2020 International Conference in Mathematics, Computer Engineering and Computer Science (ICMCECS) 978-1-7281-3126-9/20/$31.00 ©2020 IEEE 10.1109/ICMCECS47690.2020.240883

Multiplexing Gains Through Clustering In Cloud Radio Access Network

Umasabor Nelson
Department Of Telecommunication Engineering
School Of Electrical Engineering And Technology
Federal University Of Technology
Minna, Nigeria
umasabor.pg813285@st.futminna.edu.ng

Bala Salihu
Department Of Telecommunication Engineering
School Of Electrical Engineering And Technology
Federal University Of Technology
Minna, Nigeria
salbala@futminna.edu.ng

Nathaniel Salawu
Department Of Telecommunication Engineering
School Of Electrical Engineering And Technology
Federal University Of Technology
Minna, Nigeria
salawunathaniel@gmail.com

Suleiman Zubair
Department Of Telecommunication Engineering
School Of Electrical Engineering And Technology
Federal University Of Technology
Minna, Nigeria
zubairman@futminna.edu.ng

ABSTRACT—**There is an increase in the energy consumption and cost of base station deployment due to increase in the number of devices which require a dense deployment of base stations to handle the resultant data traffic. C-RAN (Cloud Radio Access Network) was proposed to remedy the problem of increasing data traffic volume and reduce the capital expenditure and operating expenditure of Mobile Network operators. Cloud Radio Access Network (C-RAN) allows for network resources to be shared amongst several base stations thereby reducing cost. By using different clustering algorithms such as K-means, Hierarchical and Gaussian Mixture Models to cluster these base stations there is a reduction in the needed network resources and this reduces cost. Capacity Utility and cost of deployment are the metrics used in making a comparative analysis of the different clustering algorithms used in this work. From evaluation of the methodology, it showed that the Hierarchical clustering algorithm had a Capacity Utility of 0.0012, Gaussian Mixture Models had 0.0035 and K-means with 0.0044 and when you compare this with the Capacity Utility before clustering of 0.63 it can be seen that the Hierarchical clustering algorithm had reduced the needed network resources better than Gaussian Mixture Models and K-means. The 3 clustering algorithms were also able to reduce the number of needed base stations from 182 to 80, thereby reducing Cost of deployment.**

Keywords-Clustering Algorithm, Statistical Multiplexing Gain, Mobile Network.

I. INTRODUCTION

In recent years, due to increase in personal devices and data-hungry mobile applications, the demand for an efficient and always available high data rate radio access network has increased [1]. The radio access network part of a base station is one of the most important and about 80% of the capital expenditure is spent on it [2]. Moreover operating expenditure like power consumption and maintenance is also increasing [3]. C-RAN is an architecture to solve the problem of increase in capital and operating expenditure [4] . C-RAN was initially proposed by [5] with further details in [2].

In C-RAN there is the centralization of baseband processing of statistically varying traffic loads from different base stations as shown in Fig. 1, these loads are multiplexed and the processing resources are also shared. A base station is provisioned to serve the maximum traffic load but due to variations in traffic load throughout the day, there may be times of underutilization. The peak of this multiplexed load across several base stations is seen to be lower than the sum of individual peak loads [6]. This lowered peak load leads to reduction in processing resources and this leads to the statistical multiplexing gain of baseband processing resources as the bandwidth requirement for each user compensates for each other due to the variability of the data traffic. Multiplexing gain has the following benefits: Reduced cost, Reduced energy consumption, Better spectrum utilization, Better resource utilization, Scalability.

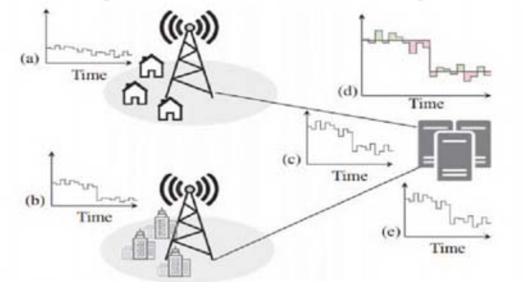

Fig. 1 A C-RAN with 2 RRHs.

