Smart Operation of 6G RAN supported by Autonomous Transport Networks

Shaoxuan Wang, Marc Ruiz*, and Luis Velasco
Optical Communications Group (GCO), Universitat Politècnica de Catalunya (UPC), Barcelona, Spain
e-mail: marc.ruiz-ramirez@upc.edu

ABSTRACT
Autonomous operation of future networks will require tight coordination of Radio Access Networks (RAN) and fixed transport networks operation, in order to achieve smart end-to-end operation. Since transport network operation allows dynamic capacity allocation management, dynamic RAN operation is also enabled and supported. In this regard, one key topic is to adapt RAN capacity by switching on/off base stations (BSs) following users demand, which reduces energy consumption while guaranteeing quality of service. In this paper, we present an approach that consists in processing RAN monitoring data by means of machine learning algorithms to decide actions to switch on/off BSs in order to guarantee cell coverage with minimum capacity requirements.

Keywords: B5G/6G, smart RAN, AI, Machine Learning

1. INTRODUCTION

Future radio access networks (RAN) will operate with massive and heterogeneous deployments and end-to-end (e2e) connectivity in support of diverse B5G/6G use cases. At the same time, energy efficiency and consumption will be a major design criterion in 6G along with other metrics such as capacity, peak data rate, latency, and reliability. Thus, cost-effective networks require solutions providing high adaptability that allow providing just the right capacity, thus eliminating overprovisioning and wasting. This requires near-real-time control that can be supported through closed control loops exploiting zero-touch and intent-based networking paradigms to enable smart operation of both RAN and fixed transport segments [1].

A key operational objective in dense and heterogeneous RAN is to reduce energy consumption. This can be achieved by managing the number of active base stations (BS) that are required to support the current user equipment (UE) traffic requirements [2]. The key question is when and which BSs in the RAN should be activated/de-activated in order to achieve remarkable energy consumption reduction, as well as maintaining committed Quality of Service (QoS) of end users.

Among different next generation RAN architectures, O-RAN is promising great potential for flexible and dynamic RAN operation [3]. Thus, the RAN intelligent controller (RIC) architecture is defined as a set of components and interfaces that allow collecting RAN monitoring data from several RAN functions, process them by means of artificial intelligence (AI) and machine learning (ML) procedures running at both near-real-time and/or non-real-time scopes, and implement directives to dynamically configure RAN resources.

In this paper, we rely on O-RAN capabilities and propose a methodology running in the RIC that combines supervised and unsupervised ML models with rule-based procedures for smart RAN operation. In particular, different features that are continuously monitored in 6G RAN are processed in order to detect in which cells the QoS of UEs is worsening, thus triggering the activation of neighboring BSs. On the other hand, those BSs covering cells with high QoS are candidate to be de-activated for energy consumption minimization purposes.

2. SMART RAN OPERATION

Fig. 1a shows the RAN scenario considered in this work. Without loss of generality, we assume that a given RAN contains a number of BSs providing coverage and connectivity to a number of *cells* (not depicted for the sake of clarity). We assume that macro BSs (MBS) provide full coverage within its area and provide the minimum capacity to absorb users' traffic, whereas micro BSs (µBSs) complement the capacity of the MBS within a limited number of neighboring cells. We assume that µBSs provide two operational modes: (*i*) *active*, where the µBS is switched on and fully operational, and (*ii*) *sleep*, where the µBS is switched off. The RAN area serves a number of UEs that belong to a mix of services including enhanced mobile broadband (eMBB), ultra-reliable low-latency communications (URLLC), and massive Internet of Things (mIoT) [4]. Fig. 1a also depicts an illustrative scenario happening at time t, where the current BSs status and UE demand is creating some saturation that is affecting the QoS of some services (represented by colored gauges). Note that there are few BSs that are switched off, which enables the possibility to activate them and offload part of the UE traffic causing such congestion. However, it is necessary to know where are the UEs having worse performance and if coincides, activate BSs covering the cells where those affected UEs are. This will increase the overall energy consumption, unless some active BSs that are covering cells with low UE demand and/or high QoS could be de-activated.

