall provide various options over Non-RT RIC platform and rApps to enable energy saving features which includes RAN NFs deployed over O-Cloud, O-Cloud Infrastructure, O-Cloud underneath hardware.

1.2 Entities/Resources Involved in this Use Case

1) SMO (NFO / FOCOM / OAM)
    a) Support interpretation and enforcement of policies from Non-RT RIC
    b) Provide necessary control to O-Cloud elements (IMS and DMS) based on policy from Non-RT RIC
2) rApp and Non-RT RIC framework
    a) Retrieve necessary performance, configuration, and other data from O-Cloud over O2 interface to define and update policies for energy saving to SMO anchored functionalities.
3) O-Cloud (IMS and DMS)
    a) O-Cloud (IMS and DMS) to read control from SMO anchored functions and enforce changes to respective nodes or Hardware underneath.
    b) Provide feedback to SMO post control enforcement.

1.3 High Level Solutions

Step 1) rApp to subscribe FACPS data over O1 and O2 from E2 nodes and O-Cloud (IMS / DMS) respectively such as CPU Utilization for NFs, bare metal power consumption, CPU load, memory load, CPU frequency etc.

Step 2) rApp to train AI/ML model or through pre-defined algorithm with data retrieved over O1 and O2, then it shall produce policy intent for energy saving in different schemes for respective node. Example of policies intent: CPU frequency change, CPU load reductions, Suspension of ongoing processes in O-Cloud, relocation of application to another set of servers, putting the un-used servers to sleep state etc.

Step 3) rApps passes the intent to NFO/FOCOM

Step 4) NFO/FOCOM to translate policy intent from Non-RT RIC and enforce controls to O-Cloud

Step 5) O-Cloud to enforce changes as dictated by NFO/FOCOM to respective nodes and shall provide feedback to SMO or Non-RT RIC upon successful or failure of policy implementation.

Step 6) rApp to monitor performance / configuration changes of RAN functions, NFs and O-Cloud Infrastructure for closed loop evaluation of policy.

2.0 Use Case 2: RAN Network and Infrastructure Load Balancing

[0032] This use case provides the motivation, description, and requirements for Non-RT RIC and O2 interface to support O-Cloud resource load management and E2 traffic load management based on O1 and O2 performance, configuration, and fault management data.

2.1 Background and Goal of this Use Case

[0033] Balancing traffic allows operators to distribute load across different cells of networks and provides better user experience in terms throughput, latency, etc.

[0034] However, related art traffic steering use cases defined in O-RAN are limited to RAN traffic management and do not have any requirement or coordination with the O-Cloud to perform load management with respect to Network Functions.

[0035] This use case is to provide requirement for Non-RT RIC, O2, SMO, and O-Cloud to enable rApp to perform load balancing in both ways (i.e., RAN side and O-Cloud side). That is, this RAN Network and Infrastructure Load Balancing use case will make load balancing complete and more efficient from users' and operators' perspectives.

2.2 Entities/Resources Involved in this Use Case

1) SMO (NFO / FOCOM / OAM)

a) Support interpretation and enforcement of policies from Non-RT RIC

b) Provide necessary control to O-Cloud elements (IMS and DMS) based on policy from Non-RT RIC

2) rApp and Non-RT RIC framework

a) Retrieve necessary performance, configuration, and other data from O-Cloud over O2 interface and O1 Interface to define and update policies for energy saving to SMO anchored functionalities.

3) O-Cloud (IMS and DMS)

a) O-Cloud (IMS and DMS) to read control from SMO anchored functions and enforce changes to respective nodes or Hardware underneath.

b) Provide feedback to SMO post control enforcement.

2.3 High Level Solutions

Step 1) rApp to subscribe FACPS data over O1 and O2 from E2 nodes and O-Cloud (IMS / DMS) respectively such as User mobility performance data, traffic distribution amongst cells, Neighbor cell measurement data, NFs configuration, NFs Inventory and telemetry.

Step 2) rApp to train AI/ML model or through pre-defined algorithm with data retrieved over O1 and O2, then it shall produce policy or configuration changes over O1 or O2 based on below,

a. Traffic Steering Policy: If NFs load is high and target cell available for mobility

b. Scale In Policy to NFO: If no target for cell, then Scale In NFs to accommodate load.

c. Traffic Steering and Scale Out Policy: If NFs are less loaded then move user to other cell and perform scale out of NFs.

