Data Analysis and Energy Consumption Prediction in a Cloud-Fog RAN Environment

Matias Romário Pinheiro dos Santos*, Rodrigo Izidoro Tinini†, Gustavo Bittencourt Figueiredo*, Daniel Macêdo Batista†
*Universidade Federal da Bahia –UFBA; E-mail: matiasromario@ieee.org, gustavo@dcc.ufba.com
†Universidade de São Paulo – USP; E-mail: {rtinini, batista}@ime.usp.br

***Abstract*—The extraction of information from data collected in a myriad of environments provides unprecedented opportunities for a big range of actions such as decision making and better resource management. Benefits from its processes are relatively large for many network domains such as protocol design, hybrid architectures redesign, and resource management and optimization. Time series or historical data series provide can be used in several ways like pattern analysis and prediction support, making it an important support tool for managers to develop goals and objectives focused on their business. The goal of this paper is to discuss the potential of data analysis in hybrid Cloud-Fog Radio Access Networks (CF-RAN) scenarios and present results of applications of the data in the process of prediction energy consumption. In particular, we analysed the knowledge data extraction of some metrics with a strong relationship with energy consumption and we perform a prediction by applying a deep learning algorithm using the previous four hour period to predict the next hour.**

I. Introduction

The fifth generation of mobile networks (5G) is expected to enable high volumes of user and industrial data, expansion of consumers service and also allow an increasing number of mobile devices to connect to the network. This is a result of a combination of important features proposed to 5G. One of such features is the adoption of the Cloud Radio Access Network (C-RAN) architecture. C-RAN impulses a high centralized deployment to support collaborative radio technologies, virtualization, better resource management and energy consumption reduction by decoupling BaseBand Units (BBU) from cell sites and by centralizing the baseband processing from the Remote Radio Head (RRH) into BBUs pools. Although C-RAN produces gains, centralized baseband processing imposes strict delay and high bandwidth requirements to the fronthaul due to the use of the Common Public Radio Interface (CPRI) protocol [1], [2].

To address such problems, a new architecture called Cloud-Fog RAN (CF-RAN) [3] was proposed to increase the coverage of C-RAN while limiting energy consumption. CF-RAN architecture relies on the Fog and Cloud computing paradigms and on the Network Function Virtualization (NFV) technology. In CF-RAN, local processing nodes called fog nodes are placed closer to users and activated according to the network demand to alleviate both the cloud and fronthaul workloads. However, in spite of these benefits, problems associated with network resources activation as a result of traffic demand fluctuation in mobile networks emerge [4]. Such fluctuation directly impacts the performance of CF-RAN due to the activation of different resources during a day, degrading already established services or allocating more resources than necessary. Hence, an interesting tool to enable evaluation and specification of requirements and reconfiguration of resource allocation is the combination of Data Analysis (DAS) and Machine Learning (ML) prediction capabilities.

Some DAS techniques, like Data Analytics (DA), Exploratory Data Analysis (EDA) and Data Mining (DM), emerge as key tools for improving the performance of the 5G network [5]. We claim that such benefits are extensible to CF-RAN once flexibility, low cost, and adaptability to deal with new demands can be easily adjusted by the application of DAS. Also, the benefits of the prediction capabilities for network managers can be valuable when establishing the DAs with the fitness to mine meaningful data insights by ML.

ML enables ease way to the network management process, driven by the ability to anticipate network configuration. In CF-RAN, theses benefits are varied, including the ability to promote antecipation in migration of processing nodes in order to mitigate network latency, waste of computational resource or overload issues, minimizing several problems associated with traffic fluctuation and energy-inefficiency.

Therefore, in this paper we present the use of DAS to enhance the 5G knowledge focusing on the relation of important metrics to the CF-RAN architecture; fronthaul metrics analysis using Long Short-Term Memory (LSTM), a particular kind Deep Recurrent Neural Network (DRNN), to predict energy consumption in CF-RAN using the results obtained in[6] for training, validation and testing. Results demonstrated that LSTM predictions bring a clear overview of future energy consumption trends and an arrangement overview of resources used.