In view of the above, we propose the smart RAN operation sketched in Fig. 1b. It consists in two main steps: *i*) *cell classification*, and *ii*) *µBS management*. The first procedure aims at processing RAN monitoring data

979-8-3503-7732-3/24/$31.00 ©2024 IEEE

obtained from different sources, including the radio units (RU), distributed units (DU) and centralized units (CU) of the different active BSs in the area [4]. This monitoring data contains features such as channel quality indicator (CQI), throughput, and latency, which are strongly correlated with UEs' QoS. Indeed, per-UE raw data is typically collected (including geo-location), so cell monitoring data can be easily obtained by aggregating UE monitoring data. To allow accurate monitoring, a selected set of BSs covering all the cells by, at least, 2 different BSs are always active. This cell monitoring data is then used to classify cells into different categories. Fig. 1b illustrates those categories by means of colors that highlight the need of activating or de-activating µBSs closer to such cells. In particular, the example shows that cells with bad QoS (orange/red) currently contain µBSs that are switched off, whereas cells with good QoS (green) contain an excess of active µBSs.

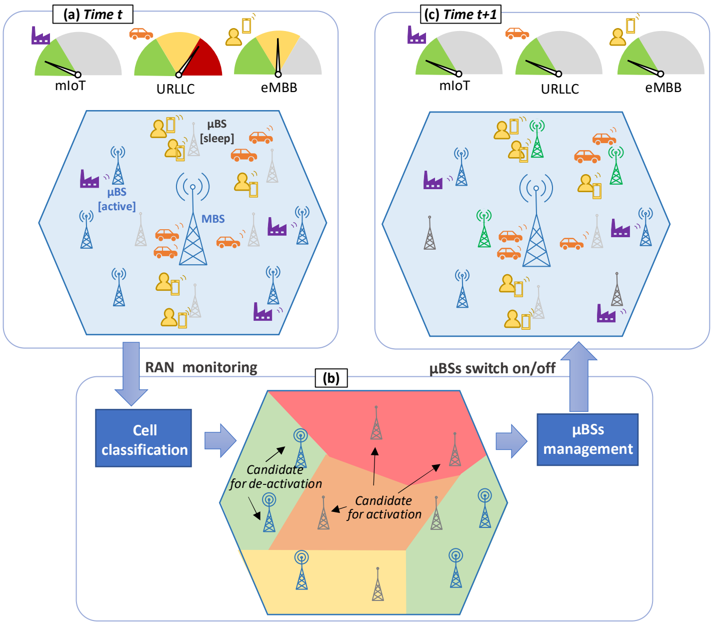

Fig. 1: Smart RAN working procedure

Thus, this classification is processed in the second step, that applies a rule-based approach to make µBSs activation/de-activation decisions. Fig. 1c illustrates the RAN status at time t+1 after applying the reconfiguration instructed by the smart RAN operation procedure. As a result of this reconfiguration, overall UEs QoS has improved with respect to previous time t.

3. CELL CLASSIFICATION AND µBS MANAGEMENT PROCEDURE

In this section, we describe the different models and algorithms involved in the smart RAN procedure sketched in Fig. 1b. The tackled scenario consists in a RAN area composed by a set I of cells that are covered by a set S of µBSs that can be switched on/off (recall that other BSs not in S are always active). The map Π contains the assignment of which cells can be covered from each µBS. Based on reference O-RAN architecture [3], we assume that the non-real time RIC can collect and process cell monitoring samples. Thus, $x_i(t)$ denotes the cell monitoring sample collected at time t in cell i. Without loss of generality and according to [5], each $x_i(t)$ sample contains an array of features that includes measurements about the number of active UEs, channel quality indicator (CQI) metrics, throughput, and latency, among others.

Let us assume that a set of cell monitoring samples $X=\{x_i(t), \forall i\in I, t\in T\}$ collected during a large time period T (e.g, a month) is available for training purposes. The first step through cell classification is to run an unsupervised ML procedure in order to identify clusters that will be afterwards used as classes in a supervised ML procedure. This clustering procedure consists in two differentiated phases. On the one hand, a Self-Organized Map (SOM) is used for dimensionality reduction of the original input [6]. On the other hand, a k-means procedure uses the output of SOM model as input to find the centroid of k clusters. In order to find the optimal number of clusters k*, we use the value that minimizes the Davies & Bouldin index (DBI) (eq. 1), which can be computed as the ratio of the sum of centroid (x') intra-cluster distances (Δ) and inter-cluster distances (δ) (eq. 2).

