[WORKSHOP]Towards Greener Networks: RApp-Based Cell Control over O-RAN Deployments

1st Given Name Surname
dept. name of organization (of Aff.)
name of organization (of Aff.)
City, Country
email address or ORCID

***Abstract*—Abstract.**

***Index Terms*—Keywords.**

I. Extra

- Blue : Things to Highlight/Expand
- Red : Questions to ask

II. Introduction

Introduction.
- The objective of this paper is to describe an rApp hosted by a Non-Real-Time RIC which controls a set of gNBs in a 5G network, which has the capability of reducing the network energy consumption during low traffic period by selectively shutting down specific sectors.
- - While a mobile network consists of multiple parts, the radio access network (RAN) is responsible for most of the energy consumption in a mobile network. ==¿ Green Future Networks – Sustainability Challenges and Initiatives in Mobile Networks by NGMN Alliance, December 2021
- A radio access network comprises of several cell sites with site infrastructure equipment and base station equipment.
- O-RAN can contribute to this effort immensely. Its disaggregated and virtualized architecture adds complexity; however, energy is the next major challenge O-RAN must overcome.
- Rimedo Labs recently released Energy Saving rApp (ES-rApp) for Non-RT RIC focusing on Massive MIMO use cases.

- The RF carrier shutdown feature (typically hosted by the SMO and non-RT RIC in O-RAN architecture) periodically checks the service load of multiple carriers and if the service load is below a specified threshold, the capacity-layers are shut off (see Figure 12). The UEs served by those carriers can camp on or access the services from the carrier providing basic coverage. When the load of the carrier providing basic coverage is higher than a specified threshold, the base station turns on the carriers that have been shut down for service provisioning.
- RF carriers can be shut down by non-RT RIC rAPPs more intelligently using information from telemetry data from E2 interface with O-RAN.
- Switching off under-loaded cells during network operation without affecting the user experience (call drops, QoS degradation, etc.) is one way to achieve RAN energy efficiency.
- A typical energy savings scenario is realized when capacity booster cells are deployed under the umbrella of cells providing basic coverage and the capacity booster cells are switched off to enter dormant mode when its capacity is no longer needed and reactivated on a need basis.
- When the booster gNB with CU/DU split decides to switch off cell(s) to the dormant state, the decision is typically made by the gNB-DU based on cell load information or by the OAM entity (non-RT RIC in O-RAN architecture). Before the cell in the gNB-DU enters into the dormant mode, the gNB-DU will send the gNB-DU configuration update message to the gNB-CU to indicate that the gNB-DU will switch off the cell subsequently sometime later.
- During the switch-off period, the gNB-CU offloads the UE to a neighboring cell and simultaneously will not admit any incoming UE to this cell being switched-off. Is this load balancing performed?
- https://networkbuilders.intel.com/solutionslibrary/a-holistic-study-of-power-consumption-and-energy-savings-strategies-for-open-vran-systems

- RF channel switch off/on
- However, the switch off/on decisions are need a lot of KPIs reporting and efficient actions so that guarantee the overall user experience. Also, there are conflicting targets between system performance and energy savings.

- Offline learning is normally preferred (including reinforcement learning which is usually performed online) due to the nature of the network environment, which is prone to misconfiguration and errors leading to outages. The proposed approach is for the model to first be trained with offline data and that the trained model then be deployed in the network for inference. - ES-rApp aims at minimizing the overall energy consumption by switching off cells that are not loaded too much, and turning on the cells if the traffic load goes high.

- If the shift of all the users from a given cell is possible (i.e., if the performance requirements of already served users can be guaranteed), it generates the appropriate policy and sends it via the A1 interface using A1-P to the Near-RT-RIC.

- First, the E2 Nodes are configured by the Service Management and Orchestration (SMO) to report the data necessary for energy-saving algorithms via the O1 Interface to the Collection and Control unit. Assuming that the Non-RT RIC and SMO are tightly coupled the NonRT RIC retrieves the collected data through internal SMO communication how???. The O-RUs are involved in this use case. The E2 Nodes need to configure them to report data through the Open RAN Fronthaul Management Plane (Open FH M-Plane) interface.

- 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.).

