

Load-Aware Dynamic RRH Assignment in Cloud Radio Access Networks

Debashisha Mishra, Amogh PC, Arun Ramamurthy, Antony Franklin A, and Bheemarjuna Reddy Tamma
 Department of Computer Science and Engineering, Indian Institute of Technology Hyderabad, India
 Email: [cs15mtech01003, cs15mtech01002, me11b005, antony.franklin, tbr]@iith.ac.in

Abstract—Due to spatio-temporal variation of mobile subscriber's data traffic requirements, traffic load experienced by base stations present at different cell sites exhibit highly dynamic behavior in traditional cellular systems. This non-uniform and dynamic traffic load leads to under utilization of the base station computing resources at cell sites. Cloud Radio Access Network (C-RAN) is an innovative architecture which addresses this issue and keeps the Total Cost of Ownership (TCO) under safe limit for cellular operators. In C-RAN, the baseband processing units (BBUs) are segregated from cell sites and are pooled in a central cloud data center thereby facilitating shared access for a set of Remote Radio Heads (RRHs) present at cell sites. In order to truly exploit the benefits of C-RAN, the BBU pool deployed in the cloud has to efficiently serve clusters of RRHs (*i.e.*, many-to-one mapping between RRHs and BBUs in the BBU pool) and thereby minimizing the required number of active BBUs.

In this work, potential benefits of C-RAN are studied by considering realistic traffic loads of base stations deployed in urban areas by using statistical models. We propose a light-weight and load-aware algorithm, Dynamic RRH Assignment (DRA), which achieves BBU pooling gain close to that of a well known First-Fit Decreasing (FFD) bin packing algorithm. Using extensive simulations, we show that DRA consumes only 25% of time on average compared to FFD for the case of urban cellular deployment of 1000 RRHs. DRA slightly overestimates the required number of active BBUs as compared to FFD by 1.7% and 1.4% for weekdays and weekends, respectively.

Keywords: Cloud-RAN, BBU pool, RRH assignment and Dynamic clustering.

I. INTRODUCTION

Cellular networks are evolving continuously with efficient and robust architectural designs in order to serve the growing need of data traffic from end users. LTE and LTE-Advanced are the latest cellular technologies offering high data rates in a cost efficient way. With rapid increase in the usage of smartphones, tablets and laptops, the data traffic requirement from end users is increasing drastically [1]. Cloud Radio Access Network (C-RAN) is a next generation cellular network architecture which has capability to meet this high demand of data traffic from end users while efficiently addressing deployment and operational challenges faced by cellular operators [2]. In C-RAN architecture, cell sites contain Remote Radio Heads (RRHs) having some part of RF circuitry while Base Band processing Units (BBUs) are pooled in a central location such as data centers using cloud computing and virtualization techniques. BBUs are the digital signal processing components which carry out compute-intensive tasks serving a set of geographically distributed RRHs. The

functional split of traditional LTE base station (a.k.a eNodeB) architecture into RRH and BBU with cloud assisted deployment options can provide considerable savings in capital expenditure (CAPEX) and operational expenditure (OPEX) for cellular operators. From implementation perspective, BBUs are installed on Virtual Machines (VMs) using hypervisors over physical computing cores. VMs dynamically share the physical processing cores present in the cloud data center [3]. The RRH and BBU components exchange baseband signals known as IQ signals over high bandwidth and low latency (front-haul) interface in order to meet the stringent real-time processing requirements of baseband signal processing in C-RAN. In addition, it provides power saving opportunities at cell sites, improves spectral efficiency, and helps in seamless convergence of multiple radio access technologies such as 4G/5G and Wi-Fi [4].

II. MOTIVATION AND RELATED WORK

In traditional cellular RAN architectures, there exists 1:1 static assignment between RRH and BBU, co-located at the cell sites. But this static 1:1 assignment is very expensive for 4G/5G cellular networks where massive number of cells are expected to be deployed for addressing mobile data crunch. In C-RAN, the basic goal is to optimally use BBU pool of cloud data center by dynamically balancing workloads which are generated at cell sites [5]. One BBU is capable of serving a finite set of RRHs (RRH cluster) and mapping between BBU and RRH cluster should be dynamic and expected to alter with varying traffic and processing loads over time. Thus, the design focus should be many-to-one adaptive mapping between RRHs and BBUs with the objective of minimizing the number of active BBUs. Figs. 1 and 2 give an example of flexible RRHs assignments to BBUs which change from one time instant to another time instant.