$$k^* = argmin_k\, DBI(k) \quad (1) \qquad DBI(k) = \frac{1}{k}\sum_{i=1}^{k} \underset{j\neq i}{max}\left\{\frac{\Delta(x'_i)+\Delta(x'_j)}{\delta\left(x'_i, x'_j\right)}\right\} \quad (2) \qquad z_s(t) = \frac{1}{|\Pi(s)|}\sum_{i\in\Pi(s)} g(f(x_i(t))) \quad (3)$$

As a result of the previous clustering, the new dataset $Y=\{<c, x'_c>_i, \forall i\in I\}$ is generated, containing for each cell i the id c and the centroid x'_c of the cluster the cell belongs to. This cluster id is the label to be predicted taking as input the centroid data. For the classifier, we use a decision tree (DT) that is trained in order to minimize the Gini impurity level, while keeping a moderated depth and minimum number of elements in each leaf node [7]. The resultant DT f is then used to predict, given an input sample $x_i(t)$, the class $c_i(t)$ the cell belongs at that specific time t, which will depend on the activity of UEs and the status of neighboring BSs.

Due to the good explainability properties of DT, an inspection of trained f is then performed in order to identify the set of cell requirements Q that each class holds. Specifically, we create the look-up table g that, for each class, assign one of the following requirements, each identified with a numerical value q: *i*) the cell requires activating a µBS (q=+1), *ii*) the cell can admit switching off a µBS (q=-1), *iii*) no changes are required (q=0). Then, we define score $z_s(t)$ as the requirement of BS s at time t, computed as the average requirement of the cells that that BS covers (eq 3). This score value needs to be eventually compared against a set of manually setup thresholds R to decide the main actions to perform in a BS, i.e, switch on, switch off, or keep the BS with its current status.

Finally, Algorithm 1 sketches the proposed smart RAN procedure that is run once the abovementioned models are trained and available for usage. It is called every time a new monitoring sample $X(t)=\{x_i(t), \forall i \in I\}$ is available and returns the set of actions $A(t)$ to perform, i.e, which µBS need to be switched on/off. After some initialization (line 1 of Algorithm 1), the numerical requirement of each cell is computed (lines 2-3). Then, for each µBS that can be dynamically managed, the score is computed according to the expression in eq. 3 (lines 4-8). The score value is then compared with thresholds R to find the potential action a (line 9). Finally, if and only if the action is different than keeping current status, the action is stored and returned (lines 10-12).

Algorithm 1 – Smart RAN procedure

IN: $X(t), f, g$ **OUT**: $A(t)$

1: $A(t) \leftarrow \emptyset$
2: **for** $i \in I$ **do**
3: $\quad q_i \leftarrow g(f(x_i(t)))$
4: **for** $s \in S$ **do**
5: $\quad z \leftarrow 0$
6: $\quad$ **for** $i \in \Pi(s)$ **do**
7: $\quad\quad z \leftarrow z + q_i$
8: $\quad z \leftarrow z \,/\, |\Pi(s)|$
9: $\quad a \leftarrow$ get_action (z, R)
10: $\quad$ **if** $a \neq$ "keep" **and** $a \neq s$.status **then**
11: $\quad\quad A(t) \leftarrow A(t) \cup$ <s.id, a>
12: **return** $A(t)$

4. ILLUSTRATIVE RESULTS

In order to evaluate the proposed approach, we used the real dataset in [5] containing 1440 monitoring data samples of 63 cells collected during T=15 days in a major European city. Every sample contains 29 different features with statistics of metrics related with number of UEs, data volume, throughput, physical resource block (PRB) utilization, handover (HO) failure, radio resource control (RRC) failures, CQI, and latency.

We start focusing on the evaluation of the two-phase clustering approach, which performance is summarized in Fig. 2. First of all, it can be observed that SOM allowed to reduce from 29 to just 9 relevant features (see legend for feature details). Using this reduced feature set, k-means clustering is executed with the DBI optimization, which results in an optimal number of clusters k*=5. The figure shows the distribution of features in each cluster, plotting the relative frequency of each of them. In view of the graphs, it can be concluded that clusters identify cells with different feature patterns, which anticipates promising usefulness to characterize cells for RAN management purposes.