The rest of this paper is organized as follows: section II presents similar related works in approach and content aspects, we considered work that addresses DAS in 5G mobile networks, several LSTM applications and their use for time series prediction; section III presents the CF-RAN architecture; section IV present the background of the simulator, the LSTM algorithm formulation and the metrics for performance evaluation considered; in section V are presented the results and discussion acquired in the DAS and prediction results from

978-1-7281-3955-5/19/$31.00

the LSTM; finally, section VI presents the research conclusions and aspects.

II. Related Works

DAS applications in 5G network scenarios present a wide range of research opportunities, such as network slicing [7] and construction of frameworks that allows end-to-end support of data analytics to improve 5G radio resource management and to enable the technology for the service-based architectures (SBAs) [5].

Authors in [8] discuss the data revolution era and technologies that make smarter environments by the use of mobile devices and mobile networks. The main task is the development of applications based on mobile cloud sensing and its architecture. Also, the authors state that 5G and big data technologies are promising techniques for applications in various domains, including mobile cloud sensing.

Authors in [7] propose a methodology to provide slicing to 5G network Mobile Network Operators (MNOs). Such methodology is based on Key Performance Indicator (KPIs) applied to a clustered large data sample. Authors have applied machine learning-based algorithms over 18 months of KPIs collected from UEs to determine number of slices to support the largest amount of applications and services.

Authors in [9] explored several methods of integration of Big Data-Drive (BDD) and network optimization to increase the QoS of User's applications. They have explored and discussed several techniques to preform data collection and analysis.

Authors in [5] propose a 5G architecture to allow end-to-end DA and also affirmed that DAS is a powerful tool to provide improvements of 5G mobile network due to the architecture prediction capabilities and the several possible statistical collection and application. Requirements of adaptability and dynamism in the architecture development and deployment are more easily managed by applying statistical mechanism from the macro analysis of data.

Recently, several works make use of LSTM algorithm in a myriad of 5G network applications. For example, authors in [10] use LSTM to predict traffic for BBU pool resource reallocations in a C-RAN architecture using a reconfigurable optical add-drop multiplexer (ROADM). Authors in [11] proposed a channel state information (CSI) estimation using a combination of convolutional neural network (CNN) and LSTM to predict CSI with high efficiency.

Time series approaches applied to prediction focusing on energy consumption have also also been proposed. In [12], authors have investigated the performance of Deep Neural Network LSTM to predict levels of energy load. The authors concentrate on the decision-making approach for future accurate energy demand predictions comparing the approach in two steps, a minute and an hour time resolution.

In this paper, we present the benefits of data analysis in a CF-RAN architecture and the use of state-of-the-art Deep Learning algorithms to efficiently forecast energy consumption. Comparing with literature [11], [10], the authors did not present the analysis and evaluation of metrics used in their decision making.

III. CF-RAN Architecture

In the CF-RAN architecture (see Fig. 1) besides the regular connection to the BBU pool located in the Cloud, the RRHs are also connected to the Fog level where several local processing nodes called fog nodes can be used to process their CPRI traffic. In this architecture, the BBUs are virtualized into vBBUs, which are responsible for performing baseband processing hosted on containers called Virtual Digital Units (VDUs). These VDUs and their hosted vBBUs are dynamically activated or deactivated through NFV as function of the network demand. In this context, fog nodes are activated to decrease the overall network latency or to alleviate the network and processing demands in the cloud and in the fronthaul, respectively [3]. Apart from the capacity differences between the Cloud and the Fog, we consider that fog nodes and cloud nodes are composed of the same components and implement identical baseband processing functions with on-demand activation and deactivation by applying the NFV paradigm.

Time and Wavelength Division Multiplexed Passive Optical Network (TWDM-PON) technology implements the fronthaul. With TWDM-PON, virtualized dedicated PONs called virtualized PON (VPON) can be created as a function of the network demand to transport baseband signals from severals RRHs to a single processing node, either in the cloud or in a fog node. If the traffic exceeds the Cloud's processing capacity, fog nodes are activated on-demand. In case there is no need for fog nodes, they remain deactivated.