First Fit Decreasing (FFD) algorithm for optimal bin packing is a very well studied optimization problem in the area of operation research and can be used for efficient clustering of a group of RRHs to a BBU [3]. The decision to include an RRH within a cluster vary over different time instants as per the load experienced from cell sites by BBU. Hence, in order to capture the traffic dynamism at a fine granularity of time, the clustering procedure has to run with a fixed periodicity of shorter time epoch. The RRH loads from cell sites are given as input to FFD in each time epoch to find optimal RRH clusters which are then assigned to BBUs in cloud assisted BBU pool.

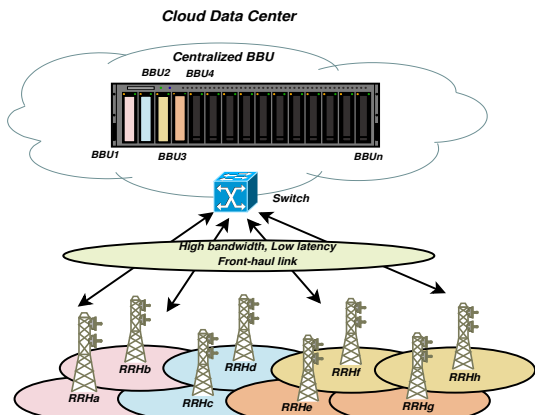


Fig. 1: Assignment of cluster $\{RRH_a, RRH_b\}$ to BBU_1 , $\{RRH_c, RRH_d\}$ to BBU_2 , $\{RRH_f, RRH_h\}$ to BBU_3 and $\{RRH_e, RRH_g\}$ to BBU_4 .

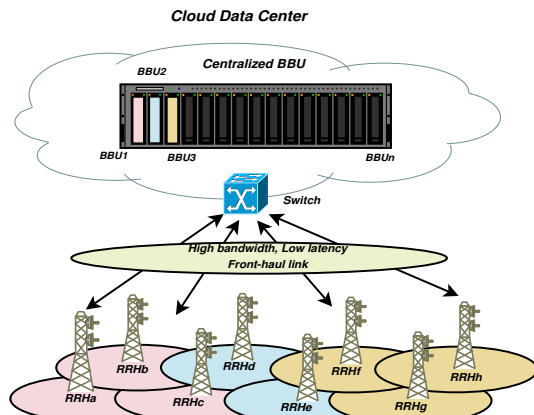


Fig. 2: Assignment of cluster $\{RRH_a, RRH_b, RRH_c\}$ to BBU_1 , $\{RRH_d, RRH_e\}$ to BBU_2 and $\{RRH_f, RRH_g, RRH_h\}$ to BBU_3 .

Authors in [6] proposed a multi-dimensional Markov model for estimating the statistical multiplexing gains of virtualized BBU pool, but it does not address the randomness exhibited by realistic systems. In [7], the authors presented an RRH clustering procedure which is based on multi-objective optimization without energy saving estimations. With the help of OPNET modeler, authors in [8] evaluated energy and cost savings in C-RAN system, but they did not address the uneven traffic conditions. They claim reduction of user data signal processing resources by a factor of 4. In [9], the authors considered a TDD C-RAN system and presented a dynamic RRH-BBU assignment scheme based on the physical resource blocks (PRBs) utilization. It also highlights coordination among base stations (BSs) to mitigate interference issue by jointly serving a user in interference region. For cooperative BS (exploiting CoMP), the authors in [10] considered maximum weight clustering and user scheduling approach to derive dynamic clusters, but their work did not consider uneven user distribution and mobility.

Our main contributions in this paper are given below.

- Towards realization of the vision of having optimal savings from many-to-one mapping of RRHs to BBU in C-RAN, we propose an efficient, light-weight, and load-aware dynamic RRH assignment (DRA) algorithm.
- We then compare performance of proposed DRA algorithm with well known FFD and quantify the potential benefits in terms of computing resource gains achieved by pooling BBU resources in C-RAN.
- Extensive simulation experiments are carried out by varying network setup from small-scale urban area having 200 RRHs to large-scale urban area having 1000 RRHs for both weekday and weekend traffic profiles.

The rest of the paper is organized as follows: We characterize spatio-temporal BS (RRH) traffic load variation at cell sites in Section III. In Section IV, the system model is discussed with minimization criteria of active BBUs as an integer programming optimization problem. The proposed DRA algorithm is given in Section V. Experimental analysis of proposed algorithm is given in Section VI. We finally conclude the paper in Section VII.