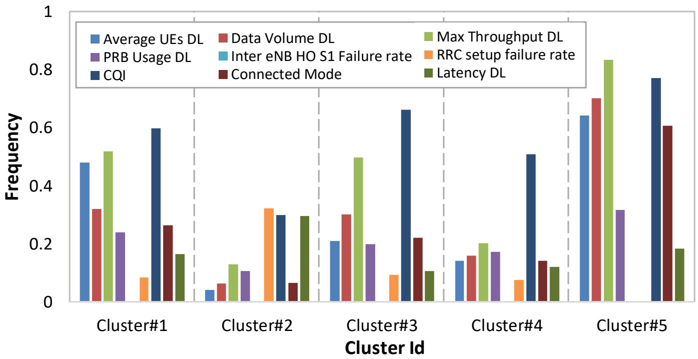

Fig. 2: Features distribution of behavior cell patterns

Taking the five clusters as true labels, a DT is trained for classification purposes. Fig. 3 shows the obtained model f, indicating the Gini obtained at every node and the number of samples that belong to each class. Note that every leaf node has Gini=0, which means that perfect classification is done. In fact, the number of leaf nodes equals the number of classes, which allows remarkable significance of every decision rule. By inspecting the obtained tree, the lookup table g in Table 1 can be easily obtained. Note that class 1 comprises cells with low throughput and latency and high CQI, showing that sufficient QoS is provided, which opens the opportunity to switch off an active neighboring µBS (q=-1). Then, classes 2 and 3 represent situations where QoS is ensured but with no excess of resources which leads to keep the current status (q=0). Finally, classes 4 and 5 represent situations that show resources saturation and poor QoS, which lead to trigger BS activation (q=1).

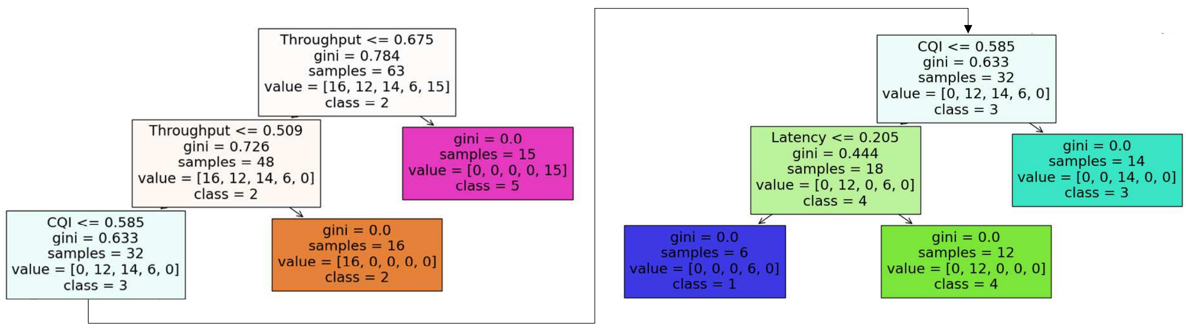

Fig. 3: Classifier f: model and accuracy

Once models f and g have been trained and validates with real data, we eventually show the application of Algorithm 1 in a simulated 6G scenario consisting in a dense urban area with 64 cells and one μBS per cell. Fig. 4 shows two different cases and four plots for each case (from left to right): *i*) the active BSs at time *t*, *ii*) the normalized number of UEs per cell, *iii*) the *q* values of every cell after processing cell monitoring data, and *iv*) the active BSs at time *t*+1, after actions *A*(*t*) are implemented. In the first case (Fig. 4a), the initial BS map contains only those BSs that are always active (which represents 50% of μBS). Due to the concentration of UEs demand in the upper left part of the map, some cells in that region require BS activation. As a result of this, 8 BSs that were in sleep mode at time *t* are switched on and become active at *t*+1.

Table.1 Look-up table *g*

Class	Features Summary	q
1	Low Throughput High CQI Low Latency	**-1**
2	Medium Throughput	0
3	Low Throughput High CQI	0
4	Low CQI High Latency	**+1**
5	High Throughput	**+1**

In the second case (Fig. 4b), we start from the previous map and emulate a mobility of the UEs towards a different region in the map where they are now concentrated. This movement reduces the UE demand in top left cells, while increasing it in the bottom right ones. The *q* values catch that mobility, as well as the need to move BS capacity following the UE demand. Thus, 6 BSs in the top left region are switched off, while 12 BSs in the bottom right area are switched on. Note that the increasing RAN capacity will require coordination with fixed transport networks to properly adapt e2e connectivity capacity in support of new active BSs [4].