Fig. 1 is showing the application of BBU resource multiplexing to short-term (in frames) and long-term (in hours) load fluctuations. (a) and (b) load at each RRH. (c) Aggregate load at the BBU. (d) long term multiplexing. (e) short term multiplexing. [7]

Long-term multiplexing is the capacity to adjust to slow variations and this slow variation is the change in traffic load during the course of a day. This changes occur in the minute to hour scale. Short-term multiplexing is the capacity to adjust to fast variations usually in the millisecond to seconds scale [7].

A. Cloud ran architecture

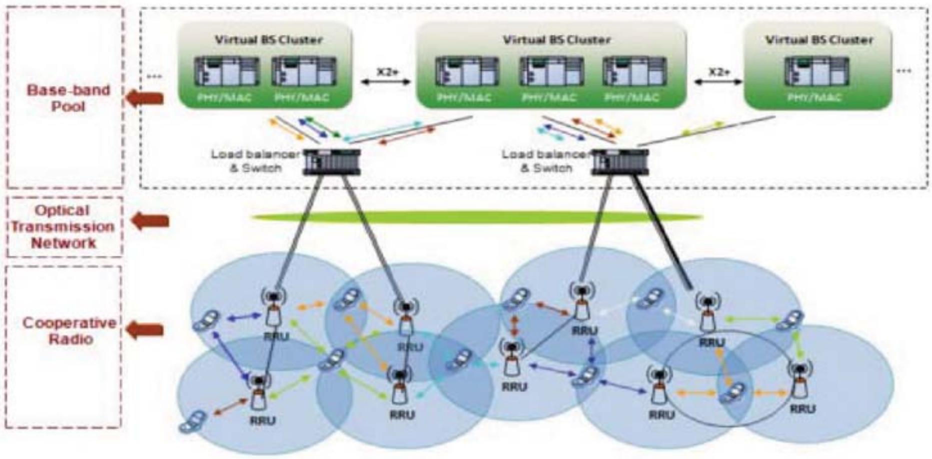

Fig. 2 Cloud RAN Architecture. [2]

C-RAN is a concept of combining the applications of cloud computing on mobile network radio access network. C-RAN was proposed to address the deficiencies of the traditional radio access network such as limited capacity, insufficient expendability, low utilization. [8] C-RAN is a network architecture where there is a separation of the Baseband units (BBUs) and Remote Radio Heads (RRHs) as shown in Fig. 2. The BBUs are responsible for baseband processing such as coding, modulation, Fast Fourier Transform (FFT) and RRHs for radio functionalities such as digital processing frequency filtering and power amplification [4]. Baseband processing are concentrated into the cloud and their functions shared among various base stations in a BBU pool, which allows it to adapt to dynamic traffic patterns and use network resources more efficiently. This allows for the use of fewer BBUs thereby reducing capital and operating expenditure. It also reduces energy consumption; it improves scalability because new BBUs can be added easily. The third part of the C-RAN is the fronthaul transport infrastructure which provides connection between the Remote Radio Heads and Baseband Units. An Optical transport network is used to achieve this, but actualization of this is a very big challenge due to the requirements of this interface such as high bandwidth (tens of Gbps), low jitter, low latency and cost. One major drawback in the achievement of C-RAN is the limited fronthaul capacity. The fronthaul is the connection between the Remote Radio Heads (RRHs) and the Baseband Units (BBUs). The signals sent from the RRHs to the BBUs can be analog or digitized [9]. The digitized or IQ (In-phase/Quadrature-phase) transmission is carried out using the CPRI (Common Public Radio Interface) protocol [10]. For C-Ran to be achieved there is need for a fronthaul connection with high bandwidth and low latency but in reality the available fronthaul is both capacity and time-delay constrained [11]. [9-11] have used different techniques to try to solve this problem.