III. CHARACTERIZATION OF TRAFFIC LOADS AT RRH

The spatio-temporal variation of distribution of mobile terminals (User Equipments or UEs) with diverse data traffic requirements is an important factor contributing to the network load on cell sites. Based upon a massive, fine grained data set collected from a large-scale cellular operator, Wang et. al. in [11] quantitatively characterized the inhomogeneity in spatio-temporal user distribution and data traffic. The studied cellular traffic followed a trimodal distribution comprising compound exponential, power-law, and exponential distributions. After collecting a huge data set from a real-world 3G cellular system, Nan et. al. in [12] provided statistical models for time varying average throughputs per cell as well as the instantaneous throughput in each cell which is very useful in simulating various time-space dependent data traffic patterns.

Temporal Load Pattern: The hourly aggregate traffic load of a cellular network follows a periodic behavior with relatively high load during daytime and lowest load during midnight as most of the users will be sleeping and their devices are inactive. In the time series, the load on individual base stations do not follow any periodicity, however the trend is consistent with diurnal activity patterns of human beings.

Spatial Load Pattern: The spatial randomness and distribution of traffic loads across different RRHs tend to have greater geographical disparity. The core of a city or urban region shows larger peak and moderate traffic fluctuations than that of the rural areas which follow relatively slower fluctuations. Similarly, the residential zones tend to be active in off-hours (nights, weekends and holidays) while business or office areas are active during daytime in weekdays.

IV. SYSTEM MODEL AND PROBLEM FORMULATION

A. RRH Load Model

For modeling spatio-temporal variation in traffic loads of RRHs, we adopted the modeling functions given in [12], where individual loads of RRHs at a given time instant follow an exponential distribution. The rate parameter λ of this exponential distribution is a time-varying function which is represented as a Gaussian Mixture Model (GMM).

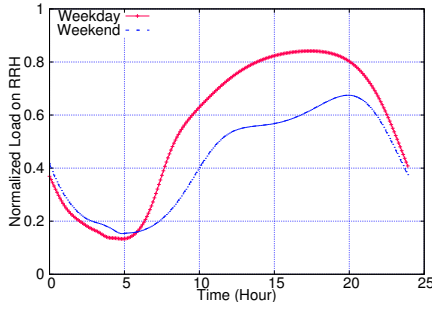


Fig. 3: Time varying mean RRH load.

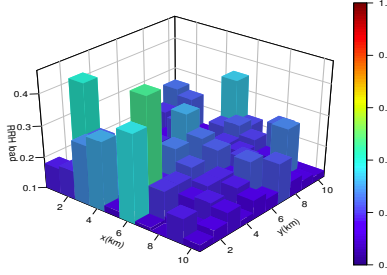


Fig. 4: Weekday spatial plot at 22:00 hour.

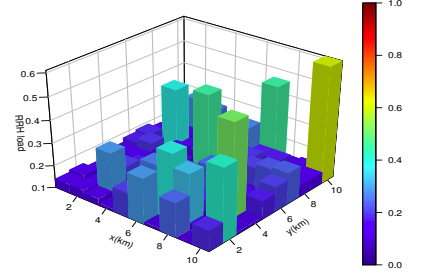


Fig. 5: Weekend spatial plot at 22:00 hour.

The probability density function of an exponential distribution is expressed by

$$f(t) = \lambda e^{-\lambda t}, t \geq 0 \quad (1)$$

where λ is the rate parameter and mean value is $\frac{1}{\lambda}$. The GMM used for modeling of time-varying rate parameter is given by

$$\lambda = \sum_{i=1}^n a_i e^{-\left(\frac{t-b_i}{c_i}\right)^2} \quad (2)$$

where a_i is the amplitude, b_i is centroid location, c_i is the peak width, n represents number of peaks in data series and t is any time instant in 24 hours of the day. Using these two models, we generated snapshots of spatial loads on each of RRHs of whole network under study for a given time instant. Let l_i be the load on i^{th} RRH, then, value of l_i is an exponential random variable with mean value of $\frac{1}{\lambda}$. Fig. 3 shows normalized mean values of traffic load considering loads from all the spatially distributed RRHs under study for 24 hour interval. As seen from the figure, load varies differently for weekdays and weekends. Figs. 4 and 5 show 3D spatial distribution plots of normalized RRH load (e.g., range of [0,1]) in a 10 km \times 10 km grid layout of urban region with 100 RRHs at 22:00 hour of the day for both weekday and weekend, respectively with more peaks on weekend than weekday.