Step 3) SMO to translate policy from Non-RT RIC and enforce controls to O-Cloud

Step 4) O-Cloud to enforce changes as dictated by SMO to respective nodes and shall provide feedback to SMO or Non-RT RIC upon successful or failure of policy implementation.

Step 5) Near RT RIC to enforce policy received from Non-RT RIC

Step 6) rApp to monitor performance of UE, RAN Functions, NFs and O-Cloud Infrastructure for evaluation of policy.

3.0 Use Case 3: Dynamic Allocation and Optimization of O-Cloud Resources for Network Slicing

[0036] This use case provides the motivation, description, and requirements for Non- RT RIC and O2 interface to support optimization of O-Cloud resources for particular network slice in coordination with RAN resources based on O1 and O2 performance, configuration, fault management data and intent defined by operators.

3.1 Background and Goal of this Use Case

[0037] Network slicing helps network providers slice a single network into multiple virtual networks built on a single infrastructure setup. Network slicing overlays these multiple virtual networks on top of a shared network. Moreover, through network slicing, each of the sliced virtual networks can be configured separately with different network attributes such as logical topology, security rules, and performance characteristics to cater to different use-cases. However, each of these sliced virtual networks is built on O-Cloud infrastructure. The features and requirements of network slices would vary semi-statically or dynamically, which leads to optimizing not only RRM resources but also O-Cloud infrastructure and deployment resources as well.

[0038] This use case is to enable rApp to influence NSSMF (Network Slice Subnet Management Function) / NFO for optimization or rebalancing of O-Cloud infrastructure resources (scale, heal, etc.) through policy to achieve a desired SLA (service level agreement) of a particular network slice.

3.2 Entities/Resources Involved in this Use Case

1) SMO (NFO / FOCOM / OAM) and NSSMF
    a) Support interpretation and enforcement of policies from Non-RT RIC
    b) Provide necessary control to O-Cloud elements (IMS and DMS) based on policy from Non-RT RIC
2) rApp and Non-RT RIC framework
    a) Retrieve necessary performance, configuration, and other data from O-Cloud over O2 interface and O1 Interface to define and update policies for optimization or Rebalancing of O-Cloud infrastructure resources of particular network slice to SMO anchored functionalities.
3) O-Cloud (IMS and DMS)
    a) O-Cloud (IMS and DMS) to read control from SMO anchored functions and enforce changes to respective nodes or hardware underneath.
    b) Provide feedback to SMO post control enforcement.

3.3 High Level Solutions

Step 1) rApp to subscribe FACPS data over O1 and O2 from E2 nodes and O-Cloud (IMS / DMS) respectively, such as network slicing related FCAPS data (Slice utilization, Slice throughput, Slice availability, etc.), Network slicing NFs LCM data.

Step 2) rApp to train AI/ML model or through pre-defined algorithm with data retrieved, then it shall produce policy for NSSMF or NFO to perform optimization in various sub-use cases

a. Scale-In Policy to NFO: If there is more requirement of NFs in terms of compute/control or memory, then Scale-In NFs to accommodate load.

b. Scale-Out Policy to NFO: If there is reduced requirement of NFs in terms of compute/control or memory, then Scale-Out NFs to accommodate load.

c. AAL-LPU Profiling: Selection or optimization of various AAL-LPU profiles for acceleration of specific functions based on use cases requirements such as URLLC, eMBB, eMTC .

Step 3) SMO/NSSMF to translate policy from Non-RT RIC and enforce controls to O-Cloud

Step 4) O-Cloud to enforce changes as dictated by SMO to respective nodes and shall provide feedback to SMO or Non-RT RIC upon successful or failure of policy implementation.

Step 5) Near-RT RIC to enforce policy received from Non-RT RIC

Step 6) rApp to monitor performance of UE, Network Slice, RAN Functions, NFs and O-Cloud Infrastructure for evaluation of policy.

4.0 Use Case 4: O-Cloud Resource Management and SLA Assurance for RAN Sharing

[0039] This use case provides the motivation, description, and requirements for O-Cloud resource management and SLA assurance for RAN Sharing. RAN Sharing enables multiple operators to share the same O-RAN infrastructure, while allowing them to remotely configure and control the shared resources via remote O1, O2 and E2 interfaces.