B. Clustering algorithm

Clustering is a method of grouping data in machine learning. It is used to sort data into groups, data that have attributes that are alike (have similar characteristics) are grouped together and vice versa[14] . Clustering is commonly used for statistical data analysis. Some applications of clustering are grouping of authors or music by genre, customers based on purchases. Base station remote radio heads (RRHs) can also be clustered by similar traffic patterns, and this clustering leads to savings in energy and capital expenditure as it reduces the number of needed mobile infrastructure (Base stations). Mobile network operators over-dimension network resources due to lack of pre-knowledge of the data traffic volume that will be needed during peak usage, base station resources may be underutilized as usage of base station resources vary throughout the day and this can be reduced by clustering of base stations effectively [15]. Clustering in C-RAN is important because it allows for the increase in capacity utility, and this increase is as a result of the BBUs sharing the data computations of several RRHs in different times of the day [16]. Some examples of clustering algorithms are the K-Means clustering algorithm which is a popular algorithm which is used to segment data into k groups[17]. Hierarchical clustering can either be agglomerative which merges objects into clusters and these clusters continue to merge until it forms just a single cluster or divisive which does the opposite [18]. In Gaussian mixture models we identify the clusters by their mean, covariance and size of the cluster. These parameters are identified for each data point and the probability that it belongs to a cluster is estimated [19].

II. PREVIOUS WORKS

Authors in [16] proposed a Cloud IQ framework to cluster a set of base stations based on similar compute platform to meet their real-time processing needs. Traffic logs of 21 cell sites of a Wideband Code Division Multiple Access (WCDMA) network in a dense urban area was used to calculate the processing load and the result of resource pooling across several base stations.

[20] Proposed the use of a multi-dimensional Markov model to investigate the statistical multiplexing gain of a BBU pool, this model captured the changes and restrictions imposed by both the radio and computational resources. This model showed that BBU pools can result in 75% multiplexing gains with 50 BBUs. For the problem of User Association, which is determining which RRH users will connect to, and the problem of RRH clustering, which is determining which RRH are going to be grouped together to achieve statistical multiplexing gain, [21] proposed a solution which was a framework to jointly optimize the User association and RRH clustering to reduce total time-response of the network and power demand of the network. Fronthaul capacity in C-RAN is finite and this becomes a major hurdle due to the large amounts of data as a result of the explosion in mobile devices. [22] Has proposed the use of queueing theory and traffic model to achieve multiplexing gain which is the reduction of the required fronthaul capacity. The multiplexing gain will depend on the difference of the data traffic. Results showed that higher traffic difference led to higher reduction in the required fronthaul capacity.

[23] proposed a deep-learning-based framework using a Multivariate LSTM (Long Short-Term Memory) model to forecast the traffic patterns of Remote Radio Heads' and these forecasted traffic patterns are used to form clusters of similar Remote Radio Heads' into BBUs. A DDCA (Distance-Constrained Complementarity-Aware) algorithm was also proposed to create adequate clustering schemes and this showed an improvement in the network capacity and reduction in deployment cost.

III. METHODOLOGY

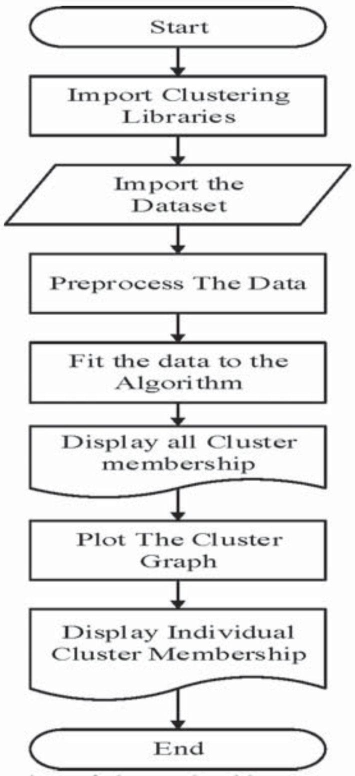

Fig. 3 Flowchart of cluster algorithm.

The flowchart in fig.3 above shows the process in achieving the clustering of RRHs by the similarities of their data. The tool used is the Python programming language with several libraries.

The data used in this paper was extracted from [24]. In the dataset the CellID is used to represent a RRH and Internet traffic as the Data traffic volume.

The first step is to import the clustering libraries like pandas, numpy, sklearn; matplotlib followed by importing the dataset, the dataset is preprocessed to remove missing information to avoid errors. During preprocessing the specific features of the dataset to be clustered is chosen. The chosen features are fit to the clustering algorithm to be used which performs

the clustering on the dataset. After the clustering is done, it outputs all the cluster membership.
For easier analysis, the graph showing all cluster membership is plotted and the cluster membership for each cluster is displayed. For each cluster, the mean of the data traffic is calculated and the number of cells in each cluster is also found out. This is clearly shown in Table I.
The information in Table I is used to evaluate our work, we focused on two metrics: Capacity utility and Cost of deployment which can be calculated using these formulas[23]:

$$Utility(P) = mean_{C_k} U(C_k) \tag{1}$$

$$Cost(P) = \sum_{k=1}^{K} \left| \{C_k\} \right| \tag{2}$$

Where $P = \{C_1 .... Ck\}$ is the clustering scheme.