B. BBU Load Consideration

The processing load incurred by a UE at a BBU in the BBU pool in a given Transmission Time Interval (TTI) depends on the number of Resource Blocks (RBs) allocated to that UE and their corresponding Modulation and Coding Scheme (MCS) values [3]. Hence, the overall processing load of an RRH on BBU in a given TTI is the summation of processing loads of all the UEs connected to that RRH in that TTI. We define theoretical peak processing load of RRH, assigned to one of BBUs in the BBU pool, by considering a single UE which is allocated all of the RBs with highest MCS value in one TTI. Based on specifications of VMs (e.g., number of processor cores, their clock speed, memory capacity, bandwidth, etc) hosting a BBU pool in the cloud, a single BBU may be capable of processing peak processing loads of one or more RRHs. Since, most of the times, traffic load at RRHs is not at its peak, we could assign even more number of RRHs to a given BBU

in the BBU pool. Note that BBU peak capacity is different from that of RRH peak load. Cellular operators, by applying efficient dynamic RRH assignment schemes, could scale down or scale up number of active BBUs required in the BBU pool.

C. Energy Model

Depending upon the load generated at cell sites, we adopt an energy consumption model given in [13] to estimate energy savings and the results are presented later in Section VI. The total power consumed at any active BBU at time instant t is

$$P_{BBU_t} = P_{BB} + \sum_{i=1}^n P_{RRH_i} \quad (3)$$

where P_{BB} is power consumed by that particular BBU and $\sum_{i=1}^n P_{RRH_i}$ is sum of the power consumed by all the RRHs associated with that particular BBU at time t . The power consumed by a specified RRH is given by

$$P_{RRH} = \left(\frac{P_r}{E_{pa}}\right) + (P_{rf} \times N_{tx}) \quad (4)$$

where P_r is the radiated power, E_{pa} is the power amplifier efficiency, P_{rf} is the power used by RF circuits and N_{tx} is the number of transceiver antennas.

D. Cluster Model

The RRH clustering and assignment problem can be formulated as a generalization of classical bin packing optimization problem where the BBUs imitate the bins and RRHs form the item set. Its solution is NP-hard [14] and various heuristics and approximation procedures are applied to find a near optimal solution. FFD is one of the most efficient heuristics where the number of used bins are bounded by $\left(\frac{11}{9} \times OPT + \frac{6}{9}\right)$ with OPT being the optimal number of bins [15]. The objectives of the RRH clustering procedure are mentioned below.

- Maximize the resource utilization of each BBU by mapping as many RRHs as possible to each BBU
- Minimize total number of BBUs to be switched-ON (utilized) for processing user requests coming from RRH clusters in the network
- Identify the subset of RRHs (termed as RRH cluster) to be assigned to each active BBU in the BBU pool

Table I presents list of variables used in problem formulation of RRH allocation to BBUs as an integer programming model.

TABLE I: Glossary

Notation	Definition
N	Set of all RRHs
M	Set of all BBUs in BBU Pool
l_i	Processing load incurred by RRH i
z_m	1 if BBU m is active; otherwise 0
y_{im}	1 if RRH i is associated with BBU m ; otherwise 0
l_{max}	Peak BBU capacity

Optimization Model:

Objective Function: Minimize $\sum_{m=1}^{|M|} z_m$.

Constraints:

$$\sum_{m=1}^{|M|} y_{im} = 1, \quad \forall i \in N \quad (5)$$

$$\sum_{i=1}^{|N|} y_{im} \times l_i \leq l_{max} \times z_m, \quad \forall m \in M \quad (6)$$

The constraint in Eqn. 5 ensures the fact that each RRH is associated with exactly one BBU, and Eqn. 6 ensures that the sum of loads from RRHs associated to a BBU does not exceed the BBU peak capacity. Using the RRH load model given in Section III, we propose a novel heuristic method, DRA in Section V and highlight its benefits in comparison with FFD.

V. PROPOSED DRA ALGORITHM

In this work, we propose a light-weight, load aware, dynamic RRH assignment (DRA) algorithm which performs very close to FFD in terms of BBU savings. The basic principle of this algorithm is to offload one or more RRHs (known as *Candidate_RRH(s)*) from an overloaded BBU to a less loaded BBU with enough available computation capacity to accommodate the incoming RRH(s). As one BBU serves a cluster of RRHs, this change in RRH assignment among BBUs results in formation of new cluster from that of previous clusters. In Section II, we already discussed the demerits of 1:1 static mapping between RRHs and BBUs, thus shifting our attention to exploit the resource savings in many-to-one dynamic mapping between RRHs and BBUs.