4.1 Background and Goal of this Use Case

[0040] RAN sharing provides the possibility for an operator (Operator A) to share a network infrastructure with another operator (Operator B). The challenge here is to enable

Operator B to configure and control resources in an infrastructure that is owned by Operator A. In that scenario, the adoption of the O-RAN architecture can facilitate the remote control and configuration of such VNFs. Indeed, Operator B can monitor and control the remote O-DU via the Near-RT RIC of site B, using a “remote” E2 interface, while all the remote VNF configuration procedures can be handled by a specific rApp of Non-RT RIC, located at the SMO of each operator.

[0041] The objective of this use case is to make rApp read SLA from Operator B via some remote interface which defines various requirements for operator B on infrastructure of Operator A, thereby optimizing O-Cloud and RAN resources to achieve expected SLA through Policy mechanism.

4.2 Entities/Resources Involved in this Use Case

1) SMO (NFO / FOCOM / OAM)
    a) Support interpretation and enforcement of policies from Non-RT RIC
    b) Provide necessary control to O-Cloud elements (IMS and DMS) based on policy from Non-RT RIC
2) rApp and Non-RT RIC framework
    a) Retrieve necessary performance, configuration, and other data from O-Cloud over O2 interface and O1 Interface to define and update policies for optimization or Rebalancing of O-Cloud infrastructure resources of particular network slice to SMO anchored functionalities.
3) Near-RT RIC framework
    a) Support interpretation and execution of intents and policies from Non-RT RIC to derive the RAN Sharing optimization at RAN level in terms of expected behavior.

b) Sending use case performance report to Non-RT RIC for evaluation and optimization

4) O-Cloud (IMS and DMS)

a) O-Cloud (IMS and DMS) to read control from SMO anchored functions and enforce changes to respective nodes or hardware underneath.

b) Provide feedback to SMO post control enforcement.

5) RAN Nodes

a) Support network state and UE performance report with required granularity to SMO over O1 interface.

b) Support policy enforcement based on messages from A1/E2, which are expected to influence RRM behavior

4.3 High Level Solutions

Step 1) rApp to subscribe FACPS data over O1 and O2 from E2 nodes and O-Cloud (IMS / DMS) respectively, such as Network slicing related FCAPS data (Slice utilization, Slice throughput, Slice availability etc.), Network slicing NFs LCM data.

Step 2) rApp to train AI/ML model or through pre-defined algorithm with data retrieved, then it shall produce policy for NSSMF or NFO to perform optimization in various sub-use cases

a. Scale-In Policy to NFO: If there is more requirement of NFs in terms of compute/control or memory, then Scale-In NFs to accommodate load.

b. Scale-Out Policy to NFO: If there is reduced requirement of NFs in terms of compute/control or memory, then Scale-Out NFs to accommodate load.

c. AAL-LPU Profiling: Selection or optimization of various AAL-LPU profiles for acceleration of specific functions based on use cases requirements such as URLLC, eMBB, eMTC .

Step 3) SMO/NSSMF to translate policy from Non-RT RIC and enforce controls to O-Cloud

Step 4) O-Cloud to enforce changes as dictated by SMO to respective nodes and shall provide feedback to SMO or Non-RT RIC upon successful or failure of policy implementation.

Step 5) Near RT RIC to enforce policy received from Non-RT RIC

Step 6) rApp to monitor performance of UE, Network Slice, RAN Functions, NFs and O-Cloud Infrastructure for evaluation of policy.

5.0 Use Case 5: Infrastructure Network Elements Healing

[0042] This use case provides the motivation, description, and requirements for network healing, allowing operators to define different healing scenarios for various O-RAN elements such as network functions deployed over O-Cloud and O-Cloud Infrastructure.

5.1 Background and Goal of this Use Case

[0043] Network Healing is becoming the key for running the cloud native operations in the mobile network for mobile operators. Network healing involves monitoring the network elements' fault and performance data, receiving the fault events that mark the undesired state of the network elements, analyzing the defined policy action for that fault, correlating with the topology of events, and passing the right policy action intent to bring the network application live again. Certain network healing features are already deployed in network equipment, but are limited and vendor specific.

[0044] This use case shall provide various options over Non-RT RIC platform and rApps to enable network element healing features which includes RAN network functions deployed over O-Cloud and O-Cloud Infrastructure.