- $$U(C) = \left(\frac{meanf(C)}{|B|}\right)^{-in\frac{meanf(C)}{|B|}} \tag{3}$$
  is to calculate the capacity utility of a BBU B connected to a Cluster C.
- $meanf(C)$ is the mean aggregated traffic volume of cluster C and $|B|$ is the BBU capacity. The BBU capacity of a cluster used in this work is based on empirical experiments carried out by[23].

We compared our metrics for before and after clustering.

IV. RESULTS

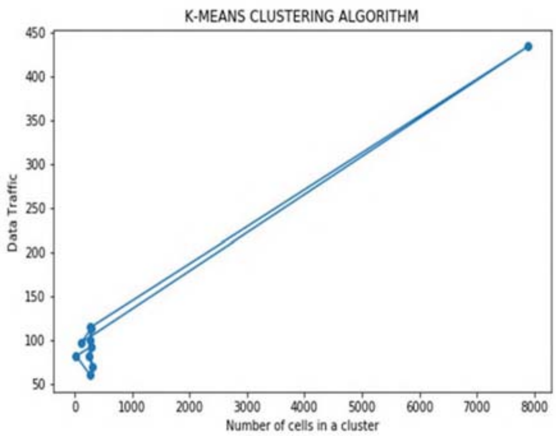

Fig. 4 K-Means Cluster

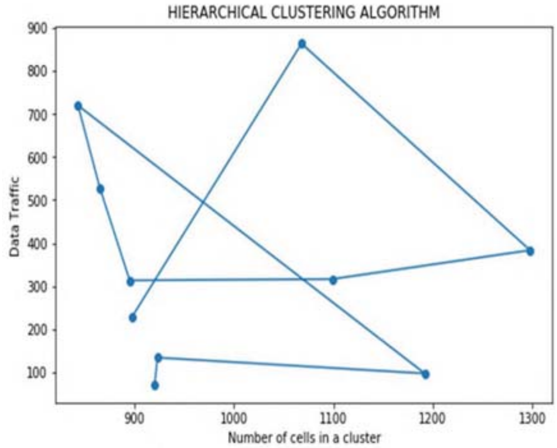

Fig. 5 Hierarchical Cluster

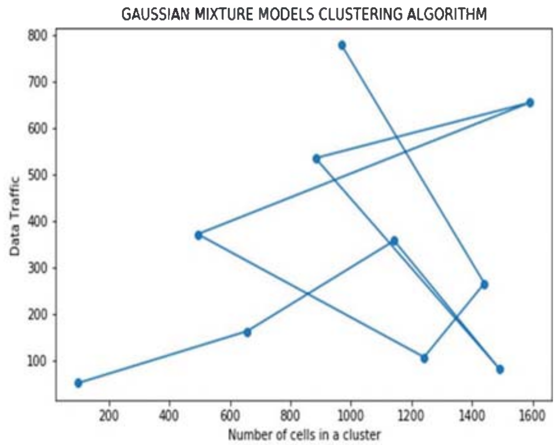

Fig. 6 Gaussian Mixture

Fig. 4, Fig.5, Fig. 6 shows the results for k-means, hierarchical and Gaussian Mixture clustering algorithm respectively and how they were able to effectively cluster cells by their mean Data Traffic

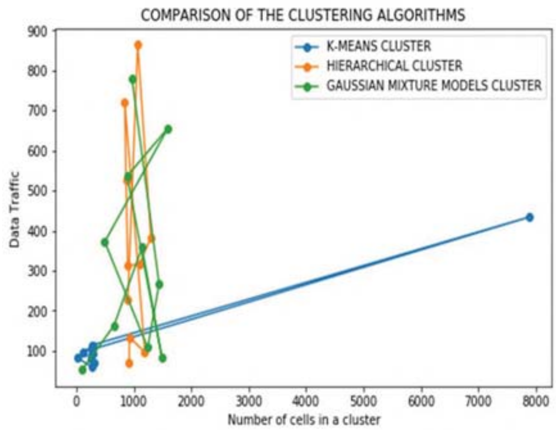

Fig. 7 A Comparison of the Clustering Algorithms.