Assume δt is the periodicity of cluster formation. For capturing traffic dynamism when $\delta t \rightarrow 0$, the many-to-one RRH clustering algorithm must be applied to RRH loads at each interval δt . FFD is one of the baseline and very efficient many-to-one clustering algorithms, but the running time is of higher order of magnitude for a large number of RRHs making it infeasible for smaller value of δt . In other words, clustering with FFD leverages frequent re-clustering to obtain the BBU resource gain. Our proposed approach eliminates the unnecessary re-clustering by checking the offered load to BBU and dynamically assigns only eligible RRH(s) from overloaded BBU to less loaded BBU. From implementation perspective, each cluster is a finite group of RRHs and set of all such clusters form a partition of the RRH set

N . Let us denote partition P_{mod} as the partition containing clusters which need reassignment of RRHs due to change in RRH load. *spill_load* for a given cluster is the excess spill amount which is the difference between the sum of processing loads for constituent RRHs in that cluster and maximum load capacity of cluster. Mathematically, assuming there are r number of RRHs associated with a given cluster, $spill_load = ((\sum_{i=1}^r l_i) - l_{max})$ where l_i is the load on i^{th} associated RRH and l_{max} is the maximum load a BBU can handle. For a given cluster, if $spill_load > 0$, it is known as *spill_cluster* and $spill_load \leq 0$ is called *non_spill_cluster*. *Candidate_RRH* is an associated RRH in a *spill_cluster* whose removal will convert it to a *non_spill_cluster*. We describe an *Offload_Selection* subroutine in the following subsection which takes care of selecting eligible *Candidate_RRH*. Algorithm 1 summarizes the proposed DRA scheme.

Algorithm 1 : DRA Algorithm

- 1: Based on *spill_load*, classify the clusters in P_{mod} as *spill_cluster* and *non_spill_cluster*
 - 2: In case, no *spill_cluster* is found, just return P_{mod}
 - 3: In case, all clusters are *spill_cluster*, perform FFD on processing loads contained in P_{mod} into BBU. This situation is too unrealistic to occur in practice when $\delta t \rightarrow 0$
 - 4: For each *spill_cluster*, find *Candidate_RRH(s)* to offload using *Offload_Selection* subroutine
 - 5: Find a *non_spill_cluster* with enough resources to accommodate *Candidate_RRH*, perform RRH assignment to it
 - 6: If no such *non_spill_cluster*, get a new BBU from BBU pool and perform the RRH assignment to the new BBU
 - 7: Apply merge procedure on *non_spill_clusters* using FFD and return new partition P_{new}
-

Numerical Illustration: Table II shows three different clusters at time t with all RRH loads normalized in the range of $[0, 1]$. Let us denote this as partition P of RRH set N . The total load experience by each of the clusters is less than the peak capacity of BBU. At time $(t + \delta t)$, assume some of the associated RRH loads varied due to traffic inhomogeneity shown as clusters of partition P_{mod} .

TABLE II: Numerical Illustration

Partition P at time t
$C_1 = \{0.71, 0.14, 0.08\}, \sum C_1 = 0.93$
$C_2 = \{0.56, 0.19\}, \sum C_2 = 0.75$
$C_3 = \{0.47, 0.25, 0.11\}, \sum C_3 = 0.83$
$C_4 = \{0.24, 0.48\}, \sum C_4 = 0.72$
Partition P_{mod} at time $(t + \delta t)$
$C_{1mod} = \{0.56, 0.14, 0.40\}, \sum C_{1mod} = 1.10$ (<i>spill_cluster</i>)
$C_{2mod} = \{0.31, 0.48\}, \sum C_{2mod} = 0.79$ (<i>non_spill_cluster</i>)
$C_{3mod} = \{0.21, 0.39, 0.80\}, \sum C_{3mod} = 1.40$ (<i>spill_cluster</i>)
$C_{4mod} = \{0.11, 0.24\}, \sum C_{4mod} = 0.35$ (<i>non_spill_cluster</i>)
Partition P_{new} at time $(t + \delta t)$ after DRA algorithm
$C_{1new} = \{0.56, 0.40\}, \sum C_{1new} = 0.96$
$C_{2new} = \{0.31, 0.48, 0.14\}, \sum C_{2new} = 0.93$
$C_{3new} = \{0.21, 0.39\}, \sum C_{3new} = 0.60$
$C_{4new} = \{0.11, 0.24\}, \sum C_{4new} = 0.35$
$C_{5new} = \{0.80\}, \sum C_{5new} = 0.80$
After FFD, $C_{3,4new} = \{0.21, 0.39, 0.11, 0.24\}, \sum C_{3,4new} = 0.95$

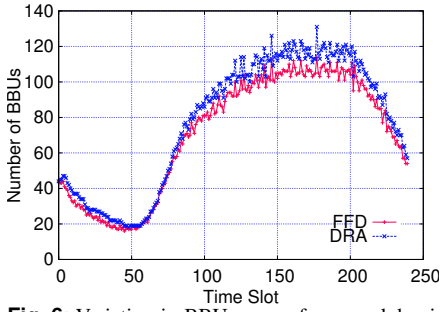


Fig. 6: Variation in BBUs usage for a weekday in a C-RAN with 1000 RRHs.