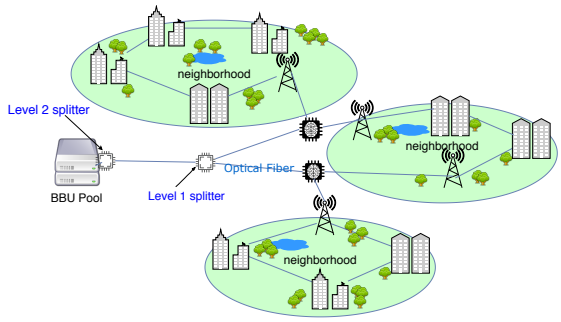

Fig. 1: CF-RAN Architecture

IV. Data Analysis Framework

DAS is performed to monitor and predict the network parameters and the behavior of CF-RAN attributes. Firstly, we collected and generated a dataset in *.csv format with several fronthaul metrics. Then, we decompose the data based on prediction or analysis functionalities regarding some priorities on different levels. Below, we present the main metrics analyzed:

Session and UE Measures: related to mediations that allow the prediction and analysis of the Users Equipment's (UEs) context (prediction of some QoS metrics, for

example). Also, it is associated with connection requests and resources availability.

Energy consumption Measure: the energy consumption measurements are associated with the active nodes and their cost. The dataset contains the total of active VDUs, the number of fog nodes used, redistribution of traffic and others.

Network Measure: This measure is associated with the application domain. In the fronthaul, we measure the total of available bandwidth, radio resource availability, total traffic load, processing load and others.

Performance Evaluation Measure: this analysis is associated with the number of request failure, a total of requisition processing blocked, blocking probability rate and others.

A. Long Short-Term Memory Algorithm

Long Short-Term Memory (LSTM), introduced by [13], is a complex Recurrent Neural Network (RNN) used in a plethora of applications such as voice recognition and text translation performed, for example, by Google [14]. Unlike Feed Forward Neural Network, LSTM and RRNs use the actual input data and past knowledge to improve the results. RNNs have several numbers of applications, including prediction of time series data based on previous time samples to predict the future configuration [15] and BBUs traffic prediction for resources reallocation [10]. This algorithm allows inferring analyzes on sequential or ordered data with long term relationships.

According to [16], LSTM has commonly in its composition a cell ($c^{(t)}$), an output gate ($o^{(t)}$), an input gate ($i^{(t)}$) and a forget gate ($f^{(t)}$). The cell contains values and the gates manages the information in the cell. The LSTM algorithm can be explained as follows:

$$z^{(t)} = g\left(W_z x^{(t)} + R_z \hat{y}^{(t-1)} + b_z\right) \quad (1)$$

$$i^{(t)} = \left(W_i x^{(t)} + R_i \hat{y}^{(t-1)} + p_i \odot c^{(t-1)} + b_i\right) \quad (2)$$

$$f^{(t)} = \left(W_f x^{(t)} + R_f \hat{y}^{(t-1)} + p_f \odot c^{(t-1)} + b_f\right) \quad (3)$$

$$c^{(t)} = i^{(t)} \odot z^{(t)} + f^{(t)} \odot c^{(t-1)} \quad (4)$$

$$o^{(t)} = \left(W_o x^{(t)} + R_o \hat{y}^{(t-1)} + p_o \odot c^{(t)} + b_o\right) \quad (5)$$

$$\hat{y}^{(t)} = o^{(t)} \odot h\left(c^{(t)}\right) \quad (6)$$

We can interpret the main components in this formulation as follows:

$z^{(t)}$: the equation 1 is the input and it is produced based on the current input and previous output;

$i^{(t)}$: the equation 2 is also know as input gate; it determines the amount of input to be retained in the cell state $c^{(t)}$;

$c^{(t)}$: the equation 4 is related with the current cell state. It is performed based on the $z^{(t)}$ update and the previous state;

$\hat{y}^{(t)}$: the equation 6 is the output and represents the analysis of the impact of the cell state in the output;

$f^{(t)}$: the equation 3 the forget gate and represents the amount of previous state that have to be removed or passed;

$o^{(t)}$: the equation 5 is the output gate and represents the analysis of the next hidden state, previous inputs.