5.2 Entities/Resources Involved in this Use Case

1) SMO (NFO / FOCOM / OAM)

a) Support interpretation and enforcement of policies from Non-RT RIC

b) Provide necessary control to O-Cloud elements (IMS and DMS) based on policy intent from Non-RT RIC

2) rApp and Non-RT RIC framework

a) Retrieve necessary fault, performance data from NFs over O1 interface, and from O-Cloud over O2 interface to define and update desired policy intent for healing to SMO anchored functionalities.

3) O-Cloud (IMS and DMS)

a) O-Cloud (IMS and DMS) to read control from SMO anchored functions and enforce changes to respective nodes or Hardware underneath.

b) Provide feedback to SMO post control enforcement.

5.3 High Level Solutions

Step 1) rApp to subscribe FACPS data over O1 and O2 from E2 nodes and O-Cloud (IMS / DMS) respectively such as O-Cloud fault / performance data, NFs fault/performance data, for example NF_faulted, NF_deployment_failed, Hardware_failure etc

Step 2) rApp to train AI/ML model or through pre-defined algorithm with data retrieved over O1 and O2, then it shall produce desired policy intent for healing in different schemes for respective node.

a) Example-1: If an application goes down and fault event of the respective application will be received by rApps, then rApps will evaluate the possible root cause based on defined algorithms and will produce the policy action intent such as Restart the application, relocate the application to different server, or increase the resource limitation for the application, And such desired policy intent will be passed

to the NFO (within SMO), which will take the required action towards the application deployed on the o-cloud.

b) Example-2: If a hardware goes down and fault event of the respective hardware will be received by rApps, then rApps will evaluate the possible root cause based on defined algorithms and will produce the policy action intent such as reboot the hardware and relocate the applications to different server, and if even after rebooting the hardware, the issue persists, then raise the hardware maintenance ticket in the incident management module.

Step 3) rApps passes the intent to NFO/FOCOM.

Step 4) NFO/FOCOM to translate the policy intent from Non-RT RIC to the APIs and will take that desired action, such as restart of the hardware, relocating the applications to different server.

Step 5) O-Cloud to enforce changes as dictated by NFO/FOCOM to respective applications / hardware and shall provide feedback to SMO or Non-RT RIC upon successful or failure of policy implementation.

Step 6) rApp to monitor performance / configuration changes of RAN NFs and O-Cloud Infrastructure for closed loop evaluation of policy and take follow-up actions such as raising the incident ticket.

6.0 Use Case 6: Infrastructure Network Troubleshooting

[0045] This use case provides the motivation, description, and requirements for root cause analysis in the mobile network allowing operators to find out the root cause of the fault events and performance degradations of network elements based on AI/ML algorithms and correlation of data from network functions and O-Cloud Infrastructure.

6.1 Background and Goal of this Use Case

[0046] Due to complex virtual application deployment scenarios, finding the root cause manually is very time consuming process and could cause network outage for long periods of time when there is delay in fixing the issue. Automatically finding the root cause involves monitoring the network element's fault and performance data, performing the root cause analysis using AI/ML algorithms and correlation of data from various layers, drilling down to exact cause of the problem, and taking appropriate action.

[0047] This use case shall provide various options over Non-RT RIC platform and rApps to enable network element problem's root cause analysis which includes RAN network functions deployed over O-Cloud and O-Cloud Infrastructure.

6.2 Entities/Resources Involved in this Use Case

1) SMO (NFO / FOCOM / OAM)
    a) Support interpretation and enforcement of policies from Non-RT RIC
    b) Provide necessary control to O-Cloud elements (IMS and DMS) based on policy intent from Non-RT RIC
2) Incident Management Tool
    a) A tool where network incidents are raised and respective operational teams act thereon.
3) rApp and Non-RT RIC framework
    a) Retrieve necessary fault, performance data from NFs over O1 interface, and from O-Cloud over O2 interface to define and update desired policy intent for healing to SMO anchored functionalities.
4) O-Cloud (IMS and DMS)

a) O-Cloud (IMS and DMS) to read control from SMO anchored functions and enforce changes to respective nodes or hardware underneath.

b) Provide feedback to SMO post control enforcement.