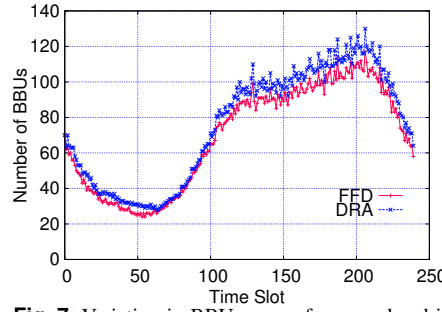


Fig. 7: Variation in BBUs usage for a weekend in a C-RAN with 1000 RRHs.

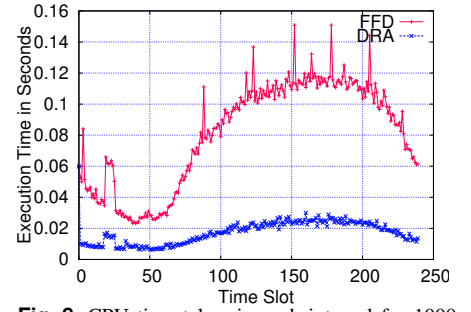


Fig. 8: CPU time taken in each interval for 1000 RRHs.

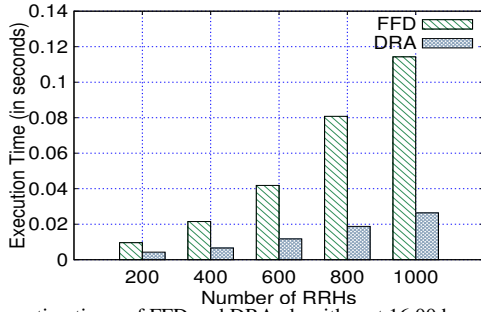


Fig. 9: Execution times of FFD and DRA algorithm at 16.00 hour on a weekday w.r.t varying number of RRHs.

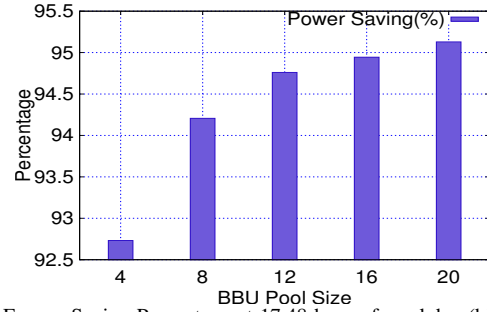


Fig. 10: Energy Saving Percentage at 17.48 hour of weekday (high load) with varying BBU pool size

DRA procedure takes first spill_cluster C_{1mod} and calculates the $spill_load$ which is $(1.10 - 1) = 0.10$. For spill amount of 0.10, DRA finds suitable *Candidate_RRH* to be offloaded from C_{1mod} by *Offload_Selection* subroutine. In this case, RRH with normalized load 0.14 is the *Candidate_RRH*, which needs re-assignment to another non_spill_cluster. Cluster C_{2mod} is suitable non_spill_cluster capable of accommodating additional normalized load of $(1 - 0.79) = 0.21$ without reaching peak capacity of BBU. Hence, RRH with normalized load 0.14 can be re-assigned to C_{2mod} . After this assignment both C_{1mod} and C_{2mod} are non_spill_clusters. For spill_cluster C_{3mod} , $spill_load$ is 0.40 and *Candidate_RRH* is 0.80. As no non_spill_cluster can accommodate this load, DRA switches on a fresh BBU C_{5new} . This process repeats for all the spill_clusters until each of them converts to non_spill_clusters. The resulting clusters after this process may not fully utilize all processing capacity of BBUs because RRHs are locally re-assigned. The DRA scheme applies FFD (Step 7 of DRA algorithm) on all resulting non_spill_clusters to ensure minimum number of clusters (hence minimum number of BBUs), each with higher possible utilization. In this illustration, C_{3new} and C_{4new} are not optimal clusters, because their sum is below 1. Hence, FFD packing outputs a single cluster $C_{3,4new}$ with total assigned load of 0.95. The final partition of RRH set contains 4 clusters as shown in Table II and are given by $C_{1new}, C_{2new}, C_{3,4new}$ and C_{5new} .