TABLE I. RESULTS OF CLUSTER ALGORITHMS.

CLUSTER	CLUSTER ALGORITHMS					
	K-MEANS		HIERARCHICAL		GAUSSIAN MIXTURE MODELS	
	NUMBER OF CELLS	MEAN OF DATA TRAFFIC	NUMBER OF CELLS	MEAN OF DATA TRAFFIC	NUMBER OF CELLS	MEAN OF DATA TRAFFIC
0	278	99.82	897	226.82	969	778.88
1	304	69.17	1068	863.92	1441	266.54
2	254	81.40	1298	383.33	1242	107.91
3	276	114.26	1099	316.25	494	372.35
4	7893	434.07	895	313.21	1591	655.26
5	119	96.97	865	528.19	883	535.81
6	293	113.47	843	719.93	1491	82.49
7	294	92.38	1192	96.76	1143	358.17
8	27	81.81	923	133.66	657	163.65
9	272	59.68	920	68.98	99	52.24

TABLE II RESULTS OF CAPACITY UTILITY AND COST OF DEPLOYMENT

CLUSTERING ALGORITHM	CAPACITY UTILITY	COST OF DEPLOYMENT
BEFORE CLUSTERING	0.63	182
K-MEANS	0.0044	80
HIERARCHICAL	0.0012	80
GAUSSIAN MIXTURE MODEL	0.0035	80

Table I shows the results of the cluster algorithms based on the number of cells in a cluster by their mean data Traffic and Table II shows the results for the various clustering algorithms and how they performed based on capacity utility and cost of deployment, 10 clusters was formed for each clustering algorithm and the BBU capacity is 8 for each cluster[23]. From the table it is seen that there was a reduction in capacity utility and cost after clustering was done, it is also noticeable that hierarchical clustering algorithm performed best in comparison to the other used clustering algorithms. The cost of deployment reduced from 182 to 80 which means that the area covered by 182 RRHs can be effectively covered by 80 RRHs.

V. CONCLUSION

As a result of increase in personal devices and data-hungry mobile applications, the demand for an efficient and always available high data rate radio access network has increased. The radio access network part of a base station is one of the most important and about 80% of the capital expenditure is spent on it. Mobile network Operators are looking for ways to cut down this cost. Cloud Radio Access Network (C-RAN) is an architecture to solve the problem of increase in capital and operating expenditure. In C-RAN there is the centralization of baseband processing of statistically varying traffic loads from different base stations, these loads are multiplexed and the processing resources are also shared. These multiplexed loads, lead to considerable gains such as reduction in cost and energy consumption. Some base station resources may be underutilized as usage of base station resources vary throughout the day and this can be reduced by clustering of base stations effectively and exploiting their multiplexing gains. This work presented a way to effectively cluster data using K-means, Hierarchical and Gaussian Mixture Models clustering algorithms to achieve multiplexing gains such as improved capacity utility and reduced cost in C-RAN. A comparative analysis was also done to determine the optimal C-RAN clustering algorithm based on the following metrics: Capacity utility and Cost of deployment. The methodology for the clustering algorithm was implemented using Python programming language and results obtained showed that the hierarchical clustering algorithm outperformed the other clustering algorithms.

ACKNOWLEDGEMENT

We would like to thank the Department of Telecommunication Engineering of the Federal University of Technology, Minna for all the support and constructive criticism towards this work, and also thank the reviewers for their corrective suggestion.

REFERENCES

[1] C. Visual and N. Index, "Cisco Visual Networking Index: Global Mobile Data About the Cisco Visual Networking Index Q. Why did Cisco develop the Cisco Visual Networking Index ™ (Cisco VNI ™) Forecast?," pp. 2017–2022, 2017.

[2] C. Chen, "C-RAN: The road towards green radio access network," *White Pap.*, vol. 0, 2011.