A. *Key Questions*

Key questions to answer to implement above:

- Key question 1 - When to turn off and on the cells?
- Key question 2 - Which cells to turn on and off?
- Key question 3 - what is the goodness measure of particular energy saving decision

III. Related Works

Network traffic prediction has always been a largely explored subject in networking, with a flurry of recent proposals ushered in by the recent development of machine and deep learning tools. Such deep learning-based algorithms have recently been explored to find potential representations of network traffic flows for all types of networks, including Internet, cellular, etc. We first categorize cellular traffic problems into two main types – temporal prediction problems and spatiotemporal prediction problems. Modelling the traffic flow through a node exclusively as a time series is an example of the temporal approach towards network traffic prediction [11]. High traffic on a given node in a cellular network often implies a high load on the other nearby nodes. Taking the traffic flow of nearby nodes and other external factors into consideration when modelling is known as the spatiotemporal approach to network traffic prediction. Spatiotemporal approaches are found to give slightly more accurate forecasts [12].

Both types of problems can be formulated as supervised learning problems with a difference being in the form of feature representation. In the temporal approach, the collected traffic data can be represented as a univariate time series and the prediction for the values in the future time steps is based on the historical data of the past time steps. In [13], Clemente et Al used Naive Bayes classification and the Holt-Winters method to perform the temporal network forecasting in real time Clemenete et Al first performed systematic preprocessing to reduce bias by selecting the cells with less missing data occurrences, which was then selected to train the classifies to allocate the cells between predictable and non- predictable, taking into account previous traffic forecast error.

Building upon the temporal approach, Zhang et al. [14] presented a new technique for traffic forecasting that takes advantage of the tremendous capabilities of a deep convolutional neural network by treating traffic data as images. The spatial and temporal variability of cell traffic is well captured within the dimensions of the images. The experiments show that our proposed model is applicable and effective. Even with the ease of machine learning implementations, regression based models have been found to be fairly accurate, as proven by Yu et Al in [15]. In [15], Yu et Al applied a switching ARIMA model to learn the patterns present in traffic flow series, where the variability of duration is introduced and the sigmoid function describes the relation between the duration of the time series and the transition probability of the patterns. The MGCN-LSTM model, presented in [16] by Len et Al, was a spatial-temporal traffic prediction model which implemented a multi-graph convolutional network (MGCN) to capture spatial features, and a multi-channel long short-term memory (LSTM) to recognise the temporal patterns among short-term, daily, and weekly periodic data. The proposed model was found to greatly outperform commonly implemented algorithms such as ARIMA, LSTM and ConvLSTM.

Hybrid models can handle a variety of data types and structures, making them ideal for diverse applications along with combining the best features of different methodologies. This very principle is proven by Kuber et Al in [17] which proposes a linear ensemble model composed of three separate sub-models. Each sub-model is used to predict the traffic load in terms of time, space and historical pattern respectively, handling one dimension particularly. Different methodologies such as time series analysis, linear regression and regression tree are applied to the sub-models, which is aggregated and found to perform comparable to a ResNet-based CNN model. Another approach for the same is highlighted in [18] Tian et Al. The approach involves analysing the chaotic property of network traffic by analyzing the chaos characteristics of the network data. [18] proposes a neural network optimization method based on efficient global search capability of quantum genetic algorithm and based on the study of artificial neural networks, wavelet transform theory and quantum genetic algorithm. The proposed quantum genetic artificial neural network model can predict the network traffic more accurately compared to a similarly implemented ARMA model.

- This implementation is akin to a traffic steering xApp, as it involves offloading traffic from low-load cells to other cells, ensuring that as many RUs as possible can enter a low-power sleep mode. The handovers and traffic redistribution are predicted and managed in real-time by the xApp, enhancing the overall energy efficiency of the network. ==¿ https://www.diva-portal.org/smash/get/diva2:1765998/FULLTEXT01.pdf

- Ericcson, Rimodo Labs, Juniper Networks, VMWare have preexisiting rApps on the market, but have not described them very well.

- During the ES-rApp operation, it gets information from the O1 interface about cells available in the network, their type (macro cell or small cell), and the PRB usage of each cell. ES-rApp collects such data during predefined times and calculates average O-DU PRB usage in the time domain. Periodically, the rApp can make decisions about enabling or disabling one of the cells. ES-rApp will enable a cell in case of congestion (high PRB usage) observed in at least one cell that is currently enabled. ES-rApp will disable a cell in case of average PRB usage below some threshold. If none of the situations occurs, the ES-rApp continues observation. ==¿ https://rimedolabs.com/blog/es-rapp-control-over-ts-xapp-in-oran/

- VMWare and VIAVI ==¿ https://www.virtualexhibition.o-ran.org/classic/generation/2024/category/intelligent-ran-control-demonstrations/sub/intelligent-control/394

- Juniper's implementation ==¿ https://blogs.juniper.net/en-us/service-provider-transformation/delivering-on-the-o-ran-promise-with-juniper-networks-ran-intelligent-controller-ric

last 1 para for the novelty of our paper compared to existing work. should i write anything here? *something like:* To the best of our knowledge, comparisons between ARIMA, Prophet, LSTM trained on a single dataset but applied in various conditions have not been well studied in the literature