B. Index of Performance in Experimental Results

Metrics performance assessment used to evaluate the model benefits and drawbacks are: mean absolute error (MAE), median absolute error (MADE), coefficient of determination (R^2) and root mean square error (RMSE).

Root Mean Square Error (RMSE): to evaluate the loss function in regressions. Calculate this metric by the root of the sum of the square distances between the predicted and the real value.

$$\sqrt{\frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}{T}}$$

T represent the time observed and $\hat{y}_t$ represents the predicted value in time t of y_t real value.

MAE: also used for regression models. This metric is associated with the sum of the absolute differences between predictions and real values.

$$\frac{\sum_{i=1}^{n} \hat{y}_i - y_i}{n}$$

$\hat{y}_i$ represent the predicted value of the y_i real value.

R^2 **Score**: also applied for regression models. This metric is associated to the analysis of agreement of the values observed by the model. That's mean that the larger the R^2 (near to 1), more explanatory the model is.

$$\sum_{i=1}^{n} (\hat{y}_i - y)^2$$

n is the observation length, $\hat{y}_i$ is the predicted value of y_i and the y is the mean of the observations.

V. Results

In this section, we explore the effects of data analysis in the average fronthaul metrics and in the analysis for an LSTM prediction. We consider a CF-RAN architecture with one Cloud and up to two fog nodes with maximum range of 20 km extension from RRH to fog and 20 km from Fog to Cloud. Also, we consider the following power consumption parameters for evaluation (see Tab. I):

TABLE I: Power Consumption Parameters. Based on [3]

Element	Value
Cloud	600 watts
Fog node	300 watts
VDU Cloud	100 watts
VDU Fog	50 watts
vBBU	20 watts
Line Card	5 watts
OLT	100 watts

Power consumption collected and used for prediction is presented in figure 2. Important information is about the stand-alone consumption that is equal to $600W$, but it begins, at the time 0h, equals to $0W$ in every single day.

We calculate the power consumption by collecting the consumption of all active processing elements in the environment, such as VDUs, processing nodes, LCs activated and vBBUs.

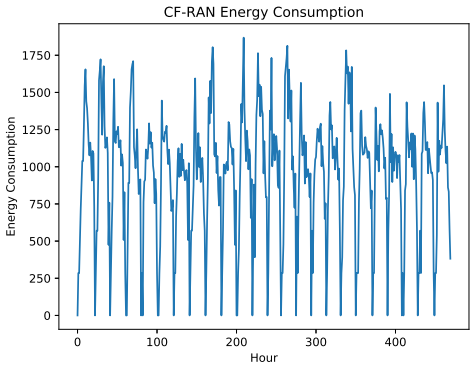

Fig. 2: Normalizes Plot of Energy Consumption in CF-RAN

In this paper, LSTM network was performed using 80% for training and validation, and the last 20% used for testing. In the training process, we used batch sizes of 12 and 1000 epochs for optimization. We present the metrics of performance evaluation used in subsection IV-B.

A. DAS Analysis Results

Among the analyzes performed in this process, we used the correlation matrix in the process of identifying metrics with a strong relationship with the energy consumption and, besides that, only identified metrics were considered for the rest of DAS steps. The metrics used are rearranged as present in Tab. II.