6.3 High Level Solutions

Step 1) rApp to subscribe FACPS data over O1 and O2 from E2 nodes and O-Cloud (IMS / DMS) respectively such as O-Cloud fault / performance data, NFs fault/performance data, for example CPU_utilization, Memory_utilization, fan_failure, NF_faulted, NF_deployment_failed, Hardware_failure, etc

Step 2) rApp to train AI/ML model or through pre-defined algorithm with data retrieved over O1 and O2, then it shall correlate the data and drill down to exact root cause schemes for respective node.

a) Example-1: If an application performance is degraded below a threshold, this performance KPI will be received by rApps, then rApps will evaluate the possible root cause based on defined algorithms, like CPU utilization on the O-Cloud, whether there is over utilization of CPU, memory, checking how many applications are deployed on the same cloud, correlating performance of all applications on the same cloud and making out that over allocation of CPU is the root cause, raising the incident of the root cause and suggesting an action to add more compute to the O-Cloud.

b) Example-2: If an application is not reachable, then rApps will evaluate the possible root cause by finding the O-Cloud where this application is deployed, checking whether there is issue in the o-cloud infrastructure by correlating with the relevant fault events from the infrastructure, and raising a ticket in the incident manager for this hardware

issue and passing a policy intent to the NFO/FOCOM to deploy the application on different o-cloud.

Step 3) rApps passes the intent to NFO/FOCOM or rApps raises the incident ticket for root cause.

Step 4) NFO/FOCOM to translate the policy intent from Non-RT RIC to the APIs and will take that desired action, And if incident is raised in the incident management module, operational team will be notified and they will immediately act on the root cause to solve the problem.

Step 5) rApp to monitor performance / configuration changes of RAN NFs and O-Cloud Infrastructure for correlations of data over the time.

[0048] The foregoing disclosure provides illustration and description, but is not intended to be exhaustive or to limit the implementations to the precise form disclosed. Modifications and variations are possible in light of the above disclosure or may be acquired from practice of the implementations. Further, one or more features from one or more embodiments may be combined with one or more features of one or more other embodiments described above.

[0049] Some embodiments may relate to a system, a method, and/or a computer readable medium at any possible technical detail level of integration. Further, one or more of the above components described above may be implemented as instructions stored on a computer readable medium and executable by at least one processor (and/or may include at least one processor). The computer readable medium may include a computer-readable non-transitory storage medium (or media) having computer readable program instructions thereon for causing a processor to carry out operations.

[0050] The computer readable storage medium can be a tangible device that can retain and store instructions for use by an instruction execution device. The computer readable storage

medium may be, for example, but is not limited to, an electronic storage device, a magnetic storage device, an optical storage device, an electromagnetic storage device, a semiconductor storage device, or any suitable combination of the foregoing. A non-exhaustive list of more specific examples of the computer readable storage medium includes the following: a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), a static random access memory (SRAM), a portable compact disc read-only memory (CD-ROM), a digital versatile disk (DVD), a memory stick, a floppy disk, a mechanically encoded device such as punch-cards or raised structures in a groove having instructions recorded thereon, and any suitable combination of the foregoing. A computer readable storage medium, as used herein, is not to be construed as being transitory signals per se, such as radio waves or other freely propagating electromagnetic waves, electromagnetic waves propagating through a waveguide or other transmission media (e.g., light pulses passing through a fiber-optic cable), or electrical signals transmitted through a wire.

[0051] Computer readable program instructions described herein can be downloaded to respective computing/processing devices from a computer readable storage medium or to an external computer or external storage device via a network, for example, the Internet, a local area network, a wide area network and/or a wireless network. The network may comprise copper transmission cables, optical transmission fibers, wireless transmission, routers, firewalls, switches, gateway computers and/or edge servers. A network adapter card or network interface in each computing/processing device receives computer readable program instructions from the network and forwards the computer readable program instructions for storage in a computer readable storage medium within the respective computing/processing device.

[0052] Computer readable program code/instructions for carrying out operations may be

assembler instructions, instruction-set-architecture (ISA) instructions, machine instructions, machine dependent instructions, microcode, firmware instructions, state-setting data, configuration data for integrated circuitry, or either source code or object code written in any combination of one or more programming languages, including an object oriented programming language such as Smalltalk, C++, or the like, and procedural programming languages, such as the "C" programming language or similar programming languages. The computer readable program instructions may execute entirely on the user's computer, partly on the user's computer, as a stand-alone software package, partly on the user's computer and partly on a remote computer or entirely on the remote computer or server. In the latter scenario, the remote computer may be connected to the user's computer through any type of network, including a local area network (LAN) or a wide area network (WAN), or the connection may be made to an external computer (for example, through the Internet using an Internet Service Provider). In some embodiments, electronic circuitry including, for example, programmable logic circuitry, field-programmable gate arrays (FPGA), or programmable logic arrays (PLA) may execute the computer readable program instructions by utilizing state information of the computer readable program instructions to personalize the electronic circuitry, in order to perform aspects or operations.