Subroutine - Offload_Selection: The decision of removing a *Candidate_RRH* from a spilled cluster is essential for stability of the system and convergence towards optimal clustering. DRA algorithm ensures this characteristic by *Offload_Selection* subroutine which selects an offload candidate. There are two possibilities of selecting an offload candidate

based on the value of $spill_load$ in spill_cluster S . Let $max(S)$ denotes the maximum RRH load value present in cluster S . If $spill_load \leq max(S)$, choose RRH in S whose processing load is just greater than or equal to $spill_load$. Otherwise, sort the associated RRH loads in non-increasing order and remove RRHs sequentially (the highest load first) one after another until the cluster S becomes non_spill_cluster.

Asymptotic Analysis: The asymptotic analysis of time complexity for both FFD and proposed DRA scheme are summarized in Table III. For FFD best case, all RRH loads have to be sorted in non-increasing order, thus taking $O(|N|log|N|)$ time. In FFD worst case, due to high RRH load, each RRH can consume one BBU with $O(|N|^2)$ number of comparisons. In average case, there is at most $O(|N|^2)$ comparisons in FFD.

In best case, DRA does not need any reassignment, but it has to scan every RRH load to ensure that no spill_cluster exists. Thus, it's running time is $O(|N|)$. In worst case, all the spill_clusters may demand switching on a new BBU resulting total N clusters. Hence, step 7 of the DRA scheme may apply FFD to N clusters taking $O(|N|^2)$ comparisons. In average case, suppose DRA forms k clusters out of N RRHs, then on average $(\frac{k}{2})$ clusters are of spill type and $(\frac{k}{2})$ are of non_spill type. Hence, the running time for RRH re-assignment encompass $O(|k|^2)$ for deriving final cluster along with $O(|N|)$ RRH comparisons.

TABLE III: Asymptotic complexity analysis.

Test Input	FFD	DRA	Scenario
Best Case	$O(N log N)$	$O(N)$	No spill_cluster
Average Case	$O(N ^2)$	$O(N + k ^2)$	spill_clusters and non_spill_clusters
Worst Case	$O(N ^2)$	$O(N ^2)$	All spill_clusters

VI. EXPERIMENTAL SETUP AND PERFORMANCE RESULTS

Our simulation setup considers downlink transmission processing of C-RAN, with weekday and weekend traffic profiles spanning 24 hours with time domain sampling of 10 samples per hour. Thus, in 24 hours of a day, we collected a total of 240 samples where sampling periodicity $\delta t = 6$ minutes. In spatial domain, we evaluated a large scale urban network with 200 to 1000 RRHs. Each independent run of simulation is associated with 10 seeds and results are averaged. The simulations are carried out on a commodity hardware having Ubuntu 64-bit Linux distribution on x86, Intel 4 core, 1.7 GHz processor. Other simulation parameters are summarized in Table IV.

We compared the system performance in terms of the computational resource gain (number of active BBUs) obtained from RRH clustering along with the power consumption trend for FFD and proposed DRA scheme. In Section III, we highlighted the exponential distribution of RRH load in spatial domain. Using the same distribution model, individual RRH loads are generated which vary over space and time dimensions. We obtain the processing load on BBU for a given RRH by assuming that the load is distributed equally across all the UEs connected to this RRH. Also, we assume that the processing capacity of the BBUs is equal to the processing load of the peak RRH load. To quantify gains, these loads are given as inputs to the proposed DRA and baseline FFD schemes.

Figs. 6 and 7 show the BBU gains for weekday and weekend, respectively for a large scale urban cellular network setup containing 1000 RRHs where BBU gain is the savings in number of BBUs needed compared to 1:1 and baseline FFD schemes. On average, compared with 1:1 RRH to BBU mapping scheme, DRA reduces the required number of BBUs by 87%. It is observed that DRA over estimates FFD by 1.7% and 1.4% of the total number of BBUs on weekday and weekend, respectively. We then compared the total running time of the algorithm for 1000 RRHs, which is shown in Fig. 8. The comparison of execution time of FFD and DRA in seconds at 16.00 hours of the day by varying the number of RRHs between 200 and 1000 is presented in Fig. 9. With increase in number of RRHs, C-RAN needs to deploy more number of BBUs for serving them and hence there is an expected increase in execution time of algorithms. Fig. 10 shows that more than 90% energy savings for C-RAN could be realized with varying size of BBU pool (4 to 20 BBUs) using DRA scheme compared to conventional RAN. Larger pool size offers more energy saving opportunities due to shared infrastructure maintenance and cooling.