TABLE II: Summary of the metrics collected

Summary			
A	Energy	**J**	Activated dus
B	RRH redistribution	**K**	Avg. total. allocated
C	Total requested	**L**	Avg. time. inc. batch
D	Average act. switch	**M**	Avg. service availability
E	Avg. act. nodes	**N**	Avg. lambda usage
F	Avg. act. lambdas	**O**	RRHs available
G	Avg. act. dus	**P**	Arrival rate
H	Migrations	**Q**	Time
I	Actived lambdas		

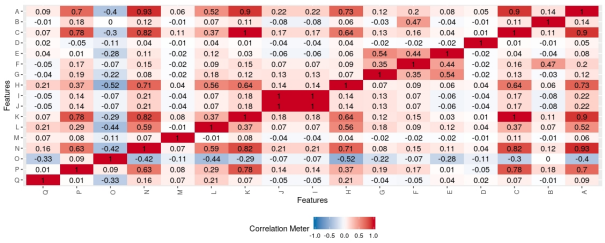

Fig. 3: Plot of Correlation Data

The correlation matrix is a standardization measure that establishes the relationship between two variables indicating their strength and their direction through the linear relationship between each other. Moreover, it is also possible to make this identification of this correlation by observing the pattern in a scatter diagram.

In Fig. 3, we identified a strong relationship between the energy cost and the following metrics: arrival rate; average lambda usage; average total allocated; migrations; total request. The arrival rate, for example, that is associated with the queuing theory and to the average arrival rate of the traffic in individual nodes of the network, its observed that the total amount of network traffic at the moment, when being at the peak, starts to activate the fog nodes to relieve the fronthaul, implying direct higher consumption of energy. Another example is the migration, associated with total redistributions of traffic between fogs or cloud fog, which occurs on-demand, which also implies a considerable increase in energy consumption.

Scatter plot is considered a useful tool for performing a series of data comparisons, allowing a visual demonstration of various data relationships to check for association with strength, direction and correlation.

In Fig. 4, we compared metrics identified in the matrix of correlations group by energy. Also, we get a better observation regarding the correlation already presented in Fig. 3. The values of the metrics grow with the increase of the energy values, which enables affirming that have a positive correlation. Besides, it is observed that the lambda usage, total allocated and the total request has a linear relationship with energy and enables affirming that it has a concordance. On the other hand, In arrival rate and migrations, there is a correlation, but the variables are independent of energy.

In Fig.5, arranged metrics in the matrix of correlations are

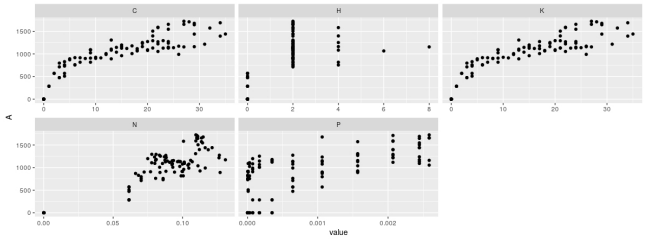

Fig. 4: Scatter Plot

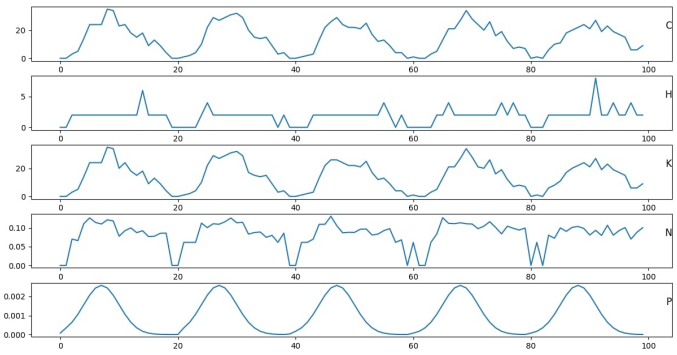

Fig. 5: Data Behaviour

represented by its data behavior in the dataset. We observed a strong association between total requests and avg. total allocated since the blocking probability is low. Fig. 5 also demonstrates the behavior of metrics within the dataset. For example, Arrival rate follows a Gaussian behavior given its direct relation to the generated traffic, which also follows a Gaussian behavior.