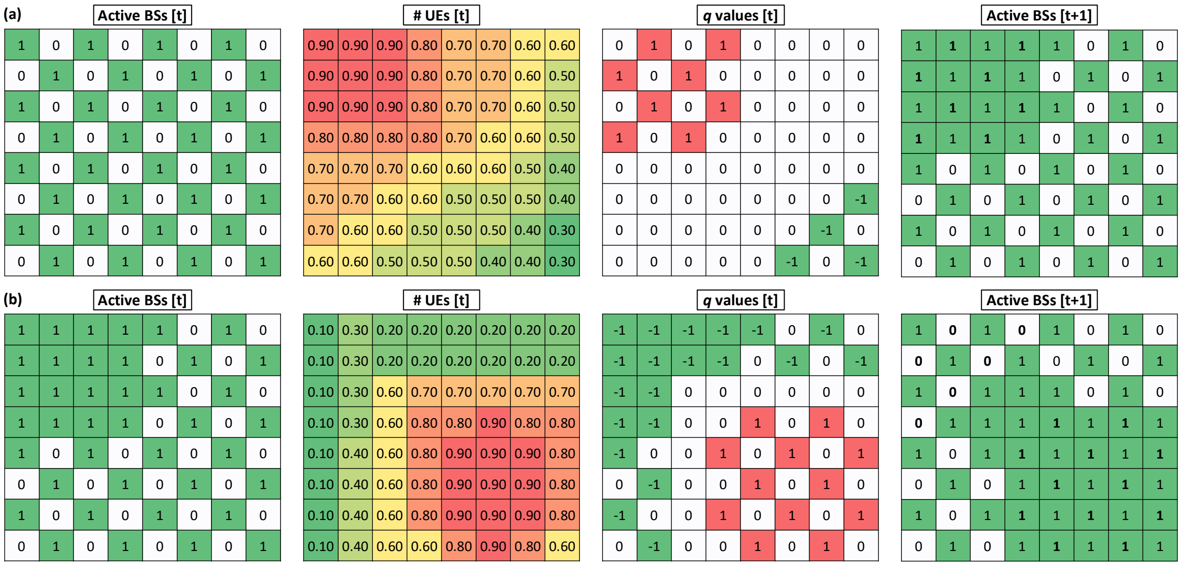

Fig. 4: Illustrative results in a dense urban scenario

5. CONCLUSIONS

A smart RAN operation procedure suitable for e2e autonomous network operation has been proposed. The overall approach is based on ML models and processes RAN monitoring data to decide which BSs need to be switched on/off in order to reduce RAN energy consumption while guaranteeing UEs QoS. ML models have been trained and tested with real data and the complete procedure has been evaluated in a simulated 6G scenario.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the Smart Networks and Services Joint Undertaking under the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101096120 (SEASON), and the MICINN IBON (PID2020-114135RB-I00) projects and from the ICREA Institution.

REFERENCES

[1] L. Velasco *et al*., "End-to-End Intent-Based Networking," IEEE Comm. Mag., vol. 59, pp. 106-112, 2021.

[2] Y. Zhu and S. Wang, “Joint Traffic Prediction and Base Station Sleeping for Energy Saving in Cellular Networks,” in Proc. IEEE Int. Conf. on Communications (ICC), 2021.

[3] O-RAN ALLIANCE WG1, “O-ran architecture description, v.8.0,” Tech. Rep., 03 2023.

[4] S. Wang, M. Ruiz, and L. Velasco, "Context-based e2e Autonomous Operation in B5G Networks," MDPI Sensors, vol. 24, pp. 1-25, 2024.

[5] S. Wang *et al*.,” Extracting Cell Patterns from High-dimensional Radio Network Performance Datasets using Self-Organizing Maps and K-means Clustering”, IEEE Access, vol.9, pp.2169-3536, 2021.

[6] Y. Liming and L. Tianyao,” A K-means Optimized Self-Organizing Map Neural Network Video Recommendation Framework,” in Proc. IEEE Int. Conf. on Electronic Technology, Communication and Information (ICETCI),2023

[7] F. Posch *et al*.,” A Classifier for Aerial Users in 5G Networks”, in Proc. IEEE Globecom Workshops (GC Wkshps),2023.