VII. CONCLUSIONS AND FUTURE WORK

In this paper, we analyzed and quantified the BBU resource savings and time complexity measures of DRA in contrast to FFD considering spatio-temporal traffic variations from base stations. DRA is capable of reducing the time complexity of clustering procedure as compared to FFD in an urban network setup and offers quite close BBU resource savings. The savings trend follows a diurnal human traffic pattern which validates

TABLE IV: Simulation Parameters

Parameter	Value
Number of RRHs	200 to 1000
Sampling periodicity	6 minutes
Traffic duration	24 Hours
Total number of samples	240
Traffic profile	Weekday, Weekend
Geographical region	Urban
RRH load range	[0,1] (0 to 100%)
Maximum Load on BBU	1 (100%)
Spatial load distribution	Exponential
Time-varying rate parameter	Gaussian Mixture Model

our proposed scheme. As part of ongoing work, we aim to define various dependent factors such as UE position, cell edge constraints, BS cooperation (e.g., CoMP) in the processing load characterization of RRH for quantifying savings.

ACKNOWLEDGMENT

This work is supported by the project "Converged Cloud Communication Technologies", Deity, Govt. of India.

REFERENCES

- [1] Cisco, "Visual networking index, global mobile data traffic forecast update," *White Paper*, 2014.
- [2] "C-RAN - The road towards green RAN," *China Mobile White Paper*, vol. 2, 2011.
- [3] S. Bhaumik, S. P. Chandrabose, M. K. Jataprolu, G. Kumar, A. Muralidhar, P. Polakos, V. Srinivasan, and T. Woo, "CloudIQ: a framework for processing base stations in a data center," in *Proceedings of MOBICOM*, pp. 125–136, ACM, 2012.
- [4] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for mobile networks - a technology overview," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 1, pp. 405–426, 2014.
- [5] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das, "Understanding traffic dynamics in cellular data networks," in *Proceedings IEEE INFOCOM*, pp. 882–890, 2011.
- [6] J. Liu, S. Zhou, J. Gong, Z. Niu, and S. Xu, "On the statistical multiplexing gain of virtual base station pools," in *IEEE GLOBECOM*, pp. 2283–2288, 2014.
- [7] X. Chen, N. Li, J. Wang, C. Xing, L. Sun, and M. Lei, "A dynamic clustering algorithm design for C-RAN based on multi-objective optimization theory," in *IEEE VTC Spring*, pp. 1–5, 2014.
- [8] A. Checko, H. L. Christiansen, and M. S. Berger, "Evaluation of energy and cost savings in mobile Cloud RAN," *OPNETWORK*, 2013.
- [9] D. Zhu and M. Lei, "Traffic and interference-aware dynamic BBU-RRU mapping in C-RAN TDD with cross-subframe coordinated scheduling/beamforming," in *IEEE ICC Workshops*, pp. 884–889, 2013.
- [10] Y. Du and G. de Veciana, "Wireless networks without edges: Dynamic radio resource clustering and user scheduling," in *Proceedings IEEE INFOCOM*, pp. 1321–1329, 2014.
- [11] H. Wang, J. Ding, Y. Li, P. Hui, J. Yuan, and D. Jin, "Characterizing the spatio-temporal inhomogeneity of mobile traffic in large-scale cellular data networks," *HOTPOST '15*, pp. 19–24, ACM, 2015.
- [12] E. Nan, X. Chu, W. Guo, and J. Zhang, "User data traffic analysis for 3G cellular networks," in *CHINACOM*, pp. 468–472, IEEE, 2013.
- [13] M. Khan, R. Alhumaima, and H. Al-Raweshidy, "Reducing energy consumption by dynamic resource allocation in c-ran," in *EuCNC*, pp. 169–174, IEEE, 2015.
- [14] E. Falkenauer and A. Delchambre, "A genetic algorithm for bin packing and line balancing," in *International Conference on Robotics and Automation*, pp. 1186–1192, IEEE, 1992.
- [15] G. Dósa, "The tight bound of first fit decreasing bin-packing algorithm is $FFD(I) \leq 11/9OPT(I) + 6/9$," in *Combinatorics, Algorithms, Probabilistic and Experimental Methodologies*, Springer, 2007.