[3] G. Ye Li, C. Xiong, C. Yang, S. Zhang, Y. Chen, and S. Xu, "Energy-Efficient Wireless Communications:

Tutorial, Survey, and Open Issues," *IEEE Wirel. Commun.*, no. December, pp. 28–35, 2011.

[4] H. Lehrmann, M. Stübert, A. Checko, H. L. Christiansen, and Y. Yan, "Cloud RAN for Mobile Networks - a Technology Overview," 2014.

[5] Y. Lin, L. Shao, Z. Zhu, Q. Wang, and R. K. Sabhikhi, "Wireless network cloud: Architecture and system requirements," *IBM J. Res. Dev.*, vol. 54, no. 1, pp. 4:1-4:12, 2010.

[6] I. Rish, T. J. Watson, and Y. Heights, "IBM Research Report," *Science (80-.).*, vol. 22230, no. October, pp. 1–10, 2001.

[7] A. E. Kalor, M. I. Agurto, N. K. Pratas, J. J. Nielsen, and P. Popovski, "Statistical multiplexing of computations in C-RAN with tradeoffs in latency and energy," *2017 IEEE Int. Conf. Commun. Work. ICC Work. 2017*, vol. 1, pp. 772–777, 2017.

[8] M. Chen, Y. Zhang, L. Hu, and T. Taleb, "Cloud-based Wireless Network : Virtualized , Reconfigurable , Smart Wireless Network to Enable 5G Technologies," 2015.

[9] J. Melzer, "Cloud Radio Access Networks," pp. 163–174, 2013.

[10] Common Public Radio Interface, "CPRI Specification V7.0," *Stand. Doc. Specif.*, vol. 0, p. 128, 2015.

[11] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: Insights and challenges," *IEEE Wirel. Commun.*, vol. 22, no. 2, pp. 152–160, 2015.

[12] L. Ramalho *et al.*, "An LPC-based fronthaul compression scheme," *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 318–321, 2017.

[13] J. Lorca and L. Cucala, "Lossless compression technique for the fronthaul of LTE/LTE-advanced cloud-RAN architectures," pp. 1–9, 2013.

[14] P. Fränti and S. Sieranoja, "How much can k-means be improved by using better initialization and repeats?," *Pattern Recognit.*, vol. 93, pp. 95–112, 2019.

[15] A. Checko, H. Holm, and H. Christiansen, "Optimizing small cell deployment by the use of C-RANs," *Eur. Wirel. 2014; 20th Eur. Wirel. Conf. Proc.*, no. c, pp. 1–6, 2014.

[16] S. Bhaumik *et al.*, "CloudIQ : A Framework for Processing Base Stations in a Data Center," pp. 125–136, 2012.

[17] N. Arunkumar *et al.*, "K-Means clustering and neural network for object detecting and identifying abnormality of brain tumor," *Soft Comput.*, 2018.

[18] T. W. Liao, "Clustering of time series data — a survey," vol. 38, pp. 1857–1874, 2005.

[19] R. Sridharan, "Gaussian mixture models and the EM Algorithm (Lecture Notes - MIT CSAIL)," p. 11, 2017.

[20] J. Liu, S. Zhou, J. Gong, Z. Niu, and S. Xu, "Statistical multiplexing gain analysis of heterogeneous virtual base station pools in cloud radio access networks," *IEEE Trans. Wirel. Commun.*, vol. 15, no. 8, pp. 5681–5694, 2016.

[21] H. Taleb, M. El Helou, S. Lahoud, K. Khawam, and S. Martin, "An Efficient Heuristic for Joint User Association and RRH Clustering in Cloud Radio Access Networks," *2018 25th Int. Conf. Telecommun. ICT 2018*, pp. 8–14, 2018.

[22] J. K. Chaudhary, J. Bartelt, and G. Fettweis, "Statistical multiplexing in fronthaul-constrained massive MIMO," *EuCNC 2017 - Eur. Conf. Networks Commun.*, 2017.

[23] L. Chen, D. Yang, D. Zhang, C. Wang, J. Li, and T. M. T. Nguyen, "Deep mobile traffic forecast and complementary base station clustering for C-RAN optimization," *J. Netw. Comput. Appl.*, vol. 121, pp. 59–69, 2018.

[24] G. Barlacchi *et al.*, "A multi-source dataset of urban life in the city of Milan and the Province of Trentino," pp. 1–15, 2015.