Boxplot of data is another useful tool that provides a complementary measure of the data character's perspective for discretionary data (outliers), as well as visual support of position, dispersion, and symmetry. Outliers were identified in the migration considering the values upper to 5 as discrepant (see Fig. 6). Also, we identified a small variation in the arrival rate, avg. lambda usage and migrations data, if compared to the other two metrics that have more varied values. The maximum identified in the boxplot for avg. total allocated and the total request is equal. Furthermore, a slight asymmetry was identified that implies in a majority located data in the low side.

Next boxplot (see Fig. 7) demonstrate a box for each continuous features based on one feature selected for grouping. In this case, we selected an energy-based data grouping. Thus, we identified several important observation of the way the interaction of energy and the metrics occur. For example, a direct relationship between the total number of generated requests that increases and directly imposed the energy consumption that grows too. Another observation is that higher values of energy are directly related to higher values of the metrics, wich giving a perception that they are directly associated. Higher energy consumption is also associated with higher arrival rates.

Arrival rate, total requested and avg. total allocated reinforce previous observations about the direct relationship of these

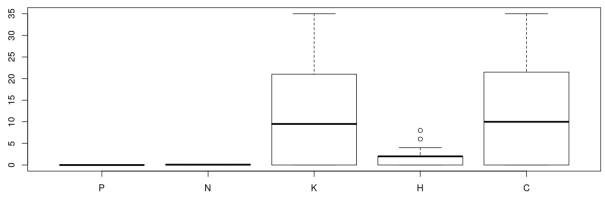

Fig. 6: Data BoxPlot

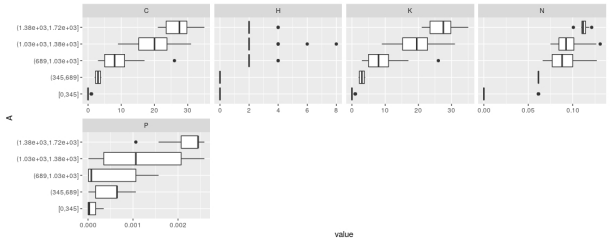

Fig. 7: BoxPlot Grouped by Energy

metrics to the energy consumption.

B. Energy Consumption Prediction with LSTM

We collected, trained and compared data in the test process with those based on [6]. As a first experiment, a regression-based recurrent networks utilizing an LSTM architecture trained using Truncated Back-propagation Through Time (TBPTT), Adam optimization, batch size of 12, MSE loss function and a number of previous time steps in 4 to predict the next time period, corresponding four hours, was implemented to provide first results presented in Fig. 8(a).

For regularizing the results, we use a Dropout rate of 20% for a probabilistic exclusion method in LSTM trains process focusing on the weight update to provide overfitting reduction.

For the experiments, the model obtains an MAE in $95{:}58W$, RMSE $237{:}13W$, maximum error in $531{:}4W$ and a R^2 Score in $0{:}87$, corresponding to 87% of adjustment.

As a second experiment, a stack-based LSTM architecture trained using TBPTT, Adam optimization, batch size of 12, MSE loss function and several previous time steps in 4 to predict the next period, corresponding four hours, was implemented to provide results presented in Fig. 8(b).

As a result in its prediction, we get an MAE in $25{:}34W$, was acquired an RMSE in $98{:}71W$, the maximum error in $165{:}76$ and R^2 Score in $0{:}95$, corresponding to 95% of adjustment.

If comparing the two results presented, it observed a big capacity of energy values prediction even in a dynamic network traffic request environment. The stack-based LSTM presents better results but with a long time of execution even with the Dropout. With this perspective, the benefits of the results are considerable, for example, a perceptual of reduction in the MAE and R^2 Score gain in 8%.

VI. Conclusion

We investigated a two-level CF-RAN architecture data behavior focusing on the prediction of energy consumption in a dynamic scenario. As a result, we presented the CF-RAN scenario and presented six metrics with a strong relationship with energy consumption in a set of many generated. Also, we presented an analysis of these metrics and how they relate to energy wast. Also, we presented a time series forecast problem of energy consumption in a CF-RAN dynamic scenario using two LSTM deep recurrent neural network architecture, we compared these approaches of LSTM architecture and compared its results employing energy prediction.