IV. Algorithm Description

A. ES rApp

Energy Saving rApp.
- The rApp is data driven in the sense that it does not incorporate a rules-based logic but determines the rules which meet the target objective based on the input data and network configuration. The rApp has the following components:
- [PULAK] More information on each of these sub 1. The Radio Database, which is a geospatial database, where data is indexed using the latitude, longitude, and altitude and includes the clutter information. For a mobile network, this database is initialized with the network inventory information and the predicted RF power (downlink) for each pixel from all the sectors in the vicinity greater than a predefined threshold (Pth) using a Radio Link budget simulator using the local clutter data. All the timestamped measurement reports from the gNBs are stored in the database as a sliding window with preconfigured depth (td). This Radio Database can also be external to the rApp so that it can be shared across multiple rApps. It can also import data from RF link simulators and drive tests through an external interface.
2. The Traffic Predictor, which predicts the net traffic volume, Percentage PRB utilization and the number of active UEs as a function of time for each sector based on historical data and the previous measurements stored. Based on the historical inputs ARIMA/SERIMA Algorithm is used to predict those (all the 3 bullet points above) values for the future. The input information is expected in the time interval of 15 minutes.
3. The Coverage Predictor, which predicts the overlap of coverage of each sector with respect to the adjacent sectors. The Coverage Predictor also updates the link level prediction model based on the measured values.
4. The Energy Saving Decision Entity: Simulation which uses the results of the Traffic Predictor and the Coverage Predictor to determine the sectors in the network, which can be shut down with the least impact in the services. The Simulator evaluates the impact of the Energy Saving Decision Entity before configuring the network through the SMO or EMS
4. The Anomaly Detection Entity, which matches the real-time measurements and the results of the Traffic and Coverage Predictor to find anomalies. If there is a major deviation of the measured and predicted values, the energy saving decision making process is suspended.

A Dashboard for visualization of the Radio Mapping Database is also used. *How does this link to the TS xApp?*

B. Algorithm

[IMP] - how do you decide whether the system is to be considered for shutdown or bringup?
- use of SINR?
- where does the prediction part come into play?

V. Methodology

Methodology.

A. Model Selection

Model Selection.

B. Data Selection

Data Selection.

C. Performance Metrics

Performance Metrics.

VI. Experiments and Results

CQI + total throughput + energy consumption.

- [This should otherwise describe the overall system architecture and where does the rApp reside and with which other components it interacts with.]
- Use Case Description - [This section shall describe the use

Algorithm 1: CellShutdown Procedure

```
Data: Lat-long of all the transmitters (antenna characteristics, antenna height, etc.)
Data: Field strength at any point from all the transmitters (using CloudRF)
Data: Entire geography subdivided into smaller areas (30m x 30m) each represented by a pixel obtained from CloudRF map
1  for each pixel do
2      Find the RX power from each BS (more than a certain threshold);
3      Find the strongest BS;
4      Calculate SINR = power of strongest cell / (sum of all remaining cells' power + noise power);
5  end
6  for each BS do
7      Find the list of neighbors;
8      Find the count of pixels where:;
9        I. Serving BS is dominant;
10       II. Neighbor is within a threshold of the serving cell;
11     This count is degree of overlap C_ij;
12     Total overlap is C_ij + C_ji;
13     Weight of each edge of the undirected graph is the degree of overlap;
14     Calculate the rank M_i = (K · ∑ degree of the node) / (Traffic volume at node (i)) (let K = 1 for now);
15 end
16 Sort the list of BS in order of descending rank; the top one is the candidate for shutdown;
17 Select the top candidate in rank list above for shutdown;
18 With this candidate being shutdown (power = 0), find the SINR distribution (in digital twin);
19 if the SINR distribution is within bounds then
20     Break the loop;
21 end
22 else
23     Continue to step 1;
24 end
```

Algorithm 2: CellBringup Procedure

```
Data: For each cell we have desired pattern for CQI distribution for each traffic pattern and time of day
for each cell do
    Measure the distance between CQI distributions observed and desired;
    if the observed is off by a certain threshold then
        Turn the cell ON;
        Recalculate the predicted CQI distribution;
        if the distribution improves then
            Keep the cell ON;
        end
        else
            Leave the cell OFF;
        end
    end
end
```

case applicable to the energy saving scheme implemented here. Please use PlantUML or equivalent to describe the interaction between several actors which achieve the end result. Refrain from giving details of operation at each actor and focus on the interactions]

VII. Conclusions

Conclusions.

VIII. Acknowledgements

Acknowledgements.