[0053] These computer readable program instructions may be provided to a processor of a general purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions, which execute via the processor of the computer or other programmable data processing apparatus, create means for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks. These computer readable program instructions may also be stored in a computer readable storage medium that can direct a computer, a programmable data processing apparatus, and/or other devices to function in

a particular manner, such that the computer readable storage medium having instructions stored therein comprises an article of manufacture including instructions which implement aspects of the function/act specified in the flowchart and/or block diagram block or blocks.

[0054] The computer readable program instructions may also be loaded onto a computer, other programmable data processing apparatus, or other device to cause a series of operational steps to be performed on the computer, other programmable apparatus or other device to produce a computer implemented process, such that the instructions which execute on the computer, other programmable apparatus, or other device implement the functions/acts specified in the flowchart and/or block diagram block or blocks.

[0055] The flowchart and block diagrams in the Figures illustrate the architecture, functionality, and operation of possible implementations of systems, methods, and computer readable media according to various embodiments. In this regard, each block in the flowchart or block diagrams may represent a module, segment, or portion of instructions, which comprises one or more executable instructions for implementing the specified logical function(s). The method, computer system, and computer readable medium may include additional blocks, fewer blocks, different blocks, or differently arranged blocks than those depicted in the Figures. In some alternative implementations, the functions noted in the blocks may occur out of the order noted in the Figures. For example, two blocks shown in succession may, in fact, be executed concurrently or substantially concurrently, or the blocks may sometimes be executed in the reverse order, depending upon the functionality involved. It will also be noted that each block of the block diagrams and/or flowchart illustration, and combinations of blocks in the block diagrams and/or flowchart illustration, can be implemented by special purpose hardware-based systems that perform the specified functions or acts or carry out combinations of special purpose hardware and

computer instructions.

[0056] It will be apparent that systems and/or methods, described herein, may be implemented in different forms of hardware, firmware, or a combination of hardware and software. The actual specialized control hardware or software code used to implement these systems and/or methods is not limiting of the implementations. Thus, the operation and behavior of the systems and/or methods were described herein without reference to specific software code—it being understood that software and hardware may be designed to implement the systems and/or methods based on the description herein.

WHAT IS CLAIMED IS:

1. An Open Radio Access Network (O-RAN) system comprising:

an O-RAN Cloud platform (O-Cloud) comprising a plurality of physical nodes configured to host virtualized network functions (VNFs), the VNFs comprising O-RAN centralized units (O-CUs) and O-RAN distributed units (O-DUs);

a Service Management and Orchestration (SMO) framework configured to manage and orchestrate the O-Cloud, the SMO comprising anchored functionalities, a non-real-time radio access network (RAN) intelligent controller (Non-RT RIC), an O1 interface for communication with the VNFs and an O2 interface for communication with Infrastructure Management Services (IMS) and Deployment Management Services (DMS) of the O-Cloud; and

a near-real-time RAN intelligent controller (Near-RT RIC) configured to perform RAN optimization using policies and data provided by the Non-RT RIC,

wherein the Non-RT RIC is configured to implement one or more applications (rApps) to:

collect O-Cloud inventory and telemetry data on node level or O-Cloud level over the O2 interface,

collect cellwise traffic data, including mobility and neighbor cell load data, over the O1 interface,

train and deploy an artificial intelligence/machine learning (AI/ML) model to generate policies based on collected O1 and O2 data, and

generate policies for the O-Cloud or VNFs based on at least one of priority, load, and energy consumption, and

wherein the anchored functionalities of the SMO framework are configured to receive and manage the policies for the O-Cloud from the Non-RT RIC, and generate optimization controls

for transmission to the O-Cloud via the O2 interface and/or transmit the policies to the IMS and DMS via the O2 interface for implementation.

2. The O-RAN system according to claim 1, wherein the anchored functionalities of the SMO framework comprise a Network Function Orchestrator (NFO) and a Federated O-Cloud Orchestration and Management (FOCOM).